



HAL
open science

Élaboration d'un modèle pour une base de textes indexée pédagogiquement pour l'enseignement des langues

Mathieu Loiseau

► **To cite this version:**

Mathieu Loiseau. Élaboration d'un modèle pour une base de textes indexée pédagogiquement pour l'enseignement des langues. Linguistique. Université Stendhal - Grenoble III, 2009. Français. NNT : . tel-00440460v3

HAL Id: tel-00440460

<https://theses.hal.science/tel-00440460v3>

Submitted on 6 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Élaboration d'un modèle pour une base de textes indexée pédagogiquement pour l'enseignement des langues

Volume 1

THÈSE

présentée et soutenue publiquement le 9 décembre 2009

pour l'obtention du

Doctorat de l'Université Stendhal Grenoble 3

(Spécialité Industries de la langue)

par

Mathieu Loiseau

Composition du jury

<i>Directeurs :</i>	Georges ANTONIADIS	LIDILEM (Université Stendhal Grenoble 3)
	Claude PONTON	LIDILEM (Université Stendhal Grenoble 3)
<i>Rapporteurs :</i>	Éric BRUILLARD	UMR STEF (ENS Cachan) – INRP
	Natalie KÜBLER	CLILLAC (Université Paris Diderot - Paris 7)
<i>Examineur :</i>	Cédrick FAIRON	CENTAL (Université Catholique de Louvain)

Mis en page avec la classe thloria.

Remerciements

Je tiens tout d'abord à remercier mes directeurs de thèses, Georges Antoniadis et Claude Ponton, pour leur patience, leurs conseils pertinents (comme par exemple la métaphore disco de Georges ou les interventions précises de Claude sur les fonctions), leur aide à la formalisation du modèle, la relecture intensive de ces dernières semaines, la gestion des articles, confs et cours invités (en Grèce et en Colombie notamment), etc.

Je voudrais ensuite remercier Natalie Kübler, Éric Bruillard et Cédrick Fairon, d'avoir accepté d'évaluer mon travail et de faire partie de mon jury.

J'aimerais également adresser de chaleureux remerciements (sans ordre particulier) :

- aux enseignants qui ont rempli mes fastidieux questionnaires, en particulier Fanny Meunier qui m'en a fait parvenir plusieurs, mais aussi les amis qui bien que pas nécessairement passionnés par mes recherches se sont ~~fait ch-~~ enquiquinés (t'as vu Roseline?) à tout remplir : Nat', Charlotte, Mag', Claude, Florent et j'en oublie ;
- à mes collègues du DIP et du LIDILEM pour leur convivialité : Roseline, Zohra, Ramona, Francis, Agnès, Jack, Saskia, Sophie, Marie-Christine, Josée et tous les autres.
- à tous mes relecteurs et consultants en consulting bibliographico-scientifique, sans ki ma tête, ressemblerait a sa. Alexia, Anna, Audrey, Aurélie, Benabri, Bubu, Cristelle, Françoise, Julien, Lucy, Marinette, Nico, Ninie, Olivier et Thomas : *on* est incontestablement reconnaissant, forcément, devant le temps que vous avez passé à discuter les problèmes de ma thèse, relire des phrases écrites en 4 fois sans qu'une seule virgule ne soit placée au bon endroit, passé en revue ma biblio, cliqué des dizaines de liens, corrigé dans le texte certaines erreurs, repris mes traductions, etc. etc. et même etc.
- à tous les sales pirates qui s'évertuent à détruire toute forme de création, et en particulier les pirates de la science qui – dénués du plus élémentaire respect ~~du travail de leurs pairs~~ des intérêts financiers de Thomson-Reuters, Elsevier, SAGE, Jstore, ACM, Springer et autres « diffuseurs de la connaissance scientifique » – m'ont laissé usurper leur identité sur des comptes dont les abonnements valent des sommes pharaoniques, ont photocopié des articles ou mis à disposition des ouvrages entiers.
- à tous ceux qui m'ont nourri pendant la phase terminale de ma rédaction : Laurence, ses boccolinos, ses cannelés et ses pizziches ; les dodus, leur foyer accueillant et leur table gastronomique ; Benjamin, ses goûts irréprochables dès qu'il s'agit de choisir un resto ou un goûter, mais aussi Noëlle, sa mère, pour ses compotes ; mes parents chez qui il me suffisait de passer un week end pour être dispensé pendant 3 jours de perdre mon temps à me restaurer ; ma petite soeur Charlotte qui s'est appliquée, à grands coups de paquets de bonbecs, à ce que mon séant puisse continuer à en faire office ; Fanny et Pantxo pour leurs pâtisseries à la rhubarbe ; et aussi Aurélie (et ses burritos), Myriam (et ses chips), Agnès (et ses madeleines), Aïcha (et ses gratins de légumes), Liséjérôme et Sebécарole (pour les soirées et pique-niques où je venais avec rien) et tous ceux qui tolèrent ma fromagephobie ;
- à mes sponsors : mes parents qui continuent à m'accorder des crédits pour du matériel informatique ; Thomas qui m'a prêté un ordinateur dans les 24h qui ont suivi le vol du mien à un moment critique ; Audrey qui m'a prêté de quoi écouter de la musique en ces circonstances ; et Agnès grâce à qui j'ai pu transporter tout ce matériel ; Cristelle qui m'a offert une magnifique imprimante.
- à ceux qui se sont penchés sur « ma propension à être attiré par le désert de Gobie mélodique » [Salhi, 2007] et ont ainsi égayé mes heures de rédaction : Mehdi (John Coltrane), Ben (Bert

Jansch, Davy Graham, Tinariwen), Cyril (Made to measure), Laurence (Björk), Pantxo (the Roots), Léon (Bloc Party), Wilfried et Audrey (Lcd Soundsystem, Tiga).

- à mes multiples conseillers de développement personnel : Ben, Audrey, Wil, Roubi, les vioques du DIP ;-) et en particulier Aurélie et Bubu qui m'ont un peu porté à bout de bras pendant certaines crises de doute.
- à mes camarades de vacances, qui m'ont sorti la tête de de l'écran : les dodus et benabri, pour nos diverses virées et surtout la prochaine que je louperai pas ; Pierre y Virginia, los helados de dulce de leche, los corderos patagónicos, los asados et leur goût de reviens-y ; Mehdi et Ben, les surfers couche-tôt de Kenitra ; Mimi et Bubu, les 2 meilleures copilotes qui aient arpenté les bitumes transylvains, évité leurs nappes et leurs bergers lubriques ; la fine équipe bordelaise (Agnès, Aïcha, Laurence, Lucy, Mourad, Myriam, Nico et Virginie), qui n'a pas insisté pour écourter nos périodes de bossage alors que l'environnement se prêtait bien au zonage en dégustant du caviar d'aubergine ; Charlotte, malgré son accent anglais un peu trop natif à mon goût ;
- aux kOpin-e-s de CEDILL, aux buveurs de poucs, aux buveurs de thé, mitonneurs de gâteaux, cobus, amis que je vois pas beaucoup (même par téléphone) : Nico, Pantxo, Thomas, Agnès, Anna, Vannina, Virginia, Alexia, Aïcha, Sandra, Marion, Auriane, Aurélie, Fanny, Ninie, Mimi, Mick, Cricri, Cristelle, Cyril, Lucy, Luc, Bubu, Dudu, Marie-Laure, Claude et d'autres encore.
- à ceux qui m'ont supporté presque quotidiennement (ivl ou irl) pendant toute ma thèse : Audrey, Wilfried, Maya, Benjamin, Myriam, Aurélie et Laurence.
- à ma famille, à qui j'ai peu rendu visite ces derniers temps : Sophie, Charlotte, Jeannette, Philippe et Nicole.

Toutes mes excuses pour tous les noms et toutes les raisons qui m'ont échappé à l'heure indue à laquelle j'écris ces remerciements. Pour finir, je voudrais exprimer ma gratitude à tous ceux qui ont subi mon irascibilité des mauvais jours et qui malgré tout continuent de m'adresser la parole (à peu près tous les noms cités ci-dessus et bien d'autres encore).

Table des matières

Conventions	1
Partie I Position du problème et étude de l'existant	3
Chapitre 1 Indexation pédagogique	5
1.1 Genèse du projet	5
1.1.1 Le projet MIRTO	5
1.1.2 Un projet autonome	6
1.2 Définition du projet : cas d'utilisation	7
1.2.1 Ajout d'un texte	8
1.2.2 Recherche d'un texte	9
1.2.3 Aide au choix d'un texte	9
1.3 Indexation pédagogique	9
1.3.1 Indexation d'objets	10
1.3.2 Langage documentaire	12
1.3.3 « Pédagogique »	15
1.4 Les textes dans l'enseignement des langues	18
1.4.1 L'approche communicative	18
1.4.2 <i>Task-based approach</i>	19
1.4.3 Autres approches et méthodes	20
1.4.4 Le texte omniprésent	23

1.4.5	Positionnement didactique	24
1.5	Organisation de notre travail	26
Chapitre 2 Instrument, corpus et ALAO		29
2.1	ALAO, perspective historique	30
2.1.1	Ordinateur et apprentissage	30
2.1.2	Définition de l'ALAO	35
2.1.3	L'ALAO <i>structural</i>	38
2.1.4	L'ALAO <i>communicatif</i>	40
2.1.5	L'ALAO <i>intégratif</i>	43
2.1.6	Les programmes de l'ALAO intégratif	44
2.1.7	Conclusion	46
2.2	Travers de l'ALAO	47
2.2.1	Conception d'une base de textes	47
2.2.2	Conception	48
2.2.3	Modélisation	53
2.2.4	Perception et usage	58
2.2.5	Conclusion	63
2.3	Corpus et enseignement des langues	64
2.3.1	Texte authentique	64
2.3.2	Corpus et collections	68
2.3.3	Outils	69
2.3.4	Instrumentation du corpus et des outils d'interrogation	70
2.3.5	Instrumentalisation du corpus et des outils d'interrogation	75
2.3.6	Une problématique différente, des outils communs	81
2.4	Des EIAH aux EIAH	84
Chapitre 3 Les standards de description de ressources pédagogiques		87
3.1	La description : un premier pas vers l'indexation	87

3.1.1	Normes et standards	88
3.1.2	Metadonnées	93
3.2	Les acteurs	94
3.2.1	AICC	94
3.2.2	La DCMI	95
3.2.3	edna et GEM	96
3.2.4	ARIADNE et IMS	97
3.2.5	IEEE – LTSC	97
3.2.6	ADLnet	98
3.2.7	ISO/IEC JTC1 SC36	98
3.3	Quels standards, pour quels objets pédagogiques ?	98
3.3.1	LOM	98
3.3.2	SCORM	101
3.3.3	OUN-EML / IMS-LD	103
3.3.4	Vers une typologie des objets pédagogiques	104
3.4	Étude du LOM	105
3.4.1	Spécifications LOM : notions générales	106
3.4.2	Éléments « non pédagogiques »	108
3.4.3	Catégorie 5, Educational	110
3.5	Les langages d’indexation de ressources pédagogiques	117
3.5.1	LOM, un modèle inadapté à notre travail ?	117
3.5.2	Les autres langages d’indexation	118
3.5.3	Conclusion	120
	Synthèse de la partie théorique	123
1	Indexation pédagogique	123
2	Le texte dans l’enseignement des langues	124
3	Notre point de vue : celui de l’ALAO	124

4	Travers de l'ALAO et influence sur notre démarche	125
5	Étude de l'existant	126
5.1	Corpus	126
5.2	Standards de description de ressources pédagogiques	127
Partie II Recueil et analyse des données		131
	Préambule : Choix d'un mode de recueil de données	133
Chapitre 4 Mise en évidence du contexte pédagogique		137
4.1	Hypothèses et Objectifs	137
4.1.1	Recherche <i>vs</i> processus de recherche	138
4.1.2	Formats	139
4.1.3	Population	140
4.1.4	Textes fabriqués	143
4.1.5	Recueils	143
4.1.6	« Recherche passive »	144
4.1.7	Utilisations différentes d'un texte	145
4.1.8	Critères	145
4.1.9	Récapitulatif	146
4.2	Protocole	149
4.2.1	Support	149
4.2.2	Conditions de passation	149
4.3	Description de la population	153
4.3.1	Langues et Pays	154
4.3.2	Description du public	159
4.3.3	Des enseignants globalement expérimentés	161
4.3.4	Méthodes et éclectisme	163
4.3.5	Réduction du contexte	166

4.4	Vers une description du processus de recherche de textes : les réponses des enseignants	167
4.4.1	Texte fabriqué et maîtrise du contenu	167
4.4.2	Recueils de textes et recherche passive	173
4.4.3	Provenances et sources	175
4.4.4	Optique de conservation d'un texte et critères de recherche	182
4.4.5	Confirmation de H_b	187
4.5	Caractérisation du processus selon différents facteurs	188
4.5.1	Codage des données	189
4.5.2	Influence de la langue enseignée	191
4.5.3	Influence du cadre de l'enseignement	192
4.5.4	Influence de l'expérience de l'enseignant	192
4.5.5	Un processus général relativement stable	193
4.6	Flux de la construction du texte comme ressource pédagogique et CP	195
4.6.1	De l'approche instrumentale au flux social du <i>faire</i>	196
4.6.2	Adaptation du concept d'Holloway à notre problématique	198
4.6.3	<i>Faires</i> en marge de notre problématique	202
4.6.4	Recherche pour une activité	205
4.6.5	Introduction du contexte pédagogique et conséquences sur la modélisation des pratiques en rapport avec la planification d'activités faisant intervenir des textes	211
4.7	Conclusion	213
Chapitre 5 Mise en évidence de l'influence du CP		215
5.1	Protocole et objectifs	215
5.1.1	Codage d'un système de génération de questionnaire	215
5.1.2	Mode de diffusion	216
5.1.3	Objectifs du questionnaire	216
5.1.4	Réponses	218

5.1.5	Récapitulatif	219
5.2	Recueils de textes	220
5.2.1	Possession et organisation de recueils	221
5.2.2	Codage des données	221
5.2.3	Non pris en compte	229
5.2.4	Des catégories à dominante liée à la situation d'apprentissage	230
5.3	Établissement de relations entre propriété de texte et des CP	231
5.3.1	Les différents types d'activité considérés	232
5.3.2	Les traits considérés	236
5.3.3	Influences de la classe d'activité	238
5.3.4	Influence du public	243
5.3.5	Synthèse des influences mises en évidence	244
5.4	Conséquences sur le contexte pédagogique et sa modélisation	245
5.4.1	Exemple de lien entre des aspects du contexte pédagogique et des propriétés des textes	246
5.4.2	Le contexte pédagogique comme paradigme	249

Partie III Proposition d'un modèle 251

Chapitre 6 Représentation des informations sur les textes, le modèle Prisme-Facette 253

6.1	Propriétés et stratégie	254
6.2	Le couple <i>facette-prisme</i>	256
6.2.1	Prisme	256
6.2.2	Facette et valeur de facette d'un texte selon un prisme	257
6.2.3	Facettes et contexte pédagogique	259
6.2.4	Prismes et contexte pédagogique	261
6.3	Sélection et prismes	263

6.3.1	La contrainte comme fonctionnalité du prisme	264
6.3.2	Prismes pour la sélection	264
6.3.3	Modification du traitement d'un prisme pour permettre la prise de décision	265
6.4	Facette <i>vs</i> métadonnées	266
6.4.1	Informations sur des ressources	266
6.4.2	Données sous-jacentes	268
6.5	L'aide à l'évaluation comme complément de la sélection	269
6.5.1	Approche moins-disante et prolongement de la problématique	270
6.5.2	Évaluation et <i>visualisations</i>	270
6.5.3	Dimension fonctionnelle de la visualisation	271
6.5.4	Vue et contexte pédagogique	272
6.5.5	Vues et compréhension du système	272
Chapitre 7 Mise en œuvre du modèle prisme-facette		275
7.1	Collection de textes	275
7.1.1	Collections et contexte pédagogique	276
7.2	La recherche de texte du point de vue de l'utilisateur-enseignant	278
7.2.1	Affichage de l'interface	280
7.2.2	Recherche de textes	280
7.2.3	Précision du CP	280
7.3	Indexation et propriétés sous-jacentes	281
7.3.1	Séparation des données et des traitements	281
7.3.2	Fonctions et pré-traitements	282
7.3.3	Prisme et SRI	285
7.3.4	Ajout d'un texte au système	285
7.3.5	Trois types de scripts	286
7.4	Architecture d'un système d'indexation pédagogique de textes	287

7.4.1	Niveau des scripts	287
7.4.2	Coordination de prismes et de visualisations	289
7.5	Conception pour une implémentation informatique du modèle	291
7.5.1	Point de départ : la facette	291
7.5.2	<i>Package</i> « Fonctions et scripts » et prisme	292
7.5.3	La sélection de textes	294
7.5.4	Évaluation	296
7.5.5	Indexation	296
7.5.6	Maquette	298
7.5.7	Conception d'un système générique	302
7.6	Ajout d'un prisme au système	302
7.6.1	Définition de la facette	302
7.6.2	Définition du pré-traitement	303
7.6.3	Définition du contexte pédagogique (et du calcul)	303
7.6.4	Cadre pour la conception de facettes	304
	Conclusion	305
	Chapitre 8 Perspectives et points de vues	307
8.1	Quels textes indexer ?	307
8.1.1	Textes tout-venants et RSS	307
8.1.2	Textes « didactisés »	308
8.2	Élaboration de prismes	308
8.2.1	Réutilisation de pré-traitements et combinaison avec de nouvelles fonctions	309
8.2.2	Le système comme source d'informations pour les chercheurs	310
8.3	L'architecture du système	311
8.3.1	Modèles de l'enseignant et de l'apprenant	311

8.3.2	Interfaces	312
8.3.3	Combinaison de prismes	313
8.3.4	Collections	314
8.4	Adaptation à d'autres utilisations	315
8.4.1	Indexation pédagogique de problèmes pour l'enseignement de l'algorithme	315
8.4.2	Plateforme, par exemple MIRTO	316
8.5	Hybridation et ALAO intégratif	316
	Acronymes	319
	Bibliographie	323
	Index	357
	Annexes	a-1
	Annexe A Questionnaire 1 : Choix et utilisation de ressources textuelles en cours de langues	a-3
	Annexe B Statistiques officielles	a-11
	Annexe C Recours au texte fabriqué (questionnaire 1)	a-13
	Annexe D Croisement « sources » et « provenance » (questionnaire 1)	a-19
	Annexe E Documentation du gestionnaire de questionnaires	a-29
	Annexe F Réponses au second questionnaire	a-39
F.1	Précisions / mise à jour des réponses au questionnaire n°1	a-39
F.2	Apports potentiels du TAL dans l'organisation de la base de textes	a-57
F.2.1	Présentation de la rubrique	a-57
F.2.2	Au niveau du mot	a-57

F.2.3	Formes fléchies	a-59
F.2.4	Au niveau de la phrase	a-60
F.3	Utilisation des textes	a-68
F.3.1	Exercices Lacunaires	a-68
F.3.2	Introduction d'une nouvelle notion grammaticale ou de nouveau vocabulaire	a-89
F.3.3	Activité de compréhension	a-100
F.3.4	Activité de traduction	a-111
F.4	Influence de la langue maternelle des apprenants et de leur niveau	a-115
F.5	Types de textes	a-116
F.5.1	Catégorisation des résultats	a-118
F.6	Suggestions	a-123
Annexe G Documentation de la classe BD de la maquette		a-127

Table des figures

1.1	Diagramme UML des cas d'utilisation pour notre prototype	8
1.2	Schéma général d'un modèle de Recherche d'Information [Géry, 2002, p. 4]	13
1.3	Un modèle de la situation d'enseignement-apprentissage [Rézeau, 2001, Figure 1.5 p. 45]	16
2.1	Principaux courants en informatique et éducation [Bruillard, 1997, Figure 1.1 p. 24]	30
2.2	« Un modèle typique de l'ALAO et les domaines scientifiques associés [Ahmad <i>et al.</i> , 1985, p. 45] » [Bannert <i>et al.</i> , 2003, Fig. 1]	53
2.3	Index bibliographique KWIC [Luhn, 1960, Figure 1 p. 290]	62
2.4	Concordances présentées par MICRO-CONCORD [Johns, 1986, p. 151]	62
2.5	SACODEYL Annotator ¹¹³	79
3.1	Le processus de normalisation (adapté de [SCTIC, 2002, p. 12–13])	89
3.2	Acteurs et standards pour la description de ressources pédagogiques	94
3.3	<i>LOM aggregation levels</i> (Descripteur 1.8, voir figure 3.4)	99
3.4	Les éléments du LOM [Hodgins & Duval, 2004, tsp. 12]	100
3.5	Illustration conceptuelle du modèle d'agrégation de contenu [SCORM, 2006a, Figure 2.1.5a, p. CAM-2-8 (26)]	103
3.6	Exemple d'activité de production à partir du discours de Martin Luther King, Jr. http://tr.im/dreamtoo	112
3.7	Exemple d'activité de vocabulaire à partir du discours de Martin Luther King, Jr. http://tr.im/dreamlac	114
3.8	Exemple d'exercice lacunaire sur l'opposition entre « will » et « shall » à partir du discours de Martin Luther King, Jr. Exercice généré avec la version de 2005 du prototype de la plateforme MIRTO.	115
3.9	Interface de recherche avancée dans le dépôt edna http://www.edna.edu.au/edna/go/search	119

4.1	Parcours et effectifs (questionnaire 1)	152
4.2	Distribution de l'expérience des enseignants (débutants et expérimentés)	162
4.3	Distribution des enseignants en fonction du nombre de méthodes sélectionnées (la courbe indique la distribution normale)	165
4.4	Comparaison des réponses à la question A.2.5 (provenances), classées par choix	176
4.5	Comparaison des réponses à la question A.2.6 (sources), classées par choix	179
4.6	Transformation du texte de son écosystème d'origine au statut de texte utilisé dans l'enseignement des langues	199
4.7	La recherche de textes vue selon le modèle de Yinger [1978]	206
5.1	Exemple d'exercice lacunaire d'inférence du sens d'un mot d'après le contexte [Gómez Casañ & Martín Viaño, 1990, p. 50]	233
5.2	Tolérance à la présence d'éléments inconnus pour chaque classe d'activité	241
5.3	Influence du contexte pédagogique sur l'assignation de propriétés au texte	246
6.1	L'apprentissage des langues [Rüschhoff, 1993, p. 7]	255
6.2	Exemples de prismes et de valeurs pour les facettes correspondantes (ici, constantes quel que soit le contexte pédagogique)	258
6.3	Exemple de texte.	259
6.4	Exemple de prisme et facettes correspondantes	260
6.5	Facettes d'un texte	262
6.6	Exemple de visualisations d'un texte pour un même contexte pédagogique	271
7.1	Exemple de séquence de recherche de textes	279
7.2	Modèle hiérarchisé de la plate-forme MIRTO [Antoniadis <i>et al.</i> , 2005b, p. 68]	281
7.3	Prismes et fonctions	283
7.4	Séquence d'ajout d'un texte au système	286
7.5	Architecture modulaire à base de prismes	288
7.6	Diagramme de classe d'une implémentation (recherche de textes)	293
7.7	Diagramme de classe d'une implémentation (indexation)	297
7.8	Problèmes clés dans la définition d'un prisme	304
8.1	Exemple de portion d'interface associée à un prisme	313
B.1	Places ouvertes aux concours de l'agrégation et du CAPES de 2002 à 2006. Données compilées à partir de [MEN, 2003, 2004, 2005, 2006, 2007].	a-12
F.1	Expérience des enseignants ayant répondu au questionnaire	a-45

F.2	Longueur d'un texte pour un exercice lacunaire de travail sur l'inférence du sens d'un mot	a-72
F.3	Longueur d'un texte pour un exercice lacunaire à « objectif linguistique »	a-77
F.4	Longueur d'un texte pour un exercice lacunaire de compréhension	a-84
F.5	Longueur d'un texte pour l'introduction d'une nouvelle notion grammaticale	a-90
F.6	Nombre d'éléments représentatifs de cette notion grammaticale dans le texte	a-92
F.7	Longueur d'un texte pour l'introduction de nouveau vocabulaire	a-94
F.8	Nombre d'éléments représentatifs de ce nouveau vocabulaire dans le texte	a-96
F.9	Longueur d'un texte pour une activité de compréhension	a-107

Liste des tableaux

1.1	Les différentes méthodes et approches utilisées par les enseignants et concepteurs de ressources ALAO d'après [Levy, 1997, p. 123]	23
2.1	« Les huit fonctions pédagogiques et leurs caractéristiques » [de Vries, 2001, p. 112]	34
2.2	De « l'ALAO communicatif » à « l'ALAO intégratif »	42
2.3	Récapitulatif des « phases de l'ALAO » de Warschauer [2000]	47
2.4	Types d'activités utilisant de corpus dans l'enseignement des langues	72
3.1	Les éléments du DCES [DCMI, 2008]	95
4.1	Objectifs des questions du premier questionnaire	148
4.2	Réponses à la question A.1.1 p. a-3 (L2)	154
4.3	Cadre et pays d'exercice pour des enquêtés qui enseignent plus d'une langue . . .	156
4.4	Réponses à la question A.1.5 p. a-3 (Pays d'exercice)	157
4.5	Réponses à la question A.1.4 p. a-3 (Cadre de l'enseignement)	159
4.6	Réponses à la question A.1.7 p. a-5 (Méthodes)	164
4.7	Réponses à la question A.2.1 p. a-5 (Usage des textes authentiques)	168
4.8	Réponses à la question A.2.2 p. a-5 (Obtention des textes fabriqués)	170
4.9	Possession et organisation d'un recueil de textes (questions A.2.4 et a) p. a-6) . .	173
4.10	Possession d'un recueil croisée avec le fait de trouver des textes par hasard	174
4.11	Indicateurs en fonction du mode de comptage pour les provenances	177
4.12	Indicateurs en fonction du mode de comptage pour les sources	180
4.13	Corrélations entre les provenances et la source « manuel »	180
4.14	Importance des lectures personnelles comme provenance croisée avec le fait de trouver des textes par hasard	181
4.15	Réponses à la question a) p. a-7 (Optique dans laquelle les enseignants conservent les textes qu'ils trouvent par hasard)	183

4.16	Possession d'un recueil croisée avec le fait de conserver des textes trouvés par hasard en attendant de trouver une manière adéquate de les utiliser	184
4.17	Conservation des textes en attendant de trouver une manière de l'utiliser croisée recherche d'un moyen d'utiliser un texte plutôt qu'un texte pour une activité . .	185
4.18	Réponses à la question A.2.9 p. a-7 (Classes de critères de recherche de texte) . .	186
4.19	112 témoignages concernant la réutilisation d'un texte dans différents contextes (questions A.2.8 et a) p. a-7)	187
5.1	Complétude des réponses des enseignants (moyenne 50 réponses par question) . .	219
5.2	Récapitulatif des réponses aux questions 11 et 12	222
5.3	Témoignage [38], réponse aux questions 11 et 12	223
5.4	Catégorisation des témoignages des enseignants répondant aux questions n°11 et 12 du second questionnaire	230
5.5	Comparaison entre les types d'activité sur le plan de la longueur des textes . . .	238
5.6	Valeurs significatives pour le test post-hoc de Tukey (Classes d'activités / longueur des textes)	239
5.7	Comparaison entre les types d'activité sur le plan du nombre d'éléments représentatifs d'une notion	240
5.8	Valeurs significatives pour le test post-hoc de Tukey (Classes d'activités / nombre d'éléments représentatifs)	240
5.9	Type d'activité croisé avec la tolérance vis-à-vis du vocabulaire inconnu	242
5.10	Type d'activité croisé avec la tolérance vis-à-vis des structures grammaticales inconnues	243
7.1	Exemple de sorties de TreeTagger	300
C.1	Réponses aux questions A.2.1, A.2.2 et A.2.3 p. a-5	a-18
F.1	Réponses à la question n° 1 du second questionnaire	a-39
F.2	Réponses à la question n° 2 du second questionnaire	a-39
F.3	Réponses à la question n° 3 du second questionnaire	a-40
F.4	Réponses à la question n° 4 du second questionnaire	a-41
F.5	Réponses à la question n° 5 du second questionnaire	a-43
F.6	Réponses à la question n° 6 du second questionnaire	a-44
F.7	Réponses à la question n° 8 du second questionnaire	a-46
F.8	Catégorisation des réponses à la question n° 8 du second questionnaire	a-46
F.9	Réponses à la question n° 9 du second questionnaire	a-47
F.10	Réponses à la question n° 10 du second questionnaire	a-47

F.11 Réponses à la question n° 11 du second questionnaire	a-49
F.12 Réponses à la question n° 12 du second questionnaire	a-55
F.13 Réponses à la question n° 16 du second questionnaire	a-57
F.14 Réponses à la question n° 15 du second questionnaire	a-57
F.15 Réponses à la question n° 14 du second questionnaire	a-58
F.16 Réponses à la question n° 17 du second questionnaire	a-58
F.17 Réponses à la question n° 18 du second questionnaire	a-58
F.18 Réponses à la question n° 19 du second questionnaire	a-59
F.19 Réponses à la question n° 20 du second questionnaire	a-59
F.20 Réponses à la question n° 21 du second questionnaire	a-59
F.21 Réponses à la question n° 22 du second questionnaire	a-60
F.22 Réponses à la question n° 23 du second questionnaire	a-60
F.23 Réponses à la question n° 30 du second questionnaire	a-61
F.24 Réponses à la question n° 24 du second questionnaire	a-61
F.25 Réponses à la question n° 25 du second questionnaire	a-64
F.26 Réponses à la question n° 26 du second questionnaire	a-64
F.27 Réponses à la question n° 27 du second questionnaire	a-66
F.28 Réponses à la question n° 28 du second questionnaire	a-66
F.29 Réponses à la question n° 29 du second questionnaire	a-67
F.30 Réponses à la question n° 31 du second questionnaire	a-68
F.31 Réponses à la question n° 32 du second questionnaire	a-69
F.32 Catégorisation des réponses à la question n° 32 du second questionnaire	a-70
F.33 Réponses à la question n° 33 du second questionnaire	a-71
F.34 Normalisation des réponses à la question n° 33 du second questionnaire	a-71
F.35 Réponses à la question n° 34 du second questionnaire	a-72
F.36 Réponses à la question n° 35 du second questionnaire	a-72
F.37 Réponses à la question n° 36 du second questionnaire	a-72
F.38 Réponses à la question n° 37 du second questionnaire	a-73
F.39 Réponses à la question n° 38 du second questionnaire	a-74
F.40 Réponses à la question n° 39 du second questionnaire	a-75
F.41 Normalisation des réponses à la question n° 39 du second questionnaire	a-76
F.42 Réponses à la question n° 40 du second questionnaire	a-78
F.43 Réponses à la question n° 41 du second questionnaire	a-78

F.44 Réponses à la question n° 42 du second questionnaire	a-79
F.45 Réponses à la question n° 43 du second questionnaire	a-79
F.46 Réponses à la question n° 44 du second questionnaire	a-82
F.47 Réponses à la question n° 45 du second questionnaire	a-83
F.48 Normalisation des réponses à la question n° 45 du second questionnaire	a-84
F.49 Réponses à la question n° 46 du second questionnaire	a-85
F.50 Réponses à la question n° 47 du second questionnaire	a-85
F.51 Réponses à la question n° 48 du second questionnaire	a-86
F.52 Réponses à la question n° 49 du second questionnaire	a-86
F.53 Réponses à la question n° 50 du second questionnaire	a-88
F.54 Réponses à la question n° 51 du second questionnaire	a-89
F.55 Réponses à la question n° 52 du second questionnaire	a-90
F.56 Réponses à la question n° 54 du second questionnaire	a-91
F.57 Réponses à la question n° 56 du second questionnaire	a-92
F.58 Réponses à la question n° 58 du second questionnaire	a-93
F.59 Réponses à la question n° 53 du second questionnaire	a-94
F.60 Réponses à la question n° 55 du second questionnaire	a-95
F.61 Réponses à la question n° 57 du second questionnaire	a-96
F.62 Réponses à la question n° 59 du second questionnaire	a-97
F.63 Réponses à la question n° 60 du second questionnaire	a-97
F.64 Réponses à la question n° 61 du second questionnaire	a-100
F.65 Réponses à la question n° 62 du second questionnaire	a-100
F.66 Réponses à la question n° 63 du second questionnaire	a-102
F.67 Classement des réponses à la question 63	a-105
F.68 Réponses à la question n° 64 du second questionnaire	a-107
F.69 Réponses à la question n° 65 du second questionnaire	a-107
F.70 Réponses à la question n° 66 du second questionnaire	a-108
F.71 Réponses à la question n° 67 du second questionnaire	a-108
F.72 Réponses à la question n° 68 du second questionnaire	a-111
F.73 Réponses à la question n° 69 du second questionnaire	a-111
F.74 Réponses à la question n° 70 du second questionnaire	a-112
F.75 Réponses à la question n° 71 du second questionnaire	a-114
F.76 Réponses à la question n° 77 du second questionnaire	a-115

F.77 Réponses à la question n° 72 du second questionnaire	a-115
F.78 Réponses à la question n° 73 du second questionnaire	a-116
F.79 Réponses à la question n° 74 du second questionnaire	a-116
F.80 Réponses à la question n° 75 du second questionnaire	a-118
F.81 Classement des réponses à la question 75	a-121
F.82 Catégorisation des réponses à la question n° 75 du second questionnaire	a-122
F.83 Réponses à la question n° 76 du second questionnaire	a-125

Conventions

Dans ce manuscrit nous avons opté pour un certain nombre de conventions d'écriture et de mise en forme. Certaines sont indiquées avant la portion du texte qui les concerne, sous forme d'encart (cf. NB 4.1 p. 154 et NB 7.1 p. 292), les autres sont explicitées ci-dessous.

1 Traductions

Toute citation d'une langue étrangère est traduite ou glosée. Pour ne pas faire perdre de temps aux lecteurs qui n'ont que faire d'une traduction (souvent imparfaite), nous avons grisé toutes les notes de bas de page qui les concernent. Une note de bas de page sera donc indiquée ainsi¹, alors qu'une traduction sera indiquée de cette manière². Quand l'auteur de la traduction n'est pas indiqué, cela signifie qu'il s'agit de notre propre traduction.

2 Bibliographie

Malgré nos efforts, nous n'avons pas toujours pu retrouver les originaux des articles ou ouvrages auxquels se référaient d'autres auteurs. Quand nous faisons références à des travaux cités, que nous n'avons pu vérifier par nous-même, nous les indiquons dans la bibliographie en gris en les liant aux ouvrages dans lesquels nous les avons retrouvés, voir par exemple [Bar-Hillel, 1964].

3 Liens

Tous les liens présents dans ce manuscrit ont été vérifiés au mois d'octobre 2009 (cf. remerciements). Seuls ceux qui étaient brisés à cette époque et que nous n'avons pu retrouver sont indiqués avec une date de dernière consultation.

4 Renvois

Le manuscrit comprend de nombreux renvois qui sont destinés à rendre plus aisée la lecture partielle sur écran.

1. Une note de bas de page.

2. Une traduction. (traduit par auteur de la traduction).

Première partie

Position du problème et étude de l'existant

Chapitre 1

Indexation pédagogique (de textes (pour l’enseignement des langues))

1.1 Genèse du projet

1.1.1 Le projet MIRTO

Le projet de recherche dont traite ce manuscrit tire ses origines du projet Multi-apprentissages Interactifs par des Recherches sur des Textes et l’Oral (MIRTO) [Antoniadis & Ponton, 2002] de l’Université Stendhal. Ce projet, commencé en 2001, a pour objectif la création d’une plateforme offrant « aux enseignants de langue la possibilité de concevoir des scénarios pédagogiques en bénéficiant pleinement d’outils TAL » [Antoniadis *et al.*, 2005b]. Il naît de la constatation d’un certain nombre de problèmes communs à beaucoup de systèmes d’Apprentissage des Langues Assisté par Ordinateur (ALAO) (sur lesquels nous reviendrons dans le prochain chapitre), qui découlent tous d’un même fait :

« le plus souvent, informaticiens et didacticiens des langues n’admettent pas la même acception du terme “langue”. Pour les uns, il s’agit d’une séquence de codes, pour les autres d’un système de formes et de concepts. Cette différence peut s’expliquer aisément lorsque l’on constate que l’informatique [seule], par définition, ne peut considérer et ne peut traiter que la forme de la langue dénuée de toute interprétation, alors que pour la didactique des langues, la forme n’existe que par ses propriétés et des concepts qu’elle est censée représenter. »

[Antoniadis *et al.*, 2005b, p. 65]

Pour Antoniadis et Ponton, une solution à ce problème réside dans l’utilisation de fonctions de Traitement Automatique des Langues (TAL), d’où la création de la plateforme MIRTO.

Dès la phase de conception, MIRTO place – comme son nom l’indique (« par des Recherches sur des Textes ») – le texte au centre de son approche. L’un des objectifs du projet est de permettre un renouvellement des activités proposées en créant plusieurs exercices d’un même type à partir de textes tout-venant. L’intégration d’un corpus à la plateforme découle directement de cette approche. Le système doit permettre de générer des exercices à partir de n’importe quel texte,

a fortiori à partir des textes d'un corpus. Permettre aux enseignants de choisir les supports de ces activités en fonction de critères relevant de leur problématique devient une nécessité et notre projet commence à la suite du lancement du projet MIRTO, à la rentrée universitaire 2002 dans le cadre d'un mémoire de Diplôme d'Études Approfondies (DEA).

Un des objectifs de l'intégration d'une base de textes au système est de proposer un système évolutif et générique : pour chaque activité il doit être possible d'en générer une autre analogue à partir d'un autre texte. Notre base doit donc pouvoir contenir et référencer des textes qui n'ont pas nécessairement été déjà utilisés dans l'enseignement des langues et l'organiser de manière à permettre aux enseignants (et éventuellement à plus long terme au système MIRTO lui-même) d'accéder à des textes qui serviront de base pour leurs activités (ou celles générées par la plateforme).

1.1.2 Un projet autonome

Si MIRTO est à l'origine de notre projet, il ne le gouverne pas. La création d'une base de textes, pour l'enseignement des langues est un projet à part entière, et l'intégration à la plateforme MIRTO, sans en avoir été exclue, n'est qu'une perspective à long terme. L'approche adoptée, dès le travail préliminaire, est de se détacher de l'unique perspective de l'utilisation dans un Système de Gestion de l'Apprentissage (SGA) et de tenter de s'appuyer sur les pratiques enseignantes dans un contexte aussi large que possible.

L'étude préliminaire effectuée en DEA avait donné lieu à 8 entretiens avec des enseignants de profils différents [Loiseau, 2003, p. 18] :

- 3 langues : anglais, espagnol, polonais ;
- entre 1 et 30 ans d'expérience de l'enseignement ;
- enseignants au collège, à l'université et en formation continue ;
- public débutant à Bac+5 (spécialistes) ;
- utilisation de l'outil informatique à titre personnel, jusqu'à la création d'activités pédagogiques avec support informatique.

Au cours de cette étude préliminaire, il est apparu que :

- tous les enseignants interrogés étaient amenés à rechercher des textes rédigés en L2¹ ;
- la recherche de textes représentait une part importante du temps consacré à préparer leurs cours.

Certains des enseignants consultés dans cette étude se constituaient d'ailleurs eux-mêmes des recueils de textes (informatisés ou non) afin de diminuer le temps alloué à cette recherche. Tous les enseignants interrogés ayant un recueil – soient 3 sujets – organisent leur recueil selon des critères différents. Si une étude préliminaire fondée sur huit sujets ne permet pas de tirer de conclusions générales, le fait que tous les sujets interrogés insistent sur le temps alloué à la recherche de matériel textuel pour leur cours – indépendamment de la langue enseignée, de leur expérience, du contexte dans lequel ils enseignent ou de leur utilisation ou non de l'outil informatique – semble indiquer que la recherche de texte fait partie des tâches du métier d'enseignant de langue² et, par conséquent, que le projet, ne doit pas se cantonner à une problématique d'ALAO.

Dans la suite de ce chapitre introductif nous répondrons à deux questions majeures dans l'optique de la réalisation de ce projet :

1. Dans cette étude, les langues enseignées sont exclusivement des L2, voir NB 1.1 p. 7
 2. Hypothèse H_a p. 138.

Nota Bene 1.1 L1 et L2

Alors que dans le langage courant, le terme consacré est « langue maternelle », opposé à la « langue étrangère », en didactique des langues, « un examen appuyé sur des sources historiques, sociolinguistiques et développementales [a mené] à une complexification du statut des langues et de leur rôle chez les sujets parlants » [Cuq, 2003, article “langue maternelle” p. 150]. Dans ce mémoire nous parlerons donc de L1 et de L2 dans l’acception de Matthey dans [Matthey, 2003, pp. 1 & 13] :

- une L1 est la (ou les) langue(s) de la socialisation première (n’excluant donc pas plusieurs L1) ;
- une L2 correspond à toute acquisition ultérieure.

Dans le cadre de notre travail, L2 est ainsi équivalent à Ln, puisque nous recouvrons sous le même terme la seconde, la troisième et la n-ième langue acquise.

Dans ce mémoire, L2 désignera aussi bien une « langue étrangère », qu’une « langue seconde » – celle du Français Langue Seconde (FLS) –, qui « bien que n’étant pas langue première, possède une ou plusieurs fonctions dans le milieu à titre de langue véhiculaire, langue de culture, langue scolaire ou deuxième langue officielle » [Mackey, 1997].

Nous avons volontairement simplifié les notions ici, sans entrer dans les débats qu’elles suscitent. Pour un point de vue sociolinguistique synthétique des notions de L1, L2, langue maternelle et langue étrangère consulter [Mackey, 1997]. Pour une étude détaillée du statut de la langue seconde, consulter [Cuq, 1991].

- un projet pour quoi ?
- un projet pour qui ?

À travers les réponses à ces deux questions naïves, nous indiquerons les lignes directrices méthodologiques qui guideront ce projet. Elles nous donneront également l’occasion de définir plus précisément notre problématique.

1.2 Définition du projet : cas d’utilisation

Dans cette section nous allons préciser la thématique de notre projet et définir les objectifs d’une base de textes définie pour l’enseignement des langues.

Nous représentons dans la figure 1.1, les utilisations potentielles d’une telle base par un utilisateur-enseignant, sous forme d’un diagramme de « cas d’utilisation » (l’un des multiples types de diagrammes de la notation *Unified Modeling Language* (UML)³). Dans [Cockburn, 1997b]⁴, Cockburn suggère plusieurs « *scopes*⁵ ». Ces *scopes* découlent de trois niveaux de lecture des diagrammes de cas d’utilisation. Nous nous intéressons ici en particulier au « *strategic scope* ». L’article de Cockburn se place principalement dans un contexte industriel de création d’un produit informatique pour une entreprise et de sa position dans la politique économique de cette dernière. En ce qui nous concerne, le « *strategic scope* » correspondra au point de vue des objectifs globaux du projet, la différence résidant dans le fait que nos objectifs sont scientifiques et non économiques. Nous le différencierons du « *system scope* », qui concerne plus particulièrement le système tel que nous le développons. Le point de vue du système est plus détaillé

3. Pour plus d’informations sur les diagrammes de cas d’utilisation consulter <http://www.agilemodeling.com/style/useCaseDiagram.htm> résumant [Ambler, 2005, p. 33–45].

4. Article initialement publié en deux parties : [Cockburn, 1997a], puis [Cockburn, 1997c]. Un « *digest* » est disponible : http://alistair.cockburn.us/index.php/Use_case_fundamentals.

5. Le terme anglais regroupe la notion de « cadre » et de « portée ». Par la suite, nous parlerons de point de vue, sans que cela soit une traduction officielle de *scope*.

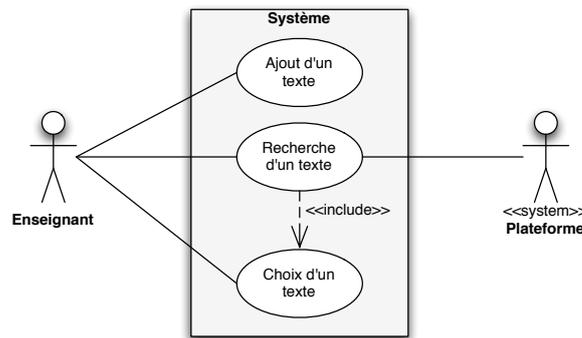


FIGURE 1.1: Diagramme UML des cas d'utilisation pour notre prototype

dans la mesure où il s'approche plus de la phase d'implémentation. Il est plus spécifique que le point de vue global (ou stratégique). Entre le *strategic scope* et le *system scope*, des limitations techniques ou temporelles liées au développement du système ou à sa nature (prototype) peuvent intervenir. Dans le cadre de ce paragraphe, nous nous plaçons principalement dans le contexte « stratégique ».

1.2.1 Ajout d'un texte

Le premier cas d'utilisation doit, notamment pour des raisons pratiques, permettre aux utilisateurs d'ajouter leurs textes à la base. La population de la base est en soit un problème complexe, qui recoupe une problématique didactique (choix des textes à ajouter), mais aussi juridique (droits d'auteur des textes). En permettant aux utilisateur d'ajouter leurs textes, nous déchargeons de cette tâche à la manière des applications « Web 2.0 »⁶ :

« La caractéristique de base de ces sites est qu'une part très significative des contenus qui les alimentent y sont "postés" par les utilisateurs eux-mêmes. Ces sites s'écartent donc de la logique traditionnelle "descendante" où les médias, dépositaires des informations, envoient celles-ci vers des lecteurs ou des spectateurs passifs ; ici, l'utilisateur est à la fois à l'origine de l'émission et à la réception des informations. Les sites du Web 2.0 jouent dès lors le rôle d'interface. »

[Wauthy, 2008]

Dans ce cas d'utilisation, « les contenus » sont les textes et le rôle d'interface de notre système est de permettre aux utilisateurs d'en ajouter. Comme cette fonctionnalité doit être accessible à l'utilisateur final du système, un enseignant, qui n'est pas nécessairement aguerri à l'utilisation de l'outil informatique, elle doit être le moins contraignante possible. Pour cela, il faudra l'automatiser au maximum.

Ce cas d'utilisation décrit une procédure d'ajout de document, mais ne dit rien quant à leur nature. Des restrictions techniques seront à définir au moment de l'implémentation sur les formats pris en charge par le système. Le fait que l'utilisateur concerné soit un enseignant augmente la probabilité d'ajout de textes éprouvés dans le contexte de l'enseignement des langues. Ce n'est cependant pas un pré-requis pour l'ajout d'un texte, le système doit être capable de traiter tout document. C'est là l'une des contraintes originales du projet, dans le sens où l'un des buts du système est de permettre aux enseignants de choisir parmi un large éventail de textes. Ce cas

6. Il est à noter que cette stratégie n'évacue pas la composante juridique du problème.

d'utilisation n'exclut d'ailleurs pas que nous jouions le rôle de l'enseignant en ajoutant nous-même des textes à la base. Cependant, selon le point de vue « stratégique » développé ici, la population de la base n'est pas un objectif en soi de notre projet, les cas d'utilisation suivants vont préciser le rôle d'interface de notre système dans le cadre de la recherche de documents.

1.2.2 Recherche d'un texte

La recherche de texte est la fonctionnalité centrale de notre travail. Notre objectif est donc de faire que l'utilisateur (acteur « enseignant » dans la figure 1.1) puisse accéder aux textes de la base et retrouver ceux qui sont les plus appropriés en vue de les intégrer à une séquence pédagogique. Le système joue ici le « rôle d'interface⁷ » entre les utilisateurs (acteur « enseignant ») et le contenu (les textes).

L'« enseignant » n'est pas l'unique acteur pour ce cas d'utilisation, une plateforme d'enseignement des langues – dans la figure 1.1, conformément à la notation UML l'utilisateur « plateforme » est indiqué comme utilisateur système avec l'étiquette «**system**» [Ambler, 2005, p. 38] – devra aussi pouvoir effectuer des recherches de textes. Le cas d'utilisation est dans cette configuration une perspective à plus long terme (cf. § 1.1.2 p. 6). Le point central de notre travail étant de proposer cette fonctionnalité à un utilisateur enseignant. En outre, nous ne pensons pas pouvoir fournir à l'utilisateur un seul et unique texte, le plus adapté, à chaque requête (voir cas d'utilisation « Aide au choix d'un texte » et la notion de stratégie « moins-disante » p. 11), de ce fait dans le cas d'un acteur « plateforme » le choix du texte sera effectué parmi les textes candidats par la plateforme elle-même.

1.2.3 Aide au choix d'un texte

Comme l'indique l'étiquette «**include**» [Ambler, 2005, p. 40] de la figure 1.1, le cas d'utilisation « choix d'un texte » est un sous-traitement du cas d'utilisation « recherche d'un texte ». Ignorer cette étape serait considérer que le choix est fait par le système pendant la recherche. Dès lors que tous les textes renvoyés par le système ne sont pas systématiquement conservés par les utilisateurs, ces derniers effectuent un choix : valider ou non la pertinence des réponses du système à leur question. Nous n'imaginons pas que le système puisse effectuer cette tâche à la place de l'utilisateur. À partir du moment où nous considérons qu'une décision a lieu après la réponse du système à la requête, nous avons le choix de considérer que l'utilisateur effectue cette tâche *offline* ou d'associer le système à ce processus d'une manière ou d'une autre. Nous choisissons cette seconde solution et considérons que le système doit pouvoir présenter à l'enseignant les textes, lui offrant la possibilité de les consulter et autant que faire se peut l'assister dans sa prise de décision.

1.3 Indexation pédagogique

Que ce soit dans une bibliothèque municipale, en utilisant un moteur de recherche ou à la fin d'un livre, le processus utilisé pour retrouver rapidement une information s'appelle l'« indexation ». Il suit le même principe qu'il s'appuie sur des fiches cartonnées ou sur un système informatique.

7. cf. [Wauthy, 2008], cité pour le cas d'utilisation « Ajout d'un texte » p. 8

Nous allons le considérer dans un premier temps dans son acception générique, ensuite d'un point de vue informatique et enfin le spécifier plus précisément.

1.3.1 Indexation d'objets

Notre définition minimale du cas d'utilisation « recherche d'un texte » suggère que notre base de textes doit être organisée : notre système doit contenir des textes et doit permettre à l'utilisateur, qu'il soit un enseignant ou un système, de récupérer des textes en fonction de caractéristiques spécifiées par lui. Le processus d'organisation des données du système s'appelle l'indexation.

Définition 1.1 (Indexation) « *Opération qui consiste à décrire et à caractériser un document à l'aide de représentations des concepts contenus dans ce document, c'est-à-dire à transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse.*⁸ » [U-Montréal, 2002]

Il est à noter que, dans cette définition, « concept contenu dans le document » est à prendre au sens large. Par exemple, une base d'articles de recherche pourrait demander que soit renseigné l'*International Standard Serial Number* (ISSN) correspondant quand celui-ci est disponible. Cependant l'ISSN est attribué à un périodique [Wikipedia, 2008d] et non à un article, qui considéré isolément ne le contient pas nécessairement. Même dans ce cas, il serait possible de retrouver le périodique dont est issu l'article et de renseigner l'ISSN. Comment doit-on considérer ce concept ? « Contenu dans le document » ? Associé au document ?

La définition, d'un point de vue plus global, décrit l'opération d'indexation, qui consiste à analyser l'objet à indexer, en « extraire des concepts » du fait de cette analyse, et enfin les exprimer selon un langage documentaire. L'agent de ces opérations n'est pas spécifié. On peut donc imaginer plusieurs configurations : l'analyse pourra être effectuée par un opérateur humain ou une machine, de même l'expression des concepts extraits dans le langage documentaire pourra être effectuée aussi bien par un humain que par une machine. Prenons l'exemple de l'ajout d'un ouvrage dans une bibliothèque, un bibliothécaire analysera l'ouvrage, en extraira certains concepts, par exemple le nom et le prénom de l'auteur, le titre, l'année, mais aussi le thème de l'ouvrage. Il saisira dans une interface prévue à cet effet les différentes informations glanées, le langage documentaire est transparent et il suffit d'entrer les valeurs obtenues. Cependant, dans la bibliothèque où il travaille, une cote est définie pour chaque ouvrage ; elle est composée de la classification Dewey suivie des 3 premières lettres du nom de l'auteur. Pour obtenir la classification Dewey, le bibliothécaire utilise l'interface WebDewey⁹ et la fournit au système de gestion de sa bibliothèque, qui calcule automatiquement la cote de l'ouvrage à partir des informations obtenues (indice Dewey et nom de l'auteur). Dans cet exemple l'analyse est effectuée par un opérateur humain et le système lui sert d'interface pour traduire ces concepts dans le langage documentaire employé par la bibliothèque. Le cas de l'indice Dewey est une solution mixte, puisque l'utilisateur effectue la traduction entre le contenu de l'ouvrage et son indice, mais il est assisté dans sa tâche par un système informatique.

Dans notre cas, l'opération d'indexation fera elle aussi intervenir l'utilisateur et le système. L'utilisateur n'est pas un documentaliste et en tant qu'enseignant, c'est avant tout le cas d'utilisation

8. Cette définition est issue de la norme Z47-102 [AFNOR, 1993, p. 512], que nous n'avons pu consulter, pour des « raisons économiques ». Elle est cependant rappelée en termes analogues dans [Bertrand *et al.*, 1996].

9. <http://www.oclc.org/dewey/versions/webdewey/default.htm>

« Recherche de textes » qui l'intéresse. Les deux parties du processus d'indexation devront donc être aussi simples et peu fastidieuses que possible et pour y parvenir elles devront être automatisées tant que faire se peut. L'analyse de certains concepts du document ne pourra être automatisée, comme par exemple l'auteur du texte ou l'année d'écriture (si tant est que ces critères s'avèrent pertinents en vue de leur exploitation dans l'enseignement des langues). Cependant toute analyse automatisable devra être prise en charge par le système. La partie d'analyse de l'indexation sera donc hybride, en ce sens que certains concepts ne pourront être gérés par le système, mais que la partie la plus fastidieuse devra être automatisée lorsque c'est possible.

a) Analyse et stratégie moins-disante

L'enjeu majeur d'une analyse automatique des documents est la fiabilité de cette analyse, l'informatique ayant tendance à perdre en fiabilité au fur et à mesure que l'on s'écarte des données de surface (le codage des données analysées). Par exemple dans le cas du texte, les caractères qui composent le fichier sont codés de manière non-ambiguë. Si l'analyse du document est un comptage de chacun des caractères (tant de 'a', tant de 'b', etc.), un ordinateur pourra s'acquitter de la tâche avec une parfaite fiabilité. Si les concepts à analyser sont moins superficiels et concernent par exemple les mots, des difficultés commencent à apparaître, à plus forte raison dans des langues comme le chinois [Li, 2003] ou le grec ancien où les mots ne sont pas séparés par des espaces. Encore plus distante de la surface, une analyse morphologique¹⁰ automatique, ne peut à l'heure actuelle être fiable à 100%¹¹. Les niveaux les plus éloignés de la surface en terme d'analyse, les niveaux sémantique et pragmatique manquent encore de fiabilité [Antoniadis, 2004, p. 89] et ne peuvent être abordés avec précision que dans des systèmes dédiés à une application particulière.

Nous n'allons pas développer d'outils d'analyse pour la réalisation de notre prototype, nous nous contenterons de réutiliser des outils existants en gardant en tête ces problèmes de fiabilité. Notre approche s'apparentera à la stratégie dite « moins-disante » proposée par [Kraif & Ponton, 2007]. Le constat de départ est proche de celui de Gamper et Knapp :

*« Artificial Intelligence offers many possibilities to improve computer-assisted language learning systems. However, the application of most of these technologies is not mature yet and still requires more research. »*¹²

[Gamper & Knapp, 2002, p. 338-339]

Notre problématique est bien différente de celle de Kraif et Ponton, qui ont développé cette stratégie moins-disante dans l'optique de systèmes d'ALAO destinés à être utilisés directement par les apprenants. Elle nous semble cependant adaptable à notre problème, dans la mesure où comme pour le diagnostic des erreurs d'apprenants, dans le cadre de l'indexation de documents, les erreurs d'analyse vont nuire aux performances du système : la précision¹³ et le rappel¹⁴ dans

10. « [Segmentation] d'un texte en unités élémentaires auxquelles sont attachées des connaissances dans le système » [Fuchs *et al.*, 1993, p. 83], par exemple « réminiscences ← nom, féminin, pluriel ».

11. L'analyseur le mieux évalué dans l'action Grammaire et Ressources pour les Analyseurs de Corpus et leur Évaluation (GRACE) proposait pour chaque *token* une étiquette et une seule, qui était correcte dans 95% des cas [Paroubek *et al.*, 1999, figure].

12. L'intelligence artificielle offre de vastes possibilités pour l'amélioration de systèmes d'ALAO. Cependant, l'utilisation de ces technologies n'est pas encore assez mature et requiert plus de recherches.

13. « La précision est le nombre de documents pertinents retrouvés rapporté au nombre de documents total proposé par le moteur de recherche pour une requête donnée. » [Wikipedia, 2008h]

14. « Le rappel est défini par le nombre de documents pertinents retrouvés au regard du nombre de documents pertinents que possède la base de données. » [Wikipedia, 2008h]

le cadre de la recherche d'information. Pour l'intégration de chaque outil au système, un équilibre sera à trouver entre la précision et le rappel, équilibre qui dépendra des concepts à extraire des documents et de la fiabilité des outils. Cependant, des outils dont la fiabilité serait trop faible perdraient tout intérêt et généreraient du bruit (précision faible) et/ou provoqueraient du silence (rappel faible).

Nous ne suggérons pas ici que notre système ne devra intégrer que des outils fiables à 100%. L'évaluation de la fiabilité doit aller au delà d'une évaluation purement quantitative et entrer dans le détail du type d'imprécisions (cf. [Lebarbé, 2002, § 3.3.4]), afin d'évaluer leur impact sur l'indexation. À terme, certaines imprécisions seront vraisemblablement compensées par la quantité de documents présents dans la base. Ces considérations ne nous affranchissent pas de la stratégie moins-disante de Kraif et Ponton : notre système ne recherche, ne choisit pas un texte en particulier, il propose à l'utilisateur une aide à la recherche et au choix de textes. C'est à l'utilisateur (enseignant ou système informatique) que revient la décision finale (cf. § 1.2 p. 7).

1.3.2 Langage documentaire

Nota Bene 1.2 Système de Recherche d'Information (SRI)

Dans cette section (1.3.2) nous nous appuyons sur la définition de modèle de recherche d'information de Géry [2002]. Sachant que tout SRI s'appuie sur un modèle de recherche d'information, la question que nous pouvons nous poser est de savoir si un système de recherche de textes pour l'enseignement des langues est un SRI.

Il en donne la définition suivante :

Définition 1.2 (Système de Recherche d'Information (SRI)) « Un SRI est un système informatique qui facilite l'accès à un ensemble de documents (*corpus*), pour permettre de retrouver ceux dont le contenu correspond le mieux à un besoin d'information d'un utilisateur. » [Géry, 2002, p. 2]

Selon cette définition une base de textes indexée pédagogiquement pour l'enseignement des langues est effectivement un SRI. En effet nous pourrions objecter que l'enseignant n'a pas pour lui-même un besoin d'information, puisque l'objectif déclaré est de permettre à l'utilisateur d'intégrer des textes à des séquences pédagogiques (cf. § 1.2.2 p. 9). Cependant, s'il ne cherche pas un support d'information pour lui-même, il s'agit bien d'informations pour les apprenants [Rüschhoff, 1993, p. 7]^a et donc son besoin d'information est gouverné par celui qu'il estime être besoin des apprenants.

^a. Pour plus de détails sur le point de vue de Rüschhoff sur les informations, voir § 6.1 p. 254.

La deuxième partie de l'opération d'indexation consiste à représenter les concepts issus de l'analyse selon les termes d'un « langage documentaire ». En amont de l'action d'indexer, il faut spécifier ce langage :

Définition 1.3 (Langage documentaire) « Un langage documentaire est un langage artificiel qui fournit une représentation formalisée et univoque des documents d'un corpus et des questions qui intéressent un groupe d'utilisateurs, afin de permettre le repérage simple des documents du corpus qui répondent aux questions de ces utilisateurs. » [Lefèvre, 2000]

Cette définition est, à notre avis, exemplaire. Elle permet de toucher à de nombreux aspects de la problématique de l'indexation et *a fortiori* de la nôtre. Nous avons établi certaines contraintes

techniques de l'indexation dans le paragraphe précédent en revenant sur le cas d'utilisation « ajout d'un texte » : du point de vue des performances, les outils d'analyse des documents devront afficher un degré minimum de fiabilité (à définir ultérieurement). Cependant, d'un point de vue qualitatif, le choix des outils dépend du langage documentaire défini pour le système qui décide des concepts à extraire.

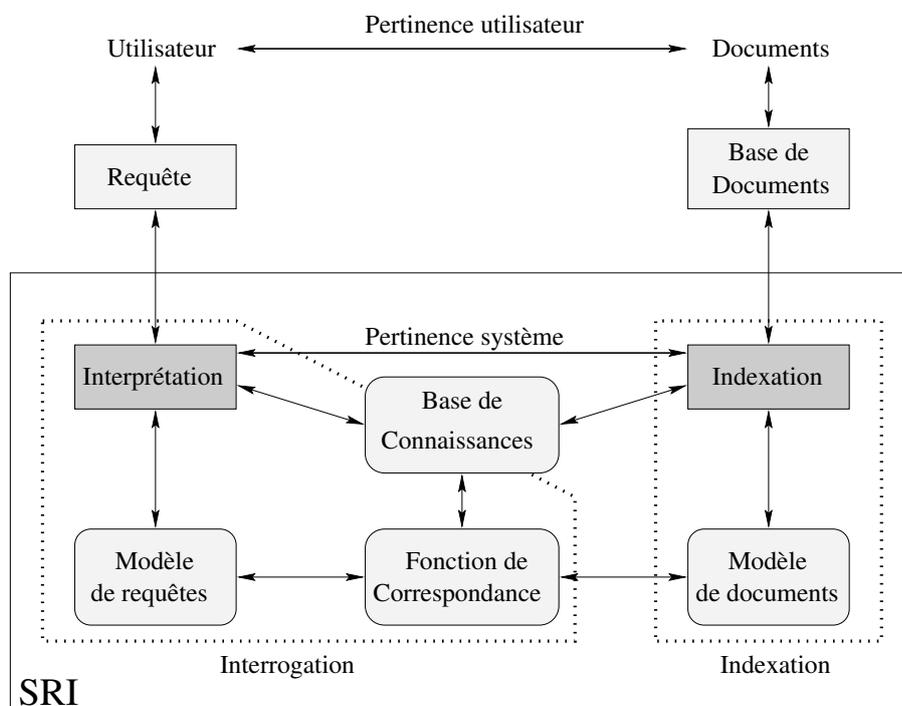


FIGURE 1.2: Schéma général d'un modèle de Recherche d'Information [Géry, 2002, p. 4]

Prise en compte du point de vue informatique, cette définition permet à elle seule de définir les rôles des principaux composants d'un modèle de recherche d'information [Géry, 2002, § 1.2.3] (cf. NB 1.2 p. 12 et figure 1.2 de la présente page). En effet, le langage documentaire, doit permettre une représentation « des documents d'un corpus ». Il doit donc permettre de décrire les concepts du document, en vue de l'indexer. Cette composante du langage documentaire représenterait dans un modèle de recherche d'information le « modèle de documents, qui correspond à la modélisation du contenu sémantique des documents, dans le formalisme de représentation de connaissances » [Géry, 2002]. Cependant ce n'est là qu'une partie de l'objectif du langage documentaire. Ce dernier doit aussi représenter « les questions qui intéressent un groupe d'utilisateurs ». C'est le « modèle de requête » du modèle de recherche d'information. Il « correspond à la modélisation du besoin d'information de l'utilisateur, dans le formalisme de représentation de connaissances » [Géry, 2002]. Enfin, le langage documentaire, pour arriver à un « repérage simple des documents du corpus qui répondent aux questions de ces usagers », doit permettre la définition d'une « fonction de correspondance », qui fera le lien entre les aspects indexation et requête. « Le système évalue la pertinence (la valeur de correspondance) des documents par rapport à la requête. La fonction de correspondance est un élément clé d'un [système de recherche d'information], car la qualité des résultats dépend de l'aptitude du système à calculer une pertinence des documents la plus proche possible du jugement de l'utilisateur » [Géry, 2002]¹⁵.

15. En suivant à la lettre la définition de Géry, il faudrait pour décrire un modèle de recherche d'information,

a) Définition du langage documentaire

La définition du langage documentaire n'est pas *stricto sensu* une sous-partie du processus d'indexation. La définition de Lefèvre (définition 1.3) fait du langage documentaire un outil de l'indexation, puisque l'analyse des documents en dépend étroitement, mais elle permet aussi de replacer l'indexation dans un contexte plus global ne perdant pas de vue son objectif : rendre accessible des documents à des usagers en fonction de leurs besoins d'informations. D'un point de vue global, la définition, aussi bien informatique que conceptuelle, d'un langage documentaire pour notre système, nous permet d'élaborer le modèle de recherche d'information dont il dépendra. Ce travail a des implications liées aux utilisateurs du système, les enseignants, que nous explorerons dans la suite de ce chapitre, mais aussi des aspects plus purement informatiques qu'il convient de traiter ici.

Comme dans tout travail de modélisation informatique, nous allons au cours de la définition du langage documentaire devoir coder de l'information analogique, issue du monde réel – relevant dans notre cas de la problématique de l'enseignement des langues – sous forme numérique.

« En général, n'importe quel type d'information ou de message, à condition qu'il soit explicitable ou mesurable, peut être traduit numériquement. »

[Lévy, 1997, p. 60]

C'est là, un point critique de notre problématique, il s'agit, pour pouvoir coder l'information, d'explicitier les caractéristiques des ressources (les textes) du point de vue des usagers du système (l'utilisation des textes dans l'enseignement des langues). Lévy explique dans la même page qu'au cours de ce processus de « numérisation », il est possible qu'une perte d'information ait lieu.

Les exemples de perte d'information lors de codage numérique dans le contexte d'une utilisation grand public de l'outil informatique sont légions : codage du son sur un CD par rapport à une bande master, transfert du CD en mp3, stockage d'une image d'un format RAW vers un format jpg, etc. Ces pertes d'information entre les données analogiques (continues) et les données numérique (discrètes) ou d'un format numérique à un autre, si elles sont maîtrisées, dépendent de l'usage qui doit être fait des ressources. Prenons l'exemple d'une affiche de film, selon qu'elle doit être stockée en vue d'une impression de haute qualité en 120cm par 160cm par un professionnel de la photographie ou affichée via un navigateur Web sur le moniteur du lecteur d'un blog de cinéphile, les besoins sont radicalement différents. La perte d'information entre l'un et l'autre dépend de ces besoins.

Dans un contexte légèrement différent, celui de la transmission d'une valeur continue, dont la nature nous importe peu ici, Feynman pointe en une phrase un point clé de la problématique de la numérisation : « Tout le secret de la transmission de la valeur S réside dans l'approximation que nous pouvons en faire » [Feynman, 2006, p. 178]. De la même manière que le son d'un CD est un approximation de celui de la bande master qui a servi à le créer, que le même son au format mp3 est une approximation de celui du CD ou que l'image RAW est une approximation du point

introduire la notion de base de connaissances : « un thésaurus, composé de concepts apparaissant dans le corpus, reliés entre eux par diverses relations (spécificité/généricité, synonymie, voir-aussi, etc.). En considérant par exemple les relations de synonymie entre les concepts, il est ainsi possible de retrouver, pour une requête composée du terme "voiture", des documents traitant de voiture ou d'automobile » [Géry, 2002]. Nous considérerons cependant cette base de connaissance, comme une composante de la fonction de correspondance, afin de simplifier la définition de modèle de recherche d'information. Ce faisant, nous négligeons certaines relations comme la spécificité/généricité, mais ces dernières supposent une ouverture du modèle de requête (vocabulaire non contrôlé), qui va à notre avis au delà du concept de système de recherche d'information au sens large du terme. Il n'est pas encore question de ce degré de complexité à ce point du travail.

de vue du photographe ayant pris le cliché, nous allons devoir nous contenter, pour décrire un texte, d'effectuer une approximation des caractéristiques qui peuvent intéresser un enseignant en vue de son utilisation dans le cadre de son cours de langue. La précision de cette approximation dépendra de la finesse avec laquelle nous pourrions isoler les informations qui entrent en jeu dans le choix d'un texte en vue de son utilisation en classe de langue et leur importance respective dans ce processus. La place centrale de l'« usager » dans la définition de langage documentaire rejoint la problématique de la numérisation, en ce sens que la précision de notre indexation dépend directement de l'approximation que nous ferons de la problématique du choix d'un texte et transitivement de l'explicitation qui pourra être faite des critères impliqués dans le choix de l'utilisateur et des besoins de ce dernier dans le cadre de l'utilisation d'une base de textes destinée aux enseignants de langue.

1.3.3 « Pédagogique »

En résumé, l'indexation revient donc à « décrire un document à l'aide de représentations des concepts contenus dans ce document », suivant un langage destiné à « fournir une représentation formalisée et univoque des documents d'un corpus et des questions des usagers » du système. La définition de langage documentaire de Lefèvre met en exergue la position centrale des utilisateurs du système dans le processus d'indexation. Il ne s'agit pas seulement de pouvoir représenter formellement chaque document de manière à pouvoir les différencier, mais aussi que le formalisme utilisé puisse représenter les questions des utilisateurs et permettre d'accéder aux objets qui répondent à ces questions. Dans notre cas, les objets sont des textes et les questions des usagers traitent de l'utilisation de ces textes dans l'enseignement des langues. Notre travail relève de ce que nous avons appelé l'« indexation pédagogique » [Loiseau *et al.*, 2005].

Dans le cadre de notre travail, lorsque nous parlons d'indexation pédagogique, nous nous référons à :

Définition 1.4 (Indexation pédagogique) *Indexation effectuée selon un langage documentaire permettant à l'utilisateur de rechercher des objets pour les utiliser dans un enseignement.*

Le terme « pédagogique », qualifiant l'indexation, est légèrement approximatif du point de vue de la littérature. Il est difficile de définir contrastivement didactique et pédagogie et nous ne nous y hasarderons pas, d'autant plus que ce n'est pas là l'objectif de notre travail. Pour ce choix terminologique, nous nous sommes appuyé sur le travail effectué par [Rézeau, 2001, chapitre 1, §2 & 3]. La plupart des travaux qu'il cite visent à décrire la « situation pédagogique » [Houssaye, 1988, p. 40] et l'influence qu'exercent les unes sur les autres, les entités¹⁶ qui la composent. Rézeau, combine le triangle de [Houssaye, 1988, p. 41] avec le modèle systémique d'Altet ([Rézeau, 2001, p. 42]) pour obtenir « un premier modèle de la situation d'enseignement-apprentissage », dont nous reprenons le schéma (cf. figure 1.3).

Le processus d'indexation est, à travers la notion de langage documentaire, centré sur l'utilisateur, dans notre cas l'enseignant. Dans la figure 1.3, nous avons donc mis en exergue ses relations avec les autres pôles : sa relation avec le savoir relèverait de la didactique alors que celle avec l'apprenant relèverait de la pédagogie. Ainsi l'objectif d'une indexation « effectuée selon un langage documentaire permettant à l'utilisateur de rechercher des objets pour les utiliser dans

16. Chez Houssaye ces entités sont l'enseignant, l'apprenant et le savoir et sont appelées « pôles »

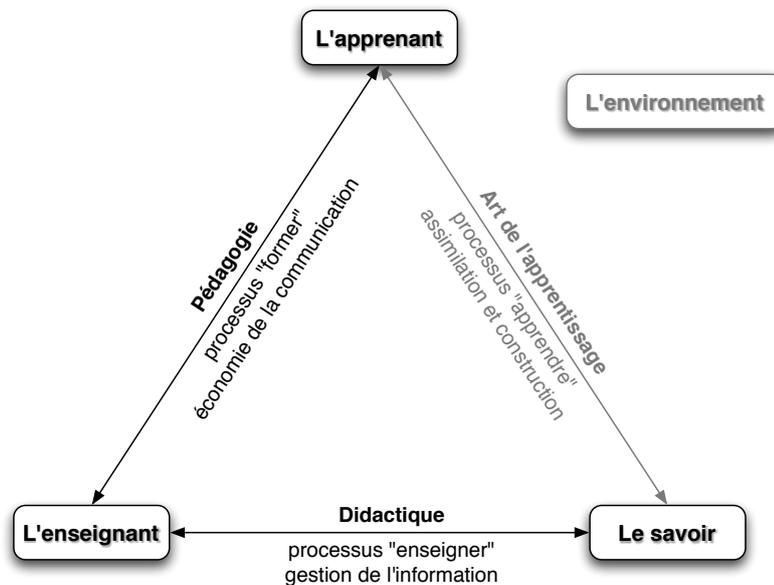


FIGURE 1.3: Un modèle de la situation d'enseignement-apprentissage [Rézeau, 2001, Figure 1.5 p. 45]

un enseignement¹⁷ » nous semble s'inscrire dans l'axe pédagogique de la figure 1.3. Dans notre cas, il s'agit de proposer aux enseignants (utilisateurs¹⁸) un système qui leur permette d'avoir accès à des textes (objets¹⁸) qu'ils pourront utiliser comme support dans le cadre de leurs cours de langue (questions des utilisateurs¹⁸).

Nos objectifs se différencient par exemple de ceux de [Tsui, 2004]. Elle a recours à des corpus comme le *British National Corpus* (BNC) ou le « *South China Morning Post*¹⁹ », qui sont interfacés avec des outils de recherche (*XML Aware Indexing and Retrieval Architecture* (XAIRA) pour le BNC / un « *Pattern Finder* » pour le second). Ces outils indexent les corpus afin de pouvoir effectuer des recherches sur leur contenu linguistique. Tsui propose de les utiliser avec des objectifs différents des nôtres : dans le cadre de la relation de l'enseignant au savoir, un savoir qu'un locuteur non-natif n'a pas nécessairement. Elle distingue dans son article trois types de situation [Tsui, 2004, p. 40] :

- usages différents de synonymes selon le contexte, que les enseignants ont des difficultés à expliquer ;
- preuves linguistiques qui entrent en contradiction avec la grammaire prescriptive telle qu'elle leur a été enseignée ;
- et, enfin, des collocations que les enseignants essaient de rationaliser.

En ce sens, l'article de Tsui propose une méthode d'accès au savoir linguistique pour les enseignants et ne pose pas la question de la transmission de ce savoir aux apprenants ni de l'utilisation éventuelle des corpus dans le contexte de la classe. La problématique traitée est résolument tournée vers ce que Chevallard appelle, au départ dans le cadre de la didactique des mathématiques, la « transposition didactique », qui « renvoie au passage du savoir savant au savoir enseigné » [Chevallard, 1982]. Dans le cas de Tsui, le savoir savant se rapporte à la L2 enseignée et l'utilisation

17. cf. définition 1.4

18. dans la définition 1.3 (Langage Documentaire) de Lefèvre

19. Issu du projet *TeleCorpora* (<http://www.telenex.hku.hk>)

des corpus doit permettre de formaliser certains aspects de la langue en vue de les transmettre.

Dans notre cas, l'indexation doit permettre à l'enseignant d'accéder à des textes utilisables dans le contexte de l'enseignement des langues, que ce soit par l'intermédiaire d'une plateforme de type MIRTO ou en cours présentiel. Nous pouvons considérer que si la décision de recourir à un texte relèverait plutôt de la « didactisation », le choix du texte lui-même concerne la relation de l'enseignant à l'apprenant, c'est le choix d'un support pour la constitution du savoir de ce dernier. Dans le paradigme défini ci-dessus, l'objectif de la recherche de l'enseignant relève de la pédagogie.

Le terme d'« indexation pédagogique » désignerait donc une indexation à objectif pédagogique. Nous introduisons ce terme du fait qu'il symbolise notre problématique, mais nous ne l'opposons pas à une éventuelle notion d'indexation « didactique », d'une part car le point de vue utilisé pour arriver à ce terme ne décrit que partiellement la situation d'enseignement-apprentissage et d'autre part car la séparation entre didactique et pédagogie n'est pas claire et établie²⁰. Pour Dabène, par exemple :

« La Didactique des Langues Étrangères se définit [...] par le va-et-vient permanent entre les réalités du terrain pédagogique (la classe de langue) et les apports de la réflexion théorique dans les domaines scientifiques concernés. »

[Dabène, 1989] citée par [Rézeau, 2001]

Selon cette remarque, notre travail de définition des critères qui peuvent être utiles à l'enseignant pour le choix d'un texte relève de la didactique des langues, alors que l'objectif de l'utilisation des textes obtenus par le biais de notre système est plus proche de la « réalité du terrain pédagogique », une composante du travail didactique.

a) Différentes indexations pédagogiques

Revenons maintenant à la définition de « langage documentaire » de Lefèvre (définition 1.3). L'introduction de la notion d'« indexation pédagogique », nous a permis de préciser quelque peu le rôle des utilisateurs de notre système. Il s'agit pour eux de récupérer des objets qu'ils pourront utiliser dans le cadre de leurs cours (en ligne ou en présentiel). Nous avons, dans la présentation du projet MIRTO et la description du cadre de nos recherches, indiqué que les objets étaient des textes. La problématique de notre travail relève donc de l'indexation pédagogique de textes. Il faut cependant préciser encore cette notion, pour qu'elle décrive adéquatement notre problématique. Notre travail revêt un aspect didactique, or toujours d'après Rézeau, « la didactique relève d'une recherche disciplinaire » [Rézeau, 2001, p. 52]. Ce point de vue pourrait être contredit, c'est d'ailleurs le cas dans le cadre d'un débat qui oppose didactique(s) disciplinaire(s) et didactique générale (en passant par didactiques spécifiques et spécialisées) [Laurin & Gaudreau, 2001, p. 17]. Cependant, l'existence de didactiques disciplinaires n'est pas contestable. Ce qui ne surprendra personne, puisque, dans la pratique, tout enseignement est lié à une discipline [Ramsden, 1992, p. 110]. Et si la didactique des langues peut s'alimenter de points de vues extérieurs²¹ elle est en même temps un champ disciplinaire à part entière. Il semble de plus tout à fait cohérent que,

20. Voir [Jonnaert & Laurin, 2001a, p. 4] pour une position du problème précise et concise, qui met en exergue le fait que la question d'une séparation entre didactique et pédagogie incite à examiner les liens qui peuvent – doivent ? – exister entre les deux.

21. Dans le paragraphe suivant celui que nous citons, Rézeau introduit d'ailleurs par des concepts issus de didactiques des mathématiques.

même si plusieurs disciplines peuvent s'intéresser aux textes, le point de vue d'un enseignant d'histoire ne sera pas le même que celui d'un enseignant de biologie ou enfin d'un enseignant de langue. Les « questions qui intéressent [le] groupe d'utilisateurs » - des enseignants - dépendent du sous-groupe disciplinaire. Notre travail est donc un travail d'indexation pédagogique de textes pour l'enseignement des langues, en opposition à une indexation pédagogique de textes pour l'enseignement de la sociologie ou de l'économie. Au sein de la didactique des langues, plusieurs approches et méthodes coexistent, chacune constituant potentiellement un point d'entrée pour une « indexation pédagogique de textes pour l'enseignement des langues ».

1.4 Les textes dans l'enseignement des langues

Une méthode particulière peut constituer un point d'entrée pour notre problématique, mais ce n'est pas l'unique objectif du rapide travail sur les principales méthodes d'enseignement des langues effectué dans ce paragraphe. Il devrait également permettre de se convaincre du fait que notre projet ne relève pas uniquement d'un usage par l'intermédiaire d'une plateforme d'apprentissage. Il ne s'agit pas ici d'effectuer un état de l'art exhaustif, mais de constater certaines préconisations en matière d'utilisation de textes dans l'enseignement des langues. Nous nous appuyons donc pour cela, sur un ouvrage généraliste [Germain, 1993] et sur une étude issue de [Levy, 1997]. Le premier ouvrage dresse un panorama des méthodes employées, ce qui est suffisant pour la question que nous nous posons : celle de l'usage des textes dans l'enseignement des langues selon la méthode. Enfin, l'étude issue du second ouvrage n'est pas aussi générale que nous l'aurions souhaité, puisqu'elle concerne principalement des enseignants de langue auteurs de ressources d'ALAO et enseignant dans des pays anglophones. Il faut donc bien garder en tête que les statistiques d'usage des différentes méthodes ne sont pas nécessairement transposables à toute population d'enseignants de langue.

1.4.1 L'approche communicative

Selon l'étude de [Levy, 1997], la méthode la plus employée est l'approche communicative (plus de 70% des enseignants y ont recours). Dans l'approche communicative, « la langue est vue avant tout comme un instrument de communication ou mieux comme un instrument d'interaction sociale. » [Germain, 1993]. Cette approche considère que la « connaissance des règles, du vocabulaire et des structures grammaticales est une condition nécessaire mais pas suffisantes à la communication » [Germain, 1993]. Elle s'appuie sur des travaux de sociolinguistique comme ceux de Hymes [1984] qui « [en établissant] un rapport entre la notion de communauté linguistique et celle de compétence, [doit] préciser que les membres d'une communauté linguistique ont en partage une compétence de deux types, un savoir linguistique et un savoir sociolinguistique ou, en d'autres termes, une connaissance conjuguée de normes de grammaire et de normes d'emploi ».

Sophie Moirand revient sur la notion de compétence de communication et la définit de la manière suivante :

« Une compétence de communication reposerait, pour moi, sur la combinaison de plusieurs composantes :

- une composante linguistique, c'est-à-dire la connaissance et l'appropriation (la capacité de les utiliser) des modèles phonétiques, lexicaux, grammaticaux et textuels du système de la langue ;

- une composante discursive, c'est-à-dire la connaissance et l'appropriation des différents types de discours et de leur organisation en fonction des paramètres de la situation de communication dans laquelle ils sont produits et interprétés ;
- une composante référentielle, c'est-à-dire la connaissance des domaines d'expérience et des objets du monde et de leurs relations ;
- une composante socioculturelle, c'est-à-dire la connaissance et l'appropriation des règles sociales et des normes d'interaction entre les individus et les institutions, la connaissance de l'histoire culturelle et des relations entre les objets sociaux.

»

[Moirand, 1982]

Pour faire acquérir aux apprenants ces compétences, « une des caractéristiques de l'approche communicative est que l'enseignant recourt, dans la mesure du possible à des documents dits authentiques, c'est-à-dire non expressément conçus pour être utilisés dans une classe de langue seconde » [Germain, 1993]. Le texte authentique occupe ainsi une place centrale dans l'approche communicative [Chatzissavidis, 2003]. En effet, il est perçu comme une source de motivation, un facteur d'autonomisation de l'apprenant vis-à-vis de son apprentissage et, enfin, il offre la possibilité de travailler sur « la mise en relation des énoncés produits avec les conditions de production » (aspects linguistiques, pragmatiques et usages sociaux)[Bérard, 1991, pp. 50–51].

1.4.2 *Task-based approach*

La deuxième méthode la plus utilisée par les enseignants et auteurs d'activités ou systèmes d'ALAO de l'étude [Levy, 1997] est la *task-based approach* (plus de 55% des réponses). En français, elle est appelée « perspective actionnelle ». C'est l'approche didactique retenue par le Cadre Européen Commun de Référence pour les Langues (CECR) [Riba, 2006], dans laquelle « on se propose de former un “acteur social” ; ce qui impliquera nécessairement [...] de le faire agir avec les autres pendant le temps de son apprentissage en lui proposant des occasions de “co-actions” dans le sens d'actions communes à finalité collective » [Puren, 2002]. « À l'approche communicative basée sur la méthode de l'interaction correspondrait une perspective interculturelle, et à l'approche actionnelle correspondra une perspective co-culturelle, basée sur le “faire-ensemble” où l'on ne se contente pas d'assumer nos différences pour créer ensemble des ressemblances » [Riba, 2006].

La perspective actionnelle s'articule autour de la notion de « tâche », dont nombreuses définitions coexistent (cf. [Nunan, 2006]). Pour Puren, une approche critique de la didactique des langues impose de faire la différence entre la « description d'un concept dans l'actualité et sa définition » [Puren, 2006, ≈ 9'15"]. Il en propose donc une définition relativement abstraite, selon laquelle une tâche est une « unité de sens dans l'activité d'enseignement et d'apprentissage » [Puren, 2006, ≈ 9'40"]. Cependant ce vers quoi nous voulons tendre ici, c'est bien la description du concept de tâche, dans le contexte de la perspective actionnelle.

L'approche d'Oxford [2006] n'est pas fondamentalement différente de celle Puren : elle passe en revue les différentes acceptions de la notion de tâche à travers le temps²² et conclut cette partie par définir le sens de tâche, dans le contexte de l'enseignement et l'apprentissage des langues : selon elle, une *tâche* est « une séquence éducative orientée résultat ou un cadre d'étude comportemental pour la recherche ou l'apprentissage présentiel²³ » [Oxford, 2006]. À notre avis

22. À travers le temps mais aussi les contextes : elle ne se borne pas à l'enseignement des langues.

23. *an outcome-oriented instructional segment or [a] behavioral framework for research or classroom learning*

cette définition très concise met bien en exergue, dans sa première partie (orientée vers le résultat) la différence que fait Puren entre une tâche dans l'approche communicative et dans la perspective actionnelle : dans la première, la tâche s'arrête quand la communication a eu lieu (acte de parole), alors que dans la seconde la tâche se termine avec la réalisation d'une action sociale [Puren, 2006].

Cette perspective s'appuie elle aussi, sur des supports – notamment textuels – (*input*) principalement authentiques :

- « D'un point de vue pédagogique, la perspective actionnelle a renforcé les principes et pratiques suivants :[...] L'introduction de textes authentiques dans la situation d'apprentissage [...] ²⁴ » [Nunan, 2006] ;
- « En tout cas, le choix des données langagières, aussi bien des enregistrements de langue parlée que des textes écrits, est d'une importance vitale. Les concepteurs de cours devraient tenter de choisir un ensemble représentatif de textes cibles types à partir de sources d'échantillons accessibles issus de la vie réelle. Ces échantillons devraient représenter les traits langagiers typiques des genres issus des communautés discursives présentes ou futures des apprenants ²⁵ » [Willis, 2000] ;
- « Après avoir choisi une thématique, il faut rassembler un certain nombre de textes sur le thème, textes dont la lecture et les informations qu'ils contiennent amènent l'enseignant à définir une mission dans ses grandes lignes » [Bourguignon, 2007].

1.4.3 Autres approches et méthodes

Nous avons présenté dans le détail les deux approches proposées ci-dessus, du fait de leur importance dans l'enseignement des langues à l'heure actuelle. Encore une fois, notre objectif ici n'est pas de dresser un état de l'art des méthodes d'enseignement des langues. Nous nous proposons donc de résumer le reste des données issues de [Levy, 1997] dans un tableau récapitulatif (tableau 1.1). Nous sommes bien conscient de l'absurdité à résumer tout un courant de pensée didactique en quelques lignes, mais nous nous en contenterons dans la mesure où les méthodes ne sont pas, en elles-mêmes, essentielles à notre argumentation.

Dans le tableau 1.1, nous regroupons sous le terme « méthode » aussi bien la notion de méthode que celle d'approche ou de perspective. Nous décrivons extrêmement succinctement les principes de la méthode et la place du texte dans les pratiques. Quant à l'« usage » ce sont les statistiques issues de [Levy, 1997].

MÉTHODE	PRINCIPES	USAGE
Approche communicative	cf. paragraphe correspondant page 18	>70%
Perspective actionnelle	cf. paragraphe correspondant page 19	>55%
<i>Formal Grammar</i>	cf. NB 1.3 p. 24	>30%

24. *Pedagogically, task-based language teaching has strengthened the following principles and practices. [...] The introduction of authentic texts into the learning situation.*

25. *In all cases, the choice of language data, both recordings of spoken language and written texts, is of vital pedagogical importance. Course designers should aim to choose a representative set of target text types from accessible real life sources-samples that reflect the typical language features of genres from the learners' present or future discourse communities.*

MÉTHODE	PRINCIPES	USAGE
Approche fonctionnelle-notionnelle	« L'approche dite fonctionnelle-notionnelle [...] renvoie à un principe d'organisation des programmes d'apprentissages des langues vivantes [...] où la priorité est donnée à la valeur communicative des éléments du langage plutôt qu'à leur valeur grammaticale et formelle. L'analyse des besoins langagiers permet de déterminer ce qui est nécessaire aux apprenants en termes de fonction du langage et d'actes de paroles, et les notions générales et spécifiques que l'apprenant devra maîtriser. » [Cuq, 2003, article "notionnel" p. 179]	≈30%
	SUPPORT TEXTUEL texte authentique cf. NB 1.4 p. 25.	
Méthode situationnelle	Dans la méthode situationnelle, l'oral est privilégié et l'écrit est dérivé de l'oral et introduit après l'oral. La langue est vue principalement comme un phénomène oral dont le cœur serait la « structure syntaxique ». L'enseignement vise à créer des automatismes (théorie béhavioriste). La notion de « situation » n'est pas un critère de sélection du contenu, mais un mode de pratique des structures orales.[Germain, 1993, p. 191–195]	>20%
	SUPPORT TEXTUEL « Le matériel est non seulement de nature livresque [mais aussi visuel]. L'élément visuel occupe une place importante. » « Le choix du contenu est de nature linguistique. » [Germain, 1993, p. 193]	
<i>Cognitive code</i>	Il s'agit d'une version améliorée de la méthode grammaire-traduction, selon laquelle l'apprentissage d'une langue est un processus d'acquisition du contrôle conscient des structures phonologiques, grammaticales et lexicales de la L2[Carroll, 1965, p. 278]. Cette acquisition s'opère par l'étude et l'analyse de ces patrons par des exercices et des explications (avant l'usage des structures) [Chastain & Woerdehoff, 1968, p. 269].	>15%
	SUPPORT TEXTUEL cf. Méthode grammaire-traduction.	
Méthode Orale	La méthode orale date du début du 20e siècle et, comme la méthode directe, s'inscrit en réaction à la grammaire-traduction. Mais la méthode orale va plus loin et proscrit l'écrit jusqu'à ce que l'apprenant acquière la capacité à reproduire couramment ce qu'il entend [Palmer, 1922, p. 24]. Les quatre aptitudes sont considérées comme dépendant de la quantité de « discours mémorisé » [Palmer, 1922, p. 23]. Cependant, d'après Palmer, il ne s'agit pas d'une « méthode complète », à laquelle il faut associer une autre méthode pour le développement des aptitudes écrites [Palmer, 1922, p. 24]. (cf. éclectisme actuellement, § 1.4.5 p. 24)	>15%

MÉTHODE	PRINCIPES		USAGE
	SUPPORT TEXTUEL	La méthode orale n'utilise pour l'essentiel pas de textes, les supports sont essentiellement des tables de substitutions, qui ne sont pas censées être fournies aux apprenants avant d'avoir été mémorisées et assimilées de manière approfondie [Lemieux, 1964, p. 323].	
Méthode Directe	« La méthode directe s'inscrit dans la foulée des mouvements ou initiatives de réformes contre la grammaire-traduction [...] orientée vers l'écrit littéraire. » Elle vise à apprendre à penser en L2 et donne la priorité à l'oral. L'enseignant démontre à l'aide de supports visuels, il ne traduit ni n'explique un contenu choisi en fonction de situations (ex : à la poste) : la L1 n'est pas utilisée en classe [Germain, 1993, p. 127–131].		≈15%
	SUPPORT TEXTUEL	L'aptitude à lire est développée après l'aptitude à parler par des activités de lecture, par exemple à haute voix. La méthode directe a contribué à populariser l'usage du texte narratif [Germain, 1993, p. 127–131].	
Méthode Grammaire-traduction	Historiquement la méthode grammaire-traduction servait principalement pour enseigner les langues classiques comme le latin ou le grec ancien. Elle considère la L2 comme un ensemble de règles et de vocabulaire. Elle s'appuie principalement sur des exercices de thème et de version. L'un des objectifs principaux est que l'apprenant puisse lire des ouvrages écrits en L2 [Germain, 1993, p. 101–107].		≈10%
	SUPPORT TEXTUEL	L'un des rôles de l'enseignant est de choisir les textes (littéraires surtout), qui, lorsqu'ils constituent le point de départ d'une leçon, jouent un rôle primordial [Germain, 1993, p. 104].	
Approche Naturelle	La méthode naturelle est fondée sur le processus d' <i>acquisition</i> d'une L2 par des adultes en milieu naturel [Germain, 1993, p. 243]. Elle peut être décrite selon 4 principes : la classe est dédiée en premier lieu à fournir des apports (selon la traduction de <i>input</i> de Cuq [2003]) en vue de l'acquisition ; l'enseignant ne parle que la L2 et les apprenants peuvent utiliser leur L1 ou la L2 (sans correction sauf en cas d'échec de la communication) ; les devoirs peuvent inclure de la grammaire formelle (NB 1.3 p. 24) et faire l'objet de corrections ; les activités peuvent être structurées, mais l'objectif est de permettre aux apprenants d'échanger des idées, effectuer des tâches et résoudre des problèmes [Krashen, 1982, p. 138].		≈10%

MÉTHODE	PRINCIPES		USAGE
	SUPPORT TEXTUEL	Outre les supports visuels (cf. méthode directe), l'approche naturelle prône l'usage de matériel authentique y compris textuel [Germain, 1993, p. 251]. Pour les niveaux plus avancés (quand les apprenants « communiquent »), l'enseignant peut orienter le choix des supports, en particulier des textes, pour l'acquisition de la grammaire [Terrell, 1977, p. 327].	
Méthode Audio-orale	La méthode audio-orale « constituait un mélange de la psychologie béhavioriste et du structuralisme linguistique qui a largement influencé l'enseignement de la grammaire grâce aux “pattern drills” ou “cadres syntaxiques”. » [Rodríguez, 2001] Elle est assez proche de la méthode directe, et comme pour elle, la L1 ne devrait pas être utilisée en classe. Elle est cependant plus orientée vers la grammaire alors que la méthode directe est plus centrée sur le vocabulaire. La langue est vue comme un ensemble de cadres syntaxiques acquis comme un ensemble d'automatismes linguistiques [Germain, 1993]		>5%
	SUPPORT TEXTUEL	Les structures sont présentées sous forme de dialogues répétés par les apprenants. Le contenu est choisi en fonction de la proximité avec la L1, du plus proche au plus éloigné (structures / phonétique) Germain [1993]. L'enseignant peut avoir recours à des textes pour des exercices de dictée ²⁶ .	
Autres	Nous avons regroupé ici les méthodes utilisées par moins de d'enseignants que la méthode Audio-orale (« <i>C.L.L.</i> , <i>T.P.R.</i> , <i>S.G.A.V.</i> , <i>Suggestopedia</i> , <i>Silent Way</i> ») et la catégorie « <i>Other</i> » (représentant à elle seule environ 20% des enseignants).		>35%

TABLE 1.1: Les différentes méthodes et approches utilisées par les enseignants et concepteurs de ressources ALAO d'après [Levy, 1997, p. 123]

1.4.4 Le texte omniprésent

À travers cette rapide étude, nous pouvons constater que le texte en tant que support pédagogique est présent dans quasiment toutes les méthodes les plus utilisées par les enseignants. Cette rapide étude n'a pas la prétention d'être exhaustive : comme nous l'avons fait remarquer la population concernée par le questionnaire travaille principalement en contexte anglophone et concerne avant tout des concepteurs d'applications d'ALAO, qui, pour la grande majorité (97,1%), enseignent les langues [Levy, 1997, p. 120]. Ces conclusions ne sont donc pas nécessairement généralisables

26. http://www.saskschools.ca/curr_content/hutt/esl/almstrat.htm

Nota Bene 1.3 Grammaire formelle

Le questionnaire de Levy propose « *Formal Grammar* » parmi les méthodes. Il ne s'agit pas ici d'une méthode proprement dite, mais plutôt de la manière de traiter l'enseignement du savoir grammatical. Sa présence dans une liste de méthodes nous paraît peu appropriée dans la mesure où il s'agit plus d'une composante de méthode. « *Formal Grammar* » signifie ici : avoir recours à des règles de grammaire formelles pour l'enseigner. Certaines méthodes, comme la méthode « Grammaire traduction » ou la méthode « audio-orale », prônent l'enseignement de la grammaire de manière explicite, d'autres imposent un apprentissage de la grammaire purement inductif (méthode « par le mouvement » par exemple). Compte-tenu de la coexistence dans les pratiques des enseignants (cf. § 1.4.5 p. 24) de diverses méthodes, la problématique du recours à la « grammaire formelle » reste la même que celle énoncée par Mossel : « [...] le grand problème pour le pédagogue est de savoir ce qui, dans la grammaire formelle, peut vraiment servir à l'acquisition de cette maîtrise et ce qui est plutôt superflu, nuisible. » [Mossel, 1960]. Cette notion de recours à la « grammaire formelle » aurait à notre avis dû figurer dans une question différente et non être présentée comme une méthode ou une approche à part entière de l'enseignement des langues, en ce sens qu'elle n'en est qu'une composante.

aux enseignants de langues qui n'utilisent pas l'ALAO. En outre, nous nous appuyons ici sur des constatations issues d'analyses des méthodes et non sur des travaux d'observation d'enseignants utilisant les méthodes en question en classe. Malgré ces réserves, nous constatons que toutes les méthodes que nous avons détaillées utilisent des textes comme support ou apport. Si pour une population d'enseignants différente, les proportions d'utilisation des méthodes pourraient changer, les méthodes employées resteraient les mêmes et cette omniprésence du texte ne serait pas altérée. Les méthodes avant tout orientées vers l'oral, où l'enseignant n'utilise que la L2 ou dans lesquelles l'écrit est dérivé de l'oral, font appel aux textes plus tard dans l'apprentissage, mais y font toujours référence (méthode situationnelle, méthode directe, approche naturelle). Enfin les méthodes ou approches les plus populaires font toutes la part belle au texte authentique (approche communicative, perspective actionnelle et approche notionnelle-fonctionnelle). L'importance du texte, qui est évidente dès lors qu'il s'agit de transmettre l'aptitude de compréhension écrite (lire), est aussi constatée dans l'acquisition des autres aptitudes notamment comme support ou apport, par exemple dans le cadre d'un enseignement inductif de la grammaire. Comme le suggère notre étude préliminaire, l'utilisation de textes semble inévitable dans l'enseignement des langues, justifiant par là même l'existence de notre projet, même dans un contexte d'enseignement uniquement présentiel.

1.4.5 Positionnement didactique

L'importance du texte dans l'enseignement des langues n'est pas la seule conclusion que nous pouvons tirer du tableau 1.1. La somme des pourcentages approximatifs de la colonne « usage » est supérieure à 310%, soit une moyenne supérieure à 3 méthodes employées par personne. Selon cette étude ([Levy, 1997]), les enseignants ne construisent pas l'intégralité de leurs cours selon une méthode unique, mais en combinent plusieurs. À partir de cette constatation, il devient difficile de choisir de définir notre système d'indexation en prenant comme point d'entrée une méthode particulière. Le choix évident entre la méthode la plus utilisée (approche communicative) et celle qui est préconisée par le conseil de l'Europe (perspective actionnelle) perd de sa pertinence : les informations à notre disposition nous indiquent pour chaque méthode combien d'enseignants l'utilisent, mais pas dans quelles conditions ou en quelles proportions. Quand un enseignant

Nota Bene 1.4 Approche notionnelle-fonctionnelle et approche communicative

Certains auteurs identifient l'approche notionnelle-fonctionnelle à l'approche communicative ou à ses prémisses ([Germain, 1993, p. 202], faisant référence à [Wilkins, 1976], ou Martinez [1996] cité par [Rézeau, 2001, p. 94]). En effet, sans développer explicitement la notion de compétence de communication, l'approche notionnelle-fonctionnelle introduit des composantes pragmatiques et sociolinguistiques à l'élaboration des cours, par « le passage d'un programme grammatical à un programme situationnel et, puisque l'apprenant fait partie de la situation d'apprentissage au moins autant que la langue, à un programme centré sur les besoins de l'apprenant » [Rézeau, 2001, p. 95]. Cependant d'autres auteurs, comme Springer [Springer, 1996, p. 161] et Rézeau [Rézeau, 2001, p. 95] (citant Springer), considèrent que les deux approches n'évoluent pas sur le même plan. L'enseignement fonctionnel opérant un renversement didactique et l'approche communicative un renversement pédagogique. L'approche notionnelle focalisant sur la préparation des cours est « critiquée [notamment] pour son désintérêt pour ce qui se passe en salle de classe » [Springer, 1996, p. 161].

La nécessité d'utilisation de supports authentiques mise en exergue dans l'approche communicative est aussi présente dans l'approche notionnelle fonctionnelle :

« L'approche notionnelle-fonctionnelle, parce qu'elle vise les pratiques langagières, doit tenir compte [...] des activités de compréhension de messages authentiques, variés et ordinaires, c'est-à-dire non purifiés ou normalisés. L'intérêt du document authentique réside dans la possibilité de prise de conscience de l'apprenant quant à l'écart entre la langue idéalisée et artificielle de l'enseignement et les représentations imparfaites de la communication quotidienne » [Springer, 1996, p. 92] (citant [Wilkins, 1976, p. 79])

déclare utiliser plusieurs méthodes, les utilise-t-il toutes au sein d'un même cours, avec le même groupe d'apprenants, en fonction des objectifs de chaque séquence? Si tel est le cas, à quelle méthode l'enseignant se réfère-t-il lorsqu'il décide d'utiliser ou lorsqu'il choisit un texte? Mais le choix de la méthode pourrait aussi dépendre du groupe d'apprenants : son niveau, ses objectifs. Auquel cas, choisir une méthode, reviendrait à choisir un public. De plus, nous savons finalement peu de choses du processus de recherche et de choix d'un texte, à quel point est-il dépendant d'une approche? En outre, les enseignants interrogés se réfèrent à plusieurs méthodes, peut être n'intègrent-ils que certains éléments (conception de la langue, rôle de l'enseignant, de l'apprenant, etc.) de chaque méthode utilisée, sans que cela soit les mêmes aspects qui soient considérés par chaque enseignant. Enfin, quand nous constatons l'absence de consensus chez les didacticiens dans leurs conceptions des liens entre plusieurs méthodes (cf. NB 1.4 p. 25), pouvons-nous considérer les pratiques liées à une méthode particulière comme homogènes entre les enseignants?

Cette constatation de la multiplicité des méthodes utilisées n'est pas nouvelle, elle est désignée dans [Puren, 1991], sous le terme d'« éclectisme ». Pour Puren, l'éclectisme ne doit pas être un « éclectisme mou » du « tout se vaut » [Puren, 1991, p. 46]. Beacco définit la notion de « méthodologie circulaire », « sur laquelle les méthodologies constituées d'origine savante, c'est-à-dire non directement issues des pratiques de classe des enseignants, viennent se greffer ou se diluer » [Beacco, 1995, p. 44]. Elle semble s'apparenter à cet « éclectisme mou ». Pour lui, un « choix éclectique de pratiques de classe signifie sélection raisonnée et non ensemble hétéroclite de techniques d'enseignement » [Beacco, 1995, p. 43]. Spielmann, quant à lui, semble penser que l'« approche éclectique » des enseignants relève plus souvent de « l'éclectisme mou » que de la « sélection raisonnée ». Il dresse un profil de « l'enseignant moyen » selon lequel ce dernier « évolue [...] dans un univers méthodologique où le temps s'est pratiquement arrêté lorsqu'il a reçu son certificat » et dont la « philosophie didactique n'a pas pour fondement une théorie

précise, ce qu'il (se) cache parfois en revendiquant une approche "éclectique" » [Spielmann, 1992, p. 910]. Bien évidemment, ce portrait stéréotypé, non fondé sur des données expérimentales, sert le propos développé dans l'article (qui le nuance peu à peu, en diminuant ainsi la violence). Nous nous garderons bien de nous prononcer sur sa validité. Cependant son point de vue, ainsi que ceux de Beacco et Puren, renforcent l'importance du questionnement suscité par notre analyse des données de [Levy, 1997]. La littérature n'est pas, à notre connaissance, capable de répondre à toutes ces questions. Tenter de le faire dépasserait de très loin la portée d'un travail préliminaire de thèse, nous avons donc employé une stratégie différente. Les méthodes n'ayant pas permis de trouver un point d'entrée clair pour notre problème, nous avons adopté une démarche empirique, qui ne demande pas que nous privilégions une méthode en particulier. Après tout, ce sont les usagers qui doivent être au centre du processus d'indexation et non les méthodes qu'ils utilisent. Il nous paraît donc tout à fait cohérent de nous appuyer sur les pratiques des enseignants pour élaborer notre modèle.

1.5 Organisation de notre travail

Notre projet de thèse, bien qu'issu du projet MIRTO, s'en démarque et son intégration à la plateforme est une perspective à plus long terme. En effet, la recherche de ressources textuelles pour l'enseignement des langues n'est pas une problématique purement liée au paradigme de l'ALAO. Une base de textes indexée pédagogiquement pour l'enseignement des langues peut être utilisée afin de rechercher des textes pour les utiliser dans des séquences pédagogiques aussi bien dans le contexte d'un travail de l'apprenant sur un ordinateur que dans le cadre d'un cours présentiel sans support informatique. Avant tout, notre travail de thèse va consister à tenter de définir un exemple d'indexation pédagogique de textes pour l'enseignement des langues. Nous avons défini dans ce chapitre ce que nous entendions par « indexation pédagogique », il nous faut maintenant passer de la notion générique à une instance. L'instance que nous voulons définir garde certaines contraintes issues du projet MIRTO, particulièrement l'indexation de documents dont l'utilisation dans l'enseignement des langues n'est pas avérée. Une indexation doit permettre aux usagers d'accéder aux textes du système « répondant à leurs questions ». La base pourra donc contenir des textes qui ne répondent à aucune question, une grande quantité de textes devant augmenter les chances d'en trouver qui satisfassent les requêtes des enseignants ; le rôle de l'indexation étant précisément de « retrouver » les documents adéquats. Pour y parvenir, nous avons conclu qu'une démarche empirique était la plus adaptée. Cette dernière doit nous permettre d'explicitier en partie les critères intervenant dans le choix des enseignants et leurs relations, afin d'en proposer un modèle et éventuellement une approximation, qui puisse être satisfaisante dans le cadre des cas d'utilisation que nous avons définis. Cette approximation devra faire l'objet de la définition d'un langage documentaire et donc d'un modèle de recherche d'information, sur lequel fonder un prototype. Mais avant de recueillir des données sur les pratiques des enseignants nous devons passer en revue plusieurs domaines connexes à notre problématique.

Dans un premier temps nous rappellerons rapidement les évolutions dans le domaine de l'ALAO, ainsi que ses travers, la notion de corpus et nous en profiterons pour affiner la description de notre problématique. Le chapitre suivant contiendra un état de l'art de la description de ressources pédagogiques. Après avoir conclu, suite à ces deux chapitres, sur l'utilisabilité des outils actuels dans le cadre de notre projet et formulé quelques hypothèses quant à l'indexation pédagogique pour l'enseignement des langues, nous consacrerons un chapitre à notre recueil de données et aux conclusions que nous avons pu en tirer. Nous proposerons ensuite un modèle conceptuel

informatisable permettant à notre avis d'arriver à une forme d'indexation pédagogique de textes pour l'enseignement des langues. Nous éprouverons la faisabilité dans le chapitre suivant, dédié à la conception d'une maquette. Enfin, nous concluons sur le travail de recherche exposé tout au long de ce manuscrit.

Chapitre 2

Instrument, corpus et ALAO

« Language learning is certainly the oldest occupation in human history (sic), although language teaching can scarcely be considered the oldest profession. We sometimes speak of military intelligence as being the second oldest profession (re-sic), but even here the language learner probably comes into the picture after the foreign defector, the apostate, or the turncoat. In our day of accelerated communications, there is a tendency to stress the recency of massive study and learning of foreign language without reflecting on the significant contributions of other generations to the promotion of this art. ¹ »

[Lemieux, 1964]

Outre l'avantage de nous épargner l'écriture de notre propre introduction et celui de proposer – dans une perspective de didactique du discours scientifique – un bon exemple de « généralisation [...] périlleuse intellectuellement ² » [de Nuchèze, 1998, p. 28], ce court paragraphe a le mérite de rappeler, non sans humour, l'importance de la prise en considération d'autres travaux pour préciser ses propres recherches. Si dans notre cas il ne s'agit pas nécessairement d'« autres générations », l'examen de domaines connexes au nôtre va nous permettre d'affiner notre problématique et notre démarche.

Notre travail de conception d'une base de textes indexée pédagogiquement étant à destination d'enseignants de langues, il semble trivial de dire qu'il relève du champs des Environnements Informatiques pour l'Apprentissage Humain (EIAH) et en particulier de celui de l'Apprentissage des Langues Assisté par Ordinateur (ALAO). Une présentation historique de ces deux domaines va montrer que ce n'est pas si évident (section 2.1 ALAO, perspective historique p. 30). Après nous être positionné dans la perspective de ces deux champs, nous nous focaliserons sur le second. Une analyse de certains travers de l'ALAO nous permettra de tirer des conséquences méthodologiques concernant la manière d'aborder le problème mais aussi le processus de conception de notre prototype (section 2.2 Travers de l'ALAO p. 47).

1. L'apprentissage des langues est certainement la plus ancienne occupation dans l'histoire humaine, même si l'on ne peut guère considérer l'enseignement des langues comme la plus ancienne profession. On parle parfois du renseignement militaire comme le deuxième plus ancien métier, mais même dans ce contexte, l'apprenant des langues n'intègre sans doute le tableau qu'après le transfuge, l'apostat ou le renégat. En notre époque de communication accélérée, il y a une tendance à insister sur la nouveauté de l'apprentissage et de l'étude massifs des langues étrangères sans réfléchir aux contributions d'autres générations à la promotion de cet art.

2. Peut-être avons-nous, en voulant éviter cet écueil discursif, commis nos propres erreurs concernant un principe de base régissant l'usage de la citation... [Boch & Grossmann, 2001, p. 110]

Enfin, puisque notre problématique concerne l'organisation d'un ensemble de textes, nous nous intéresserons aux corpus et plus particulièrement à leurs usages dans l'enseignement des langues, dans une perspective de mise à profit de résultats théoriques pour l'indexation pédagogique de textes (section 2.3 Corpus et enseignement des langues p. 64).

2.1 ALAO, perspective historique

Pour nous positionner par rapport aux domaines de « l'apprentissage avec les ordinateurs », nous avons choisi une perspective historique, qui doit nous permettre de passer en revue la plupart des grands types de systèmes développés. En toute logique, avant de tracer une chronologie de l'ALAO, nous allons en consulter les racines.

2.1.1 Ordinateur et apprentissage

Étant donné l'objet de notre travail, notre présentation historique des EIAH a avant tout pour but de contextualiser la chronologie de l'ALAO que nous détaillerons dans la section suivante (§ 2.1.2 p. 35). Notre présentation sera donc succincte. La figure 2.1, empruntée à [Bruillard, 1997], vise ainsi à donner une vision globale de la chronologie.

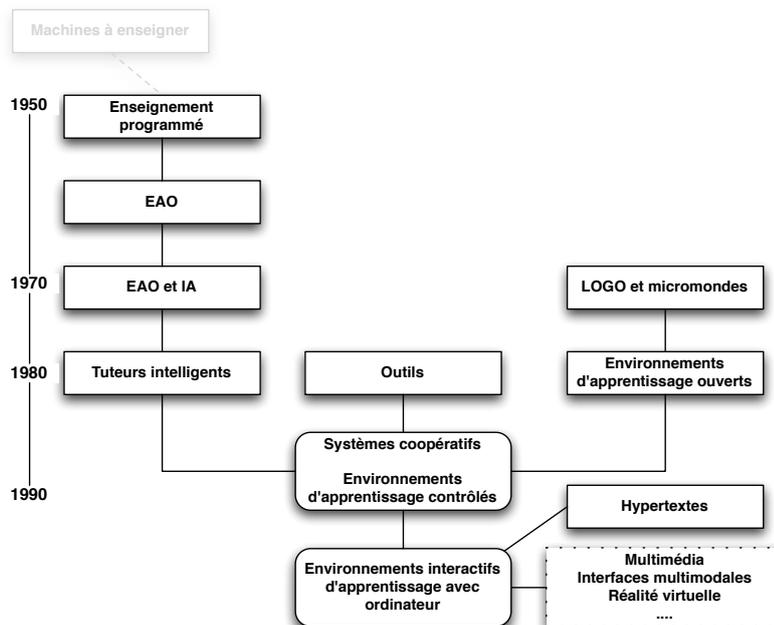


FIGURE 2.1: Principaux courants en informatique et éducation [Bruillard, 1997, Figure 1.1 p. 24]

a) L'enseignement programmé

Le premier courant recensé par la figure (2.1) est celui de l'enseignement programmé, issu des théories behavioristes et dans lequel « l'apprentissage n'est pas une simple connexion stimulus-

réponse, mais son efficacité résulterait des contingences de renforcements (nature, délais et fréquences) » [de Vries & Baillé, 2006, p. 31]. Cependant, comme aurait pu le faire remarquer Lemieux (cf. p. 29), ce courant n'est pas né par génération spontanée, mais suite à une période où les machines à enseigner ont proliféré [Bruillard, 1997, p. 33]. Il est par ailleurs possible de remonter jusqu'en 1809, pour y trouver une machine à enseigner la lecture [Antoniadis, 2008, p. 43]. La transition se situe principalement dans le déplacement de l'attention du matériel vers le programme, notamment sous l'impulsion de B.F. Skinner :

« L'enseignement programmé se définit avant tout comme un courant de recherche. Il utilise des machines à enseigner mais ce ne sont que des instruments, elles n'enseignent pas ; ce sont les programmes supportés par les appareils qui contiennent les éléments pédagogiques et servent à instruire. [...] Cette distinction entre programme et machine est toujours fortement réaffirmée. Le programme [...] correspond à la division en unités élémentaires de la matière à enseigner, l'articulation de ces unités, leur formulation, la progression des difficultés. La machine est le dispositif qui met en œuvre le programme. Ceci permet de considérer comme "machine" une liasse de feuilles ronéotypées ou un livre particulier. »

[Bruillard, 1997, p. 36]

Quatre principes de base sous-tendent l'enseignement programmé [de Montmollin, 1975, p. 5]³ :

- « la recherche d'un ordre efficace de présentation des concepts » (principe de structuration de la matière à enseigner, cf. définition de programme dans la citation ci-dessus) ;
- « l'adaptation permanente aux difficultés d'assimilation de l'élève » (principe d'adaptation) ;
- « la participation active de l'élève » (principe de stimulation) ;
- « la correction immédiate et point par point de l'acquis » (principe de contrôle).

Il y a deux grands types de programmes, respectant ces principes : les programmes dits « linéaires » (Skinner) et les programmes « ramifiés » (Crowder) [Rézeau, 2001, pp. 243–247]. Dans l'approche de Skinner, l'enseignant doit concevoir la séquence de manière à fournir à l'apprenant toutes sortes d'indices le menant à formuler, par lui-même, la bonne réponse (sans passer par un QCM, donc). Pour Crowder, au contraire, il ne faut pas éviter l'erreur. La progression s'effectue aussi pas à pas, mais propose un parcours adapté aux réponses de l'apprenant, selon qu'il effectue une erreur, qu'il répond correctement ou qu'il ne sait pas [Demaizière & Dubuisson, 1992, p. 40 & 41].

b) L'EAO et ses « mutations »

Aux débuts de l'Enseignement Assisté par Ordinateur (EAO) (années 60), l'ordinateur tel qu'il est utilisé joue le rôle des machines à enseigner et ne se démarque pas vraiment de l'enseignement programmé. Une séquence se joue selon un déroulement cyclique : « présentation d'informations et d'une question, réponse de l'apprenant, analyse de la réponse puis continuation ou branchement à une autre partie du cours » [Bruillard, 1997, p. 71]. Demaizière et Dubuisson distinguent au delà de ce fonctionnement typique d'une session d'EAO, genre tutoriel, d'autres familles de systèmes : exercices, tests et jeux⁴ [Demaizière & Dubuisson, 1992, p. 44–49].

Peu à peu, le domaine et son appellation évoluent, devenant Enseignement Intelligemment Assisté par Ordinateur (EIAO) quand, dans les années 70, des techniques d'intelligence artificielle sont

3. Dont sont tirés tous les passages entre guillemets de la liste.

4. Les autres systèmes qu'elles distinguent tendent à relever plus spécifiquement d'approches différentes.

intégrées aux logiciels avec comme objectif de créer des tuteurs artificiels pleinement adaptatifs. Pour ce faire, « il faut implanter des connaissances en machine pour les trois composantes qui interagissent dans la formation : le sujet, l'élève et le professeur » [Bruillard, 1997, p. 70]. Ces composantes rappellent les pôles définis dans le triangle de Houssaye (cf. figure 1.3 p. 16), la composante sujet représentant le pôle savoir. L'architecture des tuteurs intelligents est constituée de quatre modules principaux (cf. [Wenger, 1987]) : le module expert (composante sujet / pôle savoir), le module pédagogue (composante professeur), le modèle de l'élève et l'interface. Cette dernière peut s'apparenter à l'environnement dans le triangle de Houssaye [Houssaye, 1988, p. 41], ce qui tendrait à montrer l'ambition de ce type de système, qui veut modéliser la situation d'enseignement/apprentissage dans son ensemble.

c) Environnements d'apprentissage ouverts

C'est avec les environnements d'apprentissage ouverts que les travaux sur les interfaces prennent une autre dimension. Ils viennent d'une remise en question des principes behavioristes et cognitivistes qui sous-tendent tutoriels, exercices et autres tuteurs intelligents et de l'influence, entre autre, des théories piagétienne.

Les micromondes apparaissent dans les années 70. Il n'y a pas de définition patentée de la notion de micromonde [Bruillard, 1997, p. 140], nous les décrivons comme des « environnements de découverte fondés sur une approche constructiviste de l'apprentissage » [Zampa, 2003, p. 85]. Dans ces environnements, l'élève manipule et construit des objets ayant des propriétés, à la fois d'objets formels du domaine et d'objets du monde réel. Par exemple, l'enfant utilisateur de LOGO (langage de programmation utilisé comme outil de formation [Papert, 1981]) peut simuler les mouvements de la tortue avec son propre corps (objets du monde réel) mais ses trajectoires dessinent à l'écran des figures (objets formels du domaine de la géométrie).

Les logiciels de simulation se rapprochent des micromondes en ce sens qu'ils définissent un environnement dans lequel l'apprenant a l'initiative du déroulement, son rôle étant de manipuler des valeurs et autres paramètres d'une situation « [imitant] une partie de la réalité[...], d'observer et d'interpréter les résultats » [de Vries, 2001, p. 111].

Enfin, dans cette catégorie Bruillard ajoute les « outils ». « Les outils permettant de modéliser des situations sont très nombreux : d'une part les programmes classiques disponibles sur un ordinateur [...] traitements de textes, tableurs [etc.] et, d'autre part, les programmes plus spécialisés comme les logiciels de calcul formel ou les systèmes experts » [Bruillard, 1997, p. 170]. Toujours d'après Bruillard, sur un axe allant de l'expression à l'exploration, les outils relèvent de l'expression alors que les logiciels de simulation relèvent de l'exploration, entre les deux se trouvant divers types de micromondes.

d) Hypermédias

Les « hypermédias » sont dédiés à une autre forme d'exploration, ils n'imitent pas la réalité comme les simulateurs, mais regroupent des connaissances d'un domaine, liant entre eux des concepts. Ils tirent leurs origines de pratiques déjà mises au point pour le papier au Moyen Âge [Bernard, 1995, p. 323]. Pour une mention plus explicite du concept de l'hypertextualité (sans

que le terme ne soit employé), il faut remonter jusqu'en 1945, et à l'article « fondateur⁵ » de Vannevar Bush et sa description du *memex*, dispositif – resté virtuel – d'extension de la mémoire :

« *It affords an immediate step [...] to associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. [...] Thereafter, at any time, when one of these items is in view, the other can be instantly recalled merely by tapping a button [...].*⁶ »

[Bush, 1945, p. 8]

Dans le contexte des hypermédias dans l'EAO, « le rôle de l'ordinateur est de fournir un espace d'exploration correspondant aux concepts et relations d'une matière et la tâche proposée à l'élève est d'explorer cet espace » [de Vries, 2001, p. 110].

e) Environnements collaboratifs

Les logiciels d'apprentissage collaboratifs ou de *Computer-Supported Collaborative Learning* (CSCL) dérivent d'un champ plus large d'étude, les systèmes de travail collaboratif : « *computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment.*⁷ » [Ellis et al., 1991, p. 40]. Pour Lin Hsiao, alors que dans le domaine du travail collaboratif, l'accent est particulièrement porté sur les techniques de communication, dans l'apprentissage collaboratif c'est le contenu de la communication qui est au centre de la problématique [Lin Hsiao, 1996]. En effet, si le système peut avoir plusieurs rôles, l'objectif reste le même : la construction des connaissances par les élèves du fait de leurs interactions. Le système permet de créer un contexte propice à l'apprentissage de la communication sur un domaine précis, sous forme d'une « activité authentique équivalent à l'échange entre personnes exerçant le même métier » [de Vries, 2001, p. 112].

f) Récapitulatif

En guise de conclusion de cet historique succinct, nous proposons de récapituler les différentes fonctions des logiciels d'apprentissage que nous avons abordés sous la forme d'un tableau en récapitulant les principales caractéristiques (tableau 2.1). Nous l'empruntons à de Vries :

FONCTION PÉDAGOGIQUE	TYPE DE LOGICIEL	THÉORIE	TÂCHE	CONNAISSANCES
Présenter de l'information	tutoriel	cognitiviste	lire	présentation ordonnée
Dispenser des exercices	exercices ré-pétés	behavioriste	faire des exercices	association

5. « Le rôle fondateur attribué à Bush dans l'histoire de l'hypertexte est lié à son article intitulé “*As we may think*”, dont les idées ont considérablement influencé les chercheurs » [Bruillard, 1997, p. 230]

6. [...] Il permet une avancée immédiate pour l'*indexation associative*, dont l'idée centrale est la possibilité, à partir de n'importe quel objet, d'en sélectionner immédiatement un autre. C'est la fonctionnalité essentielle du memex. [explication du processus de liaison de 2 objets] Après cela, à n'importe quel moment, quand l'un de ces objets est en consultation, l'autre peut être retrouvé instantanément par une simple pression d'un bouton [...].

7. Des systèmes informatiques destinés à proposer une interface vers un environnement partagé pour soutenir des groupes de personnes engagés dans une tâche commune (ou un but commun).

FONCTION PÉDAGOGIQUE	TYPE DE LOGICIEL	THÉORIE	TÂCHE	CONNAISSANCES
Véritablement enseigner	tuteur intelligent	cognitiviste	dialoguer	représentation
Captiver l'attention et la motivation de l'élève	jeu éducatif	principalement behavioriste	jouer	
Fournir un espace d'exploration	hypermédia	cognitiviste constructiviste	explorer	présentation en accès libre
Fournir un environnement pour la découverte de lois naturelles	simulation	constructiviste cognition située	manipuler, observer	modélisation
Fournir un environnement pour la découverte de domaines abstraits	micro-monde	constructiviste	construire	matérialisation
Fournir un espace d'échange entre élèves	apprentissage collaboratif	cognition située	discuter	construction de l'élève

TABLE 2.1: « Les huit fonctions pédagogiques et leurs caractéristiques » [de Vries, 2001, p. 112]

g) EIAH : définition du domaine et prise de position

À l'heure actuelle et depuis la fin des années 90, le champ scientifique dont sont issus ces différents types de logiciel est désigné par le « concept intégrateur d'EIAH » [Grandbastien & Labat, 2006, p. 19], [de Vries, 2001, p. 106], [Vivet, 1997, p. 12]⁸. Grandbastien et Labat en donnent une définition large :

Définition 2.1 (EIAH) « *Le champ scientifique des EIAH correspond aux travaux focalisés sur les environnements informatiques dont la finalité explicite est de susciter et d'accompagner l'apprentissage humain, c'est-à-dire la construction de connaissances chez un apprenant. Ce type d'environnement mobilise des agents humains (élève, enseignant, tuteur) et artificiels (agents informatiques, qui peuvent eux aussi tenir différents rôles) et leur offre des situations d'interaction, localement ou à travers les réseaux informatiques, ainsi que des conditions d'accès à des ressources formatives (humaines et/ou médiatisées), ici encore locales ou distribuées. Il comprend les questions scientifiques et technologiques soulevées par la conception, la réalisation et l'évaluation de ces environnements, ainsi que la compréhension de leurs impacts sur la connaissance, la personne et la société.* » [Grandbastien & Labat, 2006, p. 18]

Cette définition est double : elle décrit d'une part le champ scientifique des EIAH et d'autre part ce qu'est un EIAH. Nous pouvons tout d'abord remarquer que la définition d'un EIAH est en accord avec les fonctionnalités représentées dans le tableau 2.1, puisque chacune des fonctions

⁸. Comme en témoigne le fait que la conférence EIAH 2003 ait succédé aux journées Environnement Interactif d'Apprentissage avec Ordinateur (EIAO₂) de 2001 organisées par l'ATIEF. <http://www.inrp.fr/atief/accueil.htm>

pédagogiques présentées fait intervenir au minimum un agent humain élève et un agent informatique (système, dont les rôles sont différents selon la fonction) et leur « offre des situations d'interactions ». Enfin, la finalité des logiciels présentés est clairement de « susciter et d'accompagner l'apprentissage humain ». Qu'en déduire concernant notre système ? Il mobilise un agent humain enseignant et un agent informatique (le système lui-même), qui interagissent afin que l'agent informatique assiste l'agent humain dans son choix de ressources brutes à utiliser dans la formation. Cependant, il nous semble que la « suscitation et l'accompagnement de la construction de connaissances chez l'apprenant » (cf. définition 2.1) n'est pas une finalité *explicite* de notre système. En effet, implicitement les ressources obtenues par l'enseignant sont destinées à la construction de connaissances chez l'apprenant, mais seulement après que l'enseignant les ait adaptées, didactisées et incluses dans une séquence pédagogique. L'usage prescrit de notre système (ses fonctions constituantes) est un usage par l'enseignant, qui fait la jonction entre le système et l'apprentissage humain. En ce sens, notre système n'est pas un EIAH. Il ne s'inscrit d'ailleurs dans aucune des fonctionnalités pédagogiques définies par de Vries (cf. tableau 2.1).

Notre travail n'est en revanche pas dissocié du champ des EIAH : c'est la conception de l'EIAH Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO), qui est à l'origine de notre projet. La problématique de l'aide au choix des textes a été soulevée par l'architecture même de la plateforme MIRTO et par cette nécessité d'organiser les textes qui serviront de support aux activités. Si à court terme la réalisation de MIRTO peut se passer de notre projet (et inversement, cf. § 1.1.2 p. 6), il fait partie intégrante de sa phase de conception. Enfin à notre sens, la problématique de l'indexation pédagogique de textes pour l'enseignement des langues relève des EIAH au même titre qu'elle relève de la didactique des langues et de la modélisation informatique (cf. chapitre 1). Les informations issues de la recherche des critères de choix des textes par les enseignants peuvent, même indépendamment du projet MIRTO, intervenir dans la conception du modèle pédagogue (cf. § *L'EAO et ses « mutations »* p. 31) dans un EIAH dédié à l'enseignement des langues.

2.1.2 Définition de l'ALAO

Comme nous l'avons vu rapidement dans le paragraphe précédent, notamment à travers le tableau 2.1, la recherche en EIAH est indissociable des théories de l'enseignement qui sous-tendent les formes de logiciels développés. L'application de ces théories à l'enseignement, relève tout d'abord de la didactique, qui est une recherche disciplinaire (cf. § *Différentes indexations pédagogiques* p. 17). L'utilisation d'EIAH dans l'apprentissage des langues s'appelle l'ALAO. Pour définir ce champ scientifique, nous passerons par son équivalent anglophone : *Computer Assisted Language Learning* (CALL), que Levy définit simplement.

Définition 2.2 (ALAO) « *The search for and study of applications of the computer in language teaching and learning.*⁹ » [Levy, 1997, p. 1]

Cette définition est cohérente par rapport à celle du champ des EIAH, et pour cause : Levy explique, tout de suite après, la genèse du *CALL*, ce qui permet de mettre en perspective les champs de l'ALAO et des EIAH :

« *The name is a fairly recent one: the existence of CALL in the academic literature has been recognizable for about the last thirty years. The subject is interdisciplinary*

9. La recherche et l'étude des applications de l'ordinateur dans l'enseignement et l'apprentissage des langues.

*in nature, and it has evolved out of early efforts to find ways of using the computer for teaching or for instructional purposes across a wide variety of subject areas, with the weight of knowledge and breadth of application in language learning ultimately resulting in a more specialized field of study.*¹⁰ »

[Levy, 1997, p. 1]

Tout comme Heift et Schulze [Heift & Schulze, 2007, p. 10], Chanier insiste aussi sur cette interdisciplinarité du champ de l'ALAO, qui d'après lui participe à le définir :

« L'ALAO est un domaine de recherche et développement qui intéresse plusieurs disciplines intervenant dans le champ des sciences cognitives :

- la linguistique (dans son acceptation large comprenant la linguistique théorique, appliquée, la didactique des langues),
- la linguistique-informatique,
- l'informatique (en particulier l'Intelligence Artificielle (IA)),
- la psycholinguistique.

»

[Chanier, 1995, p. 4]

Cependant, quand il définit l'ALAO, Chanier parle, à l'instar de Chappelle avec le sigle *Computer Applications in Second Language Acquisition* (CASLA) [Chappelle, 2001, p. 2], d'« Acquisition des Langues Assistée par Ordinateur ». Il rappelle, sans prendre parti, les différences qui peuvent exister entre « acquisition » et « apprentissage » :

« Nous utilisons à la fois les termes “acquisition” et “apprentissage”. L'acquisition est quelquefois différenciée de l'apprentissage lorsqu'on souhaite discriminer les processus cognitifs qui les sous-tendent. Le terme “acquisition” est alors utilisé pour signifier qu'une langue est assimilée souvent inconsciemment dans certaines conditions d'exposition (ou encore d'immersion), alors que le terme “apprentissage” renvoie à l'étude consciente d'une langue. »

[Chanier, 1995, p. 5]

Cette distinction entre l'acquisition et l'apprentissage met en exergue, d'une part l'interdisciplinarité de l'ALAO, mais aussi l'une des particularités de l'enseignement des langues par rapport à d'autres disciplines. Dans l'enseignement des langues, l'objet de l'apprentissage en est aussi le médium. C'est aussi le point de vue de Dargirolle, pour qui « [la] langue étant à la fois moyen et fin de cet enseignement, parler la langue et parler de la langue sont les pôles entre lesquels évoluent constamment enseignants et apprenants » [Dargirolle, 1999, p. 148]. L'influence de la xénité¹¹ entre L1 et L2 sur l'apprentissage de cette dernière va dans le sens d'une notion de matériau linguistique et de son importance pour l'enseignement des langues :

- « Selon que la langue enseignée est perçue par l'élève comme proche ou lointaine, les processus d'apprentissage mis en jeu ne seront pas exactement les mêmes. » [Dabène, 1975, pp. 51–52] ;

10. Le nom est en soi assez récent : l'existence du CALL dans la littérature est identifiable depuis une trentaine d'années. Le domaine est interdisciplinaire par nature. Il est né des premiers efforts pour trouver des façons d'utiliser l'ordinateur dans l'enseignement ou en vue d'aider à la formation dans une grande variété de domaines. Le poids des connaissances et la quantité d'applications en apprentissage des langues ont fini par résulter en un champ d'étude plus spécialisé. Dans la perspective historique de la définition, nous ne traduisons pas le sigle CALL par ALAO, d'une part car « aucun [sigle francophone] ne s'est imposé à l'égal de l'acronyme anglo-saxon CALL » [Rézeau, 2001, p. 224], mais aussi plus simplement car les deux sigles ne sont pas apparus en même temps. Par la suite nous emploierons sans discernement l'un ou l'autre.

11. La notion de xénité, introduite par Weinrich dans [Weinrich, 1989, p. 56], est définie par Pallotti de la manière suivante : « la proximité ou l'extranéité d'une langue par rapport à une autre » [Pallotti, 2002].

- « Cette proximité [entre les langues romanes permet] en effet entre les locuteurs concernés une relative inter compréhension et facilite en partie leurs apprentissages réciproques. » [Dabène, 1994, p. 41].

Bien que discutée ([Nevins *et al.*, 2007]/[Everett, 2007]), la notion de grammaire universelle et avec elle, celle d'universaux linguistiques¹² [Wikipedia, 2008e] appuierait cette hypothèse d'un objet d'apprentissage particulier par essence dans l'ALAO. Pour Bangs, la particularité de l'enseignement des langues par rapport à d'autres tient à sa nature spécifique : il vise l'acquisition d'un savoir-faire, au lieu d'être centré sur un savoir, comme pour la majorité des disciplines du *e-learning* [Bangs, 2002, p. 21]. Il s'agit d'un savoir-faire acquis spontanément par quasiment tout être humain, dans le cadre de la L1, mais dont l'apprentissage explicite est loin de fonctionner aussi systématiquement dans le cadre de la L2. Quelle qu'en soit la raison, le statut à part de l'enseignement des langues semble consensuel :

« *However strongly psychologists, linguists, and language teachers may disagree on other issues, the overwhelming majority support the view that learning a second language differs in significant ways from learning anything else.*¹³ »

[Hubbard, 1987, p. 229]

Peut-être est-ce à la particularité de son objet que l'on doit l'apparition d'un véritable champ disciplinaire (l'ALAO) au sein de celui des EIAH. En tout état de cause, les praticiens de l'ALAO revendiquent ce statut à part :

« *CALL is already on the cutting edge of technology: it can also be on the cutting edge of methodology. In order to achieve this, it is necessary to start with the concept that second-language learning is essentially different from other types of learning[...].*¹⁴ »

[Hubbard, 1987, p. 252]

Pour décrire l'évolution de l'ALAO, nous nous appuyons sur les travaux de Warschauer et Healey [Warschauer, 1996; Warschauer & Healey, 1998; Warschauer, 2000], qui isolent trois phases dans l'histoire de l'ALAO. Pour eux, ces phases traduisent d'une part des connaissances technologiques, mais aussi des points de vues pédagogiques. Elles ne sont pas strictement exclusives. Il n'y a pas de rupture nette entre elles et l'embrassement d'une nouvelle phase ne signifie pas nécessairement le rejet complet de la précédente. C'est d'après Bax, l'une des faiblesses de l'analyse de Warschauer et Healey [Bax, 2003, p. 16], qui préfère parler d'« approches » [Bax, 2003, p. 20]. D'après lui, le terme de « phase » ne rend pas suffisamment compte du fait que la description des pratiques associées à une période donnée peuvent faire intervenir des aspects représentatifs de phases « antérieures ». Nous avons malgré tout conservé ce terme, car Bax ne rejette pas une perspective historique [Bax, 2003, p. 22], que nous avons choisie comme point de vue pour ce paragraphe. Malgré toutes les réserves que la chronologie de Warschauer et Healey peut susciter, nous avons décidé de la présenter, quitte à mentionner certains des reproches qui lui sont faits. Comme le fait remarquer Delcloque¹⁵ dans « *History of CALL* », il nous semble que ce paradigme est utile pour montrer que l'évolution de l'ALAO reflète celles de la technologie, de la linguistique et des

12. Kirby *et al.* considèrent la grammaire universelle comme un biais d'apprentissage introduit par l'apprenant lui-même à la tâche d'acquisition de la langue, dont le processus d'apprentissage itératif aurait fait émerger des universaux linguistiques du fait de la succession de générations d'apprenants [Kirby *et al.*, 2004]

13. Peu importe l'intensité des différends entre psychologues, linguistes et enseignants de langues sur d'autres sujets, l'écrasante majorité d'entre eux s'accorde pour dire que l'apprentissage d'une L2 est significativement différent de tout autre apprentissage.

14. L'ALAO est d'ores et déjà à la pointe de la technologie ; il peut être également à la pointe de la méthodologie. Afin d'y parvenir, il faut partir de la notion que l'apprentissage d'une seconde langue est, par essence, *différent* des autres types d'enseignement.

15. pourtant assez critique face à ce type de démarche historique [Bax, 2003, p. 14]

sciences de l'éducation [Delcloque, 2000, p. 93].

2.1.3 L'ALAO *structural*

Cette phase, que Warschauer appelle en 1996 [Warschauer, 1996] et 1998 avec Healey [Warschauer & Healey, 1998] « béhavioriste », change de nom dans ses travaux ultérieurs [Warschauer, 2000] devenant « structurale » – Bax souligne cette incohérence dans son article [Bax, 2003, p. 15]. Elle correspond à la branche « tuteurs » dans l'historique des EIAH (années 60/70, cf. figure 2.1), aussi bien du point de vue temporel que du point de vue du type de logiciels. Sur cette période, l'ALAO adapte les technologies de l'EAO à sa problématique et produit principalement des tutoriels et des exercices, parfois considérés comme « bêtes et méchants » - comme en témoigne le sobriquet qui leur est parfois donné en anglais : « *drill & kill* » dérivé péjoratif de « *drill & practice* »¹⁶. La logique derrière ces programmes est la suivante :

«

- *Repeated exposure to the same material is beneficial or even essential to learning*¹⁷ ;
- *A computer is ideal for carrying out repeated drills, since the machine does not get bored with presenting the same material and since it can provide immediate non-judgmental feed-back*¹⁸ ;
- *A computer can present such material on an individualized basis, allowing students to proceed at their own pace and freeing up class time for other activities*¹⁹.

»

[Warschauer, 1996]

À cette époque, l'ordinateur est encore extrêmement coûteux et propose des interfaces minimales. Les principaux projets nécessitent des moyens considérables et s'exportent peu [Rézeau, 1999, p. 62]. Nous pouvons cependant citer deux de ces gros projets américains : en 1969 la NSF²⁰ subventionne *Programmed Logic for Automated Teaching Operations* (PLATO) IV (cf. NB 2.1 p. 39) et *Time-shared Interactive Computer Controlled Information Television* (TICCIT) [Jones, 1995]²¹. Les activités proposées par ceux-ci sont cohérentes avec les initiatives de l'époque décrites brièvement dans le paragraphe L'EAO et ses « mutations » (p. 31). Notamment dans le cas de l'allemand, TICCIT s'affranchit un peu de l'enseignement programmé en laissant l'apprenant avancer dans la séquence d'apprentissage à son rythme. La séquence est découpée en segments, chacun consacré à un point de grammaire. Au sein d'un segment l'apprenant a le choix entre 4 rubriques : l'objectif, la règle concernée, une série d'exemples de la règle concernée et des exercices. Les concepteurs ont longuement réfléchi l'ordre dans lequel les notions sont abordées et l'apprenant est censé le respecter, même s'il est libre de choisir ses rubriques, voire de passer directement au segment suivant s'il s'estime connaître l'objet de la leçon [Jones, 1995, pp. 90–96]. On reste donc plus proche de l'enseignement programmé que de l'hypertexte. De son côté,

16. *Drill* : exercice (dans le contexte militaire et pour les exercices de sécurité – ex : fire drill – des tâches répétitives) / *practice* : pratique, entraînement (pratique d'un fait de langue) / *kill* : tuer

17. L'exposition répétée au même matériel est bénéfique, voire essentielle à l'apprentissage ;

18. Un ordinateur est idéal pour dispenser des exercices répétés, car la machine ne se lasse pas de présenter le même matériel et peut fournir un feed-back, qui ne soit ressenti comme un jugement ;

19. Un ordinateur peut présenter le même matériel de manière individualisée, permettant aux apprenants d'avancer à leur propre rythme, libérant ainsi du temps en classe pour d'autres activités.

20. « La *National Science Foundation* (NSF) est une agence gouvernementale indépendante des États-Unis d'Amérique, pour soutenir financièrement la recherche scientifique fondamentale. » [Wikipedia, 2008f]

21. PLATO : University of Illinois at Urbana-Champaign / TICCIT : Brigham Young University.

Nota Bene 2.1 PLATO IV

Le projet PLATO, et *a fortiori*, sa génération la plus marquante, PLATO IV, s'appuyait sur un ordinateur central effectuant les traitements pour des terminaux. Le développement matériel de PLATO était faramineux, permettant à l'échelle des quelques institutions, ayant pu acheter leurs systèmes, des technologies qui ne verraient le jour que longtemps après sur les ordinateurs personnels. Plus de 20 ans avant l'invention du *World Wide Web* [Berners-Lee, 1989], étudiants et enseignants pouvaient se laisser des « notes » entre Urbana (Illinois) et Hawaii [Hart, 1995, p. 35], constituant ainsi une sorte de service d'e-mail restreint aux utilisateurs de PLATO. Un écran tactile permettait l'usage d'interfaces graphiques [Hart, 1995, pp. 27, 35] alors que le concept même de souris n'était breveté que depuis quelques années [Englebart, 1970]. Sans avoir la capacité de numériser du son, le système était interfacé avec des bandes sonores à « accès direct ^a » [Hart, 1995, p. 26]. Enfin un projecteur de microfiche [Hart, 1995, p. 26], relativement peu robuste, était associé au système. Au delà de ses caractéristiques techniques impressionnantes, PLATO était dédié à l'EAO. Outre pour l'enseignement des langues PLATO était utilisé pour la biologie, la physique et la formation au métier d'infirmière. TUTOR, un langage auteur, a été développé expressément pour PLATO. Pour produire des ressources, les utilisateurs pouvaient soit utiliser des « *paradigms* » (formats d'instructions préfabriqués), soit coder directement avec TUTOR, solution plus puissante, mais aussi plus complexe à maîtriser. Comme l'explique [Levy, 1997, p. 17], PLATO montre bien l'influence des technologies (matérielles et logicielles) sur la modélisation en ALAO :

« *The necessity of programming in TUTOR cut us off from most of the computational linguistics required by more sophisticated CALL designs and largely restricted us to variations on drill and practice.* ^b »

[Hart, 1995, p. 36]

Peu importantes ses limitations, PLATO marque une avancée décisive dans le monde de l'ALAO notamment par le fait qu'il était le premier projet à coordonner les efforts d'enseignants et de techniciens pour le développement d'activités d'ALAO ^c [Levy, 1997, pp. 17–18].

a. Un dispositif fait un « accès direct [à un élément] lorsqu'il écrit ou qu'il lit l'élément en se rendant directement à l'endroit où l'élément doit être écrit ou lu sans écrire ou lire les éléments précédents » [Wikipedia, 2008a]. ex : *Accès à des morceaux : cassette audio (séquentiel) / CD (direct).*

b. La contrainte de programmer en TUTOR nous a coupé de la plupart des notions de linguistique computationnelle nécessaires à des fonctionnalités d'ALAO plus sophistiquées et nous a cantonné largement à des variations sur le thème *drill and practice*

c. « *It was a ground-breaking project in numerous ways, and whatever its limitations, it was the first project to engage language teachers and technical staff in the development of CALL materials in a coordinated way.* »

PLATO propose une boîte à outils aux enseignants qui veulent créer des activités, cette boîte à outils est particulièrement orientée question-réponse-feed-back (cf. NB 2.1 p. 39). PLATO était un projet à grande échelle, proposant aux utilisateurs (enseignants ou apprenants) non seulement une grande quantité de supports de cours mais aussi l'accès à des technologies réservées uniquement à des institutions spécialisées.

Ces projets déclinent devant les nouvelles avancées pédagogiques du domaine de l'enseignement des langues, quand l'approche communicative commence à s'imposer [Hart, 1995, p. 32]. Cependant, comme nous le faisons remarquer en début de paragraphe (en citant [Warschauer, 1996]), il n'y a pas de rupture nette entre les différentes phases de l'ALAO :

« *[W]e wanted to devote more class time to real communication activities and believed that the TICCIT lab, by taking over the major part of grammar instruction, would*

*free up our class time for genuine language practice.*²² »

[Jones, 1995, p. 91]

Il ne s'agit pas bien sûr de l'approche communicative *per se*, mais la phrase montre une certaine prise en compte des évolutions dans le champ de la didactique des langues. D'ailleurs l'approche communicative n'a pas sonné le glas de ces projets : Levy fait remarquer que l'un comme l'autre sont encore accessibles sous d'autres avatars (micro-TICCIT ou TenCORE, une implémentation TUTOR pour PC [Hart, 1995, p. 33]), ce qui montre que ces systèmes étaient capables d'évoluer. C'est l'essor de l'ordinateur personnel qui enterre ces projets. Pour un coût beaucoup plus faible, le PC donne, dès la fin des années 80, des capacités de calcul suffisantes pour faire fonctionner des activités du type de PLATO et TICCIT (TICCIT s'est adapté plus facilement à ces changements matériels, TUTOR étant intimement lié à l'architecture du système PLATO).

2.1.4 L'ALAO *communicatif*

À la fin des années 70 et au début des années 80, alors que l'approche behavioriste de l'enseignement des langues a été remise en question (ex : [Chomsky, 1959]), l'approche communicative émerge. Adaptée à l'ALAO, cette dernière indique que les logiciels fournissant des activités de type « *drill & practice* » ne proposent pas assez de situations de communication authentiques et n'ont de ce fait que peu d'intérêt [Warschauer, 1996; Warschauer & Healey, 1998]. Bax critique cette appellation d'« ALAO communicatif », car pour lui ce terme n'est pas un descripteur adéquat des logiciels disponibles dans les années 80²³ [Bax, 2003, p. 18]. Il n'a pas tort à ce sujet, cependant sa remise en question de la phase ou approche est tempérée par Jung [Jung, 2005, p. 9], qui a pris au pied de la lettre sa recommandation d'« explorer les pratiques réelles de l'époque²⁴ » [Bax, 2003, p. 17]. En effet, Jung suggère que si les logiciels n'étaient pas spécialement tournés vers la communication, l'usage qui en était fait était peut-être plus en accord avec les préceptes de l'approche communicative, comme en témoigne [Jones, 1986]. Enfin, en relisant l'ouvrage sur lequel Warschauer se fonde pour définir l'ALAO *communicatif*, on peut constater que le terme y est déjà évoqué à plusieurs reprises [Underwood, 1984, chap. 6]. Underwood prévient cependant qu'il n'emploie pas le terme « communicatif » dans son sens le plus strict [Underwood, 1984, note 1, p. 67]. Pour lui la communication requiert l'introduction de nouvelles informations issues du monde réel, en ce sens les activités de ce que nous appelons dans ce paragraphe l'ALAO communicatif doivent, toujours d'après Underwood, « faire sens²⁵ », mais pas nécessairement être « communicatives ». Nous employons ce terme d'ALAO communicatif selon l'acception d'Underwood – même si comme le fait remarquer Bax, on ne peut pas considérer que l'ALAO ait tourné le dos à l'approche communicative par la suite²⁶ [Bax, 2003, p. 16]. C'est emploi est à notre avis justifié par le fait que :

22. Nous voulions consacrer plus de temps en classe à de réelles activités de communication et pensions que le laboratoire TICCIT, en prenant en charge la majeure partie de l'enseignement de la grammaire, libérerait du temps en classe pour une véritable pratique de la langue.

23. « *this must raise the suspicion that the term "communicative" may not be appropriate to the software or the uses of CALL during the 1980s at all* »

24. « *explore actual practice at the time* »

25. Les traductions consacrées du terme « *meaningful* » sont, selon le contexte, « significatif », « sérieux » (<http://dictionary.reverso.net/english-french/meaningful> ou « constructif » (Harrap's shorter). Dans ce cas précis, il s'agit d'activités au cours desquelles la production linguistique a un sens : « l'apprenant doit comprendre ce qu'il dit, mais le contenu est déjà connu. » [Underwood, 1984, note 1, p. 67]

26. « *language teaching in general still operates today very much within a communicative framework in many teaching contexts, so it is confusing for teachers to hear that "communicative CALL" is no longer with us* »

« *Part of the history of CALL is the fact that CALL is not a grassroots movement, but rather a top-down movement which trickles down from tertiary education institutes into secondary schools.*²⁷ »

[Jung, 2005, p. 9]

Si les dates proposées par Warschauer sont contestables [Bax, 2003, p. 16], le fait est que c'est au cours des années 80 que des propositions émergent sur les principes que doivent suivre les logiciels d'ALAO pour être plus en accord avec l'approche communicative. Nous en avons recensé un certain nombre dans le tableau 2.2. Nous l'avons intitulé « De "l'ALAO communicatif" à "l'ALAO intégratif" », en utilisant un point de vue d'observateur, dans une perspective d'intégration au découpage de l'historique de l'ALAO de Warschauer. Du point de vue des auteurs, il s'agit plutôt de préceptes d'un ALAO communicatif conceptuel (Underwood), ou de l'ALAO à l'heure de l'approche communicative (Oxford).

Dans la première colonne, nous reprenons les principes d'Underwood [Underwood, 1984, p. 52–54], tels que cités par Warschauer [Warschauer, 1996]. Ce dernier résume 13 principes en 7 points : les points A-4 (jugement/évaluation) et A-6 (usage de la L2) regroupent chacun deux principes énoncés par Underwood et les principes traitant de travers de l'ALAO structural ont été ignorés. La seconde colonne reprend des principes cités dans [Chanier, 1995], qui semblent issus de la contribution de Oxford dans un forum, auquel nous n'avons malheureusement pas pu accéder.

Les principes cités dans [Warschauer, 1996; Warschauer & Healey, 1998] sont bien antérieurs à ceux cités dans [Chanier, 1995], qui proposent un point de vue plus technique, prenant en compte des détails d'implémentation. Par exemple, l'élément B-4 mentionne explicitement le modèle de l'apprenant, un des modules principaux des tuteurs intelligents (cf. § *L'EAO et ses « mutations »* p. 31). Nous n'impliquons pas ici que les *desiderata* cités par Chanier [1995] suggèrent que l'ALAO doit s'articuler autour de tuteurs intelligents - comme le montre l'élément B-6 -, mais que les propositions de Oxford bénéficient d'un certain recul par rapport à celles d'Underwood, et qu'elles semblent s'orienter vers la phase de l'« ALAO intégratif » tel que décrit par Warschauer.

Au cours de la période située entre les propositions de Underwood et celles d'Oxford, avec le développement du micro-ordinateur, des « enseignants de langue-programmeurs²⁸ » peuvent maintenant développer leurs propres logiciels, ce qui n'était pas possible à l'époque des systèmes s'articulant autour d'un ordinateur central (cf. NB 2.1 p. 39). Les enseignants-programmeurs ont trois pistes pour la conception d'activités : l'utilisation d'un langage de programmation général (type BASIC ou Pascal), d'un langage auteur, dédié à la conception d'activité d'enseignement (sur le modèle de TUTOR, utilisable sur un ordinateur personnel, comme PILOT²⁹) ou d'un système auteur qui peut être pur (pas de programmation nécessaire) ou non [Wyatt, 1983].

Parmi les activités réalisables par un enseignant de langue-programmeur, une famille de programmes voit le jour au début des années 80 : les systèmes de reconstruction de texte [Levy, 1997, p. 25]. Ces programmes suivent le modèle de *storyboard* de Johns et Higgins [Higgins & Johns, 1984, p. 57]. D'après Brett, ce type d'activité satisfait au moins les « prémisses » A-6 et B-3 [Brett, 1994]³⁰, auxquels nous pouvons ajouter le A-1 et le A-2. Enfin, d'après [Zeilinger-Trier,

27. Une part de l'histoire de l'ALAO réside dans le fait qu'il ne s'agit pas d'un mouvement issu de la base, mais plutôt un champ disciplinaire provenant de l'enseignement universitaire infiltrant peu à peu l'enseignement secondaire.

28. « *individual language teacher-programmers* » [Levy, 1997, p. 25]

29. PILOT est décrit dans [Wyatt, 1983]. [Wikipedia, 2008g] en donne un aperçu rapide et un lien vers une implémentation open source.

30. cité dans [Levy, 1997, p. 26]

<p>« Prémises de l'ALAO communicatif » de [Underwood, 1984, p. 52–54] cités dans [Warschauer, 1996]</p>	<p><i>Desiderata</i> pour l'ALAO sélectionnés dans [Oxford, 1993] dans [Chanier, 1995, p. 14–15]</p>
<p>« <i>[Communicative CALL:]</i> » ^a</p> <p>A-1 « <i>focuses more on using forms rather than on the forms themselves</i> » ^b ;</p> <p>A-2 « <i>teaches grammar implicitly rather than explicitly</i> » ^c ;</p> <p>A-3 « <i>allows and encourages students to generate original utterances rather than just manipulate prefabricated language</i> » ^d ;</p> <p>A-4 « <i>does not judge and evaluate everything the students [do] nor reward them with congratulatory messages, lights, or bells</i> » ^e ;</p> <p>A-5 « <i>avoids telling students they are wrong and is flexible to a variety of student responses</i> » ^f ;</p> <p>A-6 « <i>uses the target language exclusively and creates an environment in which using the target language feels natural, both on and off the screen</i> » ^g ;</p> <p>A-7 « <i>will never try to do anything that a book can do just as well</i> » ^h.</p> <hr/> <p>a. [L'ALAO communicatif :]</p> <p>b. porte davantage l'attention sur l'usage des formes que sur les formes elles-mêmes ;</p> <p>c. enseigne la grammaire implicitement plutôt qu'explicitement ;</p> <p>d. permet aux étudiants de générer des énoncés originaux, plutôt que de se contenter de manipuler de la langue préfabriquée ;</p> <p>e. ne juge pas ni n'évalue tout ce que font les apprenants, ni ne les récompense avec des messages, des lumières ou des sons de félicitation ;</p> <p>f. évite de dire aux étudiants qu'ils ont tort et se montre capable de prendre en charge des réponses d'étudiants diverses ;</p> <p>g. utilise exclusivement la L2 et crée un environnement dans lequel l'usage de la L2 est naturel, aussi bien dans l'utilisation de l'ordinateur qu'à côté ;</p> <p>h. n'essaiera jamais de faire ce qui peut tout aussi bien être fait par un livre. (nos traductions)</p>	<p>B-1 « La compétence communicative doit être la pierre angulaire de l'ALAO » ;</p> <p>B-2 « L'ALAO doit fournir une assistance linguistique appropriée aux besoins courants de l'apprenant. Le concept d'assistance en situation d'apprentissage renvoie à ce que Vygotsky appelait la "zone proximale de développement" » ;</p> <p>B-3 « L'ALAO doit offrir un "Input" linguistique riche et authentique » ;</p> <p>B-4 « Le modèle de l'apprenant en ALAO doit être construit en intégrant des styles différents d'apprentissage » ;</p> <p>B-5 « Le contenu linguistique offert par un environnement ALAO est plus facilement assimilé s'il est présenté sous forme d'associations, elles-mêmes organisées autour de thèmes intéressants et pertinents, et autour de tâches proches de celles rencontrées en situation réelle de communication en L2 » ;</p> <p>B-6 « L'ALAO doit impliquer des types différents d'interaction, ne se limitant pas à celles de type tuteur-élève » ;</p> <p>B-7 « L'ALAO doit fournir un système de correction d'erreurs utile et pertinent en concordance avec les besoins changeants de l'apprenant » ;</p> <p>B-8 « L'ALAO doit impliquer toutes les aptitudes langagières appropriées (écriture, lecture, parler, écoute) et étudier comment la mobilisation de certaines capacités peut aider le développement d'autres » ;</p> <p>B-9 « L'ALAO doit former les utilisateurs à devenir progressivement des apprenants autonomes en leur enseignant explicitement les stratégies d'apprentissage en langues appropriées ».</p>

TABLE 2.2: De « l'ALAO communicatif » à « l'ALAO intégratif »

2002, p. 85] le critère A-7 est satisfait. Les systèmes de reconstruction de texte, tout comme les activités de lecture (*paced reading*) et les jeux linguistiques (cf. § *L'EAO et ses « mutations »* p. 31) laissent l'ordinateur dans le rôle de tuteur.

D'autres rôles existent pour l'ordinateur dans ce contexte d'ALAO communicatif. Il peut être stimulus de production langagière (écrite ou orale), le type de logiciel privilégié pour ce genre d'interaction est la simulation. Nous avons aussi vu qu'un logiciel dans lequel l'ordinateur joue le rôle de tuteur peut aussi servir de stimulus (reconstruction de texte en groupe par exemple). Enfin, comme dans les environnements d'apprentissage ouverts (cf. § *Environnements d'apprentissage ouverts* p. 32), l'ordinateur est utilisé comme outil. Dans le cas de l'ALAO, un outil ne doit pas nécessairement fournir un apport linguistique, mais est utilisé pour effectuer des tâches linguistiques [Rogerson-Revell, 2007, p. 59]; par exemple un correcteur orthographique dans un traitement de texte.

2.1.5 L'ALAO intégratif

L'ALAO communicatif porte en lui les premiers éléments de la phase suivante, que Warschauer [1996] appelle la phase de l'ALAO intégratif, qui dépend directement de deux avancées technologiques : le multimédia, et à travers lui la notion d'hypermédia (cf. § *Hypermédiats* p. 32) d'une part et le développement des réseaux et d'internet d'autre part (cf. § *Environnements collaboratifs* p. 33). Warschauer et Healey expliquent le terme « *integrative* », par deux types d'intégration [Warschauer & Healey, 1998] :

- Intégration, comme dans « approche intégrée de l'enseignement d'une langue », qui consiste, selon l'objet de l'intégration, en le fait de ne pas cloisonner les aptitudes (cf. [Poussard, 2000, p. 77]), les compétences de communication ou les points de vues sur la langue (cf. [LeBlanc, 1989]) ;
- mais aussi d'un point de vue technologique, en intégrant plus complètement la technologie dans le processus d'apprentissage. (cf. exemple de la tâche proposée pour *the italian project* dans [Levy & Kennedy, 2004, p. 54] ou dans un contexte de formation professionnelle [Demai-zière & Dubuisson, 1992, pp. 239–241]).

Les *desiderata* cités par Chanier (cf. tableau 2.2) sont en quelque sorte représentatifs de la transition progressive de l'ALAO communicatif à l'ALAO intégratif : alors que la compétence de communication reste la pierre angulaire de l'ALAO (*desideratum* B-1), il doit impliquer toutes les aptitudes (*desideratum* B-8 et premier type d'intégration de Warschauer et Healey). Ils sont en ce sens représentatifs de l'une des objections de Bax à Warschauer selon laquelle l'« approche intégrée³¹ » ne s'affranchit pas de l'approche communicative [Bax, 2003, p. 20]. Enfin le *desideratum* B-5 relèverait plus du second type d'intégration, même si ce n'est qu'en filigrane : ce qui est suggéré par ce précepte est le fait que l'environnement doit avoir un rôle proche de celui qu'il aurait dans une situation réelle de communication. Enfin, le lien entre cet alinéa et l'ALAO intégratif est plus évident de par son recours à la notion de tâche, qui rappelle la perspective actionnelle (cf. § 1.4.2 p. 19) représentative de cette nouvelle phase de l'ALAO [Warschauer & Healey, 1998]. Le lien entre la tâche et l'ALAO intégratif est plus évident dans des travaux comme ceux de Goodfellow *et al.* [1999], qui considèrent les propriétés intégratives offertes par la tâche pour pallier les insuffisances de certains avatars de l'approche communicative :

31. Pour son analyse, Bax n'utilise pas le terme de phase et parle pour son analyse d'« approches ». Il se réfère à l'équivalent, moyennant quelques mises au point, de la phase intégrative de Warschauer, sous le terme d'« *Integrated approach* ».

« [Communicativeness] itself is now less of an all-inclusive goal for language pedagogy than it once was. In Little's view [Little, 1997], the "dominant form" of the communicative approach (by which he means a focus on spoken communication at the expense of both writing and grammatical knowledge), fails to develop any kind of language awareness and does not foster learner autonomy. [The] technology needs to support a range of learning activity, not simply increased opportunity for communication. This includes both productive activity intended to increase learners' facility with the target language, and reflective activities designed to develop their language awareness.³² »

[Goodfellow *et al.*, 1999]

2.1.6 Les programmes de l'ALAO intégratif

C'est pour cette phase que les critiques de Bax [Bax, 2003, p. 16] à l'égard de la chronologie de Warschauer sont les plus fortes. Il soulève notamment des problèmes de dates, car pour lui nous sommes toujours dans le paradigme qu'il appelle « approche ouverte » (qui pourrait être assimilée chez Warschauer à la phase communicative de l'ALAO). Cependant, son point de vue est très axé sur les produits finaux, alors que celui de Warschauer semble plus centré sur les concepts et propositions. Cette différence se retrouve dans les termes choisis par l'un et l'autre. Le terme intégratif choisi par Warschauer témoigne de l'*intention* d'intégration (affectant de manière représentative les technologies développées). Le choix de Bax de parler d'« ALAO intégré », témoigne d'une intégration *effective*, indissociable d'une « normalisation » de l'outil informatique. Dans ce contexte, « normalisation » signifie que la technologie fait à tel point partie du mode de vie des gens qu'elle devient invisible³³[Bax, 2003, p. 25]. Un tel niveau d'intégration rendrait suranné l'usage même du sigle ALAO dans la mesure où nous ne parlons ni d'ALAS (Apprentissage des Langues Assisté par Stylo), ni d'ALAL (Apprentissage des Langues Assisté par Livre) [Bax, 2003, p. 23]. Nous sommes effectivement encore éloigné de ce niveau d'intégration – selon la deuxième forme exprimée par Warschauer (intégration de la technologie dans le processus d'apprentissage, cf. § L'ALAO intégratif p. 43). Et ce d'autant plus loin que, comme le fait remarquer Rogerson-Revell, il est difficile d'imaginer l'informatique se stabilisant autant que les technologies « stylos » et « livres », dans un futur proche [Rogerson-Revell, 2007, p. 58]. Il nous semble que ce que Warschauer appelle la phase de l'ALAO intégratif témoigne d'apports théoriques et peut être associé à des nouveautés technologiques (Multimédia et Réseaux). Ces apports et nouveautés sont nécessaires à la normalisation suggérée par Bax, notamment à travers le développement de nouveaux types de systèmes d'ALAO ou de nouvelles utilisations de systèmes existants.

Pour Levy et Stockwell [Levy & Stockwell, 2006, p. 3–5] :

« *CALL is about design, development, and evaluation, as well as research and practice conceived around a ready-made product or generic application.*³⁴ »

32. La capacité à communiquer est maintenant moins un but inclusif pour la pédagogie des langues que par le passé. D'après Little [1997], la « forme dominante » de l'approche communicative (par laquelle il entend l'accent placé sur la communication orale aux dépens à la fois de l'écrit et du savoir grammatical) ne parvient pas à faire développer aux apprenants une conscience du langage et ne favorise pas l'autonomie de l'apprenant. La technologie doit soutenir une gamme d'activités d'apprentissage, et pas uniquement fournir de plus amples opportunités de communication. Cela comprend aussi bien des activités de production censées augmenter l'aisance avec la L2, que des activités réflexives conçues pour développer leur conscience du langage.

33. « *The technology is so integrated into our lives that it becomes invisible* »

34. L'ALAO est un champ disciplinaire centré sur la conception, le développement et l'évaluation, ainsi que la recherche et les pratiques conçues autour de programmes clé en main ou génériques

[Levy & Stockwell, 2006, p. 3]

Ils séparent donc les programmes en deux types et adoptent ensuite le point de vue du concepteur. En effet, selon que l'utilisateur du programme est concepteur ou veut se servir du programme avec ses apprenants, la notion de généricité n'aura pas le même sens. À défaut d'une terminologie consacrée, dans la suite du paragraphe nous appellerons ces points de vue respectivement *concepteur* et *utilisateur*. Un programme générique, du point de vue utilisateur, sera un programme qui n'est pas expressément dédié à l'enseignement mais auquel des usages pédagogiques ont été trouvés (il s'agit des outils décrits dans le § *Environnements d'apprentissage ouverts* p. 32). Dans le domaine de l'ALAO, nous avons cité l'exemple d'un logiciel de traitement de textes, Levy et Stockwell parlent aussi de logiciels de messagerie électronique. Selon ce point de vue utilisateur, des logiciels clé en main seront des sites Web éducatifs, des CD-ROM et autres programmes d'ALAO qui existent depuis les phases précédentes de l'ALAO et qui continuent à être développés aujourd'hui ; ce qui tend à montrer, une fois de plus que les phases de Warschauer ne sont pas nettement séparées.

Si nous adoptons maintenant le point de vue du concepteur, la notion même de généricité subit un glissement sémantique : tout système auteur (*authorware*) est générique, mais il peut être un système auteur dédié à l'ALAO ou être plus lié à une technologie qu'à une application. Par exemple, un concepteur de ressources d'apprentissage pourra les développer avec des logiciels complètement génériques (liés à une technologie) comme des éditeurs Web (comme SeaMonkey³⁵, pour citer un exemple de logiciel libre) ou encore des logiciels de traitement de texte ou de présentation (comme OpenOffice.org Write et Impress³⁶) [Levy & Stockwell, 2006, p. 3]. Mais il pourra aussi se tourner vers des systèmes auteurs plus spécialisés (liés à une application), comme *Hot potatoes*³⁷, qui proposent des modèles d'activités, paramétrables par le concepteur. Des systèmes auteur comme *Hot Potatoes* ne sont pas des systèmes auteurs purs au sens de [Wyatt, 1983, p. 37], puisqu'un utilisateur connaissant le javascript pourra modifier les activités créées. Les ressources produites sont en général des sites Web qui diffèrent dans leurs objectifs et leurs points de vue. Ils peuvent être destinés à des publics divers : ciblés pour un niveau donné ou prévus pour proposer des ressources à des publics plus larges. Ils peuvent être épurés ou étonnamment complets, intégrant une large gamme d'activités faisant intervenir des technologies variées, comme par exemple le site de la BBC « *Learn Welsh* »³⁸ [Levy & Stockwell, 2006, p. 3]. Enfin, ils peuvent être tout à fait novateurs de par leur approche pédagogique (cf. § *Corpus pédagogique* p. 77) ou ne se focaliser que sur des exercices de type *drill & practice*.

Cependant le programme symbole de l'ALAO intégratif serait plutôt le Système de Gestion de l'Apprentissage (SGA) (*Learning Management System (LMS)* en anglais), qui élargit, pour cette phase de l'ALAO, le rôle de l'ordinateur à la gestion des données et des messages [Rogerson-Revell, 2007, p. 60]. Via un SGA, il est possible d'organiser de grandes quantités de données (cours, activités, faisant intervenir aussi bien du texte que de l'image ou du son), notre travail s'inscrit directement dans ce type de rôle. La composante « gestion des messages » traite, quant à elle, de la possibilité de constituer des communautés d'enseignants ou d'apprenants, qui peuvent non seulement être modélisés par le système (profil de l'apprenant), mais aussi communiquer entre eux et avec leurs enseignants dans le cadre d'activités pédagogiques ou de manière autonome (CSCL, cf. § *Environnements collaboratifs* p. 33, ou d'un point de vue plus général

35. <http://www.seamonkey-project.org/>

36. <http://www.openoffice.org/index.html>

37. Voir [Ebbrell, 1998], pour un rappel synthétique des principales fonctionnalités de *Hot Potatoes* et un lien vers la page officielle (<http://hotpot.uvic.ca/>)

38. <http://www.bbc.co.uk/learnwelsh/>

Communication Médiatisée par Ordinateur (CMO)). Au delà de la place des outils dans le processus d'apprentissage (cf. § 2.1.5 p. 43), ce pendant technologique propose des outils offrant eux-mêmes d'en intégrer d'autres. Bien sûr, en fonction des tâches effectuées par les apprenants, cette prise en charge d'outils divers participe au fait de ne pas cloisonner les objectifs de chaque tâche à une seule aptitude ou compétence de communication et d'inclure la technologie au processus d'apprentissage. Mais pour y parvenir, ils *intègrent* des ressources qui n'ont pas nécessairement été conçues spécialement pour ou avec le SGA. Par exemple, Moodle³⁹ ne propose pas uniquement de développer des ressources via la plate-forme elle-même, mais aussi d'intégrer différents types de fichiers (ex : HTML, PDF, jpeg, mp3, mov, etc.) [Cole & Foster, 2008, p. 44], des liens vers des pages Web externes, mais aussi des modules SCORM (cf. § 3.3.2 p. 101), contenant des activités qui peuvent communiquer avec la plate-forme, notamment pour mettre à jour des profils d'apprenants [Cole & Foster, 2008, p. 17]. Moodle permet aussi l'inclusion d'activités créées avec Hot Potatoes ou TexToys⁴⁰, grâce au module développé par Gordon Bateson⁴¹. Cette forme d'intégration – de diverses technologies – n'est pas généralisée à tous les SGA, mais est de plus en plus présente dans la problématique du domaine (cf. chapitre 3). Levy et Stockwell la désignent par le terme d'« hybridation » [Levy & Stockwell, 2006, p. 209]. Ce terme regroupe donc aussi bien l'inclusion de ressources développées en dehors du SGA que l'utilisation de médiums différents, comme par exemple des cours hybrides combinant CD-ROM et SGA ([Rogerson-Revell, 2007, p. 62] ou *Language in The Toolbox* [Depover *et al.*, 2000, p. 7]) ou une méthode papier et des prolongements sur internet [Mangenot, 2002].

2.1.7 Conclusion

Le champ disciplinaire de l'ALAO, bien qu'autonome, existe par ses interactions avec d'autres domaines plus ou moins proches. Les champs disciplinaires influençant le plus directement l'ALAO sont la didactique des langues, les EIAH et l'informatique au sens large du terme. En effet, en proposant une description de chacune des phases définies par Warschauer [1996], nous avons pu constater que chaque phase pouvait être principalement associée à un courant théorique de la didactique des langues et à un type de technologie en particulier (cf. tableau 2.3). Ainsi, l'ALAO structural s'appuie sur des architectures centralisées et sur la méthode audio-orale. Les systèmes réalisés s'inscrivent dans la famille des tuteurs et suivent donc le champ des EIAH (qui ne porte pas encore ce nom). Avec l'arrivée de l'approche communicative et de l'ordinateur personnel, les systèmes produits changent et de nouvelles utilisations voient le jour ; l'ordinateur comme outil, par exemple. Peu à peu cet ALAO communicatif devient intégratif avec l'arrivée de la perspective actionnelle et le développement de l'hypermédia et de la CMO, et avec eux, le développement de plate-formes d'enseignement des langues.

Notre travail s'inscrit dans cette dernière mouvance. En effet, c'est un outil qui est avant tout destiné à être utilisé directement par les enseignants, aussi bien dans un contexte de cours présentiel traditionnel, que dans un contexte d'ALAO. Mais dans cette perspective intégrative, il pourrait tout à fait être ajouté à un SGA, comme outil autonome (comme un dictionnaire), d'autant plus que les fonctionnalités de notre outil correspondent tout à fait aux prérogatives de gestion des données des systèmes actuels. Il peut aussi être intégré, et permettre à l'enseignant d'utiliser directement les textes sélectionnés avec des modules auteurs comme TexToys. Enfin, notre prototype pourrait faire l'objet d'un degré d'intégration à une plate-forme encore supérieur :

39. <http://moodle.org/>

40. Système auteur dédié à l'ALAO (<http://www.cict.co.uk/software/textoys/index.htm>)

41. <http://docs.moodle.org/en/Hotpot>

PHASE DE L'ALAO	COURANT THÉORIQUE	TECHNOLOGIE
Structural	Courant béhavioriste, méthode audio-orale (cf. tableau 1.1 p. 23)	Ordinateur central et terminaux
Communicatif	Approche communicative (cf. § 1.4.1 p. 18)	Ordinateur personnel
Intégratif	Perspective actionnelle (cf. § 1.4.2 p. 19)	Réseaux, Multimédia (et hypermédia)

TABLE 2.3: Récapitulatif des « phases de l'ALAO » de Warschauer [2000]

dans le cadre de l'inclusion à MIRTO, la plateforme doit pouvoir formuler des requêtes et traiter les résultats. Cela suppose l'interaction avec les modules de scénarisation ou de gestion des apprenants, afin que le système soit à même de proposer une requête adaptée et d'adopter une stratégie de sélection des textes à partir des résultats.

Si notre projet n'est pas un EIAH *per se*, il s'inscrit non seulement dans la problématique du champ, mais il fait partie intégrante de la logique intégrative de l'ALAO.

Nous allons maintenant considérer, par le biais de travers constatés, certains aspects méthodologiques de la conception de systèmes dédiés à l'ALAO.

2.2 Travers de l'ALAO

Nous avons pu constater que la transition entre deux phases de l'ALAO n'était pas abrupte et que les phases ne s'excluent pas nécessairement. En effet, en revenant au tableau 2.2, force est de constater que les *desiderata* cités par Chanier, relèvent finalement autant de ce que Warschauer appelle l'ALAO intégratif que de l'ALAO communicatif, tout en revendiquant la filiation de l'approche communicative (*desideratum* B-1). Cependant, ces transitions marquent aussi un certain rejet des préceptes ou erreurs des phases précédentes, comme le rejet de l'approche de l'apprentissage des langues qui sous-tend l'ALAO structural, par l'ALAO communicatif. Dans ce paragraphe, nous allons passer en revue certains des écueils dont la prise de conscience a fait évoluer l'ALAO, et en tirer des conséquences quant à notre travail.

2.2.1 Conception d'une base de textes

Pour expliciter l'objectif de ce paragraphe, nous souhaitons revenir à une phrase de Levy et Stockwell que nous avons déjà citée sans nous y attarder (cf. note ³⁴ p. 44) :

« *CALL is about design, development, and evaluation, as well as research and practice conceived around a ready-made product or generic application.* ⁴² »

[Levy & Stockwell, 2006, p. 3]

Cet énoncé propose à notre avis une décomposition très pertinente des différents problèmes de l'ALAO. En liant tous ces aspects dans une phrase aussi concise, les auteurs mettent en

42. L'ALAO traite de la conception, du développement et de l'évaluation, ainsi que de la recherche et des pratiques conçues autour de programmes clé en main ou génériques

exergue les interrelations entre ces différents problèmes. En effet, le développement d'un système d'ALAO dépend étroitement de sa phase de conception. Toute cette phase de création (conception + développement) est elle-même indissociable de l'évaluation des systèmes d'ALAO, comme en témoigne le fait que des travaux d'évaluation débouchent sur des recommandations quant au processus de création de systèmes d'ALAO (comme ceux de Hubbard [1987, 1996] ou de [Chappelle, 2001]). Ranguin souscrit d'ailleurs au point de vue de Reeves & Marlino quand ils signalent que « l'évaluation n'est pas une fin en soi ; c'est un moyen pour améliorer la prise de décision » (cité par [Ranguin, 1995, p. 152]). Enfin ce « triptyque conception, développement, évaluation » s'alimente des recherches et pratiques autour des programmes qui en sont issus.

Ayant rappelé l'évolution du champ de l'ALAO, nous allons expliciter ici les composantes de notre projet. Comme l'indique le titre de notre thèse, il s'agit avant tout d'un travail de conception (contenant une phase de développement). Si nous avons choisi cette entrée des « travers de l'ALAO » pour revenir sur notre problématique, c'est parce qu'elle a été définie en grande partie en réaction à certains d'entre eux. La base de textes que nous concevons n'est pas en soi un EIAH, elle n'est pas destinée à être utilisée directement par les apprenants (même si nous avons pu montrer que dans une vision intégrative de l'ALAO, notre problématique relève directement de ce champ disciplinaire). Nous n'allons donc pas pouvoir l'évaluer à l'aune de grilles prévues pour décrire des systèmes complets. Nous n'allons pas présenter ici une revue exhaustive des travaux d'évaluation de ressources d'ALAO, mais nous focaliser sur les aspects qui ont le plus influencé notre projet, sous le prisme de certains travers clés.

2.2.2 Conception

Avec l'ordinateur personnel, apparaît la possibilité pour les enseignants de langues de concevoir eux-mêmes leurs ressources d'ALAO (cf. note ²⁸ p. 41). La plupart du temps, les enseignants et formateurs qui s'attellent à ce type de tâche n'ont pas suivi de formation en informatique. Ils produisent des ressources destinées spécialement à leurs apprenants, qui sont la plupart du temps bien adaptées à ces derniers. Cependant, ces enseignants-programmeurs, le plus souvent autodidactes, travaillent principalement seuls. Ils construisent en général des ressources de petite taille, très ciblées, dont la diffusion large et la commercialisation sont le plus souvent impossibles [Demaizière & Dubuisson, 1992, p. 249]. « Les industriels, comme les universitaires ou les enseignants du primaire ou du secondaire se sont alors tournés vers la création d'équipes de conception et de réalisation permettant de réunir les compétences nécessaires sur des projets de moyenne et grande taille[, ce qui rend] incontournables les questions d'organisation » [Demaizière & Dubuisson, 1992, p. 250].

a) Des carences didactiques

Contrairement aux enseignants-programmeurs, les industriels disposent d'un véritable savoir-faire informatique mais, conformément aux avertissements de Cros [1965]⁴³, ils ont tendance à

43. Cités par [Bruillard, 1997, p. 33] : « Il est malheureusement possible de présenter sous les apparences de la plus haute technicité des résultats très hâtivement élaborés, sans base méthodologique sérieuse. Et ceci ne manquera pas de se produire si un effort officiel n'est pas fait. Par simple imitation des travaux américains, des sociétés privées, plus ou moins sérieuses, proposeront sur le marché de l'enseignement leurs nouvelles techniques présentées comme des recettes magiques. Ce qui ne manquera pas de provoquer (ou d'accentuer) une réaction de défense de la part du principal intéressé : le corps enseignant. »

laisser les « prouesses techniques prendre le pas sur des principes didactiques solides⁴⁴ », comme le fait remarquer Decoo à propos d'un logiciel particulier (Le Français Par Ordinateur) [Decoo, 1984, p. 45]. Le cas de ce logiciel n'est pas isolé :

« [Given] the high cost of creating pedagogically sound programs, it is hard to see how money could be made, and so it becomes clearer as to why there are not hundreds of publishers out there rushing to offer us quality language training over the Internet, and also why so much of what is there is unsound. »⁴⁵

[Bangs, 2002, p. 27]

Ce type de constatation n'est pas sans rappeler la situation des machines à enseigner telle que décrite par Skinner [1963] 40 ans auparavant. Pour lui, « l'éducation est une grande industrie » dans laquelle prospèrent « des forces commerciales perturbatrices » avec des produits « qui ne sont pas toujours au point » [Skinner, 1963, p. 170]. Les « gadgeteurs » qui produisent des machines « [n'appliquant] pas des principes psychologiques valides » [Skinner, 1963, p. 171] trouvent leur équivalent chez Bangs et Cantos Gómez. Ils les évoquent, en des termes moins dépréciatifs, parmi les facteurs qui expliquent que malgré le temps écoulé depuis les principes énoncés par Underwood (cf. tableau 2.2), nous ne soyons pas submergés de programmes d'apprentissage des langues de qualité : ils parlent de « sauts technologiques qui laissent la pédagogie dans leur sillage⁴⁶ » [Bangs & Cantos Gómez, 2004, p. 226–227]. Bouldoires effectue, quant à lui, un constat proche dans le domaine des EIAH. Il parle d'une « politique de l'offre [accompagnée] d'un discours promotionnel préconstitué » [Bouldoires, 1997, p. 130]. Il fait allusion au fait que souvent les produits développés ne se préoccupent que peu des besoins et pratiques potentielles des usagers (« demande ») et se concentrent sur une approche consumériste techniciste. Cette dernière fait des aspects techniques le premier argument de vente et ne rend aucunement compte des usages potentiels des systèmes [Bouldoires, 1997, p. 129–130].

Ces remarques ne sont pas nouvelles et suscitent des recommandations quant au processus de conception. Hubbard, par exemple, suggère que les acteurs de l'ALAO devraient attacher plus d'importance au champs de l'acquisition d'une langue seconde et de la didactique des langues, pour appliquer leurs avancées à la conception, l'évaluation et l'utilisation des systèmes⁴⁷ [Hubbard, 1987, p. 252]. Le consensus actuel prône un travail concerté de spécialistes de l'enseignement et de spécialistes de l'informatique, comme en témoigne le point de vue de Bangs, pour qui la plupart des spécialistes de l'ALAO a fini par comprendre que c'est l'aspect pédagogique qui doit guider la création de matériel, pas la technologie. « Le point de départ ne doit jamais être la technologie [...] mais des principes didactiques⁴⁸ » [Bangs, 2002, p. 24].

44. « *technical prowesses take precedence over sound didactic principles* »

45. Étant donné le coût élevé de la création de programmes ayant de *solides bases pédagogiques*, il devient difficile de voir comment faire de l'argent avec. À partir de ce constat, on comprend bien pourquoi il n'y a pas des centaines d'éditeurs se précipitant pour nous proposer des formations de langues de qualité sur internet et aussi pourquoi une si grande partie de ce qui est disponible ne s'appuie sur aucune base solide.

46. « *Technological leaps leaving pedagogy in their wake* »

47. « *It is clear that many CALL practitioners, reviewers and authors need to pay more attention to second-language acquisition and to teaching methodology in CALL software design, evaluation and use.* »

48. « *Most of us have learnt over the years that it should be the pedagogy which drives materials creation, not the technology – even though there are those who do not seem to have learnt this message yet! The starting point should never be the technology, whether we are talking about platform or software, but the learning design principles.* »

b) Un manque de maîtrise de la technologie

Cependant, cela ne signifie pas que, dans le processus de conception de matériel pédagogique, seuls les spécialistes informatiques aient besoin de comprendre mieux les autres points de vues mis en jeux. Comme le signale Godwin-Jones, dans le cas des technologies Web, les avancées décisives dans le domaine technologique ne doivent pas se faire sans la participation des enseignants, qui doivent « être impliqués directement dans le fait de repousser les frontières de la technologie pour aider à trouver de nouvelles méthodes pour améliorer l'apprentissage des étudiants⁴⁹ » [Godwin-Jones, 2000, p. 10]. Cet avis est partagé par Allum, qui constate que les logiciels développés par les enseignants avec de solides bases pédagogiques échouent souvent en termes d'utilisabilité et suggère qu'ils s'appuient plus systématiquement sur des principes de base d'Interface Homme-Machine (IHM) dans leurs activités de conception. Ce point de vue n'exclut pas celui selon lequel les logiciels professionnels pêchent quant à eux souvent en terme d'adéquation avec les principes de certains enseignants⁵⁰ [Allum, 2001, p. 146].

c) Collaboration de deux pôles

Dans cette succession de remarques, deux types de compétences, centraux pour la conception d'applications d'ALAO, sont mis en évidence : compétences didactiques et compétences techniques. Ils ne remettent aucunement en question l'invocation des multiples champs disciplinaires par Chanier, ou Heift et Schulze (cf. § 2.1.2 p. 35). De la même manière que les compétences informatiques mises en oeuvre peuvent faire appel à l'IHM, les compétences didactiques utilisées pour la conception d'une application peuvent faire intervenir des principes de psychologie, d'acquisition d'une langue seconde. Ces aspects didactiques et techniques renvoient directement au nom même du champ disciplinaire, l'ALAO. À la lumière de ce sigle, la remarque de Bangs concernant la préséance de la composante didactique sur la composante technologique (cf. note⁴⁸ p. 49) semble couler de source : l'apprentissage des langues (compétences didactiques) est *assisté* par l'ordinateur (compétences techniques). Cependant, d'après la littérature, un travail supplémentaire semble devoir être fait dans ce sens.

Nous parlons ici de compétences et non d'acteurs impliqués :

- tout d'abord car historiquement nous avons pu voir que certains acteurs jouaient parfois plusieurs rôles comme les enseignants-programmeurs, mais aussi dans le cas de projets industriels parfois développés principalement par des informaticiens ;
- mais avant tout en considération des solutions proposées par différents chercheurs.

Les travaux que nous avons cités (§ 2.2.2 p. 48) font intervenir sans les nommer deux principaux pôles de compétences, qui correspondraient aux deux grands types que nous avons évoqués ici. Il est bien sûr possible d'adopter un grain de précision beaucoup plus fin, comme celui développé par Demaizières et Dubuisson pour l'EAO [Demaizières & Dubuisson, 1992, pp.257–261] :

1. expertise de contenu et compétences didactiques (contenus, structuration des contenus, présentation) ;
2. compétences pédagogiques (scénarisation, types d'activités) ;

49. « *We need to have teachers directly involved in pushing the boundaries of the technologies in order to help search out new methods to enhance student learning* »

50. « *Teacher-designed, pedagogically sound software often fails in terms of usability, while highly usable, professionally made software is often based on ideas that are not in accord with those of a particular curriculum or teacher.* »

3. compétences EAO (connaissance des médias et outils informatiques disponibles dans le domaine) ;
4. compétences en ergonomie des systèmes d'EAO (interactions entre apprenant et ordinateur) ;
5. compétences médiatiques (intégration de ressources multimédia au sein du système) ;
6. compétences de médiatisation EAO (implantation : les auteurs parlent de « médiatisation EAO », car l'implémentation peut faire intervenir des systèmes auteurs dédiés à l'EAO, mais il s'agit ici de la partie la plus technique du travail).

Les compétences que nous appelons didactiques sont avant tout les compétences 1 (contenus, didactiques) et 2 (pédagogiques) de la liste, alors que les compétences techniques s'étendent de l'élément 3 à l'élément 6 (EAO, ergonomie, médiatiques, médiatisation). Ce que nous avons appelé les compétences didactiques sont en général fournies par des enseignants, des formateurs ou des didacticiens alors que les compétences techniques relèvent d'ingénieurs, d'informaticiens, spécialistes des problématiques d'apprentissage des langues, le plus souvent. Bangs aborde le problème sous un angle plus proche du génie logiciel, selon lequel les compétences listées pourraient être placées sur un continuum entre le « quoi faire » et le « comment faire » : le « quoi faire » relevant de la conception (compétences didactiques avant tout) et le « comment faire » de son implémentation (surtout les compétences techniques). Pour lui, c'est sur cette répartition des compétences entre les acteurs qu'un effort doit être fait par les deux partis :

« Part of the problem may be the broken link between pedagogical design and interactive programming, but even if the creators of a project or course need to hand over the technologists for the actual creation of the routines, this should always be subject to a rigorous set of design documents which specify the exact nature of the interactions to be employed [...]. [There] is a clear need for the whole area of instructional design to be promoted—and in “both directions”. On the one hand, pedagogical designers need to understand the interactions and other technical possibilities at their disposal, and which they can use to put appropriate learning theories into practice. On the other hand, the technology community needs to understand what the language learning community really requires from the application of the new technologies. The gap may be narrowing, but it still exists and ways should be sought to bridge it in the near future.⁵¹ »

[Bangs, 2003, p. 95]

En évoquant le besoin des enseignants et concepteurs d'acquérir plus de compétences techniques (en particulier celles qui correspondraient au 3^e ensemble de la liste ci-dessus : médias / outils informatiques présents dans le domaine [Demaizière & Dubuisson, 1992, pp.257–261]), il rejoint Rogerson-Revell [2007] et Allum [2001]. Nous retrouvons en ce point l'approche intégrative au

51. Une partie du problème réside peut-être dans le lien brisé entre conception pédagogique et programmation des interactions, mais même si les créateurs d'un projet ou d'un cours doivent laisser la main aux technologues pour la création des routines, cela devrait toujours faire l'objet d'un ensemble de documents de conception rigoureux qui spécifie la nature exacte des interactions à employer. Il y a un besoin clair de promotion de la zone de la conception de ressources éducationnelles, et ce dans les « deux sens ». D'une part, les concepteurs pédagogiques doivent comprendre les interactions et les autres possibilités à leur disposition, qu'ils peuvent utiliser pour mettre en application les théories de l'apprentissage adéquates. D'autre part, la communauté des spécialistes de la technologie doit comprendre ce dont la communauté de l'apprentissage des langues a réellement besoin en ce qui concerne l'utilisation des nouvelles technologies. Le fossé tend à se réduire, mais il existe toujours, et des moyens de le traverser doivent être cherchés.

centre de la problématique de l'ALAO (cf. § 2.1.5 p. 43). Pour intégrer au mieux l'outil informatique dans le processus d'apprentissage, il paraît en effet primordial que les détenteurs des compétences didactiques, qui peuvent être ceux qui les mettent en place dans le cadre de la formation (enseignants et formateurs), acquièrent une compréhension suffisante de l'outil informatique pour l'intégrer au mieux dans l'offre de formation qu'ils représentent.

Notre travail relève quant à lui de la seconde partie de « l'autre direction ». Notre point de vue est celui d'un représentant de ceux que Bangs appellent les technologues. Il appuie la nécessité de notre approche empirique dans l'optique de prendre en compte les connaissances didactiques du domaine. De par nos objectifs, le travail que nous effectuons est avant tout un travail de conception et de modélisation. Prises dans une perspective de création d'outil, nos décisions technologiques doivent s'appuyer sur des connaissances didactiques pour répondre aux besoins des utilisateurs. Ce sont eux les détenteurs des compétences didactiques, qui décideront de l'intégration de l'outil à leurs pratiques. Il nous faut donc adapter le plus possible la conception de notre outil à ces dernières.

d) Processus

Depover *et al.* [2000] abordent ce problème différemment. Ils ne l'adressent pas uniquement sous l'angle des acteurs, mais aussi sous celui du processus de conception. Ils recommandent ainsi un cycle de développement en plusieurs phases dans lesquelles interviennent deux classes d'acteurs : clients et concepteurs. Dans le cas présenté dans l'article, et d'une certaine manière dans le nôtre, les détenteurs des compétences pédagogiques ne sont pas tous des concepteurs. Certains font partie de cette classe d'acteurs, mais les autres sont considérés comme clients, non pas dans un sens monétaire mais plutôt dans un sens de commanditaire. N'ayant de commanditaire, pour adapter ce modèle à notre problème nous devons considérer les utilisateurs comme clients. En effet, c'est à eux qu'est destiné notre système. Si les utilisateurs finaux ne sont pas considérés comme clients dans le modèle de Depover *et al.*, c'est que les projets présentés sont plus « traditionnellement ⁵² » ancrés dans la problématique de l'ALAO, en tant que champ dans lequel sont conçus des EIAH dont les utilisateurs sont les apprenants. Considérer les enseignants comme client nous paraît cependant cohérent d'après l'étude préliminaire que nous avons effectuée, selon laquelle la recherche de supports textuels est une tâche à part entière du travail des enseignants (cf. § 1.1.2 p. 6 / [Loiseau, 2003]).

Depover *et al.* parlent pour désigner leur méthodologie de « design incrémentiel et itératif » [Depover *et al.*, 2000, p. 4]. À chaque étape, des échanges avec le client sont suggérés (itératif) selon des modalités qui dépendent des objectifs de l'étape et de l'état de développement du produit (incrémentiel). Le cycle de vie comporte trois étapes : « pré-adoption, implantation et routinisation » [Depover *et al.*, 2000, p. 16], dont seules les deux premières nous concernent. La phase de pré-adoption consiste selon eux en la définition d'une idée mobilisatrice (base de texte indexée pédagogiquement pour l'enseignement des langues, dans notre cas), l'établissement d'un cahier des charges, puis d'une structure modulaire [Depover *et al.*, 2000, pp. 5–6], avant de commencer la phase d'implantation par un prototype qui doit servir de support à un dialogue avec les clients afin de l'améliorer [Depover *et al.*, 2000, p. 8]. Nous généralisons ici les concepts

52. À quoi on peut opposer le point de vue de Murray et Barnes selon qui : « *Whilst some software may provide opportunities for pupils to work more independently, in most contexts the teacher remains the primary director of learning.* » [Murray & Barnes, 1998, p. 251] Alors que certains logiciels offrent la possibilité aux élèves de travailler indépendamment, dans la plupart des contextes, c'est toujours principalement l'enseignant qui dirige l'apprentissage. (traduit par Lucy Garnier).

présentés dans l'article qui servent le développement d'applications dont les utilisateurs finaux sont les apprenants. Dans notre cas, les fonctionnalités ont été définies *a priori* en fonction des besoins du projet MIRTO et de notre étude préliminaire (cf. chapitre 1). Cependant, ces fonctionnalités constituent des lignes directrices à confronter aux pratiques des enseignants, à partir desquelles nous définirons un modèle. Enfin, le cadre de ce travail s'arrête à la première itération de la phase d'implantation, avec un premier prototype. Étant donné la quantité de travail pour y parvenir, nous n'allons pas poursuivre au delà de cette première phase. Dans l'optique de parvenir, à plus long terme, à un produit utilisable, le modèle et le prototype doivent se montrer évolutifs et ainsi permettre la prise en compte des conclusions qui pourront être tirées des échanges avec les clients, après le développement du prototype (phase de test à grande échelle par exemple). Cependant l'objectif principal de notre travail est la définition d'un modèle, dont nous devons montrer la faisabilité avec un prototype.

2.2.3 Modélisation

Les modèles de développement que nous avons présentés visent à cerner le rôle de chacun. D'autres modèles existent, comme celui présenté dans [Ahmad *et al.*, 1985] (cf. figure 2.2 citée par [Bannert *et al.*, 2003]), qui relève plus d'un positionnement du champ disciplinaire de l'ALAO que des compétences impliquées dans la réalisation d'un projet. La présence dans ce modèle du Traitement Automatique des Langues (TAL), sous forme de « Linguistique Computationnelle/Traduction automatique », est intéressante selon notre point de vue des travers de l'ALAO : il met en exergue la nécessité de rendre l'ordinateur capable d'« interpréter » des éléments de sens du langage (que ce soit dans les ressources ou les réponses d'apprenants).

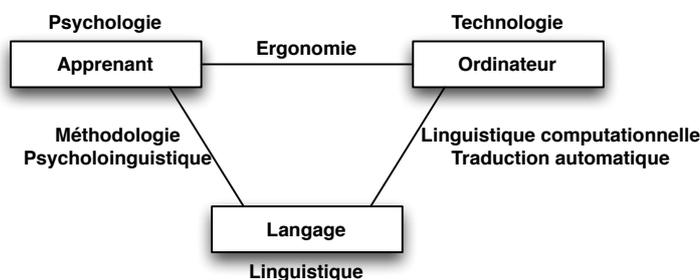


FIGURE 2.2: « Un modèle typique de l'ALAO et les domaines scientifiques associés [Ahmad *et al.*, 1985, p. 45] » [Bannert *et al.*, 2003, Fig. 1]

a) Analyse de logiciels

Antoniadis & Ponton [2002] proposent une enquête ayant porté sur 289 logiciels d'apprentissage des langues, édités entre 1991 et 2000. Après un travail de classement, un échantillon représentatif de 31 logiciels a été isolé pour une étude plus détaillée. Antoniadis [2004] en fournit certains résultats avant de les analyser. Dans presque l'intégralité des logiciels testés, le point de vue adopté pour considérer la langue est extrêmement réducteur. Rézeau, quant à lui, parle d'analyse des réponses extrêmement « minimale » pour la majeure partie des logiciels produits dans les années 90 [Rézeau, 2001, p. 365]. Nous pouvons citer comme exemple représentatif de ce point de vue, le « syndrome du blanc » [Antoniadis *et al.*, 2005b], c'est-à-dire le fait que certains logiciels,

même parmi ceux développés à la fin des années 90 n'étaient pas capables de traiter l'espace : *la□□bouillotte* n'est pas considérée comme *la□bouillotte*⁵³ [Antoniadis, 2004, § III].

Cet exemple est symptomatique du point de vue adopté par les logiciels. Ils ne considèrent pas la langue en tant que telle, les réponses de l'apprenant sont considérées comme des séquences de caractères, sans qu'aucun sens ne leur soit associé. Dans le cas de cet exemple, il ne s'agit même pas de savoir que *la* est un article défini, qui est féminin car il est lié au nom *bouillotte*, mais juste de considérer l'espace, comme ayant un sens différent des autres caractères, du fait qu'un ou plusieurs espaces ont le même sens, ce qui n'est pas le cas pour une lettre par exemple. Ce type de raisonnement peut être adopté avec des exemples moins simplistes. Si une phrase du type « *The fairy tale you like is The Shepherdess and the Chimney Sweep* » est destinée à montrer l'omission du pronom relatif en anglais, ses autres nombreuses propriétés, comme le fait que le sujet de *like* soit *you* ou que *is* soit une forme du verbe *to be* au présent, ne seront pas considérées. Dans ce contexte, une réponse d'apprenant du type « *The fairy tale that you like is The Shepherdess and the Chimney Sweep* » pourra être considérée fautive, sans autre explication [Antoniadis *et al.*, 2005a]. Blanchard *et al.* [2009] font état de réponses considérées fautes à cause de majuscules absentes (ex : la france) ou ajoutées à tort (ex : le Lapin). Cette série d'exemples explicite la citation du paragraphe 1.1.1 (p. 5) :

« le plus souvent, informaticiens et didacticiens des langues n'admettent pas la même acception du terme "langue". Pour les uns, il s'agit d'une séquence de codes, pour les autres d'un système de formes et de concepts. Cette différence peut s'expliquer aisément lorsque l'on constate que l'informatique [seule], par définition, ne peut considérer et ne peut traiter que la forme de la langue dénuée de toute interprétation, alors que pour la didactique des langues, la forme n'existe que par ses propriétés et des concepts qu'elle est censée représenter. »

[Antoniadis *et al.*, 2005b, p. 65]

Dans un premier temps nous pouvons noter, qu'Antoniadis *et al.* opposent, conformément à ce que nous indiquions dans le paragraphe 2.2.2 p. 48, le point de vue du détenteur des compétences didactiques à celui du détenteur des compétences techniques, même si ce n'est pas dans un contexte de processus de constitution de ressources ALAO. Il s'agit plutôt ici d'une forme de constat descriptif opposant le point de vue des spécialistes de l'enseignement des langues (enseignants et didacticiens), indépendamment de leur utilisation de l'outil informatique à celui des créateurs de ressources ALAO. Les auteurs suggèrent ici, qu'au moins en ce qui concerne la langue, les technologues se doivent de comprendre mieux le point de vue des membres de la communauté de l'apprentissage des langues (cf. citation de Bangs [2003], p. 51).

b) Constats

Pour en revenir au contenu de la citation, cette incapacité à considérer la forme de la langue par ses propriétés et les concepts qu'elle est censée représenter, la pauvreté du sens associé aux séquences linguistiques, quelle qu'en soit la longueur, est l'une des limites des systèmes d'ALAO constatées dans [Antoniadis *et al.*, 2005a]. Murray et Barnes, dans leur étude des apports du multimédia pour l'apprentissage de la grammaire et des structures linguistiques, se rapprochent de cette analyse. Ils déplorent l'absence fréquente de sens associé aux structures présentée par les différents logiciels :

53. □ représente le caractère *espace*.

« *The language learning process itself and the way language “works” should be highlighted in the software to aid the learner in his or her analysis of language patterns. The software should provide learners with an aid towards internalising and using structures accurately, not just recognising and producing unanalysed “chunks” of language.* ⁵⁴ »

[Murray & Barnes, 1998, p. 256]

La seconde limite concerne la rigidité des logiciels d'ALAO. Même au-delà des logiciels clé en main, qui proposent des séquences prédéfinies qui ne peuvent être modifiées, les systèmes auteurs de type *Hot Potatoes* permettent de définir facilement des activités, mais enferment l'enseignant dans des modèles pédagogiques pré-établis et peu altérables. Les ressources à partir desquelles sont créées les activités ne sont pas interchangeables : à une activité correspondra un texte et un seul, un modèle pédagogique et un seul. Sans utiliser d'autres types de connaissances et compétences, il est difficilement imaginable d'aller plus loin dans la modélisation didactique [Antoniadis *et al.*, 2005a]. Cette rigidité est perçue aussi par les utilisateurs de systèmes auteur, qui, selon une étude effectuée par Delcloque & Bramoullé en 2001, reprochent aux outils existants leur manque de flexibilité [Delcloque & Bramoullé, 2001, p. 283]. Ils généralisent en quelque sorte les remarques faites pour l'apprentissage lexical par Selva et Chanier : « les matériaux sont créés par les concepteurs, les épreuves ne sont pas renouvelables et la sélection des mots sur lesquels portent les exercices ne se fait pas toujours en fonction du travail de l'apprenant et de ce qu'il a réellement acquis » [Selva & Chanier, 2000, § 1]. Enfin les auteurs remarquent que les logiciels, tels qu'ils sont conçus, forcent les enseignants (utilisateurs) ayant peu ou pas de compétences informatiques à manipuler des concepts qui n'appartiennent pas à leur problématique ⁵⁵ [Antoniadis *et al.*, 2005a]. Même sans revenir à l'époque où tout système d'ALAO demandait de la part des concepteurs des compétences en programmation (ALAO structural. Cf. § 2.1.3 p. 38), la conception de cours interactifs requiert certaines connaissances en informatique.

c) Le TAL pour l'ALAO

Pour les auteurs, les limites citées peuvent être dépassées par l'utilisation du TAL (Antoniadis & Ponton [2002, 2004]; Antoniadis *et al.* [2004a,b, 2005a,b, 2007b]; Kraif *et al.* [2004]; Antoniadis [2004]). En effet, concernant le manque de sens associé à la langue, il ne s'agit pas de suggérer que, grâce au TAL l'ordinateur sera capable de « comprendre » n'importe quelle séquence langagière, mais des outils comme un analyseur morphologique, ou un correcteur orthographique, peuvent combler certaines lacunes. Pour Bangs le feed-back est un élément primordial de l'ALAO, car il permet à l'apprenant de ne pas se sentir jugé sur sa production (cf. § 2.1.3 p. 38), encore faut-il que le système puisse lui signifier pourquoi il considère que l'apprenant a fait une erreur [Bangs, 2002, p. 26 & 28]. L'analyse des réponses d'apprenants (description des erreurs) et le feed-back sont les problématiques où l'utilité de ce type d'outils est la plus évidente. Heift & Schulze [2007] proposent un panorama des travaux sur le sujet dans leurs chapitres 3 et 4.

En dehors du contexte des réponses d'apprenants, les informations de sens ajoutées par des ou-

54. Le processus d'apprentissage des langues lui-même et la manière dont la langue « marche » devrait être mis en valeur dans le logiciel pour aider l'apprenant dans son analyse des motifs langagiers. Le logiciel devrait fournir à l'apprenant une aide pour l'internalisation et l'usage correct des structures, pas seulement la reconnaissance et la production de « morceaux » de langue non analysés.

55. « *the third problem concerns the fact that software applications are computer sciences oriented. This orientation forces the language teachers with no or little computing skills to manipulate concepts which do not belong to their set of problems.* »

Nota Bene 2.2 Flexion, Forme, Lemme, Lexème, Lexie, etc.

Lorsque nous parlons de certains outils TAL, il peut nous arriver d'avoir recours à la terminologie du domaine que nous allons rappeler rapidement ici. Nous nous inscrivons principalement dans la théorie sens-texte.

Pour définir la plupart de ces notions nous allons recourir à la notion de *mots-formes*. Polguère définit le terme en toute rigueur dans son ouvrage à travers les propriétés d'« autonomie de fonctionnement » et de « cohésion interne ». Pour l'usage que nous allons en faire nous nous contenterons de la définition instinctive qu'il en donne : « formes linguistiques qui sont séparées à l'écrit par des espaces ou des marques de ponctuation » [Polguère, 2003, p. 43]. Ainsi, *chanson* et *chansons* sont deux mots-formes.

« Une *lexie*, aussi appelée *unité lexicale*, est un regroupement 1) de mots-formes ou 2) de constructions linguistiques qui ne se distinguent que par la flexion. Dans le premier cas, il s'agit de *lexèmes* et dans le second cas, de locutions. Chaque lexie (lexème ou locution) est associée à un sens donné. Que l'on retrouve dans le signifié de chacun des signes (mots-formes ou constructions linguistiques) auxquels elle correspond » [Polguère, 2002] cité par [Mangeot-Lerebours *et al.*, 2003]. Par exemple, *chanson* est un lexème (et donc une lexie) regroupant les deux formes que nous avons indiquées ci-dessus. *Avoir la dalle* est un exemple de lexie de type locution.

Quand nous parlons de *lemme*, nous nous référons à « une forme arbitraire, mais conventionnelle, abstraite sur l'ensemble des formes d'un [lexème] » [CNTRL, 2008]. Cette forme conventionnelle, qui est utilisée comme entrée dans les dictionnaires papiers, est aussi appelée *forme canonique* (cf. [Polguère, 2003, p. 89]).

Enfin, « la *flexion* est [...] un mécanisme morphologique qui consiste en la combinaison d'un radical et d'un affixe – appelé *affixe flexionnel* – ayant les trois propriétés suivantes :

1. son signifié est très général, plutôt abstrait et appartient nécessairement à un petit ensemble de signifiés mutuellement exclusifs appelé *catégorie flexionnelle* [...];
2. l'expression de sa catégorie flexionnelle est imposée par la langue [...];
3. sa combinaison avec le radical d'une lexie donne un mot-forme associé à la même lexie.

[...] Toutes les langues n'ont pas la même richesse flexionnelle » [Polguère, 2003, p. 60]. Par exemple, en français, dans le mot-forme *manges*, l'affixe flexionnel *-es* appartient à la catégorie flexionnelle de la 2^e personne du singulier de l'indicatif présent des verbes du premier groupe (et d'aucune autre catégorie flexionnelle) est associé au radical verbal *mang-*. *Manges* – tout comme *mangerai* ou *mangeâtes* – est donc une des *formes fléchies* de la lexie dont le lemme est *manger*.

tils TAL peuvent aussi trouver une utilité pour s'attaquer à la seconde limite que nous avons indiquée : la rigidité des systèmes ALAO. Nous pouvons, pour illustrer ce point de vue, nous appuyer sur le système ALEXIA, un « environnement informatique d'aide à l'apprentissage lexical du français langue seconde » [Selva & Chanier, 2000, § 1]. Le système est dédié au thème du travail. Les auteurs expliquent comment avec les informations ajoutées par des traitements de base du TAL, ils arrivent à proposer un système qui permette la génération semi-automatique d'activités. Les activités dont il s'agit ici sont des exercices lacunaires, proposant à l'utilisateur de reconstituer des phrases avec les mots manquants. La création de tels exercices nécessite donc un découpage des textes en phrases. Les textes ne sont ainsi plus considérés comme des suites de caractères mais comme des ensembles de phrases. Les activités décrites dans l'article proposent à l'apprenant la liste des mots qui doivent être retrouvés, mais sous forme de lemme (cf. NB 2.2 p. 56). Une analyse morphologique⁵⁶ permet donc d'obtenir pour chaque mot son lemme et d'es-

56. Un analyseur morphologique permet d'obtenir une représentation morphologique du texte. Une telle représentation indique pour chaque mot « sa forme de base (*forme canonique* ou *lemme*), sa catégorie syntaxique et

camoter les formes fléchies concernées. Ce qui empêche le système de générer automatiquement des activités est le fait que l'analyse morphologique s'appuie, comme son nom l'indique, sur la morphologie et non sur le sens. Ainsi, pour reprendre un exemple cité dans l'article, dans la langue française, il existe au moins deux lexies *boîte*. La première serait l'acception de « boîte » dans le sens de réceptacle (*Elle trie ses articles en les rangeant dans des boîtes*) et la seconde serait « boîte » au sens d'entreprise (et nous omettons les sens « discothèque » et « gamelle »). Mais, du point de vue de la morphologie, ces deux lexies sont indifférenciables, puisque associées exactement au même ensemble de formes fléchies (avec les mêmes traits morphologiques). Au contraire, un analyseur morphosyntaxique – prenant en compte le contexte – pourra faire la différence entre la lexie *manger* dans le sens *manger une pomme*, et la lexie dans l'acception *Son hôtesse lui accommode son manger*⁵⁷, puisque la différence de sens entre les deux lexies s'exprime ici par des ensembles de formes fléchies différentes. C'est ce genre de limitation qui a poussé les concepteurs d' ALEXIA à sélectionner *a priori* des ensembles de phrases contenant les lexies relevant du monde du travail, rendant la génération d'activités semi-automatique [Selva & Chanier, 2000, § 6]. Avec MIRTO, Antoniadis *et al.* cherchent à systématiser ce type de fonctionnement par une architecture qui permette à des équipes concepteurs d'activités (cumulant compétences didactiques et techniques) d'ajouter des scripts de génération automatique d'activités configurables par des enseignants (n'ayant pas nécessairement de compétences techniques), s'attaquant ainsi à la troisième limite citée [Antoniadis *et al.*, 2005b]. Pour y parvenir, le TAL n'est pas suffisant, ils s'appuient sur une structure modulaire et sur un travail de modélisation des activités faisant travailler de concert informaticiens et didacticiens.

d) TAL et indexation pédagogique

Pour la réalisation de notre base de textes, il paraît primordial de prendre en compte ces limites perçues dans les systèmes d'ALAO. L'absence de sens associé aux formes empêcherait toute recherche prenant en compte le contenu des ressources textuelles, autrement que par des recherches de sous-chaines de textes. Plus nous pourrions ajouter d'informations aux textes et plus les recherches pourront être précises. Il convient cependant de ne pas oublier la troisième limite que nous avons reprise : les enseignants, s'ils travaillent sur l'apprentissage de la langue, n'ont pas pour autant un point de vue de linguiste et la manière d'interpréter les informations ajoutées aux textes doit s'appuyer sur leur(s) problématique(s). Il ne suffit pas d'ajouter des informations issues d'outils TAL à des textes pour parvenir à les indexer pédagogiquement. Une partie du problème est de décider quelles informations sont pertinentes pour les enseignants et sous quelle forme, ce qui nous ramène à la nécessité d'évaluer leurs besoins. Enfin, pour notre système comme pour la recherche en ALAO en général, l'évolutivité est une caractéristique nécessaire. Comme dans le processus de développement de Depover *et al.* [2000], c'est par une démarche itérative que nous pouvons espérer arriver à des formes d'indexation pédagogique satisfaisantes pour les enseignants, par d'incessants retours entre la théorie (conceptualisation, modélisation de l'indexation pédagogique) et la pratique (confrontation de la théorie aux pratiques des enseignants). Un tel processus, dans le domaine de l'informatique, *a fortiori* de l'ALAO, n'est possible que par la conception de systèmes évolutifs.

Pour appréhender la problématique de l'utilisation du TAL dans l'ALAO, nous reprendrons les questions posées par [Zock, 1996, p. 1002] :

les informations lexicales qui lui sont liées » [Bouillon *et al.*, 1998].

57. <http://fr.wiktionary.org/wiki/manger>

1. « *what can current Natural Language Processing (NLP) technology contribute to CALL ?*⁵⁸ »
2. « *how can this technology meet the demands of pedagogical theory for communicative language teaching in a natural environment?*⁵⁹ »
3. « *what can NLP-based systems teach us about language acquisition, linguistic theory and NATURAL language processing in general?*⁶⁰ »
4. « *what effect can a domain like CALL, or the involved disciplines have on the development of NLP technology?*⁶¹ »
5. « *what lessons have been, or can be learned by looking at the available CALL systems?*⁶² »

Notre travail, dans le cadre de cette thèse, se concentre principalement sur les deux premières questions. Il ne s'agit pas de développer de nouveaux modèles ou de nouveaux systèmes pour le TAL (question 4) mais d'examiner les possibilités offertes par les outils qu'il propose (question 1). Ces possibilités d'analyse, et donc d'informations, devront être prises en compte dans la modélisation des textes et de leur contenu, mais à la lumière de la manière dont elles pourront être interprétées pour remplir un besoin d'origine didactique. Que ce soit pour l'ALAO en général, ou notre travail en particulier, le TAL ne doit pas être vu comme la panacée, mais comme l'un des champs disciplinaires qui touchent au domaine (comme l'IHM ou la sociolinguistique, par exemple [Heift & Schulze, 2007, p. 10]), proposant des outils et des concepts donnant accès à des informations qui pourront être utilisées par l'ordinateur à destination des utilisateurs (enseignants, concepteurs ou apprenants).

2.2.4 Perception et usage

Lors de la journée d'étude TALAL⁶³, un chercheur, semble-t-il extrêmement expérimenté, prit la parole après une communication pour signifier avec véhémence qu'après avoir « réinventé la langue dans notre coin avec le TAL », nous autres, informaticiens, nous apprêtions à faire de même avec l'ALAO. La réponse qui lui fut faite, se contenta de reprendre le point de vue du communicant, qui disait vouloir travailler en accord avec les enseignants de langue. Mais finalement, la question portait plus sur la perception de la technologie en vue de son évaluation et de son usage, que sur la « réinvention » de l'objet inhérente à la tâche de modélisation, qui vise à représenter, présenter à nouveau des objets pour les rendre manipulables.

Que ce soit le TAL du point de vue des concepteurs, ou l'ALAO en général du point de vue des utilisateurs, un « bon » usage des technologies semble passer par le dépassement des affects qui leurs sont liés.

58. Quelles technologies TAL existantes peuvent contribuer à l'ALAO ?

59. Comment ces technologies peuvent-elles accéder aux demandes des théories pédagogiques pour l'enseignement des langues dans un environnement naturel ?

60. Que peuvent nous enseigner les systèmes utilisant le TAL sur l'acquisition langagières, les théories linguistiques et le traitement des langues NATURELLES en général ?

61. Quels effets est-ce qu'un domaine comme l'ALAO ou ses disciplines connexes peuvent avoir sur le développement de technologies TAL ?

62. Quelles leçons ont été ou peuvent être tirées d'une étude des systèmes d'ALAO disponibles ? N.B. : Les remarques exposées dans ce paragraphe (§ *Modélisation* p. 53), répondent partiellement à cette question.

63. http://alsic.u-strasbg.fr/v07/lamadon/alsic_v07_12-poi1.htm

a) Limites du TAL

L'exemple du système ALEXIA (cf. § *Le TAL pour l'ALAO* p. 55), tel que nous l'avons décrit expose certaines limites du TAL. Alors qu'en 1954, quand le champ débutait avec l'expérience « Georgetown-IBM », les chercheurs impliqués suggéraient que dans un délai de quelques années, le problème de la traduction automatique serait réglé [Wikipedia, 2008c], la communauté TAL, malgré un « penchant pour l'*hubris* » [Kraif *et al.*, 2004], est bien plus consciente de ses limites (au moins actuelles) et plus encline à suivre le conseil bien connu de Bar-Hillel de faire un usage judicieux et modeste des aides automatisées⁶⁴ [Bar-Hillel, 1964, p. 183]. Cet usage modeste et judicieux nous ramène à la stratégie moins-disante que nous évoquions dans le paragraphe a) (p. 11). Cette stratégie passe par la compréhension des limites des outils, spécialement dans le contexte de l'ALAO où l'un des objectifs de l'usage du TAL est précisément d'éviter les rétroactions erronées. Nous pouvons par exemple citer certains travaux (Kraif & Ponton [2007]; Blanchard *et al.* [2009]), qui visent à proposer des méthodes pour analyser les réponses des étudiants, prenant en compte ces limitations. Plutôt que de tenter d'analyser des productions libres d'apprenants, leur approche est de restreindre le contexte afin de rendre plus fiables les analyses effectuées et, par voie de conséquence, produire un feed-back plus adapté. Une stratégie qu'ils résument en un *leitmotiv* : « Faire peu et simplement, mais faire bien⁶⁵ » [Blanchard *et al.*, 2009]. Ils rejoignent d'une certaine manière la démarche de Selva & Chanier [2000] qui, conscients que « la désambiguïsation sémantique ne peut se faire automatiquement[, les] programmes qui lèvent les ambiguïtés [n'étant] pas assez fiables », décident de procéder manuellement pour cette étape du traitement. Même si dans notre cas les erreurs potentielles seront moins lourdes de conséquences, puisqu'avant d'être proposé aux apprenants chaque texte passera par le filtre de l'enseignant, le manque de fiabilité des outils peut nuire à l'utilisation du système, affectant directement précision et rappel. Il nous faut donc, nous aussi, jongler avec « bruit et silence » (cf. Kraif & Ponton [2007]; Blanchard *et al.* [2009]).

b) *Fear, awe and 'wow'*

Il est tentant d'établir un parallèle entre la manière dont les enseignants perçoivent les systèmes d'ALAO et les attentes qu'a pu susciter le TAL. L'avis des spécialistes de l'ALAO à l'égard des systèmes existants est globalement négatif :

*« Multimedia language learning materials comprise, as with all learning resources, some gems amongst a lot of mediocrity. »*⁶⁶

[Murray & Barnes, 1998, p. 259]

En introduction à des considérations méthodologiques, Hubbard parle d'un haut niveau de mécontentement chez les enseignants de langues par rapport aux systèmes existants⁶⁷ [Hubbard, 1996]. Malgré tout, ces mêmes enseignants, quand ils sont programmeurs, sont d'après Shepherd, conquis *a priori* par les aspects techniques des outils. Cela débouche sur l'apparition de programmes qui reflètent une méthodologie qui serait rejetée par les enseignants qui ont écrit les

64. « *judicious and modest use of mechanical aids* »

65. « *Doing few and simply, but doing well* »

66. Le matériel multimedia d'apprentissage des langues, comprend comme c'est le cas pour toute ressource d'apprentissage, quelques bijoux parmi un flot de médiocrité.

67. « *Given the high level of dissatisfaction many teachers have for much of the present courseware* », cité dans [Rézeau, 2001, p. 324]

programmes s'ils les voyaient sur papier⁶⁸ [Shepherd, 1987]. Bax désigne ce rapport à l'ALAO par les termes « *fear* » et « *awe* »⁶⁹ [Bax, 2003, p. 25]. Cette attitude ambivalente de crainte et d'admiration est pour Bax une étape dans le processus de normalisation/intégration⁷⁰ d'une technologie. Il considère que ce comportement de *fear and awe* est l'attitude dominante en 2003 vis-à-vis de l'ALAO.

Ce concept de *fear and awe* est très cohérent avec le « "wow" factor⁷¹ » de Murray & Barnes [1998], qui veut décrire aussi bien les réactions initiales d'utilisateurs (enseignants/apprenants) extrêmement positives que celles qui sont extrêmement négatives. Pour les auteurs, cette évaluation instinctive peut affecter l'opinion globale qu'aura l'utilisateur du programme à moyen ou long terme [Murray & Barnes, 1998, p. 250]⁷². Le point de vue de Murray et Barnes est celui de l'évaluation. Il s'agit de faire prendre conscience aux enseignants du « facteur "waou" » et de proposer ensuite un cadre d'évaluation alors que Bax veut décrire l'attitude des utilisateurs envers les systèmes d'ALAO. Dans un cas comme dans l'autre, ils expliquent que l'un des éléments intervenant dans ces attitudes dépend directement des attentes des utilisateurs vis-à-vis de l'outil informatique dans le contexte de l'enseignement des langues :

« [Some of the] causes can originate from teacher/learner hostility (technophobia), a high learning curve [...], and disillusionment after anticipation of an educational panacea.⁷³ »

[Murray & Barnes, 1998, p. 250]

« Instead of addressing the role which the software could play within the wider classroom context (a small role, but a useful one) their expectation seemed to be that it should either do everything and replace current technologies such as dictionaries and even the teacher, or it was not useful.⁷⁴ »

[Bax, 2003, p. 25]

Le rôle que peut jouer l'ALAO dans le contexte de la classe pour Bax (petit, mais utile) n'est pas sans rappeler le *Leitmotiv* emprunté à Blanchard *et al.* et leur stratégie moins-disante.

68. « *appearance of programs reflecting methodology which would be rejected by the very teachers who wrote the programs if they were seen in print* » cité dans [Rézeau, 2001, p. 322]

69. *fear* : peur / *awe* : difficile à traduire, admiration et fascination, mais dans l'expression figée *fear and awe*, *awe* peut vouloir dire « respect, mêlé de crainte » (Collins : <http://dictionary.reverso.net/english-french/awe>).

70. Nous ajoutons ici le terme intégration, car il parle de normalisation dans le sens de « devenir normal ». Dans ce contexte, une technologie est « normalisée », quand elle est tellement intégrée à nos vies qu'elle est devenue invisible : « *The technology is so integrated into our lives that it becomes invisible* » [Bax, 2003, p. 25], comme le livre, que nous citons précédemment (cf. § 2.1.6 p. 44). Il ne parle pas ici du processus de création d'une norme (*standardization* en anglais) que nous discutons plus loin (cf. § 3.1.1 p. 88).

71. facteur « waou »

72. « *The "wow" factor encompasses both extremely positive and extremely negative initial reactions in the user (teacher/learner) towards a software package. This immediate, instinctive evaluation can colour the user's opinion of the program as a whole, even on a medium- to long-term basis.* »

73. Certaines causes peuvent provenir de l'hostilité (technophobie), de l'enseignant/apprenant, des difficultés de prise en main du logiciel, ou de la désillusion qui suit l'attente d'une panacée éducative.

74. Au lieu d'adresser le problème du rôle que doit jouer un logiciel dans le contexte plus large de la classe (un petit rôle, mais un rôle utile), leurs attentes semble être que le logiciel doit soit tout faire et remplacer les technologies actuelles, telles que les dictionnaires ou même l'ordinateur, ou alors qu'il n'est pas utile.

c) Approche instrumentale

Ces deux points de vues sont analogues mais dans des contextes différents ; Bax s'intéresse à l'usage de l'ALAO en classe alors que Blanchard, Kraif et Ponton se focalisent sur l'usage du TAL dans la conception de ressources ALAO. Cependant leurs recommandations sont les mêmes : comprendre les outils, leurs limites et leurs points forts et en choisir des usages qui focalisent sur leurs points forts en essayant d'en éviter les faiblesses. Ce qui nous amène à la question des usages : « Il y a souvent [...] un décalage entre l'usage prescrit et l'usage réel » [Boulidoires, 1998]. Vitalis étaye ce point de vue, en parlant, lui, d'« usage social » :

« La formation d'un usage social prend du temps et ne correspond pas toujours à ce qu'avaient imaginé les promoteurs de systèmes. De manière plus subtile et plus dissimulée, les pratiques d'utilisation au quotidien nous mettent en présence d'un ensemble de manières de faire particulières, avec leurs ruses et leurs braconnages. Les analyses sont ici particulièrement difficiles dans la mesure où ce type d'activité ne se signale pas au grand jour par des produits particuliers mais par un art d'utiliser ceux qui lui sont offerts ou imposés. »

[Vitalis, 1993, p. 37]

Contamines *et al.* se centrent plus sur les pratiques du point de vue de l'outil. Ils reprennent l'approche instrumentale définie par Rabardel qui « repose sur la distinction fondamentale existante entre, d'une part, l'objet matériel [*i.e.* l'outil ou l'artefact] et, d'autre part, l'objet matériel inscrit effectivement et efficacement dans un usage [*i.e.* l'instrument] » [Contamines *et al.*, 2003, p. 162]. Un logiciel d'ALAO est un artefact et, comme à tout artefact, un mode opératoire lui est associé : des fonctionnalités prévues par les concepteurs. Ce sont les « fonctions constitutives » de l'objet ; elles correspondent à son usage prescrit. Des fonctions constituées émergent ensuite à l'usage. Le processus au cours duquel un artefact devient un instrument s'appelle la « genèse instrumentale », qui peut se faire selon deux processus interdépendants :

- l'instrumentation : « [...] c'est l'utilisateur qui évolue, qui apprend. Ce sont les schèmes d'usage qui évoluent, se transforment, sont créés, s'incorporent aux schèmes déjà existants » [Contamines *et al.*, 2003, p. 163]. Elle correspond aux usages sociaux de Vitalis.
- l'instrumentalisation : « [...] c'est l'artefact qui évolue. Ce sont l'émergence de nouvelles propriétés fonctionnelles pour l'artefact - fonctions constituées - et l'institution des fonctions constitutives par le sujet » [Contamines *et al.*, 2003, p. 163].

Un exemple de genèse instrumentale pourrait être celui des concordanciers. Nous allons pour les besoins de cet exemple réduire l'histoire à son minimum : nous adoptons de nombreux raccourcis comme le fait de ne pas parler de linguistique de corpus. Mais l'objectif ici n'est pas de retracer l'histoire des concordanciers, des concordances de la Bible à nos jours, mais bien de donner un exemple d'instrument à travers le prisme de la genèse instrumentale.

La dénomination *KeyWord In Context* (KWIC) est attribuée à Luhn [Manning & Schütze, 1999, p. 35], qui l'utilisait dans le contexte de l'indexation d'articles scientifiques [Luhn, 1960]. L'idée venait du fait qu'un processus automatique d'indexation permettrait de gagner en efficacité dans l'optique de pouvoir suivre la cadence des productions scientifiques. L'indexation automatique à partir de mots-clés présents dans le texte était automatisable ; Luhn suit donc cette piste. Le sens de ces mots-clés pouvait être affecté par le contexte dans lequel ils étaient employés, Luhn introduit donc la notation KWIC afin de donner accès à un expert du domaine à l'acception adéquate de chaque terme. L'index se présente comme dans la figure 2.3 où, pour chaque mot présent dans l'index, un contexte est fourni. L'ordinateur n'est pas chargé ici de « comprendre » la langue, mais de fournir à l'utilisateur les informations nécessaires à la compréhension. Pour

une utilisation comme index, Luhn suggère une présentation limitée à 60 caractères.

KEYWORD-IN-CONTEXT BIBLIOGRAPHICAL INDEX

COULOMB EXCHANGE ENERGY FROM SHELL-MODEL WAV	1719
OF ATOMIC AND MOLECULAR EXCITATION OF PROTONS IN HELIUM II B	0011
EXCITATION BY A TRAPPED-ELECTRON THE	0150
THERMAL EXCITATIONS IN LIQUID HE3.	1465
ENERGIES OF GROUND AND EXCITED NUCLEAR CONFIGURATIONS IN TH	0452
EXCITED STATES OF V51 AND CR53.	1691
4-PLUS EXCITED STATE IN OSMIUM-188.	1717
INTERNAL PHOTOEFFECT AND EXCITON DIFFUSION IN CADMIUM AND ZIN	0123
OF THE CONTRIBUTION OF EXCITONS TO THE COMPLEX DIELECTRIC	1555
THERMAL EXPANSION OF SOME CRYSTALS WITH THE	0136
ENERGY LEVELS IN F18 FROM THE N14/ALPHA, ALPHA/N14 AND	0547
ON FROM AL27-PLUS-P AND F19-PLUS-P.	0239
TIC MEASUREMENTS OF THE FE-CR SPINELS.	1603
BARIUM FERRATE III.	0326
MAGNETOSTATIC MODES IN FERRIMAGNETIC SPHERES.	0059
NICKEL-IRON FERRITE.	0397

FIGURE 2.3: Index bibliographique KWIC [Luhn, 1960, Figure 1 p. 290]

L'usage d'un système d'indexation pour n'indexer qu'un mot et ainsi obtenir tous ses contextes est un « braconnage » correspondant à un processus d'instrumentation : les enseignants de langue prennent conscience de l'intérêt de cette notation pour l'enseignement du vocabulaire par exemple [Johns, 1986, p. 160] en insistant sur le développement de stratégies pour deviner le sens de mots inconnus à partir d'indices contextuels. Ils s'approprient l'artefact par l'usage et en font ainsi un instrument.

Modifier l'outil d'indexation de Luhn pour en faire un véritable concordancier, permettant d'accéder à toutes les occurrences d'une forme dans un corpus et proposant des lignes non plus de 60 caractères mais de 74 ou 130 caractères (cf. figure 2.4), serait une instrumentalisation de l'artefact original.

on	
1) Purex process depends heavily on	the design of the solvent-extraction appa
2) ssing plants. An improvement on	both the extraction column and the mixer-
3) centrifugal separator mounted on	the same shaft as the mixing vanes Typic
4) nium. They designed the plant on	the principle that the equipment inside t
5) essing plant, this time based on	Purex solvent instead of Butex. With a ca
6) 0 tons of spent fuel per year on	a site owned by the state of New York in
7) on of the chain-reacting pile on	December 2, 1942, and the explosion of th
8) n of the first plutonium bomb on	July 16, 1945. The important legacy of H
9) located at standard positions on	the inside and near the top of the canyon
10) led the length of the canyons on	rails. The crane operator, protected by h
11) s all the uranium. This play on	oxidation states gave rise to the name Re
12) raction process to be applied on	a large scale. The Redox process, with He
as	
1) backbone of structural engineering. A beam can be defined as a structural member that offers resistance to bending caused by app	
2) e simple types of beams. Beams should not be just regarded as rectangular cross-section bars of wood or metal. A beam can be of	
3) refore be a layer which is the same length in the bent beam as in the unbent beam. This layer, or surface, is called the neutral	
4) Because this neutral surface is the same length under load as with no load it is under no strain and hence no stress. Thus while	
5) aces are the highest stress regions it is necessary to have as much material in these regions as possible if the beam is to bend a	
6) s it is necessary to have as much material in these regions as possible if the beam is to bend as little as possible under the act	
7) aterial in these regions as possible if the beam is to bend as little as possible under the action of a load. There is no great n	
8) these regions as possible if the beam is to bend as little as possible under the action of a load. There is no great need for ea	
9) here is no great need for material near the neutral surface as there is very little stress there. Hence the widespread use of the	
10) metallic form needed for weapons. If plutonium is ever used as fuel in nuclear power plants, plutonium oxide would be the preferre	
11) spent fuels from reactors using highly enriched uranium such as the reactors of nuclear submarines) in which only traces of plutoni	
12) at Hanford and Savannah River in being designed for direct (as opposed to remote) maintenance. The process equipment must be cheap	

FIGURE 2.4: Concordances présentées par MICRO-CONCORD [Johns, 1986, p. 151]

Bien sûr, il ne s'agit pas réellement du même outil matériel entre les articles de Luhn [1960] et Johns [1986]. Nous passons sous silence les développements en lexicographie et construction de dictionnaires (Collins Cobuild), en stylistique et recherche littéraire, ou en linguistique [Flowerdew, 1993, p. 87]. Nous ne nous en servons que comme exemple de genèse instrumentale, pour clarifier les définitions données ci-dessus. La genèse instrumentale étend le champ instrumental

de l'artefact. Ce dernier couvre, dans cet exemple peu rigoureux, aussi bien l'indexation d'articles scientifiques que l'enseignement du vocabulaire. En effet, le champ instrumental représente « l'ensemble des valeurs fonctionnelles et subjectives que l'artefact peut potentiellement prendre au sein de l'activité d'un individu » [Contamines *et al.*, 2003, p. 164].

d) Perception et instrumentation

L'approche instrumentale décrite par Contamines *et al.* [2003] vise à prendre en compte l'évolution des outils en ne les considérant pas comme figés, mais comme des artefacts qui existent à travers les usages qui en sont faits. Si nous revenons à l'historique des EIAH décrit plus tôt, ce point de vue a été adopté à travers l'utilisation de l'ordinateur comme outil (cf. § c) p. 32). Il a été largement développé dans la communauté de l'ALAO, notamment dans les années 80 : Higgins [1986] ou Jones [1986], qui cite un certain nombre de logiciels, pas forcément conçus pour l'enseignement des langues, mais en propose des utilisations pertinentes dans le contexte de cet enseignement et conclut avec les lignes directrices suivantes :

*« Try it and see what happens. Don't pre-judge.
Don't expect the program to do all the work.
If things don't work out, don't automatically blame the program. The problem may lie elsewhere.
Above all, use your imagination.⁷⁵ »*

[Jones, 1986, p. 178]

Ces préceptes correspondent au point de vue défendu par Bax concernant le rôle que doit avoir l'ALAO dans la classe. Ils sont cohérents avec le fait que la genèse instrumentale des outils de l'ALAO passe par une meilleure intégration de l'outil informatique qui doit permettre de dépasser le « *Wow* » factor » et autre « *fear and awe* », condition préalable à une réelle appropriation des outils.

2.2.5 Conclusion

L'étude de certains travers de l'ALAO permet d'établir que le problème de conception qui est le nôtre – la création d'une base de textes indexée pédagogiquement pour l'enseignement des langues – devait s'appuyer sur un processus faisant intervenir les enseignants. Elle confirme, du point de vue de l'ALAO, la démarche empirique que nous avons adoptée au regard d'éléments de didactique des langues, nécessaire à la prise en compte de la problématique des enseignants. Cette dernière demande de considérer les textes comme de la langue, et non uniquement comme une suite de caractères dénuée de sens. Pour ce faire, le TAL est l'un des outils à notre disposition, il ne doit pas éclipser les autres ni être perçu comme une solution miracle dont la seule intégration à notre système permettra d'éviter tous les écueils, comme en témoigne le constat de Gamper et Knapp sur l'intelligence artificielle cité dans le chapitre 1 :

« Artificial Intelligence offers many possibilities to improve computer-assisted language learning systems. However, the application of most of these technologies is not

75. Essayez, et voyez ce qui se passe. Ne pré-jugez pas.

Ne vous attendez pas à ce que le programme fasse tout le travail.

Si cela ne fonctionne pas, ne rejetez pas systématiquement la faute sur le programme. Le problème peut se situer ailleurs.

Par dessus tout, utilisez votre imagination.

*mature yet and still requires more research.*⁷⁶ »

[Gamper & Knapp, 2002, p. 338-339]

Cependant, le TAL semble être un atout pertinent dès lors que l'on est suffisamment conscient de ses limites pour en faire, dans notre cas, un instrument au service de l'indexation pédagogique (instrumentation – son instrumentalisation correspondrait plus à l'effet que peut avoir l'ALAO sur le TAL, Zock [1996]). La compréhension des limites de l'outil est aussi une condition nécessaire pour la bonne utilisation de la base de textes et, éventuellement à plus long terme, le développement de fonctions constituées, dans une perspective d'intégration des outils aux pratiques en classe suggérée par Bax [2003]. C'est pourquoi la stratégie moins-disante doit être claire dès les premières phases de tests du prototype. Cependant, le travail effectué dans le cadre de cette thèse s'arrête à la première itération du processus de développement, après la réalisation d'une maquette utilisant le modèle que devons concevoir.

2.3 Corpus et enseignement des langues

Les conclusions que nous avons tirées jusqu'ici, bien qu'essentielles pour la suite de notre travail d'un point de vue méthodologique, pourraient être adaptées à bien d'autres problématiques de l'ALAO que la nôtre. Dans ce paragraphe, nous nous concentrons plus précisément sur la place des textes dans notre travail, à travers un survol de la problématique de l'usage des corpus dans l'enseignement des langues. Nous irons, après avoir exploré rapidement la notion d'authenticité, dans la direction des corpus, avant d'en constater l'instrumentation dans l'enseignement des langues. Ceci nous donnera l'occasion de revenir sur l'exemple des concordanciers, puis de constater les premières instrumentalisations. Nous tenterons de nous positionner par rapport à ces différentes initiatives.

2.3.1 Texte authentique

« La mode des documents authentiques s'est développée, en didactique des langues, essentiellement en réaction à la relative "artificialité" des dialogues des manuels audio-oraux et audiovisuels » [Besse, 1987, p. 184]. Comme nous le faisons remarquer dans notre chapitre introductif (§ 1.4.1 p. 18), le texte authentique, sans en être l'apanage (cf. tableau 1.1), occupe une place centrale dans l'approche communicative. L'avènement du texte authentique est intimement lié au développement de cette approche⁷⁷ [Little & Singleton, 1988, p. 1], la plus utilisée (cf. tableau 1.1). L'importance que revêt le texte authentique s'impose d'autant plus que l'approche communicative n'est pas la seule à y avoir recours (cf. tableau 1.1). Une base de textes semble donc être un moyen de faciliter l'accès aux textes authentiques par les enseignants. Même si nous avons déjà stipulé que le contenu éventuel de la base ne relevait pas strictement de notre problématique (cf. § *Ajout d'un texte* p. 8), les « questions du groupe d'utilisateur » par rapport aux textes, et parmi elles celle de l'authenticité, sont au centre de nos préoccupations (cf. § *Recherche d'un texte* p. 9). En effet, les références vantant les avantages du texte authentique par rapport à d'autres formes (sans pour autant exclure ces dernières) sont nombreuses ; une rapide recherche bibliographique en donne des exemples :

76. Traduction : voir note ¹² p. 11.

77. « *The rise of the "authentic text" is closely associated with the development of communicative approaches to language teaching.* »

- « *Experience suggests that learners find [“authentic” materials] more interesting than traditional course books and thus more motivating*⁷⁸ » [Little & Singleton, 1988, p. 1] ;
- « *[Research] into first and second language acquisition and language processing suggests that “authentic texts” will more effectively foster and sustain unconscious acquisition than invented texts*⁷⁹ » [Little & Singleton, 1988, p. 1] ;
- « Mais on peut considérer aussi que le texte original [authentique] est plus formateur qu’un texte adapté, parfois même plus aisément lisible. » [Vigner, 2001, p. 54] ;
- « *[In] principle “authentic texts” provide an ideal focus for the development of learner autonomy.*⁸⁰ » [Little & Singleton, 1988, p. 2] ;
- « *Proponents of authentic materials point out that classroom texts and dialogues do not adequately prepare learners for coping with the language they hear and read in the real world outside the classroom*⁸¹ » [Nunan, 1989, p. 54].

Il faut signaler que si l’usage de textes authentiques est très en vogue, les mêmes chercheurs concèdent volontiers qu’ils ne dispensent pas de l’usage de matériel plus traditionnel [Little & Singleton, 1988] ou qu’ils ne sont pas toujours utilisables [Nunan, 1989, p. 138]. Si nous nous intéressons en premier lieu à cette caractéristique, c’est qu’elle représente un cas d’application concret de pratiques existantes :

*« El potencial de Internet en la enseñanza ha sido y es objeto de numerosos debates y un buen número de profesores ven en él actualmente una herramienta de acceso a la información y a un material lingüístico auténtico. »*⁸²

[Sanz, 2003, p. 142]

Dans ce contexte, le but de notre travail est de proposer un modèle prenant en compte la problématique de l’enseignant (recherche de ressources authentiques pour l’enseignement des langues). Bien sûr, il dépendra des ressources indexées et nous ne pouvons en cela comparer l’objet de notre travail avec la quantité d’information accessible par internet⁸³. Cependant pour un ensemble de textes restreint, notre modèle doit adapter les possibilités de recherches aux enseignants. Enfin, ce point d’entrée va nous permettre d’aborder les corpus et avec eux, de nous positionner par rapport aux systèmes d’ALAO s’approchant le plus de notre travail.

a) Définition

Dans le paragraphe *L’approche communicative* (p. 18), nous reprenons une définition de document authentique de Germain [1993] : « non expressément conçu pour être [utilisé] dans une classe de langue ». Lorsqu’il définit cette notion, l’authenticité n’est pas au centre de son propos ; Germain propose une définition rapide pour que le lecteur comprenne globalement de quoi il s’agit. Elle n’est cependant pas suffisante pour définir précisément la notion. Pour être claire, elle devrait

78. L’expérience suggère que les apprenants trouvent le matériel « authentique » plus intéressant que les manuels traditionnels et de ce fait plus motivants.

79. Les recherches en acquisition des L1 et L2 et en sciences cognitives suggèrent que des textes authentiques vont favoriser et éveiller plus efficacement l’acquisition inconsciente que des textes inventés.

80. En principe, les textes authentiques fournissent une perspective idéale pour le développement de l’autonomie de l’apprenant.

81. Les défenseurs du matériel authentique font remarquer que les textes et dialogues de la classe ne préparent pas correctement les apprenants à faire face avec le langage qu’ils entendent et lisent dans le monde réel en dehors de la classe

82. Le potentiel d’Internet dans l’enseignement a été, et est, l’objet de nombreux débats et bon nombre d’enseignants y voient un outil d’accès à l’information et à un matériel linguistique authentique.

83. Même si en théorie, il pourrait permettre d’indexer les textes qu’il contient.

préciser ce que signifie concevoir un document. Un texte proposé par un enseignant dans une version écourtée est-il authentique ? Dans quelle mesure un texte dont certaines phrases ont été modifiées pour le rendre accessible aux apprenants est-il authentique ? Est-ce l'écriture du texte par son auteur ou la dernière modification qui fait office de conception ? Fournir le texte avec des ajouts, comme une liste de vocabulaire, affecte-t-il l'authenticité du texte ?

Pour éclaircir la situation, Taylor [1994] distingue plusieurs types d'authenticité :

- l'authenticité de la langue ;
- l'authenticité de la tâche ;
- et l'authenticité de la situation.

D'après [Tatsuki, 2006], l'authenticité de la langue de Taylor correspondrait à deux des quatre types d'authenticité de Breen : l'authenticité du texte utilisé comme *input* pour les apprenants⁸⁴ et l'authenticité des interprétations des apprenants de tels textes⁸⁵ [Breen, 1985, p. 61]. Les deux autres types se retrouvent chez Taylor en des termes analogues.

Comme le signale Taylor, il faut spécifier le type d'authenticité dont il s'agit, sinon la position de Hutchinson et Waters s'impose, aussi bien dans le domaine de l'apprentissage d'une L2 en général que dans le domaine de la langue de spécialité pour lequel elle a été spécifiée [Taylor, 1994, p. 2] :

« [*“Authentic”*] usually carries the sense of “taken from the target situation and, therefore, not originally constructed for language teaching purposes”. In reality this is a contradiction of the term authentic. Authenticity is not a characteristic of the text in itself: it is a feature of a text in a particular context. [...] A text can only be truly authentic, in other words, in the context for which it was originally written. Since in English for Specific Purposes (ESP) any text is automatically removed from its original context, there can be no such thing as an authentic text in ESP.⁸⁶ »

[Hutchinson & Waters, 1987, p. 159]

Besse fait le même constat : « le document d'origine authentique est toujours plus ou moins “désauthentifé” par l'usage didactique qu'on en fait, puisqu'il est choisi et traité à des fins qui ne sont pas celles pour lesquelles il a été primitivement produit » [Besse, 1987, p. 183]. Le fait de ne pas spécifier le type d'authenticité concerné légitime toutes ces critiques, ainsi que celle de Bouchard :

« Il est vrai que le texte non fabriqué à des fins pédagogiques se réalise sous la forme d'un objet textuel ayant des caractéristiques matérielles précises qui influent, et sur sa lisibilité et sur son interprétation. Mais il est vrai aussi que, dès qu'on “sort” un tel texte de la situation de réception pour laquelle il est fait, on évacue une part importante de son authenticité. »

[Bouchard, 1990, p. 42]

La définition que donne Germain correspond donc à l'authenticité de la langue et plus particulièrement à « l'authenticité du texte utilisé comme *input* pour les apprenants⁸⁴ » [Breen, 1985, p. 61]. Elle est en accord avec la définition informelle que donne Nunan à authentique pour décrire

84. « *Authenticity of text used as input data for learners* »

85. « *Authenticity of the learners own intreprétation of such texts* »

86. « Authentique » revêt en général le sens de « tiré de la situation cible, et de ce fait, non expressément conçu avec un objectif d'enseignement des langues ». En réalité, c'est une contradiction du terme authentique. L'authenticité n'est pas une caractéristique du texte lui-même : c'est un trait d'un texte dans un contexte précis. [...] En d'autres termes, un texte ne peut être vraiment authentique que dans son contexte original. Dans la mesure où, en anglais langue de spécialité, tout texte est automatiquement enlevé de son contexte d'origine, aucun texte authentique ne peut y exister.

tout matériel pédagogique : « en règle générale, authentique dans ce cas là correspond à n'importe quel matériau qui n'a pas été spécifiquement produit pour l'enseignement des langues⁸⁷ » [Nunan, 1989, p. 54].

Dans la suite de ce mémoire, nous nous référerons à cette définition quand nous emploierons le terme « authentique » pour parler d'une ressource pédagogique. À l'instar de Vigner, nous différencions le texte authentique du texte fabriqué, qui comme son nom l'indique est écrit en vue de son utilisation dans l'enseignement des langues [Vigner, 2001, p. 53]. Mais entre les deux, il dégage plusieurs caractéristiques qui font que le texte n'est ni authentique, ni à proprement parler fabriqué. Un texte d'origine authentique peut, toujours dans cette optique d'authenticité de l'*input*, être altéré dans sa forme avant d'être utilisé par les apprenants. Ces altérations peuvent être de deux types : l'abrègement et l'adaptation. Dans le premier cas il s'agit d'une opposition entre texte intégral et texte abrégé [Vigner, 2001, p. 54]. Il s'agit toujours d'instances attestées de la langue, mais le texte n'est plus présenté sous forme d'« unités rhétoriques complexes » [Widdowson, 1981, p. 93]. L'adaptation d'un texte peut consister à « [le réduire] (on supprime certaines phrases, certains groupes) ou par la traduction, s'il s'agit d'un texte au départ écrit dans une autre langue que [la L2 enseignée] » [Vigner, 2001, p. 54]. Les « versions simplifiées » de Widdowson intègrent ces possibilités. Elles sont « des passages dérivés d'exemples réels de discours au moyen d'un procédé de substitution lexicale et syntaxique ». L'auteur les décrit en termes de traduction mais d'un autre genre : « En effet, elles intègrent [des gloses] afin de produire une version qui est censée correspondre au niveau de compétence linguistique de l'apprenant. Il s'agit donc essentiellement d'une sorte de traduction de l'usage propre à l'apprenant » [Widdowson, 1981, p. 102].

b) Authenticité et ALAO

Ayant défini ce que nous entendions par « authentique » dans le cadre de notre problématique, celle des textes pour l'enseignement des langues, nous pouvons aborder la question dans le cadre de l'ALAO. Dans la base d'articles concernant l'ALAO de Jung, 98% des instances du descripteur « authenticité » proviennent d'articles écrits après 1992 [Jung, 2005, p. 12]. En 2000, Harris constate que malgré l'importance prise par les ressources authentiques comme source d'*input* pour l'acquisition d'une L2, elles ont brillé par leur absence des systèmes d'ALAO [Harris, 2000, p. 4]. Cela n'exclut pas l'authenticité de la recherche du domaine, comme le signale sa présence dans la base d'articles de Jung ou les recommandations de Murray et Barnes qui signalent que beaucoup de textes utilisés dans des logiciels étaient peu convaincants. Pour eux, le but devrait être d'utiliser des textes aussi authentiques que possible⁸⁸[Murray & Barnes, 1998, p. 253]. Notons tout d'abord que la formulation suggère, tout comme la remarque de Vigner selon laquelle « le texte adapté est plus proche du texte authentique que le texte fabriqué » [Vigner, 2001, p. 54], de voir l'authenticité non comme une propriété binaire, mais comme un continuum. Ils rejoignent en cela des travaux cités dans [Harris, 2000, p. 6] (Davies, Lautamatti).

De notre point de vue, cette remarque de Murray et Barnes, met en exergue l'un des rôles potentiels d'une base de textes indexée pédagogiquement pour l'enseignement des langues, à savoir la possibilité de fournir des textes authentiques à des systèmes d'ALAO (cf. § 1.2.2 p. 9,

87. « A rule of thumb for authentic here is any material which has not been specifically produced for the purposes of language teaching. »

88. « Many texts used in software have been rather contrived; the aim should be that the texts are as near authentic as possible ».

utilisateur «system»). La plupart des systèmes d'ALAO qui intègrent du matériel authentique le font par le biais des corpus. Nous allons constater dans les prochains paragraphes en quoi notre problématique se démarque de ces domaines.

2.3.2 Corpus et collections

L'usage des corpus dans l'enseignement des langues a suivi son apparition en linguistique. Nous abordons donc dans un premier temps la notion dans ce contexte, en commençant par une définition avant de constater les fonctionnalités proposées par l'informatique.

a) Définition

Définition 2.3 (Corpus) « *A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.*⁸⁹ » [Sinclair, 1996, p. 4]

À l'aune de la définition de Sinclair, certaines grandes collections de documents, utilisées par la communauté du TAL pour mettre au point ses traitements, usurpent en quelque sorte l'appellation de « Corpus » [Habert *et al.*, 1997, p. 143]. Dans les premières communications sur le projet MIRTO, l'entité qui a donné naissance au présent travail s'appelait corpus (Antoniadis & Ponton [2002, 2004]). Cette appellation n'est plus justifiée d'après la définition de Sinclair. En effet, cette dernière porte avant tout sur la constitution du corpus, par rapport à son usage. Bien évidemment, elle n'ignore pas que la constitution se fait avec un but précis, mais elle se focalise sur le processus. Et la constitution du fond documentaire que nous allons organiser et interfacer ne suit pas ces directives. Le choix que nous avons fait de nous décharger de la composition de la collection de textes et de nous concentrer sur l'indexation des ressources nous extrait à lui seul de la problématique du corpus telle que définie par Sinclair. En effet, non seulement nous n'avons pas défini de critère linguistique pour la sélection des textes (le seul critère pourrait être « le texte est, de l'avis d'au moins un enseignant, utilisable dans le contexte de l'enseignement des langues »), mais en plus l'objectif de la constitution de la base n'est pas fondamentalement de servir d'échantillon de la langue. Il s'agit plutôt d'un fond documentaire, dont l'organisation doit permettre à l'utilisateur de retrouver parmi de nombreux textes les plus adaptés à ses besoins. Cette problématique est centrée sur l'organisation des textes, mais, encore une fois, nous nous écartons de la définition de Sinclair. S'il s'agit bien de définir des critères d'organisation explicites, ils sont avant tout pédagogiques et non linguistiques. Bien sûr, il y a de fortes présomptions selon lesquelles les critères pédagogiques s'appuieront d'une manière ou d'une autre sur des critères linguistiques, cependant notre système n'est pas destiné à des linguistes, mais à des enseignants de langue. Enfin, l'unité centrale de notre système est le texte. Il se démarque en cela des collections utilisées en TAL pour mettre au point des traitements : « on serait tenté de voir là “du texte” » [Habert *et al.*, 1998, p. 35]. L'objectif n'est pas de fournir à l'utilisateur uniquement des « données langagières ». Ces dernières peuvent être utilisées pour faire la sélection la plus appropriée dans la collection globale, mais l'objectif final est de proposer un texte à l'utilisateur.

89. « Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage. » (traduit par Habert *et al.* [1997], p. 11).

b) Collection de textes

Dans la suite de ce manuscrit, pour désigner l'ensemble des textes indexés par notre base nous ne parlerons donc pas de corpus, mais de collection (de textes) :

« *Words such as collection and archive refer to sets of texts that do not need to be selected or do not need to be ordered or the selection and/or ordering do not need to be on linguistic criteria. They are therefore quite unlike corpora.*⁹⁰ »

[Sinclair, 1996, p. 5]

En nous éloignant de la problématique des corpus, les contraintes sur la taille de la collection s'en trouvent diminuées :

« Il est [...] impossible de définir ce qu'est un "bon" corpus en faisant abstraction de l'objectif visé. Néanmoins, on s'attend à ce que certaines exigences soient respectées dans tous les cas : la taille des données doit être suffisante (par souci de représentativité), elles doivent être diversifiées, et leur origine clairement mémorisée (les coordonnées des documents primaires sont conservées). »

[Habert *et al.*, 1998, p. 35]

Bien sûr, plus il y aura de textes différents (taille suffisante, données diversifiées) et plus il y aura de chances qu'il existe, dans la collection, un texte satisfaisant pleinement les besoins de l'enseignant et un nombre trop faible de textes réduira à néant l'intérêt du système du point de vue de l'utilisateur. Mais strictement, il n'y a pas de minimum au nombre de textes que le système doit référencer : indiquer que le système ne référence aucun texte satisfaisant une demande donnée de l'utilisateur est une réponse adéquate à une question (si c'est effectivement le cas). Notre problématique est celle de l'indexation pédagogique de textes pour l'enseignement des langues et non celle du peuplement d'une collection de textes. En revanche, nous devons, nous aussi, garder « clairement mémorisées » les origines des documents, car ce sont les textes qui doivent être fournis à l'utilisateur.

2.3.3 Outils

Qu'il s'agisse de corpus ou de collection, c'est le pendant électronique de ces ensembles de textes qui nous intéresse. Tout d'abord, il peut s'agir de corpus annotés. Ces annotations jouent le rôle que nous avons désigné pour le TAL dans l'ALAO (§ *Le TAL pour l'ALAO* p. 55) : à savoir fournir à l'ordinateur, et par son intermédiaire à l'utilisateur, des informations supplémentaires sur le contenu des textes. Ces annotations peuvent être ajoutées automatiquement (ce qui ne signifie pas qu'il n'y ait pas de vérification humaine ensuite) ou manuellement (ce qui ne signifie pas qu'elles ne soient pas aidées d'une manière ou d'une autre par des outils informatiques) [Habert *et al.*, 1997, p. 11]. Parmi ces outils informatiques d'annotation de corpus figurent d'ailleurs des outils TAL.

De la même manière que le choix des textes dépend des objectifs poursuivis, l'annotation doit être faite dans un but précis. Par exemple, l'étude menée par Biber sur les genres textuels avait pour objectif d'utiliser des outils de statistique multidimensionnelle, afin d'essayer de caractériser des textes en fonction de 41 traits lexicaux et syntaxiques [Biber, 1986, p. 387]. D'après Habert, « l'étiquetage mis en oeuvre par Biber s'éloigne de l'étiquetage morpho-syntaxique pratiqué en

90. Des mots comme « collection » ou « archive » se réfèrent à des ensembles de textes qui n'ont pas besoin d'être sélectionnés ou qui n'ont pas besoin d'être ordonnés ou pour lesquels il n'est pas nécessaire que la sélection et/ou l'ordonnement dépende de critères linguistiques. Ils sont donc bien différents des corpus.

général. Il est partiel et partial. Il est “inéquitable” : il s’intéresse à des fonctionnements linguistiques très spécifiques qu’il analyse en détail tandis qu’il en laisse d’autres dans l’ombre » [Habert *et al.*, 1997, p. 29]. Le choix des traits analysés avait été effectué en fonction de travaux antérieurs [Biber, 1986, pp. 387–388], seuls les traits pertinents pour l’analyse ont été annotés. Dans cet exemple, le corpus est annoté pour une analyse uniquement quantitative, mais l’annotation peut aussi permettre une analyse qualitative. Par exemple, Koteyko suggère que dans le domaine de l’analyse du discours, les paraphrases, qu’elle estime facilement repérables avec les outils de recherche disponibles avec la plupart des logiciels d’analyse de corpus, permettent, avec un corpus constitué dans cet objectif, une étude diachronique des liens intertextuels [Koteyko, 2006, p. 151]. Schmied donne un autre exemple d’analyse qualitative, dans le cadre d’un travail sur les constructions relatives en anglais. La première partie du travail consiste en le repérage d’occurrences des catégories utilisées dans son analyse. Elles sont sélectionnées dans le corpus, pour être discutées comme variables, selon les valeurs qu’elles peuvent prendre [Schmied, 1993, p. 86]. Dans la suite de Schmied, McEnery et Wilson remarquent que l’analyse qualitative est souvent une prémisses d’une analyse quantitative : « avant que les phénomènes linguistiques soient classifiés et comptés, les catégories de la classification doivent être identifiées⁹¹ » [McEnery & Wilson, 2001, p. 76].

Les travaux de Biber que nous avons cités montrent un usage de l’ordinateur spécialisé dans le repérage et le dénombrement, alors que ceux de Koteyko se concentrent sur le repérage. Dans un cas comme dans l’autre, l’ordinateur s’appuie sur les informations dont il dispose : les annotations. Ces dernières sont elles aussi soumises aux limites de l’informatique et du TAL, en ce sens qu’elles dépendent de la fiabilité des outils utilisés. Plus l’objet de l’analyse s’éloigne de la surface et plus les traitements manquent de fiabilité, d’où l’utilisation d’outils comme aide à l’annotation supervisée par un opérateur humain. Il peut s’agir d’étiquettes morpho-syntaxiques associées aux mots ou encore des représentations syntaxiques décorant les phrases. D’autres outils permettent d’aligner des corpus multilingues. Les corpus peuvent aussi être utilisés pour des études diachroniques ou pour la constitution de ressources lexicales. Mais nous n’allons pas passer en revue tous les types d’outils disponibles pour la linguistique de corpus, notre objectif étant de montrer que, comme pour notre travail, le choix des outils et des méthodes est intimement lié aux objectifs poursuivis.

Ayant présenté quelques éléments basiques de la problématique de la constitution des corpus, nous allons maintenant nous intéresser à leur utilisation dans l’enseignement des langues. En effet, comme l’indique Teubert, la didactique des langues s’intéresse depuis longtemps aux corpus [Teubert, 2005, p. 2]. D’après Granger *et al.* depuis l’article de Francis [1967], qui analysait les conséquences éventuelles du corpus Brown pour l’enseignement de l’anglais comme L2, le lien entre la linguistique de corpus et l’enseignement s’est renforcé graduellement, les corpus devenant un composant majeur de la scène pédagogique [Granger *et al.*, 2007, p. 253].

2.3.4 Instrumentation du corpus et des outils d’interrogation

Le fait que notre travail concerne une collection de textes et non un corpus ne nous dispense pas d’examiner les utilisations de ces derniers dans l’enseignement des langues, et ce pour deux raisons :

- certains aspects de ces usages, comme l’annotation des textes, peuvent être pertinents dans

91. « before linguistic phenomena are classified and counted, the categories for classification must first be identified »

notre cas ;

- les « corpus » en question ne sont pas forcément des corpus au sens de Sinclair du terme.

Dans une perspective instrumentale, nous allons tout d'abord constater des usages des corpus, qui relèvent principalement de l'instrumentation, avant de nous pencher sur des exemples d'instrumentalisation. Ces deux composantes du processus de genèse instrumentale ne sont pas mutuellement exclusives. Nous allons, dans un premier temps, examiner certaines utilisations de corpus et d'outils d'interrogation comme matériel pédagogique. Ils n'ont pas nécessairement été conçus pour l'enseignement des langues, c'est dans l'usage qu'ils sont adaptés à l'enseignement.

Revenons, avec plus de rigueur (cf. § c) p. 61), sur l'usage des concordanciers en classe de langue, qui nous a servi d'exemple pour le processus de genèse instrumentale. Cet usage, bien qu'attribué à Peter Roe en 1969 [McEnery & Wilson, 1997, p. 12] (cité par [Chambers, 2007a, p. 249]), ne s'est développé que bien plus tard, en grande partie sous l'impulsion de Johns qui lui donne le nom de *Data Driven Learning* (DDL)⁹²[Johns, 2002, p. 107]. Dans le DDL, l'objectif est « d'éliminer l'intermédiaire [c'est-à-dire l'enseignant] dans la mesure du possible pour donner à l'apprenant un accès direct aux données » [Boulton, 2007, p. 38] (citant [Johns, 1991, p. 30]). Les activités de DDL sont principalement tournées vers la « frontière collocationnelle entre syntaxe et lexique⁹³ » [Johns, 2002, p. 109]. Les principales activités du DDL sont des instrumentations d'artefacts regroupant aussi bien les données (corpus) que les outils d'exploitation de ces dernières (concordanciers) ; cela ne signifie pas qu'il n'y ait pas de choix des outils. Flowerdew exprime certains aspects de cette problématique : choix du concordancier en fonction de certains critères comme la vitesse de recherche et le choix des données sur lesquelles faire marcher le concordancier selon qu'il s'agit de travail grammatical ou lexical pour une langue de spécialité ou non [Flowerdew, 1993, pp. 88–90].

Aston recense cinq principaux types d'activités d'enseignement des langues, ayant recours aux corpus [Aston, 2002, p. 10]. Ils recourent les propositions faites par Johns :

TYPE D'ACTIVITÉ [Aston, 2002, p. 10]	DESCRIPTION
Activités centrées sur la forme ⁹⁴	« Visant à établir et utiliser les usages de certains traits linguistiques ⁹⁵ » [Aston, 2002, p. 10]. Par exemple le travail à cheval sur la syntaxe et le lexique décrit par Johns : classification par les apprenants des collocations de prépositions à partir d'exemples issus du concordancier [Johns, 1986, possibilité 1 p. 160]. Flowerdew parle des activités possibles en termes d'approches inductives et déductives. Les sorties d'un concordancier pouvant être utilisées comme exemplier d'une règle énoncée (approche déductive), mais aussi pour faire élaborer à l'apprenant des règles à partir d'exemples (approche inductive)[Flowerdew, 1993, p. 95].

92. « apprentissage à partir des données » (traduit par [Boulton, 2007], p. 37).

93. « *It is on the ["collocational border" between syntax and lexis] that DDL methods seem to be most effective.* »

94. Dans l'ordre du tableau « *form-focused activity, meaning-focused activity, skill-focused activity, reference activity, browsing activity* »

95. « *aiming to establish and practice the use of particular linguistic features* »

TYPE D'ACTIVITÉ [Aston, 2002, p. 10]	DESCRIPTION
Activités centrées sur le sens	Pour établir du sens dans certains contextes. Comme par exemple les activités sur le vocabulaire décrites par Johns : utiliser les contextes multiples offerts par les concordancier pour émettre et tester des hypothèses de sens dépendant des contextes [Johns, 1986, possibilité 1 p. 160].
Activités centrées sur les savoir-faire	« Visant à développer des compétences et des stratégies de lecture, en particulier du type “ <i>bottom-up</i> ” [Brodine, 2001] ⁹⁶ » [Aston, 2002, p. 10]. En effet, Brodine oppose les stratégies ‘ <i>top-down</i> ’, faisant intervenir les connaissances déjà établies du lecteur pour comprendre les énoncés produits par l’auteur du texte, aux stratégies ‘ <i>bottom-up</i> ’ qui partent des unités linguistiques et construisent le sens à partir de là [Brodine, 2001, p. 139–140]. Les activités que nous avons citées comme centrées sur le sens peuvent aussi satisfaire cette description dans la mesure où Johns met en exergue l’importance du développement de stratégies pour apprendre à deviner le sens des mots à partir d’indices contextuels [Johns, 1986, possibilité 1 p. 160].
Activités où les corpus servent de référence	Le corpus est utilisé comme support dans des tâches impliquant d’autres textes (aide à l’écriture, la lecture ou la traduction). Par exemple, Johns suggère l’usage du corpus comme aide à la correction pour des devoirs d’apprenant, où l’enseignant se contente de suggérer à l’apprenant que des recherches avec un concordancier suffiront à lui expliquer son erreur. L’apprenant utilise ensuite le concordancier en autonomie [Johns, 1986, possibilité 1 p. 160].
Activités d’exploration	« Les apprenants alternent entre les différents types d’activités dans des explorations spontanées des corpus ⁹⁷ » [Aston, 2002, p. 10]. Ce type d’activité suppose que l’apprenant est autonome, répond à ses propres questions et explore la langue « au petit bonheur la chance » comme on lit un dictionnaire [Flowerdew, 1993, p. 95].

TABLE 2.4: Types d’activités utilisant de corpus dans l’enseignement des langues

Selon le corpus utilisé, des informations supplémentaires peuvent être disponibles. Ainsi un concordancier peut interroger des corpus alignés [Lamy & Klarskov Mortensen, 2007, section 1.3]. C’est-à-dire des corpus dont chaque texte est disponible en plusieurs langues (une ou plusieurs traductions de chaque texte) et pour lesquels « il existe un système de mise en relation entre segments du texte de “grain équivalent” : sections, paragraphes, phrases » [Habert *et al.*, 1997, p. 135]. Des activités pouvant s’inscrire dans ce que Johns appelle le DDL peuvent être proposées : analyse de concordances [St. John, 2001] [Johns, 2002, p. 115] et exercices lacunaires [Johns, 2002, p. 116], montrant l’influence du contexte sur la traduction de tel ou tel phénomène

96. « *aiming to develop particular reading skills and strategies, particularly of a “bottom-up” variety* [Brodine, 2001] »

97. « *learners alternate between the previous types of activity in serendipitous explorations of the corpus* »

langagier.

Johns propose également un type d'activité qui constitue en quelque sorte une instrumentalisation des données mais pas nécessairement des outils d'exploitation. Il s'agit de constituer un corpus de travaux d'apprenants et de comparer avec eux les tournures utilisées dans leurs corpus avec celles d'un corpus de spécialité correspondant au thème de l'activité (brochure commerciale destinée à la vente de produits dans le cas de l'expérience relatée)[Johns, 1986, pp. 160–161]. Encore une fois, il s'agit de confronter les conceptions de la langue des apprenants à des usages attestés.

a) Un texte ou du texte

La première constatation que nous pouvons faire quant aux usages des corpus relevant globalement du DDL présentés ci-dessus, est que l'utilisateur final reste toujours l'apprenant. Qu'il travaille en autonomie (cf. « activités d'explorations » dans le tableau 2.4), que le travail soit guidé par l'enseignant ou que l'enseignant fournisse les concordances sous forme d'imprimés, c'est toujours l'apprenant qui les manipule. Dans notre cas, les fonctions constituantes stipulent que les utilisateurs finaux sont des enseignants qui recherchent des textes à utiliser dans le cadre de leurs cours (constitution de matériel pédagogique). Flowerdew propose certains usages qui sortent de la liste définie par Aston et du cadre strict du DDL : il s'agit d'utilisations par les enseignants dans leur tâche de préparation⁹⁸. Tout d'abord comme source d'informations pour l'appréciation de l'usage, notamment pour les enseignants qui n'enseignent pas leur L1. Il prône en cela l'approche défendue par Tsui [2004], dont nous avons parlé dans le chapitre introductif (cf. § 1.3.3 p. 16). Il parle des concordanciers comme source d'exemples attestés pour la présentation de phénomènes langagiers, notamment contrastifs. Et enfin, il suggère une utilisation comme *input* pour le développement de matériel pédagogique [Flowerdew, 1993, pp. 91–93], par exemple pour des exercices lacunaires (voir également [Johns, 2002, pp. 108, 116]).

Que ce soit dans ce type de situation, ou dans celle de l'usage des corpus dans la classe, une différence majeure persiste avec notre problématique. L'usage qui est fait des corpus ne considère plus le texte comme unité : le travail s'articulant avant tout autour de concordances (cf. figure 2.3 et figure 2.4), il perd quelque peu de vue une forme d'authenticité de la langue (cf. types d'authenticité issus de Taylor [1994] dans § *Définition* p. 66). Les données langagières proposées aux apprenants sont effectivement des usages attestés, mais l'authenticité de la dimension interprétative de ces données langagières par les apprenants (cf. note⁸⁵ p. 66) n'est plus représentée :

*« [Corpus] language may be considered as “genuine”, in the sense of attested products of use (cf. [Widdowson, 2003, p. 93]), but not “authentic” for the purpose of language teaching/learning. For the purpose of teaching, language samples provided by corpora still need to be “authenticised” by the learners. In other words, they must be re-contextualised in a pedagogical setting to make them relevant for specific classroom purposes and thus make them real for the learners. »*⁹⁹

[Kaltenböck & Mehlmauer-Larcher, 2005, p. 70]

98. Notons ici, que quand la préparation concerne la constitution de matériel pédagogique, ce dernier peut être destiné à des activités de DDL.

99. La langue dans les corpus peut être considérée comme « véritable », dans le sens de produits attestés de l'usage, mais pas comme « authentique » dans un but d'enseignement/apprentissage des langues. Pour l'enseignement, les échantillons de langue fournis par les corpus doivent être « authentifiés » par les apprenants. En d'autres termes, ils doivent être re-contextualisés dans un environnement pédagogique pour être pertinents dans des objectifs spécifiques à la classe et ainsi les rendre réels pour les apprenants.

En effet, pour les auteurs, les corpus fournissent du co-texte et non du contexte [Kaltenböck & Mehlmauer-Larcher, 2005, p. 69]. La contextualisation est à la charge des apprenants et peut être plus ou moins ardue en fonction des informations disponibles concernant chaque texte, comme le signale Partington :

« *we tend to know nothing about the author of the message of a concordance line and their illocutionary intentions, maybe very little about the intended audience and the circumstances in which the message was produced. Concordance data is as decontextualised as any linguistic information could [possibly] be[...].*¹⁰⁰ »

[Partington, 1998, p. 145]

Dans l'utilisation des corpus prônée dans les exemples donnés ci-dessus, c'est le corpus qui fait office de matériel pédagogique et non les textes qu'il contient. Ces derniers perdent finalement de leur autonomie d'entités indépendantes (*des* textes) et se trouvent regroupés en un matériau : *du* texte. Dans ces activités, ce qui est perdu en authenticité de la langue est gagné en authenticité de la tâche et éventuellement de la situation (cf. types d'authenticité de Taylor [1994] 66). Comme le signalent les mêmes Kaltenböck et Mehlmauer-Larcher, les apprenants effectuent des tâches de chercheur [Kaltenböck & Mehlmauer-Larcher, 2005, p. 78] (authenticité de la tâche) : le travail de recontextualisation des énoncés est le même que celui que devrait faire un linguiste. En outre, les discussions méta-linguistiques peuvent être menées en L2 (authenticité de la situation de communication).

Si, à première vue, notre problématique peut paraître proche des utilisations de type DDL, il en ressort que par la manière même de considérer le texte, elle s'en démarque tout à fait. Dans les exemples que nous avons cités, le contenu des corpus est avant tout *du* texte, un ensemble de données langagières dont l'articulation en textes n'est que peu utilisée, du fait des objectifs poursuivis et des propriétés des outils :

- « *It is on the ["collocational border" between syntax and lexis] that DDL methods seem to be most effective*¹⁰¹ » [Johns, 2002, p. 109] ;
- « *The focus [of the activities] was on grammar/usage as this tends to lend itself to a corpus approach*¹⁰² » [Boulton, 2008, p. 39].

Malgré ces incompatibilités de point de vue, cette problématique de la contextualisation des données issues d'un corpus sera à terme à prendre en considération : notre problème premier est de permettre à l'utilisateur (enseignant) de pouvoir trouver des textes à utiliser dans le cadre de la classe. Cependant, pour être utilisables dans une perspective communicationnelle, les textes devront pouvoir être recontextualisés par les apprenants. Ne s'agissant pas de concordances (matériau textuel), la tâche sera plus aisée. À terme, il ne suffira plus uniquement de proposer des traits permettant de retrouver les textes, il faut aussi proposer des descripteurs qui facilitent leur recontextualisation.

100. On a tendance à ne rien savoir de l'auteur du message d'une ligne de concordance et de son intention illocutoire, à ne savoir que très peu de chose du public auquel l'énoncé était destiné et des circonstances dans lesquelles le message était produit. Les concordances sont aussi décontextualisées qu'une information linguistique peut l'être.

101. c'est à la frontière collocationnelle entre syntaxe et lexique que les méthodes de DDL semblent les plus efficaces

102. [Dans les activités], l'accent était mis sur la grammaire / l'usage, car il semble que cela se prête bien à une approche avec corpus

2.3.5 Instrumentalisation du corpus et des outils d'interrogation

Les usages des corpus tels que nous les avons vus ne couvrent pas l'intégralité de leur champ instrumental dans le cadre de la didactique des langues. Tous ces usages constituent en premier lieu des instrumentations, qui réutilisent des données et des outils créés pour la linguistique de corpus. Il existe cependant d'autres pistes d'utilisation de corpus dans l'enseignement qui passent par une instrumentalisation, à l'instar de l'activité relatée dans Johns [1986], que nous citons page 73. Elle constitue en quelque sorte une expérience de l'usage d'un « corpus » d'apprenants dans l'enseignement des langues, alors que le champ disciplinaire n'en est qu'à ses balbutiements (cf. NB 2.3 p. 75).

Nota Bene 2.3 Corpus d'apprenants

Les corpus d'apprenants contiennent, comme leur nom l'indique, des textes écrits par des locuteurs d'une langue qui n'est pas leur L1. La L2 peut s'avérer être une langue seconde, étrangère ou officielle (comme l'anglais pour l'Inde) selon la terminologie employée par Granger [Granger, 2002, pp. 8–9]. Ce domaine de recherche à part entière naît à la fin des années 80, mais surtout au début des années 90^a et veut analyser des productions d'apprenants avec les outils de la linguistique de corpus offrant ainsi aux chercheurs en acquisition d'une langue seconde et en didactique des langues une nouvelle perspective d'étude de données jusque là peu analysées [Granger, 2002, p. 6]. Ils créent alors leur propres corpus, ce qui soulève de nouvelles questions (problème de l'authenticité dans le cadre de textes d'apprenants, condition de recueil des données) [Granger, 2002, p. 8–9] et demandent de nouveaux types d'annotation : annotation des erreurs, informations sur le locuteur (L1, niveau, genre) [Pravec, 2002, p. 94] (instrumentalisation).

Même si des corpus d'apprenants peuvent être utilisés dans le cadre de la classe, comme le montrent des expériences comme celles de Johns [Johns, 1986, pp. 161–162] et Seidlhofer [2002], l'objectif reste en premier lieu la description de la langue de l'apprenant avec un point de vue d'acquisition d'une langue seconde et l'amélioration de l'enseignement des langues étrangères [Granger, 2002, p. 4]. Un corpus d'apprenants permet d'aller au-delà des analyses possibles en étudiant du texte produit par des locuteurs natifs : non pas se contenter des informations, certes précieuses de fréquence d'usage des mots, phrases et structures, mais tenter d'en savoir plus sur leur difficulté pour tel ou tel profil d'apprenants [Granger, 2002, p. 21]. En outre, les corpus d'apprenants (annotés) peuvent être utilisés par les apprenants pour susciter des conversations concernant l'édition de textes [Granger *et al.*, 2007, p. 259] ou pour élaborer et tester les systèmes de diagnostic d'erreurs et de feed-back [Granger *et al.*, 2007, p. 266].

^a cf. bibliographie du domaine de Sylviane Granger : <http://tr.im/bibGranger>

Les instrumentalisations dont nous allons parler sont des développements très récents qui s'appuient sur des travaux publiés à la toute fin des années 90 et qui ne prennent corps qu'au milieu des années 2000. Elles proviennent du constat que :

« [A] careful look around the many different places where languages are learnt and taught makes clear that corpora, while being the “buzzword” in language research departments, are still far from being part of mainstream teaching practice, if not *terra incognita altogether*.¹⁰³ »

[Braun, 2005, p. 48]

103. Un tour d'horizon rigoureux des nombreux endroits où les langues sont apprises et enseignées montre avec clarté que les corpus, bien qu'étant le mot clé à la mode dans le monde de la recherche, sont loin de faire partie des pratiques courantes d'enseignement, pour ne pas dire une *terra incognita*.

Pour Braun, la sous-utilisation des corpus en classe de langue, en dehors des centres étroitement liés au monde de la recherche, provient d'un déficit de formation, qui commencerait à peine à être comblé [Braun, 2007, p. 308]. Au delà de problèmes de méthodologie, comme la difficulté de l'interprétation de concordances, Braun suggère que c'est aussi par leurs contenus que les corpus doivent s'adapter aux contraintes du programme. Elle rejoint en cela l'analyse d'Aston pour qui des corpus constitués par les enseignants et/ou les apprenants (corpus « faits-maison ») pourraient s'avérer « plus adéquats dans une perspective d'apprentissage que des corpus pré-compilés, en ce sens qu'ils peuvent être conçus spécialement pour s'adapter aux connaissances et centres d'intérêts des apprenants¹⁰⁴ » [Aston, 2002, p. 10]. Il sépare ces corpus à construire en deux types : corpus construits à partir de textes avec lesquels les apprenants sont familiers (textes étudiés ou écrits par les apprenants) et des corpus de textes d'un type donné ou illustrant un domaine d'usage, certains traits linguistiques. Parmi différentes stratégies de constitutions de corpus adaptés à l'enseignement, Aston propose des exemples de création de sous-corpus du *British National Corpus* (BNC). Il conclut que les sous-corpus peuvent satisfaire des propriétés parmi celles qui l'ont motivé à proposer l'usage de corpus « faits maison ». En effet, les sous-corpus peuvent fournir des quantités de données plus petites et donc plus gérables mais aussi plus homogènes que de grands corpus. Le fait de diminuer la représentativité des concordances proposées à partir d'un sous-corpus permet en outre pour lui d'en faciliter l'interprétation [Aston, 2002, p. 22]. Il pense que le travail sur sous-corpus facilite, notamment via sa plus grande homogénéité, la formulation d'hypothèses que les apprenants peuvent ensuite tester sur un corpus complet [Aston, 2002, p. 23], un point de vue explicité dans [Aston, 1997, § *From small to large*]. Enfin, il suggère que la création et la sélection par les apprenants de leurs propres sous-corpus pour certaines tâches particulières, leur permet de mieux appréhender les corpus qu'ils ne connaissent pas et même pour ceux qui ne se destinent pas à utiliser les corpus dans leur vie professionnelle de développer une certaine « sensibilité aux variations de genre et de registres » [Aston, 2002, p. 23].

La nécessité de contraindre la quantité¹⁰⁵ et le type de documents dans le corpus se retrouve chez Braun, qui pousse plus loin la logique que Aston en prônant la constitution de corpus « pédagogiques ». Le type de démarche proposé par Aston pouvait relever de ce que nous avons appelé une instrumentalisation des données pour un artefact comprenant le corpus et son système d'interrogation (quand il ne s'agit pas d'un sous-corpus qui, dans le cadre des exemples donnés, ne représente qu'un usage détourné de l'interface d'interrogation du BNC pour les besoins de l'enseignement des langues). Braun va plus loin dans l'instrumentalisation puisqu'elle ne suggère pas uniquement d'utiliser des données adaptées aux étudiants avec des outils de linguistique de corpus, mais en plus de créer ses propres outils d'exploitation dans le cadre de l'enseignement des langues. Les adaptations qu'elle suggère et les questions qu'elle pose touchent tous les aspects de la création d'un corpus et de son système d'interrogation :

- choix des textes à ajouter : taille et contenu (cf. ci-dessus) ;
- annotations linguistiques des textes trop complexes pour un contexte d'enseignement des langues [Braun, 2005, p. 50] ;
- nécessité de considérer du discours (cf. § *Un texte ou du texte* p. 73) et non simplement du texte [Braun, 2005, p. 52] (un exemple de fonctionnalité allant dans ce sens serait la possibilité d'accéder au contenu du corpus par une lecture « horizontale », qu'elle oppose à la lecture « verticale » proposée par la notation KWIC ou l'analyse de fréquences [Braun, 2005, p. 54]) ;

104. « “[H]ome made” corpora may be more appropriate for learning purposes than pre-compiled ones, insofar as they can be specifically targeted to the learner’s knowledge and concerns; »

105. cf. Braun citant [Aston, 1997, p. 54] [Braun, 2005, p. 49]

- enrichissement du corpus avec des activités ou des tâches toutes prêtes [Braun, 2005, pp. 53 & 55].

a) Corpus pédagogique

Dans le cadre de l'adaptation des corpus à la problématique de l'enseignement des langues, Braun se rapproche de la problématique de l'indexation pédagogique quand elle discute les conséquences d'une approche orientée vers le discours sur l'annotation des corpus :

« *In the creation of a pedagogically relevant corpus, a discourse-based approach will support the corpus developers in identifying not only relevant formal units such as individual means of expression, grammatical structures, larger text passages, but also the discourse functions they fulfil, the context-specific meanings they carry and the topics they relate to. [This] type of corpus analysis will be the main source for pedagogic annotation, which will, to a large extent, have to be manual and will often focus on units beyond concordance lines or sentences. The major task of the annotation is to support pedagogically motivated corpus queries.*¹⁰⁶ »

[Braun, 2005, p. 54]

Braun emploie le terme d'« annotation pédagogique », ce qui consiste en l'analyse des textes et en l'ajout d'informations, relevant de la problématique de la didactique des langues, qui soient interprétables par la machine. En nous plaçant du point de vue du processus d'indexation tel que nous l'avons expliqué dans le chapitre 1, ces informations seraient des exemples de « concepts contenus dans les documents » (cf. § 1.3.1 p. 10). De même, les requêtes de corpus « motivées par des problèmes pédagogiques » ne sont pas sans rappeler les questions des usagers dans la définition du langage documentaire de Lefèvre (définition 1.3 p. 12). En expliquant l'opération en ces termes, la proximité avec l'indexation pédagogique est indéniable. Il faut cependant garder à l'esprit l'objectif de cette annotation, car si les utilisateurs d'un « corpus pédagogique » pourraient dans certains cas être les mêmes personnes que ceux de notre système, les rôles qu'ils jouent et leurs objectifs sont différents.

Comme en témoignent son système *English Language Interview Corpus as a Second-Language Application* (ELISA) [Braun *et al.*, 2004] et le projet qui l'a suivi, *System Aided Compilation and Open Distribution of European Youth Language* (SACODEYL)[Widmann *et al.*, 2008, p. 326], la citation ci-dessus est à replacer dans le contexte de la création d'un corpus en tant que matériel pédagogique et non comme un outil destiné aux enseignants pour la recherche de textes comme dans notre cas. Dans le cadre d'ELISA, une grosse partie du travail était dédiée à la « médiation pédagogique du corpus¹⁰⁷ », qui par là-même s'inscrit dans une perspective intégrative de l'ALAO telle que nous l'avons décrite précédemment (cf. § 2.1.5 p. 43 et § 2.1.6 p. 44). En effet, la démarche vise à inclure diverses compétences dans une même tâche (compréhension orale, réflexion métalinguistique) tout en intégrant la technologie au processus d'apprentissage (utili-

106. Dans l'optique de la création d'un corpus pertinent d'un point de vue pédagogique, une approche fondée sur le discours va aider les concepteurs de corpus à identifier non seulement des unités formelles cohérentes comme des moyens individuels d'expression, des structures grammaticales ou des parties de textes plus grandes, mais aussi les fonction discursives qu'ils remplissent, le sens qu'ils communiquent dans ce contexte précis et les thèmes auxquels ils sont liés. Ce type d'analyse de corpus sera la source principale de l'*annotation pédagogique* qui devra être en grande partie manuelle et se concentrera souvent sur des unités allant au delà des lignes de concordance ou de la phrase. La tâche principale de l'annotation est de permettre des requêtes de corpus *motivées par des problèmes pédagogiques*. (traduit par Lucy Garnier).

107. « *Another requirement concerns the pedagogical mediation of corpora.* »

sation d'outils comme les concordanciers) [Braun, 2007, p. 310–311]. Le corpus pédagogique, tel que défini par Braun, est en cela tout à fait représentatif de la phase que Warschauer désigne comme intégrative :

- il doit regrouper des textes relativement homogènes, choisis pour s'inscrire dans le programme des étudiants, annotés en conséquence ;
- il doit proposer des outils pour retrouver des sections de textes en fonction des annotations et les proposer sous des représentations adaptées (« horizontales » et « verticales ») ;
- il doit également être enrichi d'activités [Braun, 2007, p. 309].

Du fait de la quantité d'aspects traités, finalement peu d'informations sont disponibles sur le processus d'annotation et les raisons des choix des propriétés qui seront étiquetées dans les textes. Dans ces deux projets, les étiquettes associées à ces propriétés sont appelées catégories. Dans [Braun, 2005, p. 56], des pistes sont proposées :

- « catégories liées au contenu (thèmes, mots-clés)¹⁰⁸ » ;
- « catégories liées à la L2 (*i.e.* propriétés lexicales, grammaticales, pragmatiques et discursives)¹⁰⁹ » ;
- « catégories liées à l'apprenant (niveau, connaissances requises, compétences qui peuvent être travaillées, défis et difficultés)¹¹⁰ » ;

b) Annotation pédagogique

Dans le cadre des projets ELISA et SACODEYL, les « catégories » pressenties ne sont pas plus détaillées. Même dans des travaux ultérieurs s'appuyant très largement sur ceux-ci, le processus d'annotation et les raisons ayant présidé au choix des catégories sont relativement peu explicités¹¹¹, même si les catégories elles-mêmes gagnent en précision : 5 types de catégories (thèmes, caractéristiques grammaticales, caractéristiques lexicales, organisation textuelle, variation/style, niveau Cadre Européen Commun de Référence pour les Langues (CECR)), avec pour chacune d'entre elles des exemples [Widmann *et al.*, 2008, p. 324]. La stratégie quant à l'annotation pédagogique de SACODEYL¹¹² prend le parti de laisser les enseignants maîtres de l'annotation. Un étiquetage préalable est effectué pour les vidéos incluses dans le corpus, mais dans les articles que nous avons pu consulter, une place très faible est accordée à l'explication des catégories retenues et à la manière de les utiliser dans l'annotation. En revanche l'article de Widmann *et al.* est bien plus détaillé quant aux conséquences de cette dernière. Chaque catégorie présente dans le corpus permet d'accéder aux sections la contenant ou aux mots qu'elle décrit [Widmann *et al.*, 2008, pp. 323–326]. L'accent est mis sur le fait que chaque catégorie utilisée à l'annotation peut être réutilisée ultérieurement dans la recherche [Widmann *et al.*, 2008, pp. 323] et non sur le choix des catégories par défaut. Les catégories proposées dans l'« annotateur¹¹³ » sont très calquées sur la langue anglaise (cf. figure 2.5) sans être transposables au français ou à l'espagnol, pourtant présents dans le projet [Widmann *et al.*, 2008, pp. 321]. Les travaux d'Alcaraz et Pérez-Paredes éclairent le point de vue de SACODEYL : au lieu de réutiliser des outils et *a fortiori* des anno-

108. « *content-related categories (topics, keywords)* »

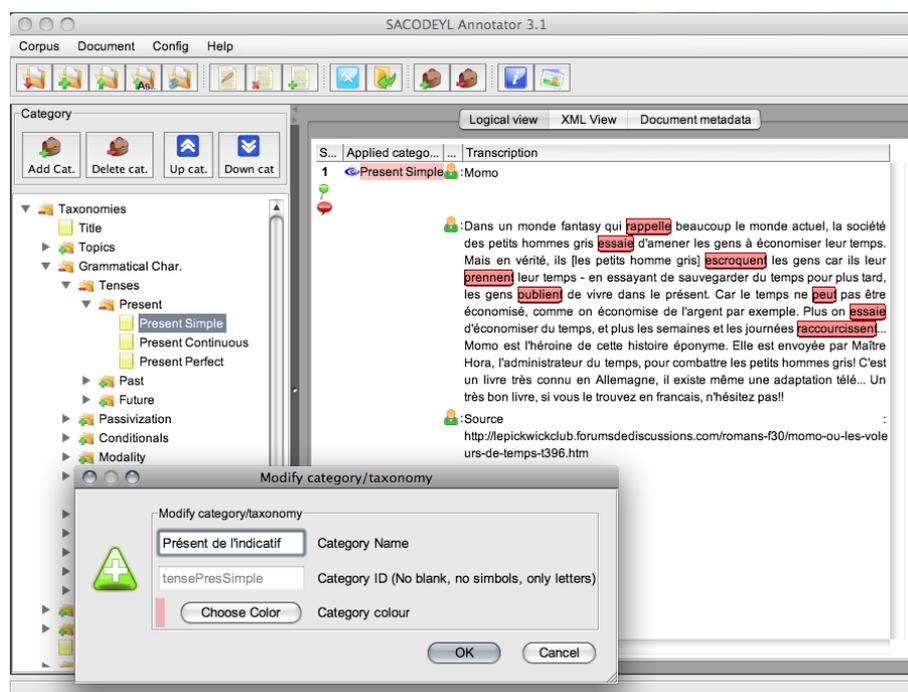
109. « *L2-related categories (ie lexical, grammatical, pragmatic and discourse properties)* »

110. « *learner-related categories (level of proficiency, relevant knowledge requirements, skills which can be practised, challenges and difficulties)* »

111. « *All corpora have been annotated with the needs of language learners in mind[.]* » [Widmann *et al.*, 2008, p. 321]

112. <http://www.um.es/sacodeyl/>

113. <http://www.um.es/sacodeyl/en/pages/software.htm#annotator> pour le téléchargement du système, disponible pour Linux, Windows et Mac .

FIGURE 2.5: SACODEYL Annotator¹¹³

tations destinées à des chercheurs, Alcaraz et Pérez-Paredes voient les enseignants comme « des guides et des éclaireurs », orientant les apprenants dans un corpus pédagogique. L'annotation de ce dernier, pilotée par les enseignants, est vue comme l'une des manifestations de leur rôle de guidage [Alcaraz & Pérez Paredes, 2008, p. 29]. Conscients de la variété des points de vue que des enseignants de différents horizons peuvent apporter [Alcaraz & Pérez Paredes, 2008, p. 29], l'annotateur SACODEYL est complètement générique de ce point de vue et peut être adapté à différents schémas d'annotation (taxonomies dans la terminologie du projet), d'où la possibilité :

- de renommer une catégorie sans modifier l'annotation¹¹⁴ ;
- supprimer, créer et ordonner des catégories (panneau de gauche de la figure 2.5).

Encore une fois, nos objectifs et contraintes nous différencient des travaux cités. Dans [Alcaraz & Pérez Paredes, 2008], les auteurs proposent une étude sur les annotations effectuées par deux enseignants. Si, comme eux, nous pensons que l'annotation et l'indexation pédagogiques ne peuvent s'obtenir qu'en travaillant avec des enseignants, dans notre cas nous ne pouvons demander aux enseignants d'annoter les textes. Tout d'abord, comme nous le faisons remarquer par la citation de Braun (p. 75), les corpus sont très peu utilisés dans le secondaire. Or, pour Chambers, la formation des enseignants a été identifiée, par Conrad et Sinclair notamment, comme un domaine clé pour l'introduction de nouveaux développements dans l'apprentissage des langues dans le primaire et le secondaire. Pour elle, il n'est pas complètement plausible que les organismes de formation des enseignants intègrent l'usage des corpus¹¹⁵ [Chambers, 2007b,

114. Dans la figure 2.5, nous renommeons la catégorie « *Present Simple* » en « *Présent de l'indicatif* » sans pour autant perdre les annotations effectuées).

115. « *Language teacher education has been identified as a key area for introducing new developments in language learning at primary and secondary level, and one could thus envisage a situation where corpus data would be increasingly used at these levels of education, introduced by teachers who had received the necessary training as part of the programme qualifying them as teachers. As such programmes typically already have a very full schedule, however, this scenario cannot be seen as entirely plausible.* »

p. 13]. Or l'un des parti-pris de notre travail est de nous appuyer sur une démarche empirique, sans privilégier une approche par rapport à une autre. Nous ne pouvons donc pas attendre des utilisateurs-enseignants qu'ils soient familiers avec la problématique de l'annotation de textes, que ce soit dans le contexte de la linguistique de corpus ou de l'indexation. En effet, dans le cadre du logiciel d'annotation du projet SACODEYL, le corpus est considéré comme du matériel pédagogique à la disposition des apprenants sous la supervision des enseignants. Ces derniers sont, de par leur implication dans une pédagogie ayant recours aux corpus, plus concernés par la problématique de l'enrichissement des textes à destination de systèmes informatiques. Si un bon usage d'une base de textes indexée pédagogiquement pour l'enseignement des langues passe par une certaine compréhension de l'outil, elle ne doit pas être un pré-requis à l'utilisation de celle-ci. Le fait que, dans SACODEYL, le corpus soit annoté par l'enseignant pour lui-même, dispense d'une cohérence inter-annotateurs. Au contraire, l'indexation de documents en est directement tributaire. Pour Luhn, la personne qui effectue la tâche d'indexation influence non seulement l'approche choisie, mais aussi l'interprétation du matériel à indexer.

« While there may be differences of opinion as to the effectiveness of this or that scheme, the important fact seems to be that any reasonable scheme of ordering, if understood, will save time in locating information. ¹¹⁶ »

[Luhn, 1960, p.289]

Ce point de vue ne suggère pas que le schéma d'indexation n'est pas important. Dans notre cas, l'approche choisie est une approche centrée sur l'enseignement des langues. Mais quel que soit le schéma, pour « faire gagner du temps dans la localisation de l'information », le système de description doit être cohérent. Quel que soit le domaine, des informations réunies selon un protocole de description trop peu formalisé diminuera sensiblement les conclusions qui pourront être tirées de ces informations, qu'il s'agisse de recherche de documents ou d'analyse d'archives :

« Les autorités de l'époque ont enregistré des noms de lieux à côté des patronymes d'un grand nombre de pirates capturés (trois cent quarante-six). Le problème est que nous ignorons la relation réelle qu'ils entretiennent avec l'endroit mentionné. Est-ce un lieu de naissance ? Un lieu d'où viennent des ancêtres ? Un lieu où il (ou elle) a grandi ? Est-ce plutôt le port d'attache, celui dont le navigateur part à chaque fois ? Une communauté comprenant une famille, des amis et des voisins ? Ou bien est-ce simplement le dernier port dont est parti le marin, une ville portuaire parmi une suite infinie d'autres villes portuaires où il a reçu sa paye et l'a dépensée comme "un seigneur pendant six semaines", avant de reprendre le large de nouveau ? Peut-être s'agit-il simplement d'un lieu décidé au dernier moment afin de tromper les autorités et ainsi protéger les êtres aimés de l'ignominie d'une exécution imminente. »

[Rediker, 2008, pp. 97-98]

Cet exemple peut sembler très éloigné de notre problématique, mais nous pouvons rapprocher l'annotateur du représentant de l'autorité du XVII^e siècle (sans s'attarder sur le fait que la comparaison n'est pas flatteuse) ; quant à la ressource à décrire elle serait le pirate appréhendé (ne nous attardons pas non plus sur la déshumanisation du pirate). Il y a dans ce processus de recueil de données deux problèmes. Quelles sont les directives de l'autorité visant à rendre ces informations cohérentes ? Les autorités veulent-elles savoir quelles sont les racines des pirates ou le dernier endroit d'où ils arrivent afin de retracer leurs mouvements ? Sans directives précises, les informations notées par le représentant de l'autorité sont inutilisables. Enfin ce représentant,

116. Bien qu'il puisse y avoir des différences d'opinion concernant l'efficacité de tel ou tel schéma, le fait important semble être que tout schéma d'indexation sensé, s'il est bien compris, fera gagner du temps dans la localisation de l'information.

confronté au pirate, devra prendre la décision de croire ce qu'il dit ou d'essayer de croiser les informations avec d'autres sources pour avoir une idée précise. Encore une fois, selon le représentant de l'autorité, l'annotateur dans notre cas, les informations recueillies peuvent être radicalement différentes. C'est pourquoi il est possible de demander aux utilisateurs de remplir la base de données des informations générales non ambiguës, mais une analyse de chaque texte devrait, pour être utilisable, s'effectuer selon un ensemble de consignes extrêmement précises. Une telle tâche serait beaucoup trop fastidieuse et dissuaderait de tout ajout de texte (cas d'utilisation « ajout d'un texte », § 1.2.1 p. 8). L'analyse du contenu des documents devra dans notre cas – contrairement aux approches choisies dans la problématique des corpus pédagogiques [Braun, 2005, p. 54][Alcaraz & Pérez Paredes, 2008] – s'appuyer sur des traitements automatiques (avec la prise en compte de leur limites, cf. § a) p. 59). Bien évidemment, ces différences ne nous dispensent pas de prendre en compte les conclusions de ces travaux pour la définition de critères pour l'indexation pédagogique. Si nous ne l'avons pas fait pendant la partie empirique de notre travail, c'est parce qu'ils n'avaient pas encore été écrits (notre recueil de données s'est terminé au printemps 2006, alors que la seconde réunion du projet SACODEYL n'avait pas encore eu lieu¹¹⁷). Quoi qu'il en soit, les données fournies dans [Alcaraz & Pérez Paredes, 2008] sont avant tout quantitatives. À cette phase de notre travail, nous aurions eu besoin de données qualitatives concernant les catégories. Ce type de données sera peut être disponible ultérieurement dans le cadre de ce projet, néanmoins, à la fin de l'année 2008, elles n'ont pas été publiées.

2.3.6 Une problématique différente, des outils communs

Alors que les premières expérimentations de l'usage de corpus et de concordanciers dans l'enseignement des langues semblent dater de la fin des années 60, elles ne se répandent (dans le monde de la recherche) qu'à la fin des années 80, notamment sous l'impulsion de Johns. Nous pouvons remarquer que cette diffusion coïncide avec l'émergence de la problématique de l'authenticité – qu'elle concerne les ressources, le discours, la tâche ou la situation – via l'avènement de l'approche communicative. Son intérêt s'accroît encore avec l'apparition de la perspective actionnelle. Elle pourrait aussi être associée aux avancées technologiques qui augmentent non seulement les capacités de calculs mais aussi de stockage des ordinateurs personnels, permettant à n'importe quelle machine personnelle de contenir un corpus. L'article de Johns sur micro-concord donne une idée des difficultés qui pouvaient dissuader un enseignant d'avoir recours à un corpus en 1986 : le système pouvait rechercher des mots dans des fichiers textes qui pouvaient faire jusqu'à 34 ko, mais qui, dans ce cas, étaient en général de moins de 20 ko¹¹⁸, à cause de limitations du logiciel de traitement de texte utilisé [Johns, 1986, p. 153] et permettait d'imprimer des concordances en colonnes de 74 ou 130 caractères sur une imprimante compatible Epson [Johns, 1986, p. 151]. Encore une fois, nous pouvons constater, même à un niveau très ponctuel, que l'utilisation d'une technologie dans l'enseignement dépend d'avancées technologiques et pédagogiques (cf. tableau 2.3).

Dans le cadre de l'usage des corpus dans l'enseignement des langues, la miniaturisation des composants et le gain de puissance des ordinateurs personnels permet, dans d'autres domaines (linguistique de corpus), la constitution de ressources textuelles de grande taille dans lesquelles des recherches de texte sont possibles. C'est sans faute d'accord que nous écrivons le mot « texte » : tels qu'ils sont utilisés dans l'enseignement des langues, les corpus donnent accès à *du* texte en

117. <http://www.um.es/sacodeyl/common/events.htm>

118. À titre de comparaison, les fichiers textes utilisés pour générer ce manuscrit totalisent 2,8 Mo, soit près de 150 fois plus.

tant que matériau et non à *des* textes en tant qu'objets. Ce sont les corpus, leurs contenus, leurs systèmes d'interrogation ou leurs sorties (sous forme de concordances, la plupart du temps) qui sont considérés comme du matériel pédagogique ; de l'*input*, indissociable du corpus lui-même. Les concordances prennent leur valeur par le fait qu'elles sont issues d'un corpus, représentatif de la langue. C'est aussi le cas dans les travaux concernant les corpus pédagogiques. En cela, ces usages diffèrent des objectifs de notre base de textes.

Notre travail a plus trait au processus de repérage de ressources par les enseignants qui, selon Álvarez, s'effectue selon le processus suivant [Álvarez, 2007, p. 22] :

1. « Búsqueda de recursos » ;
2. « Selección de la evaluación » ;
3. « Evaluación » ;
4. « Presentación » ¹¹⁹.

Le travail d'Álvarez est en général centré sur des ressources pédagogiques non brutes. Si, dans le détail de chaque phase, des différences apparaissent, la séquence dans sa globalité semble cohérente pour les ressources brutes. Dans le cadre d'une base de textes, une partie du choix de l'évaluation est décidée *a priori* pour être implantée dans le système. À chaque recherche de l'utilisateur, il effectue une partie de l'évaluation des ressources de la base avant de laisser le soin à ce dernier de terminer le processus en choisissant parmi les textes candidats celui ou ceux à utiliser dans le cadre du cours. Malheureusement, les travaux sur l'usage des corpus dans l'enseignement des langues sont très spécifiques à une démarche donnée, que nous ne voulons pas privilégier par rapport à d'autres démarches plus « traditionnelles » : le corpus (et les outils de gestion) constituent le matériel pédagogique et non une interface vers ce matériel. C'est le cas dans le DDL mais aussi dans le cadre des corpus pédagogiques, pour lesquels les aspects les plus aboutis concernent surtout le recueil des données et la présentation des résultats. Les travaux sur la description des ressources ne sont pas encore suffisamment avancés pour que nous puissions nous appuyer dessus. Dans l'approche de Braun, et plus généralement avec le projet SACODEYL, le fait que le contenu soit très contrôlé et augmenté de nombreuses fonctionnalités (alignement vidéo, activités et tâches proposées aux enseignants, interface adaptée à l'usage dans la classe) constitue en soi une valeur ajoutée pédagogique conséquente. Ils n'évident pas pour autant la dimension de l'annotation qui commence à être étudiée en détail. Notre approche est différente, de par les objectifs mêmes de notre projet : il s'agit pour nous de proposer un système de requête compatible avec la gestion d'une collection de textes conséquente.

Dans le cas général, la valeur des systèmes d'information est largement déterminée par le contenu et l'organisation de la base de données, plus que par la structure des interactions préconçues [Benjamin & Blunt, 1992, p. 12]. Selon ce point de vue, nous avons décidé de nous extraire de la problématique de la constitution des ressources pour fournir un cadre de travail permettant une organisation propice à la recherche de données pour l'enseignement des langues. Cette décision s'appuie notamment sur le point de vue de Roberts et Bosse-Andrieu : « *A corpus is only as good as the query system you have to consult it* ¹²⁰ » [Roberts & Bosse-Andrieu, 2006]. Cela ne signifie pas que la constitution du corpus n'a pas d'importance, les questions de cohérence et de représentativité des données par rapport à la langue ne sont pas ignorées. Cependant, ce que nous entendons par cette citation est que même si les ressources disponibles sont parfaitement adéquates, la qualité du corpus peut, malgré tout, pâtir d'un système de requête inadapté.

119. Recherche de ressources ; sélection du mode d'évaluation ; Évaluation ; Présentation

120. Un corpus est seulement aussi bon que le système de requête que vous avez pour le consulter.

Le système se place comme interface entre la quantité de ressources (collection de textes) et l'adéquation de quelques textes avec les besoins de l'enseignant. Le but est d'obtenir la qualité des supports à travers une indexation cohérente et adaptée aux questions des enseignants.

a) Des travaux proches, une approche différente

Nous pouvons illustrer la remarque de Benjamin et Blunt à travers les travaux que nous avons trouvés qui ressemblent le plus à notre problématique : ceux de Netter d'une part et ceux de Scharoff *et al.* d'autre part.

Netter est une enseignante de Français Langue Étrangère (FLE) qui a créé en 1995 le site ClicNet qui « édite et localise des ressources virtuelles en français pour les étudiants, les enseignants de français langue étrangère (FLE) ou langue seconde (FLS), et tous ceux qui s'intéressent aux cultures, aux arts et aux littératures francophones » [Netter, 2008]. Son approche est centrée sur le contenu dans la mesure où c'est un travail d'édition qui est fait, les recherches sont gérées par le système Atomz¹²¹ qui n'adapte pas ses critères à la problématique de l'enseignement des langues. Notre approche rejoint plus celle de Sharoff *et al.* : la recherche de l'« aiguille dans la botte de foin » [Sharoff *et al.*, 2008]. Pour eux, la botte de foin est le Web. Le travail est donc articulé non pas autour du regroupement de ressources mais de leur recherche. Ces travaux arrivent eux aussi après la majeure partie de notre travail et apparaissent ici dans un souci de complétude (et non d'exhaustivité). Les auteurs considèrent les textes disponibles trop axés vers la littérature ou dans certains cas (spécialisés dans des domaines dont l'évolution est rapide) simplement périssables [Sharoff *et al.*, 2008, p. 268]. Ils rejoignent en cela Little, pour qui le texte authentique tel qu'il est défini dans l'enseignement est par nature éphémère [Little & Singleton, 1988, p. 2]. Pour pallier ce problème, ils proposent d'indexer automatiquement certaines ressources Web. Leur stratégie consiste en un recensement des techniques de catégorisation automatique de textes, dont ils tentent ensuite d'utiliser les plus appropriées [Sharoff *et al.*, 2008, p. 269].

Cette démarche rappelle quelque peu celle adoptée dans le projet ALICE [Levin *et al.*, 1991], une plate-forme d'apprentissage des langues et de conception de programmes d'« instruction des langues intelligemment assistée par ordinateur¹²² » [Levin *et al.*, 1991, p. 28]. Ce système qui, comme l'indique la période à laquelle il a été développé, est influencé par les tuteurs intelligents (cf. § b) p. 31) veut intégrer des techniques d'IA dans un environnement proposant des ressources de :

« (1) *syntactic parsing and pattern matching with error detection*, (2) *morphological analysis of inflected words and generation of inflectional paradigms*, (3) *on-line dictionaries and other reference material*, (4) *indexing and retrieval of examples from corpora such as videodiscs, stored audio material, or on-line texts*, and (5) *an authoring system that allows teachers to combine and configure ALICE tools and resources into instructional programs*.¹²³ »

[Levin *et al.*, 1991, p. 28]

121. <http://atomz.com/>

122. *Intelligent Computer Assisted Language Instruction (ICALI)*

123. (1) analyse syntaxique et filtrage par motifs (terme traduit d'après l'article [Wikipedia, 2008b]) avec détection d'erreur, (2) analyse morphologique de formes fléchies et génération de paradigmes flexionnels, (3) dictionnaires en ligne et autres matériel de référence, (4) indexation et recherche d'exemples dans un corpus comme des vidéo-disques, du matériel audio ou des textes en-ligne et (5) un système auteur qui permet aux enseignants de combiner et configurer des outils et des ressources d'ALICE pour en faire des programmes d'enseignement des langues.

La partie qui nous intéresse est celle désignée par le (4) dans cette énumération. Tout d'abord, il faut constater que la problématique est beaucoup plus vaste que la nôtre : l'indexation ne se cantonne pas à des textes écrits, mais peut concerner de l'oral, de la vidéo, ainsi que ce que les auteurs appellent « meta-texte » : un texte décrivant la ressource en question [Levin *et al.*, 1991, p. 49]. Ensuite, ce travail était prospectif en 1991 et nous n'avons trouvé d'autres articles au sujet de cette composante d'ALICE. D'après l'article que nous avons pu consulter, la démarche employée est une démarche d'instrumentation d'un outil de recherche d'information (textes) : CLARIT [Levin *et al.*, 1991, p. 50]. CLARIT est un système de recherche d'informations qui a participé plusieurs fois à la conférence sur la recherche de textes TREC¹²⁴ [Evans & Lefferts, 1995]. C'est un système de recherche générique, qui n'a pas été spécialement conçu pour l'enseignement des langues. De ce fait, la démarche proposée ne justifie pas d'un point de vue didactique le choix de l'indexation, sur laquelle nous ne pouvons pas avoir de recul quant à l'adéquation par rapport aux besoins des enseignants.

Notre démarche, peut être moins prompte à offrir des résultats, diffère en ce sens que nous ne partons pas des outils et des techniques, mais des besoins des enseignants. Les outils que nous utiliserons seront vraisemblablement du même ordre que ceux employés dans les autres systèmes. Cependant notre préoccupation est d'y avoir recours d'une manière qui corresponde le mieux possible aux besoins des enseignants, plutôt que de demander aux utilisateurs-enseignants d'adapter leurs requêtes aux outils.

2.4 Des EIAH aux EIAH

À travers trois grands axes (histoire et travers de l'ALAO, corpus), nous avons tenté de nous positionner vis-à-vis de champs disciplinaires proches de notre problématique, mais dont elle ne relève pas toujours de manière évidente.

Un bref recensement des différents types d'EIAH nous a permis de statuer sur le fait qu'une base de textes indexée pédagogiquement pour l'enseignement des langues n'est pas en soi un EIAH ; et ce, même si sa problématique n'est pas séparée de ce champ, de par les fonctions constituantes qu'elle aura à terme (intégration dans MIRTO), mais aussi par les implications que peuvent avoir des recherches sur les critères objectifs intervenant dans le choix de textes (modèle pédagogue, cf. § *EIAH : définition du domaine et prise de position* p. 35). Enfin, comme le signalent Contamines *et al.* (ou Vitalis en d'autres termes), les fonctions constituantes ne sont pas les seules définissant le champ instrumental d'un artefact (cf. § *Approche instrumentale* p. 61). Or, le fait que notre base de textes soit conçue à destination des enseignants n'en interdit pas des usages détournés : « C'est [l'usager] qui mène véritablement le jeu à travers la description des phénomènes d'appropriation, de détournement, de piratage, de résistance, voire de rejet » [Vitalis, 1993, p. 36]. Par définition, les fonctions constituées ne sont pas de notre ressort. Pourtant, créer un outil compréhensible et évolutif peut favoriser les processus d'instrumentation et d'instrumentalisation.

Que les usages sociaux de notre système impliquent plus directement l'apprenant ou non, notre travail s'inscrit plus systématiquement dans le champ de l'ALAO, tel qu'il est défini par Levy : « La recherche et l'étude des applications de l'ordinateur dans l'enseignement et l'apprentissage des langues » [Levy, 1997, p. 1] (cf. définition 2.2). L'ALAO, tout comme le champ des EIAH, évolue principalement de concert avec la technologie et la didactique. De la même manière que la didactique est un savoir disciplinaire (cf. § *Différentes indexations pédagogiques* p. 17), l'ALAO

124. Text REtrieval Conference : <http://trec.nist.gov/>

constitue un champ à part entière, de par son objet d'étude : l'enseignement des langues. Une étude chronologique montre que l'ALAO s'alimente des évolutions des EIAH, tout en ayant sa propre histoire. Warschauer et Healey en proposent une analyse qui l'articule en trois phases [Warschauer & Healey, 1998; Warschauer, 1996, 2000]. Cette analyse est sujette à certaines critiques mais la plupart des spécialistes du domaine lui reconnaissent l'intérêt du paradigme proposé par Warschauer et Healy pour décrire l'évolution du domaine en lien avec la technologie et la pédagogie [Delcloque & Bramoullé, 2001; Rogerson-Revell, 2007; Bax, 2003]. Au delà de cet intérêt, cette chronologie nous semble tout à fait cohérente au moins du point de vue des apports théoriques, même si du point de vue de la production de logiciels d'ALAO les critiques de Bax quant à la validité des phases nous paraissent recevables.

Dans cette chronologie, la période actuelle est étiquetée selon le paradigme « phase de l'ALAO intégratif », rapport à des objectifs intégratifs dans la sphère pédagogique (non cloisonnement des compétences travaillées) et dans la technologie (intégration de la technologie dans les pratiques). Ce qualificatif d'intégratif peut également être utilisé pour décrire le développement de plateformes de type SGA dont la conception modulaire doit permettre d'intégrer différents types de ressources. Nous pouvons citer aussi les travaux sur les corpus pédagogiques [Braun, 2007; Alcaraz & Pérez Paredes, 2008; Widmann *et al.*, 2008], qui veulent intégrer dans les systèmes développés aussi bien la constitution de ressources et la population du corpus, mais aussi des interfaces de présentation des résultats adaptées à l'enseignement ainsi que des tâches et activités toutes prêtes. De tels systèmes proposent toutes les formes d'intégration que nous avons évoquées. De notre côté, nous travaillons, comme nous l'indiquions, sur un éventuel composant au sein d'un SGA. Malgré le fait qu'il ne soit pas conçu pour être utilisé par des apprenants, il est tout à fait représentatif des nouvelles problématiques qui apparaissent avec l'ALAO intégratif (et les SGA modulaires) parmi lesquelles la classification de ressources.

Cette volonté intégrative des corpus pédagogiques rend la problématique poursuivie extrêmement vaste et les points de vue multiples : ce n'est que récemment que certaines des équipes concernées ont pu se concentrer sur l'aspect annotation pédagogique dont aurait profité le plus notre travail. Cependant l'examen de certains travers de l'ALAO nous a permis de tirer quelques conclusions. Pour pouvoir considérer le problème de la recherche de textes d'un point de vue didactique, il faut dépasser la vision purement informatique du texte et aller au delà de la séquence de caractères. Les outils qui permettent ce genre de traitement sont des outils TAL qui, du fait de la complexité de l'objet qu'ils analysent, la langue, pèchent parfois par manque de fiabilité ou de robustesse. Tout comme l'utilisation des logiciels d'ALAO est tributaire de la compréhension des outils par enseignants et apprenants, la manipulation de ces outils ne peut se faire efficacement qu'en évitant le phénomène de « *fear and awe* » décrit par Bax [2003]. Une bonne intégration des outils à un système d'ALAO demande donc d'en comprendre précisément les limites. Enfin, trop de systèmes ont été conçus en partant de la technologie pour aller vers l'enseignement, sans que les concepteurs ne s'intéressent précisément au point de vue des enseignants. Le processus de conception d'un système d'ALAO doit refléter la dualité de sa problématique, telle qu'elle est mise en évidence par la chronologie de Warschauer et Healey. Il doit cumuler les points de vues technologique et didactique, à travers une réelle communication entre ces deux pôles. Le processus incrémentiel et itératif décrit par Depover *et al.* [2000] constitue à notre sens une piste de conception nécessaire, même si nous ne pouvons pas, dans le cadre de cette thèse, en effectuer toutes les itérations. Au delà du cycle de vie du prototype, ce besoin de superposer technique et didactique justifie la démarche empirique que nous avons choisie.

Finalement, cette démarche empirique est celle qui nous a rapproché de la problématique des

corpus dans l'enseignement des langues. Ils semblaient constituer les entités les plus proches de notre base de texte : des textes réunis au sein d'une même entité, augmentés d'annotations permettant d'accéder à des faits de langue authentiques. Ils étaient de plus utilisés par des enseignants de langue depuis un certain nombre d'années. Cependant, c'est en partie aussi cette démarche empirique qui nous éloigne des corpus. Bien évidemment, le simple fait que les corpus traitent plus particulièrement *du* texte et non *des* textes nous en distancie d'ores et déjà. Cependant, c'est d'une part le fait que les corpus sont encore peu utilisés et surtout le fait que les recherches du domaine ne se soient penchées que très récemment sur les nécessités « d'annotation pédagogique » qui nous en éloignent à cette phase de notre travail.

Le plus spécifique (corpus pédagogique) s'est révélé un domaine très prometteur, mais pas suffisamment avancé dans la problématique de l'annotation pédagogique pour nous servir de base pour l'indexation pédagogique. Nous avons décidé de nous en retourner au plus générique : le domaine des EIAH dans lequel le développement des SGA a mené à une véritable réflexion sur la description des ressources pédagogiques dans une optique de normalisation devant permettre de faciliter l'intégration des ressources à un système, ainsi que l'échange de ressources d'un système à l'autre. Dans le chapitre suivant nous allons aborder ces initiatives du point de vue de la description de ressources brutes, plus précisément de textes pour l'enseignement des langues.

Chapitre 3

Les standards de description de ressources pédagogiques

3.1 La description : un premier pas vers l'indexation

Comme le suggère l'étroite relation entre la notion de langage documentaire et celle d'indexation, cette dernière repose sur une description précise des ressources. La définition de Lefèvre (cf. définition 1.3 p. 12), citée dans notre chapitre introductif, met justement en exergue cette nécessité de décrire les textes en des termes univoques et relevant de la problématique des enseignants de langue :

« Un langage documentaire est un langage artificiel qui fournit une représentation formalisée et univoque des documents d'un corpus et des questions qui intéressent un groupe d'utilisateurs, afin de permettre le repérage simple des documents du corpus qui répondent aux questions de ces utilisateurs. »

[Lefèvre, 2000]

Une fois qu'un tel langage est défini, qu'une description univoque de chaque ressource est possible, une partie conséquente du travail d'indexation est effectuée (reste à implanter ce langage documentaire dans un système performant d'interrogation de la base de textes et bien sûr à ajouter des ressources au système).

C'est afin d'avoir des indices de ce que doit être l'indexation pédagogique pour l'enseignement des langues, que nous avons décidé de nous intéresser à une problématique plus générale que la nôtre. Comme nous allons le voir dans ce chapitre, l'idée de décrire des ressources pédagogiques en termes didactiques n'est pas nouvelle : de nombreuses organisations se sont penchées sur le problème dans une optique de normalisation. Nous allons donc procéder à un inventaire des principaux acteurs, en décrivant les liens qui existent entre eux. Nous nous focaliserons ensuite sur les initiatives les plus pertinentes dans le cadre de nos recherches, afin de pouvoir établir les compatibilités et inadéquations entre ces travaux et notre problématique. Mais avant de rentrer dans le détail des standards proposés, nous allons introduire certaines notions, nécessaires à la compréhension de ce qui suit.

3.1.1 Normes et standards¹

a) Le rôle de la normalisation

Le rôle principal de la normalisation est de permettre l'interopérabilité. Comme le souligne Duval à travers l'exemple du Web [Duval, 2004], le développement des standards *HyperText Transfer Protocol* (HTTP), *Unified Resource Locator* (URL) et *HyperText Mark-up Language* (HTML) permettent de lire avec n'importe quel navigateur les pages Web écrites avec n'importe quel éditeur, stockées sur n'importe quel type de serveur capable de communiquer « en HTTP », laissant ainsi utilisateurs, développeurs et hébergeurs libres de choisir leur matériel et logiciels.

Adapté au domaine de l'enseignement et de l'apprentissage, le processus de normalisation aura pour objectif d'assurer « l'interopérabilité et la réutilisation des ressources et outils afin que les systèmes de formation en ligne soient le plus ouverts possibles et que les données et contenus soient les plus portables possibles » [Vidal *et al.*, 2004, p. 2²]. Simard, cité dans [SCTIC, 2002, p. 7], donne 5 objectifs dont l'interopérabilité et à la réutilisabilité :

1. l'accessibilité : « permettre la recherche, l'identification, l'accès et la livraison de contenus et composantes de formation en ligne de façon distribuée » ;
2. l'interopérabilité : « permettre l'utilisation de contenus et composantes développés par une organisation sur une plateforme donnée par d'autres organisations sur d'autres plateformes » ;
3. la réutilisabilité : « permettre la réutilisation des contenus et composantes à différentes fins, dans différentes applications, dans différents produits, dans différents contextes et via différents modes d'accès » ;
4. la durabilité : « permettre aux contenus et composantes d'affronter les changements technologiques sans la nécessité d'une réingénierie ou d'un re-développement » ;
5. l'adaptabilité : « permettre la modulation sur mesure des contenus et des composantes ».

Notre intérêt pour les standards de ressources pédagogiques provient en premier lieu de ce que Simard appelle l'accessibilité, puisque c'est précisément la mission première de notre prototype : permettre de rechercher, identifier, accéder à des textes, puis les récupérer pour les utiliser dans l'enseignement des langues.

b) Le processus de normalisation

Pour atteindre ces objectifs, la normalisation est décrite – toujours par Simard, à nouveau cité dans [SCTIC, 2002, p. 12–13] – comme un processus en cinq phases³ (figure 3.1).

Simard qualifie la phase d'initialisation (phase 1 dans figure 3.1) d'embryonnaire ; pour lui, elle correspond à l'élaboration d'un cahier des charges. Au cours de la phase suivante (phase 2 dans figure 3.1) des groupes et consortiums commencent à émerger, sans que cela soit systématique. C'est au cours de la phase d'évaluation (phase 3 dans figure 3.1) que les collaborations apparaissent ou se consolident autour de projets pilotes et de prototypes. Les spécifications sont donc

1. Les termes « norme » et « standard » sont utilisés selon leur sens intuitif dans un premier temps et sont définis dans le paragraphe « Définitions » en se fondant sur le processus que nous décrivons dans le paragraphe correspondant.

2. Du document électronique, doit correspondre à la page 49 de l'ouvrage papier.

3. Certains des intitulés des noms de phases sont empruntés à [Pernin, 2006].

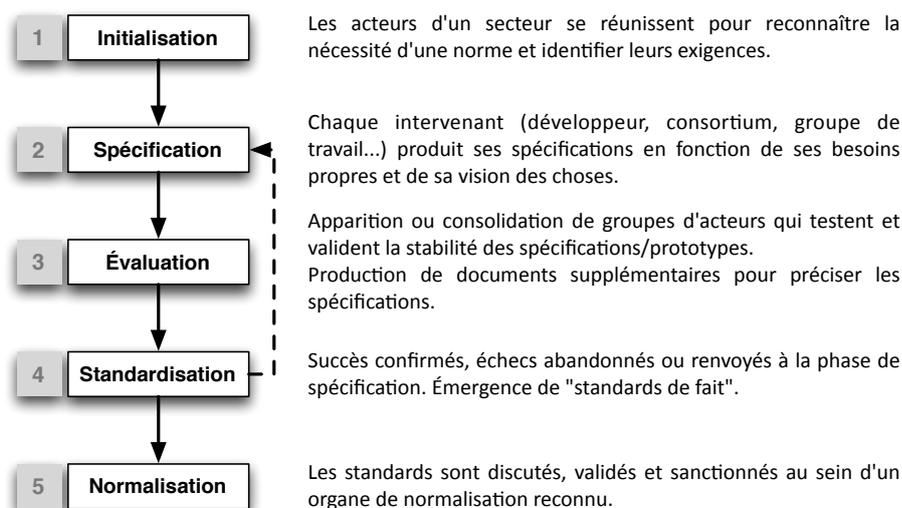


FIGURE 3.1: Le processus de normalisation (adapté de [SCTIC, 2002, p. 12–13])

testées dans la « réalité concrète » et donnent naissance à des documents complémentaires relatifs à la mise en œuvre des standards. Pour des raisons économiques, les collaborations se renforcent, de larges consortiums se forment et le positionnement stratégique s'intensifie en vue de la phase de standardisation (phase 4 dans figure 3.1). Dans cette phase, les acquis de l'expérience débouchent sur l'émergence de modèles dominants (« standards de fait »). Des organes d'accréditation ou de certification apparaissent ou se spécialisent. Enfin (phase 5 dans figure 3.1), les standards venus à maturité sont discutés, validés et sanctionnés par une instance de normalisation reconnue sur le plan national, régional ou international. La normalisation s'effectue par consensus, selon un processus ouvert visant à assurer une haute précision. Il est à noter que consensus n'est pas unanimité (cf. définition 3.1).

Définition 3.1 (Consensus) « Accord général caractérisé par l'absence d'opposition ferme à l'encontre de l'essentiel du sujet émanant d'une partie importante des intérêts en jeu et par un processus de recherche de prise en considération des vues de toutes les parties concernées et de rapprochement des positions divergentes éventuelles. » [ISO, 2004]

c) Définitions

Revenons à l'exemple de Duval, qui est d'autant plus intéressant pour nous que le terme « standard » en anglais (langue dans laquelle l'article a été écrit) recoupe trois notions différentes en français : « recommandation », « norme » et « standard ».

Pernin se sert de la description du processus de normalisation (figure 3.1) pour différencier ces notions.

Définition 3.2 (Recommandation) « Une recommandation est un ensemble de règles développées et proposées par un organisme pour apporter une solution à une question technique ou méthodologique. » [Pernin, 2006, p. 204]

Définition 3.3 (Standard) « Un standard est un ensemble de recommandations développées et préconisées par un groupe représentatif d'utilisateurs réunis autour d'une instance particulière. » [Pernin, 2006, p. 204]

Définition 3.4 (Norme) « Une norme est un ensemble de règles de conformité qui sont édictées par un organisme de normalisation, dont la portée est internationale (comme l'International Organization for Standardization (ISO)), régionale ou nationale (tel que l'Agence Française de NORmalisation (AFNOR) au niveau français). » [Pernin, 2006, p. 204]

Pernin (toujours dans le même article), tout comme Vidal *et al.*, fait référence à la notion de « standard de fait » qui désigne « les modèles qui s'imposent d'eux-mêmes » [Vidal *et al.*, 2004, p. 2⁴].

d) Enjeux de la normalisation : quelques exemples

À la lumière de ces définitions, les « standards » du Web seraient finalement plus des « recommandations » dans la mesure où ils sont produits par un consortium, sans être préconisés par une instance particulière regroupant des utilisateurs. Certains auteurs ne font pas la distinction entre standard et recommandation, [Guidon, 2001], dans ce cas les recommandations *World Wide Web Consortium* (W3C) seront des standards. Ils ne constituent assurément pas une « norme » puisque le W3C qui les édite n'est pas un organisme de normalisation. Quiconque a déjà essayé d'écrire un site Web faisant intervenir des feuilles de style *Cascading Style Sheet* (CSS)⁵ qui fonctionne aussi bien pour Mozilla Firefox, Apple Safari et Microsoft Internet Explorer en sera aisément convaincu, tant ces trois acteurs suivent différemment les recommandations du W3C.

Ces finesses dans le respect des recommandations du W3C illustrent l'importance, pour les acteurs d'un domaine – notamment les industriels – du processus de normalisation.

Prenons l'exemple de Microsoft, qui grâce entre autre à une politique de « vente liée⁶ », s'est trouvée en position de quasi-monopole sur les navigateurs Web. Selon certaines estimations, *Internet Explorer* (IE)⁷ a représenté jusqu'à 96% du marché des navigateurs⁸, pic atteint en 2002. En 1997, deux ans après la sortie de la première version d'IE, James Love et Ralph Nader prédisent :

« Si Microsoft réussit ainsi à déloger Netscape, il pourra utiliser son monopole pour contrôler les futurs standards de l'édition, des services d'information et du commerce électronique. »

[Love & Nader, 1997]

Ils expliquent l'une des stratégies de Microsoft pour arriver à ses fins en prenant l'exemple du langage Java :

« Microsoft cherche donc à neutraliser Java en utilisant la stratégie qu'il applique toujours et qui consiste à ajouter des fonctionnalités à sa propre version d'un standard commun. Si suffisamment de programmeurs utilisent ces nouvelles fonctions,

4. Du document électronique, doit correspondre à la page 49 de l'ouvrage papier.

5. D'autres recommandations issues du W3C <http://www.w3.org/TR/CSS21/>

6. Microsoft a été condamnée devant le tribunal européen pour la vente liée de *Windows Media Player*. <http://www.monde-diplomatique.fr/carnet/2007-09-17-Microsoft-condamnee>

7. Le navigateur Web de Microsoft

8. <http://www.thecounter.com/stats/2002/October/browser.php>
http://www.onestat.com/html/aboutus_pressbox4.html

leurs produits Java ne fonctionneront plus que sur les ordinateurs dotés du système d'exploitation de Microsoft. »

[Love & Nader, 1997]

Nous pouvons aisément faire un parallèle entre cette stratégie et ce qui s'est produit avec les standards du Web. Microsoft, du fait de sa position de quasi-monopole, avait tout intérêt à avoir une lecture très personnelle des recommandations⁹ : les développeurs de sites Web, qui veulent rendre leurs données accessibles au plus grand nombre, s'adaptent à cette lecture, perpétuant ainsi le monopole de Microsoft. Plus grand sera le nombre de sites adaptés à cette lecture des recommandations et pas aux autres, plus l'utilisateur aura intérêt, s'il ne veut pas voir sa navigation affectée par des incompatibilités et autres problèmes d'affichage, à adopter le navigateur prenant en compte ces finesses d'interprétation des recommandations... La lecture des recommandations de l'industriel devenant ainsi un standard de fait. Depuis 2003, cette stratégie perd de son efficacité et avec elle la position de standard de fait de la lecture Microsoft des recommandations W3C. Mozilla Firefox, soutenu par la communauté du logiciel libre, prend graduellement des parts de marché à Microsoft (aujourd'hui Microsoft ne représente avec IE « plus que » 73% de parts de marché contre 19% à son concurrent ouvert Firefox¹⁰). Par essence cette communauté suit beaucoup plus scrupuleusement les recommandations du W3C, mais nous ne réduirons pas cette inversion de tendance uniquement à ces aspects de norme au sens large du terme.

Si la décision de Microsoft de suivre plus strictement les recommandations du W3C dans la prochaine version d'IE (version 8)¹¹ devrait faire évoluer la situation dans ce cas particulier, cet exemple montre que l'interopérabilité, n'est pas une conséquence automatique de l'existence d'un standard (ou de recommandations ici). Il souligne le fait que la norme est indissociable de son utilisation et de son appropriation et par conséquent des intérêts et objectifs des acteurs du domaine.

Le processus de normalisation n'est pas non plus exempt de ce type de considération. Il est censé assurer la justesse et l'ouverture, constatés par Duval¹² [Duval, 2004] dans le domaine de la description de ressources pédagogiques. La bataille autour de la normalisation du format ouvert de documents *Office Open XML* (OOXML) de Microsoft montre bien que cette justesse et cette ouverture ne sont pas systématiques.

En l'occurrence, certains observateurs suggèrent que Microsoft veut faire valider son format OOXML – alors qu'une norme ISO existe déjà pour le format *Open Document Format* (ODF) – pour ne pas perdre les marchés gouvernementaux des pays dont les appels d'offres imposent des technologies ouvertes¹³. D'après les opposants à la standardisation du format, OOXML n'est pas réellement ouvert en ce sens que la complexité de la documentation (6000 pages) ne le rend complètement implantable que par Microsoft¹⁴. Enfin les opposants dénoncent un certain nombre

9. Et il ne s'agit pas nécessairement de fonctionnalités supplémentaires comme le suggèrent Love et Nader pour Java. Les spécifications CSS 2 (<http://www.w3.org/TR/CSS2/visuren.html>) font par exemple état, dès 1998, du positionnement fixe d'éléments, qui ne sera pris en charge par IE qu'à partir de la version 7 sortie fin 2006.

10. statistiques de juillet 2008 <http://marketshare.hitslink.com/report.aspx?qprid=1>

11. <http://www.microsoft.com/windows/products/winfamily/ie/ie8/readiness/DevelopersNew.htm>

12. Duval est un acteur très présent de ce processus de normalisation. Il est impliqué auprès de nombreux acteurs y participant comme ARIADNE, IEEE(LOM), CEN, ainsi que Prolearn (<http://www.prolearn-project.org/>) [CEN, 2008]

13. <http://blogs.zdnet.com/microsoft/?p=247>

14. http://blog.wired.com/monkeybites/2007/02/microsofts_ooxm.html (Page consultée en août 2008).

d'irrégularités dans le processus de normalisation¹⁵ participant ainsi à la bataille de lobbying¹⁶ qui anime ce processus.

e) Conséquences

Malgré un processus de standardisation apparemment clair et ouvert, il suffit de considérer l'importance que revêt actuellement la formation dans la société occidentale¹⁷ pour imaginer les enjeux de la normalisation dans le domaine de la Formation Ouverte À Distance (FOAD).

« Les acteurs en présence visent à relever des défis de nature différente, parfois en synergie mais représentant des points de vue divergents voire contradictoires. »

[Pernin, 2003, p. 181]

Les points de vue des différents acteurs influenceront donc les choix dans la conception des standards, qui une fois établis et répandus peuvent affecter les pratiques des enseignants et formateurs. Nous pouvons notamment citer l'incompatibilité du *Learning Object Metadata* (LOM) avec l'approche instrumentale décrite dans [Contamines *et al.*, 2003].

Ces considérations nous confortent dans l'idée que notre projet ne doit pas nécessairement s'inscrire dans un standard. Si parmi les standards étudiés, un (ou plusieurs) semble(nt) s'accorder avec notre conception de l'indexation pédagogique, nous le(s) réutiliserons. Mais en aucun cas nous ne modifierons notre approche dans l'unique objectif de satisfaire un standard. En ce sens, l'approche de Lebarbé et Meynard [Lebarbé & Meynard, 2008] dans le projet Corpus Littéraire et Linguistique assisté par des outils d'Intelligence Artificielle (CLELIA)¹⁸ vis à vis de la *Text Encoding Initiative* (TEI) nous semble pertinente. Les auteurs considèrent le standard comme un pivot permettant l'interaction entre les différents langages de description, chacun ayant ses spécificités mais étant convertible (avec quelques pertes parfois) vers le pivot. Lebarbé et Meynard rejoignent à ce sujet l'analyse de Burnard qu'ils citent dans leur article [Burnard, 2001]. Nous sommes cependant moins contraint dans le cadre de notre projet de rendre les textes contenus dans notre prototype indexables par d'autres systèmes. Alors que dans le cadre de CLELIA, ils ont voulu « assurer l'interopérabilité entre [leur] base documentaire et un certain nombre de bases existantes » [Lebarbé & Meynard, 2008, p. 324], la considération de ce type d'interaction est, pour nous, avant tout prospective. Si cette possibilité d'interopérabilité constituerait un plus, ce n'est pas l'objectif principal du projet. Il se pourrait d'ailleurs que les standards considérés soient plus aptes à décrire notre système en tant que ressource pédagogique, que les textes qu'il indexera.

Il est par contre primordial d'avoir la volonté de voir dans quelles mesures ces standards ne proposent pas déjà une forme d'indexation pédagogique pour l'enseignement des langues. Loin de nous l'idée de rejeter en bloc la notion de standard, et ce d'autant plus que sans elle, pas de réutilisabilité des ressources pédagogiques. Nous tentons juste de replacer ici dans son contexte la

15. <http://www.nooxml.org/irregularities>
<http://www.groklaw.net/article.php?story=20070829070630660>

16. <http://blogs.zdnet.com/microsoft/?p=1301>

17. Bouldoires parle d'ailleurs de « marché "juteux" de l'enseignement » [Bouldoires, 1997, p. 129]. À titre indicatif, d'après l'Institut National de la Statistique et des Études Économiques (INSEE) les dépenses en formation professionnelle continue et l'apprentissage en France représentaient en 2005 25,9 milliards d'euros [INSEE, 2008b] et en 2006 les dépenses pour l'éducation représentaient 6,8% du Produit Intérieur Brut (PIB), c'est à dire 121,4 milliards d'euros [INSEE, 2008a].

18. Plateforme destinée à la numérisation des manuscrits de Stendhal. Elle permet de transcrire, éditer et reclasser les feuillets préalablement scannés des manuscrits de Stendhal

problématique de la normalisation et ses éventuels effets de bord, afin de considérer les standards étudiés pour ce qu'ils sont : le fruit de la réflexion d'une communauté d'acteurs, ayant chacun sa perception de l'enseignement, sa problématique et ses objectifs.

3.1.2 Metadonnées

La plupart des standards présentés dans ce chapitre s'appuient sur la notion de « métadonnées », qui n'est pas réservée à la description de ressources pédagogiques. C'est une notion générique qui intervient dans divers domaines (notamment dans les systèmes d'information) et qui signifie littéralement « données sur les données ». Berners-Lee préconise que ces données soient « interprétables par une machine » [Berners-Lee, 1997], ce qui est essentiel dans l'optique de l'indexation (même si ce n'est pas le seul usage possible de métadonnées interprétables par une machine, cf. balise méta HTML du codage du fichier texte). La définition suivante nous paraît donc satisfaisante :

Définition 3.5 (Métadonnées) « *Les métadonnées sont des informations sur des ressources web (ou autres) compréhensibles par les êtres humains et traitables par des logiciels.* » [Bourda, 2002, 116–117]

La description de ressources pédagogiques semble convenir particulièrement bien à ce type de problématique, puisqu'il s'agit de proposer des données (description) concernant d'autres données (les ressources pédagogiques).

Les métadonnées considérées sont en général structurées autour d'un schéma, de contraintes d'usage et parfois de lignes directrices d'implantation, qui peuvent nécessiter d'être adaptés à certaines problématiques, comme le suggère Lynch à travers la notion de profil :

« *Profiles are basically customizations of the standard to particular communities of implementors with common applications requirements*¹⁹. »

[Lynch, 1997]

Duval sous-entend l'objectif d'adaptation à une application dans le terme même « profil d'application » et en propose une définition plus formelle centrée sur la manière d'atteindre cet objectif.

Définition 3.6 (Profil d'application) « *An application profile is an assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema.*²⁰ » [Duval et al., 2002]

Nous commencerons par présenter les standards que nous avons étudiés, avant d'analyser en détail les plus proches de notre problématique. Cette étude doit nous permettre de préparer en partie notre recueil de données, dont l'un des objectifs est de conclure définitivement quant à l'utilisabilité de ces standards dans notre problématique.

19. Un *profil* est finalement une personnalisation d'un standard à une communauté donnée de développeurs ayant des besoins équivalents en termes d'application.

20. Un *profil d'application* est un assemblage d'éléments de métadonnées issues d'un ou plusieurs schéma de métadonnées, combinés dans un schéma composite.

3.2 Les acteurs

Il existe plusieurs acteurs, souvent liés entre eux, qui s'intéressent à la description de ressources pédagogiques. Il est aisé de s'y perdre, de ne pas savoir quel standard étudier pour quel problème. Il faut cependant noter le travail du Comité Européen de Normalisation (CEN), qui via le *Learning Technologies Standards Observatory* (LTSO), recense les principaux standards. La plupart des informations de description des normes de ce chapitre, issues directement des spécifications des standards concernés, se trouvent résumées dans l'« observatory content » [CEN, 2008] du CEN-LTSO, qui est régulièrement mis à jour.

Dans la figure 3.2, nous essayons de donner une vision globale des principaux acteurs et standards du domaine. De ce fait, elle recense des entités de natures différentes :

- consortiums et autres organisations ;
- standards.

Les entités sont représentées par leur logo et les standards sont indiqués en noir sous l'entité correspondante, quand ils ne portent pas le même nom. Les relations peuvent aussi bien marquer l'influence d'un standard sur un autre au niveau de la conception (relation « découle des travaux »), que son intégration dans un autre, ou enfin une interopérabilité (relations « profil d'application » et « accord de compatibilité »). Nous pensons cependant que cette hétérogénéité est nécessaire pour appréhender la situation.

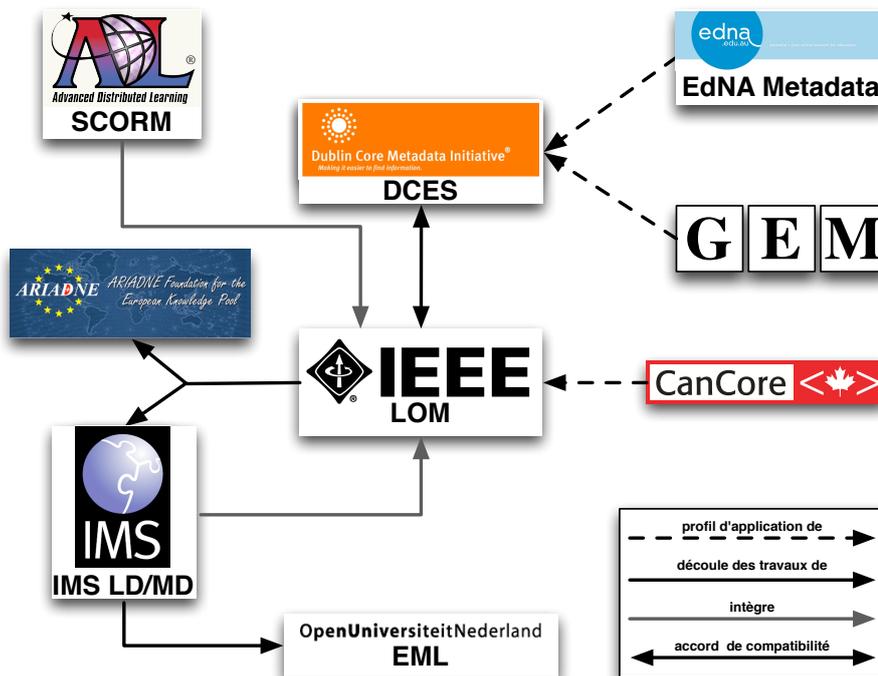


FIGURE 3.2: Acteurs et standards pour la description de ressources pédagogiques

3.2.1 AICC

Parmi les instances liées à la description de ressources pédagogiques, le premier organisme important à apparaître fut l'*Aviation Industry Computer-Based Training (CBT) Committee* (AICC).

Cette initiative industrielle touche à la standardisation dans le domaines des technologies de l'éducation depuis 1988. Elle vise à définir « des méthodes et des lignes directrices communes pour développer des formations assistées par ordinateur dans le domaine de l'aéronautique » [Pernin, 2006, p. 205]. Cette association s'est intéressée, dans les années 90, au catalogage de ressources pédagogiques pour produire en 1996 la première version du « Digital Electronic Library System » [AICC, 2002]. Active dans des projets avec IEEE P1484-LTSC, CEN-LTSC, IMS, ADL ou le groupe de travail de normalisation ISO/IEC JTC1 SC36, l'AICC a produit en 2006 un profil d'application du LOM [AICC, 2006]. L'association n'apparaît pas dans la figure 3.2 du fait de sa spécialisation dans le domaine de l'aéronautique.

3.2.2 La DCMI

Parmi les instances liées à la description de ressources pédagogiques, le premier organisme important à apparaître fut la *Dublin Core Metadata Initiative* (DCMI). Elle tire son origine de la seconde International World Wide Web Conference de Chicago où une discussion portant sur le Web et le manque de marquage sémantique a donné naissance à une journée d'étude sur le sujet à Dublin, Ohio en 1995 [CEN, 2008]. La DCMI est le résultat de cette journée. Elle est devenue une organisation tentant de promouvoir l'adoption de standards de métadonnées compatibles entre eux.

La DCMI a produit le *Dublin Core Element Set* (DCES), un ensemble de 15 descripteurs généraux :

ÉLÉMENT	DESCRIPTION
Title	Un nom donné à la ressource
Creator	Une entité à l'origine de la création du contenu de la ressource
Subject	Un sujet du contenu de la ressource
Description	Une description du contenu de la ressource
Publisher	Une entité responsable de la diffusion de la ressource, dans sa forme actuelle
Contributor	Une entité qui a contribué à la création du contenu de la ressource
Date	Une date associée avec un événement dans le cycle de vie de la ressource
Type	La nature ou le genre du contenu de la ressource
Format	La matérialisation physique ou numérique de la ressource
Identifiant	Une référence non ambiguë à la ressource selon un contexte donné
Source	Une référence à une ressource dont dérive la ressource en question
Language	Une langue du contenu intellectuel de la ressource
Relation	Une référence vers une ressource qui lui est liée
Coverage	La portée ou la couverture spatio-temporelle du contenu de la ressource
Rights	Des informations sur les droits sur et au sujet de la ressource

TABLE 3.1: Les éléments du DCES [DCMI, 2008]

Ces descripteurs sont amenés à être utilisés dans des contextes plus spécifiques :

« *This model allows different communities to use the DC elements for core descriptive information, and allowing domain specific extensions which make sense within a*

*more limited arena.*²¹ »

[Hillmann, 2005]

Dans [Duval *et al.*, 2002], deux méthodes (non-exclusives) d'adaptation d'un modèle de métadonnées à un domaine spécifique sont présentées : l'extension et le raffinement. L'extension correspond à l'ajout d'éléments qui n'existent pas dans le schéma original, alors que le raffinement consiste à ajouter des qualificatifs à des éléments existants. Les principes grammaticaux [DCMI, 2003] de la DCMI spécifient que les raffinements doivent se faire selon le « *dumb-down principle* ». Ce dernier spécifie que tout qualificatif doit pouvoir être ignoré et ce, sans que la valeur ne perde son sens. Un qualificatif doit raffiner un élément et non en changer la portée. Par exemple la DCMI propose comme raffinement de l'élément *Description* les qualificatifs *Table of content* et *Abstract*. Quelle que soit la valeur associée à l'élément (ex : 1. Description / 2. Acteurs / 3. LOM / 4. Conclusion), le qualificatif²² (*Abstract* ou *Table des matières*) ne sera pas nécessaire pour donner un sens à la valeur.

C'est dans ce contexte d'extension et de raffinement que le DCES est utilisé dans la communauté du e-learning. Un groupe d'étude travaille d'ailleurs sur l'utilisation des éléments du DCES (et leurs extensions) dans le monde de l'éducation au sein même de la DCMI²³. Ce groupe fait intervenir des membres des équipes *the Getaway to Educational Material* (GEM) et *Educational Network of Australia* (edna). Nous pouvons également faire remarquer l'existence d'un « Memorandum of Understanding between the Dublin Core Metadata Initiative and the IEEE Learning Technology Standards Committee²⁴ » [DCMI, 2000] qui stipule que les deux organisations maintiendraient la compatibilité entre leurs standards.

3.2.3 edna et GEM

edna²⁵ est un réseau communautaire australien de l'éducation et de la formation, qui regroupe aussi bien des acteurs privés que publics, dans tous les domaines de la formation et de l'éducation, avec pour objectif de soutenir l'interopérabilité dans tous les secteurs de l'éducation et de la formation en Australie dans le domaine de la recherche et de la gestion de ressources en ligne [Loiseau, 2004]. Le site de edna²⁶ héberge un certain nombre de ressources pédagogiques décrites selon le profil d'application du DCES qu'ils ont défini et offre la possibilité d'effectuer des recherches sur ces métadonnées.

Avant de produire en 2006 le edna Resources – metadata application profile [edna, 2006], l'organisation avait produit le EdNA Metadata Standard dont la dernière version date de 2002 [edna, 2002].

Le GEM est un projet financé par le "Department of Education" des Etats-Unis, dans le cadre duquel une première version d'un autre profil d'application du DCES [GEM, 2001] a été développée. Après quoi un outil (GemCat) permettant d'annoter et d'exploiter des documents selon la

21. Ce modèle permet aux différentes communautés d'utilisateurs de se servir des éléments DC pour les informations descriptives de base tout en autorisant les extensions spécifiques à un domaine et qui sont pertinentes pour un public moins large.

22. décrivant dans cet exemple la forme de la *Description*

23. <http://dublincore.org/groups/education/>

24. Organisme responsable du LOM (cf. § 3.2.5 p. 97)

25. L'acronyme edna a changé avec le temps, passant de « EdNA » à « edna ». Pour désigner l'institution, nous utiliserons l'acronyme le plus récent. Pour les titres de documents, nous conservons l'acronyme utilisé dans la publication.

26. <http://www.edna.edu.au/>

syntaxe et les spécifications d'utilisation et d'implémentation du standard [Lowe, 1999] a été réalisé. Depuis le profil d'application a évolué vers sa seconde version [GEM, 2004a]. Le site de l'organisation propose un catalogue de ressources indexé en utilisant leur profil d'application ²⁷.

Ces entités proposent toutes deux un profil d'application du DCES (cf. figure 3.2), qu'elles réutilisent dans un catalogue de ressources éducatives.

3.2.4 ARIADNE et IMS

Peu de temps après la création de la DCMI – en 1996 – le projet européen *Alliance of Remote Instructional Authoring & Distribution Networks for Europe* (ARIADNE) voit le jour et entame un travail sur le développement d'outils de production de ressources éducatives et la gestion de ces dernières, produisant en 1999 le schéma d'annotation ARIADNE Educational Metadata [Duval *et al.*, 2000].

À la même époque, l'association nord-américaine EDUCAUSE ²⁸ (anciennement EDUCOM), propose un ensemble de métadonnées fondé sur le DCES visant à l'adapter au domaine de l'éducation. EDUCAUSE lance le consortium « *Instructional Management Systems (IMS) Global Learning Consortium* », regroupant des institutions de l'enseignement supérieur et des industriels. Il vise à définir des standards pour l'interopérabilité des applications et services d'apprentissage distribués. Les travaux sur le DCES se concrétisent en 1999 avec la première version de [IMS, 2006] ²⁹.

« *Today, the main results of the IMS are in the fields of metadata, content packaging, test definition, and student and group profiling and management.* ³⁰ »

[CEN, 2008, p. 23]

ARIADNE ³¹ et IMS ³² ont participé activement à l'élaboration du standard LOM (cf. figure 3.2).

3.2.5 IEEE – LTSC

L'*Institute of Electrical and Electronics Engineers, Inc* (IEEE) est une association professionnelle à but non lucratif reconnue dans le domaine de l'élaboration de standards technologiques. Elle lance en 1997 le comité *Learning Technology Standards Committee* (LTSC) ³³, spécialisé dans les technologies de l'éducation. Parmi les groupes de travail de ce comité, le « wg12 » à l'origine du standard LOM.

27. <http://www.thegateway.org/help/search-help/browse>

28. <http://www.educause.edu>

29. <http://imsglobal.org/metadata/>

30. Aujourd'hui les principaux résultats d'IMS concernent les métadonnées, l'agrégation de contenus, la définition de tests et la gestion et la modélisation des profils des apprenants (et groupes d'apprenants).

31. « *We are also active in localization and internationalization issues, both in the IEEE/LTSC and the European Committee for Standardization Learning Technology Workshop* » [Duval *et al.*, 2001, p. 76–77].

32. « *IMS member organizations have been key contributors to the Working Draft 6.1 Learning Object Metadata (LOM) document / « ARIADNE, together with the IMS Consortium, has been one of the contributors to the different versions of LOM » »* [CEN, 2008, p. 135]

33. anciennement appelé P1484.

3.2.6 ADLnet

Le consortium états-unien *Advanced Distributed Learning* (ADL) fut créé à l'initiative du *Department of Defense* (DoD) avec comme objectifs :

« (1) promouvoir l'utilisation de l'apprentissage basé sur les technologies et le web en particulier, (2) fournir un modèle de référence permettant de garantir la qualité des contenus en termes de réutilisabilité, accessibilité, pérennité, interopérabilité et (3) fournir une base solide pour des investissements dans le domaine. »

[Pernin, 2004, p. 2]

Le consortium est à l'origine du modèle *Sharable Content Object Reference Model* (SCORM) qui réutilise le LOM [SCORM, 2006a, section 4].

3.2.7 ISO/IEC JTC1 SC36

L'organe ISO responsable de la normalisation dans le domaine des technologies de l'éducation est le ISO/IEC JTC1 SC36³⁴. Tous les organismes présentés ci-dessus y participent plus ou moins directement, via leur relations avec les autres acteurs, notamment IEEE, IMS et DCMI [ISO, 2007, p. 8–9]. Ce rapide tour d'horizon des standards nous a permis d'éclaircir certaines des relations qui existent entre eux, en entrant plus avant dans le contenu de certains standards, nous allons découvrir les autres.

3.3 Quels standards, pour quels objets pédagogiques ?

Chacun des acteurs présentés ci-dessus propose donc un ou plusieurs standards liés aux technologies de l'éducation, chacun avec ses propres objectifs ou sa vision. Notre problème est très précis : décrire pédagogiquement des textes pour l'enseignement des langues. Nous n'allons créer, dans le cadre de ce travail, ni activité, ni cours. Il nous faut donc analyser les modèles dont les objectifs sont le plus en adéquation avec les nôtres. Pour y parvenir nous nous appuyerons sur la notion d'objet pédagogique ou *learning object*, pour laquelle chaque standard a sa propre définition. [Pernin, 2003] souligne cet état de fait et propose une typologie des modèles en fonction de leur perception de la notion d'objet pédagogique. Il s'appuie pour cela sur trois standards : LOM, SCORM et *Open Universiteit Nederland* (OUN) *Educational Modelling Language* (EML). Ayant présenté les instances dans la partie précédente, nous allons décrire ici rapidement ces standards, afin de clarifier cette typologie.

3.3.1 LOM

Le *Learning Object Metadata* (LOM) est un standard de métadonnées dont la vocation est de décrire des objets pédagogiques de natures très différentes.

Définition 3.7 (Objet pédagogique (dans le contexte de LOM)) « *Any entity -digital or non-digital- that may be used for learning, education or training*³⁵. » [IEEE, 2002, p. 5]

34. « 36 subcommittee of the first joint International Standardization Organization and International Electrotechnical Commission Committee » [CEN, 2008, p. 23]

35. Toute entité numérique ou non, qui peut être utilisée pour l'apprentissage, l'enseignement ou la formation.

Cette définition est par nécessité très large : il suffit à une ressource de pouvoir être dans l'apprentissage pour être qualifiée d'objet pédagogique. En effet, il serait absurde de priver une activité conçue pour l'enseignement de ce statut. Inversement, cette définition ne fait pas de différence entre une ressource potentiellement utilisable dans l'apprentissage et une ressource qui a effectivement été utilisée, en quelque sorte un objet pédagogique « avéré ».

Ce n'est pas là le seul axe selon lequel la définition est inclusive. Pour appréhender ces objets de niveaux très différents le LOM fait intervenir des « niveaux d'agrégation ». Ces niveaux d'agrégation sont numérotés de 1 à 4, ils sont indiqués en blanc sur fond noir dans la figure 3.3.

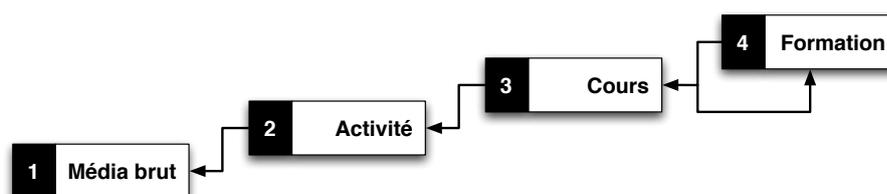


FIGURE 3.3: LOM aggregation levels (Descripteur 1.8, voir figure 3.4)

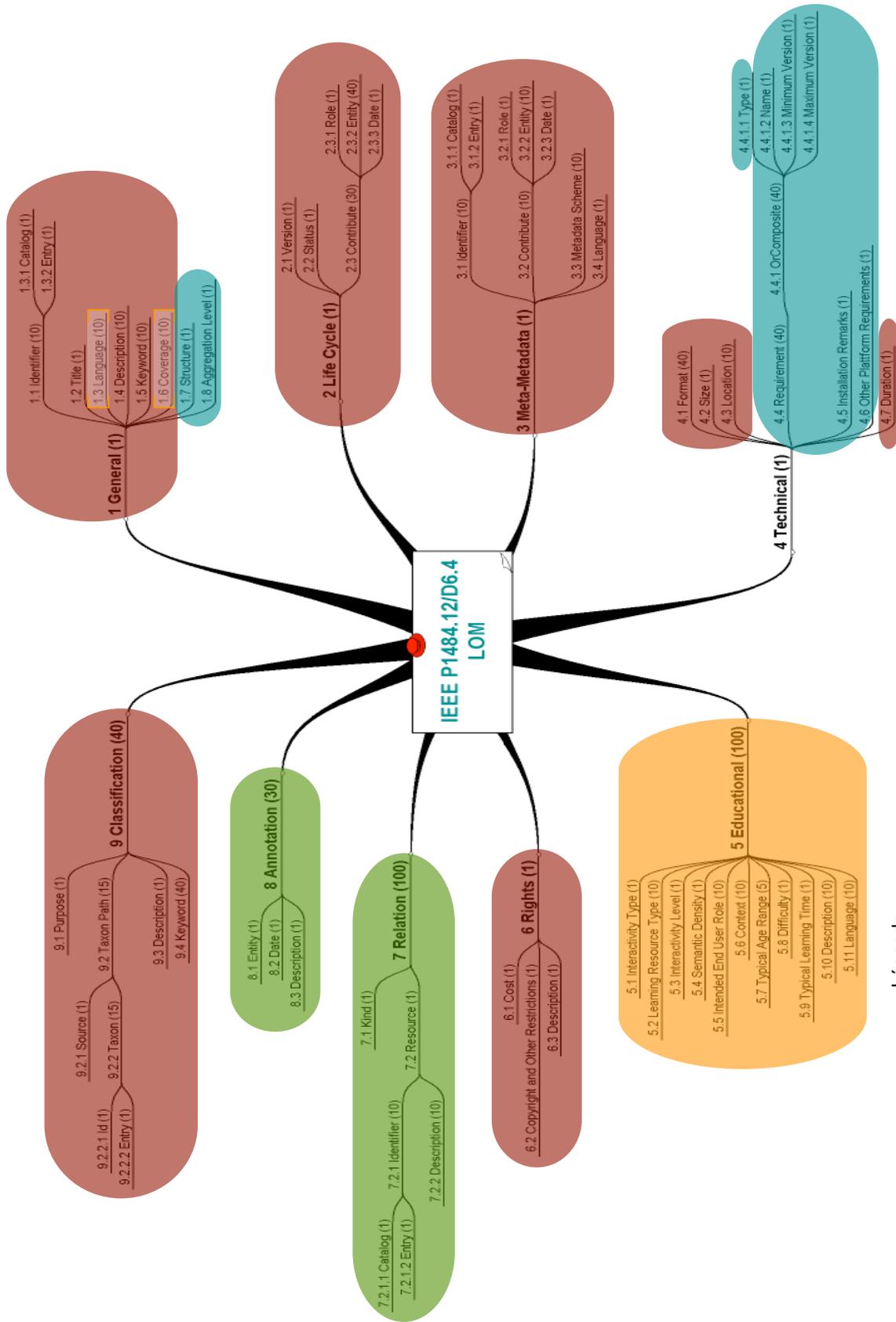
Le niveau 1 est celui des ressources brutes : une image, une vidéo sont des objets pédagogiques de niveau d'agrégation 1. C'est le niveau auquel nous travaillons : les textes contenus dans la base, sont des textes bruts non exploités pédagogiquement. Pour notre problème précis, nous pouvons nous extraire du débat de savoir si tout texte est un objet pédagogique selon la définition du LOM. En effet, les ressources brutes textuelles gagnent le statut d'objet pédagogique par le fait qu'au moins un enseignant les considère utilisables dans l'enseignement. Or, étant donné son mode de remplissage (cf. § 1.2.1 p. 8) – tout texte de la base est censé avoir été ajouté par un enseignant, qui l'aura considéré utilisable –, tout texte de notre base de données sera un objet pédagogique selon la définition du LOM.

Les objets de niveau supérieur peuvent faire référence à, ou contenir des objets de niveau inférieur. Les objets de niveau 2, par exemple des activités, pourront réutiliser des ressources décrites comme de niveau 1. [IEEE, 2002, p. 15] donne l'exemple d'une image de la Joconde comme ressource de niveau 1 et d'une leçon contenant cette image comme ressource de niveau 2. De la même manière, un cours complet sur un sujet donné (ressource de niveau 3) pourra contenir un ensemble d'activités de niveau 2 et une formation (niveau 4) regroupera différents cours (niveau 3). Un cursus complet (niveau 4) pourra contenir récursivement d'autres ressources elle-mêmes de niveau 4. Par exemple, un cursus de licence regroupera les cursus des 3 années de licence, soient 6 objets pédagogiques de niveau 4 (un par semestre).

Le LOM permet donc de décrire les objets pédagogiques, dans une acception très large de la notion, avec une série de 68 éléments (dont 10 sont des éléments composites³⁶) regroupés en 9 catégories [IEEE, 2002] (cf. figure 3.4).

Le LOM propose un cadre très générique qui doit permettre de décrire des objets pédagogiques en vue de leur indexation. Nous reviendrons plus loin sur le contenu de ces descripteurs (cf. § 3.4 p. 105).

³⁶. Les *Aggregate Data Elements* sont des éléments regroupant eux-mêmes d'autres éléments. La numérotation des éléments du LOM prend en compte la profondeur de l'élément dans la hiérarchie globale.



Légende :

- À étudier**
- Au delà de l'indexation pédagogique**
- Non discriminant**
- Ne relevant pas de l'indexation pédagogique**

Overview of LOM draft 6.4

The numbers in parenthesis show the multiplicity of the element. Numbers greater than 1 indicate the smallest permitted maximum of entries an implementation must allow. This mind map was prepared by Thomas Herrmann, Teleteach GmbH, Germany. Please send any comments to th@teleteach.de

FIGURE 3.4: Les éléments du LOM [Hodgins & Duval, 2004, tsp. 12]

3.3.2 SCORM

L'approche de ADL avec *Sharable Content Object Reference Model* (SCORM) est légèrement différente. L'objectif poursuivi est de permettre le contrôle d'objets pédagogiques sur le Web [Pernin, 2003]. En revenant aux rôles de la normalisation isolés par Simard [SCTIC, 2002], alors que le LOM se focalise sur les objectifs d'accessibilité et de réutilisabilité, le SCORM, du fait de sa problématique militaire « d'instruction à tout moment, n'importe où³⁷ », est consacré en premier lieu à l'adaptabilité et la durabilité, même si les autres composantes sont aussi revendiquées. Dans [SCORM, 2006b], une autre accessibilité est introduite, celle des prix (« affordability »), qui consiste à augmenter l'efficacité et la productivité en réduisant le temps et le coût nécessaire pour délivrer l'instruction. Le « SCORM overview » insiste sur le concept fondateur de « Web-based assumption » selon lequel « le Web fournit la meilleure opportunité de maximiser l'accès et la réutilisation des contenus pédagogiques³⁸ » [SCORM, 2006b, p. SCORM-1-5 (13)].

Pour y parvenir, le SCORM s'appuie sur quatre recommandations (« books » dans la terminologie SCORM) :

- la vue d'ensemble SCORM [2006b] est un document qui introduit, d'un point de vue global, les concepts utilisés dans les autres « livres » ;
- le *Content Aggregation Model* (CAM) SCORM [2006a] : le Modèle d'Agrégation de Contenu « fournit un guide pour l'identification des ressources de base et leur agrégation dans un contenu structuré de formation » [Pernin, 2003, p. 189] ;
- le *Run Time Environment* (RTE) : l'Environnement d'Exécution « fournit un guide pour l'exécution des contenus, la communication et le suivi des activités dans l'environnement web. Ce modèle [...] permet aux développeurs de Système de Gestion de Formation de disposer d'un cadre pratique pour intégrer des objets pédagogiques » [Pernin, 2003, p. 189] ;
- le *Sequencing and Navigation* (SN) : *séquencement et de navigation* définit des méthodes pour représenter de manière cohérente le déroulement d'une expérience d'apprentissage selon le séquencement d'activité attendu d'un Système de Gestion de l'Apprentissage (SGA) conforme au SCORM, ainsi que les fonctionnalités et comportements à implanter dans les dits SGA pour traiter ces informations de séquencement lors de l'exécution [SCORM, 2006b].

a) RTE et SN

Les « livres » de SCORM, notamment les deux derniers de notre liste, réutilisent des standards déjà développés : le RTE réutilise des fonctionnalités définies par l'AICC, quant au SN, il s'appuie sur des spécifications IMS (« Simple Sequencing Information and Behavior³⁹ »). Nous ne détaillerons pas plus ici, les spécifications relatives au RTE et au SN : elles relèvent de la création de séquences pédagogiques et de leur implémentation dans des SGA. Notre objectif principal dans le cadre de ce travail est, rappelons-le, l'indexation pédagogique de textes pour l'enseignement

37. « *“anytime, anywhere” instruction is essential to maintaining military readiness in the information age, where future forces and their support activities need to be highly adaptive to address threats effectively and rapidly.* » L'instruction à tout moment, n'importe où est essentielle au maintien d'une armée opérationnelle à l'ère de l'information, où les forces futures et leurs activités de soutien se doivent d'être hautement adaptable pour gérer efficacement et rapidement les menaces. (Août 2008 : <http://www.adlnet.gov/about/history.aspx> depuis déplacé vers : <http://www.adlnet.org/help/Pages/CommonQuestions.aspx>).

38. « *the Web provides the best opportunity to maximize access to and reuse of learning content* »

39. <http://www.imsglobal.org/simplesequencing/index.html>

des langues. Dans l'état actuel de l'avancement de nos travaux, les préoccupations au centre de ces spécifications semblent prématurées.

b) CAM

SCORM propose trois niveaux de composants (cf. figure 3.5) :

- le niveau élémentaire, équivalent du premier niveau d'agrégation de LOM. « La brique élémentaire d'une ressource pédagogique⁴⁰ » [SCORM, 2006a, p. CAM-2-3], un *asset* (éléments violets dans la figure 3.5) est une représentation électronique d'un média textuel, sonore, visuel ou de quelque autre type, qui peut être affichée par un client Web et présentée à un apprenant. Cette contrainte « numérique » restreint la notion d'objet pédagogique du LOM (qui concerne « une entité, numérique ou non »). Plusieurs *assets* peuvent être regroupés pour en créer un autre.
- le niveau supérieur, *Sharable Content Object* (SCO) (cubes bleus dans la figure 3.5) ou objet de contenu partageable, est une collection d'un ou plusieurs *asset(s)* qui utilise le RTE pour communiquer avec un SGA. C'est finalement cet aspect qui différencie un SCO d'un *asset*.
- l'agrégat de contenu ou *Content Aggregation* est « un ensemble de ressources pédagogiques structuré de façon cohérente au sein d'une entité de plus haut niveau » [Pernin, 2003, p. 189]. Comme on peut le voir sur la figure 3.5, une ressource peut aussi bien être un *asset* qu'un SCO (d'où leur regroupement au sein d'un même cadre dans la figure 3.5). L'agrégat de contenu regroupe les informations de navigation et d'enchaînement des activités. Les activités sont elles-mêmes composées récursivement d'autres activités (sous-activités), cette organisation correspondant en quelque sorte aux niveaux d'agrégation 2 à 4 du LOM. Dans la figure 3.5, l'organisation est considérée comme une activité. Elle représente l'utilisation prescrite du contenu à travers des unités structurées d'instructions, les *items*, eux-mêmes considérés comme des activités et décomposables en sous-activités. Le SN donne donc des outils pour décrire le déroulement de la séquence d'apprentissage de l'agrégat de contenu, alors que le RTE modélise l'interaction des SCO avec le SGA.

c) Les métadonnées dans le CAM

Une fois les ressources constituées elles sont assemblées en *packages* (« paquets de contenu ») suivant un profil d'application des spécifications IMS Content Packaging⁴¹, qui suggère d'inclure un manifeste dans chaque paquet de contenu. Le manifeste contient entre autres informations des métadonnées, dont le type dépend de la nature du paquet [SCORM, 2006a, Table 3.5.3a⁴²]. Pour des composants du modèle de contenu (*Content Model Components* : Assets, SCOs et Activités), SCORM recommande l'usage du LOM ou d'un profil d'application [SCORM, 2006a, § 3.4.1.5.⁴³] (cf. figure 3.2).

Nous n'irons donc pas plus loin dans l'étude de SCORM et nous focaliserons sur le LOM, la problématique du séquençement des textes au sein de séries d'activités dépassant de loin les objectifs de notre travail.

40. « *The Asset is the basic building block of a learning resource* »

41. <http://www.imsglobal.org/content/packaging/>

42. p. CAM-3-57 et CAM-3-58 (83 et 84)

43. p. CAM-3-53 (49)

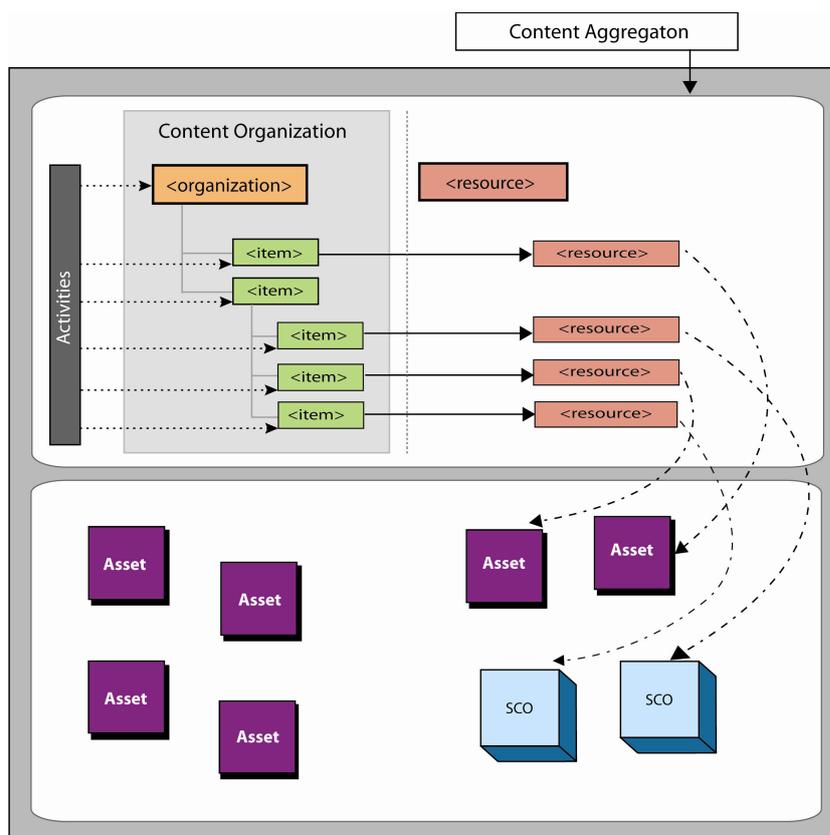


FIGURE 3.5: Illustration conceptuelle du modèle d'agrégation de contenu [SCORM, 2006a, Figure 2.1.5a, p. CAM-2-8 (26)]

3.3.3 OUN-EML / IMS-LD

Enfin, l'OUN-*Educational Modelling Language* (EML) aborde le sujet avec une autre approche, qui s'appuie sur un point de vue différent sur la question des objets pédagogiques. Koper, dans son discours inaugural de l'*Open Universiteit Nederland* (OUN) [Koper, 2000] suggère que les objets de connaissance ne sont pas le concept clé d'un SGA. Il propose, au lieu de décrire simplement les ressources, de décrire les situations effectives d'apprentissage. Les activités se trouvent au centre du modèle dont l'unité d'apprentissage est le principal élément de structuration.

Le modèle a très largement conditionné la création du *IMS-Learning Design* (IMS-LD)⁴⁴ : « *EML is no longer under development. EML was taken as a base to develop the IMS Learning Design specification.* »⁴⁵

Nous ne détaillerons pas ici ce modèle. EML se démarque du LOM, qui a un point de vue généraliste dont notre problématique pourrait être considérée comme un cas particulier. De même, alors que notre travail relève finalement de la description de certaines ressources dans SCORM, EML est centré en premier lieu sur les activités utilisant les ressources : leurs pré-

44. <http://www.imsglobal.org/learningdesign/>

45. EML n'est plus en développement, il a été choisi pour servir de fondement pour développer les spécifications *IMS-Learning Design* (IMS-LD) (cf. figure 3.2 p. 94) <http://www.learningnetworks.org/?q=EML>

requis, leurs objectifs, leur déroulement, le point de vue des apprenants, etc. Si le point de vue centré sur l'activité d'EML mérite d'être mentionné, tant il constitue une avancée dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH)⁴⁶, nous ne pourrions *a priori* guère nous préoccuper des questions centrales de la modélisation EML dans le cadre de notre travail.

3.3.4 Vers une typologie des objets pédagogiques

Revenons à ce qui nous a mené à présenter plus ou moins succinctement ces trois standards : la définition contrastive des objets pédagogiques donnée par Pernin. Cette dernière s'appuie sur les trois types de modèles dont les trois exemples que nous venons de citer sont représentatifs [Pernin, 2003, Figure 10, p. 205].

- les langages de modélisation pédagogique (comme OUN-EML), qui à travers le point de vue de l'ingénierie pédagogique, tentent de favoriser la conception de solutions de formation variées et ouvertes ;
- les modèles de mise en œuvre informatique (ex : SCORM), une approche d'ingénierie des composants logiciels, visant à :
 - fournir des solutions opérationnelles et pérennes ;
 - favoriser la réutilisabilité des composants ;
- et les langages d'indexation, dont la problématique de gestion des systèmes d'information est la plus apte à proposer une solution à notre problème, en favorisant le catalogage, la recherche et la réutilisation d'objets.

Ces différences dans les problématiques sont pour Pernin autant de points de vue des objets pédagogiques. Il propose de s'en servir pour pallier l'imprécision de la notion du LOM :

Définition 3.8 (Objet pédagogique) « *Un objet pédagogique est une entité numérique ou non, abstraite ou concrète, qui peut être utilisée, réutilisée ou référencée lors d'une formation. Il existe trois principales classes d'objets pédagogiques :*

- *les Unités d'Apprentissage qui permettent de structurer la formation et de l'organiser dans l'espace et le temps ;*
- *les Activités pédagogiques qui définissent les modalités précises d'acquisition, de validation, de communication d'une ou plusieurs connaissances ;*
- *les Ressources pédagogiques, physiques ou numériques nécessaires à la réalisation des activités.*

» [Pernin, 2003, p. 206]

Nous ne suggérons pas ici, qu'il y ait équivalence entre le type de modèle et d'objet pédagogique. Il est tout à fait possible de considérer les unités d'apprentissage sous l'angle de l'indexation. Il y a cependant un lien entre la problématique de chaque modèle et sa définition de l'objet pédagogique, ce qui a amené Pernin à proposer ces typologies.

Plus récemment, Balatsoukas *et al.* ont eux aussi examiné les différentes définitions d'objet pédagogique. Leurs points d'entrée pour leur étude contrastive étaient le niveau d'agrégation et la granularité de la notion d'objet pédagogique dans la littérature [Balatsoukas *et al.*, 2008]. Cette approche est assez cohérente avec la définition de Pernin dont les auteurs n'avaient visiblement pas connaissance. Ils opposent un point de vue « objectiviste », selon lequel « un type d'objet

46. <http://zope.cetis.ac.uk/content/20011015103421>

explicite, mesurable et standard est défini⁴⁷ » et un point de vue « relativiste » qui « symbolise les approches « ouvertes » qui considèrent n'importe quel niveau de granularité comme un objet pédagogique⁴⁸ ». Ils revendiquent cependant le fait de ne pas proposer de définition, de ne pas apporter de solution au point de vue « objectiviste », arguant du manque d'homogénéité de la notion dans la littérature, que ce soit par rapport à la granularité ou au niveau d'agrégation. Si la définition de Pernin semble pouvoir répondre à ce problème, elle n'a pas été suffisamment analysée, par exemple du point de vue de ses conséquences sur les propriétés des objets pour être considérée comme telle. À plus forte raison dans le cadre de l'approche défendue par Balatsoukas *et al.*, pour qui la notion de définition se doit d'être standardisée (cf. définition du point de vue « objectiviste ») et donc reprise par les instances en question.

La non-standardisation de la définition ci-dessus, ne nous empêche pas de la réutiliser. Dans notre cas, il n'y a pas de confusion possible, la classe d'objet concernée est bien la troisième, celle des ressources, que nous traitons du point de vue de l'indexation. Les ressources sont des textes bruts, qui une fois récupérés par un utilisateur enseignant (ou à plus long terme un SGA) seront utilisés dans une activité pédagogique.

Dans la suite de ce chapitre nous étudierons donc les principaux « langages d'indexation » du domaine de la description de ressources pédagogiques afin d'évaluer leur adéquation avec notre problématique. Cette étude va nous permettre de revenir sur les imprécisions de la notion d'objet pédagogique, puis de formuler des hypothèses que nous testerons dans notre partie expérimentale (chapitres 4 et 5).

3.4 Étude du LOM

Nous avons expliqué pourquoi nous nous centrons sur l'analyse de langages d'indexation de ressources pédagogiques. En outre, l'objectif ici n'est pas de suivre un standard et de l'adapter tant que faire se peut à la notion d'indexation pédagogique, mais bien de voir en quoi la composante « éducative » de ces modèles les différencie de modèles généralistes de description de ressources, comme par exemple le DCES. En fonction des conclusions de cette étude, nous verrons dans quelles conditions ces standards sont propices à l'indexation pédagogique de textes pour l'enseignement des langues ; si la réutilisation d'un de ces standards (en l'état ou sous forme de profil d'application) est une piste de solution viable à notre problème ; ou si nous devons poursuivre une autre direction. Nous accorderons donc une attention toute particulière à la composante pédagogique de ces standards.

Le LOM, considéré comme standard de fait dans SCORM, réutilisé dans IMS metadata (cf. figure 3.2 p. 94), s'appuyant sur ARIADNE metadata, compatible avec le DCES et enfin ayant donné lieu à de nombreux profils d'application (dont CanCore ou la norme française LOMFR⁴⁹), s'impose comme un modèle central dans le panorama des métadonnées pédagogiques. Nous commencerons donc notre analyse par lui.

Nous avons déjà expliqué la définition d'« objet pédagogique » selon LOM⁵⁰, ainsi que la notion de niveau d'agrégation ; les textes étant des objets pédagogiques de niveau d'agrégation 1, pour l'enseignement des langues.

47. *an explicit, measurable, and standard type of learning object is defined.*

48. *symbolizes the "open" approaches that treat any level of granularity as a learning object.*

49. <http://www.lom-fr.org/>

50. « toute entité numérique ou non, qui peut être utilisée pour l'apprentissage, l'enseignement ou la formation. »

Avant d'entrer dans le détail des éléments du LOM nous allons expliquer quelques notions générales de LOM.

3.4.1 Spécifications LOM : notions générales

Les spécifications LOM [IEEE, 2002] proposent donc une définition de la notion d'objet pédagogique (définition 3.7) et une série de 68 éléments, groupés en neuf catégories pour les décrire (cf. § 3.3.1 p. 98). Pour chaque élément, le schéma LOM spécifie des attributs : « **nr**⁵¹, **name**, **explanation**, **size**, **order**, **example** » et pour les éléments terminaux (non composites) : « **value space** » et « **datatype** ». Certains attributs ne nécessitent pas plus d'explication (nom, explication, exemple) nous reviendrons par contre sur les autres.

a) Size

La « taille » concerne le nombre de valeurs autorisées : certains éléments ne tolèrent qu'une valeur et une seule, mais ce n'est pas le cas de tous.

En soit, cet attribut n'est pas compliqué, les spécifications LOM introduisent cependant la notion de *smallest permitted maximum*, qui relève de l'implantation du standard. Prenons le descripteur 1.3, **Language** dont le « *smallest permitted maximum* » est 10. Cela signifie qu'une implantation du LOM peut limiter le nombre de langues autorisées dans une ressource, mais cette valeur maximale doit être supérieure ou égale à 10.

b) Order

L'ensemble des valeurs de l'élément est-il ordonné ? Cet attribut n'a pas de sens quand **size** vaut 1. Quand l'ensemble de valeurs est ordonné, la relation d'ordre est indiquée dans l'explication de l'élément.

ex : 3.2.2, **Entity**, « *Les identifiants et informations concernant les entités (i.e., personnes, organisations) ayant contribué à cet enregistrement de métadonnées*⁵². *Les entités seront inscrites en ordre décroissant de pertinence*⁵³ ». Un exemple de ce que peut être cet ordre de pertinence, serait de dire que la personne responsable de la création de l'enregistrement de métadonnées a eu un rôle plus déterminant que la personne responsable de la validation de l'enregistrement. Les spécifications ne donnent pas plus d'information sur cette notion de « pertinence ».

c) Value Space

L'ensemble de valeurs autorisées pour l'élément : un vocabulaire défini dans LOM ou une référence explicite à un autre standard. Par exemple, tous les éléments « langue » – 1.3, **Language** (de l'objet), 3.4, **Language** (de l'enregistrement de métadonnées) et 5.11, **Language** (des utilisateurs

51. l'identifiant numérique de l'élément (cf. figure 3.4)

52. Cet élément se trouve dans la catégorie 3, **meta-metadata**, les données sur les métadonnées.

53. *The identification of and information about entities (i.e., people, organizations) contributing to this metadata instance. The entities shall be ordered as most relevant first.*

finaux de l'objet pédagogique) – se réfèrent aux normes ISO 639:1988⁵⁴ pour la langue elle-même et ISO 3166-1:1997⁵⁵ pour le pays (ou subdivision du pays), permettant ainsi de spécifier plus précisément la langue (si l'annotateur l'estime nécessaire ou pertinent). L'utilisation conjointe de ces deux normes permet par exemple de spécifier que la langue employée est le français du Canada.

d) Datatype

Le type de données attendu. LOM définit 6 types de données : **LangString** (une chaîne de caractères associée à la langue dans laquelle elle est écrite), **DateTime** (un point dans le temps, chaîne de caractères), **Duration** (un intervalle de temps, chaîne de caractères), **Vocabulary** (choix parmi une liste de valeurs, avec éventuellement la source de la liste de valeurs), **CharacterString** (chaîne de caractères), **undefined** (indéfini, non utilisé dans les spécifications).

e) Conformité

Les spécifications définissent aussi la notion de conformité au LOM :

- il n'est pas obligatoire de renseigner tous les éléments LOM, une instance LOM vide est conforme ;
- si le standard n'est pas étendu, l'instance sera strictement conforme ;
- sinon elle sera conforme à condition qu'aucun des éléments ajoutés ne remplace un élément LOM.

Un enregistrement de métadonnées comprenant un élément « auteur », ne sera pas strictement conforme, puisque cet élément remplacerait une acception de l'élément 2.3, **Contribute** (cf. § 3.4.2).

f) Éléments

Nous pourrions donc réutiliser le LOM même si certains éléments ne sont pas pertinents dans le cadre de notre travail. Dans cette éventualité, il y a cependant de fortes chances que nous sortions du contexte de stricte conformité.

Tous les éléments du LOM sont listés dans la figure 3.4. Conformément à notre point de vue sur l'indexation pédagogique, nous n'allons pas tous les passer en revue, mais considérer les plus pertinents pour notre problème.

Dans la suite de ce chapitre, nous nous référerons aux éléments du LOM par leur numéro et nom, sans nécessairement les replacer dans le contexte global du modèle LOM. Par exemple, si nous nous référons à l'élément 1.1, **Identifier**, nous ne reviendrons pas nécessairement sur le fait qu'il se décompose en 1.1.1, **Catalog** et 1.1.2, **Entry** (et *vice versa*). La figure 3.4 contenant tous les éléments numérotés donne ces informations contextuelles.

54. http://www.iso.org/iso/catalogue_detail?csnumber=4766 (La norme ISO 639:1988 a été rendu obsolète par sa version suivante ISO 639-1:2002)

55. http://www.iso.org/iso/fr/catalogue_detail?csnumber=24591 (Elle aussi dépréciée et remplacée par sa version suivante ISO 3166-1:2006)

3.4.2 Éléments « non pédagogiques »

Nous dressons ici une liste des éléments « non pédagogiques ». Nous les appelons ainsi en opposition à la catégorie 5, **Educational** et du fait qu'ils ne relèvent pas directement de notre problématique. Ils peuvent s'en écarter pour plusieurs raisons, nous les avons regroupés en conséquence (cf. figure 3.4). Il n'est pas question ici de dresser une typologie ; le but est avant tout, de passer en revue les éléments, afin de nous focaliser sur ceux qui traitent de l'indexation pédagogique des ressources. Nos « catégories » ont comme unique but de clarifier notre propos.

a) Éléments ne relevant pas de l'indexation pédagogique

Le LOM comporte un certain nombre d'éléments qui ont trait à la description de ressources, sans que ce soit expressément sous l'angle de la didactique. Bien qu'ayant leur importance dans le processus de description des ressources, elles ne nous feront pas avancer dans l'aspect pédagogique de cette description.

Par exemple la catégorie 9, **Classification** pourrait nous permettre d'intégrer certains des critères que nous estimerons relevant de l'indexation pédagogique. Les spécifications LOM ne nous donnent cependant aucun indice du contenu éventuel de cet élément, de la taxonomie à employer pour l'indexation pédagogique pour l'enseignement des langues, si ce n'est une manière de le coder pour améliorer l'interopérabilité du système. De la même manière, les champs 1.4, **Description** et 1.5, **Keyword** contiennent des informations dans l'optique de l'utilisation de la ressource dans l'enseignement. Cependant, leur contenu est complètement libre. Sans lignes directrices, ils ne constituent pas une entrée vers un langage documentaire pour l'indexation pédagogique.

Les informations contenues dans 3, **Meta-Metadata** doivent permettre à un système (éventuellement autre que le nôtre) de traiter les informations contenues dans le reste des métadonnées.

4.3, **Location** permet d'accéder à la ressource à partir de l'enregistrement de métadonnées. Cet élément trouve son utilité quand ces dernières sont stockées à part. Il relève, à l'instar de 4.1, **Format**, de contraintes techniques et non d'une éventuelle description pédagogique de l'élément (au même titre que d'autres éléments de la catégorie 4, **Technical**, que nous avons considéré comme « non-discriminants »).

Le problème des droits des textes (6, **Rights**) est un problème légal, s'il doit être adressé, à plus long terme, dans l'optique d'un prototype public, il concerne à notre avis plus l'entité mettant à disposition les ressources (gestion des accès) que l'indexation pédagogique elle-même.

D'autres éléments relèvent plutôt d'une indexation bibliographique standard, comme l'identifiant de la ressource (1.1, **Identifier**), le titre de la ressource (1.2, **Title**) ou l'auteur (2.3.2, **Entity**, quand 2.3.1, **Role** est *auteur*). Le contenu de 2, **Lifecycle**, pour les autres valeurs de 2.3.1, **Role**, contient des informations de suivi de la ressource. Nous aurions aussi pu considérer que 2.2, **Status** était non discriminant du fait de son espace de valeurs (« *draft, final, revised, unavailable*⁵⁶ ») : les textes contenus dans la base sont tous censé être dans leur version finale (voir 7, **Relation** dans le paragraphe *au delà de l'indexation pédagogique*). Mais notre souci n'est pas ici d'être exhaustif.

Certaines informations sont tout simplement inadaptées à notre problème, notamment 4.7, **Duration** qui concerne la « durée d'un objet pédagogique, lu à la vitesse prévue ». Dans le

56. brouillon, final, révisé, indisponible

cas d'une vidéo qui risque d'être interrompue lors de son exploitation pédagogique ce descripteur aurait un sens, dans le cas d'un texte... Quel temps, un locuteur natif en situation authentique de communication doit-il mettre pour lire un texte ? À partir de quel degré de compréhension un texte peut-il être considéré comme lu ?

4.2, **Size**, la taille en octets de la ressource, est très peu pertinent pour un texte. Dans le contexte technologique actuel, n'importe quelle machine sera à même de lire un fichier texte (quel que soit le format utilisé). Une taille en nombre de mots aurait en revanche pu s'avérer pertinente.

Au contraire, des informations comme la langue (1.3, **Language**) et les circonstances de son écriture (lieu, époque, contexte de publication : 1.6, **Coverage**) semblent pertinentes pour notre problématique : le choix d'un texte en portugais du Portugal ou du Brésil peut relever d'une décision pédagogique, tout comme celui d'un mode d'emploi par rapport à un article de journal. Mais elles existent déjà dans des systèmes plus généralistes comme le *Corpus de Referencia del Español Actual* (CREA) de la *Real Academia Española*⁵⁷. Ces éléments, sans être spécifiques à notre problématique, occupent un statut particulier puisqu'en fonction des usages des enseignants, ils pourront être intégrés à notre acception de l'indexation pédagogique. Nous ne les détaillons pas plus, car ils ne différencient pas LOM et sa composante pédagogique d'autres problématiques, par exemple linguistique.

b) Éléments non-discriminants

Il s'agit d'éléments du LOM qui ne permettront pas de différencier un texte d'un autre. Dans le contexte de notre travail, les objets auront tous certaines caractéristiques communes, comme le fait d'être des ressources brutes (cf. 1.8, **Aggregation Level** = 1) ou d'être, *a priori* tous lisibles avec un navigateur Web (cf. 4.4, **Requirement**). Le problème le plus probable concernerait l'encodage des caractères. De nos jours la plupart des navigateurs sont capables de prendre en charge de nombreux encodages et l'avènement d'Unicode tend à limiter ce problème. De plus l'encodage du texte est typiquement le genre d'aspect technique qui doit être transparent à l'utilisateur-enseignant de langue. Dans l'optique de l'indexation pédagogique de textes, ce critère ne peut être considéré. Il représente précisément le genre de dérive de l'Apprentissage des Langues Assisté par Ordinateur (ALAO) que nous avons dénoncé dans le chapitre précédent : à savoir la nécessité pour l'enseignant d'adapter sa problématique à celle de l'informatique.

En plus des éléments déjà nommés, nous considérons comme non-discriminants :

- 1.7, **Structure** ;
- 4.5, **Installation remarks** ;
- 4.6, **Other Platform Requirements**.

c) Au delà de l'indexation pédagogique

Certains éléments ont trait à des aspects dépassant notre problématique, mais qui relèveraient de l'utilisation d'une plateforme intégrant une base de textes indexée pédagogiquement pour l'enseignement des langues. Une telle plateforme pourraient introduire des aspects collaboratifs. Sans aller jusqu'à créer des activités, un élément comme 7, **Relation** permettrait de lier différents textes issus d'un même texte source. En effet, même à partir d'un texte authentique, un enseignant peut être amené à faire quelques modifications : Vigné mentionne la possibilité

57. <http://corpus.rae.es/creanet.html>

d'adapter (modification du vocabulaire ou des tournures) ou d'abrégé un texte [Vigner, 2001, p. 54]. Dans ce cas, cette nouvelle version n'en est pas une au sens de LOM (2, **Life Cycle**), elle ne remplace pas l'objet pédagogique mais en constitue un nouveau, dédié à un public ou un type d'activité différent. 8, **Annotation** permettrait aux enseignants de faire état de leur expérience d'enseignement pour chaque texte. Si, à long terme ce type d'informations améliorerait probablement le système, ces fonctionnalités vont bien au delà de ce que nous pouvons réaliser dans le contexte de ce travail.

3.4.3 Catégorie 5, Educational

La catégorie dite pédagogique de LOM comporte 11 éléments que nous allons confronter à un exemple précis, celui d'un texte utilisable dans l'enseignement des langues. Nous avons choisi comme texte la transcription du fameux discours de Martin Luther King, Jr. du 28 août 1963 au Lincoln Memorial⁵⁸. Ce choix relève de notre expérience d'apprenant de l'anglais et donc de la certitude que ce texte peut être employé dans l'enseignement des langues (puisque nous y avons été confronté en tant qu'apprenant à plusieurs reprises).

En nous appuyant sur notre expérience d'apprenant et en essayant de décrire le texte d'un point de vue pédagogique, comme un enseignant pourraient le faire, nous nous plaçons en porte-à-faux vis à vis du postulat d'objectivité de la méthode scientifique. Valcke en rappelle, en introduction de [Valcke, 1974], certaines conditions épistémologiques parmi lesquelles « l'affirmation d'une distinction radicale entre le sujet qui observe et l'objet observé ».

Cependant, nous pensons que le choix de ce texte, malgré les biais qu'il introduit, est judicieux pour plusieurs raisons. Premièrement, parce que du fait de sa célébrité, il est probable que nous puissions trouver des utilisations avérées du texte dans l'enseignement des langues. Ensuite parce qu'il va nous permettre d'introduire une hypothèse issue de nos recherches préliminaires, effectuées en Diplôme d'Études Approfondies (DEA) [Loiseau, 2003]⁵⁹. Et enfin, parce que nous n'allons pas tirer d'autre conclusion de cet exemple que la nécessité de confronter nos hypothèses aux pratiques des enseignants.

La catégorie 5, **Educational** « décrit les caractéristiques éducatives ou pédagogiques clés de l'objet d'apprentissage⁶⁰ ». De nombreux ensembles de propriétés « éducatives » peuvent correspondre à un seul objet pédagogique, le *smallest permitted maximum* étant 100 éléments pour la catégorie.

a) Élément 5.1, Interactivity type

Pour décrire le « mode d'apprentissage prédominant supporté par l'objet d'apprentissage⁶¹ », les spécifications LOM [IEEE, 2002] proposent un vocabulaire de trois valeurs : « *active*, *expositive*, *mixed* ». L'apprentissage « actif » réside dans le fait que l'objet « induit une action productive de l'apprenant » en opposition à l'apprentissage « expositif » « qui se produit quand le travail de l'apprenant consiste principalement à absorber le contenu auquel il est exposé ». L'apprentissage « mixte » consistant en un mélange des deux.

58. « I have a dream » : <http://www.americanrhetoric.com/speeches/mlkhaveadream.htm>

59. voir § 3.5 p. 117

60. *This category describes the key educational or pedagogic characteristics of this learning object.*

61. *Predominant mode of learning supported by this learning activity*

La première remarque que nous ferons sur cet élément concerne la formulation des types d'apprentissage, qui sans faire un procès d'intention à l'IEEE, pourrait être comprise comme un jugement de valeur sur l'apprentissage selon qu'il est « actif », et donc de qualité⁶², ou au contraire « expositif »⁶³. Nous revenons avec une telle description aux enjeux, pour ne pas dire dangers, du dépôt d'une norme qui évaluerait l'intérêt d'une ressource pédagogique en termes de création technologique reléguant au second plan toute la modélisation pédagogique sous-jacente. En effet, l'activité de l'apprenant n'est pas en soi un gage de qualité de l'apprentissage, elle doit s'appuyer sur une réelle réflexion didactique ; en particulier dans un domaine comme l'ALAO, où le médium de diffusion du savoir a historiquement souvent été privilégié par rapport à la modélisation didactique (voir notamment Bangs & Cantos Gómez [2004] dans le paragraphe § *Des carences didactiques* p. 48).

Une description ambiguë Indépendamment de cette interprétation qui pourrait avoir trait à notre lecture de locuteur non natif de l'anglais⁶⁴, l'explication de l'élément ne coïncide pas avec les exemples qui sont donnés, en effet la phrase clef de cette description est à notre avis⁶⁵ :

« *An active learning object prompts the learner [...] for some [...] productive action or decision, not necessarily performed within the learning object's framework*⁶⁶. »

[IEEE, 2002, p. 24]

Cette description manque de clarté dans la mesure où les exemples de documents « expositifs » donnés vont à l'encontre de la partie que nous avons indiquée en gras. En effet, même si les exemples donnés proposent chacun un média avec les actions de l'apprenant correspondantes, les médias de la catégorie « expositive » ne se retrouvent jamais dans l'autre catégorie même avec d'autres actions. Par exemple, un questionnaire où l'apprenant choisit ou écrit des réponses est un objet actif, une vidéo que l'apprenant peut visionner, rembobiner, commencer et arrêter est un document expositif. Mais que dire d'une vidéo que l'apprenant peut visionner, rembobiner, commencer et arrêter pour répondre sur papier aux questions qui lui sont posées dans le cadre même de la vidéo ? Un document hypertexte que l'apprenant lit et dans lequel il navigue est expositif alors qu'un document hypermedia contenant une applet de simulation est mixte. En quelques recherches sur internet, nous avons trouvé ce document (figure 3.6) auquel pourrait renvoyer la transcription initiale du discours de Martin Luther King, Jr. Le document hypertextuel ainsi conçu est à notre avis tout à fait « actif », il semble l'être d'après l'explication de l'élément, mais une analyse fondée sur les exemples, privilégiant le média le classerait dans l'autre catégorie, nuisant ainsi, à défaut de consignes claires, à la cohérence d'un système indexant les ressources.

Choix d'un point de vue Outre cette difficulté dans l'annotation, la description de l'élément met en cause la définition d'objet pédagogique. En effet, comme nous l'avons fait remarquer, elle

62. « *productive action* » / « *prompts semantically meaningful input* » / « *e.g., learning by doing* » : action productive / demande à l'utilisateur des réponses sémantiquement significatives / ex : apprentissage par la pratique

63. « *the learner's job mainly consists of absorbing the content exposed to him* » / « *does not prompt any meaningful input* » / « *e.g., passive learning* » : le travail de l'apprenant consiste principalement à absorber le contenu auquel il est exposé / ne demande aucune interaction significative / ex : apprentissage passif

64. du fait de la connotation positive du terme « productif » en français et de l'accentuation de la négativité d'une expression du type « ne demande aucune interaction significative »

65. C'est nous qui mettons en gras.

66. Un objet pédagogique *actif* demande à l'apprenant une action ou une décision productive, qui ne doit pas nécessairement être exécutée dans le cadre même de cet objet

STAMP ON BLACK HISTORY

Speech

Directions: Create your own "I Have a Dream Too!" speech by filling in the blanks

"I Have a Dream Too!"

I have a dream that one day this nation will _____

I have a dream that one day _____

I have a dream that one day _____

I have a dream that _____

I have a dream today.

I have a dream that one day _____

I have a dream today.

I have a dream that one day _____

This is my hope and faith. With this faith we will be able to _____

This will be the day when _____

When we let freedom ring, when we let it ring from every village and every hamlet, from every state and every city, we will be able to speed up that day when all of God's children, black men and white men, Jews and Gentiles, Protestants and Catholics, will be able to join hands and sing the words of the old Negro spiritual, "Free at last! Free at last! Thank God Almighty, we are free at last!"

FIGURE 3.6: Exemple d'activité de production à partir du discours de Martin Luther King, Jr.
<http://tr.im/dreamtoo>

stipule que l'incitation à l'« activité » de l'apprenant doit être intrinsèque à l'objet. Dans le cadre de ressources brutes – et nous ne parlons pas ici uniquement de textes dans le contexte de l'enseignement des langues – ces ressources ne sont pas nécessairement dédiées à l'enseignement au moment de leur création. C'est même précisément là, la caractéristique d'un texte authentique. Se placer du point de vue de l'auteur du texte ou de la ressource brute (cf. exemple récurrent de la Joconde donné dans les spécifications LOM) interdirait finalement toute description du point de vue éducatif, et ce, alors que la définition de l'objet pédagogique les inclut dans le modèle, du fait qu'elles « peuvent être utilisées dans la formation ». N'ayant pas été créées pour l'enseignement, elles n'ont aucun mode d'apprentissage sous-jacent associé. Nous nous placerons donc du point de vue de l'enseignant qui estime la ressource utilisable dans le contexte de l'apprentissage. L'usage que l'enseignant en fait est dans ce cas l'objet de l'élément 5.1, **Interactivity Type**, même si cela va à l'encontre de la description faite de l'élément. Si nous ne choisissons pas ce point de vue, l'analyse du reste des éléments 5, **Educational** serait superflue pour tous les textes authentiques, puisqu'ils sont par définition dépourvus, à leur création, de portée pédagogique. À moins d'utiliser, comme le suggère [Widdowson, 1981] en présentant l'approche communicative, des textes relatifs à l'enseignement d'autres matières, écrits en L2. Dans ce cas les propriétés pédagogiques sont celles de la matière concernée et ne relèvent pas de l'enseignement des langues. L'unique élément qui aurait une réelle portée pédagogique serait 8, **Annotation** (et éventuellement 9, **Classification**, si la classification choisie est de l'ordre de l'utilisation dans l'enseignement des langues) et nous les avons écartées toutes deux, puisque leurs descriptions respectives ne nous donnent pas d'information concernant leur usage dans l'optique d'une indexation pédagogique telle que nous l'avons définie. En prenant le point de vue de l'auteur du texte, le seul cas où des textes pourraient être pertinemment décrits est celui des textes construits

– non authentiques –, puisque leur auteur serait l’enseignant qui les utilise ensuite dans son cours. Cependant nous entrerions dans une démarche excluant d’office l’approche communicative, qui est la méthode la plus utilisée (cf. tableaux 1.1 p. 23 et 4.6 p. 164.) ; ce qui irait à l’encontre de notre démarche empirique.

Il est possible que le point de vue choisi ne soit pas celui des spécifications LOM, ce qui n’est pas évident au vu de la définition d’objet pédagogique. Mais quand bien même nous nous tromperions, notre point de vue permet de confronter des descripteurs généraux à notre problématique.

b) Élément 5.2, Learning Resource Type

Cet élément est lui aussi assorti d’un vocabulaire. L’élément permet l’utilisation de plusieurs descripteurs, le type le plus dominant devant être cité en premier parmi : « *exercise, simulation, questionnaire, diagram, figure, graph, index, slide, table, narrative text, exam, experiment, problem statement, self assessment, lecture* ⁶⁷ ». Dans [Viéville & de La Passadière, 2003, p. 137], les auteurs font remarquer que le vocabulaire contient « des informations à la fois sur la forme du document (diapositive, table, index...) et sur l’utilisation pédagogique que l’on peut en faire (exercice, simulation, examen...) ». Ils font aussi remarquer que l’élément est ordonné : « on peut alors se demander sur quels critères un auteur va pouvoir établir cet ordre ».

Même indépendamment de ces remarques, les termes du vocabulaire semblent parfois mal choisis. Prenons l’exemple de « *narrative text* ». Si le descripteur est « texte narratif » et non « texte », cela signifie que la composante narrative est importante. Cette dernière est présente dans le discours de Martin Luther King, Jr., mais ce n’est pas la composante centrale de ce discours. La narration y est une figure rhétorique. Le texte pourrait être considéré comme une conférence/cours magistral (« *lecture* »), mais ce n’est pas vraiment le cas non plus.

Dans le cas d’une ressource textuelle du type « mode d’emploi ⁶⁸ », le vocabulaire est encore moins adapté.

c) Élément 5.3, Interactivity Level

Le niveau d’interactivité est décrit selon une échelle à cinq degrés (« *very low, low, medium, high, very high* »), dont la signification est intrinsèquement dépendante des usages d’une communauté d’utilisateurs, comme le signalent les spécifications. Dans ce cas, « l’interactivité se réfère au degré avec lequel l’apprenant peut influencer l’aspect ou le comportement de l’objet d’apprentissage ⁶⁹ » [IEEE, 2002, p. 25].

Outre le problème d’associer un sens aux termes du vocabulaire, la difficulté, qui va devenir récurrente dans le contexte de l’indexation pédagogique de notre texte sera d’anticiper les différents usages potentiels du discours de Martin Luther King, Jr. Nous avons, dans la figure 3.6, proposé une utilisation avérée du texte. Le recours à un moteur de recherche, nous a permis d’en trouver rapidement deux autres. Un exercice de vocabulaire, destiné à des apprenants de L1 anglais dans

67. *exercice, simulation, questionnaire, diagramme, figure, graphique, index, transparent, tableau, texte narratif, examen, expérience, problématique, autoévaluation, conférence*

68. D’après Sophie Moirand, des « consignes écrites de matériels professionnels » constituent des « documents dans leurs fonctionnement réels » propice à une utilisation dans l’approche communicative [Moirand, 1982, p. 48].

69. *Interactivity in this context refers to the degree to which the learner can influence the aspect or behavior of the learning object.*

le contexte d'un travail sur les droits civiques (figure 3.7) et l'inclusion potentielle « en fonction du niveau des élèves » dans une séquence traitant du thème de la ségrégation en première année de Brevet d'Études Professionnelles (BEP)⁷⁰. Enfin, pour poursuivre cette liste non exhaus-

I have a Dream
by Martin Luther King, Jr.

Hint Can you guess the words to the 'Dream' speech? If you get stuck, click on the hint button.

The following are several excerpts of the speech delivered by Dr. King on the steps at the Lincoln Memorial in Washington D.C. on August 28, 1963. To read the full text, just click here for the [full text](#).

...I have a [] that one day this nation will rise up and live out the true [] of its creed: "We hold these [] to be self-evident: that all men are created equal."
 ...I have a dream that one day on the red hills of Georgia the sons of former slaves and the [] of [] slaveowners will be able to [] down together at a table of []. I
 have a dream that one day even the [] of Mississippi, a desert state, sweltering with the heat of injustice and [], will be transformed into an oasis of [] and justice. I have a
 [] that my four children will one day live in a [] where they will not be [] by the [] of their skin but by the content of their []. I have a dream today.
 ...Let [] ring from every hill and every molehill of Mississippi. From [] mountainside, let freedom []. When we let [] ring, when we let it ring from every []
 and every hamlet, from every state and every [], we will be able to speed up that day when all of [] children, black men and [] men, Jews and [], Protestants and
 [], will
 be able to join [] and sing in the [] of the old Negro [], "Free at [] free at last! thank God [], we are [] at last!"

Check

[Go to Mystery Boxes](#) [Go back to previous page](#)

FIGURE 3.7: Exemple d'activité de vocabulaire à partir du discours de Martin Luther King, Jr. <http://tr.im/dreamlac>.

tive des usages possibles du discours du Lincoln Memorial, nous pouvons proposer nous-même un exercice structural, généré avec Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO), sur l'opposition entre « *will* » et « *shall* », qui d'après [Adamczewski & Gabilan, 1993, p. 53], fait partie de la problématique de la maîtrise de ces deux modes⁷¹ (cf. figure 3.8).

Le niveau d'interactivité, pour les seules utilisations possibles trouvées ici, pourrait par exemple être « *very low* » pour l'exercice de production et l'inclusion dans la séquence de BEP, « *medium* » pour l'activité de vocabulaire, et enfin « *high* » pour l'activité issue de MIRTO, puisque chacun des mots est cliquable afin d'en donner la nature. Il est à noter que cette même activité imprimée sur papier serait peut-être d'un niveau d'interactivité « *low* ». Rien que pour ces utilisations, qui ne représentent vraisemblablement pas tous les usages possibles du texte dans l'enseignement des langues, nous aurions besoin de trois ou quatre éléments de la catégorie 5, Educational, un par utilisation, à chacun correspondant une valeur de 5.3, Interactivity Level.

d) Élément 5.4, Semantic Density

La densité sémantique est évaluée à l'aide de la même échelle à cinq degrés que le niveau d'interactivité. Comme lui, il dépend intrinsèquement des usages d'une communauté d'utilisateurs, limitant par là même les objectifs d'accessibilité et d'interopérabilité. Là encore, la description est assez floue et surtout assez peu adaptée à des ressources brutes. Tous les exemples donnés considèrent la densité sémantique du point de vue de l'objectif de la leçon. Or, comme nous l'avons dit, une ressource brute n'a pas nécessairement d'objectif intrinsèque. La mise en œuvre de cet élément dépendra donc de l'activité qui l'utilise. Par exemple, considérant le discours de Martin Luther King, Jr. sous l'angle de l'opposition entre « *will* » (27 occurrences) et « *shall* » (5

70. <http://lve.scola.ac-paris.fr/anglais/segregation.php>

71. Comme cette utilisation du texte n'est pas avérée, il se peut que les usages de « *will* » et « *shall* » ne soient pas parfaits pour l'exercice. Elle sert juste d'illustration ici.

Remplir les blancs avec le modal approprié ("will" ou "shall")

I say to you today, my friends, that in spite of the difficulties and frustrations of the moment, I still have a dream. It is a dream deeply rooted in the American dream.

I have a dream that one day this nation rise up and live out the true meaning of its creed: "We hold these truths to be self-evident: that all men are created equal."

I have a dream that one day on the red hills of Georgia the sons of former slaves and the sons of former slaveowners be able to sit down together at a table of brotherhood.

I have a dream that one day even the state of Mississippi, a desert state, sweltering with the heat of injustice and oppression, be transformed into an oasis of freedom and justice.

I have a dream that my four children one day live in a nation where they not be judged by the color of their skin but by the content of their character.

I have a dream today.

I have a dream that one day the state of Alabama, whose governor's lips are presently dripping with the words of interposition and nullification, be transformed into a situation where little black boys and black girls be able to join hands with little white boys and white girls and walk together as sisters and brothers.

I have a dream today.

I have a dream that one day every valley be exalted, every hill and mountain be made low, the rough places be made plain, and the crooked places be made straight, and the glory of the Lord be revealed, and all flesh see it together.

This is our hope. This is the faith with which I return to the South. With this faith we be able to hew out of the mountain of despair a stone of hope. With this faith we be able to transform the jangling discords of our nation into a beautiful symphony of brotherhood. With this faith we be able to work together, to pray together, to struggle together, to go to jail together, to stand up for **freedom** together, knowing that we be free one day.

freedom
noun
noun
sg

FIGURE 3.8: Exemple d'exercice lacunaire sur l'opposition entre « will » et « shall » à partir du discours de Martin Luther King, Jr.

Exercice généré avec la version de 2005 du prototype de la plateforme MIRTO.

occurrences, dont 4 dans une phrase contenant 2 occurrences de *will*), le texte est beaucoup plus dense que sous celui de l'opposition entre « *would* » (2 occurrences) et « *should* » (0 occurrence).

e) Élément 5.5, Intended End User Role

Cet élément est associé à un vocabulaire ayant 4 valeurs : « *teacher, author, learner, manager* ». Dans le cadre de notre problématique, ce sont les cas d'utilisation du système qui dicteront le remplissage de cet élément. Les textes sont destinés à être utilisés par des enseignants dans le cadre de leurs cours. Toutefois, le choix de l'intégration d'un texte à un objet pédagogique de granularité moindre fait de l'enseignant un auteur (d'activité, d'une séquence). Enfin dans son utilisation du texte dans une séquence, l'enseignant-auteur va confronter les apprenants au texte. Le LOM fixant le *smallest permitted maximum* de cet élément à 10⁷², cet élément, non-discriminant pour notre problématique, aurait toujours pour valeur : « *author, teacher, learner* ».

Pour préciser les compétences développées ou requises par l'apprenant, le système propose l'utilisation de la catégorie 9, **Classification**.

f) Élément 5.6, Context

Le contexte d'utilisation représente l'environnement dans lequel la ressource sera utilisée. Un vocabulaire de 4 mots est à nouveau proposé : « *school, higher education, training, other*⁷³ ». Les spécifications suggèrent l'utilisation d'un raffinement de l'ensemble des valeurs (cf. § 3.2.2

⁷². ce qui est assez étrange pour un vocabulaire de 4 mots, ce *smallest permitted maximum* est donc fixé dans l'optique de l'extension du vocabulaire.

⁷³. enseignement primaire et secondaire, enseignement supérieur, formation continue, autre

p. 95), et l'utilisation conjointe des deux vocabulaires, ce qui est tout à fait cohérent étant donnée l'influence du système éducatif de chaque pays sur un tel élément.

Ici encore, le problème des différentes utilisations potentielles d'un même texte se pose.

g) **Élément 5.7, Typical Age Range**

Comme son nom l'indique, cet élément concerne l'âge typique de l'utilisateur de la ressource. Dans notre triptyque « auteur-enseignant-apprenant » (cf. § 5.5, **Intended End User Role**), nous considérerons que l'utilisateur concerné est l'apprenant, sans quoi l'élément n'aurait que très peu d'intérêt. Le niveau des apprenants est censé être plus ou moins pris en compte par la notion de contexte (et de difficulté), dans notre cas l'information pertinente serait donc l'âge du public pour lequel le texte est adéquat, de par son contenu, indépendamment de son usage dans l'enseignement des langues.

h) **Éléments 5.8, Difficulty, 5.9, Typical Learning Time, 5.10, Description et 5.11, Language**

Nous traitons ensemble ces quatre éléments pour plusieurs raisons : nous pensons avoir proposé une critique suffisamment détaillée jusqu'ici pour avoir déjà une idée des problèmes potentiels liés à l'utilisation de LOM dans notre travail. Ces problèmes relèvent principalement de la multiplicité des annotations possibles pour notre unique ressource, le discours de Martin Luther King, Jr. au Lincoln Memorial. Traiter ensemble ces éléments va nous permettre d'illustrer plus précisément cette problématique.

5.8, Difficulty concerne la difficulté pour le public « type » de la ressource et s'exprime sur une échelle à 5 niveaux : « *very easy, easy, medium, difficult, very difficult* ». La durée typique d'apprentissage est une approximation du temps nécessaire au public type pour effectuer l'activité liée à l'objet pédagogique. Ce n'est pas ici une approximation que de parler d'une activité liée à l'objet, puisque il s'agit d'une des rares occurrences – dans cette catégorie **5, Educational**⁷⁴ – de mention explicite du fait que le travail lié à l'objet peut être effectué dans le cadre de l'objet lui-même mais aussi avec l'objet⁷⁵ [IEEE, 2002, p. 30]. La description est, comme son nom l'indique, un commentaire sur comment utiliser l'objet pédagogique (ce qui confirme notre hypothèse de nous placer du point de vue de l'enseignant qui juge l'objet utilisable dans le contexte de l'enseignement et non dans celui de l'auteur du texte). Enfin, la langue est celle des utilisateurs finaux de la ressource et non nécessairement celle de la ressource.

5.10, Description est donc l'élément grâce auquel nous pourrions indiquer que le discours de Martin Luther King, Jr. peut être utilisé dans différents contextes. À chaque contexte correspondra un enregistrement **5, Educational**, avec les valeurs des autres champs correspondant à cet usage. Comme le font remarquer les spécifications LOM⁷⁶, les éléments **5.8, Difficulty** et **5.9, Typical Learning Time** dépendent intimement d'autres éléments. En effet, la difficulté d'un texte, d'une activité ou d'une ressource dépendra directement du public auquel il est destiné, décrit par les éléments **5.6, Educational Context** et **5.7, Typical Age Range**. Nous prétendons

74. voir aussi § **Élément 5.1, Interactivity type** p. 110.

75. « *to work through this learning object* », puis « *to work with the learning object* ».

76. « *can be characterized by data elements 5.6: Educational.Context and 5.7: Educational.TypicalAgeRange* » [IEEE, 2002, p. 29 et 30], notes ajoutées aux éléments **5.8, Difficulty** et **5.9, Typical Learning Time**.

qu'il dépendra tout aussi directement de la « description ». Selon que le texte est donné tel quel ou qu'il est préconisé de l'intégrer à un moment précis d'une séquence pédagogique qui doit préparer à sa compréhension, la difficulté sera directement affectée, et avec elle la durée de l'activité. Enfin la L1 des apprenants aura une influence sur la difficulté et le temps d'apprentissage d'une ressource :

« On admettra aisément que la relation de xénité peut varier considérablement : les langues sont plus ou moins étrangères les unes aux autres et le sont, par conséquent, pour leurs locuteurs respectifs. »

[Dabène, 1995]

Ces éléments mettent en exergue la difficulté à proposer une annotation exhaustive des textes dont il semble que les usages potentiels puissent être nombreux.

3.5 Les langages d'indexation de ressources pédagogiques

De par la nature des différents standards, nous avons pu nous focaliser sur les langages d'indexation de ressources pédagogiques. Dans ce contexte, le LOM fait figure de standard central, nous l'avons donc revu en détail. Notre étude soulève un certain nombre de questions relatives à notre travail et un certain nombre de commentaires généraux, qu'il nous faut mettre en perspective avec les autres langages d'indexation de ressources pédagogiques, si nous voulons pouvoir en tirer des conclusions.

3.5.1 LOM, un modèle inadapté à notre travail ?

a) Point de vue

À travers la description de l'élément 5.1, *Interactivity Type* nous avons pu soulever le problème du point de vue de l'annotation, qui a par ailleurs été relevé dans un cadre plus général. Contamines *et al.* soulignent que le schéma LOM devrait permettre de distinguer si les informations inscrites dans les éléments de données de la catégorie 5, *Educational* reflètent la logique d'utilisation de la ressource pour son auteur, pour une personne qui propose la ressource ou si elles sont une expression de la logique d'usage rapportée par un utilisateur de la ressource souhaitant la recommander. « Car la manière d'interpréter les informations contenues dans cette catégorie va dépendre du statut et des motifs de la personne qui a inscrit ces informations » [Contamines *et al.*, 2003, p. 170].

Ce problème est encore plus flagrant dans le cas des ressources brutes puisque l'auteur de la ressource peut tout à fait n'avoir aucune logique pédagogique pour l'utilisation de la ressource qu'il a produite. Pernin fait d'ailleurs remarquer « une certaine incohérence entre la définition générique des objets pédagogiques proposée par IEEE et les éléments permettant de les décrire (en particulier, il est complexe de décrire des entités non numériques) » [Pernin, 2004, p. 2]. Même si sa remarque vise particulièrement les catégories plus techniques, spécialement inadaptées aux ressources non numériques – nous avons notamment fait remarquer l'inadéquation totale des éléments 4.7, *Duration* et 4.2, *Size* pour décrire des textes – elle peut être étendue aux ressources brutes, notamment par rapport à cette notion de point de vue de l'annotation.

b) Types de données

Comme nous l'avons mentionné au cours de l'analyse, les vocabulaires proposés pour les valeurs des éléments de la catégorie 5, **Educational** ne sont pas toujours adaptés et parfois flous tant qu'ils ne s'inscrivent pas dans les pratiques d'une communauté d'utilisateurs. À quoi Pernin ajoute que des éléments comme 5.3, **Interactivity Level** et 5.4, **Semantic Density** « semblent difficiles à renseigner et à exploiter. Qui va décider de déclarer une ressource “à densité sémantique très faible” et “à niveau d'interactivité très bas” ? » demande-t-il [Pernin, 2003, p. 188].

c) Un cadre d'utilisation très large

L'exemple que nous avons choisi soulève un certain nombre de questions de par la quantité d'utilisations pédagogiques dont il peut faire l'objet. Les métadonnées doivent permettre de décrire des propriétés intrinsèques aux objets. Or, comme nous avons pu le faire remarquer, dans le cadre de notre exemple, plusieurs n-uplets de valeurs sont possibles et nous n'avons aucune certitude quant à l'exhaustivité de l'annotation qui serait ainsi effectuée. Alors que dans le cadre d'une activité d'apprentissage constituée dans un objectif précis, la majeure partie des propriétés pédagogiques de l'objet en seront des fonctions constituantes (usage prescrit) et pourraient être considérées comme des propriétés intrinsèques de l'objet, nous pouvons nous demander si une telle description peut exister pour des ressources brutes qui n'ont pas nécessairement été créées pour être utilisées dans l'enseignement ou l'apprentissage. Si, comme semble l'indiquer notre exemple, les textes peuvent avoir de nombreux usages, sans que l'un ne s'impose par rapport aux autres, une description exhaustive de ce type de ressource avec les éléments du LOM paraît difficile. Pour Pernin :

« le modèle proposé souffre d'un certain nombre d'imprécisions ou d'ambiguïtés, parmi lesquelles nous pouvons souligner [...] la volonté d'intégrer au sein d'un même modèle des entités de niveau conceptuellement très différent : les ressources nécessaires à la mise en place d'activités pédagogiques et les activités elles-mêmes. »

[Pernin, 2004, p. 2]

L'objectif de notre travail n'est pas de proposer une collection de ressources accessibles selon le prisme de l'usage qui en a été fait par des utilisateurs, mais une base de textes qui permette aux utilisateurs de formuler des requêtes selon des termes relevant de leur problématique et qui, dans l'idéal, donnerait accès à toutes les ressources de la base qui puissent satisfaire ces requêtes.

3.5.2 Les autres langages d'indexation

Certaines des conclusions que nous avons tiré de l'étude du LOM sont extensibles aux autres standards présentés dans la figure 3.2. En effet, l'annexe B des spécifications du LOM [IEEE, 2002, p. 44] rappelle la compatibilité entre le LOM et le DCES en indiquant pour chaque élément du DCES les éléments du LOM correspondant. À l'exception de **DC.Type**, pour lequel l'élément 5.2, **Learning Resource Type** joue le rôle de vocabulaire de valeurs aucun des éléments dits pédagogiques du LOM n'est représenté. L'étude de edna et GEM peut donc se résumer aux éléments ajoutés par leurs profils d'application respectifs ; les autres éléments relevant de la DCMI.

a) edna

FIGURE 3.9: Interface de recherche avancée dans le dépôt edna
<http://www.edna.edu.au/edna/go/search>

Le profil d'application edna [2006] a recours à quatre espaces de nommage⁷⁷ : DC [DCMI, 2008], EDNA, AGLS [AGLS, 2002] et AD. Le premier concerne les éléments repris au DCES. Le second concerne, comme son nom l'indique, les éléments définis par edna. *Australian Government Locator Service* (AGLS) est un standard de métadonnées que les « départements et agences gouvernementaux [australiens] peuvent utiliser pour améliorer la visibilité et l'accessibilité de leurs services Web⁷⁸ ». Et AD est utilisé pour des éléments générés automatiquement pour décrire les métadonnées dans le système d'administration edna [edna, 2006, p. 2]. Les six éléments de l'espace de nommage EDNA sont particulièrement liés à l'ajout de ressources au *repository* edna, dans lequel les recherches s'effectuent par mots-clés, élément par élément (cf. figure 3.9). Avec des éléments comme `EDNA.Review` et `EDNA.Reviewer`, l'organisation accorde une importance toute particulière au suivi des ressources.

edna se différencie du LOM par une approche plus pratique, implantée autour du dépôt de ressources edna et par un accent mis sur l'évaluation des ressources, sans apporter de nouvel élément à notre réflexion cependant.

77. ou espace de noms : « En informatique, un espace de noms est une notion permettant de lever une ambiguïté sur des termes qui pourraient être homonymes sans cela. Il est matérialisé par un préfixe identifiant de manière unique la signification d'un terme. Au sein d'un même espace de noms, il n'y a pas d'homonymes. » [Wikipedia, 2009b]

Dans le cadre d'un profil d'application, il ne s'agit pas seulement de problèmes d'homonymie, mais de savoir à quelle spécification, quel schéma, se rapporte chaque élément.

78. "The AGLS Metadata Standard is a set of 19 descriptive elements that government departments and agencies can use to improve the visibility and accessibility of their web services".

<http://www.naa.gov.au/records-management/publications/AGLS-Element.aspx>

b) GEM

Le GEM propose lui aussi un profil d'application du DCES [GEM, 2004a]. Ce profil d'application propose trois versions de GEM : une version minimale (« Gateway Lite »), une version intermédiaire (« Gateway Full ») et la version complète (GEM). Le profil d'application propose des raffinements d'éléments DCES, ainsi que des extensions, parmi lesquelles l'extension MARC⁷⁹, proposée par la DCMI [DCMI, 2005]. Pour chaque élément, une équivalence avec le LOM est indiquée (si elle existe, ce qui est le cas la plupart du temps). Le GEM spécifie par exemple un raffinement **prerequisites** qui doit contenir les pré-requis pour l'utilisation d'une ressource ou une extension **instructionalMethod** (« Un processus, utilisé pour engendrer des connaissances, des attitudes ou aptitudes, que la ressource est conçue pour développer⁸⁰ » [GEM, 2004b]). Aucune de ces notions n'existent dans LOM. Elles indiquent un point de vue plus orienté vers les pratiques pédagogiques. Toutefois, si nous désirions décrire notre document selon ce schéma, les mêmes problèmes que ceux survenus avec le LOM se révéleraient : chacun de ces éléments ne semble prendre sa valeur qu'en fonction de l'activité réutilisant le texte considéré.

L'influence de la méthode d'enseignement sur l'utilisation du texte semble évidente dans la mesure où même si elle peut convenir à diverses méthodes, rien n'empêche d'en indiquer plusieurs. Cela ne remet cependant pas en question le fait que le choix de l'utilisation du texte s'inscrit dans une démarche pédagogique de l'enseignant, qui sera difficile à spécifier *a priori* de manière exhaustive pour tous les usages possibles de la ressource. Quant aux pré-requis, il nous semble qu'ils peuvent être différents selon l'utilisation d'un même texte : par exemple la séquence de BEP, en opposition à l'utilisation du texte de Martin Luther King dans l'activité sur l'opposition entre « will » et « shall ». La première suppose un travail sur la ségrégation et des connaissances en vocabulaire, mais elle ne demande pas d'avoir travaillé sur l'opposition entre les modaux « will » et « shall » dont la maîtrise ne semble pas nécessaire à la compréhension du texte.

3.5.3 Conclusion

D'après nos remarques, qui recoupent celles d'autres chercheurs (cf. § 3.5.1 p. 117), le LOM souffre de certaines insuffisances structurelles, qui peuvent le rendre difficile d'utilisation. Il propose une structuration très détaillée (avec quels éléments décrire les objets), mais reste assez vague quant à comment l'utiliser (comment remplir les éléments) : ensembles de valeurs parfois inadaptés ou complexes, pas de spécification précise du point de vue de la description, notion d'objet pédagogique plus large que les spécifications ne le permettent réellement. Ces insuffisances nuisent à l'interopérabilité des ressources ainsi annotées, mais aussi à l'accessibilité, puisque les différences inter-annotateurs diminuent la cohérence d'une collection de ressources.

Pour pallier ce type de problème, des profils d'application peuvent être définis, comme le fait par exemple CanCore avec le LOM (cf. figure 3.2 p. 94) : « CanCore vise à simplifier et à interpréter [le LOM] pour faciliter le travail de conception, d'élaboration et d'indexation des personnes chargées de la mise en application et des concepteurs de fiche » [Friesen *et al.*, 2003].

Dans certains cas, les profils d'applications ont pour objectif d'adapter un standard à une problématique donnée, comme edna et GEM qui adaptent la DCMI à la problématique de l'apprentissage. Ces standards, bien que plus directement applicatifs – puisque l'un comme l'autre utilisés dans une banque de ressources pédagogiques –, ne résolvent pas les problèmes posés par

79. <http://www.loc.gov/marc/>

80. *A process, used to engender knowledge, attitudes and skills, that the resource is designed to support.*

notre étude du LOM. Bien sûr, les standards étudiés sont très généraux et ne permettent pas une indexation pédagogique de textes pour l'enseignement des langues. Mais en amont de ce manque de spécificité, ces problèmes découlent de la nature même des objets que nous voulons décrire. Alors que les descripteurs pédagogiques utilisés par les différents standards pourraient pour la plupart être pertinents pour l'indexation de ressources textuelles pour l'enseignement des langues, le fait d'avoir affaire à des ressources brutes semble interdire de considérer les propriétés comme intrinsèques des objets.

Nous formulons donc l'hypothèse suivante : un texte donné peut être utilisé avec des objectifs différents dans des conditions différentes⁸¹. Cette hypothèse semble relever du bon sens. Elle est d'ailleurs présente, dans un cadre bien plus générique, dans [Balatsoukas *et al.*, 2008] : les auteurs introduisent la notion de « spectre des objets pédagogiques » qui part de la granularité-agrégation la plus faible (données brutes) à la plus élevée (cours). D'après eux, plus un objet a une granularité-agrégation élevée, moins la ressource est réutilisable. Ceci s'expliquant par le fait qu'en avançant dans le « spectre », plus la contextualisation de la ressource est forte, moins elle est adaptable. Sans valider notre hypothèse, cette remarque tend à la conforter. Malgré son apparente évidence, elle est lourde de conséquences. Selon cette hypothèse, le cas du texte de Martin Luther King, Jr., que nous avons utilisé comme support tout au long de ce chapitre, n'est pas un cas isolé. Elle entérine l'influence des propriétés dites pédagogiques des standards étudiés les unes sur les autres rendant ainsi une description manuelle exhaustive des textes très difficile.

a) Propriétés intrinsèques et extrinsèques...

[Recker & Wiley, 2001] reconnaissent l'impact que les propriétés « pédagogiques » peuvent avoir les unes sur les autres. Ils suggèrent que pour qu'une structure de métadonnées soit efficace, elle doit intégrer aussi bien des données qui font autorité (*authoritative*), que d'autres données⁸². Les premières sont des informations « persistantes et non falsifiables », alors que les autres décrivent des aspects contextuels et non définitifs des objets pédagogiques. Cette utilisation du terme « *authoritative* » provient du fait que ces aspects contextuels regroupent selon eux les utilisations des objets pédagogiques, leur inclusion dans un contexte d'apprentissage. Contrairement aux aspects « *authoritative* », qui sont plutôt destinés à être remplis par un administrateur-descripteur⁸³, les aspects « non *authoritative* » représentent le point de vue d'un enseignant utilisateur de la ressource et par lui, celui d'une communauté d'utilisateurs. Ils rejoignent ici le point de vue de Contamines *et al.* [Contamines *et al.*, 2003] que nous avons déjà cité⁸⁴.

Ils ont cependant une approche différente, plus centrée sur les données, dont ils poursuivent la catégorisation. Ils désignent ainsi deux types de propriétés : les propriétés intrinsèques qui, comme leur nom l'indiquent, relèvent de l'objet lui-même et sont dérivables du simple fait d'avoir la ressource en sa possession, et les propriétés extrinsèques qui décrivent le contexte dans lequel la ressource est utilisée⁸⁵ [Recker & Wiley, 2001, p. 260].

81. Hypothèse H_b p. 138.

82. Qui ne font pas autorité, donc.

83. Il s'agit ici d'un rôle, qui peut être tenu par un enseignant

84. Recker et Wiley rappellent précisément le passage de l'article de Contamines *et al.* que nous avons cité en § a) p. 117 quand il font remarquer que « l'information de savoir qui a contribué est aussi importante que la contribution elle-même » *Information about who contributed is as important as the contribution itself* [Recker & Wiley, 2001, p. 262].

85. « *Intrinsic properties are those that are derivable by simply having the resource at hand [...], whereas extrinsic properties describe the context in which the resource is used.* »

b) ...selon une approche différente

Dans la suite de leur article, Recker et Wiley expliquent comment avec des informations sur les utilisateurs et leurs appréciations des diverses ressources, ils établissent pour chacun un ensemble d'utilisateurs voisins. Ils peuvent ensuite extrapoler à partir de ces derniers l'appréciation du premier d'une ressource qu'il n'a pas encore utilisé. Ce mécanisme est fréquemment utilisé, par exemple dans des applications « Web 2.0 » comme last.fm⁸⁶, qui va nous servir d'analogie. Dans last.fm, chaque utilisateur envoie au serveur la liste des chansons qu'il écoute. À partir de ses habitudes, le système va lui désigner des utilisateurs voisins et en consultant ce que ces utilisateurs voisins écoutent, il pourra dresser une liste de suggestions. Mais le serveur n'analyse pas la musique elle-même, les albums d'un artiste qui n'aurait été écouté par aucun utilisateur ne seront conseillés à personne. Le serveur n'utilise pas ces statistiques de lecture pour les confronter aux données intrinsèques des morceaux (la durée, le spectrogramme par exemple). L'analyse est purement fondée sur les données extrinsèques. C'est le même type de procédé qui est développé par Recker et Wiley. Les autres informations contextuelles (les informations non appréciatives) peuvent apporter quelques finesses à l'algorithme de la même manière que last.fm pourrait, si la question des genres musicaux n'était pas si complexe, proposer des groupes de voisins différents selon les genres musicaux abordés⁸⁷.

Nous ne discutons pas ici l'efficacité de cette approche, elle ne convient cependant pas à notre problème. Outre le fait qu'une telle approche nécessite un grand nombre d'utilisateurs pour être efficace, nous voulons pouvoir proposer à l'utilisateur les textes qui satisfont sa requête, peu importe qu'ils aient ou non été utilisés par d'autres enseignants au préalable. Encore une fois, il ne s'agit pas de remettre en question l'approche de Recker et Wiley, que nous n'excluons pas d'intégrer, à long terme, pour améliorer la fiabilité du système que ce soit uniquement par les données extrinsèques, ou comme nous le suggérons dans notre exemple musical en croisant données extrinsèques et intrinsèques.

Dans le cadre de ce travail, nous nous bornerons cependant à une approche, s'appuyant sur cette notion de données intrinsèques et extrinsèques, visant à donner un sens aux données extrinsèques du point de vue des propriétés intrinsèques des textes. Le Traitement Automatique des Langues (TAL) peut permettre d'obtenir des informations qui nous aideront à « sémantiser didactiquement » les informations intrinsèques des textes, mais avant de se demander dans quelle mesure, la première étape est de confirmer notre hypothèse d'utilisations multiples des textes, qui conditionnera ce que nous traiterons comme données intrinsèques et extrinsèques. Après quoi il nous faudra tenter d'isoler des critères d'indexation pédagogique de textes, que l'étude des standards ci-dessus n'a pas pu nous fournir du fait d'un point de vue de l'objet pédagogique trop générique pour être utilisable en l'état pour la description de ressources textuelles pour l'enseignement des langues. Ensuite, en fonction des critères isolés et de la stratégie adoptée pour les indexer, nous pourrions prendre position quant à l'implantation d'un système *ad hoc* ou l'intégration d'un standard de description de ressources pédagogiques, via un profil d'application adapté à notre domaine. Ces deux solutions ne sont pas mutuellement exclusives.

86. <http://last.fm> L'utilisation que la chaîne de télévision CBS – qui non contente d'avoir sa propre maison de disques CBS records, a racheté last.fm en 2007 – peut faire de ces informations constituerait probablement une utilisation très « intéressante » de ce type d'algorithme (cf. [Wauthy, 2008]). Ce n'est cependant pas l'objet de notre exemple.

87. Du type l'utilisateur U_1 a des habitudes d'écoute comparables à celles des utilisateurs U_2 et U_3 en ce qui concerne le Jazz, mais est bien plus proches des utilisateurs U_4 et U_5 pour le rock.

Synthèse de la partie théorique

Avant d'entrer dans le détail de notre expérimentation, nous nous proposons de retracer en quelques pages notre cheminement. Cet « interlude » *inter-chapitres* doit rappeler aux lecteurs des 122 pages qui précèdent nos principales conclusions et se veut un point d'entrée vers la suite du mémoire pour les autres.

1 Indexation pédagogique

Notre thèse a pour objectif de proposer un modèle pour ce que nous avons appelé l'*indexation pédagogique de textes pour l'enseignement des langues* et d'en montrer la faisabilité par l'implémentation d'un prototype (cf. § 1.2 p. 7). Elle est issue de la réflexion autour du projet Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO), un projet de plate-forme utilisant des outils de Traitement Automatique des Langues (TAL) pour générer automatiquement des activités d'apprentissage des langues à partir de textes tout-venants. Cependant notre travail (et la réalisation de la base de textes qui l'accompagne) s'en démarque dans sa réalisation, que ce soit du point de vue de la réflexion ou du développement, même si à terme un rapprochement est possible. En effet, le problème que nous posons touche aux pratiques des enseignants concernant la recherche de textes, indépendamment de l'usage de MIRTO. À terme, notre système pourrait être un module de MIRTO, mais ce n'est pas l'objectif de notre travail dans le cadre de cette thèse : nous nous concentrons sur 3 cas d'utilisation :

- ajout d'un texte à la base (cf. § 1.2.1 p. 8) ;
- recherche d'un texte, selon des critères relevant de la problématique de la didactique des langues, en vue de l'utiliser dans le contexte de l'enseignement des langues (cf. § 1.2.2 p. 9) ;
- aide au choix d'un texte (cf. § 1.2.3 p. 9).

Tous ces cas d'utilisation dépendent directement de la manière dont les textes seront indexés par le système.

Le choix du terme d'« indexation pédagogique » n'est pas neutre (cf. § 1.3.3 p. 15) : il prend le parti que cette tâche (choix du texte) relève plus de la relation entre l'enseignant et l'apprenant qu'entre l'enseignant et le savoir. En effet, d'après la description de la « situation pédagogique » par Rézeau [2001] que nous évoquons, une indexation didactique relèverait plus d'un usage permettant à l'enseignant d'accéder au savoir en vue de la planification du cours que du choix d'un texte pour faire acquérir ce savoir aux apprenants. Alors que dans notre cas, le problème consiste à choisir des textes pour transmettre ce savoir. Nous avons conclu que l'indexation pédagogique pouvait être définie de la manière suivante : « Indexation effectuée selon un langage documentaire permettant à l'utilisateur de rechercher des objets pour les utiliser dans un enseignement » (définition 1.4 p. 15 s'appuyant sur les notions d'indexation - définition 1.1 p. 10 - et de langage

documentaire - définition 1.3 p. 12). Dans notre cas, il s'agit donc d'indexation pédagogique *pour l'enseignement des langues*.

2 Le texte dans l'enseignement des langues

Ayant posé le problème central de notre travail, en vue de nous positionner par rapport à la didactique des langues, nous avons examiné les méthodes d'enseignement les plus employées et tenté de voir la place occupée par le texte au sein de celles-ci (cf. § 1.4 p. 18 et tableau 1.1). La conclusion de cette étude était double :

1. le texte en tant que support pédagogique est présent dans la quasi-intégralité des méthodes les plus populaires auprès des enseignants.
2. les enseignants de langue ne se cantonnent pas à une seule méthode ou approche (cf. § 1.4.5 p. 24). Ce fonctionnement est connu dans la littérature sous le terme d'éclectisme.

La conjonction de ces deux constatations rend le choix d'une méthode d'enseignement comme point d'entrée vers l'indexation pédagogique hasardeux, même si l'approche communicative semble prédominer (cf. [Levy, 1997, p. 123]). À partir du moment où les enseignants font appel à différentes méthodes, comment savoir laquelle intervient au moment du choix du texte (et ce d'autant plus que la plupart des méthodes font appel au texte écrit) ? De plus, nous n'avons que peu d'informations sur la manière dont les enseignants recherchent et choisissent leurs textes, et donc sur la manière dont la méthode d'enseignement choisie influence les stratégies mises en œuvre. Nous en avons conclu que, si nous voulions caractériser le rapport des enseignants au texte, il était nécessaire d'adopter une approche empirique.

3 Notre point de vue : celui de l'ALAO

Après avoir placé notre problématique vis-à-vis de l'informatique et de la didactique des langues, nous avons tenté de préciser le point de vue selon lequel nous abordons le problème : celui de l'Apprentissage des Langues Assisté par Ordinateur (ALAO). Pour cela, nous avons commencé par examiner le champ disciplinaire des Environnements Informatiques pour l'Apprentissage Humain (EIAH) qui traite des « travaux focalisés sur les environnements informatiques dont la finalité explicite est de susciter et d'accompagner l'apprentissage humain » [Grandbastien & Labat, 2006, p. 18]. Conformément aux remarques selon lesquelles la didactique est un savoir disciplinaire (cf. citation de Rézeau [2001] dans § *Différentes indexations pédagogiques* p. 17) et celles qui concernent le statut particulier de l'enseignement des langues parmi les autres enseignements (cf. § 2.1.2 p. 35), un véritable champ disciplinaire issu des EIAH s'est construit autour du cas particulier de l'apprentissage des langues : l'ALAO.

Pour nous positionner dans ce champ disciplinaire, nous avons eu recours à une présentation historique largement inspirée de la chronologie en trois phases de Warschauer & Healey [1998] :

1. l'ALAO *structural*, qui fait la part belle aux exercices de type *drill & practice* (cf. § 2.1.3 p. 38) ;
2. l'ALAO *communicatif*, dont les préceptes s'inspirent de l'approche communicative (cf. § 2.1.4 p. 40) ;
3. l'ALAO *intégratif*, qui s'exprime selon Warschauer et Healey par deux types d'intégration :

- le décloisonnement des aptitudes travaillées dans l'apprentissage ;
- l'intégration de la technologie au processus d'apprentissage ;
- auxquels nous ajoutons, l'hybridation des systèmes, c'est-à-dire proposer des systèmes qui combinent les supports (de l'activité ou de création de l'activité).

(cf. § 2.1.5 et 2.1.6 pp. 43 à 46).

Dans la mesure où notre prototype n'est pas destiné aux apprenants mais aux enseignants, nous avons pu constater qu'il était difficile de situer précisément notre problème au sein de l'ALAO et des EIAH. Cependant, dans le contexte d'un ALAO intégratif, notre système peut tout à fait être vu comme un éventuel module pris en charge par un Système de Gestion de l'Apprentissage (SGA).

Au cours de cette présentation, nous avons pu exposer certaines des réserves qui sont émises à l'égard de la chronologie exposée (3 phases de l'ALAO), avant de conclure que, si elles sont recevables, elles ont tendance à adopter un point de vue très strict et centré sur les systèmes, qui n'est pas celui des auteurs. Même si nous avons pu voir grâce à ces phases que le développement de l'ALAO en tant que champ disciplinaire est intimement lié aux avancées didactiques et technologiques¹, l'offre de systèmes d'ALAO ne suit pas toujours parfaitement la théorie. Après une époque où le dilettantisme était la norme, un processus de professionnalisation s'est mis en place, les enjeux économiques de la formation ont pris le dessus et la majeure partie des systèmes d'ALAO est le fait d'entreprises, qui ne sont pas toujours en phase avec les développements théoriques du moment.

4 Travers de l'ALAO et influence sur notre démarche

Cette distance entre la majeure partie des productions et les avancées théoriques est probablement l'une des raisons qui font que pour Murray et Barnes, « *Multimedia language learning materials comprise, as with all learning resources, some gems amongst a lot of mediocrity.*² » [Murray & Barnes, 1998, p. 259]. En effet, comme l'évoque l'intitulé du champ, la conception d'un système d'ALAO s'articule autour de deux grands pôles de compétences : didactiques et techniques. Les spécialistes semblent s'accorder pour dire qu'un dialogue doit se systématiser entre ces deux pôles. Les tenants des compétences didactiques doivent connaître les possibilités offertes par la technologie pour concevoir des applications de qualité. À l'inverse, le point de départ ne doit jamais être la technologie, mais des principes didactiques [Bangs, 2002, p. 24] (cf. note⁴⁸ p. 49). En réaction à ce type de travers, nous nous inspirons du « design incrémentiel et itératif » de Depover *et al.* [2000]. À chaque étape, des échanges avec le client sont suggérés (aspect itératif) selon des modalités qui dépendent des objectifs de l'étape et de l'état de développement du produit (aspect incrémentiel). Comme notre projet vise à développer un prototype, nous n'effectuerons qu'une itération, cependant nous devons prendre en compte les pratiques des enseignants pour ne pas adopter un point de vue strictement technique. Nous revenons ainsi à la démarche empirique déjà proposée en réaction à l'étude des différentes méthodes d'enseignement.

Au delà du processus de conception lui-même, le point de vue techniciste qui a présidé à la création de nombreux programmes ignore, du fait même de cette non prise en compte du point de vue des enseignants, certains outils. C'est là l'un des constats ayant mené à la création du

1. en toute cohérence avec le champ des EIAH.

2. Le matériel multimedia d'apprentissage des langues, comprend comme c'est le cas pour toute ressource d'apprentissage, quelques bijoux parmi un flot de médiocrité. (cf. note⁶⁶ p. 59)

projet MIRTO : la manière de considérer la langue dans les applications informatiques (cf. § 2.2.3 p. 53). L'absence de sens associé à la forme dans le cadre de toute application n'ayant recours à aucune forme de TAL, nuit à son adéquation avec les besoins des enseignants [Antoniadis *et al.*, 2005b]. Les insuffisances sont particulièrement criantes dès qu'il s'agit d'évaluation et de feedback, mais elles ne s'y cantonnent pas. Le fait de ne considérer que des « suites de caractères » et non des propriétés ou des concepts est aussi en partie responsable de la rigidité des logiciels en question.

Bien évidemment, le fait d'utiliser du TAL dans un système d'ALAO ne rend pas ce dernier évolutif et conforme aux besoins des enseignants. En tant qu'outil, pour être intégré avec succès, le TAL doit être compris, ce qui signifie connaître ses possibilités et appréhender ses limites (cf. § a) p. 59), spécialement dans le contexte de l'apprentissage. Nous avons donc décidé de nous inspirer de la stratégie moins-disante décrite par Ponton et Kraif (cf. § a) p. 11 & § a) p. 59).

Les questions posées par la manière dont le TAL a pu susciter des attentes allant au delà de ses possibilités réelles comportent des similarités avec la manière dont les enseignants sont amenés à appréhender les systèmes d'ALAO (cf. § *Fear, awe and 'wow'* p. 59). Cette constatation nous a mené à considérer ces différents outils à travers le même prisme : l'« approche instrumentale » (cf. § *Approche instrumentale* p. 61). Selon cette approche, il ne faut pas uniquement aborder les outils en fonction de la manière dont les concepteurs les considèrent mais aussi prendre en compte la problématique des usages. De notre point de vue (cf. § 2.2.5 p. 63), ces considérations valident le recours à la stratégie moins-disante, la nécessité d'une démarche empirique et la consultation des futurs utilisateurs du système, les enseignants. Elles permettent de mettre en perspective notre travail qui, de par notre position de concepteur, ne concerne que des usages prescrits qui seront à confronter avec les usages réels. Enfin, elles s'avèrent être un outil intéressant pour l'étude de l'existant.

5 Étude de l'existant

Jusqu'ici nous avons positionné notre problématique vis-à-vis de champs disciplinaires (didactique des langues, systèmes d'information, EIAH, ALAO), ce qui nous a permis de tirer certaines conclusions méthodologiques. Nous nous intéressons ici de plus près aux problématiques qui jouxtent la nôtre :

5.1 Corpus

En tout premier lieu, l'étude de l'usage de corpus dans l'enseignement des langues nous a mené à nous poser la question de l'authenticité. Il est apparu qu'il y avait divers points de vue à considérer. Lorsque nous employons le terme « texte authentique » nous nous placerons désormais dans le cadre de l'authenticité de la langue (cf. § 2.3.1 p. 64).

Bien qu'ils relèvent surtout de la linguistique, les corpus se rapprochent de notre problématique en ceci qu'ils réunissent des textes et peuvent faciliter l'accès à du matériel authentique. Cependant une étude de la notion de corpus (cf. définition 2.3 p. 68), nous en distancie, pour constater que notre travail traite de collection de textes (cf. § *Collection de textes* p. 69). Nous avons malgré tout décidé d'examiner les usages des corpus dans l'enseignement des langues qui, au moins d'un point de vue théorique se multiplient, afin de voir dans quelle mesure nous pouvons nous en inspirer.

Les usages les plus courants de corpus dans l'enseignement des langues en sont, selon la terminologie de l'approche instrumentale, des instrumentations (cf. § 2.3.4 p. 70) : alors que les corpus et les outils d'interrogation (principalement des concordanciers, dans ce cas) ne sont pas destinés à l'enseignement, c'est par l'usage qu'ils sont adaptés à des problématiques didactiques. Ces usages relèvent en grande partie de ce que Johns a appelé le *Data Driven Learning* (DDL) [Johns, 2002, p. 107]. Ils sont variés mais visent en particulier à placer les apprenants en position d'acquérir des notions à la croisée entre lexique et grammaire de manière inductive. Ils ne relèvent pas réellement de notre problématique, car ils ont tendance à considérer les ressources textuelles comme un ensemble de données langagières, *du* texte, et non *des* textes à appréhender dans leur ensemble.

Nous nous sommes ensuite intéressé aux instrumentalisations des corpus (cf. § 2.3.5 p. 75). Toujours selon le concept de genèse instrumentale (cf. § *Approche instrumentale* p. 61), l'instrumentalisation correspond au fait de modifier les outils. Le fait de créer des corpus pédagogiques, pourrait être attribué à Braun qui constate que malgré la popularité des corpus chez les didacticiens, les usages en classe sont très rares [Braun, 2005]. Les instrumentalisations peuvent concerner les corpus, qui doivent être adaptés aux apprenants : corpus plus petits, contenu plus homogène, annotation plus adaptée. Cependant, toujours d'après Braun, la notion de corpus pédagogique exige aussi un travail sur les outils d'interrogation, notamment à travers la possibilité de proposer des activités. La plupart des recherches dans le domaine sont très récentes, particulièrement la partie qui nous concerne le plus, à savoir quoi annoter, quoi rechercher dans les corpus (ultérieures à 2007). Parmi les travaux les plus pertinents pour notre problématique, nous pouvons citer ceux de Sharoff *et al.* [2008]. Nous avons néanmoins choisi un chemin différent : plutôt que de partir des outils, nous tentons de partir des utilisateurs-enseignants, ce qui, dans ce cas précis, est probablement une solution moins rapide.

À ce moment de notre réflexion, nous avons pu nous positionner par rapport aux champs disciplinaires des EIAH et de l'ALAO. Nous avons pu en tirer des conclusions méthodologiques. Nous avons pu nous interroger sur les notions de corpus et de texte authentique, précisant ainsi notre problématique. Toutefois, les travaux sur les corpus pédagogiques ne nous ont pas fourni de raccourci vers la notion d'indexation pédagogique. Nous nous sommes donc tourné vers les nombreux travaux concernant la description de ressources pédagogiques (cf. chapitre 3).

5.2 Standards de description de ressources pédagogiques

Avant d'entrer dans le détail des principaux standards de description de ressources pédagogiques, nous avons commencé par recenser les objectifs de la normalisation et les enjeux du processus lui-même. Parmi les multiples objectifs revendiqués (accessibilité, interopérabilité, réutilisabilité, durabilité et adaptabilité, cf. § 3.1.1 p. 88), celui qui nous préoccupe avant tout est l'accessibilité, puisque c'est précisément la fonction de l'indexation pédagogique.

Un examen rapide des stratégies de Microsoft concernant la normalisation (cf. § *Enjeux de la normalisation : quelques exemples* p. 90) montre que :

- le processus n'est pas en soi une garantie d'atteinte des objectifs susnommés ;
- la norme est indissociable de son utilisation et de son appropriation par les acteurs du domaine.

Et si dans le domaine des technologies de l'apprentissage, on ne peut comparer la situation d'aucune structure à celle de Microsoft, les enjeux financiers ainsi que les intérêts et objectifs des différents acteurs peuvent diverger (cf. § *Conséquences* p. 92). Notre conclusion est donc que

si nous pouvons permettre une meilleure interopérabilité entre notre système et d'autres grâce à des normes, nous y aurons recours. Mais pour cela, encore faut-il qu'elles soient aptes à traiter notre problème. Il n'est pas question d'adapter notre problématique pour rentrer dans le cadre d'une norme.

Après avoir passé en revue les acteurs, les principaux standards et les liens qu'ils entretiennent (cf. figure 3.2), nous nous sommes focalisés sur trois d'entre eux (*Learning Object Metadata* (LOM), *Sharable Content Object Reference Model* (SCORM), *OUN-Educational Modelling Language* (EML) / *IMS-Learning Design* (IMS-LD), cf. § 3.3 p. 98) pour constater, à l'instar de [Pernin, 2003], qu'ils avaient chacun un point de vue différent sur les ressources qu'ils sont amenés à décrire (cf. § 3.3.4 p. 104) :

- modélisation pédagogique (EML) ;
- mise en œuvre informatique (SCORM) ;
- indexation (LOM).

Notre point de vue est donc celui du LOM et des autres standards du même type. Ils s'articulent autour de la notion de « métadonnées » : des « informations sur des ressources [...], compréhensibles par les êtres humains et traitables par des logiciels » [Bourda, 2002, 116–117]. Les métadonnées s'articulent en général autour d'un schéma, qui peut parfois être adapté à une problématique donnée par le biais d'un profil d'application.

Après avoir statué sur le fait que les textes de notre base étaient des objets pédagogiques selon la définition du LOM (cf. § 3.3.1 p. 98), nous avons étudié les concepts centraux du LOM et passé en revue ses éléments les plus pertinents pour notre problématique (cf. § 3.4 p. 105). Il en découle une série de constatations (cf. § 3.5.1 p. 117). Les spécifications LOM sont parfois floues : le point de vue à adopter lors de l'annotation n'est pas toujours clair, les valeurs à utiliser pour décrire les éléments peuvent être discutées dans leur formulation (« niveau d'interactivité très bas » [Pernin, 2003, p. 188]) ou, de l'aveu des auteurs du standard, ne prennent leur sens qu'au sein d'une communauté d'utilisateurs [IEEE, 2002, p. 25–26]³. Toutefois, le problème majeur pour notre travail est, à notre avis, l'un de ceux soulignés par Pernin :

« la volonté d'intégrer au sein d'un même modèle des entités de niveau conceptuellement très différent : les ressources nécessaires à la mise en place d'activités pédagogiques et les activités elles-mêmes. »

[Pernin, 2004, p. 2]

Les autres problèmes signalés pourraient être résolus par la définition d'un profil d'application (cf. § 3.5.3 p. 120). Cependant, les standards les plus répandus – il ne s'agit pas seulement du LOM, mais aussi d'*Educational Network of Australia* (edna), du *the Getaway to Educational Material* (GEM) et par transitivité de la *Dublin Core Metadata Initiative* (DCMI) – ne semblent pas aptes à décrire pédagogiquement les ressources brutes. Ce point de vue a été corroboré par Balatsoukas *et al.* [2008] pour qui, plus le niveau d'agrégation des ressources est bas, plus ce qu'ils appellent le « spectre de l'objet⁴ », est large. Ce simple constat rend une description exhaustive des caractéristiques pédagogiques plus complexe, pour ne pas dire impossible.

Toutefois, cet article n'a été écrit que bien après que notre travail sur les standards de description de ressources pédagogiques ait été terminé. À ce moment de nos recherches, nous avons formulé une hypothèse qui pourrait être considérée comme un cas particulier de la conclusion de Balatsoukas *et al.*, selon laquelle un texte donné pouvait être utilisé dans plusieurs contextes

3. Il nous semble que cet avertissement aurait pu être inscrit pour un plus grand nombre d'éléments.

4. La notion de « spectre de l'objet » chez Balatsoukas *et al.* [2008] pourrait s'apparenter à son champ instrumental dans l'approche instrumentale (cf. § c) p. 61).

radicalement différents. Validée cette hypothèse risquait de rendre l'usage de ces standards impossible, nous poussant à nous poser la question de la nature des propriétés concernées. Dans cette optique, les travaux de Recker & Wiley [2001] sont primordiaux. Ils permettent de faire la différence entre propriétés intrinsèques (« persistantes et non falsifiables ») et extrinsèques (décrivant des aspects contextuels et non définitifs). Cependant, leur solution fondée uniquement sur les usages ne correspond pas parfaitement à nos objectifs qui sont de pouvoir proposer à un enseignant un texte qui répond à sa requête même si aucun enseignant ne l'a décrit comme tel auparavant. Notre objectif pour la suite de cette thèse était donc de vérifier l'hypothèse majeure formulée ci-dessus et de tenter d'analyser et modéliser le rapport qui peut exister entre les propriétés pédagogiques intrinsèques des textes et leurs propriétés extrinsèques.

Deuxième partie

Recueil et analyse des données

Préambule : Choix d'un mode de recueil de données

Le chapitre précédent nous a mené à nous interroger sur la possibilité d'utiliser un standard de description de ressources pédagogiques dans le cadre de l'indexation pédagogique de textes pour l'enseignement des langues. En effet, le texte, à l'instar d'autres ressources brutes, semble difficile à cataloguer pédagogiquement en utilisant les standards existants du fait de champs instrumentaux trop larges. Si, conformément à l'hypothèse que nous avons formulée, les textes peuvent être utilisés dans des contextes différents, alors, considérer les propriétés pédagogiques comme intrinsèques peut poser problème : une description exhaustive paraît dès lors hors d'atteinte. Dans ce cas, la seule solution utilisant les standards « pédagogiques » reviendrait à une annotation, par les enseignants, d'usages avérés des textes, ce qui nous éloigne de la notion d'indexation pédagogique telle que nous l'avons définie. En effet, notre définition (cf. définition 1.4 p. 15) s'appuie sur la notion de langage documentaire qui doit permettre de décrire les éléments suffisamment précisément (cf. définition 1.3 p. 12) pour « permettre le repérage simple des documents du corpus qui répondent aux questions de ces usagers » [Lefèvre, 2000]. Un fonctionnement fondé uniquement sur les utilisations avérées des textes interdirait la formulation de questions du type : « un texte qui *pourrait* être utilisé dans le contexte... ». Les seules questions possibles seraient « un texte qui *a déjà été utilisé par un enseignant* dans le contexte... ».

L'objectif central de ce chapitre est donc de tester notre hypothèse à travers une étude des pratiques déclarées des enseignants (cf. § 1.4.5 p. 24, § 2.2.5 p. 63 et § 2.4 p. 84), et de tirer les conséquences de sa validation (proposition d'une stratégie) ou invalidation (utilisation d'un standard ou définition d'un profil d'application).

Pour y parvenir, nous articulerons notre réflexion autour de deux questionnaires que nous traiterons séparément. Nous procéderons chronologiquement en commençant par le premier questionnaire. Nous montrerons ensuite comment l'analyse des résultats de celui-ci nous a mené au second, avant de conclure sur cette partie empirique et les concepts qui en découlent.

D'un point de vue chronologique, nous avons commencé à réfléchir à notre premier questionnaire en parallèle de notre travail sur les standards pédagogiques. Il s'alimente de certaines questions issues des entretiens que nous avons menés dans une étude préliminaire exploratoire [Loiseau, 2003].

1 Étude préliminaire

Cette étude préliminaire s'articule autour de 8 entretiens semi-dirigés avec des enseignants de langue de profils différents :

- langues : anglais, espagnol, polonais ;
- expérience de l'enseignement : entre 1 et 30 ans ;
- cadre de l'enseignement : collège, Université et centre de formation continue ;
- public : débutant à Bac+5 (spécialistes) ;
- littératie informatique (cf. NB 3.1 p. 134) : utilisation de l'outil informatique à titre personnel, jusqu'à la création d'activités pédagogiques avec support informatique.

L'hétérogénéité de la population visait à recueillir les pratiques déclarées les plus diverses possibles, même si le nombre peu important d'enquêtés a interdit toute généralisation. Cette étude nous a permis d'effectuer certaines constatations, qui soulèvent autant de questions :

- Certains enseignants – dans le cadre de cette étude, surtout des enseignants d'anglais – lisent

Nota Bene 3.1 Littératie informatique

Dans le cadre d'études ayant trait aux pratiques d'utilisateurs de l'outil informatique, il est souvent question du degré de connaissance des utilisateurs concernant ces « nouveaux outils ». Nous parlerons de cette maîtrise de l'outil informatique en utilisant le terme *littératie informatique*, qui est l'une des traductions proposées pour le terme anglais de *computer literacy*. L'expression *alphabétisation informatique* est employée dans plusieurs articles : [Larose *et al.*, 2002] ou [Lusalusa *et al.*, 2000]. Nous lui préférons toutefois l'expression *littératie informatique* trouvée chez Reffay & Chanier [2005]. Puisque, comme le souligne Mangenot [2003], la notion de *literacy* est traduite par différents termes comme « alphabétisation, littératie ou littéracie », qui sont utilisés dans différents contextes, il nous a semblé plus pertinent d'en utiliser un qui ne soit pas couramment employé dans d'autres contextes.

des textes écrits dans la langue enseignée pour leur propre intérêt, et peuvent décider de les retenir pour leur cours alors que ce n'était pas là l'objectif de la lecture. Nous avons appelé cette pratique « recherche passive » [Loiseau, 2003, p. 46]⁵ ;

- certains enseignants disposent de recueils de textes, qui sont parfois organisés, constituant ainsi un point d'entrée vers les caractéristiques qu'ils considèrent pertinentes dans leurs choix de textes [Loiseau, 2003, p. 47] ;
- certains modifient parfois les textes (cf. pratiques auxquelles faisait référence Vigner [2001], cf. p. 67), ce qui pose la question des usages des textes authentiques par rapport aux textes fabriqués ;
- le processus de recherche ne repose pas uniquement sur des critères tangibles et précis, certains témoignages faisant explicitement référence à l'« intuition » [Loiseau, 2003, p. 58] ;
- il semble que certains critères en influencent d'autres (la quantité de structures inconnues ne faisant pas l'objet de la leçon semblant influencée par des critères comme la L1 des apprenants, leur niveau ou le type d'activité) [Loiseau, 2003, p. 59].

2 Pourquoi des questionnaires

Les objectifs de cette étude étaient multiples. Outre le fait d'établir un lien avec une communauté d'utilisateurs potentiels, elle devait nous permettre d'entamer une réflexion sur les pratiques déclarées des enseignants concernant la recherche de textes, mais aussi sur la manière dont ils considèrent le texte. Suite à cette enquête exploratoire, nous avons souhaité confronter certaines de nos conclusions à un plus grand nombre d'enquêtés pour apporter une teneur quantitative à nos remarques [Ghiglione & Matalon, 1978, p. 93]. Le schéma traditionnel dans cette situation est de s'orienter vers un questionnaire. Si nous avons emprunté ce chemin, c'est aussi parce que les alternatives, à savoir l'observation et l'expérimentation [Angers, 2000, pp. 74–79], sont peu réalistes avec nos objectifs de recherche.

a) Observation en situation

L'une des objections à une observation directe du processus de recherche/choix de texte, pourrait être qu'il s'agit d'une activité difficile à délimiter, un processus de longue haleine, qui va se prolonger jusque dans les loisirs de certains enseignants. C'est ce que nous avons pu constater

5. C'est le terme employé dans le mémoire, mais il est possible que nous ne l'ayons pas dénommée ainsi nous-même. Le terme de « recherche passive » est peut être issu des témoignages. Il n'est peut être pas le plus adapté, mais nous le conservons ici dans un souci de cohérence.

au cours de nos entretiens avec la notion de « recherche passive ». Cette absence de délimitation entre le temps de travail et le temps personnel rend une observation en situation peu appropriée, en particulier pour un traitement quantitatif.

b) Auto-observation

Dans ce type d'observation indirecte, il s'agit de demander à l'enquêté de jouer le rôle d'observateur, en lui signalant les éléments à prendre en compte. Cette méthode peut par exemple être utilisée en ergonomie, comme dans l'expérience détaillée par Patesson [2007] (« L'auto-observation de l'activité au téléphone »). Il nous a cependant semblé que notre problème était plus difficile à traiter de la sorte, surtout en raison des contraintes imposées aux enquêtés. Pour avoir des éléments de réponse à nos questions, il aurait fallu que chaque enseignant participant s'engage à répondre à une série de questions à chaque fois qu'il recherchait ou trouvait un texte (nous avons déjà signalé qu'un texte pouvait être trouvé sans être cherché). L'enseignant aurait dû, pour chaque recherche, spécifier ce qu'il cherchait, chacun des textes non retenus, la raison pour laquelle ils n'étaient pas retenus, ainsi que les textes retenus (avec justification). Un tel corpus de données serait une avancée décisive pour notre problématique, mais cela représente un travail difficile à attendre d'un grand nombre de sujets. En considérant les difficultés rencontrées pour réunir nos questionnaires, le fait de ne pas avoir choisi un protocole aussi fastidieux s'est avéré être une bonne décision⁶.

Enfin, il ne faut pas perdre de vue que l'une de nos principales hypothèses quant à l'utilisation ou non d'un standard de description de ressources pédagogiques concernait la possibilité d'utiliser un texte donné dans plusieurs contextes différents. Pour prouver une telle hypothèse avec ce protocole, il aurait fallu retrouver plusieurs fois le même texte parmi ceux retenus par les enseignants dans des contextes différents, ce qui est peu probable.

c) Expérimentation

Monter une expérimentation à ce stade de notre travail, avant la proposition d'un modèle ou la réalisation d'un prototype, surtout prospectif, nous a semblé prématuré. Nous avons beaucoup trop de variables potentielles que nous ne connaissons que peu en terme d'importance ou d'effet les unes sur les autres. Un travail de catégorisation des propriétés doit être effectué pour pouvoir choisir des textes suffisamment représentatifs de chacune d'entre elles et *a fortiori* constater leur effet auprès d'enseignants. Mais, nous disposons de peu d'information sur la manière dont ces derniers considèrent le problème.

3 Conséquences sur les pratiques

Nous avons donc décidé de nous orienter vers un questionnaire, ce qui a des conséquences directes sur l'objet de notre recueil de données. Si nous pouvions parler de « pratiques » quand nous adressions le problème dans le cas général, à partir du moment où nous parlons de nos données, il nous faut adopter une précision supérieure. On oppose en général « pratiques déclarées » et « pratiques effectives ». Comme leur nom l'indique les premières se réfèrent aux pratiques telles que les enseignants eux-mêmes en rendent compte. Les secondes résultent d'une procédure

6. cf. § 4.2.2 p. 149.

d'observation. Au terme de pratiques « effectives », nous préférons, à l'instar de Clanet [2007], celui de pratiques « constatées » qui rend mieux compte de la différence de point de vue entre les deux, sans donner une teinte de « vérité » supérieure aux pratiques effectives :

« Écartons tout de suite la prétention d'un accès aux "vraies" pratiques. Quelle que soit la qualité de la méthodologie qu'utilise le chercheur, son accès aux pratiques est dépendant de cette méthodologie et en possède les limites. Les pratiques observées ne sont jamais que les pratiques constatées sous les conditions de l'observation. Parler de pratiques effectives ne peut être un raccourci acceptable que dans la mesure où il ne conforte pas l'illusion d'une connaissance directe et totale des pratiques. »

[Bru, 2004, p. 283]

En effet, nous pouvons transposer ici le « paradoxe de l'observateur » de Labov, à l'observation de pratiques enseignantes :

« Nous en arrivons ainsi au *paradoxe de l'observateur* : le but de la recherche linguistique au sein de la communauté est de découvrir comment les gens parlent quand on ne les observe pas systématiquement ; mais la seule façon d'y parvenir est de les observer systématiquement. »

[Labov, 1976, p. 289–290]

La situation est identique dans la mesure où l'objectif est d'accéder aux pratiques habituelles des enseignants et, pour y parvenir, l'observation donne une tournure tout à fait exceptionnelle au cours, par la mise en place même d'un protocole d'observation.

Nous ne contestons aucunement ici l'intérêt d'avoir accès aux pratiques constatées des enseignants, cependant nous tenons à faire valoir que les pratiques déclarées sur lesquelles nous allons travailler n'en constituent pas un parent pauvre. À long terme, les informations que pourraient nous apporter l'observation nous permettraient très probablement d'avoir une vision plus complète et précise du problème. Cependant, dans une perspective compréhensive et explicative, des pratiques déclarées semblent adéquates [Bru, 2002, p. 70]. Ramsden [1992] propose un modèle de l'enseignement qui appuierait ce point de vue. Ce modèle concerne l'enseignement supérieur et n'a pas de prétention universaliste. Pour lui, l'action d'enseignement – les pratiques⁷ – est issue de théories éducatives vues par le prisme du contexte dans lequel elle a lieu et d'une réflexion sur l'expérience de l'enseignant [Ramsden, 1992, p. 118–119], c'est à dire la perception qu'il a de ses propres pratiques.

7. Il ne s'agit pas ici de pratiques déclarées ou constatées, puisqu'il n'est pas question ici d'observation ou même de description.

Chapitre 4

Mise en évidence du contexte pédagogique

À ce stade de nos investigations, il nous a paru pertinent d'avoir recours à l'expertise des enseignants en leur demandant, sous forme de questionnaire, une réflexion sur leurs pratiques déclarées¹. Avant de détailler davantage notre protocole, nous allons commencer par recenser nos hypothèses.

4.1 Hypothèses et Objectifs

Avec ce questionnaire, nous nous intéressons aux pratiques des enseignants quant à la recherche et au choix de textes. D'un point de vue théorique, cette tâche s'inscrit dans la planification qui se décompose en 3 phases : « Jackson [1968] [...] distingue la phase préactive ou postactive de la phase interactive : l'enseignant planifie ses cours avant la leçon (préaction) mais aussi entre deux leçons en fonction des leçons qui précèdent (régulation postactive) ; l'interaction en classe comprend elle-même des opérations cognitives de planification du moment qui suit en fonction de ce qui vient de se passer² » [Tochon, 1989, p. 23]. Pour être plus précis, la recherche et le choix de textes est un sous-problème de la phase préactive de la planification. À ce titre, il s'inscrit dans un processus plus large. Plusieurs modèles adoptant des points de vue différents visant à décrire la planification coexistent. Par exemple, Yinger [1978] voit la planification comme un processus en trois phases : définition du problème, formulation/solution du problème et implémentation/évaluation/routinisation [Yinger, 1978, p. 28]. Que ce soit dans la phase de définition ou dans celle de formulation/solution du problème, les supports pédagogiques font partie des variables qui interagissent et influencent les prises de décisions [Yinger, 1978, p. 22, 30, 38].

Notre point de vue concerne donc un point très précis de la planification, interagissant avec et influencé par d'autres aspects du processus. La prise en compte des décisions effectuées en amont de la recherche ou du choix d'un texte se fera par le biais de variables qui n'auront pas vocation à décrire le processus par lequel elles ont été définies mais la situation dans laquelle aura lieu la tâche qui nous intéresse. Le processus qui nous intéresse est donc la recherche et le choix de

1. cf. § 3 p. 135.

2. Ces trois phases sont largement reprises, par exemple par Yinger [1978, p. 1], Dargirolle [1999, pp. 143–144] ou Charlier & Charlier [1998, p. 16].

textes et si nous ne tentons pas de décrire toute la planification, nous adopterons un point de vue générique, qui va s'articuler autour d'une série d'hypothèses.

4.1.1 Recherche *vs* processus de recherche

Nos hypothèses de travail sont les suivantes :

- H_a**. l'une des tâches des enseignants de langue est la recherche de textes ;
- H_b**. un texte donné peut être utilisé avec des objectifs différents dans des conditions différentes (cf. § 3.5.3 p. 120) ;
- H_c**. la recherche de textes se fait en premier lieu dans l'optique d'une activité concrète ;
- H_d**. les enseignants pratiquent couramment l'éclectisme (cf. § 1.4.5 p. 24) ;
- H_e**. la langue enseignée (L2) n'a pas d'effet sur le processus général de recherche de textes ;
- H_f**. le pays dans lequel l'enseignant exerce n'a pas d'effet sur le processus général de recherche de textes ;
- H_g**. le public n'a pas d'effet sur le processus général de recherche de textes ;
- H_h**. l'expérience de l'enseignant n'a pas d'effet sur le processus général de recherche de textes ;
- H_i**. le degré de littératie informatique (cf. NB 3.1 p. 134) des enseignants n'a pas d'effet sur le processus général de recherche de textes.

Avant d'entrer dans le détail des autres hypothèses, nous rappelons que H_b et H_d ont déjà été discutées, la seconde étant même un résultat théorique que nous voulons comparer à nos données. Les hypothèses H_b et H_c sont intimement liées. En effet, le poids de H_b dépend grandement du fait que H_c soit vérifiée : plus l'activité finale est importante dans le processus de recherche de textes, plus H_b aura de lourdes conséquences sur l'indexation pédagogique de textes pour l'enseignement des langues.

L'hypothèse H_a a pour but d'entériner la remarque que nous faisons dans notre partie introductive (cf. § 1.1.2 p. 6) selon laquelle les enseignants de langue étaient amenés à rechercher des textes pour leur cours. Généraliser cette remarque attesterait de la légitimité de notre problématique non seulement d'un point de vue de recherche de connaissance (pour lequel ce n'est pas strictement nécessaire), mais aussi d'un point de vue utilitariste, cher à certains décideurs et autres tenants de l'innovation [MEDEF & CCIP, 2004].

Ensuite, il faut bien comprendre ce que nous entendons par « processus général de recherche », sans quoi les hypothèses H_e à H_i paraissent totalement farfelues, en particulier à la lumière des travaux que nous avons pu citer précédemment, mais aussi aux yeux de la logique même. Nous ne voulons pas dire que les facteurs énoncés sont dépourvus d'influence sur la recherche de textes. En effet, comprise dans ce sens, l'hypothèse H_e, pour ne citer que celle-ci, est tout bonnement absurde. Bien évidemment, la langue enseignée a au moins pour effet sur la recherche que l'enseignant prospectera vraisemblablement, dans l'écrasante majorité des cas³, des textes écrits dans la langue enseignée. Cependant, il n'est pas dit que le processus de recherche de textes soit radicalement différent pour un enseignant d'anglais et pour un enseignant d'espagnol. Dans le cadre de nos hypothèses, c'est-à-dire, en règle générale, pour les questions qui donneront lieu

3. Mais gardons nous bien de généraliser, par exemple, ce n'est pas nécessairement le cas pour un exercice de thème.

à des analyses quantitatives, c'est avant tout le processus, dans son acception la plus générique, qui nous intéresse. Malgré tout, ce n'est pas le seul point de vue que nous adoptons, certaines questions reflètent une démarche plus prospective dans l'optique de dégager certains critères ou certaines familles de critères intervenant dans les recherches.

L'hypothèse H_i précise un peu plus la différence que nous voulons faire entre la recherche elle-même et le processus général. Pour l'illustrer, nous nous appuyons sur un témoignage issu de notre étude préliminaire : « "Même dans *El País*, [...] ou dans la presse qui vient d'Amérique latine, le dimanche, il y a des histoires [...] qui sont [très] littéraires" » [Loiseau, 2003, p. 58]. Ce témoignage indique une stratégie utilisée par un enseignant pour trouver des nouvelles littéraires. Cette stratégie, dont l'objectif est très précis, consiste en le choix de journaux qui sont pourtant, à première vue, associés à d'autres genre textuels. Dans l'absolu, il est intéressant d'avoir ce genre d'information, cependant, ce n'est pas l'objet des hypothèses que nous avons formulées. Sinon, l'hypothèse H_i n'aurait pas de sens. À partir du moment où, pour un type de recherche donné, certains enseignants ont recours à Internet, le degré de littératie informatique a nécessairement un effet, puisque ceux qui ne savent pas utiliser un navigateur passeront complètement outre ce type de pratiques. Nous avons donc pris soin de parler de processus général, pour montrer que nous désirons prendre de la distance vis-à-vis du processus. En considérant la stratégie sus-mentionnée à un niveau *méta*, en la reformulant de manière plus générique – « choisir la source en fonction de ce qu'on veut y trouver » –, nous nous approchons du problème qui nous intéresse. Avec un tel grain, il n'est plus fondamentalement incohérent de faire l'hypothèse H_i (dans notre exemple, Internet devient une source de données parmi d'autres), mais surtout nous nous rapprochons d'une réponse concernant la problématique générale de la recherche de textes qui, dans le cas de notre exemple, va dans le sens de notre hypothèse H_c à savoir une recherche de textes dans l'optique d'une activité concrète.

Avant de faire le tour des questions posées en spécifiant lesquelles concernent le processus général et lesquelles sont d'ordre prospectif, nous allons faire un détour par la forme prise par nos questions.

4.1.2 Formats

Selon les suggestions de Ghiglione et Matalon, nous commençons notre questionnaire par un certain nombre de questions ouvertes (6) afin de ne pas donner l'impression que nous n'attendons « que des réponses brèves, sans détails et [que nous ne nous intéressons] qu'à ce qui entre dans des cadres pré-établis » [Ghiglione & Matalon, 1978, p. 99]. Nous essayons d'en savoir plus sur les enseignants, sans entrer dans des questions d'ordre trop privé (cf. § 4.1.3 p. 140). Le questionnaire, outre une section « remarques/suivi », comportait 20 questions : 8 questions ouvertes, 8 questions fermées et 4 questions que nous qualifierons de semi-fermées sur lesquelles nous allons revenir.

Selon de Singly [1992], les stratégies d'usages des questions ouvertes et fermées sont liées à deux approches, l'une ayant vocation à tester des hypothèses précises (caractérisation du processus général), et l'autre « plus attentive à la complexité du réel » (dimension prospective). Pour améliorer les questions fermées, outre la conduite d'entretiens préalables, il suggère de :

« Prévoir, dans la liste des réponses, une catégorie "autres" avec de la place pour ces réponses libres "imprévues". Lorsque, au moment du dépouillement, il y a beaucoup de réponses "autres", sont créées de nouvelles catégories. Mais cette technique n'est pas idéale, car rien ne garantit que la liberté de proposer des réponses "autres" puisse être appropriée également pour tous les individus. Cette solution est surtout intéressante

lorsque les questionnaires sont remplis par les personnes interrogées elles-mêmes. »
[de Singly, 1992, p. 69]

Il y a, dans notre travail, une forte dimension prospective et si, dans certains cas, il nous faut éviter l'ambiguïté autant que possible (cf. § 4.1.7 p. 145), nous devons également tenter d'obtenir la vision la plus complète possible. L'usage de questions « semi-fermées » s'est ainsi imposé de lui-même ; il est cependant confirmé par la théorie (cf. de Singly [1992]). Il est aussi en accord avec le mode de passation choisi selon lequel les enquêtés remplissent eux-mêmes les questionnaires (cf. § 4.2.2 p. 149). Nous développons plus loin la question de notre échantillon (cf. § 4.1.3 p. 140 et § a) p. 150), cependant nous pouvons indiquer ici que le niveau d'étude des enquêtés – des enseignants de langue – est élevé, et d'après Schuman et Presser, plus le niveau d'éducation des enquêtés est élevé, moins les résultats sont tributaires du type de question choisi (ouvertes ou fermées) [Schuman & Presser, 1979, pp. 710–711]. Bien évidemment, pour que cette remarque ait un sens, il faut que le travail de préparation du questionnaire (par exemple des entretiens préliminaires) ait été fait avec sérieux, puisque si les choix proposés dans les questions fermées sont inadaptés, les résultats obtenus risquent de ne pas être équivalents à ceux issus de questions ouvertes.

Prise dans l'autre sens et en ce qui concerne notre population, cette remarque tend à valider le fait de laisser la possibilité à l'enquêté d'aller au delà des réponses que nous proposons. Cependant, dans certains cas, il nous a semblé que les choix ajoutés risquaient d'être marginaux par rapport à ceux que nous proposons, c'est l'une des raisons qui ont fait que nous demandions aux enquêtés d'ordonner leurs choix (voir § e) p. 142 et § 4.1.6 p. 144).

Ces différentes formes de question étaient rendues disponibles aux enquêtés à travers plusieurs supports⁴ :

- « hors-ligne » : version électronique et papier à nous renvoyer ;
- « en-ligne » : formulaire électronique *PHP: Hypertext Preprocessor* (PHP)/MySQL.

Cet état de fait est lié à la population de l'étude, comme nous commençons à l'expliquer dans la section suivante.

4.1.3 Population

La partie théorique de notre travail (la réflexion concernant la notion d'indexation pédagogique de textes dans l'optique de leur utilisation dans l'enseignement des langues) n'entre à aucun moment dans une spécialisation pour une sous-problématique précise (ex : *indexation pédagogique de textes pour l'enseignement de l'allemand LV2 dans les collèges français*). L'objectif est ici de dégager des mécanismes les plus généraux possibles. Nous allons donc tenter de confronter les pratiques à certains facteurs. Nous ne nous appuyerons pas ici sur les traditionnels indicateurs des « déterminants sociaux » évoqués par de Singly :

« les variables dont les intitulés renvoient directement à une désignation biologique – le sexe et l'âge – même si leur conception veut rompre avec une telle définition ; les variables servant à approcher le montant des capitaux, sociaux, culturels, économiques des individus interrogés ; les variables indiquant le mode d'organisation de la vie privée dans laquelle les personnes sont insérées. »

[de Singly, 1992, p. 48]

4. Nous revenons plus en détail sur les différents supports du questionnaire : § 4.2.1 p. 149.

Dans le cadre de notre étude, nous n'allons pas tenter de définir une indexation pédagogique qui dépende de ces variables puisque :

- nous traitons l'impact de l'âge sous une autre forme (expérience, cf. § d) p. 142) ;
- les études concernant le sexe comme facteur ne montrent que peu de différences entre les pratiques [Grasha, 2002; Starbuck, 2003] et les représentations [Goodwin & Stevens, 1993] respectives des enseignants et des enseignantes ;
- tous les enquêtés font partie de la même catégorie socio-professionnelle ;
- il ne s'agit pas d'une étude sociologique.

Nos hypothèses nécessitent néanmoins d'explorer d'autres facteurs pour décrire les enseignants qui répondent à notre questionnaire.

a) Langue(s) enseignée(s)

Le premier facteur auquel nous voulons confronter les pratiques concerne la langue enseignée (cf. question A.1.1 p. a-3 / hypothèse H_e). Nous devons savoir si le modèle auquel nous aboutirons est un modèle générique pour l'enseignement des langues ou s'il s'agit d'un modèle spécifique à l'enseignement d'une L2 donnée. Nous avons tenté de couvrir le plus de langues possibles, pour deux raisons essentielles :

- moins nous avons de langues représentées, moins nous pourrions savoir si la langue a un effet sur le processus de recherche de textes ;
- si le processus est lié à la langue enseignée, plus nous aurons de langues, plus nous aurons de pratiques différentes à constater ;

b) Pays

Dans [Herrerias, 2001], un panorama européen des effets des politiques éducatives pour les langues est présenté. La partie qui a attiré notre attention est celle qui traite de l'exemple du Luxembourg où « les langues étrangères ne jouent pas le rôle classique qu'on leur attribue dans le système éducatif dans la plupart des pays » [Herrerias, 2001, p. 322].

Sans entrer dans une étude comparée du statut de l'enseignement des langues selon les pays⁵, il semble donc tout à fait cohérent que la culture de l'enseignement des langues puisse influencer sur le processus de recherche de textes. Nous formulons cependant l'hypothèse que le processus profond n'est pas affecté par ce facteur. Si la culture éducative d'un système national peut influencer les pratiques, il faut considérer que l'institution et le rapport à la langue que peuvent avoir les apprenants qui la fréquentent peuvent l'influencer aussi.

c) Public

La finalité de notre étude n'est pas de créer un outil utilisable uniquement par des enseignants s'adressant à un public donné. Par public (cf. hypothèse H_g), nous faisons aussi bien référence à un niveau, à des caractéristiques propres aux apprenants comme leur âge (cf. question A.1.6 p. a-4) ou une L1 (cf. question A.1.2), qu'à une institution (cf. question A.1.4). Cette description

5. pour lequel un bon point de départ pourrait être [Herrerias, 1998] qui fait un état des lieux pour chacun des pays de l'Union Européenne (à la date de sa parution : 15 pays).

s'inspire en quelque sorte des descripteurs des principaux standards⁶ à ceci près que nous ne les considérons pas ici comme des propriétés des textes, mais comme un facteur qui pourrait influencer le processus de recherche de textes des enseignants.

d) Expérience

Plutôt que de considérer l'âge des enseignants, il nous paraît plus pertinent de savoir depuis combien de temps ils exercent ce métier (cf. question A.1.3 p. a-3 / hypothèse H_h). En effet, comme nous le verrons plus en détail dans la section 4.3.3, la littérature décrit la carrière des enseignants à travers des phases qui dépendent directement du nombre d'années qu'ils ont enseigné et non en fonction de leur âge.

e) Littératie informatique⁷

Dans leur conception, ces questionnaires n'étaient pas uniquement dédiés aux utilisateurs potentiels de notre prototype. Nous étudions ici les stratégies de recherche de textes développées par les enseignants, que ce processus fasse intervenir des ressources informatiques ou des moyens plus traditionnels. Nous ne nous adressons donc pas uniquement aux enseignants qui utilisent couramment l'outil informatique et avons conçu une version papier du questionnaire⁸.

En outre, il nous avait paru intéressant de savoir quelles étaient les sources les plus utilisées par les enseignants (cf. question A.2.6 p. a-6), notamment pour voir la place occupée par Internet dans cette liste et pour estimer le degré d'utilisation de l'ordinateur dans les pratiques actuelles de recherche de textes. C'est un indicateur que nous avons choisi pour évaluer le niveau de littératie informatique des enquêtés, puisque c'est cet aspect, plus que la maîtrise d'autres logiciels, qui nous importe dans le contexte de notre travail qui touche à la recherche de documents. Nous aurions pu préciser ce niveau en demandant aux enseignants s'ils utilisaient régulièrement des bases de données, mais nous nous sommes contenté, peut être à tort, de cet indicateur.

Pour gagner en précision, nous demandions aux enseignants d'ordonner les sources, de manière à savoir lesquelles étaient privilégiées. En outre, nous avons ici une de nos questions « semi-fermées » pour laquelle les réponses ajoutées risquaient d'être marginales par rapport aux autres (en termes d'importance au moins). Il y a donc deux aspects à la question A.2.6 : elle permet d'une part de caractériser la population, mais elle a aussi pour fonction de recenser les sources auxquelles les enseignants font appel.

f) Méthodes

Plus que de croiser les méthodes employées (question A.1.7 p. a-5) avec les processus de recherche de textes, nous voulions confronter la population de notre étude à la constatation de l'éclectisme que nous avons pu faire à travers les réponses au questionnaire de Levy [1997, p. 123] (cf. § 1.4.5 p. 24). En effet, il est probable que dans les autres facteurs choisis, certains témoignages soient à cheval sur plusieurs catégories (enquêtés enseignant plusieurs langues). Cependant, ces cas seront

6. cf. notion d'*audience* dans le LOM (éléments 5.6, Context et 5.7, Typical Age Range [IEEE, 2002, p. 29]), GEM (cf. Audience – <http://tr.im/GEMAudience>[GEM, 2004c]) ou edna (cf. Audience, [edna, 2006, p. 7]).

7. Cf. NB 3.1 p. 134.

8. <http://mathieu.loiseau.free.fr/bdtp/q1/QBaseDeTextes1.rtf> .

vraisemblablement plus marginaux que dans le cas des méthodes. Si l'hypothèse d'éclectisme est vérifiée (cf. H_d p. 138), la multiplicité des méthodes associée à chaque témoignage sera une norme. À partir de cette constatation, il devient difficile de catégoriser les réponses en fonction des méthodes employées puisque les réserves indiquées dans le paragraphe 1.4.5 (p. 24) resteront valides⁹.

La liste des méthodes est issue de [Germain, 1993] et comporte de ce fait un grand oubli ; la perspective actionnelle, qui prend réellement son essor avec le Cadre Européen Commun de Référence pour les Langues (CECR)¹⁰ ([Conseil de l'Europe, 2000]) qui est ultérieur à l'ouvrage utilisé pour établir la liste.

4.1.4 Textes fabriqués

Les questions que nous posons ont pour but d'appréhender un peu mieux le processus de recherche de documents en vue de définir un modèle qui soit représentatif des questions que se posent les enseignants. Le point d'entrée des textes fabriqués ou de l'opposition entre texte authentique et texte fabriqué est une question dont le grain est plus fin que celui du strict processus. À travers cette problématique, nous espérons obtenir des informations sur les critères qui entrent en jeu dans le choix des textes. Suite à l'avènement de l'approche communicative et l'état de grâce du texte authentique, le fait d'avoir recours à des textes fabriqués a de grandes chances d'être une décision consciente et raisonnée. Il nous a donc semblé pertinent de poser des questions aux enseignants dans ce sens : les raisons qui font qu'ils ont recours à des textes fabriqués seront vraisemblablement des caractéristiques qui entrent en compte dans le choix d'un document (cf. question A.2.3 p. a-5). À ce titre, cette dernière question fait partie des questions plus prospectives du questionnaire.

Dans l'optique du contenu de notre système, bien qu'il ne soit pas de notre ressort dans le cadre de ce travail (cf. § 1.2.1 p. 8), il nous a paru pertinent de savoir si les enseignants sont amenés à rechercher des textes fabriqués ou s'ils les écrivent systématiquement eux-mêmes (cf. question A.2.2). Cette série de questions doit aussi nous permettre de constater l'importance de la recherche de textes au sein des pratiques enseignantes (cf. H_a).

4.1.5 Recueils

Au cours de notre enquête préliminaire, il est apparu que 3 enseignants (sur les 8 de l'échantillon) possédaient leur propre « banque de textes » qu'ils consultaient pour chercher des supports d'activité. Ces recueils étaient la plupart du temps des recueils papier, mais pas uniquement. Le point le plus intéressant de cette partie de l'étude consistait à constater que souvent ces collections avaient une organisation bien précise qui reflétait ce que les enseignants attendaient de ces textes. Les 3 enseignants utilisaient un système de classement différent [Loiseau, 2003, p. 47]. Nous nous sommes donc attaché à tenter de recenser les pratiques des enseignants interrogés quant aux critères d'organisation d'éventuels recueils (cf. question A.2.4 p. a-6 et associées), qui pourront nous aider à définir des critères de recherche.

La question des recueils est emblématique de l'intérêt que peut avoir un travail sur les pratiques déclarées. Ce qui nous intéresse à ce stade de notre travail, c'est le système de classement que

9. Méthode associée à un public ou à des compétences ? Quelle méthode intervient pour quel choix de texte ?

10. Au moins en France.

l'enseignant avait l'intention de mettre en place : quels critères lui ont paru pertinents pour l'organisation de son recueil ? Dans ce cas, peu nous importe que le « faire » ne soit pas directement décidé par le « dire sur le faire » (cf. [Clanet, 2001, p. 328]), puisque plus que la manière dont les enseignants classent leur recueil, c'est la manière dont ils veulent le classer qui est pertinente pour tenter de comprendre quel critères entrent en jeu. La question des difficultés à utiliser les critères ainsi définis et des raisons qui ont mené à passer outre seront, elles aussi, pertinentes, mais à plus long terme.

4.1.6 « Recherche passive »

Dans notre série d'entretiens, la possession d'un recueil était en règle générale associée à la pratique que nous avons appelée « recherche passive », qui n'était bien évidemment jamais l'unique méthode de recherche de documents. Les enseignants qui la pratiquaient expliquaient que, dans le cadre de leurs loisirs, ils lisaient dans la L2 enseignée, et que parfois un texte leur paraissait propice à l'apprentissage. Dans ce cas, ils le remisaient et l'ajoutaient à leur recueil.

Notons ici que ce type de pratique, où l'enseignant effectue une tâche de planification dans le cadre d'une toute autre activité, n'est pas spécifique à l'enseignement des langues :

« Planning may be formal, as when a teacher prepares a lesson plan [...], or informal, including the usually invisible thinking that a teacher does while shopping, driving home from work, or eating lunch. As long as what a teacher is doing aids in preparing a framework for guiding future action, it counts as planning. ¹¹ »

[Clark & Yinger, 1980, p. 6]

Nous nous intéressons à la question de la recherche passive en deux phases, tout d'abord avec la question A.2.5 (p. a-6) dans laquelle nous tentons d'examiner l'importance relative de 3 comportements présents dans les entretiens réalisés :

- « Recherches pour une activité précise » : l'enseignant se lance dans le processus de recherche alors qu'il sait le type d'activité qu'il veut mettre en place (recherche comme processus de planification explicite) ;
- « Textes repérés au cours de lectures personnelles » : il s'agit d'une recherche passive qui suggère que l'enseignant réutilise les textes repérés au hasard de ses lectures (recherche comme processus de planification implicite) ;
- « Imposés par le programme » : il n'y a pas de réelle recherche de textes, il peut s'agir de l'utilisation d'une activité toute faite, ou alors le choix porte sur la manière dont présenter le texte aux apprenants.

Cette question était « semi-fermée », car c'est précisément le type de question au centre de notre problématique, pour laquelle nous ne voulons pas oublier de possibilités (cf. H_c). Afin de tenter d'apprécier l'importance de ces solutions les unes par rapport aux autres, nous demandons aux enseignants de les ordonner.

Nous revenons ensuite sur la recherche passive dans la suite du questionnaire avec la séquence de questions A.2.7 (p. a-6). Dans cette question, nous demandons explicitement aux enseignants s'ils y ont recours, avant de leur faire préciser ce qu'ils ont tendance à associer aux textes qui leur paraissent pertinents :

11. La planification peut être formelle, comme quand un enseignant prépare le plan d'une leçon [...], ou informel, y compris la réflexion habituellement invisible que l'enseignant mène quand il fait ses courses, rentre chez lui en voiture ou mange son déjeuner. Tant que ce que l'enseignant fait *l'aide à définir un cadre pour guider ses actions futures*, cela compte comme de la planification.

- une activité ;
- un public ;
- ou s'ils peuvent garder des textes sans objectif précis, en attendant de trouver un moyen de les utiliser.

4.1.7 Utilisations différentes d'un texte

Cette question précisant les pratiques des enseignants ayant recours à la recherche passive (cf. question A.2.7 p. a-6) constitue la plus grosse maladresse de ce questionnaire. En plus des trois propositions précédentes, se trouve celle qui doit nous permettre de confirmer l'une des hypothèses majeures que nous formulons : « Si le texte est bon il sera utilisable dans de nombreux contextes radicalement différents ». La première erreur de cette proposition réside dans le jugement de valeur du texte et l'implication mettant en lien sa qualité et son utilisabilité dans différents contextes. Si l'hypothèse était « confirmée » par ce biais, elle n'aurait pas réellement de poids, en ce sens qu'il s'agirait d'un trait des textes de « qualité ». Une qualité que nous serions bien à mal d'évaluer. Le problème majeur de cette question est de mettre sur un même plan la description de propriétés du texte et des pratiques d'enseignant. Cette proposition ne complète pas réellement les autres : le fait qu'un enseignant garde un texte pour une activité précise ou un public précis ne veut pas dire qu'il n'est pas utilisable dans des contextes différents. Cela peut signifier que l'enseignant n'enseigne que dans un seul des contextes où lui même pourrait le trouver utilisable.

Malgré tout, cette question est restée en l'état : nous n'avons pris conscience de ces défauts qu'après un certain nombre de témoignages que nous voulions pouvoir conserver. Cependant, l'importance de l'hypothèse que nous tentions de valider nous a poussé à ajouter deux questions (questions A.2.8 et a) p. a-7) après le recueil de 18 témoignages de manière à avoir des données exploitables (cf. § c) p. 151). Dans cette série de deux questions, nous demandons dans un premier temps aux enquêtés de réfléchir à la possibilité d'utiliser un texte dans différents contextes. Dans le cas où ils répondent par l'affirmative, nous leur demandons si cela leur est arrivé, ce qui donnerait plus de poids à leurs témoignages.

4.1.8 Critères

Enfin, la dernière question posée (question A.2.9 p. a-7) a pour but de tenter de nous orienter pour la suite de notre travail vers des classes de propriétés qui interviennent dans la recherche du texte. Nous avons isolé 4 types de propriétés :

- le but de l'activité en termes de pédagogie¹² ;
- le type d'activité ;
- le public¹³ ;
- l'intérêt du texte.

Lorsque nous parlons d'« intérêt indépendamment du contexte scolaire », l'imprécision du terme est intentionnelle. Nous faisons référence à des aspects du processus de recherche de textes qui paraissent difficilement formalisables, en grande partie car ils sont à peine conscients. En effet, « les enseignants n'ont pas toujours une façon rationnelle d'expliquer leur action » [Tochon,

12. Le terme « compétence » nous a semblé trop emprunt d'approche communicative, pour être utilisé dans une formule du type « les compétences à développer ».

13. Dans « public », nous incluons son niveau, ses intérêts, son âge, etc..

1989, p. 31], ce qui est apparu clairement dans notre étude préliminaire, notamment à travers le témoignage suivant :

« “Je crois qu’avec l’expérience on trouve toujours. Tu sais ce que tu vas trouver dans *Vocabulaire*, ce que tu vas trouver dans le bouquin de grammaire, ce que tu vas trouver par Internet. J’ai beaucoup d’intuition toujours, pour trouver les textes.” »

[Loiseau, 2003, p. 58]

Enfin, nous avons décidé de ne pas expliciter davantage nos catégories. Il nous a semblé qu’en les détaillant, nous pousserions les enseignants à avoir des réflexions du type que celles que nous nous sommes faites sur les propriétés du *Learning Object Metadata* (LOM) quant à l’influence que peut avoir une propriété sur une autre : par exemple, les compétences à développer au cours d’une activité risquent d’influencer le déroulement de l’activité. Or, ce qui nous importait ici, était que les enseignants communiquent sur leur ressenti et non pas qu’ils effectuent un travail de formalisation qui les pousserait à cocher des catégories auxquelles ils n’ont pas la sensation de prêter attention. Cette question donne la possibilité à l’enseignant de proposer lui-même ses solutions, ce qui devait nous permettre de voir s’il se retrouvait dans les catégories proposées. Enfin, comme il s’agit de la dernière question avant la section « commentaires », nous avons postulé que si elle laissait un sentiment de « mauvaise question », les enseignants n’auraient pas le temps de l’oublier et nous le feraient savoir (cf. § 4.2 p. 149).

4.1.9 Récapitulatif

La problématique n’est pas ici de dégager explicitement des critères ou des propriétés permettant d’arriver à une indexation pédagogique. Nos premiers objectifs concernent plutôt la définition d’un cadre générique englobant les pratiques des enseignants ainsi que l’extraction de classes de facteurs qui peuvent entrer en jeu dans le processus de recherche et de choix de textes. Nous passons d’ailleurs par des recherches de textes qui n’en sont pas réellement. Dans le cas des textes fabriqués, par exemple, nous considérons de la même manière les préoccupations des enseignants qui construisent eux-mêmes le texte, et celles de ceux qui les recherchent. Dans le cas des recueils de textes et de la recherche passive (cf. § 4.1.6 p. 144), la constitution du recueil échappe quelque peu à notre problématique dans la mesure où il ne s’agit pas de partir d’une « question » et de rechercher des textes qui y répondent, mais d’un texte dont l’enseignant se dit qu’il pourra ultérieurement répondre à une (ou plusieurs) question(s). Nous tentons de caractériser plus précisément le processus de recherche de textes. Toutefois, notre point de vue n’est ni celui d’un psychologue, ni celui d’un cognitiviste¹⁴ ; le but final est avant tout d’en tirer quelques conclusions quant aux fondements d’une forme d’indexation pédagogique. Nous espérons y parvenir à travers une alternance de questions générales à croiser avec certains facteurs et de questions prospectives au grain plus fin. Bien évidemment, toutes ces conclusions dépendront en grande partie de la confirmation de l’hypothèse selon laquelle un texte pourra être utilisé dans différentes situations.

Afin de clarifier le statut de chaque question et de donner une vision globale du questionnaire, nous les récapitulons dans tableau 4.1 (cf. page suivante). Nous associons à chaque question, ou séquence de questions, la ou les hypothèses concernées et un « type ». Nous utilisons deux variétés de descripteurs dans le type. Le premier concerne le « fond » :

- prospective : il s’agit d’une question dont les réponses devront être traitées, au moins en partie, qualitativement, afin de déboucher sur des critères ou des familles de critères qui entrent

14. Même si ces points de vue auraient été précieux.

en jeux dans la recherche ou le choix de textes ;

- variable : dans nos analyse quantitatives, qui visent en général à confirmer une hypothèse, nous faisons varier des facteurs afin d'établir leur relation (ou leur absence de relation) avec les phénomènes décrits. À travers ce que nous appelons « variables », nous tentons de décrire les enquêtés (enseignants), quitte à le faire à travers une autre population : les apprenants ;
- hypothèse : les questions dont les réponses constituent à elles seules des réponses à des hypothèses (certaines hypothèses, comme H_e ou H_f , ne peuvent être testées qu'en croisant les résultats de plusieurs questions alors que d'autres, comme H_b dépendent de réponses à une ou plusieurs questions sans faire intervenir d'autres facteurs) ;
- processus : les questions dont l'objectif est de recueillir des informations sur le processus de recherche de textes au sens large du terme (cf. § 4.1.1 p. 138). Le terme est quelque peu ambigu, car les questions prospectives visent elles aussi à décrire le processus, mais en se focalisant sur des points plus spécifiques et ironiquement avec moins de précision dans l'optique d'établir un fonctionnement généralisable (que nous ne pouvons espérer approcher avec nos questions prospectives).

Le second descripteur concerne quant à lui la forme de la question :

- fermée ;
- ouverte ;
- semi-fermée (cf. § 4.1.2 p. 139).

Les commentaires sont inclus directement dans le tableau pour en faciliter la lecture.

QUESTION ¹⁵	HYPOTHÈSE	TYPE	COMMENTAIRE
A.1.1	H_e	variable / ouverte et semi-fermée	Cette variable concerne directement les enquêtés et peut être croisée avec les questions concernant le processus. Il s'agit en quelque sorte d'une fausse question ouverte puisque les réponses sont incluses dans un ensemble fini, celui des langues du monde. Dans la version papier (la première version créée), il s'agit d'une question ouverte. Dans un souci de simplicité de traitement des données, nous avons proposé aux enseignants les langues les plus communes dans un menu déroulant et laissé une case « autre ».
A.1.2	H_g	variable / ouverte et semi-fermée	Un des différents indicateurs utilisés pour décrire le public des enseignants. Cf. ci-dessus pour la forme (question A.1.1).
A.1.3	H_h	variable / ouverte	Description des enquêtés, à croiser avec les questions concernant le processus.
A.1.4	H_g	variable / ouverte	Il s'agit du contexte d'enseignement mais nous l'avons considéré ici comme une caractéristique des apprenants, dans la mesure où les enseignants de notre étude préliminaire se référaient couramment à leurs étudiants par l'intitulé de la filière [Loiseau, 2003, pp. 42–43, 47–48, 55].

15. cf. annexe A p. a-3

QUESTION	HYPOTHÈSE	TYPE	COMMENTAIRE
A.1.5	H_f	variable / ouverte	L'influence des instances nationales sur la manière de considérer la langue s'exprime-t-elle dans le processus de recherche de textes ?
A.1.6	H_g	variable / ouverte	Complète le contexte d'enseignement en décrivant le public avec deux critères supplémentaires : âge et niveau (en plus de ceux qui décrivent l'institution).
A.1.7	H_d	variable / semi-fermée	Méthodes d'enseignement (description de la population).
A.2.1	H_a	hypothèse / fermée	La confirmation de l'hypothèse dépend aussi des réponses à la question A.2.2 (cf. ci-dessous).
A.2.2	H_a	hypothèse / fermée	Complète la question A.2.1 (cf. ci-dessus).
A.2.3		prospective / ouverte	Les raisons des utilisations des textes fabriqués pourraient expliquer certains critères.
A.2.4	H_e, H_f, H_g, H_h & H_i	processus / fermée	Le fait d'avoir un recueil ou non fait partie du processus de recherche de textes d'un point de vue stratégique : il pourrait être associé à la « recherche passive » (cf. § 4.1.6 p. 144).
a)	H_e, H_f, H_g, H_h & H_i	processus / fermée	Permet de préciser la question précédente.
b)		prospective / ouverte	Les critères organisationnels des recueils sont potentiellement des critères de recherche.
A.2.5	H_c, H_e, H_f, H_g, H_h & H_i	processus / semi-fermée (ordonnée)	Modes de recherche (cf. § 4.1.6 p. 144).
A.2.6	H_i	facteur + prospective / semi-fermée (ordonnée)	Recensement des sources potentielles et évaluation du degré de littératie informatique (selon un facteur unique).
A.2.7	H_e, H_f, H_g, H_h & H_i	processus / fermée	« Recherche passive ».
a)	$H_c, H_e, H_f, H_g, H_h, H_i$ (& H_b)	processus / fermée	Stratégies (objectifs) et erreur de formulation (cf. § 4.1.7 p. 145).
A.2.8	H_b	hypothèse / fermée	Ajoutée en cours de passation.
a)	H_b	hypothèse / fermée	Ajoutée en cours de passation. Confirmation de la question précédente.
A.2.9		prospective / semi-fermée	Classes de critères.

TABLE 4.1: Objectifs des questions du premier questionnaire

Outre les préoccupations d'un ordre strictement scientifique, l'un des objectifs de ce questionnaire était d'établir le contact avec une communauté d'enseignants qui semblait s'intéresser aux problématiques de la recherche en didactique et qui s'impliquerait éventuellement dans les itérations ultérieures du processus (cf. § *Processus* p. 52 et question A.4 p. a-8).

4.2 Protocole

Avant d'entrer dans la présentation et l'analyse des résultats, nous nous proposons d'expliquer comment les données ont été recueillies.

4.2.1 Support

Afin de toucher le plus de monde possible (cf. § 4.1.3 p. 140), nous avons décidé de diffuser notre questionnaire sous plusieurs formes : papier et électronique. Deux formats électroniques étaient disponibles : un fichier RTF et un site Web généré en PHP, interfacé avec une base de données MySQL.

Le format *Rich Text Format* (RTF) est un format ouvert, pris en charge par différents logiciels sur plusieurs plateformes [Wikipedia, 2009g], il nous a donc paru approprié pour diffuser notre questionnaire sous forme de fichier électronique. Pour la version en ligne du questionnaire¹⁶, l'utilisation de *PHP: Hypertext Preprocessor* (PHP) s'est imposée d'elle même. Nous avons besoin de créer des pages dynamiques : du fait des sections A.2.7, A.2.4 et A.2.8 du questionnaire, il s'agit d'un questionnaire « en arbre » dans lequel certaines questions sont posées en fonction des réponses aux précédentes [Ghiglione & Matalon, 1978, p. 96]. Cet aspect aurait pu être géré en Javascript, mais PHP est capable aussi d'adresser le problème (même s'il faut passer par un chargement de page dans le cas de ce dernier). De plus, PHP s'adaptait bien à la problématique du stockage des données avec la prise en charge de bases de données MySQL. En outre, PHP est un langage de programmation *open source* [PHP, 2009], largement disponible chez différents hébergeurs, dont celui que nous avons utilisé.

La version en ligne du questionnaire pouvant servir d'interface avec notre base de données, nous l'avons utilisée en tant que support référence, et avons saisi les témoignages reçus sous d'autres formes (versions papier et RTF)

4.2.2 Conditions de passation

Le fait d'avoir plusieurs supports pour un même questionnaire tendra à nuire à l'homogénéité des déroulements de la passation. Puisque dans un cas, une interaction est possible entre l'enquêté et l'enquêteur (par l'intermédiaire d'un programme informatique) et dans l'autre l'enquêté remplit en autonomie son questionnaire. Pour préciser ces légères différences, il nous faut détailler le processus de diffusion.

16. À l'heure où nous écrivons ces lignes, une version « coquille vide » – pas de stockage des réponses, toutes questions affichées d'office – du questionnaire est accessible à l'adresse suivante : <http://mathieu.loiseau.free.fr/bdtp/q1/index.php>

a) Échantillon et sous-échantillons

Une première vague de diffusion des questionnaires concernait une quarantaine d'enseignants contactés directement. La majeure partie des versions papier distribuées l'ont été à ce moment du travail. Ces enseignants faisaient partie de nos connaissances personnelles ou alors ils travaillaient à l'Université Stendhal. Nous avons pensé à un mode de diffusion selon lequel les personnes ayant répondu à l'enquête la transmettraient à leurs connaissances.

Nous avons exclu d'emblée un réel questionnement sur le terrain qui nous aurait accaparé trop longtemps ou aurait demandé un budget que nous n'avions pas. Connaissant certains enquêtés potentiels, nous avons la certitude de compter dans notre échantillon un certain nombre de personnes qui n'utilisent que peu l'outil informatique (cf. § 4.1.3 p. 140). En ne dépendant pas exclusivement d'une version informatisée du questionnaire, nous espérions diversifier les niveaux de « littératie informatique ».

Ce mode de diffusion avait débuté sous les meilleurs auspices, une enseignante réunissant pour nous douze témoignages au cours d'une séance de formation d'enseignants dans une Université Belge. Par la suite, nous appellerons ces témoignages le sous-échantillon « He1o ». Le recours à un terme mystique pour décrire ce bon début prend tout son sens en constatant qu'outre ces témoignages, le nombre de réponses reçues au bout de deux mois (18, dont 14 versions papier) était bien inférieur aux demandes envoyées et attestait de la naïveté de l'idée.

Finalement, l'écrasante majorité des réponses est intervenue après l'ajout de la série de questions concernant les usages multiples des textes (cf. § 4.1.7 p. 145). Peu après cet ajout, nous nous sommes mis à contacter des sites dédiés à l'enseignement (des langues) : le café pédagogique¹⁷, framonde¹⁸, franc-parler¹⁹, primlangues²⁰. Notons toutefois que l'appel lancé sur le site du café pédagogique est celui qui nous a amené le plus de témoignages : le nombre de réponses au questionnaire est passé d'une trentaine à une centaine dans les heures qui ont suivi la publication de notre annonce²¹.

Le problème du succès de cette annonce est qu'elle a dissipé toute velléité de représentativité de la population générale des enseignants du point de vue de la littératie informatique. Nous ne pourrions donc tester l'hypothèse H_i . En effet, la grande majorité des enseignants qui répondent à notre questionnaire maîtrisent, du fait même de la méthode de contact et du déroulement de la passation, au minimum la navigation sur internet. La vitesse à laquelle les réponses se sont accumulées à la publication, tendrait à inférer sur l'usage de listes de diffusion, voire de flux *Really Simple Syndication* (RSS). L'objectif de différenciation était de comparer les pratiques entre les enseignants qui utilisent couramment internet et les autres. Pour ne pas introduire de biais du fait de la constitution de l'échantillon [Ghiglione & Matalon, 1978, p. 48], nous considérerons que la population est composée d'enseignants qui, du fait de leur connaissance de l'outil informatique, sont susceptibles d'utiliser le nôtre²².

17. <http://www.cafepedagogique.net/> (enseignement, généraliste).

18. Liste de diffusion accessible via le site du réseau Observatoire Du Français et des langues nationales (ODF₂) : <http://www.odf.auf.org/> (enseignement des langues).

19. <http://www.francparler.org/> (enseignement des langues).

20. <http://www.primlangues.education.fr/php/> (enseignement des langues).

21. <http://www.cafepedagogique.org/disci/primaire/60.php>

22. Ce constat est d'ailleurs fortement corroboré par les résultats (cf. § b) p. 178), selon lesquels internet est utilisé comme source de textes par 123 des 130 (cf. § b) p. 178) enseignants interrogés (pour 74 d'entre eux, il s'agit même de l'une des deux sources les plus utilisées).

b) Version papier / Version « en ligne »

La version papier a été relativement peu utilisée (15 réponses sur 130²³) et uniquement dans la première vague de diffusion du questionnaire. Ayant transmis en main propre la majeure partie de ces questionnaires, nous avons fourni aux enquêtés une explication succincte de l'objectif de nos travaux. Nous encourageons les enquêtés à nous contacter pour toute remarque ou suggestion en fin de questionnaire. Cependant, tout le déroulement s'effectuait en autonomie, sans interaction, puisque nous n'assistions jamais au remplissage du questionnaire.

La version « en ligne » du questionnaire reprenait les mêmes questions que la version papier. L'enquêté débutait sur une page d'accueil résumant rapidement les objectifs de la recherche et les encourageant à nous contacter en cas de suggestion ou problème.

Une courte page d'aide était accessible par le menu à tout moment du questionnaire. Elle renvoyait vers un site expliquant les principales méthodes d'enseignement des langues et la différence entre texte authentique et texte construit²⁴ afin que les enseignants ne soient pas rebutés par la terminologie. En un sens, cette aide va à l'encontre de « la constance des conditions de passation ». En considérant l'ordinateur comme un enquêteur, il explique certaines questions à une partie des sujets en leur fournissant une aide à laquelle les sujets remplissant les questionnaires fournis par un autre enquêteur (formulaire papier) n'ont pas nécessairement accès [Ghiglione & Matalon, 1978, p. 96].

Le formulaire fonctionnait par question, l'utilisateur ne pouvant avancer à la question suivante que s'il – ou elle – avait rempli la précédente. Une aide contextuelle était fournie, le cas échéant, indiquant la question à laquelle aucune réponse n'avait été donnée. Les témoignages n'étaient enregistrés qu'après avoir répondu à la dernière question. Cette pratique avait l'avantage de nous permettre de n'avoir que des témoignages complets et tous comparables entre eux²⁵. Au contraire, l'analyse des « non-réponses » aurait pu nous fournir des résultats auxquels nous n'avons pas eu accès, comme par exemple si des questions sont perçues comme opposées aux principes des enseignants. Ce type d'information aurait éventuellement pu nous en apprendre plus sur les pratiques et attentes des enseignants [Mucchielli, 1990, p. 53].

c) Influence du protocole sur les résultats

Ce questionnaire, proposé sous plusieurs formes afin d'obtenir les témoignages les plus variés possible, a finalement perdu cette dimension généraliste du fait du relatif échec de la première vague de diffusion. Il s'est aussi accompagné de plusieurs erreurs qui ont affecté les données :

- pas de trace des témoignages incomplets, qui auraient pu être une source d'information ;
- une question ajoutée après le recueil de certaines réponses, affectant ainsi la possibilité de comparer certains témoignages ;
- des conditions de passations légèrement différentes d'un mode de recueil à l'autre.

Pour mieux comprendre la situation, nous nous proposons de les résumer sous forme de figure.

La figure 4.1 (cf. page suivante) concerne les réponses avant traitement des données (pouvant mener à l'exclusion de certains témoignages). Il recense plusieurs types d'entités :

23. Nous avons reçu 132 réponses, mais deux se sont avérées être des témoignages doublés.

24. Nous les appelons dans le cadre de ce questionnaire « textes pédagogiques ».

25. Aux erreurs mentionnées plus tôt près (cf. § 4.1.7 p. 145).

- questions (rectangles blancs²⁶), les parcours prévus sont indiqués par des flèches fines ;
- contexte de réponse (rectangles foncés), ils qualifient les questions auxquelles ils sont accolés ;
- nombre de réponses (rectangles clairs), qualifient le parcours effectif (flèches épaisses) associé.

Les parcours prévus sont indiqués par des flèches fines et les flèches en pointillés désignent un parcours linéaire entre les deux questions indiquées (se référer à l'annexe A). Les contextes permettent de préciser différentes conditions dans lesquelles les sujets ont pu aborder certaines questions. Enfin les nombres de réponses décrivent les parcours effectifs (flèches épaisses dont la noirceur est « proportionnelle » au nombre de réponses).

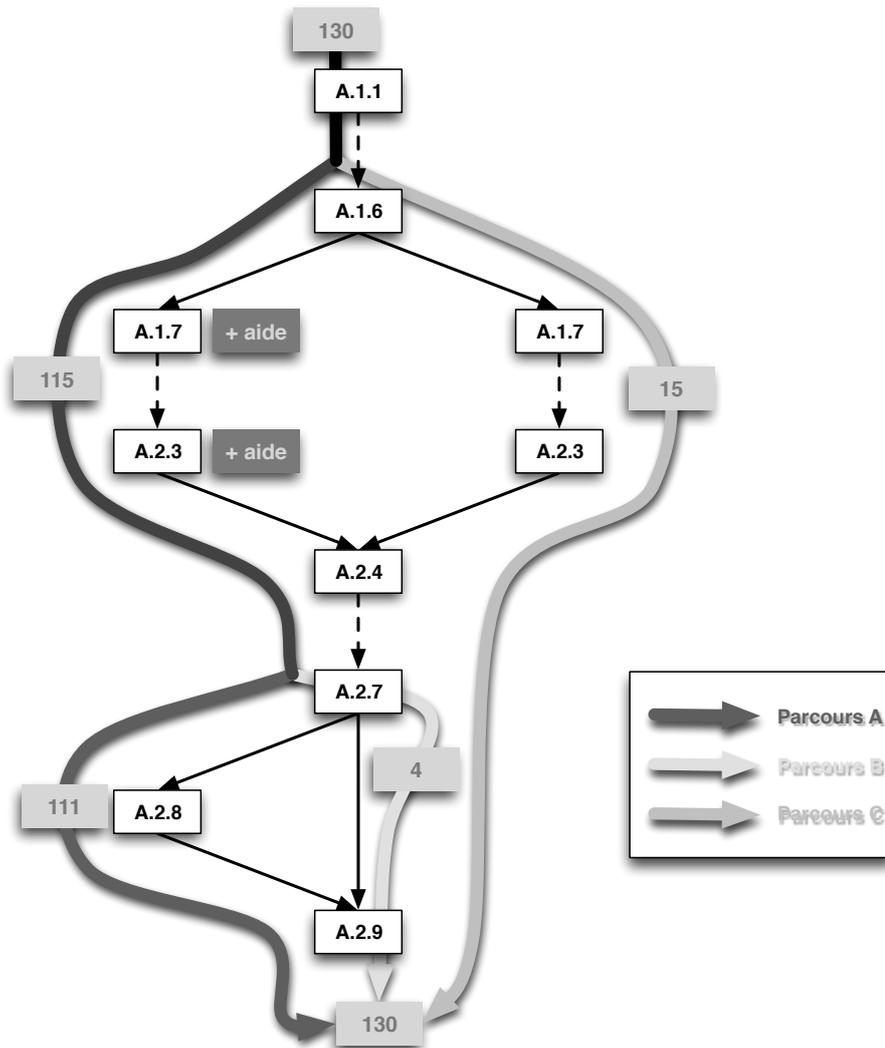


FIGURE 4.1: Parcours et effectifs (questionnaire 1)

Nous avons donc obtenu un total de 130 réponses, qui se répartissent en trois parcours. Il est à noter que si les parcours ne se rejoignent pas sur la figure c'est en réaction aux considérations de Ghiglione & Matalon [1978] concernant la « constance des conditions de passation » (cf. § *Version papier / Version « en ligne »* p. 151). Ainsi, les boites correspondant au numéro des questions donnent une représentation exacte des énoncés. Cependant, le contexte est considéré différent

26. La numérotation des questions correspond à celle de l'annexe A p. a-3

pour une même question selon le parcours qui a été effectué par l'enquêté : les 15 témoignages aux réponses A.1.7 à A.2.3 effectués sans possibilité de bénéficier d'une aide (parcours C²⁷) sont séparés des autres parcours même si les enquêtés répondent ensuite à la même séquence de questions A.2.4 à a) que les 4 enseignants dont les témoignages électroniques ont été recueillis avant l'ajout de les questions A.2.8 et a). En toute rigueur, il aurait fallu extraire un témoignage de ce parcours C : une enseignante a rempli la version papier après que nous ayons effectué l'ajout de questions. Cette dernière n'a donc pas bénéficié d'aide, mais elle a répondu aux questions sur les usages multiples des textes. Pour ne pas surcharger le schéma, nous la laissons dans le parcours C qu'elle a suivi dans les mêmes conditions que les autres enseignants jusqu'à la question a).

Nous avons donc 115 témoignages effectués dans les mêmes conditions jusqu'à la question A.2.8 (Parcours A et B²⁷). Pour les mêmes raisons, la dernière question (question A.2.9 p. a-7) ne peut pas être considérée comme remplie strictement dans les mêmes conditions pour tous les 115, puisque dans un cas les enquêtés ont répondu à une à deux questions²⁸ de plus que dans l'autre. Pour savoir si les réponses à la question A.2.9 sont influencées par la section A.2.8, il faudrait avoir plus de trois sujets n'ayant pas répondu à cette dernière (parcours B²⁷).

N'ayant pas créé de système de trace du comportement des enquêtés, nous ne pouvons savoir dans combien de cas ils ont eu recours à l'aide. Nous nous retrouvons donc avec trois ensembles de témoignages (qui s'additionnent pour arriver à 130 réponses, cf. figure 4.1). Enfin, la figure 4.1 nous montre que sur l'essentiel du questionnaire, deux de ces ensembles sont confondus.

d) Codage

Avant de voir comment nous prendrons en compte ces différents ensembles dans le traitement des résultats, une phase de codage de l'information est nécessaire (cf. [de Singly, 1992, p. 88]). Le codage est pratiquement effectué *a priori* dans le cadre des questions fermées dans la mesure où l'utilisateur ne peut pas sortir du cadre des propositions effectuées par l'enquêteur. Dans le cadre des questions ouvertes ou de celles que nous avons appelées « semi-fermées », un codage et une catégorisation des réponses seront nécessaires, qu'il s'agisse d'une analyse quantitative des réponses qui ont été données (regroupement, dénombrement, croisement avec d'autres facteurs) ou d'une analyse qualitative. En effet, même sans se poser la question de la fréquence d'une réponse, un travail de regroupement est nécessaire afin ne de pas considérer comme différentes des propositions qui sont essentiellement identiques. Dans les sections suivantes, nous nous attacherons donc à expliciter nos choix de groupements des réponses des enseignants ainsi que les autres manipulations sur les données que nous effectuerons. Nous en profiterons pour exposer les statistiques descriptives de notre questionnaire au fur et à mesure de notre traitement des données.

4.3 Description de la population

Les premières questions – question A.1.1 p. a-3 à question A.1.7 – visaient à nous en apprendre plus sur la population concernée par notre étude. Le traitement de ces données est nécessaire au test des hypothèses H_e , H_f , H_g et H_h ²⁹. Dans cette section, nous considérons notre public selon

27. cf. figure 4.1 p 152.

28. La seconde question n'est posée que pour certaines réponses à la première.

29. p. 138.

Nota Bene 4.1 Conventions de notation (témoignages)

Dans la suite du mémoire, lorsque nous nous référerons au témoignage d'un enseignant, nous l'appellerons par son identifiant dans la base de données qui a servi à stocker les témoignages, sous la forme suivante [11].

Nous nous excusons auprès des enseignants pour cette formulation impersonnelle, cependant inventer des prénoms pour environ 150 personnes aurait été pour le moins fastidieux.

4 points de vue :

- langues et pays ;
- public ;
- expérience ;
- méthodes.

4.3.1 Langues et Pays

Nous avons regroupé les réponses dans le tableau 4.2 (cf. de la présente page). Ce dernier propose des statistiques globales, mais il offre aussi le détail par parcours (cf. § c) p. 151) : pour chacune des 12 langues enseignées présentes dans l'étude, le tableau recense le nombre de témoignages qui font état de cette langue et ramène ce nombre à l'effectif total³⁰.

LANGUE 2	Pourcentage de l'effectif total		Parcours ²⁷ (Effectifs)			
	TOTAL (%)		TOTAL	A	B	C
Allemand		10.8%	14	11	0	3
Anglais		40%	52	35	4	13
Arabe		1.5%	2	2	0	0
Espagnol		36.9%	48	46	0	2
Français		10.8%	14	14	0	0
Grec		0.8%	1	1	0	0
Italien		5.4%	7	7	0	0
Lingala		0.8%	1	1	0	0
Néerlandais		8.5%	11	1	0	10
Occitan		1.5%	2	2	0	0
Portugais		0.8%	1	1	0	0
Russe		0.8%	1	1	0	0
BILAN	Langues		154	122	4	28
	Effectifs		130	111	4	15
	L2 enseignées par enquêté (moyenne)		1.18	1.1	1	1.87

TABLE 4.2: Réponses à la question A.1.1 p. a-3 (L2)

30. D'où le fait que la somme des pourcentage atteigne 118.

a) Enseignement de plusieurs langues

La première remarque que nous pouvons faire concerne le nombre de langues enseignées ramené aux enseignants. Pour cette analyse nous ne comparons pas les langues elles-mêmes³¹, nous regardons combien de langues enseigne chaque enquêté et les additionnons pour avoir une idée de combien de langues enseigne chaque enseignant. De ce fait, dans ce paragraphe, si un sujet enseigne l'anglais et l'allemand et un autre l'anglais et le russe, nous comptons 4 langues et non 3.

Dans ce contexte, malgré un effectif de 130 enquêtés, au total 154 langues sont enseignées par la population de notre questionnaire. Ainsi, chaque enseignant enseigne en moyenne 1,18 langues.

Un phénomène que nous pouvons décrire en d'autres termes :

- 18 enseignants sur 130 enseignent plus d'une langue ($\approx 14\%$);
- 12 de ces 18 sont les témoignages du sous-échantillon **He1o** (cf. § a) p. 150);
- ils représentent à eux douze 25 langues ($\approx 2,1$ langues/enquêté);
- les 3 autres représentant du parcours C n'enseignent qu'une langue;
- sans le sous-échantillon **He1o**, le nombre de sujets qui enseignent plusieurs langues diminue substantiellement : 5% des enseignants et 1,09 langues enseignées par enquêté;
- tous les enseignants qui enseignent plus d'une langue en enseignent 2 sauf trois, qui en enseignent respectivement 3, 4 et 5.

Nous n'avons pas assez de cas pour adopter une approche quantitative, cependant nous pouvons en faire une description plus fine (cf. tableau 4.3 page suivante).

Le tableau 4.3 présente l'intégralité des cas rencontrés, dont la grande majorité concerne des couples de L2 de type langue étrangère - langue seconde (cf. NB 1.1 p. 7); c'est le cas pour tous les éléments du sous-échantillon **He1o** (témoignages 1 à 12), même si ces témoignages sont issus de la région de Louvain-la-Neuve, une ville de la communauté française, où l'anglais jouit d'un statut particulier puisqu'il fait partie des choix proposés comme « deuxième langue enseignée » comme le néerlandais ou l'allemand [Lecocq, 2008, p. 27]. Nous pouvons considérer de la même manière le couple de L2 du témoignage [24] dans lequel le français est enseigné à des étudiants étrangers en France. C'est aussi le cas de l'enseignant [131] qui enseigne le Français Langue Étrangère (FLE), mais aussi l'espagnol à des francophones résidant au Mexique (langue seconde). Enfin, les cas des enseignants [134] et [120] semblent s'apparenter aux deux précédents, mais nous n'avons pas les informations pour l'affirmer. Restent ensuite 2 cas ([52] et [115]). Dans les commentaires du témoignage [52], l'enseignant précisait que le français était enseigné à des apprenants étrangers et les autres langues à des apprenants français. Seul le témoignage [115] ne combine pas des langues de statuts différents.

Parmi les sujets qui enseignent plusieurs langues, certains commentent le questionnaire et n'hésitent pas à expliquer les difficultés qu'ils ont eues à le remplir. Par exemple, l'enseignant [1] indique ne pas avoir compris la séquence de questions concernant les textes fabriqués (cf. § 4.4.1 p. 167). Nous pouvons aussi citer l'enseignant [24] qui reproche à l'étude de se focaliser sur les textes, alors que d'autres supports lui paraissent tout aussi importants (cf. § b) p. 168). Toutefois, aucun ne déplore l'impossibilité de contextualiser ses réponses en fonction de la langue enseignée.

Il semblerait alors que, si dans la plupart des cas énoncés ici, les langues enseignées ont des statuts différents, les enquêtés n'ont jamais éprouvé le besoin de dissocier la description de leurs

31. Nous y revenons dans le § c) p. 157.

ID	LANGUES	CADRE	PAYS
[120]	Espagnol Français Grec Italien	Alliance française	Argentine
[131]	Espagnol Français	Collège (3e), lycée, université, centre de langues, entreprise	Mexique
[134]	Français Portugais	Lycée/collège de la 7ème année au la 12ème (correspondant de la 4ème au BAC en France)	Portugal
[52]	Allemand Anglais Espagnol Italien Français	Chambre de commerce	France
[24]	Anglais Français	Université : licence/master1	France
[115]	Allemand Anglais	Maternelle : moyenne - grande section	France
[2]	Allemand Anglais	Collège général	Belgique
[10]	Allemand Anglais Néerlandais	Collège	Belgique
[1]	Anglais Néerlandais	Athénée royal, 1ère et 2ème générale	Belgique
[3]	Anglais Néerlandais	Secondaire supérieur	Belgique
[4]	Anglais Néerlandais	Lycée	Belgique
[5]	Anglais Néerlandais	Collège / enseignement général	Belgique
[6]	Anglais Néerlandais	Collège / enseignement général	Belgique
[7]	Anglais Néerlandais	Lycée	Belgique
[8]	Anglais Néerlandais	Collège (1, 2, 3, 4, 5, 6)	Belgique
[9]	Anglais Néerlandais	Athénée royal / degré inférieur	Belgique
[11]	Anglais Néerlandais	Athénée royal, secondaire supérieur / technique	Belgique
[12]	Anglais Néerlandais	Enseignement secondaire supérieur général et technique (5e et 6e)	Belgique

TABLE 4.3: Cadre et pays d'exercice pour des enquêtés qui enseignent plus d'une langue

pratiques selon la langue considérée ou son statut. Ces observations semblent donc aller dans le sens de l'hypothèse H_e .

b) Pays et distribution

Du point de vue du pays d'exercice, les témoignages des enseignants se répartissent selon les proportions indiquées dans le tableau 4.4 page suivante.

Nous avons choisi de re-coder deux témoignages, l'un indiquant « France et TOM » [83] et l'autre « Provence - Occitanie » [129]. Si nous avons voulu constituer une carte de France de l'enseignement de l'occitan, il aurait été pertinent de garder les spécificités du second témoignage. Cependant, l'effet que nous tentions d'évaluer était celui des politiques linguistiques nationales, que nous ne pouvons de toute façon tester avec des sous-groupes d'un seul témoignage. Finalement, la répartition n'est pas adéquate pour évaluer l'effet du pays. Malgré les envois à des listes de diffusions internationales, les réponses proviennent principalement de France : France et Belgique réunissent plus de 93% des réponses. Le fait que le questionnaire n'existe qu'en français a logiquement nui à la diffusion, mais cette répartition est aussi représentative du nombre de réponses issues du café pédagogique. Du fait de cette distribution, nous ne pourrions donc tester l'hypothèse H_f .

PAYS	Pourcentage de l'effectif total	Parcours ²⁷ (Effectifs)			
	TOTAL (%)	TOTAL	A	B	C
Angleterre	0.8%	1	1	0	0
Argentine	0.8%	1	1	0	0
Belgique	10.8%	14	2	0	12
Congo	0.8%	1	1	0	0
Espagne	1.5%	2	2	0	0
France	82.3%	107	100	4	3
Mexique	1.5%	2	2	0	0
Portugal	1.5%	2	2	0	0
TOTAL		130	111	4	15

TABLE 4.4: Réponses à la question A.1.5 p. a-3 (Pays d'exercice)

c) Une répartition « variée »

La répartition des langues dans le tableau 4.2 p. 154 montre une disparité entre les langues les plus enseignées (anglais et espagnol), 3 ou 4 langues « intermédiaires » (français, allemand, néerlandais et éventuellement italien, dernière langue dont l'effectif représente plus de 5% des témoignages) et 6 langues présentes dans un ou deux témoignages sur 130.

Nous ne sommes pas parvenus à trouver des statistiques concernant le nombre d'enseignants en fonction de la langue enseignée en France, d'où provient la plupart des témoignages. Nous nous sommes donc tourné vers la formation des enseignants et les places ouvertes aux concours de l'agrégation et du Certificat d'Aptitude au Professorat de l'Enseignement du Second degré (CAPES). Cette mesure n'est pas très fiable, puisqu'il s'agit d'instantanés de recrutement. Pour être rigoureux, il aurait fallu examiner les statistiques sur les 35 dernières années. En outre, même si nous avions toutes ces statistiques il nous manquerait au moins une partie des enseignants du supérieur et une partie de ceux qui s'inscrivent dans le cadre d'un centre de formation.

Nous avons donc pris arbitrairement les statistiques des concours sur les 5 dernières années disponibles 2002 [MEN, 2003, p. 255], 2003 [MEN, 2004, p. 263], 2004 [MEN, 2005, p. 287], 2005 [MEN, 2006, p. 297] et 2006 [MEN, 2007, p. 311]. Sur cette période, l'anglais est passé graduellement de 53,09% à 61,01%³² des postes ouverts au concours (en augmentant chaque année). Les points sont principalement grappillés au détriment de l'espagnol, de l'allemand et de la catégorie « autres »³³ restant relativement stables en proportion (en effet, le nombre de postes est globalement en baisse, malgré un sursaut en 2005). Globalement, sur ces 5 années, l'anglais représente 57% des postes, l'espagnol 27,5% et l'allemand 8,55%. En comparant ces chiffres à ceux du tableau 4.2, nous pouvons constater que malgré l'apparente surreprésentation de l'anglais et de l'espagnol, l'ensemble de L2 de notre échantillon est bien plus diversifié que la tendance actuelle dans le système scolaire français, et cela reste le cas même en excluant le sous-échantillon He10 et les enseignants de FLE. Ainsi, même si nous ne pourrions pas valider pour toute langue l'hypothèse H_e , nos données présentent une certaine « variété » compte-tenu des tendances actuelles.

32. cf. annexe B p. a-11.

33. Les langues représentées dans ces documents sont l'anglais, l'espagnol et l'allemand, toutes les autres se trouvent regroupées dans la catégorie « autres ».

d) Couples L1-L2

L'hypothèse H_g concerne les multiples dimensions de la notion de public, parmi lesquelles la L1 des apprenants (cf. § c) p. 141). Cependant, considérer les L1 de manière autonome semble une fausse piste, du fait de deux constatations.

La première concerne nos données. Du fait des conditions de recueil des données (la dépendance très flagrante au site « café pédagogique »³⁴), le français est la L1 la plus représentée. Sur 192 L1 sélectionnées³⁵, nous dénombrons 101 occurrences du français. La deuxième langue la plus fréquente dans notre étude, l'espagnol, ne concerne plus que 18 cas sur 192. Seules 4 langues, sur les 17 représentées dépassent les 10 réponses.

Cependant, la raison la plus pertinente à nos yeux provient du fait que la L1 ne peut être testée indépendamment de la L2. Comme nous le faisons remarquer à travers la notion de xénité, son impact dans l'enseignement des langues dépend d'un couple L1-L2 et non de l'une ou l'autre prise seule. Cette dimension se retrouve d'ailleurs dans les commentaires de certains témoignages ayant sélectionné une ou plusieurs L2 et plusieurs L1 : « Les langues suivantes sont enseignées à des francophones : Allemand, Anglais, Espagnol, Italien. Le Français par contre est enseigné aux étrangers dont la nationalité est listée [(15 langues)] » [52].

À considérer des couples L1-L2, nous n'avons plus 17 cas possibles (nombre de L1 différentes présentes dans nos données) mais 64 (combinaisons L1-L2). Parmi ces 64 couples L1-L2, seuls les 13 plus fréquents existent en plus de deux occurrences. La L1 de 4 des 5 cas les plus fréquents est le français, ce qui représente 107 témoignages sur 226.

Nous nous sommes donc tournés vers les travaux de Robert [2004], qui définit quatre classes de couples L1-L2 [Robert, 2004, p. 501–502] :

- langues voisines : « langue cible apparentée avec forte intercompréhension linguistique et références culturelles communes » [Robert, 2004, p. 501] ;
- langues proches : apparentées, avec intercompréhension, mais sans référents culturels communs ;
- relativement éloignées : partiellement apparentées ;
- très lointaines : non apparentées.

Robert conclut enfin que, pour les débutants, chaque classe de couple L1-L2 devrait se voir associer une pédagogie propre, alors que pour les niveaux les plus avancés, la définition d'une pédagogie des langues voisines et une autre des langues très lointaines seraient suffisantes. Ces deux pédagogies pourraient être mixées pour les cas intermédiaires [Robert, 2004, p. 510].

L'existence de ces quatre niveaux de xénité linguistique est, de l'aveu même de Robert, « arbitraire » [Robert, 2004, p. 502] et, si elle est très propice à la réflexion, elle ne peut être utilisée telle quelle pour classer nos couples L1-L2. Les exemples choisis explicitent les définitions, mais un travail reste à faire quant aux critères pour obtenir une réelle typologie dont l'utilisation fournirait des résultats cohérents et reproductibles. Par exemple, la notion de « références culturelles communes » est difficile à mettre en place dans un cadre s'appuyant strictement sur un découpage national des langues. Pour l'auteur, le catalan et le castillan sont des langues voisines alors que le catalan et le français sont des langues proches. La raison invoquée est que tout Catalan d'Espagne connaît les horaires des repas en Espagne. Il n'est pas certain que cette justification

34. § a) p. 150.

35. Chaque enseignant pouvait avoir des classes hétérogènes du point de vue de la L1 des apprenants ou avoir plusieurs classes homogènes de L1 différentes. Le questionnaire lui laissait donc la possibilité de saisir plusieurs langues.

soit acceptée par tous les Catalans de France³⁶. À la lumière de ces éléments culturels, il paraît difficile de considérer l'apprentissage « du catalan (d'Espagne) pour les Français » de la même manière pour tous les Français, quelle que soit la région.

Du fait de tous ces facteurs (nécessité de considérer des couples de langues, typologie difficile à mettre en œuvre, données difficilement exploitables), nous avons décidé de ne pas tenter de confirmer ou d'infirmer l'hypothèse H_g du point de vue de ce critère de la langue. En outre, parmi les 22 enseignants qui ont déclaré être confrontés à plusieurs L1, aucun ne nous a fait remarquer de difficulté à remplir le questionnaire ou n'a tenté de contextualiser ses réponses selon ce critère.

4.3.2 Description du public

Nous avons constaté, à travers la question de la L1, que la description du public pouvait s'avérer problématique. Cela a également été le cas avec les 3 autres indicateurs que nous avons choisis : le cadre d'enseignement – qui, d'après l'étude de Starbuck [2003], était un facteur influençant les pratiques des enseignants –, l'âge et le niveau.

a) Cadre

Le tableau 4.3 montrait, sur un échantillon de 18 cas, la diversité des cas rencontrés. Cette dernière tient en particulier aux systèmes éducatifs des pays. Les enseignants indiquent parfois des équivalents avec la France³⁷. Afin de pouvoir comparer les résultats, nous nous sommes appuyé sur le système éducatif français et avons ramené les cadres à une ou plusieurs de sept catégories (cf. tableau 4.5 p. 159). Dans le cas des collèges, lycées et athénées belges (cf. NB 4.2 p. 160), nous avons utilisé l'intitulé de la classe pour décider de la catégorie à laquelle les ajouter (collège ou lycée).

CADRE	Pourcentage de l'effectif total		Parcours ²⁷ (Effectifs)			
		TOTAL (%)	TOTAL	A	B	C
Maternelle	<input type="checkbox"/>	2.3%	3	3	0	0
Primaire	<input type="checkbox"/>	9.3%	12	12	0	0
Secondaire	Collège	<input type="checkbox"/>	63	56	1	6
	Lycée	<input type="checkbox"/>	56	46	0	10
	Tout le secondaire	<input type="checkbox"/>	101	86	1	14
Supérieur (non-spécialistes)	<input type="checkbox"/>	19.4%	25	21	3	1
Supérieur (spécialistes)	<input type="checkbox"/>	3.9%	5	5	0	0
Centre de formation	<input type="checkbox"/>	7.8%	10	10	0	0
TOTAL		Effectifs	129	110	4	15

TABLE 4.5: Réponses à la question A.1.4 p. a-3 (Cadre de l'enseignement)

Le tableau 4.5 nécessite quelques explications. Premièrement, l'effectif total est 129 et non 130. Cela provient d'un témoignage supprimé d'un enseignant qui enseigne à « tous » les niveaux

36. Le département des Pyrénées-Orientales a approuvé le 10 décembre 2007 une charte en faveur du catalan, qui stipule dans son préambule : « la langue catalane, née il y a plus de mille ans, constitue un des piliers de notre identité, du patrimoine et de la richesse du département des Pyrénées-Orientales (Catalunya Nord) » [CGal66, 2007].

37. cf. [134] dans le tableau 4.3 p. 156.

Nota Bene 4.2 Système éducatif belge [Draelants & Maroy, 2002]

La plupart des témoignages recueillis en dehors de la France provenaient de Belgique (sous-échantillon He10). Outre les politiques linguistiques, qui mettent en exergue les situations différentes des trois communautés, il nous a été difficile de classer certains témoignages du fait de l'existence d'institutions plus ou moins proches du système français : collèges, lycées et athénées. Très schématiquement – nous n'avons pas la prétention de décrire tout un système éducatif en quelques lignes –, le système éducatif belge repose sur plusieurs principes fondamentaux :

- la liberté d'enseignement (liberté pour quiconque d'ouvrir une école), qui aboutit par la loi de 1959 à la liberté des pouvoirs organisateurs ;
- la liberté de choix des parents ;
- la gratuité de l'enseignement.

Les écoles dépendent donc de Pouvoirs Organisateurs (PO) qui sont « l'autorité, la ou les personne(s) physique(s) ou morale(s) qui [...] assume(nt) la responsabilité [de l'établissement d'enseignement] » (p. 14). Selon que ce PO est une personne de droit public ou de droit privé on parlera, dans la communauté française, respectivement d'école « officielle » et d'école « libre » (p. 15). Les PO des écoles libres sont à 98% affiliés au « réseau libre catholique ». Ce qui nous mène aux athénées, qui d'après la Wikipédia sont des « écoles secondaires (laïques par opposition aux collèges) » [Wikipedia, 2009a]. En effet, qu'il s'agisse de collèges, de lycées ou d'athénées, tous sont des établissements secondaires, qui vont, en Belgique, de la 1^{ère} à la 6^e (équivalents respectifs de la 5^e et de la terminale).

dans « tous » les cadres. Ne connaissant pas le référentiel, il nous est difficile d'estimer cette « totalité ». La ligne « tout le secondaire » vise à corriger le côté un peu artificiel de la répartition « à la française » : un enseignant belge qui enseigne, dans la même institution, à des classes qui correspondraient, en France, au collège et au lycée, voit son témoignage dédoublé. Cette ligne permet de voir que les enquêtés qui enseignent dans le secondaire ne représentent pas 92,2%³⁸ de nos effectifs, mais « seulement » 78,3%. Toutefois, pour les analyses croisées, nous garderons notre découpage originel, et ne considérerons que les témoignages d'enquêtés exerçant en France pour ne pas introduire de biais, qui refléterait autant le pays d'exercice que l'institution.

b) Simplification de la variable

Pour les autres composantes de la variable « public » que nous avons définie, le niveau et l'âge, nous avons été confronté à un certain nombre de difficultés dans le traitement des réponses qui nous ont poussé à ne considérer que le cadre.

Nous n'avions pas contraint les réponses, laissant les enseignants libres d'utiliser l'échelle de leur choix. Le référentiel du CECR est très peu utilisé : seulement 2 témoignages y font référence sur les 130 du questionnaire. Les échelles employées sont les suivantes :

- débutant, faux-débutant, intermédiaire/moyen, avancé/confirmé, (supérieur), spécialistes ;
- nombre d'années d'étude de la langue ;
- statut de la langue (LV1/LV2/LV3) ;
- intitulé de la formation (ex : CAP, BACPro, 4^e, seconde à Bac+5, etc.)
- CECR [Conseil de l'Europe, 2000].

Les différentes échelles sont parfois mélangées³⁹, mais elles existent toutes en autonomie, c'est-

38. = $effectif_{collège} + effectif_{lycée}$

39. ex : LV1 ; de débutants à +4 ans.

à-dire sans référence aux autres échelles. Du fait de l'hétérogénéité des descriptions, nous ne prendrons pas non plus le niveau en compte dans nos indicateurs. En effet, le niveau tel que décrit par les enseignants dépend en règle générale de l'institution dans laquelle ils enseignent. Cette dernière donne même parfois plus d'indications sur le niveau que les informations décrivant le niveau elles-mêmes. C'est par exemple le cas quand le niveau n'est décrit que par un statut et que c'est le cadre qui permet de savoir s'il s'agit d'une LV2 au collège ou au lycée.

De la même manière, dans nos réponses, l'âge est toujours dépendant de l'institution. Cependant, les enseignants ont une manière différente de répondre. Souvent, ils donnent l'âge que les apprenants sont « censés » avoir dans le cadre de la formation qu'ils dispensent. Ils prennent parfois en compte le cas de redoublants. Certains éludent ce problème en ayant recours à la classe pour décrire l'âge des apprenants (« de 4ème à terminale » [62]). Enfin quand il s'agit de formation pour adultes, certains donnent un intervalle fermé (« 8.....78 » [83] ou « 20-55 ans » [91]) mais il peut arriver que l'intervalle soit ouvert (« >15 » [131]). Dans ces cas là, il est difficile de prendre en compte la borne supérieure de l'intervalle.

Nous avons supprimé 3 composantes de ce que nous avons appelé le public : niveau, âge et L1. Nous ne pouvons donc tester l'hypothèse H_g telle que nous l'avions formulée initialement. Au lieu de considérer le public, nous nous focaliserons plutôt sur les données que nous avons recueillies et qui en sont le plus proches : le cadre.

Nous formulons donc une nouvelle hypothèse que nous testerons en lieu et place de H_g :

H_{g2} . le cadre de l'enseignement n'a pas d'effet sur le processus général de recherche de textes.

H_{g2} est donc issue de H_g , mais elle en est différente, puisque nous n'avons pas de preuve que le public peut être modélisé adéquatement par l'institution à laquelle il est associé. Les réponses des enseignants suggèrent cependant un lien entre les deux.

4.3.3 Des enseignants globalement expérimentés

Outre des informations géographiques et institutionnelles, nous voulions considérer l'effet de l'expérience des enseignants sur les pratiques (cf. H_h p. 138).

Pour évaluer les effets de l'expérience des enseignants sur le processus de recherche de textes, nous avons décidé de séparer les enseignants en deux groupes : les « débutants » et les « expérimentés ». Pour y parvenir nous nous sommes appuyé sur les travaux de Lamarre [2003] qui, se référant à [Fuller, 1969; Huberman, 1989a; Katz, 1972], estime que les premières expériences de la vie d'enseignant s'articulent en trois étapes : la survie, la découverte et la consolidation pédagogique [Lamarre, 2003, p. 27]. La survie pourrait être décrite comme un « choc du réel » [Huberman, 1989a, p. 13], une étape de tâtonnement, où l'enseignant met en place des stratégies qui soulignent souvent le décalage entre ses idéaux et la réalité quotidienne de sa classe. La découverte décrit une période marquée par « un fort enthousiasme à débiter dans la profession », au cours de laquelle l'enseignant « explore, tente des expériences et s'interroge sur les matières à enseigner et le contenu des programmes » [Lamarre, 2003, p. 27]. Huberman regroupe ces deux étapes sous le terme générique « d'exploration » [Huberman, 1989a, p. 14], du fait qu'elles se vivent souvent en parallèle ; le second aspect aidant à tolérer le premier. Enfin, la « consolidation pédagogique » ou « stabilisation », qui correspond à la décision de l'engagement « définitif » de l'enseignant dans sa carrière⁴⁰ [Lamarre, 2003, p. 28]. D'un point de vue pédagogique, « la stabilisation précède

40. Cela peut correspondre à un acte administratif, comme une nomination officielle.

légèrement ou va de pair avec un sentiment accru de “maîtrise” pédagogique. Par exemple, les études de Fuller [1969] et de Burden [1981] évoquent un sentiment de confiance et de “confort”, assorti d’une plus grande décentration : on se préoccupe moins de soi-même et davantage des objectifs didactiques » [Huberman, 1989a, p. 15].

Nous allons donc séparer les enseignants en deux groupes : les enseignants « stabilisés », que nous appellerons enseignants « expérimentés » et les enseignants débutants qui se trouvent soit dans la phase d’exploration, soit dans le processus de stabilisation. Le problème est de fixer une limite entre ces deux phases puisqu’elles ne sont pas vécues de la même façon et *a fortiori* sur la même durée par tous les enseignants [Lamarre, 2003, p. 28]. Nous fixerons, dans le cadre de cette expérience, la limite à 5 années d’expérience. Cette limite est cohérente avec différents travaux :

- la stabilisation / consolidation d’un répertoire pédagogique « vient dans la fourchette 4-8 ans d’expérience » [Huberman, 1989b, p. 53];
- « *Maturity may be reached by some teachers within three years, by others in five or more* »⁴¹ [Katz, 1972, p. 7];
- dans l’étude de Burden [1981], les enseignants sont interrogés sur leur perception de leurs premières années d’enseignement. Après analyse, il regroupe les caractéristiques des témoignages en 3 groupes : première année, années 2 à 4 et année 5 et au-delà [Burden, 1981, 7–9];
- Zahorik dans son étude sur la planification adopte le même découpage [Zahorik, 1975, p. 136].

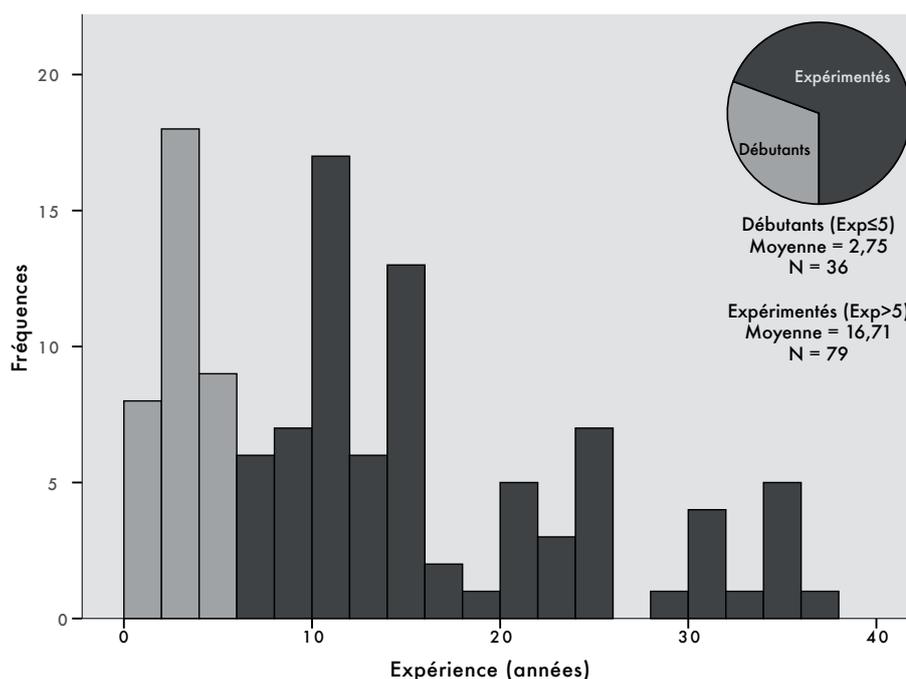


FIGURE 4.2: Distribution de l’expérience des enseignants (débutants et expérimentés)

L’histogramme de la figure 4.2 représente l’ensemble des témoignages en fonction de leur expérience. Une partie des questionnaires papier ne présentait pas de réponse et certains enseignants

41. La maturité peut être atteinte par certains enseignants au bout de 3 ans, et pour d’autres au bout de 5 ou plus. Nous ne retenons pas la phase de consolidation de Katz comme équivalent de celle d’Huberman, car ce n’est qu’au cours de la phase de maturation, que l’enseignant se considère comme tel.

ont entré des valeurs non conformes dans l'interface Web, d'où le total de 115 réponses⁴². Pour ne pas surcharger le diagramme, chaque barre de l'histogramme représente une ou deux années d'expérience. Ainsi la première représente les enseignants ayant une année d'expérience, la seconde les années 2 et 3, et la troisième les années 4 et 5. Dans les barres représentant les enseignants dits expérimentés (barres foncées), les deux premières représentent respectivement les enseignants ayant 6 ou 7 et 8 ou 9 années d'expérience, alors que la suivante ne représente que les enseignants qui enseignent depuis 10 ans. Il s'agit de la tranche d'expérience la plus représentée (17 enseignants) devant les tranches des enseignants ayant 3 et 15 ans d'expérience, qui comptent respectivement 11 et 10 éléments. Comme il était possible de s'y attendre, du fait de la définition des tranches, les enseignants expérimentés (79) sont beaucoup plus nombreux dans notre échantillon que les enseignants débutants (36)⁴³. Cependant, le nombre d'enseignants débutants n'est pas négligeable et permet d'effectuer les tests concernant l'hypothèse H_h .

4.3.4 Méthodes et éclectisme

Le dernier élément visant à décrire la population d'enseignants concerne leur recours aux méthodes d'enseignement. La question A.1.7 est posée en vue de confirmer l'hypothèse H_d . Les méthodes proposées sont issues de Germain [1993], mais la question est « semi-fermée » pour permettre aux enseignants de faire état de méthodes plus récentes ou moins connues.

Les résultats sont recensés dans le tableau 4.6. Les choix marqués d'un astérisque (*) proviennent de réponses de type « autres ». Ils ont été recodés et ne sont pas nécessairement exprimés avec les mots des enquêtés.

D'un point de vue « qualitatif », en comparant les résultats avec ceux de l'étude de Levy [1997], nous pouvons remarquer une certaine cohérence (cf. tableau 1.1 p. 23), compte-tenu du fait que les deux études ont une dizaine d'années d'écart et des différences de méthodologie : l'étude de Levy concerne avant tout des enseignants anglophones et nous n'avons pas construit notre questionnaire en vue de cette comparaison. Les méthodes proposées sont différentes : notre questionnaire ne proposait pas les choix « *formal grammar* » (cf. NB 1.3 p. 24), « *cognitive code* » ou de la méthode orale, alors que celui de Levy ne faisait pas mention de l'approche axée sur la compréhension. L'absence de la perspective actionnelle de nos résultats s'explique bien évidemment par son absence dans les choix proposés mais aussi par la continuité perçue entre l'approche communicative et la perspective actionnelle, qui est par exemple relayée dans le guide d'utilisation du CECR par Trim : « "L'apprentissage fondé sur la tâche" est, tout naturellement, une tendance forte et croissante dans le cadre de l'approche communicative » [Trim, 2002, p. 47].

Malgré toutes ces différences, dans les deux cas l'approche communicative reste de loin la méthode la plus utilisée (81,5% ici). Les différences les plus marquantes (outre les méthodes exclusives à l'un ou l'autre des questionnaires) concernent les méthodes audio-orale et grammaire-traduction, qui sont beaucoup plus populaires dans notre questionnaire⁴⁴. En dehors de ces cas, les ordres sont comparables d'une étude à l'autre.

C'est aussi le cas concernant le nombre de méthodes par enseignant situé entre 3 et 4 dans les deux études. En effet, une étude plus strictement quantitative nécessite que nous nous penchions

42. Alors que nous avons recueillis 130 témoignages.

43. Cf. légende sous forme de « camembert ».

44. ce qui pourrait s'expliquer par le *cognitive code* qui est, rappelons le, une variante de la méthode grammaire-traduction. En additionnant les proportions de ces deux méthodes dans l'étude de Levy, on retrouve le niveau de la méthode grammaire-traduction dans notre étude (cf. tableau 1.1 p. 23).

MÉTHODE	Pourcentage de l'effectif total (130)		Parcours ²⁷ (Effectifs)				
	TOTAL (%)		TOTAL	A	B	C	
Approche communicative		81.5%	106	90	4	12	
Approche axée sur la compréhension		73.8%	96	83	3	10	
Audio-orale		41.5%	54	45	1	8	
Approche fonctionnelle-notionnelle		32.3%	42	36	3	3	
Méthode situationnelle		31.5%	41	34	1	6	
Grammaire-traduction		28.5%	37	31	1	5	
Approche naturelle		13.8%	18	16	0	2	
Méthode directe		13.1%	17	14	0	3	
Approche intégrative		10.8%	14	8	1	5	
Méthode par le mouvement		9.2%	12	10	1	1	
Curriculum multidimensionnel		6.2%	8	7	1	0	
Méthode communautaire	<input type="checkbox"/>	4.6%	6	6	0	0	
Méthode des séries	<input type="checkbox"/>	3.8%	5	5	0	0	
Méthode suggestopédique	<input type="checkbox"/>	3.8%	5	3	0	2	
SGAV	<input type="checkbox"/>	3.8%	5	5	0	0	
Méthode par le silence	<input type="checkbox"/>	2.3%	3	3	0	0	
Approche cognitive*	<input type="checkbox"/>	0.8%	1	1	0	0	
Approche ludique*	<input type="checkbox"/>	0.8%	1	1	0	0	
Perspective actionnelle*	<input type="checkbox"/>	0.8%	1	1	0	0	
non pris en compte	Expérience*	<input type="checkbox"/>	0.8%	1	0	0	1
	Éclectisme*	<input type="checkbox"/>	1.6%	2	2	0	0
	Non exploitable*	<input type="checkbox"/>	2.3%	3	3	0	0
TOTAUX	Méthodes		472	399	16	57	
	Effectifs		128	109	4	15	
	Méthodes employées par enquêté		3.69	3.66	4	3.8	

TABLE 4.6: Réponses à la question A.1.7 p. a-5 (Méthodes)

sur les totaux du tableau 4.6. Ces derniers recensent le nombre de méthodes utilisées et les rapportent au nombre d'enseignants. Cependant, certains choix n'ont pas été pris en compte dans ce calcul. Premièrement, il s'agit bien d'un choix non pris en compte et non d'un témoignage. En effet, si un enseignant choisit plusieurs méthodes et ajoute un choix que nous n'avons pas jugé pertinent⁴⁵, seul le choix qui n'est pas pertinent est ignoré. De ce fait, seuls 2 témoignages ont été supprimés, pour 6 choix ignorés. Comme le résultat est un nombre de méthodes employées par enquêté, nous n'avons pas pris en compte les enseignants qui déclareraient utiliser plusieurs méthodes (« éclectisme »), sans indiquer lesquelles. Un autre exemple de témoignage non exploitable pourrait être « Cf directives de l'IG d'espagnol... » [116]. En effet, en prenant les directives de l'inspection générale, nous projeterions notre interprétation sur les pratiques de cet enseignant, il ne s'agirait donc plus de pratiques déclarées.

45. ex : « Approche axée sur la compréhension / Approche communicative / Approche intégrative / Audio-orale / Méthode situationnelle / Documents iconographiques, chansons, jeu de rôle » [104].

a) Test de l'hypothèse H_d

Une fois les réponses codées, nous nous trouvons avec la distribution de la figure 4.3. Chaque barre de l'histogramme représente le nombre d'enseignants déclarant utiliser un nombre de méthodes donné. Aucun enseignant ne déclare utiliser aucune méthode. La première barre représente donc les 14 enseignants déclarant n'utiliser qu'une seule et unique méthode.

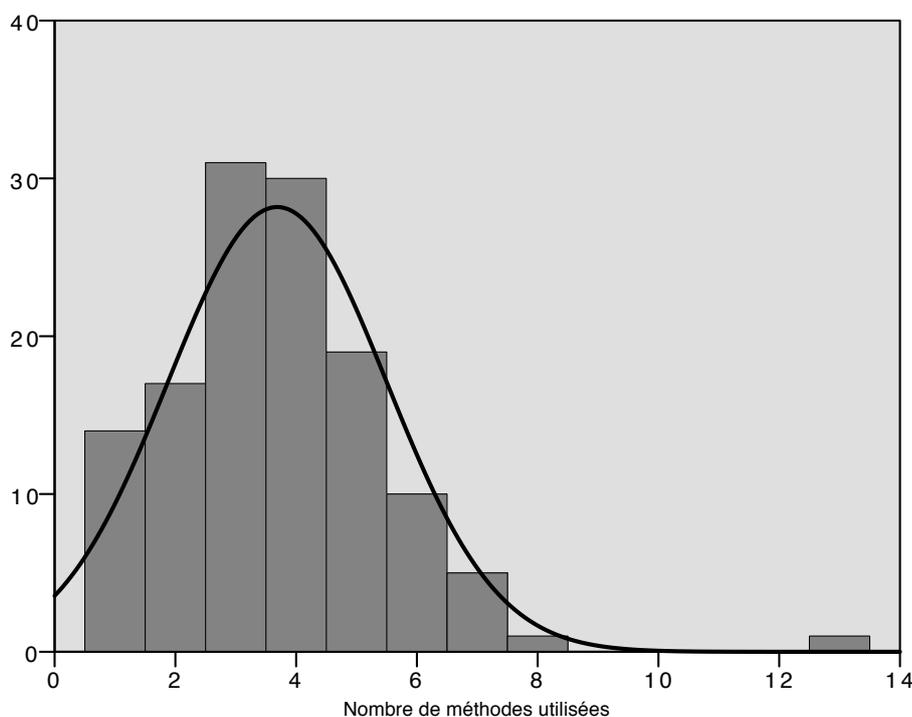


FIGURE 4.3: Distribution des enseignants en fonction du nombre de méthodes sélectionnées (la courbe indique la distribution normale)

La courbe de la figure 4.3 p. 165 représente une distribution normale calquée sur les données de l'échantillon. Un échantillon suivant une distribution Gaussienne parfaite verrait cette courbe couper le côté supérieur de chaque rectangle en son milieu.

Sans être parfaite, la distribution de notre échantillon par rapport au nombre de méthodes utilisées, nous permet d'utiliser un test-t univarié. Ce dernier est un test de moyennes et permet d'évaluer la probabilité selon laquelle un échantillon est issu d'une population dont la valeur moyenne pour une variable donnée est μ [Sheskin, 2004, p. 159].

Dans notre cas, l'hypothèse nulle, H_0 correspondra à $\mu = 1$:

H_0 . Les enseignants de langue utilisent en moyenne une seule méthode d'enseignement.

Pour confirmer notre hypothèse H_d (p. 138), l'hypothèse alternative correspondant à H_0 est la suivante :

H_1 . Les enseignants de langue utilisent en moyenne plus d'une méthode d'enseignement ($\mu > 1$ ⁴⁶).

46. Hypothèse unilatérale : nous ne postulons pas seulement une différence entre les moyennes mais une relation de supériorité.

Dans les conditions énoncées dans le paragraphe précédent, les enseignants se servent en moyenne de 3,69 méthodes⁴⁷. Or, le test *t* indique que la différence entre les moyennes est très significative⁴⁸, nous pouvons par conséquent rejeter H_0 (différence significative selon le test *t* et $3,69 > 1$). Et par là même, c'est notre hypothèse H_d qui est validée.

4.3.5 Réduction du contexte

La description de notre échantillon porte en elle les remarques que nous avons faites sur notre mode de recueil de données. En rédigeant notre questionnaire en français, nous étions voués à obtenir principalement des témoignages d'enquêtés exerçant dans des pays francophones⁴⁹ ou enseignant le FLE⁵⁰. Tout en étant conscient de ces limites, le recours à des listes de diffusion, notamment dans les réseaux FLE, nous laissait espérer une plus grande variété dans les pays d'exercice.

Du point de vue des L2 enseignées, la prépondérance de certaines langues (anglais et espagnol et dans une moindre mesure allemand, néerlandais et français) reflète les politiques linguistiques des pays les plus représentés dans l'étude : d'une part en France où l'anglais et l'espagnol sont les langues pour lesquelles le plus de places sont ouvertes au concours, et d'autre part en Belgique où du fait de l'existence de plusieurs langues officielles, les enquêtés enseignent souvent une L2, que nous qualifierons ici d'étrangère et une autre L2 qui pourrait être considérée comme officielle ou seconde. Nos données présentent cependant une distribution plus « équilibrée » que le recrutement à l'heure actuelle en France⁵¹.

Nos données sur les L2 se trouvent complètement « balkanisées » une fois croisées avec les L1, qui devaient permettre de prendre en compte le public. L'examen des données visant à décrire le public révèle que nous ne pourrions l'intégrer à nos résultats qu'à travers le prisme du cadre dans lequel l'enseignant exerce (reformulation de l'hypothèse H_g en H_{g2} p. 161).

En revanche, nos données ne posent pas de problème pour prendre en compte l'expérience des enseignants et ont confirmé notre hypothèse H_d de manière cohérente avec les données fournies par l'étude de Levy [1997] (cf. § 1.4 p. 18).

À l'exception des hypothèses H_i et H_f que nous avons dû écarter du fait de nos données (cf. § 4.2.2 p. 149 et § b) p. 156) et de l'hypothèse H_g que nous avons dû reformuler, nous pourrions tester chacune des autres hypothèses, quitte à en restreindre quelque peu la portée en fonction des informations disponibles.

47. $\sigma = 1,8$; médiane = 4.

48. $t(127) = 16,85$; $p < ,001$.

49. France : 107, Belgique : 14, Congo : 1, soient 122 témoignages sur 130 (cf. tableau 4.4 p. 157).

50. 14 témoignages sur 130 (cf. tableau 4.2 p. 154).

51. Si tant est que le mot « équilibré » puisse faire sens ici. Nous ne parlons pas d'un éventuel équilibre avec la demande des apprenants, qui peut elle-même être le reflet d'une politique linguistique nationale, qui pourrait elle-même être influencée par les attentes des tissus économiques locaux, ou quelque autre facteur social. Nous faisons seulement remarquer que, malgré des langues beaucoup plus représentées que d'autres, les disparités sont légèrement moins criantes dans nos données notamment du fait d'un fort contingent d'enseignants de l'espagnol, mais aussi d'une proportion « réduite » d'enseignants de l'anglais. Le terme est employé dans la problématique de nos recherches, pour laquelle des données idéales pour cette question (cf. H_e p. 138) auraient fait état d'une grande variété de langues, toutes représentées en nombres comparables.

4.4 Vers une description du processus de recherche de textes : les réponses des enseignants

Ayant précisé les caractéristiques des enseignants qui ont rempli notre questionnaire, nous allons maintenant examiner point par point les réponses qu'ils font à nos diverses questions afin de caractériser le processus de recherche de textes. Ce processus intervient avant que nous tentions de confirmer nos hypothèses (cf. § 4.1.1 p. 138) et a pour but de présenter les réponses des enseignants. La diversité des éléments traités donne à cette longue section un aspect « *patchwork* », elle est cependant nécessaire afin que nos conclusions, reprises dans les sections suivantes, soient explicitées. Après avoir examiné les réponses à chaque question dans le détail, nous examinerons pour certaines (cf. tableau 4.1) les influences des facteurs que nous avons décrits dans la partie précédente (cf. § 4.5 p. 188).

4.4.1 Texte fabriqué et maîtrise du contenu

Nous avons choisi le texte fabriqué, comme point d'entrée vers la recherche de textes, car le consensus autour du texte authentique peut faire penser que la décision d'utiliser un autre type de texte est une décision consciente et argumentée. Nous avons donc proposé un questionnement en 3 temps⁵² :

- proportion d'utilisation des textes authentiques (l'enseignant se positionne par rapport à son usage du texte authentique), cf. question A.2.1 p. a-5 ;
- le mode de recueil des textes fabriqués (faut-il également permettre l'ajout de textes fabriqués à un système comme celui sur lequel nous travaillons ?), cf. question A.2.2 ;
- cas où le texte fabriqué est plus adéquat que le texte authentique, cf. question A.2.3.

a) Erreur terminologique

Dans le questionnaire, nous opposons texte authentique et texte pédagogique en proposant aux enseignants les éclaircissements suivants :

Un texte authentique est un texte écrit dans sa langue par un locuteur natif dans un but purement communicatif ; un tel texte peut bien entendu être utilisé à des fins pédagogiques mais il n'aura pas été écrit dans cette perspective. Au contraire, un texte pédagogique aura été écrit par un locuteur natif ou non, avec comme objectif d'être utilisé dans une activité pédagogique. Cela ne signifie pas qu'un texte pédagogique n'aura pas de fonction communicationnelle mais que lors de son écriture, il était destiné à une utilisation dans le contexte de l'apprentissage de la langue dans laquelle il est écrit.

Cependant, en revenant plus en détail sur la notion de texte authentique, nous nous sommes aperçu que notre définition de texte pédagogique était erronée : « [l'appellation *texte pédagogique* désigne] les instructions, les programmes ou les manuels destinés à l'enseignement » [Bronckart & Plazaola Giger, 1996, p. 16]. La critique historique des manuels comme contenant un matériel linguistique fabriqué est de nature à entretenir ce type de confusion : « Le texte de manuel typique n'est pas un texte authentique, mais un texte construit pour emballer la grammaire le plus habilement possible. » [Krüger, 1981, p. 34]⁵³ S'il est à noter que les enseignants qui ont

52. L'intégralité des réponses est disponible en annexe (Annexe C p. a-13).

53. Citation tirée de [Carol *et al.*, 2004, p. 3].

répondu à notre questionnaire n'ont pas eu l'air de s'étonner de cette formulation, il n'est pas impossible que cela soit le fait de l'aide proposée pour les parcours B et C²⁷ : le seul témoignage d'un enseignant qui indique ne pas avoir compris les questions parlant de « textes pédagogiques » est issu de l'échantillon He1o ([1]), qui n'y avait pas accès.

Dans la suite de ce paragraphe, nous utiliserons donc le terme consacré de texte fabriqué, même si, comme nous l'allons voir, il ne couvre pas tous les cas de figure des réponses des enseignants.

b) La recherche de textes omniprésente

Les deux premières questions sont fermées, et la troisième (question ouverte) doit nous permettre de mieux les comprendre. Le tableau 4.7 regroupe les réponses à la question concernant la fréquence avec laquelle les enseignants utilisent des textes authentiques dans le cadre de leurs cours. L'absence d'enseignant n'y ayant jamais recours est à notre avis un bon indicateur de ce consensus autour du texte authentique, qui est encore accentué par le fait que 80% des enseignants utilisent « toujours » ou « le plus souvent » des textes authentiques.

FRÉQUENCE D'USAGE DES TEXTES AUTHENTIQUES	Pourcentage de l'effectif total		Parcours ²⁷ (Effectifs)			
	TOTAL (%)		TOTAL	A	B	C
Toujours		30%	39	37	1	1
Le plus souvent		50%	65	50	3	12
Parfois		20%	26	24	0	2
Jamais		0%	0	0	0	0
TOTAL			130	111	4	15

TABLE 4.7: Réponses à la question A.2.1 p. a-5 (Usage des textes authentiques)

Par définition, le texte authentique demande une recherche : il n'a pas été écrit en vue d'être utilisé dans l'enseignement des langues. Le fait qu'un texte dit authentique soit utilisé dans le contexte de la classe est forcément le produit d'une séquence de choix. Que ces choix concernent la source à considérer (manuel, recueil de textes personnel, quotidien, magazine, etc.) ou qu'ils permettent de choisir un texte par rapport à d'autres, c'est cette série de choix qui aboutit à détourner le texte de son objectif initial pour en faire un support pédagogique. Présenté sous cet angle, un enseignant n'a que deux solutions pour s'abstraire du problème de la recherche de textes :

- ne pas avoir recours à des textes ;
- n'utiliser que des textes écrits par ses propres soins⁵⁴.

Le premier cas est tout à fait plausible dans la mesure où certaines méthodes d'enseignement n'ont quasiment pas recours à des textes (par exemple la méthode orale, cf. tableau 1.1). Toutefois, d'après l'étude de Levy et la nôtre, la plupart des enseignants se revendiquent de méthodes ayant recours au texte. Parmi les commentaires qui nous ont été faits, aucun des quatre qui parlent de l'objet même de notre étude ne nie l'usage des textes :

- « je n'utilise pas trop de textes sauf pour les niveaux plus avancés et pour les cours sur des sujets spécifiques (commercial ou tourisme) » [13].
- « les textes ne sont qu'une partie de mes supports. Les autres ont tout autant d'importance : labo multimédia, supports audio et video... » [24].

54. Une autre alternative serait éventuellement de suivre uniquement un manuel donné, mais même dans ce cas précis, comment le manuel est-il choisi ?

- « Dommage que les questions portent sur les textes car nous utilisons de nombreux documents autres (ex : pub, CD, DVD...) [...] D'autre part, les critères de recherche d'un texte me paraissent peu nombreux et très généraux. Ce serait intéressant d'apporter des précisions supplémentaires. » [69].
- « Les textes allemands perdent vite leur intérêt, vu le contexte socio-économique et vu l'évolution de nos élèves. Il vaudrait mieux indexer des tâches à propos de textes (des "objets pédagogiques") » [128].

Pour résumer, un enseignant nous signale n'utiliser des textes qu'avec les niveaux les plus avancés (qui font donc partie de sa problématique d'enseignement). Deux enseignants nous reprochent de nous focaliser sur les textes, qui ne sont qu'un support parmi d'autres, mais ils ne nient pas l'usage de textes. Enfin le dernier témoignage nous suggère que, dans sa pratique, les textes n'ont pas une « longévité pédagogique » qui justifie leur conservation. Ce point de vue modéré par la langue enseignée, concerne particulièrement les pratiques de cet enseignant puisque comme nous allons le voir, plus de 55% des enseignants interrogés déclarent regrouper leurs textes dans un recueil personnel⁵⁵. Si tant est que le caractère éphémère des textes dans le contexte de l'enseignement des langues fût une vérité inaltérable, ce commentaire, bien que pertinent, concerne finalement une problématique qui n'exclut pas la nôtre. Si les textes perdent de leur intérêt :

- il faut en chercher plus régulièrement ;
- les textes originalement prévus pour une tâche donnée risquent de perdre de leur intérêt aux yeux de l'enseignant qui se réappropriera cette dernière. Il ou elle devra donc rechercher des textes qui lui conviennent pour les intégrer dans la séquence en question.

Nous pouvons donc conclure que, même s'il n'est pas impossible d'enseigner les langues sans avoir recours à des textes, très peu d'enseignants s'en passent effectivement. La seconde possibilité pour s'extraire de la problématique de la recherche de textes est de rédiger tout les textes utilisés en classe (cf. tableau 4.8).

Cette question du mode d'obtention des textes fabriqués (A.2.2, p. a-5) fait partie des imperfections de ce questionnaire. Elle est posée à tous les enseignants, y compris ceux qui ont déclaré toujours utiliser des textes authentiques. Certains enseignants nous font donc la remarque de cette incohérence dans la rubrique « commentaires » ou dans les réponses à la question suivante qui demandait dans quelles situations ils pouvaient être amenés à privilégier un texte fabriqué. La plupart des témoignages montrent une certaine cohérence dans la réponse à cette question, allant du presque contrit « Je n'ai pas encore utilisé de texte pédagogique » [43] au franc agacement « aucun!!!!!!!!!! » [29]. Certains enseignants, qui déclaraient toujours utiliser des textes authentiques, donnent malgré tout des exemples de situation, la plupart du temps très précises, dans lesquelles ils peuvent avoir recours à des textes fabriqués.

Cependant, l'objectif était de tenter d'évaluer la quantité d'enseignants qui pouvaient se passer complètement de toute recherche de textes dans notre échantillon. Nous ne représentons donc, dans le tableau 4.8, que les réponses des enseignants qui déclarent avoir parfois recours aux textes fabriqués. Nous avons aussi écarté le témoignage [1] qui indiquait ne pas avoir compris la question qui était posée.

Seuls 3 enseignants sur les 90 qui utilisent parfois des textes fabriqués ne font que les composer eux-mêmes⁵⁶. Les 3 enseignants qui déclarent écrire eux-mêmes les textes fabriqués qu'ils utilisent ne sont que des enseignants qui déclarent utiliser « le plus souvent » des textes authentiques. Nos données font état de 4 autres enseignants qui écrivent tous les textes fabriqués, ils ne sont

55. Cf. tableau 4.9 p. 173.

56. Ramené à l'ensemble de notre échantillon, cela ne représente plus que 2,3%.

MODE D'OBTENTION DES TEXTES FABRIQUÉS	Pourcentage de l'effectif total		Parcours ²⁷ (Effectifs)			
	TOTAL (%)		TOTAL	A	B	C
écriture	<input type="checkbox"/>	3.3%	3	3	0	0
recherche	<input type="checkbox"/>	48.9%	44	33	1	10
recherche ou écriture	<input type="checkbox"/>	47.8%	43	38	2	3
TOTAL			90	74	3	13

TABLE 4.8: Réponses à la question A.2.2 p. a-5 (Obtention des textes fabriqués)

pas inclus ici, puisqu'ils déclaraient en réponse à la question A.2.1 n'utiliser que des textes authentiques.

Finalement, nos données suggèrent que :

- le recours au texte authentique est effectivement très répandu chez les enseignants de langue ;
- les enseignants qui utilisent le plus fréquemment des textes fabriqués n'en sont pas systématiquement les auteurs.

Tous les enseignants de notre échantillon recherchent donc parfois des textes, confirmant ainsi notre hypothèse H_a (p. 138).

c) Contexte d'utilisation des textes fabriqués

L'hypothèse H_a est finalement un prétexte à une compréhension plus poussée de la recherche de textes, plus qu'un résultat en soi. Les pratiques déclarées des enseignants placent, conformément à la théorie, le texte authentique au centre de leurs préoccupations (cf. § 2.3.1 p. 64). Mais pour notre problématique, il semble plus important d'aller plus avant dans le processus de recherche de textes. Comme semblent le montrer les réponses à la question A.2.1, le choix d'avoir recours à des textes fabriqués est conscient. Nous allons donc nous centrer sur les réponses à la question A.2.3, pour tenter de comprendre ces raisons.

Les enseignants ont quasiment tous répondu à cette question. Les seuls à l'ignorer indiquaient toujours utiliser des textes authentiques⁵⁷. Cela tend à confirmer que le recours à un texte fabriqué est un choix consciemment argumenté. Ensuite, nous avons extrait de ces témoignages des catégories qui permettent de couvrir la grande majorité d'entre eux.

Contraintes externes à la classe : Mais avant d'entrer dans le détail de ces catégories, il convient de noter que le recours au texte fabriqué est avant tout un choix par défaut qui intervient faute d'avoir trouvé un texte authentique satisfaisant ou du fait de contraintes extérieures qui ne sont pas du ressort de l'enseignant : « pour faire plaisir à l'institution et aux parents, il faut "faire" le livre » [86]. L'usage d'un manuel revient dans deux autres témoignages, cependant les raisons qui mènent les enseignants à travailler avec ces textes ne relèvent, d'après nos données, que très rarement de contraintes extérieures ou institutionnelles.

Niveau des apprenants : L'une des raisons les plus souvent invoquées concerne le niveau des apprenants, jugé trop faible pour les confronter à des textes authentiques : « au début de

⁵⁷. Certains des enseignants qui indiquaient toujours utiliser des textes authentiques ont répondu à la question. Par contre, tout ceux qui l'ignoraient déclaraient n'avoir recours qu'à des textes authentiques.

l'apprentissage d'une langue et lorsque le niveau de la classe est faible » [50], « surtout pour les débutants » [74]. Très souvent dans cette catégorie de réponses, la notion de choix par défaut est exprimée par les enseignants : « quand je n'en trouve pas du niveau des élèves » [3] ou « lorsque le texte authentique comporte trop de difficultés et que je veux le mettre à la portée des élèves » [44]. Ce dernier témoignage rappelle les remarques que nous avons pu faire concernant l'une des difficultés de la notion de texte authentique, à savoir la position des textes abrégés et adaptés, ou « didactisés » [14,30], en règle générale considérés comme fabriqués par les enseignants (cf. § a) p. 67, position de Vigner [2001]).

Traits linguistiques : Certains précisent cette problématique du niveau des apprenants en pointant les caractéristiques du texte qui peuvent poser problème. Il s'agit en général de traits linguistiques : « Pour les débutants, pour ne pas les embrouiller avec trop de lexique ou de grammaire inconnus » [117]. Cependant, en règle générale, dans les témoignages à notre disposition, il est plus question de trouver certains traits linguistiques dans le texte que d'en éviter : « lorsque les objectifs linguistiques à travailler sont une priorité et qu'ils sont plus présents dans le support didactisé » [55]. Ce type de témoignage revient très régulièrement avec des formulations différentes :

- « point de langue » [33] ;
- « point grammatical précis » [38] ;
- « pour recréer une situation vue par un [document] authentique, afin d'insister et "d'enfoncer le clou", par rapport à un objectif linguistique en particulier » [51] ;
- « lorsque je veux cibler la notion » [68] ;
- « travail sur du lexique, des faits de langue précis » [98] ;
- « pour travailler sur du vocabulaire correspondant au thème étudié et approfondir ces connaissances même grammaticales » [60].

Thème : Comme en témoigne cette dernière déclamation, la présence de certaines structures et éléments lexicaux peuvent servir l'étude de « thèmes⁵⁸ ». Ils sont évoqués de plusieurs manières : « thème », bien évidemment, mais aussi sujet [10] ou situation [20]. Les enseignants qui utilisent le terme « situation », le font toujours avec un qualificatif pour signaler que c'est une famille de thèmes en particulier qui nécessite l'usage de textes fabriqués. Il s'agit de situations « quotidiennes » [19,20] ou « concrètes » [52,53] ; dans les autres cas les enseignants ne font référence qu'au thème, sans plus de précision quant à sa nature. Enfin, le thème est parfois croisé avec d'autres critères, comme nous l'avons vu avec le témoignage [60] : « adéquation exacte entre le thème du manuel et le texte que je veux faire étudier à mes élèves (dans le cadre d'une évaluation par exemple) » [96]. Ce témoignage replace l'usage d'un texte fabriqué dans un contexte lié à un type d'activité en particulier, ici une évaluation.

Type d'activité : Cela nous mène à la dernière classe de critères. De nombreux témoignages associent le recours à un texte fabriqué à certains types d'activités⁵⁹ :

- évaluation [81,96,112,133] ;
- exercices [106,119,123] ;
- révisions, bilan [36,63,91] ;
- réemploi et fixation [45,67] ;

58. Nous revenons en 4.4.2 p. 173 sur la difficulté de traiter certains témoignages invoquant les thèmes.

59. La liste de témoignages n'est pas exhaustive.

- présentation / explication d'une notion [102,108] ;
- mise au point de difficultés récurrentes [63].

Ces activités sont souvent en rapport avec des traits linguistiques : « pour expliquer un point grammatical » [108]. Parfois, la forme de l'activité n'est pas indiquée, seul son objectif est donné : « activités grammaticales » [128].

d) Maîtrise du contenu

L'étude des réponses à la séquence de questions A.2.1 à A.2.3 (p. a-5), nous ont permis de montrer que la recherche de textes fait partie intégrante de la tâche d'enseignement des langues. Nous avons pu constater que, conformément à nos attentes, le recours au texte fabriqué est un choix conscient que les enseignants sont à même de justifier (la quasi intégralité des témoignages sont exploitables).

L'examen des réponses des enseignants, nous a permis d'isoler 5 facteurs qui peuvent décider de leur usage d'un texte fabriqué :

- des contraintes externes à la classe (relevant du cadre de l'enseignement) ;
- le niveau des apprenants, qui, chez les débutants, est de nature à nuire à la possibilité de comprendre un texte authentique ;
- la recherche de traits linguistiques particuliers ;
- la recherche d'un texte correspondant à un thème en particulier ;
- le type d'activité dans laquelle le texte sera utilisé.

Tous les critères, y compris le premier, bien qu'intervenant moins directement, ont trait à la situation d'apprentissage dans laquelle sera utilisée le texte : cadre de l'enseignement, public, objectif de la leçon (traits linguistiques à travailler, thème) et moyen de l'aborder (type d'activité, thème). Nous incluons intentionnellement le thème dans les deux catégories. En effet, comme le suggère van Ek & Trim [1991] (qui est repris dans le CECR [Conseil de l'Europe, 2000, p. 45]), le thème peut influencer les « traits linguistiques » étudiés. En fonction du thème, une série de « notions spécifiques » est suggérée à l'enseignant ou au concepteur de matériel pédagogique [van Ek & Trim, 1991, p. 64]. Ces notions spécifiques sont des éléments lexicaux ou des structures grammaticales. Mais il est tout à fait imaginable, même si cela semble constituer un détournement de l'approche communicative, de partir d'une notion et de choisir un thème pour l'aborder. Ainsi, les notions spécifiques du travail issues du sous-thème « 1.10 *occupation* » (du thème « 1 *Personal identification* »⁶⁰) pourraient tout à fait être traitées dans le cadre du sous-thème « 3.2 *at work* » (du thème « 3 *Daily life*⁶¹ ») [van Ek & Trim, 1991, p. 67, 72].

Toutes les raisons évoquées permettent de décrire la situation d'enseignement. Elles peuvent, comme nous l'avons remarqué en décrivant les standards de description de ressources pédagogiques, avoir une influence les unes sur les autres. Bien sûr, chaque témoignage ne les décrit pas avec les mêmes termes ou avec les mêmes composantes. Cependant, ils tendent vers des recherches qui ont pour objectif une activité en particulier, même si cette dernière est rarement décrite intégralement dans les raisons évoquées. Ce qui n'est pas surprenant étant donnée la question qui est posée : ils donnent les critères qui, d'après leur expérience d'enseignement, influencent la décision d'avoir recours à des textes fabriqués (une propriété intrinsèque du texte selon la définition que nous avons donnée de l'authenticité, cf. § a) p. 65).

Dans certains témoignages, moins génériques et plus explicatifs dans leur formulation, la re-

60. profession / identité

61. au travail / vie quotidienne

cherche avec un objectif précis d'intégration dans une séquence est tout à fait claire : « Quand je ne trouve aucun texte authentique correspondant à la fois au thème central de la séquence et au point de langue à aborder » [102]. Ce témoignage met en exergue le fait que le texte fabriqué est la plupart du temps un choix par défaut. Le seul témoignage qui irait clairement à l'encontre de ce choix par défaut serait le [41], qui exprime un choix positif du texte fabriqué, tout en utilisant un modérateur, montrant que ce n'est pas systématique : « Je ne le préfère pas vraiment, mais il m'arrive d'utiliser un texte pédagogique si je tombe sur un qui convient bien à l'activité ou au thème que je souhaite traiter avec mes élèves » [41]. Il est à noter que ce témoignage donne également une place relativement précise au texte dans la séquence d'enseignement.

D'une certaine façon, il dit qu'un texte peut être plus ou moins adapté à une séquence, mais ne donne pas d'indice précis de la manière dont cette adéquation s'exprime. Le témoignage [136] donne une piste : « En cas d'objectif grammatical. Les occurrences sont généralement en nombre suffisant. Mais il est vrai que je pourrais les trouver dans des textes authentiques » [136]. C'est une piste très circonstanciée, qui ne s'applique que dans le cadre d'une activité à objectif grammatical. Mais c'est sa non-universalité qui en fait toute sa précision. L'enseignant explique que dans le cadre d'une telle activité, le nombre d'occurrences de la notion qui constitue l'objectif de la leçon est un critère pertinent. Le choix du texte fabriqué est encore une fois un choix par défaut, mais qui facilite la tâche de l'enseignant dans la mesure où, dans un tel texte, l'enseignant a beaucoup plus de maîtrise sur le contenu du texte, ce qui est spécialement pertinent quand le critère est d'une telle précision : le nombre d'occurrences d'une structure donnée dans un texte.

C'est avec un certain à-propos que le dernier témoignage enregistré dans le cadre de ce questionnaire résume, malgré sa précision, la plus grande partie des réponses faites à cette série de questions : le recours à un texte fabriqué est en général un choix par défaut, qui provient d'une nécessité de maîtriser son contenu en fonction d'une situation d'enseignement précise. En effet, quand il s'agit du niveau des apprenants, la maîtrise porte sur la difficulté du texte ou la quantité de structures inconnues. Elle peut porter sur le nombre d'occurrences d'une structure donnée ou de mots de vocabulaire (cf. traits linguistiques / vraisemblablement selon le type d'activité) ou sur le sujet du texte (cf. thème).

4.4.2 Recueils de textes et recherche passive

Ayant assimilé le recours à des textes fabriqués à une nécessité de contrôle du contenu à des granularités plus ou moins fines, nous poursuivons le parcours de notre questionnaire et allons examiner les réponses aux questions A.2.4 à b) (p. a-6) relatives à la constitution de recueils de textes.

a) Possession et organisation

POSSESSION D'UN RECUEIL	ORGANISATION DU RECUEIL		
	Possesseurs de recueil		Total
oui <input type="checkbox"/> 57,7%	oui <input type="checkbox"/> 60%		34,6%
	non <input type="checkbox"/> 40%		23,1%
non <input type="checkbox"/> 42,3%			

TABLE 4.9: Possession et organisation d'un recueil de textes (questions A.2.4 et a) p. a-6)

La création d'un recueil de textes, sans être une pratique généralisée, est répandue, puisque près de 60% déclarent en posséder un ; 60% de ces enseignants ont d'ailleurs développé un système pour organiser ces textes (soit 34,6% de la population totale).

En observant plus précisément le mode de classement des enseignants, nous nous sommes rendu compte que pour pouvoir réellement tirer parti de ces réponses, il nous aurait fallu des précisions supplémentaires : les témoignages, moins nombreux que dans le cas des textes fabriqués, ne permettent pas toujours de se faire une idée précise de ce qu'entend l'enseignant par les catégories qu'il donne. Par exemple, un témoignage comme « par objectifs visés » [58], n'est pas catégorisable puisque nous n'avons pas d'exemple de ce que peuvent être les objectifs de l'enseignant et encore moins des conséquences qu'ils peuvent avoir sur le classement des textes. S'agit-il d'objectifs grammaticaux, civilisationnels, communicatifs ? Et ce n'est pas le seul témoignage posant problème : à partir du moment où nous avons recueilli le témoignage suivant – « par thèmes, soit grammaticaux, soit de civilisation » [117] – le classement par thème, sans autre explication, devient ambigu. Nous avons fini par décider de laisser de côté cette question (temporairement, cf. § 5.2 p. 220).

b) Lien avec la recherche passive

Sans revenir sur nos hypothèses de départ, qui seront traitées après avoir fait un tour d'horizon des réponses, nous avons tenté de croiser la possession de recueil avec d'autres pratiques déclarées. Nous avons retenu celles qui nous semblaient les plus susceptibles d'avoir un effet, à savoir la fréquence d'utilisation de textes authentiques (cf. question A.2.1 p. a-5), les provenances des textes utilisés en cours (cf. question A.2.5 p. a-6) et enfin la possibilité de trouver des textes « par hasard » (cf. question A.2.7 p. a-6). Aucun des tests que nous avons effectués sur les deux premiers critères ne s'est avéré concluant⁶². En revanche celui concernant la question A.2.7 avec la possession d'un recueil a donné certains résultats. Nous récapitulons les effectifs des groupes dans le tableau 4.10.

		HASARD		TOTAL
		oui	non	
RECUEIL	oui	75	0	75
	non	49	6	55
TOTAL		124	6	130

TABLE 4.10: Possession d'un recueil croisée avec le fait de trouver des textes par hasard

Ce tableau permet de constater que s'il arrive à la quasi-totalité des enseignants (95,4%) de trouver par hasard un texte qu'ils réutilisent dans leur cours, aucun de ceux à qui cela n'arrive jamais ne possède de recueil. Il y aurait de ce fait une différence significative entre les réponses des enseignants qui ont un recueil de textes par rapport à ceux qui n'en utilisent pas⁶³. Cependant, le calcul tel qu'il a été effectué est contestable d'un point de vue statistique. Les conditions d'applications recommandées – moins de 20% des cases du tableau croisé avec un effectif théorique inférieur à 5⁶⁴ (et aucune avec un effectif théorique inférieur à 1) [Sheskin, 2004, p. 243] – ne

⁶². Fréquence texte authentique : $r = -0,056$; $p > 0,5$

$\chi^2(2) = 2,721$; $p > 0,25$

provenance (textes repérés au cours de lectures personnelles) : $t(128) = 0,898$; $p > 0,35$.

⁶³. $\chi^2(1) = 8,578$; $p < 0,005$.

⁶⁴. Nous connaissons au préalable le nombre de réponses à chaque question, le test portant sur la répartition.

sont pas strictement respectées. D'après Cochran [1952, p. 334], nous pouvons néanmoins nous satisfaire de la valeur « corrigée pour la continuité⁶⁵ » qui, bien que moins significative, le reste. Pour résumer, il existe une différence significative entre les réponses des possesseurs de recueils et les enseignants qui n'en utilisent pas quant à l'un des aspects de ce que nous avons appelé la « recherche passive ». Dans notre échantillon, tout enseignant qui a un recueil de textes, peut utiliser dans sa classe des textes qu'il ou elle a pu trouver par hasard⁶⁶. Dans nos données, nous constatons l'absence de témoignage « recueil sans hasard » (cf. tableau 4.10). Du fait du faible nombre de témoignages répondant « non » à la question du hasard, nous ne pouvons conclure à un pré-requis. Il est tout à fait pensable que l'un des scénarios de constitution de recueil de textes commence par la constatation que des textes propices à telle ou telle activité puissent être trouvés « par hasard », créant ainsi un besoin d'archivage pour laisser à l'enseignant le temps de traiter la problématique développée en classe au moment de leur découverte. Mais d'autres scénarios sont tout aussi probables comme, par exemple, le désir de pouvoir réutiliser des textes trouvés au cours de recherches en vue d'effectuer une activité précise en classe.

4.4.3 Provenances et sources

Ces différents scénarios (qui n'ont pas de prétention d'exhaustivité) nous ramènent à d'autres questions que nous avons posées en lien avec la recherche passive :

- question A.2.5 p. a-6, qui vise à connaître la provenance des textes utilisés, en terme de processus de recherche ;
- question A.2.6 : de quel support proviennent les textes ?
- question a) : dans quelle optique les textes trouvés par hasard sont-ils gardés ?

Nous allons examiner dans cette section les deux premières. Nous différencions ici « source » et « provenance » d'un texte, sans que la différence de sens entre les deux termes soit nécessairement issue de l'usage courant. La première désigne ici le médium qui portait la ressource textuelle. La question est donc de savoir, par exemple, si le texte a été pris dans un périodique, un livre ou sur internet. Ces alternatives mélangent des entités conceptuelles présentes sur plusieurs niveaux : un périodique peut être numérique ou non. Il eût été plus judicieux de séparer ces alternatives en deux questions, l'une concernant la source proprement dite et l'autre le support. D'autre part, nous utilisons « provenance » dans le sens étymologique suivant : « ce dont un fait provient, cause⁶⁷ ». Or ici, sans processus de recherche, découverte par hasard ou imposition, le texte ne serait pas utilisé dans l'enseignement des langues, peu importe qu'il ait été publié sous forme papier ou numérique, dans un périodique ou un roman.

a) Des recherches avec un but précis

Dans la question A.2.5 p. a-6, nous demandions aux enseignants de classer trois possibilités de provenance par ordre de fréquence, en n'indiquant que celles qu'ils utilisent. La mise en forme de

Or, notre hypothèse nulle est ici une distribution équivalente des réponses à la question concernant le hasard en fonction de la possession ou non d'un recueil. Une distribution équivalente aurait 3 personnes qui ont un recueil et qui ne trouvent jamais de textes par hasard et 3 autres qui n'ont pas de recueil et ne trouvent jamais de textes par hasard. L'effectif théorique de ces deux cases (3) est inférieur à 5.

65. $\chi^2(1) = 6,279 ; p < ,05$

66. Un test de χ^2 ne fait que comparer des effectifs, il n'établit aucune relation de précedence ou de conséquence, nous aurions tout aussi bien pu écrire qu'il y avait une différence significative entre les réponses des enseignants qui ne trouvent jamais de textes par hasard et les autres, les premiers ne possédant jamais de recueil.

67. <http://www.cnrtl.fr/lexicographie/provenance>

la question dans la version papier a fait penser à certains enquêtés que les cases étaient à cocher et non à numéroter. Dix enseignants issus du parcours C²⁷ ont fait des choix non ordonnés. Ils ont aussi été comptabilisés dans la figure 4.4. Nous avons ensuite attribué une valeur à chaque choix : un premier choix reçoit 3 points, un second choix 2, un troisième choix 1. Les éléments non sélectionnés reçoivent 0 points. Enfin pour les choix non ordonnés, nous avons alloué à chaque choix la même valeur tout en nous assurant que la somme des points alloués à un témoignage soit la même que s'ils avaient été ordonnés⁶⁸. Nous calculons donc un indicateur, pour chaque proposition, qui est la moyenne de ces « notes » reçues. Les réponses sont synthétisées dans la figure 4.4, qui regroupe deux types de données : les effectifs et la valeur de l'indicateur.

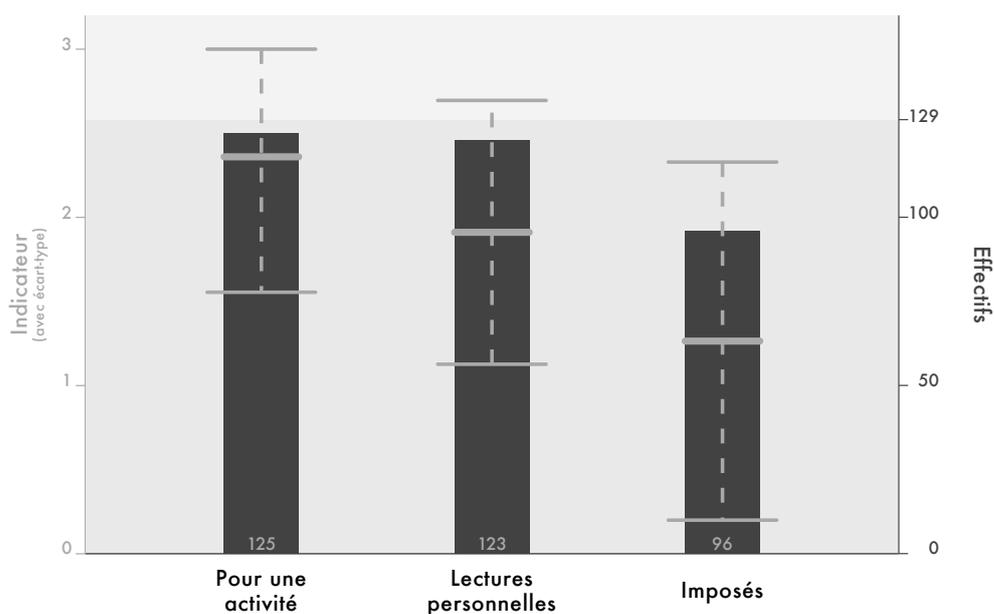


FIGURE 4.4: Comparaison des réponses à la question A.2.5 (provenances), classées par choix

L'axe vertical de gauche (clair) correspond à la moyenne de l'indicateur, dont la valeur maximale est 3 (toujours en première place) et la valeur minimale est 0 (solution jamais sélectionnée). Cette mesure dépend de l'effectif pour chaque proposition, cependant pour en avoir une vision plus claire nous les avons indiqués sous forme de barre verticale. Pour chaque proposition, nous avons donc :

- le nombre de personnes qui ont retenu ce choix (la valeur maximum est 129, un enseignant n'ayant pas répondu), sous forme de barre verticale. Les effectifs tiennent compte des choix non ordonnés ;
- la valeur moyenne de l'indicateur (trait du milieu) assortie de l'écart-type (borne supérieure et inférieure), sous forme de « moustaches ».

La proposition « imposés par le programme » sort du lot : elle est de loin la moins utilisée mais c'est aussi celle qui arrive le moins fréquemment dans les premiers choix. Et cela reste le cas, même si notre indicateur n'est calculé que parmi les enseignants qui sélectionnent ce choix, comme l'indique le tableau 4.11 page ci-contre.

68. Par exemple, si un enseignant choisit la proposition 1 et la proposition 2, nous allouons à chaque proposition 2,5 points. En effet si elles avaient été ordonnées la première aurait reçu 3 points, la seconde 2 points soit un total de 5 points pour les 2 propositions. Si un enseignant ne fait qu'un choix, il se voit de ce fait attribuer 3 points.

		EFFECTIF	INDICATEUR	ÉCART-TYPE
TOUS LES TÉMOIGNAGES	Activité	129	2,36	0,80
	Lectures perso	129	1,91	0,78
	Imposés	129	1,26	1,06
SEULEMENT SÉLECTIONNÉS	Activité	125	2,44	0,70
	Lectures perso	123	2,00	0,68
	Imposés	96	1,70	0,88
ENSEIGNANTS CHOISSANT « IMPOSÉS »	Activité	96	2,43	0,69
	Lectures perso	96	1,84	0,73
	Imposés	96	1,70	0,88

TABLE 4.11: Indicateurs en fonction du mode de comptage pour les provenances

La 2^e partie⁶⁹ du tableau contient les valeurs des indicateurs, en les calculant pour chaque choix indépendamment, en ne prenant en compte, pour chacun, que les enseignants qui les choisissent. Il s'agit de la seule section du tableau dans laquelle les effectifs varient. Nous dénombrons 96 enseignants qui comptent parmi les textes qu'ils utilisent des textes « imposés », qu'ils ne recherchent par conséquent pas. Cependant, cette source reste la moins importante, que ce soit en ne considérant pour chaque choix que les témoignages qui le sélectionnent (2^e partie du tableau) ou en ne considérant que les témoignages qui font état de textes imposés. Dans la 3^e partie du tableau 4.11, nous avons calculé les indicateurs sur le mode de la première section – les enseignants qui ne sélectionnent pas une proposition lui mettent une note de 0 – en ne considérant que les témoignages qui proposent la solution « imposés » et même dans ce cas là, cette dernière est la proposition la moins importante. Cela s'explique par le fait que cette proposition n'existe pas comme seule source de textes : 94 des 96 témoignages qui choisissent cette provenance des textes, choisissent aussi les deux autres et les deux derniers choisissent aussi la recherche pour une activité précise, qui n'est pas choisie par 4 enseignants (qui choisissent tous les quatre la solution « textes repérés au cours de lectures personnelles » comme unique provenance).

Enfin, la seule proposition « autre » qui nous soit parvenue s'intitulait « recherche continue ». Elle provenait d'un témoignage qui sélectionnait aussi la proposition des textes issus de lectures personnelles. Il doit donc s'agir d'une pratique différente de ce que nous avons appelé la « recherche passive » et qui pourrait consister en la recherche active de textes sans but précis ; un mode de recherche visant à ajouter des textes à un recueil plus que pour les utiliser directement dans la classe. Toutefois, nous n'avons pas eu la possibilité de recontacter l'enseignant concerné et toutes ces interprétations restent purement spéculatives.

Pour conclure, nous pouvons dire que les enseignants se reconnaissent dans les trois propositions (puisque même la plus marginale est sélectionnée par presque 74% des enseignants). Cependant, le fait d'ordonner les choix les met en perspective et montre que les textes issus de recherches en vue d'une activité précise sont les plus représentés chez les enseignants. Nos données s'avèrent cependant problématiques en vue de confirmer l'hypothèse H_c : elles nous permettent d'ordonner les propositions en termes d'importance, mais elles ne permettent pas de quantifier précisément les écarts entre les propositions : notre indicateur mesure les écarts de fréquence entre les différents choix et les pondère en fonction de l'importance de la source pour chaque enseignant. Toutefois,

69. Une « partie » de tableau est dans ce paragraphe, une section de 3 lignes. Le tableau 4.11 contient donc une section « tous les témoignages », une section « seulement sélectionnés » et enfin une section « Enseignants choisissant "imposés" ».

il ne fournit pas d'appréciation de la différence d'importance entre deux sources successives : si 90% des textes utilisés par un enseignant proviennent de recherches pour une activité donnée et que les 10% restants sont des textes découverts au hasard de lectures, les données seront rigoureusement identiques à celles pour un enseignant qui sélectionnerait les mêmes propositions mais dans des proportions du type 60% / 40%. Pour en revenir à l'hypothèse H_c , elle décrit une relation d'ordre et non une quantification des provenances des textes. Si nous considérons la place de la proposition « pour une activité » comme une variable, la valeur médiane de la variable est 3 pour l'indicateur (ce qui correspond au premier choix des enseignants). Le 6^e décile correspond à la valeur 2 (seconde position). Et ce n'est qu'au 9^e décile que la valeur change pour devenir 1 (choisie en troisième position par l'enseignant). La recherche pour une activité est bien la première provenance des textes d'après les enseignants.

Si nos données ne permettent pas de quantifier précisément les différences entre chaque provenance (pratiques déclarées, ordre mais pas de valeurs), elles nous permettent de montrer :

- une prépondérance des recherches en vue d'une activité précise ;
- que les textes repérés au cours de lectures personnelles, s'ils sont fréquents, le sont beaucoup moins que les premiers ;
- et enfin que si les textes imposés existent pour beaucoup d'enseignants, ils ne sont jamais l'unique provenance.

b) Source des textes : le cas des manuels

Nous avons déjà évoqué la question des sources, expliquant que notre mode de recueil de données avait diminué la représentativité des réponses à la question en termes d'importance de la source « Internet » par rapport aux autres (cf. § e) p. 142).

La question des sources a été posée sur le même mode que celle des provenances. Les mêmes erreurs ont été produites par les mêmes causes : les enseignants qui avaient fourni des choix non ordonnés à la question A.2.5 ont en général aussi fourni des réponses non ordonnées à la question A.2.6⁷⁰. Nous l'analysons dans un premier temps du point de vue qualitatif avant d'entrer dans une vision plus globale avec une analyse groupée des témoignages.

Manuels, méthodes et « autres » Les réponses « autres » les plus nombreuses (3 témoignages) proposent des « manuels » comme source. Ces trois réponses mettent en lumière une autre imprécision dans notre formulation : le terme « méthode » a été la plupart du temps compris selon le sens que nous lui avons donné dans ce contexte, cependant il aurait été plus judicieux de parler de manuel. Nous aurions pu aussi ramener la proposition « annales »^[66] à un manuel, dans la mesure où il s'agira aussi de textes identifiés comme utilisables dans l'enseignement par une institution ou des spécialistes. Cela aurait aussi pu être le cas pour la proposition « livre préparé et révisé par un groupe de professeurs du lycée »^[127], qui a aussi cette dimension de validation de l'usage du texte dans un contexte d'enseignement. Cependant, cette extension est plus délicate dans un contexte de question « semi-fermée », puisque nous voulons analyser ensuite de manière quantitative les réponses aux propositions « communes ». Or, ce rapprochement ne peut être anticipé à moins d'avoir précisément cerné les enjeux de notre travail (contrairement au rapprochement manuel-méthode).

70. Même si ce n'est pas complètement systématique, puisque 2 enseignants ordonnent leurs choix dans un cas et non dans l'autre.

D'autres formes de document textuels sont aussi proposés : courriels [14], fils de discussion de forums [13], prospectus [28], « *folletos* (documents de la vie courante) » [119]. Enfin le témoignage [31] est très détaillé, l'enseignante explique que selon le « cours » les sources diffèrent :

- traduction : « textes littéraires français (fin XIX^e, XX^e et début XXI^e siècles), des essais, biographies et articles de presse non spécialisés (pour le CAPES) » ;
- traduction spécialisée : « manuels d'utilisation, de montage, notices techniques, articles de divulgation, extraits d'ouvrages ou de presse "spécialisés" (livres de cuisine, astronomie, horlogerie, ...), essais avec terminologie spécifique à un domaine, mais aussi plaquettes publicitaires, textes administratifs, textes de normes (ISO par exemple) » ;
- « pour les cours qui concernent essentiellement un "contenu" et non un travail sur la langue tout dépend du sujet : on va du texte "littéraire" (roman, nouvelle, conte, littérature de jeunesse, ...) à l'article de presse, de l'essai au discours retranscrit (discours de Mussolini pour rhétorique par exemple), textes de lois ».

Dans ce témoignage, nous revenons à la stratégie évoquée en 4.1.1 (p. 138) : l'enseignante choisit ses sources en fonction du type de textes qu'elle pense y trouver. La source se trouve d'ailleurs parfois confondue avec le type de textes. Ce témoignage constitue malgré cela une nouvelle indication de recherche de textes pour une activité précise qui prend ici comme contrainte le public et à travers lui les objectifs pédagogiques du cours.

Point de vue global Nous présentons les réponses des enseignants concernant les sources avec les mêmes conventions que pour la figure 4.4. Elles sont synthétisées dans la figure 4.5. Nous

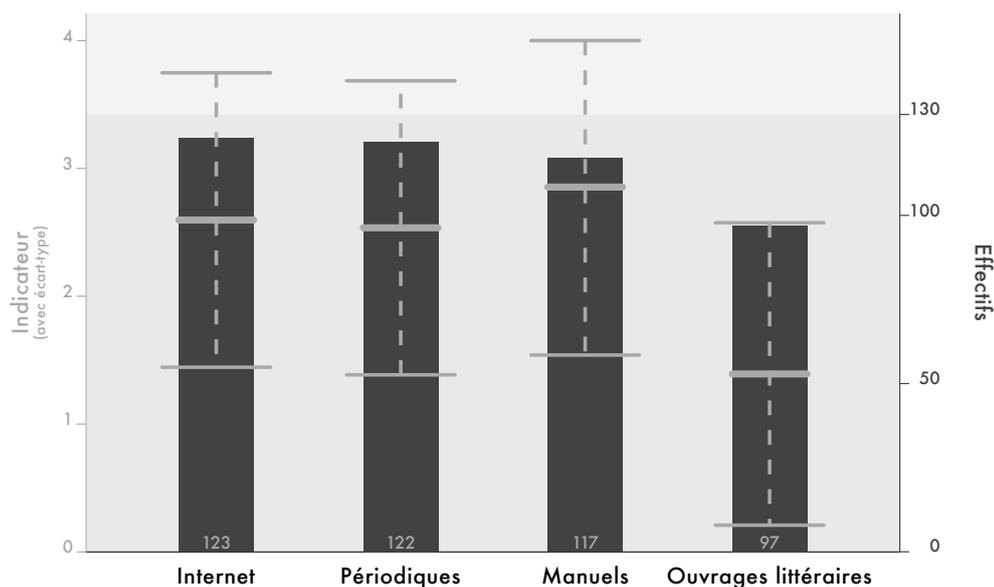


FIGURE 4.5: Comparaison des réponses à la question A.2.6 (sources), classées par choix

pouvons constater une différence majeure entre cette figure et la figure 4.4 : là où les indicateurs suivaient les effectifs quand il s'agissait de la provenance, dans le cas des sources, c'est la troisième source la plus fréquente qui a les notes les plus élevées en considérant les non-réponses comme des 0, et à plus forte raison en ne prenant en compte que les témoignages qui sélectionnent la source (cf. tableau 4.12). Cela signifie que si les sources Internet et périodiques sont les plus fréquemment invoquées, elles sont globalement moins importantes pour les enseignants que les manuels qui,

quand ils sont utilisés, constituent la première source de textes des enseignants. Malgré tout, les 3 solutions les plus retenues restent très proches selon notre indicateur. La catégorie qui sort le plus du lot est celle des ouvrages littéraires dont l'effectif est non seulement plus faible mais qui est aussi une source moins utilisée que les autres chez ceux qui y ont recours (cf. tableau 4.12).

		EFFECTIF	INDICATEUR	ÉCART-TYPE
TOUS LES TÉMOIGNAGES	Internet	130	2,60	1,15
	Périodiques	130	2,54	1,15
	Manuels	130	2,85	1,31
	Littérature	130	1,39	1,18
SEULEMENT SÉLECTIONNÉS	Internet	123	2,74	1,00
	Périodiques	122	2,70	0,98
	Manuels	117	3,17	0,95
	Littérature	97	1,87	0,99
ENSEIGNANTS CHOISSANT « LITTÉRATURE »	Internet	97	2,57	1,08
	Périodiques	97	2,53	1,06
	Manuels	97	2,90	1,30
	Littérature	97	1,87	0,99

TABLE 4.12: Indicateurs en fonction du mode de comptage pour les sources

Relation provenance-source Il est possible d'interpréter ces résultats à la lumière des éléments que nous avons déjà glanés. L'importance du manuel peut être assimilée, comme un témoignage l'indiquait, en réponse à la série de questions sur les textes fabriqués (cf. § 4.4.1 p. 167), comme le fruit d'une pression extérieure [86]. Nous avons donc testé la corrélation entre les réponses « provenance : imposé par le programme » et les réponses « source : manuels ». Pour pouvoir comparer les résultats, nous avons étendu les tests aux 3 provenances citées. Les résultats du coefficient de corrélation de Pearson sont regroupés dans le tableau 4.13 de la présente page.

VARIABLES		$n =$	r	p
Activité	Manuels	129	,001	,993
Imposés		129	,178*	,043
Lectures perso		129	-,249**	,004

*. La corrélation est significative au niveau 0,05 (bilatéral)

**. La corrélation est significative au niveau 0,01 (bilatéral)

TABLE 4.13: Corrélations entre les provenances et la source « manuel »

Tout d'abord, il convient de noter que le choix transversal de « recherche pour une activité » n'est absolument pas corrélé avec l'usage de manuels. Ensuite, une corrélation faible, mais significative, permet de dire que l'importance de la provenance « imposés » et de la source « manuels » croissent ensemble. Enfin une corrélation moyenne, et plus significative, existe entre l'importance de la provenance « textes issus de lectures personnelles » et celle de la source « manuels ». Cette corrélation négative signifie que les valeurs évoluent de manière inversée : plus la provenance « lectures personnelles » est importante et plus la source « manuels » est marginale. Inversement, plus la source « manuels » est importante et plus c'est la provenance « lectures personnelles » qui est marginale.

Ces corrélations sont observables directement dans les données (cf. annexe D) : les évolutions concordent avec les énoncés ci-dessus, même si c'est le cas de manière moins évidente pour la relation entre l'importance de la provenance « textes imposés » et celle de la source « manuels »⁷¹. Malgré tout, ces coefficients ne permettent pas de déduire des relations de cause à effet. De plus, l'indicateur que nous avons choisi et surtout le mode de recueil de données ne permettent pas une analyse fine du phénomène qui nécessiterait une observation en situation et une étude des provenances effectives des textes. Nous pouvons cependant conclure que les enseignants qui dépendent le plus de textes issus de leurs lectures personnelles sont aussi ceux qui utilisent le moins les manuels et qui dépendent le moins de textes imposés (cf. tableau 4.11 p. 177).

c) Tendances et cas général

En observant les provenances et sources de textes, nous avons pu constater que les recherches de textes en vue d'une activité précise étaient prépondérantes au sein des pratiques déclarées des enseignants. Toutefois, les deux autres provenances ont elles aussi été reconnues par les enseignants⁷², même si elles l'ont été avec un degré moindre d'importance. Les lectures personnelles devaient désigner la recherche passive et à ce titre les réponses sont cohérentes avec celles de la question 4.10, qui concernait l'éventualité de découvrir un texte au hasard et de le garder pour sa classe. Comme l'indique le tableau 4.14, les enseignants qui déclarent trouver des textes par hasard donnent une importance significativement supérieure à la provenance « lectures personnelles »⁷³.

		IMPORTANCE DES LECTURES PERSONNELLES					STATISTIQUES DE GROUPE		
		0	1	2	NO*	3	Total	Moyenne	Écart-type
HASARD	non	1	4	1	0	0	6	1	0,632
	oui	6	24	64	3	27	124	1,94	0,782

*. Réponse non ordonnée, cf. § a) p. 175

TABLE 4.14: Importance des lectures personnelles comme provenance croisée avec le fait de trouver des textes par hasard

Enfin la dernière provenance, moins importante mais revendiquée par plus de 70% des enseignants, concerne les textes qui leurs sont imposés. Ces derniers ne les recherchent par conséquent pas réellement. Cependant, l'écriture (fabriqués) ou la recherche (authentiques ou fabriqués) de ces textes imposés a nécessairement été effectuée en amont par d'autres enseignants ou des didacticiens. Par conséquent, ils partagent tous une même caractéristique : ce sont des textes dont l'usage dans un contexte d'enseignement des langues est avéré (ou à défaut prôné par des pairs). C'est une dimension qui se retrouve dans la source « manuels », dont l'importance est corrélée positivement avec celle des textes imposés. Au contraire, elle est corrélée négativement avec celle des textes issus de lectures personnelles. Ces éléments tendraient à indiquer deux tendances :

- d'une part, la recherche passive : l'enseignant, qui possède souvent un recueil, trouve fréquemment des textes au cours de ses lectures personnelles ;

71. La note maximum pour les manuels est obtenue pour les enseignants qui choisissent en deuxième position la provenance « textes imposés » (voir pp. a-25–a-26).

72. 94,6% de la population pour les « lectures personnelles » et 73,8% pour les « textes imposés ».

73. Test t d'échantillons indépendants : $t(128) = -2,892$; $p < ,005$.

- d'autre part, une recherche de textes dont l'usage dans l'enseignement des langues est avéré, qui dépend plus volontiers d'un manuel et qui s'assortit plus fréquemment de textes imposés. Cependant, les corrélations sont respectivement faible et moyenne, ces tendances ne constituent pas des groupes de pratiques mutuellement exclusifs⁷⁴. Les manuels constituent d'ailleurs la source la plus importante de textes chez les enseignants qui ne se voient jamais imposer de textes⁷⁵. Il ne faut donc pas ériger en comportement stéréotypique ces tendances, et ce à d'autant plus forte raison que ce serait ignorer les réponses transversales, les plus fréquentes, qui se retrouvent dans la quasi-intégralité des témoignages. La recherche de corrélations ne doit pas faire oublier que 96,2% des enseignants recherchent des textes pour les utiliser dans une activité précise (cf. figure 4.4 p. 176), qu'ils sont plus nombreux à trouver des textes sur Internet (première source en effectif) et dans des périodiques que dans des manuels. L'importance que revêtent ces derniers par rapport aux autres sources pourrait provenir d'une quête de contrôle sur le contenu des textes (cf. § 4.4.1 p. 167) ou, à défaut de contrôle, d'une recherche de textes dont l'usage dans l'enseignement des langues est avéré.

4.4.4 Optique de conservation d'un texte et critères de recherche

Alors que l'objectif pédagogique est le premier moteur des recherches de textes, nous ignorons si les textes retenus au hasard d'une lecture, le sont dans un but précis.

a) Des critères de conservation parfois flous

La question a) avait plusieurs objectifs (cf. tableau 4.1 p. 148) qui n'auraient pas dû cohabiter au sein de la même question (cf. § 4.1.7 p. 145). Nous nous focaliserons donc sur les réponses ayant trait à la « recherche passive », laissant de côté la proposition « Si le texte est bon il sera utilisable dans de nombreux contextes radicalement différents ».

Dans un souci de commodité, nous parlerons d'« optiques » (de conservation d'un texte) pour désigner chacune des propositions.

Le tableau 4.15 recense les 124 réponses recueillies, conformément au nombre d'enseignants ayant déclaré conserver parfois des textes découverts par hasard (cf. tableau 4.10 p. 174). Parmi ces 124 témoignages, 50 ont choisi la proposition impromptue (et 8 n'ont choisi que cette dernière). Nous verrons dans la section 4.4.5 (p. 187) que les résultats sont tout à fait différents quand la question de la réutilisation des textes est posée clairement dans un contexte approprié.

La pénultième proposition en termes de fréquence de choix concerne la conservation du texte dans l'optique de l'utiliser avec un public précis. Cette proposition est choisie par 9 enseignants de moins que « pour une activité précise », alors que la réponse la plus fréquemment retenue est « en attendant de trouver une manière adéquate de l'utiliser ». Cette dernière optique jouit d'une popularité bien supérieure aux autres : son effectif est supérieur de plus de 30% à celui la conservation d'un texte pour une activité précise (alors que 119 des enseignants qui ont répondu à cette question indiquent rechercher des textes pour une activité précise⁷⁶). L'intérêt de cette

74. Par exemple, la provenance « lectures personnelles » reste plus importante, selon notre indicateur, que la provenance « textes imposés », même en limitant l'échantillon aux enseignants qui choisissent cette dernière.

75. Importance moyenne : 2,65 / Écart-type : 1,37. Dans ce sous-échantillon, toutes les sources ont le même effectif sauf les textes littéraires. Les manuels ont la plus grande importance moyenne, même si elle est plus faible que celle pour la population totale (cf. tableau 4.12 p. 180).

76. importance moyenne = 2,34 ; écart-type = 0,842.

4.4. Vers une description du processus de recherche de textes : les réponses des enseignants 183

« OPTIQUE DANS LAQUELLE LE TEXTE EST GARDÉ »	Pourcentage de l'effectif total		Parcours ²⁷ (Effectifs)			
		TOTAL (%)	TOTAL	A	B	C
Pour une activité précise		49.2%	61	49	4	8
Pour un public précis		41.9%	52	44	3	5
En attendant de trouver une ma- nière adéquate de l'utiliser		66.1%	82	72	2	8
Si le texte est bon il sera utilisable dans de nombreux contextes radicalement différents		40.3%	50	47	1	2
TOTAUX	Optiques de conservation		245	212	10	23
	Effectifs		124	106	4	14
	Optiques de conservation de texte par enquête		1.98	2	2.5	1.64

TABLE 4.15: Réponses à la question a) p. a-7 (Optique dans laquelle les enseignants conservent les textes qu'ils trouvent par hasard)

réponse est que la question des provenances était posée dans le cas général alors qu'ici il s'agit d'une question contextualisée qui définit les pratiques dans un cas précis. Il en ressort que les deux tiers des enseignants qui trouvent des textes par hasard et les considèrent suffisamment propices à l'enseignement pour les conserver le font parfois sans savoir comment ils vont s'en servir alors que la moitié « seulement » peut associer ces textes à une activité précise et un peu plus de 4 sur 10 à un public précis.

Ce constat précise la recherche passive qui n'est pas seulement axée vers une application précise, mais qui peut être, chez la majorité des enseignants, une source de textes qu'ils conservent sans pour autant les prédestiner à une utilisation particulière. Cela nous ramène directement à notre étude préliminaire dans laquelle des enseignants faisaient explicitement référence à une forme d'intuition (cf. § 1 p. 133). Dans le cas où l'enseignant conserve le texte en attendant de trouver une manière adéquate de l'utiliser, il y a des traits qui font que ce texte lui semble propice à l'usage en classe, mais ces caractéristiques n'associent pas le texte à un public en particulier ou à une activité précise.

Le processus se déroulerait donc en deux temps pour ces textes qui sont conservés en attendant que l'enseignant trouve une manière de les utiliser. Dans un premier temps, l'enseignant repère un texte comme potentiellement utilisable dans le cadre de son enseignement. Les éléments qui font qu'il ou elle caractérise ainsi le texte pourraient nous permettre de comprendre mieux les pratiques des enseignants. Cependant, c'est en soi un problème complexe qui sort du cadre strict de notre travail : cette première phase correspondrait au remplissage de la base de textes que nous avons décidé de laisser aux enseignants. La deuxième phase consiste à trouver une manière d'utiliser le texte. Nous pouvons à ce sujet imaginer deux solutions : soit l'enseignant revient à ces textes et tente d'imaginer une séquence qui les fasse intervenir, soit quand il est à la recherche d'un texte pour une activité précise, il peut être amené à chercher parmi les textes qu'il a remisés un qui correspondrait à l'activité. La première alternative sort du cadre de nos recherches alors que la seconde s'y inscrit pleinement. Elle fait d'ailleurs partie de l'intitulé de la question A.2.9 p. a-7, qui demande aux enseignants les critères selon lesquels ils recherchent leurs textes, *même au sein de leurs recueils personnels*.

b) Des textes conservés sans recueil

Le fait de considérer les optiques dans lesquels les textes sont conservés apporte un nouvel éclairage sur le lien existant entre le fait de posséder un recueil et celui d'utiliser des textes trouvés « par hasard » (cf. § b) p. 174). En croisant la possession d'un recueil avec le fait de conserver des textes en attendant de trouver une manière adéquate de les utiliser, nous avons pu constater que 29 enseignants qui n'ont pas de recueil conservent des textes sans les prédestiner à une utilisation précise (cf. tableau 4.16).

		« EN ATTENDANT »		TOTAL
		oui	non	
RECUEIL	oui	53	22	75
	non	29	26	55
TOTAL		82	48	130

TABLE 4.16: Possession d'un recueil croisée avec le fait de conserver des textes trouvés par hasard en attendant de trouver une manière adéquate de les utiliser

Cela pose la question de ce que les enseignants considèrent comme un recueil de texte. En effet, quand les textes trouvés par hasard sont conservés pour une activité ou un public, la possession d'un recueil de texte ne paraît pas fondamentalement nécessaire (le texte ainsi trouvé peut être utilisé rapidement, sans être conservé ensuite). En revanche, la conservation en attendant de trouver une manière d'utiliser le texte, nécessite une forme de recueil, puisque le texte doit être conservé suffisamment longtemps pour être effectivement utilisé. Dès lors, en considérant cet ensemble « tampon » comme un recueil, le lien entre la possession de recueil et la pratique de la recherche passive est consolidé⁷⁷, et avec lui la description que nous faisons de la tendance correspondante (cf. § c) p. 181).

c) Recherche d'un moyen d'utiliser un texte

Parmi les réponses proposées à la question A.2.9 figure un choix que nous pourrions rapprocher de la première de ces deux alternatives : « Vous recherchez un moyen d'utiliser un texte donné plus qu'un texte pour une activité donnée ». Cependant cette solution n'est choisie que par 17% des enseignants et il n'y a pas le moindre effet entre le fait de conserver des textes en attendant de trouver un moyen de les utiliser et le fait de rechercher une manière d'utiliser un texte donné plus qu'un texte pour une activité. Les répartitions dans un tableau croisé sont exemplaires (tableau 4.17) et le test du χ^2 en atteste⁷⁸.

Ces statistiques montrent que les enseignants qui déclarent chercher un moyen d'utiliser un texte, plutôt qu'un texte pour une activité ne sont pas plus sujets à conserver un texte en attendant de trouver une manière de l'utiliser que les autres enseignants. En d'autres termes, la première des deux alternatives que nous avons énoncées sera vraisemblablement moins importante que la seconde : elle ne représente que 17% des enseignants.

77. Il n'y a plus que 20 enseignant qui n'ont pas de recueil et qui trouvent des textes par hasard (contre 104 qui ont un recueil). $\chi^2(1) = 20,192$; $p <,001$ contre $p <,005$ dans le cas précédent (cf note⁶³ p. 174).

78. $\chi^2(1) = ,003$; $p = ,954$.

		MOYEN D'UTILISER UN TEXTE		TOTAL
		oui	non	
EN ATTENDANT	oui	14	68	82
	non	7	35	42
TOTAL		21	103	124

TABLE 4.17: Conservation des textes en attendant de trouver une manière de l'utiliser croisée recherche d'un moyen d'utiliser un texte plutôt qu'un texte pour une activité

d) Des propositions cohérentes avec l'énoncé de la question

La question des critères (question A.2.9) était une question semi-fermée. Avant de traiter les réponses de manière quantitative, nous allons examiner les réponses des enseignants à la catégorie « autres ». Seuls 5 enquêtés y ont eu recours et les propositions qu'ils font constituent en général des cas particuliers des catégories proposées. C'est notamment le cas pour les propositions suivantes :

- « plaisir, intérêt culturel, intérêt des élèves » [86] / « lien avec l'actualité, une remarque des élèves ou interdisciplinarité » [63] : les propositions relèvent d'un mélange entre « l'intérêt du texte indépendamment du contexte scolaire » et « le public auquel est il est destiné » ;
- « la compétence que je veux travailler » [72] est clairement un cas particulier du « but pédagogique de l'activité dans laquelle le texte sera utilisé ».

Les trois autres cas de figures sont plus difficiles à catégoriser, ce qui met, une fois de plus, en exergue la difficulté à considérer des propriétés et même des classes de propriétés indépendamment les unes des autres dans un contexte d'enseignement :

- « Texte se rattachant au thème traité (ex : *global warming / pollution / obesity among teenagers... and so on*)⁷⁹ » [2] ;
- « En fonction du thème vu en classe » [3].

Ces deux témoignages font allusion à une sélection en fonction du thème. Quelle que soit la définition du thème (cf. § 4.4.2 p. 173), elle peut être considérée comme un choix en fonction du « but pédagogique » de l'activité qui pourrait être ramené, selon une description générique, à « travailler sur le thème T ». La formulation de la question autorise ce recodage, il est en effet demandé précisément aux enseignants de se prononcer sur le choix d'un texte. Mais la décision pédagogique de « travailler sur le thème T » peut être elle aussi être motivée par « le public auquel est destiné le texte » (cf. remarque des élèves) ou l'« intérêt en dehors du contexte scolaire » (cf. lien avec l'actualité). Comme c'est le choix du thème qui a décidé de la recherche de textes effectuée, par transitivité, il n'aurait pas été absurde de dire que dans ces cas, le texte est choisi non seulement en fonction d'un objectif pédagogique (thème T), mais aussi en fonction du public ou de l'intérêt selon ce qui a motivé le choix du thème. Il serait intéressant d'étudier le choix des thèmes par les enseignant, notamment en relation avec les choix des textes qui doivent servir de supports aux activités portant sur le thème choisi, et en particulier les influences réciproques qu'ils ont les uns sur les autres⁸⁰. Toutefois, dans le cadre de ce travail, nous devons borner notre point de vue à celui du texte, cherché dans ce cas précis selon le thème travaillé (objectif pédagogique).

79. le réchauffement de la planète / la pollution / l'obésité chez les adolescents... etc.

80. Cf. 17% d'enseignants déclarant rechercher un moyen d'utiliser un texte donné plus qu'un texte pour une activité donnée.

e) Classes de critères de recherche

CRITÈRE DE RECHERCHE	Pourcentage de l'effectif total		Parcours ²⁷ (Effectifs)			
	TOTAL (%)		TOTAL	A	B	C
Le but pédagogique de l'activité dans laquelle le texte sera utilisé	<input type="text"/>	85.4%	111	97	3	11
Le <i>modus operandi</i> de l'activité dans laquelle le texte sera utilisé	<input type="text"/>	21.5%	28	26	2	0
Le public auquel est destiné le texte	<input type="text"/>	76.9%	100	88	2	10
L'intérêt du texte indépendamment du contexte scolaire	<input type="text"/>	43.8%	57	52	2	3
Vous recherchez un moyen d'utiliser un texte donné plus qu'un texte pour une activité donnée	<input type="text"/>	16.9%	22	19	2	1
TOTAUX	Critères		318	282	11	25
	Effectifs		130	111	4	15
	Critères par enquête		2.45	2.54	2.75	1.67

TABLE 4.18: Réponses à la question A.2.9 p. a-7 (Classes de critères de recherche de texte)

L'objectif pédagogique constitue le critère le plus souvent invoqué par les enseignants pour choisir des textes. La plupart d'entre eux se dit aussi influencé par le public auquel le texte est destiné. Viennent ensuite l'intérêt du texte indépendamment du contexte scolaire (qui concerne moins de la moitié des enseignants), le *modus operandi* de l'activité (moins d'un quart des enseignants). Enfin, nous avons déjà discuté la dernière solution⁸¹ (cf. § c) p. 184).

La faible quantité de témoignages retenant le *modus operandi* de l'activité (cf. tableau 4.18) pourrait être un autre élément de l'influence des critères les uns sur les autres. En comparant les réponses qualitatives à la question concernant les situations dans lesquelles les enseignants privilégiaient des textes fabriqués (cf. annexe C p. a-13) et les réponses à cette question, nous avons pu remarquer que sur 10 enseignants qui mentionnent explicitement une forme d'activité précise comme raison de l'usage d'un texte fabriqué [71, 81, 85, 87, 90, 106, 109, 119, 123, 125], 2 seulement [71, 87] l'ont mentionnée comme critère dans la question A.2.9.

Il y a nécessairement une dimension spéculative à ce type d'explication et différentes raisons peuvent être invoquées (formulation de la question et des propositions, influence du parcours dans le questionnaire). Une explication potentielle à ce décalage est que les enseignants ont parfaitement intégré la question et répondu selon les critères qui leur semblent les plus importants. Et si le but pédagogique de l'activité influence son *modus operandi*, ils ne considèrent pas que ce soit ce dernier qui entre en ligne de compte dans leur choix de textes, mais bien leurs objectifs.

Il semblerait donc que, dans ce cadre, les enseignants considèrent les activités plus en termes de but qu'en termes de déroulement⁸².

81. recherche d'un moyen d'utiliser un texte donné plus qu'un texte pour une activité donnée.

82. Notons que la littérature indique que les enseignants « ne planifient pas par objectifs mais par activités, celles-

4.4.5 Confirmation de H_b

Pour la question précédente (critères), nous avons choisi a priori la formulation « une manière adéquate d'utiliser un texte » en opposition à « la manière adéquate d'utiliser un texte », car nous présumons qu'un texte pouvait être utilisé de plusieurs manières. Nous n'en n'avions cependant pas la confirmation à travers les déclarations des enseignants. Les réponses aux questions A.2.8 et a) ne laissent cependant aucun doute à ce sujet. La première question est une question d'opinion : les enseignants pensent-ils qu'un texte peut être utilisé dans plusieurs contextes différents. La seconde concerne en revanche directement les pratiques, elle demande à ceux qui ont répondu « oui » à la première question si cela leur est déjà arrivé. Il s'agit de savoir si la réponse à la première question est le fruit de l'expérience ou une projection. Les résultats sont présentés dans le tableau 4.19⁸³ Nous rappelons que cette question a été ajoutée en cours de passation, elle ne concerne que 112 témoignages : 111 du parcours A et un enseignant du parcours C (cf. § 4.1.3 p. 140).

TEXTE UTILISABLE DANS DIFFÉRENTS CONTEXTES	AVÉRÉE		
	Parmi les 109 « oui »		Total
oui <input type="checkbox"/> 97,3%	oui <input type="checkbox"/> 94,5%		92,0%
	non <input type="checkbox"/> 5,5%		5,4%
non <input type="checkbox"/> 2,7%			

TABLE 4.19: 112 témoignages concernant la réutilisation d'un texte dans différents contextes (questions A.2.8 et a) p. a-7)

Ces réponses confirment notre hypothèse H_b selon laquelle « un texte donné peut être utilisé avec des objectifs différents dans des conditions différentes ». Cette validation entérine le fait que les standards de description de ressources pédagogiques ne sont pas, en l'état, aptes à décrire les textes pour l'enseignement des langues (cf. § 3.5.3 p. 120). Elle est d'autant plus lourde de conséquences que les enseignants recherchent massivement les textes pour une activité précise. La difficulté de la tâche d'indexation s'en trouve augmentée : chaque texte peut être utilisé dans plusieurs contextes, pour plusieurs activités différentes, or la quasi totalité des enseignants comptent parmi les « questions qu'ils se posent quant aux textes⁸⁴ » des questions relatives aux activités dans lesquelles utiliser les textes. Il faut donc que le système puisse prendre en compte les diverses « réponses » que peut apporter chaque texte.

En vue de définir une forme d'indexation pédagogique pour l'enseignement des langues, nous allons devoir synthétiser les informations que nous avons pu recueillir sur le processus de recherche de textes, puis tenter de les mettre en perspective vis-à-vis des travaux théoriques que nous avons pu citer, en particulier du point de vue de l'opposition entre propriétés intrinsèques et extrinsèques des textes. Ceci devrait nous permettre de nous approcher d'un modèle. Mais avant de nous atteler à ces différentes tâches nous allons tenter de confirmer les hypothèses que nous énoncions dans le paragraphe 4.1.1 (p. 138) afin d'avoir toutes les données possibles sur le

ci s'adaptant mieux aux différents profils d'élèves [Shavelson & Stern, 1981] » [Tochon, 1989, p. 24]. Nous nous gardons d'apporter une conclusion définitive. Nous pouvons malgré tout nous demander si une telle constatation montre que des routines sont mises en œuvre dans la phase préactive autant que dans la phase interactive.

83. Les pourcentages sont tous calculés à partir des effectifs $((109 + 3)/112$ et $(103 + 6)/109$). Les arrondis sont à l'origine de l'erreur dans le total : $92 + 5,4 + 2,7 = 100,1$.

84. voir définition 1.3 : un langage documentaire doit permettre de décrire les objets de manière à repérer ceux qui « qui répondent aux questions [des] usagers » [Lefèvre, 2000].

processus de recherche de textes, tel qu'il nous a été décrit par les enseignants qui ont répondu à notre questionnaire.

4.5 Caractérisation du processus selon différents facteurs

À travers l'examen des réponses des enseignants, nous avons pu tester plusieurs hypothèses :

- H_a : la recherche de textes fait partie de la problématique de tous les enquêtés (cf. § b) p. 168), et donc de la majorité des enseignants ;
- H_b : un même texte peut être utilisé dans plusieurs contextes différents, ce qui a des conséquences directes sur l'indexation pédagogique de textes pour l'enseignement des langues. Selon les caractéristiques de ces contextes qui sont pertinentes pour les enseignants dans la recherche de textes, chaque objet pourra se voir attribuer plusieurs valeurs pour celles-ci (cf. § 4.4.5 p. 187). Idéalement, il faudrait que le système associe à chaque texte toutes les valeurs adéquates de chaque caractéristique. Or, nous avons montré que certaines caractéristiques avaient une influence les unes sur les autres (cf. § 3.5.3 p. 120) ;
- H_c : la première provenance des textes des enseignants est « textes issus de recherches pour une activité précise » (cf. § a) p. 175) ;
- H_d : les enseignants pratiquent couramment l'éclectisme (cf. § 4.3.4 p. 163). Avec en moyenne près de 4 méthodes employées par enseignant, parmi 19 (dont la liste aurait probablement pu être étendue), nous confirmons notre décision de ne pas articuler nos réflexions autour des méthodes ;
- H_f : l'hypothèse que le pays d'exercice n'influence pas le processus général de recherche de textes ne peut être testée (cf. § b) p. 156) ;
- H_g : l'hypothèse que le public n'influence pas le processus général de recherche de textes ne peut être testée, nous nous contenterons de tester l'influence du cadre de l'enseignement (H_{g2} p. 161) ;
- H_i : nous ne pourrions tester l'hypothèse de l'influence de la littératie informatique sur le processus, car l'échantillon est globalement homogène de ce point de vue (cf. § a) p. 150).

Nous avons pu isoler un certain nombre de faits concernant d'une part les enseignants qui ont répondu à notre enquête :

- bien que les L2 représentées soient plus diversifiées que la tendance actuelle du recrutement en France ne peut le laisser attendre, deux langues (l'anglais et l'espagnol) concentrent une majorité des témoignages (cf. § 4.3.1 p. 154) ;
- les témoignages émanent avant tout de France (cf. § b) p. 156) ;
- les enseignants sont globalement expérimentés, mais le nombre d'enseignants débutants est suffisant pour étudier l'influence de l'expérience (cf. § 4.3.3 p. 161).

Et d'autre part, leurs pratiques déclarées de recherche de textes :

- les enseignants utilisent en premier lieu des textes authentiques (cf. § b) p. 168) ;
- ils ont parfois recours à des textes fabriqués dans un souci de maîtrise du contenu de ces derniers – ce qui est tout à fait cohérent avec les situations intermédiaires de didactisation des textes – (cf. § d) p. 172) ;
- une majorité d'enseignants possède un recueil de textes. Parmi ces derniers, la plupart définit un système d'organisation (cf. tableau 4.9 p. 173) ;
- les enseignants de notre échantillon qui possèdent un recueil de textes pratiquent tous la « recherche passive » d'une certaine manière – textes trouvés par hasard – (cf. § 4.4.2 p. 174) ;
- du point de vue des sources, les manuels présentent la particularité d'être utilisés par moins

d'enseignants qu'Internet ou des périodiques tout en constituant, d'après notre indicateur, une source de textes plus prépondérante que ces derniers (cf. § a) p. 175) ;

- il existe, d'une part, une corrélation faible (positive) entre l'importance de la provenance « textes imposés » et d'autre part, une corrélation moyenne (négative) entre l'importance de la provenance « textes issus de lectures personnelles » et celle de la source « manuels » (cf. § b) p. 180) ;

- outre une tendance transversale de recherche de textes en vue de les utiliser dans une activité précise, nous avons pu isoler deux autres tendances, qui ne s'excluent pas mutuellement (cf. § c) p. 181 et § b) p. 184) :

- une forme de recherche passive, concernant principalement des enseignants qui possèdent un recueil et qui trouvent fréquemment des textes durant leurs lectures personnelles ;

- des recherches de textes dont l'usage dans l'enseignement des langues est avéré, dont les manuels (ou assimilés) constituent une source privilégiée et qui concerne fréquemment des enseignants se voyant imposer certains textes ;

- il arrive aux deux tiers des enseignants qui pratiquent la recherche passive que le texte, considéré propice à l'enseignement pour être conservé, le soit sans que l'enseignant ne sache précisément comment il s'en servira, alors que seulement la moitié des enquêtés concernés peut en conserver pour une activité ou pour un public précis (cf. § 4.4.4 p. 182) ;

- les critères de recherche les plus fréquemment invoqués sont le but pédagogique de l'activité dans laquelle le texte sera utilisé et le public auquel ce dernier est destiné (cf. § e) p. 186).

Avant de tenter d'organiser ces faits, qui à ce stade s'additionnent sans pour autant nous donner une vue générale et cohérente de la recherche de textes dans l'optique de leur usage dans l'enseignement des langues, il nous faut préciser la portée que nous pouvons leur donner. Ce n'est qu'une fois les dernières hypothèses examinées que nous entrerons dans une description générale du processus et de ses conséquences sur la recherche de textes. Les hypothèses qui restent à tester, sont les suivantes :

H_e. la langue enseignée (L2) n'a pas d'effet sur le processus général de recherche de textes ;

H_{g2}. le cadre dans lequel l'enseignant exerce n'a pas d'effet sur le processus général de recherche de textes.

H_h. l'expérience de l'enseignant n'a pas d'effet sur le processus général de recherche de textes ;

Pour confirmer ces hypothèses, nous allons reprendre les résultats qui concernent le processus général de recherche de textes – à savoir les questions A.2.4, a)⁸⁵, A.2.5⁸⁶, A.2.7 et a)⁸⁷ (cf. tableau 4.1 p. a-7) – et les examiner en séparant les réponses des enseignants selon les valeurs des facteurs. Pour nous assurer de l'homogénéité des conditions de passation, nous ne considérerons ici que les enquêtés issus du parcours A (cf. figure 4.1 p. 152).

4.5.1 Codage des données

Pour cette série de tests, nous utilisons le Λ de Wilk, le test statistique le plus utilisé pour l'analyse de variance multivariée (MANOVA) [Stevens, 2002, p. 157]. La MANOVA permet de tester l'effet de la séparation en groupes (langues, cadres, expériences) sur le processus global, que

85. Possession et organisation d'un recueil (cf. § 4.4.2 p. 173).

86. Provenances (cf. § 4.4.3 p. 175).

87. Hasard et optique pour les textes ainsi choisis (cf. tableau 4.10 p. 174 et § 4.4.4 p. 182).

nous avons considéré comme décrit par l'ensemble des réponses aux questions sus-mentionnées, recodées sous forme numérique.

Dans le cas d'une seule variable dépendante, la variance mesure sur une ligne la dispersion des valeurs autour de la moyenne. Pour une analyse concernant n variables dépendantes, la variance généralisée mesure la dispersion des données dans un espace à n dimensions [Stevens, 2002, p. 211–212]. Le Λ de Wilk est un critère inversé : plus la valeur de Λ est élevée plus l'effet est faible [Stevens, 2002, p. 212].

L'utilisation de ce test pose un problème avec nos données : les questions a) (organisation d'un recueil) et a) (optique de conservation) font suite respectivement aux questions A.2.4 (possession d'un recueil) et A.2.7 (textes trouvés au hasard) et ne sont posées aux enquêtés que s'ils avaient répondu favorablement à la première question de chaque séquence. Pour calculer une MANOVA, il faut des données numériques sans valeur manquante. Si nous conservons les données telles que nous les avons utilisées pour les analyses précédentes, le test devra ignorer les témoignages des enseignants qui déclarent ne jamais garder pour leur cours des textes trouvés au hasard (et de la même manière, ceux des enseignants qui ne déclarent pas avoir de recueil de textes). Or, une telle décision serait d'autant plus gênante, que cela reviendrait en quelque sorte à supprimer les témoignages qui nuisent à l'homogénéité des résultats pour prouver cette dernière.

Pour la question des recueils, nous pouvons grouper les résultats des deux questions sur une même échelle allant de 0 (pas de recueil) à 2 (possession d'un recueil suivant un système d'organisation). La position intermédiaire, concernant la possession d'un recueil non organisé, se voyant assignée la valeur 1.

La situation est plus complexe pour la seconde séquence de questions dans la mesure où cette dernière propose quatre choix de réponse (dont nous ne prenons en compte que trois, du fait de l'inadéquation de la quatrième⁸⁸). Sans la contrainte des valeurs manquantes, nous aurions assigné un axe⁸⁹ à chaque proposition, en assignant 0 à un témoignage n'ayant pas retenu le choix et 1 dans le cas contraire. Pour pouvoir prendre en compte les réponses des enseignants qui déclarent ne jamais utiliser en classe de textes qu'ils n'ont cherché dans cet objectif, il nous faut donc que chacune des variables associées à la question des optiques de conservation intègre la question de la conservation elle-même. L'échelle que nous utilisons, qui n'est pas l'unique solution, ajoutera la valeur -1 aux valeurs possibles pour signifier que l'enseignant n'a pas répondu à la question. Finalement, cette solution revient à considérer ces deux questions (A.2.7 et a) comme une série de trois questions de la forme suivante (l'optique de conservation étant adaptée pour chaque question) :

Vous arrive-t-il de conserver pour vos cours un texte découvert par hasard dans l'optique de l'utiliser pour une activité précise ?

- vous ne conservez jamais pour vos cours de textes découverts par hasard ;
- non ;
- oui.

L'un des problèmes de cette solution est que le choix de cette valeur va faire baisser les moyennes. Elle le fera dans les mêmes proportions pour toutes les variables dépendantes, qui resteront comparables. Du point de vue des groupes, ce codage augmente en revanche l'impact de la réponse par la négative à la question de la conservation des textes et donc la prise en compte dans les statistiques finales des témoignages, plus marginaux, allant dans ce sens.

88. « qualité intrinsèque » d'un texte, cf. § 4.4.4 p. 182.

89. Voir l'explication de la variance généralisée par Stevens [2002, p. 211–212].

4.5.2 Influence de la langue enseignée

Les effectifs au sein des enquêtés du parcours A sont indiqués dans le 4.2 p. 154. Dans cet échantillon, nous considérons de manière isolée toute langue qui représente 10% des témoignages ou plus (Allemand⁹⁰, Anglais, Espagnol, Français) et regroupons les autres sous la bannière « langues rares » (Arabe, Grec, Italien, Lingala, Néerlandais, Occitan, Portugais, Russe⁹¹). En outre, nous parlons ici de pourcentage des témoignages en prenant en compte le recodage qui a été effectué : les enquêtés qui enseignent plusieurs langues ont vu leurs témoignages doublés et ajoutés à chaque groupe concerné. En effet, n'ayant pas d'autre information sur l'importance relative des différentes langues enseignées, il nous a semblé pertinent de doubler les témoignages. La catégorie « langues rares » a nécessité que nous supprimions un témoignage par rapport aux données du 4.2. En réalité, il s'agit d'une des deux occurrences du témoignage [120] qui regroupait deux langues peu enseignées par les enquêtés : grec et italien se retrouvant dans la même catégorie, il n'était plus nécessaire de doubler le témoignage du point de vue de ces langues.

Ce découpage catégoriel pose certains problèmes étant donné que les groupes n'ont pas des effectifs homogènes : Stevens met en doute la robustesse du test quand le rapport entre la taille du plus grand groupe et du plus petit est supérieure à 1,5 [Stevens, 2002, p. 244]. Huberty et Olejnik sont plus permissifs, pour eux la limite est 2 [Huberty & Olejnik, 2006, p. 42] mais, dans un cas comme dans l'autre, nous sommes au delà de la limite (Espagnol : 46, Allemand : 11). En cas de différence substantielle dans les effectifs, ils renvoient à Hakstian *et al.* [1979]. Ces derniers expliquent en conclusion de leur article la démarche à suivre pour effectuer un test multivarié quand les effectifs diffèrent. Ils indiquent entre autres :

*« It is when $n_1 \neq n_2$ that problems arise. If the test of covariance matrix homogeneity is nonsignificant no problem exists; but if this test is significant, the user is faced with the multivariate extension of the Behrens–Fisher problem. »*⁹²

[Hakstian *et al.*, 1979, p. 1262]

Nous n'aurons donc pas à nous préoccuper de l'« extension multivariée du problème de Behrens–Fisher », puisque le test de Box (homogénéité de la matrice de covariance) appliqué à nos données n'est pas significatif⁹³.

Les résultats de la MANOVA semblent confirmer l'hypothèse H_e ⁹⁴. Ce résultat pourrait être influencé par la catégorie hybride regroupant les langues les moins fréquemment enseignées par les enquêtés. Toutefois, si nous ré-effectuons le test en excluant cette catégorie, les résultats restent cohérents (les différences entre les groupes ne sont pas significatives non plus⁹⁵).

Comme nous l'indiquions dans le paragraphe 4.3.1, nos données ne permettent pas de généraliser nos conclusions à l'enseignement de toute langue. Toutefois, dans le cadre des langues les plus enseignées en France au moins, la langue enseignée n'influencerait pas le processus général de recherche de textes que nous détaillerons ultérieurement (cf. § 4.6.4 p. 205). D'autres travaux plus centrés sur l'influence de la langue enseignée sur les pratiques (théoriques, questionnaires plus spécifiques, expérimentations, observations en situation) et à plus grande échelle (pour pou-

90. $11/121 = 9,09\%$.

91. Les langues « rares » totalisent ainsi 12,4% des témoignages.

92. C'est quand $n_1 \neq n_2$ que des problèmes surgissent. Si le test de covariance d'homogénéité de la matrice n'est pas significatif, il n'y a pas de problème; mais si le test est significatif, l'utilisateur se trouve face à l'extension multivariée du problème de Behrens–Fisher.

93. $p = ,450$.

94. $\Lambda = ,798 ; F(28, 398) = ,919 ; p = ,587$.

95. $\Lambda = ,863 ; F(21, 276) = ,692 ; p = ,841$ (test de Box : $p = ,401$).

voir intégrer plus de langues, plus différentes entre elles⁹⁶) sont nécessaires pour généraliser ces résultats. Notre point de vue ne pouvait être spécifiquement orienté sur les langues, notre problématique étant trop générale pour nous focaliser sur ce point. Toutefois, notre étude concerne des cas suffisamment nombreux et divers pour justifier le cadre, générique du point de vue de la langue, dans lequel nous évoluons à ce stade de nos recherches.

Ces conclusions ne concernent néanmoins que le processus général de recherche, il paraît tout à fait plausible que des différences entre langues apparaissent en entrant dans le cadre de critères plus précis.

4.5.3 Influence du cadre de l'enseignement

Le protocole pour le test de l'hypothèse H_{g2} , concernant le cadre dans lequel l'enseignement a lieu, reste identique. Nous allons procéder à une analyse de variance multivariée sur les mêmes questions et dont les réponses sont codées de la même manière que dans le cas des langues. La différence est que, cette fois-ci, nous utilisons comme variable indépendante et donc comme critère de groupement, le cadre. Nous avons recensé les catégories utilisées dans le tableau 4.5 (p. 159). À nouveau, nous ne considérons que les témoignages issus du parcours A. Dans un souci d'homogénéité des effectifs des groupes, nous regroupons les catégories « primaire » et « maternelle » d'une part (public jeune), et les catégories « supérieur (spécialistes) » et « centre de formation » (public adulte). Ces différents cadres possèdent des spécificités évidentes⁹⁷ et pourraient même être subdivisés⁹⁸. Mais pour ne pas ignorer complètement leurs spécificités (objectifs, organisations) par rapport aux cours pour non-spécialistes du secondaire et du supérieur, nous avons décidé de conserver ces catégories hybrides. Comme dans le cas du témoignage [120], nous devons nous assurer que les témoignages [49] et [58] ne sont pas doublés : les deux enquêtés enseignent en primaire et à l'école maternelle.

Le rapport entre la catégorie la plus représentée et la moins représentée est proche de 4⁹⁹, et donc supérieur aux recommandations évoquées dans la section précédente. Le test de Box pour l'homogénéité de la matrice de covariance n'est pas significatif¹⁰⁰. Comme pour les langues, le Λ de Wilk ne permet pas de rejeter l'hypothèse nulle du test, à savoir l'homogénéité des groupes¹⁰¹.

Encore une fois, nos conclusions peuvent être affectées par la présence de groupes hybrides. Les statistiques montrent malgré tout qu'en utilisant notre découpage en groupes, qui sont majoritairement représentatifs de la manière dont les enseignants se décrivent eux-mêmes, les différences de distribution ne sont pas significatives. Le processus de recherche de textes, tel que nous l'avons caractérisé, serait donc tout à fait comparable d'un cadre à l'autre, conformément à l'hypothèse H_{g2} .

4.5.4 Influence de l'expérience de l'enseignant

La dernier facteur à prendre en compte est l'expérience des enseignants, afin de tester l'hypothèse H_h . Dans le paragraphe 4.3.3 (p. 161), nous nous appuyions en particulier sur Huberman *et al.*

96. cf. [Robert, 2004].

97. Apprenants non lecteurs en maternelle, par exemple.

98. Centres de formations pour enfants, adultes, professionnels, migrants, etc.

99. $56/13 = 4,31$.

100. $p = ,302$.

101. $\Lambda = ,789 ; F(35, 591) = ,979 ; p = ,504$.

[1989]; Katz [1972]; Burden [1981]; Zahorik [1975], pour séparer les enseignants en deux groupes d'expérience : débutants (expérience ≤ 5) et expérimentés (expérience > 5). En ne considérant que les témoignages issus du parcours A, le rapport entre les effectifs de ces deux groupes est de 2,27. Nous effectuons donc le test de Box, pour lequel la valeur de p est faible¹⁰². Dans de telles circonstances (effectifs inégaux, violation de l'homogénéité des covariances), Neill recommande d'utiliser le critère de Pillai [Wikiversity, 2008]. D'après ce test, les différences entre échantillons ne sont pas significatives¹⁰³, même si la valeur de p est beaucoup plus faible que pour les autres facteurs testés.

Sans rigoureusement confirmer H_h , les tests effectués ne permettent pas de conclure à une différence significative entre le processus de recherche de textes chez les enseignants débutants et les enseignants confirmés.

4.5.5 Un processus général relativement stable

a) Hypothèses confirmées

Du point de vue de nos données, très génériques et manquant parfois de sujets dans les groupes les plus marginaux, les trois facteurs que nous avons décidé de tester n'auraient pas d'effet significatif sur les variables dépendantes à travers lesquelles nous avons décrit le processus global de recherche de textes. De ce point de vue, nos hypothèses H_e , H_{g2} et H_h sont confirmées.

Cependant, il faut bien prendre en compte la portée de ces conclusions. Elles concernent une quantité de variables dépendantes relativement faibles et, comme nous l'indiquons, nos données manquaient parfois de témoignages parmi les configurations les plus rares. À ce titre, la formulation originale des hypothèses avait une teneur universaliste inadaptée au processus de test de ces hypothèses. Les hypothèses, telles que nous les avons confirmées, sont donc les suivantes :

H_e' . la langue enseignée (L2), n'a pas d'effet significatif sur le processus général de recherche de textes ;

H_{g2}' . le cadre dans lequel l'enseignant exerce n'a pas d'effet significatif sur le processus général de recherche de textes ;

H_h' . l'expérience de l'enseignant n'a pas d'effet significatif sur le processus général de recherche de textes.

En outre, nous n'avons pas tenté de mesurer la taille et la puissance des effets, puisqu'ils n'étaient pas significatifs. Les données pourraient être utilisées pour préparer des études plus poussées sur les différents sujets abordés dans la mesure où une lecture plus « qualitative » des informations recueillies soulève certaines questions.

b) Questions en suspens...

Si la prise en compte des données dans leur globalité ne montre aucune différence significative, l'étude des moyennes par groupe qui, ne prenant en compte qu'une partie des données, suggère que, malgré l'homogénéité globale des pratiques, de fines différences ne sont pas exclues.

102. $p = ,003$.

103. *Pillai* : $F(14, 206) = 1,417$; $p = ,147$. La valeur de p est identique à celle qui aurait été obtenue en utilisant le Λ de Wilk (à la 3^e décimale près).

Dans le cas des cadres par exemple, certaines petites différences entre les importances des provenances mériteraient d'être creusées dans un cadre de recherche différent (toutes les comparaisons suivantes ne prennent en compte que des moyennes, elles ne prétendent pas être des vérités générales) :

- les enseignants du groupe « adultes » accordent en moyenne une moins grande importance que les autres groupes à la recherche pour une activité ;
- les enseignants qui travaillent avec des adultes (groupe « adultes » et groupe « supérieur (non-spécialistes) ») sont ceux pour qui la provenance « textes repérés au cours de lectures personnelles » est la plus importante ;
- les enseignants des groupes « enfants » et « supérieurs (non-spécialistes) » sont les moins concernés par des textes « imposés par le programme ».

Tous ces constats sont tout à fait cohérents du point de vue des stéréotypes des cadres d'enseignement. Le secondaire étant plus tributaire d'un programme que les enseignants qui travaillent avec de jeunes enfants et à l'Université avec des non-spécialistes qui évolueraient dans un cadre moins formel de ce point de vue. Les enseignants qui travaillent avec des adultes ont vraisemblablement des apprenants dont les problématiques sont plus proches des leurs et peuvent dépendre plus particulièrement de leur lectures personnelles (à moins qu'il ne s'agisse d'une question de niveau et donc de maîtrise du contenu des textes)¹⁰⁴.

Les données liées à l'expérience des enseignants sont elles aussi susceptibles de susciter des interrogations :

- en moyenne, la provenance « pour une activité » est plus importante pour les enseignants débutants ;
- inversement, la provenance « lectures personnelles » est, toujours en moyenne, plus importante pour les enseignants expérimentés.

Ces différences pourraient illustrer les différentes phases citées dans le paragraphe 4.3.3 (p. 161) :

- l'importance de la provenance « pour une activité » pourrait refléter la phase exploratoire des enseignants débutants, qui selon Lamarre [2003], s'interrogent sur les matières à enseigner, les programmes et tentent des expériences ;
- l'importance de la provenance « lectures personnelles » pourrait s'expliquer par le sentiment accru de « maîtrise pédagogique » cité par Huberman [1989a] pour la phase de stabilisation¹⁰⁵.

Toutefois, tout ceci n'est que spéculation, dans la mesure où, même si une description plus fine des phénomènes permettait de détecter des différences significatives entre les groupes, seule une étude longitudinale serait à même de dire si ces différences reflètent l'évolution des pratiques des enseignants dans leurs carrières ou l'évolution des pratiques des enseignants en opposant ceux qui sont formés aujourd'hui à ceux dont la formation a eu lieu il y a plus longtemps.

c) ...indirectement pertinentes dans notre problématique

Ces exemples de questions posées par nos données¹⁰⁶, qui n'en constituent pas une liste exhaustive, demanderaient, pour être pris en compte, des études avec un grain nettement plus fin. Elles permettraient d'évoluer à un niveau de détail dans la description des pratiques bien plus

104. Nous retrouvons, dans ces analyses, le lien entre la description du public et celle du cadre, qui en émane, cf. § 4.3.2 p. 159.

105. Cette maîtrise pédagogique pourrait être liée au développement de routines de planification (cf. [Yinger, 1979]).

106. Ce sont pour ces variables dépendantes que les différences inter-groupes sont les plus significatives... tout en ne l'étant pas ($p >, 05$).

pointu que ce qui nous importe ici. À plus long terme, ces questions peuvent resurgir. Toutefois, avant de tenter une hyperspécialisation de notre système, et donc de nouvelles études beaucoup plus précises sur chacune de ces problématiques, l'objectif était d'avoir une vision globale des pratiques. Et le fait que tous les tests que nous avons effectués ne dégagent aucune différence significative entre les groupes indique que, si des différences peuvent exister, il n'y a pas de pratique qui soit systématiquement exclue d'un groupe plutôt que d'un autre – groupes, tels que nous les avons définis. L'usage de meilleurs indicateurs prenant en compte plus de paramètres, comme les proportions d'usage des techniques, pourraient faire apparaître des différences entre les groupes, éventuellement par un affinement des groupes eux-mêmes. Mais si l'importance de chaque pratique peut fluctuer d'un groupe à l'autre, elles se retrouvent dans tous les groupes. Et c'est dans ce contexte que nous allons tenter de dresser un panorama cohérent à partir de toutes les données que nous avons pu recueillir dans ce chapitre.

4.6 Flux de la construction d'un texte comme ressource pédagogique et contexte pédagogique

Le relatif manque de données sur les pratiques de recherche et de choix des textes utilisés dans l'enseignement des langues nous a orienté vers un questionnaire prolongeant notre étude préliminaire¹⁰⁷. La forte dimension exploratoire du problème a été accompagnée d'un certain nombre de maladroresses quant à la planification du questionnaire et à la phase de recueil de données. Certaines questions n'étaient pas idéalement formulées, ce qui a eu des conséquences variables : d'un simple recodage des réponses (cf. § b) p. 178) jusqu'à une altération du questionnaire en cours de passation (cf. § 4.1.7 p. 145 ou § 4.4.4 p. 182). Les données ne nous ont pas permis d'examiner toutes les variables sous l'angle que nous avions prévu. Outre certains problèmes comme celui des L2, pas aussi diversifiées que nous l'aurions souhaité (cf. § 4.5.2 p. 191), certaines hypothèses n'ont pu être testées : H_f , H_g et H_i ¹⁰⁸. Malgré tout, les enseignants qui ont répondu étaient suffisamment nombreux et divers pour prétendre à une certaine représentativité au moins en France (cf. § c) p. 157). Le questionnaire nous a enfin permis de tirer un certain nombre de conclusions que nous avons listées dans la section 4.5 (p. 188) et de tester nos hypothèses pour arriver aux conclusions suivantes :

- H_a . l'une des tâches des enseignants de langue est la recherche de textes ;
- H_b . un texte donné peut être utilisé avec des objectifs différents dans des conditions différentes ;
- H_c . la recherche de textes se fait en premier lieu dans l'optique d'une activité concrète ;
- H_d . les enseignants pratiquent couramment l'éclectisme ;
- H_e' . la langue enseignée (L2), n'a pas d'effet significatif sur le processus général de recherche de textes ;
- H_{g2}' . le cadre dans lequel l'enseignement a lieu n'a pas d'effet significatif sur le processus général de recherche de textes ;
- H_h' . l'expérience de l'enseignant n'a pas d'effet significatif sur le processus général de recherche de textes.

107. cf. chapitre Préambule : Choix d'un mode de recueil de données p. 133.

108. H_f , cf. § b) p. 156 ; H_g , cf. § a) p. 159 ; H_i , cf. § e) p. 142.

C'est à partir de ces différents éléments que nous allons effectuer notre raisonnement visant à donner une description du processus de recherche et de choix d'un texte, qui puisse nous aider à proposer un modèle pour l'indexation pédagogique de textes pour l'enseignement des langues.

4.6.1 De l'approche instrumentale au flux social du *faire*

Quand ils introduisent l'approche instrumentale¹⁰⁹, Contamines *et al.* [2003] commencent par en décrire les fondements et expliquent qu'elle « repose sur la distinction fondamentale existante entre, d'une part, l'objet matériel et, d'autre part, l'objet matériel inscrit effectivement et efficacement dans un usage » [Contamines *et al.*, 2003, p. 162]. Bien que leur point de vue centré sur les objets soit très éloigné de celui de Holloway, cette prise en compte de la problématique de l'usage les rapprochent de sa vision du *faire* [Holloway, 2007]. Ce dernier la définit dans un contexte différent : l'action politique.

Le *faire* correspondra, pour nous, à l'action humaine, ce qui est une généralisation abusive, mais adéquate dans notre problème qui n'a pas de vellétés d'analyse de l'aliénation ou des rapports de pouvoir (cf. NB 4.3 p. 197). Holloway présente le *faire* comme une action (directement ou indirectement) sociale qui dépend du *faire* des autres, tous indissociables dans un flux global et social du *faire*. En effet, « le *faire* passé (le nôtre et celui des autres) se transforme au présent en moyen de faire » [Holloway, 2007, p. 47]. Ces moyens englobent donc l'instrument comme « objet matériel inscrit effectivement et efficacement dans un usage », qui correspondrait chez Holloway à une objectivation du *faire* subjectif des créateurs de l'instrument, dont l'existence dépend de son incorporation au flux social du *faire* par l'action d'usage qui en est faite par d'autres acteurs [Holloway, 2007, p. 50-51]. En quelque sorte la différence faite par Contamines *et al.* [2003] entre artefact et instrument pourrait se décrire chez Holloway par l'incorporation de cet objet au flux social du *faire*¹¹⁰.

Bien qu'ayant des objectifs différents, ces deux approches montrent une certaine compatibilité. Par sa nature plus proche de celle de notre travail, la vision de Contamines *et al.* est *a priori* la plus à même d'offrir un cadre qui permette d'incorporer les textes, les instruments¹¹¹ considérés ici, à une démarche sociale pour prendre en compte leurs propriétés extrinsèques, telles que définies par Recker & Wiley [2001] (cf. § a) p. 121).

« La ressource est évolutive. La genèse instrumentale est le concept permettant de rendre compte de l'aspect évolutif des instruments. Il n'y a pas de permanence de l'instrument. Une description d'une ressource ne doit donc pas être considérée comme une définition inaltérable. Au contraire, une description de la ressource ne sera toujours qu'une photographie, c'est-à-dire une coupe transversale dans l'histoire de la ressource. »

[Contamines *et al.*, 2003, p. 164]

Dans l'optique qui est la nôtre pour l'indexation pédagogique (cf. § b) p. 122), l'aspect sur lequel nous nous centrons est la description des ressources par une projection des propriétés qu'elles *peuvent* avoir pour les utilisateurs (un ensemble de propriétés qui contient celles qu'elles ont eu

109. Nous décrivons plus en détail l'approche instrumentale dans la section c) p. 61.

110. Nous y reviendrons dans la section 4.6.2 p. 198.

111. Notre système n'est dans le cadre de cette thèse qu'un artefact, puisque selon la perspective de [Depover *et al.*, 2000], nous la limitons à la production d'un premier prototype (cf. § d) p. 52); avant, donc, que ce prototype n'existe par l'usage qui en sera fait par des enseignants.

Nota Bene 4.3 Le *faire* chez Holloway

John Holloway est un économiste et philosophe marxiste qui est professeur à l'Université autonome de Puebla au Mexique. Dans Holloway [2007], « une analyse théorique et politique de ce que portent les mouvements sociaux depuis le milieu de la décennie quatre-vingt-dix - impulsés notamment par la révolte zapatiste en 1994 et par la mobilisation de Seattle en 1999 » [Wikipedia, 2009d], il introduit la notion du *faire* comme « le mouvement complet de la négation pratique, de la projection pratique au-delà du monde qui existe vers un monde radicalement différent » [Holloway, 2007, p. 46]. Il étend ensuite le *faire* à l'action humaine qui suppose une projection au delà de, dont il donne des exemples, parmi lesquels : « L'imagination du travailleur est extatique : dès le début du processus du travail, il se projette au-delà de ce qui est vers une altérité qui peut être réalisée » [Holloway, 2007, p. 47]. Cette caractérisation du *faire* selon le modèle de l'activité humaine lui permet de l'opposer ensuite au travail aliéné [Holloway, 2007, p. 61–73], alors que la notion de flux social du *faire* est, par sa rupture, le terrain de l'antagonisme entre puissance (sur) et pouvoir (de faire) [Holloway, 2007, p. 51–61].

Notons que l'importance de la projection ne se trouve pas uniquement chez les *marxistes*. D'après Clark & Yinger [1980], Herbert Simon ^a décrit dans [Simon, 1969] la planification comme la marque principale de l'activité professionnelle ^b : « *Engineering, architecture, business, law, medicine, and education are all centrally concerned with the process of design, that is, the process of devising courses of action aimed at changing existing situations into preferred ones* ^c » [Clark & Yinger, 1980, p. 14].

La tâche que nous décrivons constitue une sous-partie de la mise en place d'une séquence pédagogique. Nous allons la considérer comme un *faire* même si la projection à l'origine de l'action concernerait plutôt la tâche d'enseignement dans sa globalité. Le monde qui existe serait fait d'un groupe d'apprenants dont « l'altérité qui peut être réalisée » serait ce même groupe, fort d'une maîtrise accrue de la L2. Mais pour les besoins de notre problématique, nous réduirons la portée du *faire* à l'action de sélectionner un texte qui sera utile à la complétion d'un plan de cours ou d'une séquence d'activité.

^a. Prix Nobel d'économie et auteur de [Simon, 1997]. Il a notamment beaucoup étudié les processus de prise de décision [Wikipedia, 2009c].

^b. Et non de l'activité humaine en général. Ne connaissant pas son travail, nous ne nous hasarderons pas à tenter d'attribuer une raison à cette absence de généralisation.

^c. L'ingénierie, l'architecture, les affaires, le droit, la médecine et l'éducation sont centralement concernés par le processus de conception, c'est à dire, le processus de déterminer des plans d'action dont le but est de changer des situations existantes en des situations qui leurs sont considérées préférables.

pour d'autres utilisateurs). Cela nous écarte quelque peu de la perspective de Contamines *et al.* [2003] et de celle de Recker & Wiley [2001]. Un objectif est de pouvoir repérer des textes répondant à des questions des utilisateurs¹¹² de la forme « un texte qui *pourrait être utilisé* pour... », en plus de celles de la forme « un texte qui *a été utilisé* pour... ». Nous ne concevons pas ces deux points de vue comme antagonistes mais complémentaires. L'objectif est le même : aider l'enseignant à choisir parmi plusieurs ressources en lui en donnant une représentation. La « netteté de la photographie » de la ressource sera très probablement supérieure en décrivant l'usage effectif des textes, cependant dans l'optique de permettre à des enseignants d'en choisir un, qui en plus n'ait pas nécessairement déjà été utilisé dans la situation qui les intéresse, ce cliché n'existe pas forcément¹¹³. Pour bénéficier de plus d'opportunités de donner une fonction pédagogique pour la première fois à un objet, l'enseignant devra se contenter d'une image potentiellement moins

112. [Lefèvre, 2000], cf. définition 1.3.

113. Il n'existe que pour les ressources dont l'usage dans ce contexte est avéré par un autre enseignant.

précise, ou moins globale, mais qui peut s'avérer suffisante.

Pour parvenir à proposer ce type de d'instantané, nous ne pouvons nous abstraire de la problématique des usages : nous devons considérer les pratiques des enseignants et les mécanismes qui les régissent pour décrire les textes en fonction d'usages potentiels. La description de ces mécanismes dépend de la définition d'un cadre dans lequel l'intégrer. Notre objectif à ce stade de nos recherches est donc de proposer une représentation du processus général de recherche de textes, en vue de trouver un moyen d'y intégrer les propriétés des textes sur lesquelles nous nous appuyerons pour proposer ces instantanés de ressources. Ces derniers serviront aux utilisateurs de notre système d'accès aux textes. Ironiquement, un moyen d'arriver ultérieurement à définir ces « coupes transversales dans l'histoire des ressources¹¹⁴ » est la notion de flux social du *faire* d'Holloway. L'ironie, mais aussi la difficulté, réside dans le fait que la notion est notamment définie par Holloway pour offrir une description du changement qui ne soit pas diachronique, qui ne soit pas une comparaison de deux instantanés, mais au contraire qui permette de le comprendre comme un mouvement [Holloway, 2007, p. 49].

4.6.2 Adaptation du concept d'Holloway à notre problématique

Le concept de flux social du *faire* rend compte d'une évolution spatiale et temporelle [Holloway, 2007, p. 50], inhérente à la perception d'un mouvement. Cela rend considérablement plus difficile une représentation schématique fixe. De ce fait, la figure 4.6 page ci-contre demande un effort de représentation. Il faut l'imaginer comme une prise de vue géostationnaire d'une portion du flux social du *faire*, qui correspondrait à l'action de sélectionner un texte pour l'utiliser dans l'enseignement des langues, qu'il faut concevoir comme un mouvement au sein d'un mouvement global. Elle va nous servir de support pour notre argumentaire et sera décrite élément par élément jusqu'à la fin du chapitre (ou presque) et, à ce titre, n'a pas vocation à être comprise sans les explications qui suivent.

a) Les actions

Les actions sont à considérer dans l'espace : elles devraient être incluses dans le flux du *faire* au même titre que les textes qui en sont des objectivations. Mais si nous les avons fait figurer parmi eux, la figure aurait perdu de sa lisibilité. Elles sont donc à prendre comme positionnées au dessus du flux pour indiquer l'existence de l'action dans le flux. De ce fait, les flèches qui alimentent ces boîtes sont elles aussi des indications extérieures au flux décrivant des actions qui se passent en son sein. Elles ne vont donc pas contre le flux (ce qui n'aurait aucun sens) mais passent au dessus, seules leurs extrémités y sont intégrées. Dans la suite de cette explication, nous utiliserons donc ces méta-informations (boîtes et flèches en émanant) pour décrire ce qui se passe au sein du flux.

L'idée même de ce flux est de considérer qu'aucune action n'est fondamentalement individuelle, puisque tout *faire* dépend d'un autre *faire* antérieur. Par exemple, un texte intégré dans une activité d'enseignement des langues est l'objectivation de l'action de l'enseignant qui a planifié une activité dans laquelle figure le texte. Mais avant de voir son statut d'objet pédagogique

114. Il s'agirait plus précisément ici de coupes dans l'« histoire » de notre capacité à comprendre et modéliser les besoins des utilisateurs, qui serait mise en évidence à travers les représentations des ressources proposées en vue de modifier leur histoire.

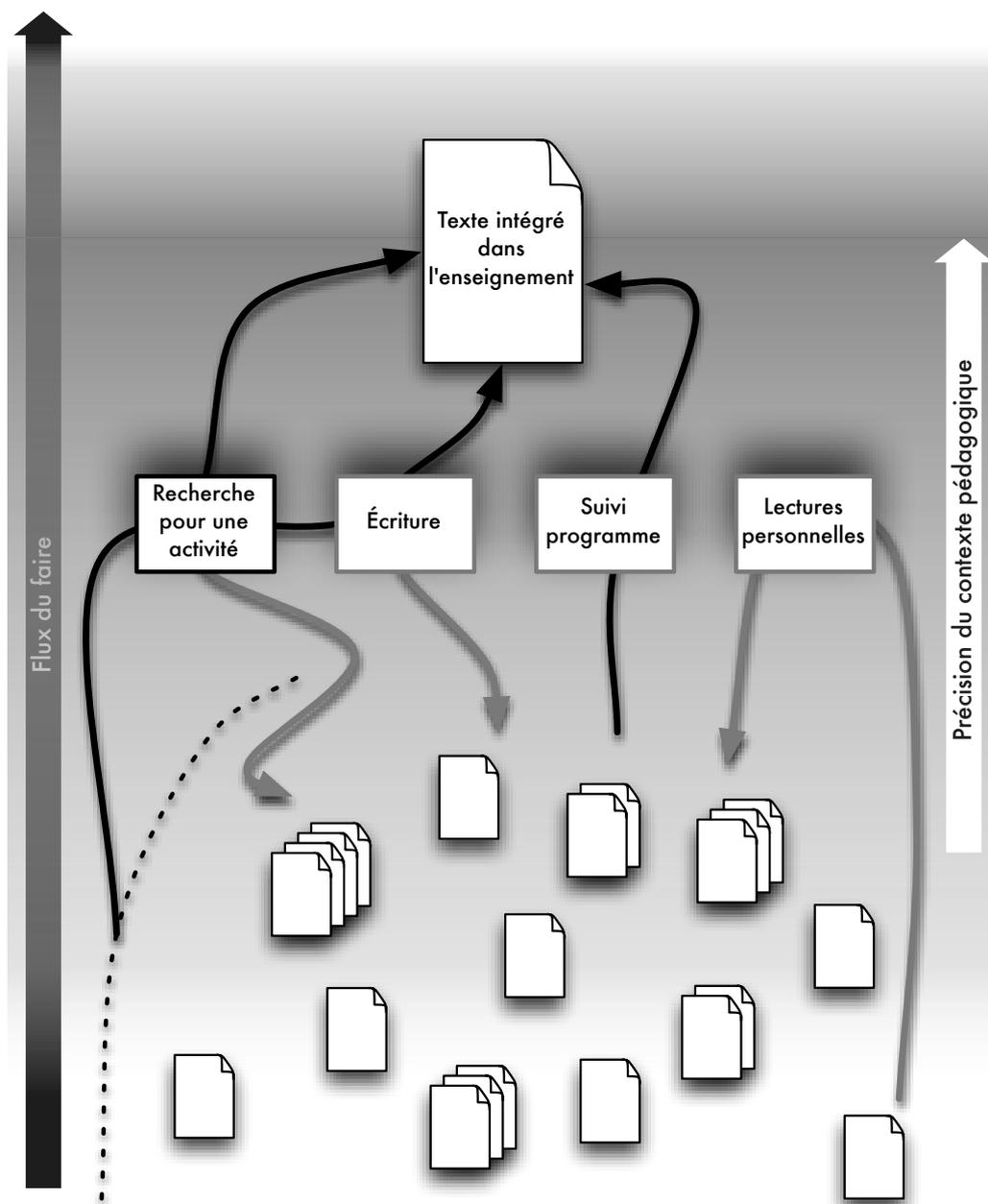


FIGURE 4.6: Transformation du texte de son écosystème d'origine au statut de texte utilisé dans l'enseignement des langues

entériné par le *faire* de l'enseignant¹¹⁵, ce texte a été écrit avec un certain objectif par son auteur ; il a été publié dans un certain contexte. Il peut ensuite avoir été ajouté à des archives, dans lesquelles l'enseignant l'aurait trouvé. Chacune de ces étapes correspond aux *faïres* de multiples acteurs, dont les actions elles-mêmes sont dépendantes du *faire* d'autres personnes. La puissance de la notion de flux du *faire* dans le contexte de notre travail réside dans le fait qu'elle reconnaît la complexité des processus, l'imbrication extrême de différents facteurs et que par l'idée de mouvement qu'elle véhicule ne nécessite pas de caractériser chaque relation.

Notre tâche descriptive concerne le mouvement en direction de l'utilisation d'un texte dans le contexte de l'enseignement des langues à travers les actions d'un enseignant. La masse diffuse du flux nous permet de reconnaître l'existence et l'importance d'autres *faïres* menant à l'objectivation de celui que nous décrivons¹¹⁶ sans pour autant devoir tous les caractériser. Les seules actions représentées dans la figure 4.6 sont celles de l'enseignant qui relèvent de la problématique de la recherche ou du choix d'un texte pour l'enseignement¹¹⁷. Nous décrivons les actions à travers leurs relations avec les objectivations des *faïres* d'autres personnes : des textes. Ces derniers suivent ainsi le point de vue d'Holloway :

« Parler du flux social du *faire* ne revient pas à nier la matérialité de ce qui est fait. [...] Quand j'écris un livre, il existe en tant qu'objet, il a une existence indépendante de la mienne, et il peut même exister quand je n'existerai plus. En ce sens, on peut dire qu'il y a une objectivation de mon *faire* subjectif, que ce qui est fait acquiert une existence séparée du *faire*, qu'il s'abstrait par lui-même du flux du *faire*. Mais cela n'est vrai que si l'on considère mon action comme un fait individuel. Considérée depuis le flux social de l'agir, l'objectivation de mon *faire* subjectif devient tout au plus éphémère. [...] L'existence du livre en tant que livre dépend du fait que toi, lecteur, tu le lises, de l'entremêlement entre ton *faire* (la lecture) et le mien (l'écriture), qui incorpore le fait (le livre) dans le flux social du *faire*. »

[Holloway, 2007, p. 50–51]

Depuis le flux social de l'agir, les artefacts au sens de Contamines *et al.* n'existent que s'ils ont subi un processus de genèse instrumentale. De la même manière, les textes ou groupes de textes existent dans la figure 4.6 parce qu'ils sont incorporés au flux par le *faire* de l'enseignant qui agit sur eux au cours de son processus de recherche.

b) Les objectivations

Pour aller plus avant dans le détail de nos conventions de notation, nous allons nous intéresser plus précisément au cas des ressources considérées, les textes. Dans notre étude des sources utilisées, nous avons constaté que les manuels, s'ils n'étaient pas la source la plus fréquemment choisie, constituaient la source la plus importante selon notre indice (cf. figure 4.5 p. 179). Les propositions « autres » nous ont mené à nous poser la question de comment les considérer. En effet, plutôt que d'aborder la question sous l'angle de la place des manuels dans la tâche de planification (cf. NB 4.4 p. 203), nous nous étions évertué à les considérer comme une source de textes à différencier des autres (cf. § b) p. 178). Le manuel n'est alors pas considéré comme un plan pré-mâché à adapter¹¹⁸, mais comme un recueil de textes. En effet, notre point de vue est

115. Cf. § 3.3.1 p. 98.

116. Le faire de l'enseignant recherchant/choisissant un texte dont l'objectivation est ce texte augmenté de la propriété d'avoir été utilisé dans l'enseignement des langues.

117. Il ne s'agit donc pas de décrire chaque action de la tâche de planification (cf. § 4.1 p. 137).

118. Voir les remarques de Yinger [1979] citées dans le NB 4.4 p. 203.

celui de la recherche de textes, si l'enseignant utilise un manuel pour mettre en place une séquence qui n'a recours à aucun texte, le manuel en question n'a pas d'intérêt pour notre problématique. Au contraire si l'enseignant met en place une séquence qui fait intervenir un texte issu d'un manuel alors, il nous semble tout à fait acceptable de considérer le manuel en question comme un recueil de textes.

Cela nous mène directement aux textes contenus dans ces manuels. Comme en témoignent Binon *et al.* [1998], il est attendu d'un manuel qu'il propose un certain nombre de textes authentiques¹¹⁹. Par définition, ces textes existaient dans un autre contexte, par exemple quand il s'agit de textes issus de périodiques [Binon *et al.*, 1998, p. 25]. Il est donc imaginable qu'un même texte utilisé dans une séquence pédagogique par deux enseignants différents provienne pour l'un d'un manuel et pour l'autre du périodique dans lequel il a initialement été publié. La différence entre les deux réside, du point de vue du processus, dans les données qui augmenteraient les textes du manuel. Nous pouvons considérer ces dernières comme des métadonnées : les consignes, les objectifs et autres informations prescriptives que l'enseignant s'appropriera, ignorera¹²⁰ ou à partir desquelles il inférera d'autres propriétés.

Tout ceci nous ramène au flux social du *faire* et plus particulièrement au *faire* passé comme moyen de faire (cf. § 4.6.1 p. 196) et aux textes comme objectivation des actions dont ils sont issus (cf. § 4.6.2 p. 198). En effet, un texte n'est pas le même objet dans son contexte de publication original et dans un manuel. Même si la mise en forme est parfaitement reproduite, il aura subi une série de transformations contextuelles qui s'exprimeront aussi bien par sa présence au sein d'une collection de textes dont des didacticiens auront estimé qu'ils sont propices à l'enseignement des langues, que par les consignes qui peuvent lui être associées.

Dans la figure 4.6 (p. 199), nous exprimons ces transformations à travers la position du texte dans le flux. Il faut tout d'abord rappeler que, du point de vue du flux social du *faire*, les objets n'existent que par leurs interactions avec les *faïres* des acteurs (les enseignants qui utilisent des textes dans des séquences pédagogiques dans notre cas). Les textes présents dans le schéma y figurent donc parce que l'enseignant dont nous décrivons l'action a été confronté à ces textes. Dans le cadre de la recherche de textes pour une activité, cela signifie que l'enseignant a été mené à se poser la question de savoir si le texte satisfaisait ses besoins ou non (et à y répondre).

La figure 4.6 contient donc un ensemble de textes et de recueils de textes (groupes de textes)¹²¹ auxquels l'enseignant est confronté dans sa « recherche pour une activité ». Ce sont donc les informations contextuelles accessibles à l'enseignant qui décident de « l'emplacement » du texte et reflètent le mouvement du texte dans le flux du *faire*, conséquence des actions de différents individus (dont l'enseignant).

Nous délimitons une sphère pédagogique, au sein de laquelle les informations contextuelles associées aux objets contiennent des propriétés « pédagogiques ». Pour faciliter la représentation, nous l'avons grisée dans la figure 4.6. Toutefois, comme les actions qui transforment les textes

119. Dans l'article, Binon *et al.* analysent 3 manuels. Parmi les critères d'évaluation de ces derniers figure la sélection des textes et l'authenticité de ces derniers [Binon *et al.*, 1998, p. 11]. Un seul manuel ne contient aucun texte authentique et les auteurs de l'étude de conclure : « Ce qui frappe, c'est l'absence totale de documents authentiques!! » [Binon *et al.*, 1998, p. 28]

120. Voir à ce sujet la citation de [Dargirolle, 1999] dans le NB 4.4 p. 203.

121. Dans le cadre de notre problématique, toutes les sources que nous avons considérées dans notre questionnaires, y compris le « programme », peuvent être ramenés à un recueil de textes. Le programme serait une collection de textes que l'enseignant doit présenter à ses apprenants au cours de l'année. Nous avons également explicité le fait de considérer un manuel comme un recueil de textes. Une œuvre littéraire peut être considérée comme un texte ou un groupe de textes, les périodiques comme des collections d'articles, etc..

qu'elle décrit, elle ne peut être délimitée nettement¹²², la définition de ce qu'est une propriété « pédagogique » n'est pas suffisamment précise.

Ayant explicité le sens du schéma, nous allons maintenant l'utiliser comme support pour décrire les actions discutées dans notre questionnaire, ce qui nous permettra de revenir sur les conventions introduites ici.

4.6.3 *Faires* en marge de notre problématique

Les éléments les plus directement représentatifs du processus de recherche de textes dans la figure 4.6 sont les différentes actions de l'enseignant qui peuvent mener à l'utilisation d'un texte dans le contexte de la classe de langue. Si elles n'ont pas de prétention d'exhaustivité¹²³, elles couvrent, d'après les réponses que les enseignants ont pu faire, une grande majorité des situations (cf. § b) p. 168).

Parmi ces actions, deux concernent plus clairement des pratiques ne relevant pas directement de la recherche de textes :

- l'écriture, puisque par définition un enseignant n'a pas recherché un texte qu'il écrit lui-même ;
- le fait de suivre le programme puisque dans ce cas l'enseignant suit des recommandations et ne recherche / choisit pas le texte utilisé.

a) Écriture et texte fabriqué

L'écriture par l'enseignant d'un texte en vue de l'utiliser en classe est indissociable de la notion de texte fabriqué, que nous avons traitée dans la section 4.4.1 (p. 167). Nous y apprenions notamment que le texte fabriqué était plus souvent recherché qu'il n'est écrit (par l'enseignant qui l'utilise). En caricaturant les résultats du tableau 4.8 p. 170, 50% des enseignants qui ont recours à des textes fabriqués ne font que les rechercher, alors que les autres 50% les écrivent ou les recherchent.

Dans la figure 4.6, les flèches représentent le *faire* de l'enseignant. Chacune est associée à la transformation d'une ressource laquelle constitue une objectivation de cette transformation. Deux flèches sortent de la boîte écriture. L'une pointe vers un texte intégré dans une activité : elle représente l'action d'écrire un texte pour y avoir recours dans l'enseignement des langues, de l'écriture à l'utilisation effective. Cependant, comme en témoigne la faible proportion d'enseignants qui écrivent tous les textes fabriqués qu'ils utilisent (< 4%), ces derniers proviennent souvent d'une recherche. Cela signifie que, dans ce flux social du *faire*, les textes fabriqués utilisés par des enseignants sont souvent déjà écrits au moment de la décision d'avoir recours à un tel texte. Ce dernier peut être issu du *faire* d'un enseignant pour lui-même (texte gardé dans un recueil d'une forme ou d'une autre) ou d'un texte écrit dans l'objectif de créer un manuel ou une autre forme de recueil institutionnalisé (cf. [127], § b) p. 178) ou non. La tâche d'écriture peut donc mener à remiser le texte écrit, en attendant qu'il soit ré-inclus¹²⁴ dans la partie du flux

122. Cf. [Holloway, 2007, p. 50].

123. Nous n'avons par exemple pas traité la situation décrite par Johns [1986] (et relatée dans ce mémoire p. 73) où l'enseignant ne choisit pas les textes mais demande aux apprenants de les écrire. Nous ne traitons pas non plus le cas où l'enseignant demande aux apprenants de choisir eux-mêmes leurs textes [Tagliante, 2006, p. 137].

124. Le texte ainsi écrit peut être utilisé directement *et* remisé (écriture comme moyen de *faire* ultérieur de l'auteur) ou partagé pour une utilisation extérieure (écriture comme moyen de *faire* ultérieur d'autres enseignants).

social du *faire* qui nous intéresse ici par l'action d'un enseignant. Nous plaçons le texte ainsi créé dans la sphère pédagogique¹²⁵, puisque c'est dans l'objectif de son utilisation en classe qu'il a été écrit, et ce, même s'il ne sera pas forcément utilisé ensuite selon les fonctions prescrites par son auteur¹²⁶.

La figure 4.6 rend aussi compte de la problématique de la « didactisation des ressources¹²⁷ » : avant d'être intégré dans une séquence d'enseignement / apprentissage, le texte issu d'une recherche¹²⁸ peut être édité, voire partiellement réécrit par l'enseignant (et pourquoi pas ensuite conservé pour une utilisation ultérieure ou partagé).

b) Suivi du programme

De la même manière que certains enseignants déclaraient avoir recours à des textes fabriqués pour des raisons parfois indépendantes de leur volonté (cf. [86], § c) p. 170), une certaine proportion d'enseignants a déclaré être soumise à des obligations dans le choix de ressources, sous forme de « textes imposés par le programme » (cf. figure 4.4 p. 176). Il s'agit de la provenance la moins importante d'après les déclarations des enseignants (cf. tableau 4.11 p. 177).

Nota Bene 4.4 Planification et manuels

Pour Yinger [1979, p. 168], l'appropriation par les enseignants des activités proposées dans les manuels est peu à peu prise en charge sous forme de « routine » et limite l'enseignant à un rôle d'exécutant, dont l'autonomie, l'initiative et le jugement sont subordonnés à ceux des auteurs du manuel. Malgré tout, d'autres études vont quelque peu à l'encontre de ce point de vue. En effet, Tochon cite Hashweh [1987] pour qui « plus l'enseignant est compétent, plus il prend de liberté dans l'utilisation des manuels » [Tochon, 1989, p. 25]. Une prise de liberté que Dargirolle exprime autrement puisque, pour elle, les différentes phases de planification « permettent à l'enseignant de modifier le programme et de le transformer par exemple en ajoutant, omettant ou changeant l'ordre des séquences d'un manuel » [Dargirolle, 1999, p. 143–144]. De plus, pour Clark & Yinger [1980], les enseignants considèrent la planification comme une opportunité d'en savoir plus sur le sujet traité et de prendre des décisions notamment quant à la complétude et la clarté du matériel pédagogique utilisé [Clark & Yinger, 1980, p. 29]. Enfin pour Boukottaya *et al.* [2009], la capacité de réutilisation d'un document dépend de la compréhension par l'enseignant de la vision générale et de la représentation de « l'Univers » des auteurs, qui est reflétée par le document en question [Boukottaya *et al.*, 2009, p. 236].

L'action correspondante dans la figure 4.6 consiste à prendre un élément issu des *faïres* de différents acteurs institutionnels – dans ces conditions, le « programme » peut être ramené à une collection de textes appartenant à la sphère pédagogique¹²⁹ constituant l'objectivation de ces *faïres* – et à le transformer en texte utilisé dans l'enseignement des langues, c'est à dire rechercher une manière de l'intégrer à une séquence pédagogique¹³⁰. Nous avons soumis ce type

125. L'extrémité de la flèche grise correspondante est dans la partie du flux que nous avons appelée la sphère pédagogique.

126. Fonctions constituantes dans l'approche instrumentale.

127. « didactiser un texte permet parfois de le rendre plus accessible, de mieux contrôler ce qu'il va apporter aux étudiants » [14] (cf. § c) p. 170).

128. Décrite plus en détails plus loin (cf. § 4.6.4 p. 205).

129. Comme l'indique l'origine de la flèche traversant la boîte « Suivi programme » dans la figure 4.6 p. 199.

130. Selon le degré de précision des informations prescriptives – programmes ou autre – des éléments de consigne sur comment intégrer les textes à des séquences pédagogiques peuvent être fournies. Quel que soit ce degré de précision l'enseignant doit se réapproprier la séquence (voir NB 4.4 p. 203).

de proposition aux enseignants, qui s'y reconnaissent beaucoup moins que dans d'autres critères (cf. § e) p. 186).

c) La recherche de textes comme un mouvement

Les actions correspondant à l'écriture¹³¹ et au suivi du programme sont celles qui concernent le moins d'enseignants. Elles nous permettent cependant d'appréhender un peu mieux l'adaptation du concept de flux social du *faire* à notre problématique.

En prenant l'exemple de l'étiquette « écriture », nous allons, à partir des différents cas traités, entrer dans le détail de la description de la tâche de recherche/choix de textes comme un flux. Nous avons vu qu'un enseignant pouvait être amené à écrire des textes pour les utiliser dans le cadre d'une séquence pédagogique (flèche noire dans la figure 4.6 p. 199). Mais ce n'est pas nécessairement la seule utilisation qui peut être faite du texte. Ce dernier peut ensuite être ré-injecté dans le flux social de l'agir par la conjonction de *faïres* qui le mènent à être utilisé ultérieurement dans l'enseignement des langues (par son auteur ou non). Une séquence d'actions correspondante pourrait par exemple consister en l'inclusion d'un texte fabriqué dans un recueil – de par les circonstances de son écriture, le texte appartient à la sphère pédagogique – puis en sa découverte au cours d'une recherche menée par un autre enseignant. Le fait que ces deux actions soient présentes simultanément dans la figure 4.6 met en évidence les conventions de représentation expliquées dans la section 4.6.2 (p. 198). La réinsertion d'un texte fabriqué dans le flux et l'action de le trouver par le biais d'une recherche ne sont pas présentés de manière séquentielle, contrairement à la didactisation d'un texte entre sa recherche et son utilisation. En effet, nous avons considéré qu'il s'agissait d'une conjonction d'actions et non d'un unique processus.

En revanche, le mouvement inhérent à la notion de flux permet de prendre en compte cette interaction de plusieurs *faïres*. Il faut considérer le schéma comme une entité qui se meut et les actions comme des étiquettes qui ne sont présentes que par notre description/observation d'un phénomène pendant un laps de temps l . Le flux contient alors toutes les actions antérieures et leurs objectivation. Cela signifie par exemple que quand nous décrirons la « recherche pour une activité¹³² », si nous nous attachons à expliciter l'action de recherche d'un texte fabriqué (donc dans la sphère pédagogique), chaque texte candidat est l'objectivation d'une occurrence antérieure de l'action d'écriture (telle que nous la décrivions plus haut). En d'autres termes, il s'agit d'une occurrence antérieure de ce même schéma pour laquelle l'observation est focalisée sur l'action d'écriture durant un laps de temps l_1 . La figure 4.6 propose une synthèse de tous ces laps de temps concernant la partie liée à l'acquisition d'un texte pour une activité au sein du *faire* de l'enseignant. Toutefois, elle doit être considérée une action à la fois, les *faïres* antérieurs sont pris en compte à travers les ressources qui en proviennent. C'est là l'intérêt de la description comme un mouvement : malgré l'énumération d'un ensemble d'actions sur lesquelles nous nous focalisons, la représentation n'est pas strictement disjonctive. Elle reconnaît la complexité du réel et l'impossibilité de prendre en compte et de décrire toutes ses composantes (cf. § 4.1.8 p. 145). Elle reconnaît l'influence de tous ces *faïres* les uns sur les autres, mais ne prétend pas permettre de tous les caractériser. Dans la figure 4.6, nous nous focalisons donc sur les actions qui mènent le plus fréquemment et directement à l'intégration d'un texte à une séquence d'enseignement, en

131. Nous n'avons pas de données sur l'importance de l'écriture comme « didactisation » d'un texte, notre remarque concerne donc en particulier l'écriture d'un texte fabriqué.

132. Voir § 4.6.4 p. 205.

les incluant dans un flux global. Nous allons donc observer deux séries d'occurrences de notre schéma pour traiter les « lectures personnelles ¹³³ » et la « recherche pour une activité » au cours desquelles nous allons examiner plus précisément cette « sphère pédagogique ».

4.6.4 Recherche pour une activité

La « provenance » la plus invoquée par les enseignants que ce soit en termes de fréquence ou d'importance est la recherche de textes pour une activité qui concerne 96,9% des enseignants (cf. figure 4.4 p. 176). Nous pouvons commencer par remarquer que ce résultat est tout à fait cohérent avec la littérature sur la planification. Le matériel pédagogique est une composante de l'activité [Charlier, 1989, p. 31], il n'est donc pas surprenant qu'elle influence la recherche et le choix des ressources. Ceci est d'autant plus cohérent que l'activité occupe une place centrale dans la tâche de planification. Clark & Yinger expliquent par exemple que les objectifs d'apprentissage sont rarement le point de départ de la planification, qui serait plutôt axée autour des apprenants et des activités ¹³⁴[Clark & Yinger, 1979, p. 15]. Comme nous le signalions dans la note ⁸² p. 186, Tochon précise ce point de vue en explicitant les relations entre apprenants et activité : « Shavelson & Stern [1981] [ont remarqué que] les enseignants n'emploient pas le modèle de Tyler [1949] qu'on leur a enseigné et qu'ils ne planifient pas par objectifs mais par activités, celles-ci s'adaptant mieux aux différents profils d'élèves. » [Tochon, 1989, p. 24].

Cette position transversale de la recherche pour une activité tendrait à donner raison aux standards de description de ressources pédagogiques que nous avons étudiés. Dans ces derniers, les éléments pédagogiques concernent en particulier des propriétés relevant de la situation d'utilisation des ressources. Cependant, le problème que nous avons soulevé, à savoir la multiplicité des utilisations potentielles d'un texte ¹³⁵, se trouve validé par le biais de la confirmation de l'hypothèse H_b (cf. § 4.4.5 p. 187). Il va donc nous falloir tenter de prendre en compte la multiplicité des utilisations potentielles des textes dans le processus de recherche.

Pour ce faire, nous allons observer le flux des actions qui impliquent les textes en considérant la recherche comme une séquence en trois étapes inspirées de [Yinger, 1978] que nous représentons dans la figure 4.7 page suivante :

- sélectionner un texte ;
- évaluer ses propriétés en fonction des besoins ;
- transformer le texte, lui assigner un contexte ;

Encore une fois ces actions sont à replacer dans un flux d'actions. Comme nous l'avons vu en 4.1 (p. 137), Yinger définit la planification comme un processus en 3 phases : définition du problème, formulation/solution du problème et implémentation/évaluation/routinisation. Nous voulons décrire ici la recherche pour une activité, ce qui signifie que la définition du problème a eu lieu et qu'il s'agit de mettre en place une activité. La troisième phase correspond à l'interaction, nous nous intéressons donc ici à la seconde phase de la planification, que Yinger visualise comme une spirale [Yinger, 1978, p. 36]. En suivant cette dernière, l'enseignant résout son problème graduellement en effectuant successivement des séquences de trois étapes (élaboration, investigation, adaptation) [Charlier, 1989, p. 24]. Chaque élaboration est la proposition d'une solution à un sous-problème qui est évaluée mentalement pendant l'investigation avant d'être adaptée au

133. Que nous décrivons dans la section c) (p. 212).

134. « *Learning objectives are seldom the starting point for planning. Instead, teachers plan around their students and around activities.* »

135. Cf. § 3.5.3 p. 120.

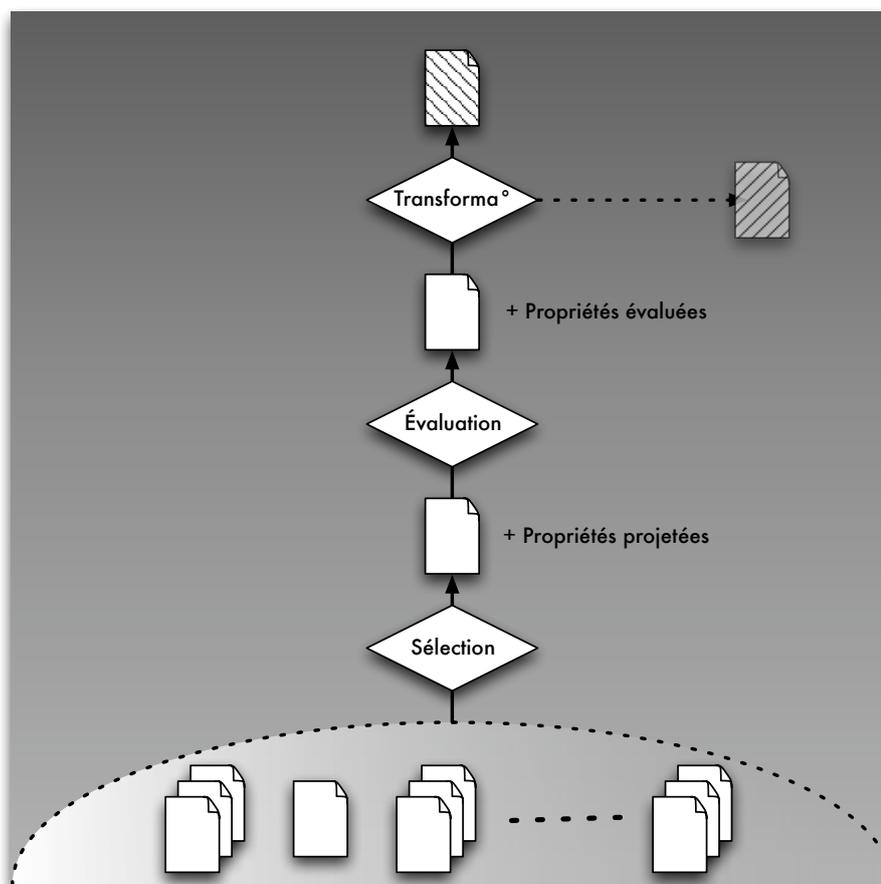


FIGURE 4.7: La recherche de textes vue selon le modèle de Yinger [1978]

problème global. La répétition de ces étapes permet de préciser graduellement le problème en vue de le résoudre finalement¹³⁶. Notre point de vue est bien plus restreint puisqu'il concerne un sous-problème, cependant son inscription dans un modèle tel que celui de Yinger suggère qu'il est influencé par des éléments plus globaux, d'autres prises de décisions et qu'il y a potentiellement plusieurs itérations de la recherche. Toutefois, nous nous attachons ici à décrire une seule itération.

a) Sélection

La première action définie ci-dessus est objectivée par un texte présent dans la figure 4.6. Comme l'indique la base en pointillés à l'origine de la flèche, lors d'une recherche pour une activité, le texte peut se trouver n'importe où dans le flux. Nous avons reproduit cette ligne en pointillés dans la figure 4.7. En effet, l'enseignant effectue sa recherche pour une activité précise. Selon ce qu'il espère trouver, il se tournera vers une source en particulier, qui lui permet de consulter un texte en vue d'évaluer ses propriétés. Le texte en question peut donc avoir été trouvé aussi bien dans un manuel (sphère pédagogique, puisque destiné à l'enseignement des langues) que dans un

136. La troisième phase correspond à l'implémentation et l'évaluation de la solution au cours de la phase d'enseignement [Charlier, 1989, p. 24].

périodique (qui, au contraire, n'appartiendrait *a priori* pas à cette sphère). Entre notre étude préliminaire et ce questionnaire, nous pouvons trouver plusieurs exemples de tels comportements :

- choix d'un auteur et d'un type de texte en fonction des structures qui sont attribuées à son style d'écriture : « “*Roald Dahl, [...] all his short stories are packed with these verbs [...] for emotion and gestures [...], that in French [require] a whole phrase [...]*.”¹³⁷ » [Loiseau, 2003, p. 58] ;
- choix d'un journal en fonction de la lisibilité présumée de ses articles : « “*There are certain news magazines, or newspapers I will never choose a text from, never ever. I will never use a text from Time or Newsweek. [...] It is packed with idiomatic expressions and what turn out to be often non-standard constructions. The Guardian is full of typing mistakes, but [...] it has a clear style.*”¹³⁸ » [Loiseau, 2003, p. 57] ;
- choix d'un numéro de périodique en fonction d'un type de texte attendu (cf. citation p. 139) ;
- choix d'un type de texte (en l'occurrence, fabriqué) par rapport à un type d'activité (cf. § c) p. 171) ;
- choix de la source en fonction du couple public-type d'activité (cf. [31], § b) p. 179).

Cette phase de sélection d'un texte à évaluer s'appuie donc sur des propriétés attribuées *a priori* (présence supposée de structures, lisibilité, type de texte, compatibilité avec un type d'activité dans nos exemples) en lien avec l'activité qui doit être mise en place. Ces propriétés peuvent être présumées (ex : *lisibilité des articles d'un journal donné*) ou avérées (ex : *statut d'objet pédagogique avéré, pour les textes d'un manuel*), elles servent en premier lieu à restreindre l'ensemble des textes à considérer.

À ce titre, elles étendent, pour ce problème, la sphère pédagogique aux textes en question, puisqu'ils sont considérés pour leurs propriétés pédagogiques présumées (cf. figure 4.7 p. 206). Ce mécanisme met en exergue :

- d'une part, les mouvements qui s'opèrent au sein du flux, sa plasticité : texte augmenté d'un ensemble de propriétés projetées ;
- et d'autre part la grande dépendance des propriétés envers le contexte : l'enseignant ne choisit pas systématiquement la même source (cf. variété des sources utilisées, § b) p. 178) à travers un choix raisonné (cf. ci-dessus).

En outre, ces dernières sont clairement influencées par le point de vue de l'enseignant sur les ressources, par son expérience, qui s'alimente par exemple des caractéristiques du public auquel il est confronté ou des interactions avec des pairs. Nous pouvons par exemple nous demander l'influence qu'aurait le fait d'enseigner l'anglais comme langue seconde ou comme langue de spécialité (pour le journalisme) sur le point de vue de l'enseignante qui expliquait l'incompatibilité de *Time* avec ses pratiques. En effet, le fait d'avoir des étudiants plus fréquemment confrontés à la langue anglaise employée en situation pourrait la rendre plus permissive aux expressions idiomatiques non standard. Les pratiques de collègues, qui s'appuieraient sur des textes issus de *Newsweek* pourraient également l'influencer.

Ce que nous voulons montrer par là, c'est la complexité des mécanismes de création des préconceptions des enseignants, qui influencent la manière dont ils perçoivent les textes. Loin de nous l'idée de porter un jugement dépréciatif sur ces préconceptions, tant elles nous semblent

137. Roald Dahl, toutes ses nouvelles sont remplies de ces verbes d'émotion et de gestes qui s'expriment en français avec des phrases entières.

138. Il y a certains magazines d'information ou certains journaux desquels je ne choisirai jamais un texte, « jamais de la vie ». Je n'utiliserai jamais un texte de *Time* ou *Newsweek*. Ils sont remplis d'expressions idiomatiques et de constructions qui s'avèrent souvent ne pas être standard. *The Guardian* est plein de fautes de frappe, mais le style est clair.

nécessaires à la tâche de recherche de textes. Un enseignant complètement « ouvert » à toute forme de texte, dépourvu de préjugés, n'aurait aucun élément qui lui permettrait d'établir une stratégie de sélection des textes à considérer. Son processus reviendrait à une phase de sélection essentiellement aléatoire.

Ainsi décrite, notre action élémentaire « sélection » est une instanciation de la phase d'élaboration du modèle de Yinger pour notre problème. Pour lui, le produit de l'élaboration est soit une solution à un sous-problème soit une facette de la définition du problème global [Yinger, 1978, p. 39]. Que l'on considère le choix d'un support comme un sous-problème ou un aspect de la planification d'une activité, le fait de proposer un texte constitue une solution potentielle. Pour Yinger, elle s'appuie sur l'exécution de routines mémorisées par l'enseignant ou la tentative d'en définir de nouvelles [Yinger, 1978, pp. 37-38]. Dans le cadre de la sélection, nous lierions la définition de ces routines à la mise en place progressive de préconceptions fondées sur l'expérience.

b) Évaluation

Une fois ce premier filtre passé, l'enseignant se trouve confronté à un texte, dont il a anticipé certaines propriétés. À travers une tâche de lecture¹³⁹, l'enseignant évalue les propriétés du texte à l'aune de sa problématique. C'est la phase que Yinger appelle l'investigation. Pour lui cette phase a deux vocations principales :

« *Investigation has two primary functions. First, it provides information about the workability of the elaboration and its success or failure as a subproblem solution. Second, it provides new knowledge and information about the planning problem based on the results of the investigations.*¹⁴⁰ »

[Yinger, 1978, p. 39]

La seconde des deux fonctions principales ne relève pas directement de notre problématique, puisque pour nous, elle concernerait les répercussions qu'a l'investigation sur la planification : l'analyse du texte par rapport au problème concerné modifie la perception de l'enseignant concernant d'autres aspects de l'activité conçue. Il pourrait s'agir par exemple de la modification du rôle des apprenants du fait du choix du texte.

En revanche la première concerne directement l'évaluation telle que nous la concevons. Elle doit répondre à la question de l'utilisabilité du texte dans l'activité planifiée. Son succès revient à considérer le texte utilisable dans ce contexte et son échec au cas contraire. Cette évaluation du succès ou de l'échec du texte comme support de l'activité correspond à la confrontation du texte aux besoins de l'enseignant, ce qui peut consister en l'évaluation des propriétés projetées sur le texte pendant la phase d'élaboration. C'est-à-dire qu'il est tout à fait imaginable que la propriété qui a servi à l'enseignant à isoler le texte « lu » au sein du flux du *faire* soit invalidée par son évaluation (ex : *présomption de lisibilité d'un article du Guardian, qui peut s'avérer inexacte dans le cas d'un texte précis dans le contexte de la préparation d'une activité en particulier pour un public donné*). Il peut également s'agir d'autres critères qui se démarquent complètement de ceux qui ont été utilisés pour la sélection. Ils peuvent être très précis, comme le suggèrent certains témoignages recueillis pour la question A.2.3 p. a-5 (textes fabriqués, cf. § 4.4.1 p. 167). Nous

139. Il y a fort à parier que le mode de lecture soit différent selon le type d'activité à mettre en place, mais toute suggestion ne serait que conjecture.

140. L'investigation a deux fonctions principales. Premièrement, elle fournit des informations sur l'efficacité de l'élaboration et sur son succès ou son échec comme solution au sous-problème. Deuxièmement, elle fournit de nouvelles connaissances et informations sur le problème de planification à partir des résultats des investigations.

pouvons citer comme exemple l'évaluation du nombre d'occurrences d'une structure grammaticale à travailler et de son adéquation avec l'objectif de l'activité (cf. témoignage [55], p. 171). Toutefois, si l'on en croit la formulation des réponses des enseignants (cf. annexe C p. a-13), l'évaluation s'effectue plus fréquemment en termes d'adaptation à un contexte d'enseignement donné (activité, public). Cette évaluation peut faire intervenir des critères plus ou moins formalisables, touchant à l'affect comme « [le] plaisir, [l']intérêt culturel, [l']intérêt des élèves »^[86] (cf. § d) p. 185), un témoignage appuyé par les 48,3% d'enseignants ayant choisi le critère « intérêt du texte indépendamment du contexte scolaire » (cf. tableau 4.18 p. 186).

À nouveau, le résultat intermédiaire est une transformation du texte par apposition de nouvelles propriétés (qui peuvent infirmer ou confirmer les propriétés projetées). Elles diffèrent de celles ajoutées à l'issue de la sélection puisqu'elles ont été vérifiées par l'enseignant au cours de sa lecture (cf. figure 4.7 p. 206). Elles sont également influencées par le contexte d'utilisation du texte, qui est donc au cœur de la description du texte.

c) Transformation

Enfin, notre dernière action élémentaire de recherche pour une activité s'apparente à la phase d'adaptation, la dernière dans la séquence définie par Yinger. C'est une phase d'intégration et de transformation qui vise à formuler et développer le problème global soit en achevant une portion de ce dernier, soit en obligeant l'enseignant à adopter un point de vue nouveau sur le problème [Yinger, 1978, p. 42].

Dans le cadre d'une recherche pour une activité, l'adaptation du problème à l'évaluation effectuée aboutit inmanquablement à une décision que nous pouvons ramener à une nouvelle transformation du texte, qui peut revêtir 3 formes :

- l'assignation d'un contexte d'utilisation correspondant à la recherche de l'enseignant (flèches noires émanant de la boîte « recherche » dans la figure 4.6) : le texte peut être transformé directement en support, d'une part de l'activité et, d'autre part, de la réflexion pour la mettre en place. Mais cette poursuite de la planification peut passer par une didactisation du texte¹⁴¹. Enfin, il n'est pas exclu que ce choix soit finalement remis en question par la suite de la planification. C'est toutefois la première de ces configurations qui nous concerne avant tout ;
- le texte ne peut pas être utilisé dans le contexte correspondant à la recherche de l'enseignant, mais il mérite d'être conservé pour une utilisation différente, un contexte différent (flèche grisée dans la figure 4.6, en pointillés dans la figure 4.7) : la perspective d'utilisation et donc les propriétés seront plus ou moins précises. Ces propriétés ne seront affinées que lorsque l'enseignant exécutera la tâche de planification dont la ressource est destinée à faire partie ;
- le texte n'est pas pertinent du point de vue de l'enseignant (aucune flèche) : le texte est rejeté.

Dans les deux derniers cas mentionnés, la recherche s'avère infructueuse et l'enseignant devra reproduire l'opération, quitte à ajuster ses critères *a priori* (ex : *recherche d'un texte fabriqué après l'échec de la recherche d'un texte authentique*, cf. [44] p. 171). Il peut même être amené à se rabattre sur un autre mode d'obtention de texte, comme l'écriture. Tous ces cas de figure reviennent précisément au nouveau point de vue que l'enseignant devra adopter sur le problème que Yinger évoquait.

141. Cf. témoignages [14,30] (cf. note ¹²⁷ p. 203). Dans la figure 4.6 (p. 199), nous avons exprimé cette possibilité avec la flèche qui relie la boîte « recherche pour une activité » et le texte intégré dans l'enseignement en transitant par la boîte « relecture ».

d) Contexte pédagogique

À travers les trois actions détaillées ci-dessus nous avons pu vérifier que la recherche de textes est un sous-problème de la planification au sens de Yinger. Cette tâche suit précisément les étapes qu'il propose pour décrire la formulation/solution du problème. Notre séquence (sélection, évaluation, assignation de nouvelles propriétés) constitueraient un « tour » dans la spirale qu'il définit pour cette seconde phase de la planification. En suivant son modèle, il n'est pas exclu que le problème traité (recherche d'un texte pour une activité) resurgisse, même en cas de succès, suite à la précision du problème global. Enfin, les travaux de Yinger suggèrent que le rôle du texte et son intégration à la séquence en cours de planification est le résultat de la résolution de multiples sous-problèmes et de remises en questions des solutions préalablement apportées. Au même titre, des modifications peuvent être apportées ensuite aux ressources et à leurs propriétés.

De notre point de vue, des propriétés sont affectées aux textes au fur et à mesure du processus de recherche, mais aussi à travers l'influence des autres sous-tâches de planification et de la manière dont l'enseignant les perçoit comme système. Peu à peu, les textes sont transformés par rapport à la sphère pédagogique, elle-même affectée par le contexte.

En effet, si nous reprenons le processus depuis son origine, le point de départ de la tâche est la formulation d'un problème : la mise en place d'une activité nécessitant un ou plusieurs textes. Une fois le problème posé, l'enseignant entre dans une phase d'élaboration qui fait appel à des routines pré-définies. Celles-ci s'expriment dans notre cas par l'assignation *a priori* de propriétés à des textes. Elles sont étroitement dépendantes du contexte d'utilisation : l'objectif de l'enseignant est de trouver une ressource potentiellement capable de constituer une solution au sous-problème « trouver un texte pour l'activité ». Ce faisant, l'enseignant commence à préciser les propriétés des textes en fonction du contexte d'utilisation. Après en avoir sélectionné un, il investit, évalue l'adéquation du texte à son problème. Cela passe par la confrontation de certaines propriétés plus ou moins concrètes du texte avec des valeurs attendues qui, là encore, reflètent le contexte dans lequel il sera utilisé. Cette étape d'évaluation aboutit ensuite à l'assignation de nouvelles propriétés à l'objet, parmi lesquelles sa qualité éventuelle de solution du problème. Toutefois, le cycle de modification du texte ne s'arrête pas au moment où il est choisi. Son adaptation au problème peut demander un travail de réécriture¹⁴² ou affecter le problème global et donc le contexte de son utilisation.

Toute cette phase de recherche correspond donc, du point de vue du texte, à une définition graduelle du contexte dans lequel il sera utilisé. Ce contexte, que nous appellerons le contexte pédagogique, semble constituer, toujours selon ce point de vue centré sur le document, un ensemble de propriétés qui seront apposées aux documents par l'enseignant à partir de caractéristiques évaluées pendant la lecture en fonction de sa perception du problème. Au fur et à mesure de la recherche, ce contexte pédagogique est précisé, mais il ne sera cristallisé qu'au moment de l'utilisation effective du texte, puisque la planification est un processus itératif dans lequel chaque résolution de problème est susceptible d'influencer la perspective globale de l'enseignant. Par conséquent, la figure 4.7 (p. 206) se termine plus tôt dans le processus que la figure 4.6 (p. 199). Le produit de la figure 4.7 est un texte auquel un contexte pédagogique provisoire a été assigné. Dans la figure 4.6, nous adoptons un point de vue en mouvement appliqué aux ressources en fonction d'un *faire* : la mise en place d'une activité. La figure intègre la recherche de texte à un processus plus large, reconnaissant ainsi les interactions entre les différentes composantes de la

142. En nous appuyant principalement sur [Vigner, 2001] et [Widdowson, 1981], nous indiquons les différents types de modifications qu'un enseignant peut opérer sur un texte (cf. § a) p. 65).

tâche globale et l'influence exercée sur ces interactions par le point de vue de l'enseignant, fruit de son expérience. La figure se termine avec la transformation ultime du texte (du point de vue de l'enseignant planifiant son cours), le texte tel qu'il est utilisé dans l'activité présentée aux apprenants.

4.6.5 Introduction du contexte pédagogique et conséquences sur la modélisation des pratiques en rapport avec la planification d'activités faisant intervenir des textes

L'étude de la recherche pour une activité nous a permis de mettre en évidence la notion de contexte pédagogique, que nous allons définir ici, avant d'examiner les conséquences de cette notion sur la description des pratiques des enseignants.

a) Définition

Définition 4.1a (Contexte pédagogique) *Ensemble des traits décrivant la situation d'enseignement.*

Notre but n'est pas ici de définir toutes les propriétés que le Contexte Pédagogique (CP) peut comprendre¹⁴³, mais plutôt d'étudier ses composantes utilisables du point de vue de la recherche de textes (cf. § d) p. 172). En effet, lorsque l'enseignant recherche un texte, il construit peu à peu un contexte pédagogique pour celui-ci. Il ne prend pas systématiquement en compte l'intégralité des informations permettant de décrire la situation d'enseignement. La décomposition en sous-problèmes permet de se focaliser sur certains aspects. Par exemple, lorsqu'un enseignant décide de se tourner vers un texte fabriqué (sélection), pense-t-il systématiquement à « l'endroit¹⁴⁴ » ?

Il nous faut donc isoler des caractéristiques de la situation d'enseignement, les relations qu'elles entretiennent entre elles et avec les propriétés intrinsèques du texte. Du fait de la multiplicité des utilisations possibles d'un texte (cf. § 4.4.5 p. 187), les propriétés du CP sont essentiellement extrinsèques quand elles sont rapportées à ce type de ressources brutes. En effet, elles reflètent le point de vue de l'enseignant qui définira la situation selon ses propres termes.

Dès lors, une indexation prenant en compte cette notion doit tenter d'évaluer jusqu'où une approximation des mécanismes régissant l'influence du CP sur les propriétés du textes est viable pour les enseignants (et la multiplicité de leurs points de vue). C'est une question qui dépasse largement le cadre de notre travail. Dans la perspective de la stratégie moins-disante que nous avons choisie (cf. § 2.2.5 p. 63), nous nous contenterons d'éléments basiques. Cependant l'objectif est de proposer un modèle évolutif qui puisse prendre en charge la complexification du problème, évidente dès lors que les propriétés décrites gagnent en abstraction.

143. Partir de l'activité et de ses 7 composantes – matériau pédagogique, endroit, structure (séquence), durée, participants, comportements acceptables, stratégie d'enseignement, contenu [Charlier, 1989, p. 31] – serait une piste dans cette direction.

144. Lieu où se déroule la séquence.

b) Place de notre système dans le processus de recherche

Avant d'examiner le contexte pédagogique au sein des pratiques des enseignants, il convient de resituer le rôle d'une base de textes indexée pédagogiquement pour l'enseignement des langues dans le processus de recherche associé. En effet, d'après les cas d'utilisation que nous avons définis (cf. § 1.2 p. 7), un tel système n'intervient pas à toutes les phases de la recherche.

Notre description de la recherche de textes s'appuie sur les actions de l'enseignant, utilisateur du système. Elles délimitent notamment la phase de sélection : c'est le fait de consulter un texte qui marque la transition entre sélection et évaluation. Rapporté aux cas d'utilisations, la première phase correspond donc à celui que nous avons appelé « recherche d'un texte » (cf. § 1.2.2 p. 9). À la lumière des éléments développés ici, ce terme est mal choisi, puisqu'à l'issue de la « recherche », l'enseignant est confronté aux textes candidats pour la première fois. Dans ce cas d'utilisation, le système effectue le premier filtre. Pour y parvenir, il s'appuie sur la conception qu'a l'enseignant du CP au moment de l'action. Dans le processus tel que nous l'avons décrit, l'enseignant choisit le texte à consulter en fonction de propriétés qu'il lui assigne *a priori*. Dans le cadre de ce processus aidé par un système informatique, il lui faut pouvoir définir un ensemble de propriétés qu'il espère trouver dans les textes. Le système s'en sert ensuite de filtre pour proposer à l'utilisateur un ensemble de textes candidats à évaluer.

La phase d'évaluation correspond au cas d'utilisation que nous avons appelé « aide au choix d'un texte » (cf. § 1.2.3 p. 9). Durant cette étape, l'enseignant accède directement aux propriétés des textes et à leurs contenus. Il procède à une « investigation » qui le mène à confronter les propriétés du texte à sa problématique pour l'aider à prendre une décision. Dans cette perspective, l'aide au choix doit lui permettre d'accéder aisément à certaines propriétés, lui facilitant ainsi le travail d'inférence qui le mènera à décider de l'utilisabilité du texte. La phase de transformation du texte ne correspond donc à aucun cas d'utilisation¹⁴⁵.

c) Le CP dans le processus de planification utilisant des textes

La notion de CP ayant émergé au cours de notre description de la recherche de textes pour une activité, pour pouvoir conclure sur l'apport de cette notion, il nous faut la confronter aux autres pratiques menant à l'intégration d'un texte à une séquence pédagogique. Avant de revenir aux pratiques déjà décrites, intéressons-nous à la recherche passive, que nous avons laissée de côté jusqu'ici.

La recherche passive, par sa dimension impromptue et liée au hasard¹⁴⁶, débute vierge de toute velléité pédagogique. La sélection du texte n'est pas liée à un problème de planification et à ce titre débute en dehors de la sphère pédagogique (cf. origine de la flèche correspondante dans la figure 4.6 p. 199). Mais au cours de sa lecture, l'enseignant considère qu'il pourrait utiliser le texte dans le cadre d'une activité. Cela correspondrait de notre point de vue à l'association du texte à une forme de CP. En effet, si le texte est conservé pour un public ou une activité (cf. § 4.4.4 p. 182), le lien avec le contexte pédagogique est évident. Cependant, même dans le cas d'un texte conservé « en attendant de trouver une manière adéquate de l'utiliser » (cf. tableau 4.15 p. 183), nous pouvons décrire l'action en termes de CP. L'enseignant lui attribue la propriété

145. L'adaptation de la vision de Recker & Wiley [2001] ou de Contamines *et al.* [2003], qui constituerait une extension du système, demanderait l'ajout d'un cas d'utilisation « prise de décision », qui répercuterait les choix des enseignants dans la description des textes.

146. Voir les sections 4.1.6 p. 144, b) p. 174 et c) p. 181.

« utilisable dans le contexte de l'enseignement des langues ». Il commence donc à décrire la situation d'enseignement puisqu'il associe la ressource à l'enseignement des langues. En outre, ce contexte est moins flou qu'il peut en avoir l'air, puisque l'enseignant conserve le texte par rapport à ses propres pratiques et par conséquent aux publics auxquels il enseigne, aux types d'activité qu'il a tendance à utiliser, aux routines qu'il a développées. Le contexte pédagogique est donc fortement lié à l'enseignant qui le définit et à son vécu. Pour en revenir à la recherche passive, qu'elle se solde par une conservation pour un public ou une activité précise ou « en attendant de trouver une manière de l'utiliser », le fait de conserver le texte peut être considéré comme une forme de contexte pédagogique de précision variable et toujours empreint de l'expérience de l'enseignant.

Comme l'inclusion à une séquence est du ressort d'une tâche de planification, la seule flèche issue de la recherche passive transforme le texte en l'incluant à la sphère pédagogique, fort de son augmentation par une forme de CP. La tâche de planification qui l'intégrera effectivement à une activité peut prendre plusieurs formes : soit le texte est retrouvé après la première formulation du problème autour d'une activité, auquel cas il s'agit d'une recherche pour une activité, soit la planification part du texte, le processus s'apparente alors au fait de suivre le programme. Dans le premier cas, l'enseignant projette, entre autres, les propriétés qu'il avait assignées au texte pendant sa « recherche passive » dans le cadre de la phase de sélection. Dans le second, la seule différence avec le suivi du programme est que la contrainte d'utiliser le texte émane de l'enseignant lui-même et non d'une institution.

Dans ces deux situations, le processus, inclus dans la complexité du flux social du *faire*, revient à définir graduellement le contexte pédagogique, dans lequel le texte sera utilisé. C'est aussi le cas de l'écriture de texte dans laquelle, l'enseignant part d'un CP relativement précis et répercute son influence à travers des propriétés du texte qu'il écrit. Le CP se traduit en un besoin de contrôle du contenu du texte (cf. § d) p. 172).

Enfin, la différence entre flèches noires et grisées de la figure 4.6 peut également être décrite selon le point de vue du contexte pédagogique. Les premières aboutissent à un CP complet, le texte dans le cadre de l'activité, alors que les secondes correspondent à une réinsertion du texte dans le flux du *planifier une activité*. Selon Holloway [2007], pour exister dans le flux de l'agir, l'objectivation du *faire* de l'enseignant, dépend de sa réutilisation par un acteur. L'inclusion d'un texte à un recueil, qui constitue une forme d'assignation de contexte pédagogique, signifie donc que ce texte existera par l'usage du recueil. Ensuite, ce sont les propriétés définies *a priori* par l'enseignant (durant l'ajout au recueil et pendant la recherche) qui décideront de sa sélection.

4.7 Conclusion

À travers les pratiques déclarées d'un échantillon de 130 enseignants, ce long chapitre nous a permis de nous intéresser aux processus qui menaient à l'inclusion de textes à des séquences d'enseignement/apprentissage. Sans avoir de prétentions d'exhaustivité ou d'universalité, cette étude nous a permis de tirer des enseignements généraux applicables aux pratiques d'enseignants de L2 différentes, exerçant dans des cadres divers et bénéficiant de degrés d'expérience variables. Les enseignants de l'étude pratiquaient couramment l'éclectisme et tous étaient menés à rechercher des textes.

L'étude nous a permis d'isoler 4 pratiques menant à l'intégration de textes à des séquences pédagogiques : la recherche pour une activité, l'écriture, le suivi du programme et la recherche

passive. Sans jamais s'exclure les deux derniers cas avaient tendance à s'opposer (cf. tableau 4.13 p. 180 et § c) p. 181). Finalement la conjonction entre ces 4 classes de pratiques, les conclusions les plus importantes de l'enquête – que sont la transversalité et la prépondérance de la recherche pour une activité (cf. § c) p. 181) et les utilisations multiples des textes (cf. § 4.4.5 p. 187)–, et le modèle de Yinger [1978], nous a mené à la définition de la notion de contexte pédagogique.

La notion n'est pas strictement nouvelle et des remarques pointant dans sa direction peuvent être trouvées dans la littérature¹⁴⁷, cependant nous n'avons pas trouvé de travaux où elle était conceptualisée. Il s'agit en général de cas particuliers de ce que nous discutons ici, comme c'est le cas à travers l'influence du contexte sur la notion d'authenticité chez Taylor : « *[Many] make the point that text authenticity is all very well, but much more important is what is done with the text. [...] Hutchinson and Waters make the point well when they say that Authenticity is not a characteristic of a text in itself: it is a feature of a text in a particular context.*¹⁴⁸ » [Taylor, 1994, p. 2]. Cette remarque montre doublement l'influence du contexte pédagogique dans la définition de certaines propriétés. En effet, elle montre que c'est le CP qui définit l'authenticité du texte. Mais en amont de l'évaluation de la propriété, sa définition est dépendante du vécu de l'enseignant, qui prend la propriété en compte. En effet, selon l'acception simplifiée de la notion d'authenticité que nous avons adoptée (cf. § a) p. 65), les textes considérés comme authentiques ne seront pas nécessairement les mêmes dans les mêmes contextes que pour Taylor.

Dans un contexte bien plus proche du nôtre, Alcaraz & Pérez Paredes touchent à la notion quand ils disent qu'un corpus motivé pédagogiquement devrait reconnaître le « cadre paramétrique » dans lequel les expériences d'apprentissage et d'enseignement ont lieu¹⁴⁹ [Alcaraz & Pérez Paredes, 2008, p. 29].

Appliqué aux pratiques d'utilisation de textes dans l'enseignement des langues, le CP permet, en incluant ces phénomènes au flux social du *faire* [Holloway, 2007], de les décrire comme des processus de précision du contexte pédagogique. Toutefois, la complexité inhérente à la notion même du flux de l'agir souligne une difficulté intrinsèque à la modélisation du contexte pédagogique qui s'alimente aussi bien de critères précis que d'éléments affectifs et non explicites pour les enseignants eux-mêmes, tous teintés de l'expérience de ces derniers.

Par conséquent, si à court terme l'enseignant fera par lui-même l'essentiel des relations entre les propriétés extrinsèques et intrinsèques, notre travail n'en est pas moins de prendre en compte le contexte pédagogique dans la tâche de recherche de textes. Pour ce faire, il nous faut définir un modèle qui permette de représenter l'influence du CP sur les propriétés intrinsèques des textes. Pour cela, nous allons nous appuyer sur un second questionnaire afin de tenter de mettre en évidence de tels mécanismes.

147. Voir par exemple Reffay *et al.* [2008]; Kern *et al.* [2004]; Chappelle [2004].

148. Nombreux sont ceux qui font remarquer que l'authenticité c'est très bien, mais que l'important c'est ce qui est fait avec le texte. Hutchinson & Waters [1987, p. 159] le font bien quand ils disent que l'authenticité n'est pas une caractéristique du texte en soi : c'est un trait du texte dans un contexte particulier.

149. « *[Any] pedagogically-motivated corpus should acknowledge the parametric framework where the learning and teaching experiences take place.* »

Chapitre 5

Mise en évidence de l'influence du CP sur certains critères de recherche/choix de textes

Au cours du chapitre précédent, nous nous sommes intéressé aux processus qui aboutissaient à l'utilisation d'un texte dans l'enseignement des langues. L'étude en question permettait de constater certaines tendances (recherche passive et recherche de textes dont l'usage dans l'enseignement des langues est avéré). Mais c'est en nous intéressant en particulier à la configuration la plus transversale au sein des pratiques, la recherche pour une activité, que nous avons pu introduire la notion de contexte pédagogique, qui correspond à la description de la situation d'enseignement (cf. définition 4.1a p. 211). Nous établissons une correspondance entre le processus d'inclusion d'un texte à une séquence pédagogique et la précision progressive du contexte pédagogique décrivant la situation dans laquelle il sera utilisé. La description de la tâche comme incorporée dans un flux et la mise en perspective avec le modèle de la planification de Yinger [1978] suggèrent que la manière dont le Contexte Pédagogique (CP) est formulé et dont il influence les traits qui entrent en compte dans le choix des textes est elle-même influencée par le vécu de l'enseignant (cf. § 4.6 p. 195).

Dans ce chapitre, notre point de vue évolue par rapport au précédent. Au lieu de tenter de décrire une tâche dans sa complexité, nous voulons nous rapprocher des propriétés considérées elles-mêmes et tenter de mettre en évidence certains aspects de la relation entre le CP et les propriétés du textes. Du point de vue général, nous passons directement à un point de vue très spécifique. L'objectif est d'avoir suffisamment d'information pour proposer ensuite un modèle informatisable pour la description pédagogique de textes. Nous nous appuyons, pour y parvenir, sur un second questionnaire.

5.1 Protocole et objectifs

5.1.1 Codage d'un système de génération de questionnaire

Notre premier questionnaire avait fait l'objet d'un certain nombre de critiques d'un point de vue technique. Il lui avait été principalement reproché de ne pas permettre la lecture de toutes

les questions avant d'y répondre, l'impossibilité de revenir sur ses réponses [17], ainsi que des séquences de questions qui ne prenaient pas suffisamment en compte les réponses précédentes (notamment sur la séquence concernant les textes pédagogiques [93]). Pour ce second questionnaire, prévu pour contenir beaucoup plus de questions que le premier, nous avons créé un système de gestion de questionnaires. En effet, plutôt que de coder strictement le questionnaire, et étant donné le nombre de questions prévues, il nous a paru plus pertinent de créer une structure générique permettant de définir des types de questions à inclure dans des pages ; le but étant d'avoir à notre disposition les données sous forme brute afin de pouvoir les traiter à notre guise.

Le système ainsi produit (dépourvu d'interface de gestion¹) utilisait *PHP: Hypertext Preprocessor* (PHP) et MySQL pour la génération des questions. Le PHP générait du Javascript pour l'affichage optionnel des questions, afin qu'un embranchement ne nécessite pas de changer de page². De ce fait, les questions étaient regroupées en fonction des pratiques qu'elles concernaient.

Parmi les options disponibles pour les questions « semi-fermées », il était possible, dans l'optique de catégoriser les réponses plus facilement et de multiplier les idées que les enseignants peuvent avoir, d'inclure dans les choix proposés aux enseignants les réponses « autres » des enseignants ayant répondu auparavant. L'option était à utiliser avec parcimonie, uniquement pour les traitements qualitatifs, dans la mesure où elle interdit tout traitement quantitatif.

Enfin, le système permettait la génération de données formatées pour R^3 , ainsi que les lignes de commande pour la génération d'histogrammes, le calcul de χ^2 et d'analyse de variance (*ANalysis Of VAriance* (ANOVA)).

5.1.2 Mode de diffusion

Pour ce second questionnaire, nous avons décidé de ne pas passer par des questionnaires papier, afin d'homogénéiser les conditions de passation. Nous avons contacté directement les enseignants ayant répondu à notre premier questionnaire (et qui nous en avaient donné l'accord) et ensuite transmis des appels à contribution selon les mêmes canaux de diffusion que ceux que nous avons utilisé pour le premier questionnaire. Le café pédagogique s'est à nouveau montré celui qui nous a amené le plus d'enquêtés⁴ (cf. § a) p. 150).

5.1.3 Objectifs du questionnaire

Nous allons dans un premier temps porter ici un regard global sur le questionnaire avant de nous focaliser sur certaines questions. Il est cependant disponible dans son intégralité en annexe⁵ (annexe F p. a-39). La séquence de questions, de titres et autres notes accessibles aux participants y ont été reproduits. Pour chaque question, un (ou plusieurs) tableau(x) en indique l'intitulé et les réponses fournies par les utilisateurs parmi les choix proposés. Les éventuels tableaux et figures

1. Voir documentation en annexe : E, p. a-29.

2. Ces procédures javascript étaient très basiques et furent créées quelque temps avant l'avènement de la librairie jQuery – http://docs.jquery.com/History_of_jQuery et ⁴ –, qui aurait sensiblement amélioré l'interface (barres de défilement) et nous aurait fait gagner beaucoup de temps dans l'implémentation.

3. Environnement statistique multiplateforme sous licence publique GNU.

<http://www.r-project.org/> / <http://www.gnu.org/licenses/gpl.html> .

4. <http://www.cafepedagogique.org/disci/langues/70.php#40> .

5. La version en ligne peut être consultée ici : <http://mathieu.loiseau.free.fr/bdtp/login.php>.

additionnels représentent les résultats après codage des réponses. La légende du tableau indique l'identifiant de chaque question. Il est à noter que cet identifiant n'est pas forcément représentatif de l'ordre dans lequel elles ont été posées.

La quantité de travail nécessaire à la mise en place du questionnaire était en rapport avec la quantité d'objectifs poursuivis. En effet, le questionnaire se voulait apte à nous apporter des réponses à court terme (sur les questions qui nous intéressent ici : la mise en évidence de relation entre le contexte pédagogique et des propriétés des textes), mais aussi à plus long terme pour le choix des propriétés à prendre en charge et sur la manière de les présenter aux utilisateurs dans le cadre du développement d'une première version publique. Deux points de vue cohabitaient donc au sein des questions. L'un était plus orienté vers la recherche alors que l'autre avait plus une approche d'ingénierie.

a) Population et précisions

La première partie du questionnaire (cf. § F.1 p. a-39) visait à en savoir plus sur les enquêtés, avant de revenir sur la question des recueils de textes.

b) Approche « *bottom-up* »

La seconde section (cf. § F.2 p. a-57) était la plus résolument tournée vers l'ingénierie. Au lieu de poser aux enseignants des questions sur leurs pratiques, il s'agit ici de questions d'opinion. Nous ne partons pas pour ces questions de leurs besoins pour les analyser et les modéliser (*top-down*), mais des outils utilisables à court terme dans un système de recherche de textes en demandant aux enseignants de s'imaginer confrontés à de telles possibilités afin de savoir lesquelles privilégier (*bottom-up*). Les questions sont pour la plupart fermées, mais quelques questions ouvertes ont été intercalées pour avoir une idée de comment ils envisagent de les utiliser. « L'effet pernicieux » de l'enquête d'opinion tel que mis en évidence par Bourdieu n'est pas présent en tant que tel (jouer sur la confusion entre réponse éthique et réponse politique en mettant « les gens en demeure de répondre à des questions qu'ils ne se sont pas posées » [Bourdieu, 2002, p. 226]). Il y a cependant une certaine pertinence, du point de vue de la recherche scientifique, à prendre en compte cette remarque. En effet, les réponses données sont à différencier de celles qui concernent les pratiques. Elles peuvent servir à formuler des hypothèses à confirmer par le biais d'études ultérieures, mais il y a un flou quant à leur capacité à traduire un réel besoin, puisque nous imposons en quelque sorte aux enseignants un point de vue centré sur les outils. Toute analyse de ces réponses se devrait donc de les replacer dans leur contexte.

c) Influence du type d'activité

La troisième section du questionnaire (cf. § F.3 p. a-68), qui s'étale sur plusieurs pages Web dans la version remplie par les enseignants⁵, vise à examiner l'influence du type d'activité sur certaines propriétés : la longueur du texte, le nombre d'éléments représentatifs (sur lesquels nous reviendrons), ainsi que la tolérance au vocabulaire et structures inconnus des apprenants. Pour chaque type d'activité concerné l'enseignant répond à une question décrivant sa vision de ces propriétés avant de s'exprimer sur les critères qu'il utilise pour choisir un texte pour ce type d'activité. Il est finalement invité à fournir des exemples de valeurs pour ces critères.

L'objectif de cette section répétitive de questions est de comparer les réponses entre elles. Nous avons toutefois inclus dans cette séquence les questions portant sur les activités de traduction pour lesquelles les questions posées sont différentes et détonnent quelque peu au sein de cet ensemble.

d) Influence du public

La section qui vient ensuite est beaucoup plus courte que celles qui précèdent (cf. § F.4 p. a-115). Elle revient sur certains des tests que nous n'avions pu mener dans notre enquête précédente (cf. § d) p. 158 et § b) p. 160), mais est adaptée au changement de point de vue. Il ne s'agit plus de savoir si le public influence le processus global de recherche, mais comment il affecte les propriétés des textes considérées par les enseignants. Nous demandons aux enseignants de s'exprimer sur l'impact que peuvent avoir les couples L1-L2 et le niveau sur la quantité de structures inconnues des apprenants présentes dans les textes.

e) Types de textes et suggestions

Enfin, le questionnaire se termine par deux questions indépendantes qui ne s'apparentent pas aux autres sections. La première concerne les typologies de textes (cf. § F.5 p. a-116), alors que la seconde recense les suggestions des enseignants (cf. § F.6 p. a-123).

f) Objectifs du chapitre

Comme nous le faisons remarquer en introduction de cette section, les objectifs de l'étude dépassent le cadre de ce qui nous occupe à ce stade de nos recherches. Dans ce chapitre, nous voulons montrer qu'il peut exister un lien entre certains aspects du contexte pédagogique et des propriétés des textes. Pour ce faire nous allons étudier quelques cas particuliers d'éléments du CP et quelques propriétés choisies *a priori*. L'objectif n'est donc pas d'examiner un trait de la situation pédagogique et de tenter de l'expliquer, mais de mettre en évidence certains jeux d'influence au sein des propriétés.

Nous n'allons donc pas utiliser ici l'intégralité des informations recueillies. Nous nous centrerons principalement sur les sections F.3 et F.4 (respectivement p. a-68 et a-115), qui traitent de deux éléments du contexte pédagogique : le type d'activité et le public. Nous reviendrons également sur les recueils de texte, puisqu'ils semblent constituer un bon indicateur des éléments pris en compte par les enseignants pour la phase de sélection (cf. figure 4.7 p. 206). En effet, le choix de passer par le recueil et d'en consulter telle ou telle section est relativement proche de notre problématique, puisqu'il s'agit de limiter le nombre de textes qui peuvent répondre à un besoin, en espérant sélectionner des textes dont l'évaluation révélera qu'ils sont adéquats pour la situation d'enseignement considérée.

5.1.4 Réponses

Alors que nous n'allons consulter, dans un premier temps, qu'une sous-partie des questions posées aux enseignants, nous les avons quand même soumis à un formulaire qui comportait jusqu'à 77 questions. D'après nos tests préalables avec quelques enseignants de notre connaissance, le

NB PAGES REMPLIES INTÉGRALEMENT	1	2	3	4	5	6	7	8	TOTAL
NB ENSEIGNANTS	34	9	6	5	4	2	2	35	97

TABLE 5.1: Complétude des réponses des enseignants (moyenne 50 réponses par question)

remplissage du questionnaire prenait environ 45 minutes, ce qui constitue d'après Ghiglione & Matalon, la limite pour « un questionnaire composé en majorité de question fermées [...] lorsque la passation se fait dans de bonnes conditions, c'est-à-dire au domicile du sujet ou dans un lieu tranquille. » [Ghiglione & Matalon, 1978, p. 100]. Notre questionnaire satisfaisait ces conditions dans la mesure où l'enseignant pouvait y répondre sur son ordinateur personnel.

Cependant, malgré nos tentatives d'expliquer notre démarche et de répondre aux questions et commentaires des enseignants au fur et à mesure du recueil de données⁶, la répétitivité du questionnaire a découragé bon nombre d'entre eux de répondre à l'intégralité du questionnaire. Les fonctionnalités du système de gestion de questionnaires liées au confort d'utilisation⁷ ont pu appuyer ce phénomène : la contribution n'est plus binaire et les réponses aux premières questions sont sauvegardées même si l'enseignant ne répond pas aux suivantes. Toutefois, une étude des non réponses est rendue possible du fait de ces fonctionnalités. Quelles qu'en soient les raisons le nombre de réponses est indiqué dans le tableau 5.1 de la présente page.

Le tableau réutilise les données utilisées pour montrer à l'utilisateur dans quelles pages il a oublié des questions, d'où l'unité employée. Le problème de cette unité est qu'elle n'a pas toujours la même valeur. Par exemple la page 2⁸ contient beaucoup plus de questions (au minimum 14 questions) que la page 8⁹ (au maximum 3 questions). De plus, quand un enseignant remplit 4 pages intégralement, il ne s'agit pas nécessairement des 4 premières. Enfin, un enquêté ayant répondu partiellement à chacune des pages, n'est pas ajouté au compte alors que quelqu'un qui n'a rempli que la première l'est.

Enfin, un effet de bord de la tentative de faciliter le remplissage pour les enseignants ayant répondu au premier questionnaire est que les questions communes aux deux questionnaires contiennent beaucoup de témoignages. Les résultats des questions communes aux deux questionnaires¹⁰ étaient directement transférées du premier vers le second, donnant dans la présentation des résultats l'illusion d'un très grand nombre de références aux premières questions par rapport aux suivantes. Ce n'est qu'à partir de la question 12 (cf. tableau F.12 p. a-55) que toutes les réponses appartiennent exclusivement à la seconde étude.

5.1.5 Récapitulatif

Notre second questionnaire avait des objectifs d'ordre tout à fait différents. Nous voulions d'une part suivre des pistes à court terme (du point de vue de nos recherches) pour avancer dans notre compréhension du contexte pédagogique. Mais en parallèle nous posions aux enseignants

6. Cf. <http://mathieu.loiseau.free.fr/bdtip/faq.php> .

7. Les améliorations de confort par rapport au premier questionnaire comprenaient la possibilité de revenir sur ses réponses, d'interrompre et de reprendre le remplissage du questionnaire, ainsi que le stockage et l'affichage de l'état d'avancement de l'enquêté par rapport à la séquence globale.

8. http://mathieu.loiseau.free.fr/bdtip/q2.02_outilsTAL.php

9. http://mathieu.loiseau.free.fr/bdtip/q2.08_types.php

10. Voir § F.1 p. a-39 (à l'exception de la question 11).

des questions qui relevaient plus de l'ingénierie ou vouées à servir de base à la mise en place de nouvelles études. Dans le cadre de ce chapitre, nous nous contenterons d'exploiter les réponses aux premières questions.

Du point de vue de ce chapitre, cette volonté de couvrir de multiples aspects nous a desservi dans le sens où le questionnaire, dont le remplissage aurait déjà été long et fastidieux sans les questions visant une exploitation à plus long terme, s'est avéré parfois dissuasif (affichage du questionnaire sans répondre, abandon en milieu de passation). Nous nous trouvons cependant avec suffisamment de réponses pour procéder aux analyses qui nous intéressent.

5.2 Recueils de textes

Dans le § 4.1.5 p. 143, nous expliquions que le fait de se pencher sur la question des recueils permettait d'en savoir plus sur les critères de recherche des enseignants. En avançant dans notre modélisation du processus de recherche, il est apparu que ces critères constituent un ensemble de propriétés assignées *a priori* par les enseignants qui leur servent d'aide dans leurs recherches de textes ultérieures pour opérer à la sélection de textes à évaluer. Cependant notre première tentative d'analyser les réponses des enseignants dans ce domaine (cf. § 4.4.2 p. 173) nous a vu nous heurter à des difficultés d'interprétation, notamment concernant la notion de thème, qui bien que revenant très souvent dans les témoignages ne semble pas toujours véhiculer le même sens. Pour tenter de dépasser cet écueil, nous avons modifié quelque peu la structure de la séquence de questions.

Dans le premier questionnaire nous posions les questions suivantes :

Question 9 : Avez-vous un recueil de texte personnel ?

(réponses cf. tableau F.9 p. a-47)

Question 10 : Ce recueil est-il organisé ?

(réponses cf. tableau F.10 p. a-47)

Question 11 : (Si oui,) Comment ce recueil est-il organisé ?

(réponses cf. tableau F.11 p. a-49)

Nous incluons également ces questions au second questionnaire, pour les enquêtés qui n'auraient pas répondu au premier, mais aussi pour permettre aux enseignants participant aux deux études de prendre en compte d'éventuels changements dans leurs pratiques. Enfin pour faciliter notre interprétation des réponses et pour avoir accès aux critères employés non seulement selon des termes génériques, mais aussi selon les valeurs qu'ils peuvent prendre, nous avons ajouté la question suivante :

Question 12 : Afin de préciser cette organisation et de donner une idée de la manière dont vous les formulez, pouvez-vous donner un exemple des catégories employées pour le classement de ce recueil ?

Si vous désirez hiérarchiser vos exemples vous pouvez le faire de la manière suivante :

Dans cet exemple, les présents simple et progressif sont regroupés sous la rubrique présent faisant elle-même partie de la rubrique grammaire.

* Civilisation

+ Civil rights movement

+ Halloween

+ Sundance film festival

* Grammaire

- + Présent
 - Présent progressif
 - Présent simple
 - + Possessifs
 - + Pronoms personnels
- (réponses cf. tableau F.12 p. a-55)

5.2.1 Possession et organisation de recueils

Avant d'entrer dans le détail de leur organisation, nous pouvons constater que les proportions sont relativement proches entre le premier et le second questionnaire du point de vue des proportions de recueils, organisés ou non. Ce qui est d'autant plus cohérent qu'un certain nombre d'enseignants ont répondu aux deux questionnaires, même si 5 enseignants « changent d'avis » entre le premier et le second questionnaire. Globalement dans le second questionnaire, les enseignants possèdent moins souvent des recueils – 50,8% contre 57,7% –, mais ces derniers sont plus souvent organisés – 63% contre 60% – (cf. tableaux F.9 p. a-47, F.10 p. a-47 et 4.9 p. 173).

5.2.2 Codage des données

Afin de pouvoir savoir quels critères de tri sont ceux qui reviennent le plus souvent, nous allons reprendre les témoignages et les catégoriser. En explicitant cette opération, nous allons dresser un panorama des réponses effectuées par les enseignants.

En nous appuyant sur les réponses des enseignants disponibles dans les tableaux tableau F.11 p. a-49 et tableau F.12 p. a-55, nous avons établi une liste de catégories selon laquelle nous avons regroupé les témoignages. Les résultats sont indiqués dans le tableau 5.2 page suivante.

Les termes employés pour chaque catégorie sont empruntés aux réponses des enseignants et ne correspondent donc pas nécessairement au sens qui leur serait donné par un didacticien ou un linguiste.

Nous indiquons dans les sections suivantes comment nous sommes arrivé à cette catégorisation des témoignages.

a) Activité

Étant donnée l'importance de l'activité dans la tâche de planification (cf. [Clark & Yinger, 1979; Tochon, 1989; Shavelson & Stern, 1981], § 4.6.4 p. 205), il n'est pas surprenant de constater qu'un certain nombre d'enseignants a recours à ce critère pour l'organisation des recueils. Il n'est cependant pas très utilisé (seulement 7 enseignants sur 57).

Quand nous parlons d'activité, il s'agirait plus de classes d'activité, puisqu'il s'agit en général d'opposition entre les quatre habiletés : « Production / Expression orale et écrite (parler et écrire) sont souvent opposées à compréhension orale et écrite (écouter et lire), la réunion de ces quatre habiletés constituant la notion générale de compétence linguistique » [Robert, 2008, p. 170]. On peut citer par exemple le témoignage [201] :

- « documents pour la compréhension écrite, classés par niveau (6^e, 5^e,...) » ;
- « idem pour la compréhension orale, l'expression orale et écrite ».

QUESTION	Critères d'organisation du recueil :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	PROPORTION
Thème		68,4% 39/57
Niveau		36,8% 21/57
Contenu ou objectif linguistique :		41,2% 24/57
• syntaxe		35,1% 20/57
• lexique		7% 4/57
• indéterminé		7% 4/57
Séquence		14% 8/57
Activité		12,3% 7/57
Type de ressource		12,3% 7/57
Compétence		8,8% 5/57
Public		5,3% 3/57
Langue		1,8% 10/57
Date d'écriture		1,8% 1/57
Objectif académique		1,8% 1/57
57 personnes ont répondu à cette question et ont choisi en moyenne 2,2 solutions par personne. Les témoignages [38,58,75,138] ne figurent pas dans les résultats ci-dessus. La répartition des témoignages entre les diverses catégories sont récapitulés dans le tableau 5.4 p. 230		

TABLE 5.2: Récapitulatif des réponses aux questions 11 et 12

Parmi les témoignages faisant état des activités, trois sont ambigus [114,131,164] : ils semblent faire état de la présence d'activités dans le recueil. En reprenant la terminologie du *Learning Object Metadata* (LOM) (cf. § 3.3.1 p. 98), dans leurs recueils se côtoieraient des objets pédagogiques de niveau d'agrégation 0 (textes) et 1 (activités). Nous les incluons donc aussi dans la catégorie « type de ressource » (voir p. 228), puisque les activités entrent en compte dans le classement, mais aussi le type d'activité : « exercice de grammaire selon les thèmes » [114] ou « élémentaire (documents simples + act. progression vocabulaire, grammaire, syntaxe) » [131].

b) Séquence

L'enseignant [131] déclare donc prendre en compte la progression comme facteur de classement, comme critère d'ordonnement. Nous dénombrons 6 enseignants qui font une mention explicite à l'intégration d'un texte dans le déroulement de la période (trimestre, semestre, année, stage, etc.). Ils s'affranchissent même parfois du terme niveau : « Progression grammaticale + réactivation du vocabulaire » [19]. Selon l'enseignant, la séquence peut concerner une séance ou une unité (« Dans chaque thème : par séquence possible : ex : un document iconographique, un texte, un document audio... » [53]) ou avoir une portée beaucoup plus globale (« Par compétences et par thèmes et par programmations (prise en compte de la progression sur l'année) » [267]).

La progression est liée à une augmentation du niveau des apprenants. Certains témoignages sont ambigus à ce sujet, par exemple : « en fonction des niveaux de difficulté, des degrés d'apprentissage » [6]. En effet ce degré d'apprentissage peut être attribué à un groupe, pour toute une année ou un semestre. Mais il peut aussi évoluer pour un groupe donné tout au long de la période

concernée par l'enseignement, la progression étant considérée comme critère du recueil.

Nous rapprochons le témoignage [75] de cette catégorie. Nous ne pouvons être certain si son tri par « unités » introduit une notion de progression ou une articulation chronologique. Toutefois, regrouper les textes par unité d'enseignement suggère que les documents au sein d'un groupe peuvent être utilisés sous forme de séquence.

Au total, 8 enseignants déclarent prendre en compte ce que nous avons appelé séquence et qui regroupe des points de vue bien différents, mais ont en commun le fait d'ordonner les textes. Il ne s'agit plus seulement de les regrouper, mais aussi de les ordonner au sein d'un groupe :

- [131] ;
- « Selon les points de langue qui seront abordés avec les élèves , *dans l'ordre de difficulté croissante* » [43] ;
- « Selon les difficultés d'apprentissage ou plutôt selon la progression de la classe » [27] ;
- « J'ai 5 classeurs classés par niveaux... I pour Primaire-initiation, II pour Collège-Lycée, et toujours du plus « basique » au plus complexe... » [109] ¹¹.

c) Niveau

Le témoignage de l'enseignant [109] fait intervenir le niveau comme critère de classement primaire. C'est le second critère de tri le plus employé (cf tableau 5.2 p. 222). Nous avons regroupé dans cette catégorie, toutes les réponses qui prenaient en compte d'une manière ou d'une autre le niveau de langue des étudiants. Dans 21 cas, l'intitulé « niveau » est clairement indiqué [2,5, 6,8,66,72,73,82,103,104,109,111,131,137,172,194,201,208,209,233,236]. Nous regroupons tous ces témoignages sous le même intitulé, même si l'échelle employée ou l'acception du niveau varie d'un témoignage à l'autre. Certains n'indiquent que « niveau » dans leur réponse, d'autres font référence au niveau d'apprentissage ([5] par exemple), au niveau d'enseignement [73], au niveau de langue [104], au « niveau de classe » [111,209], au niveau d'études [82] ou encore au « degré de d'apprentissage » [6]. D'autres, enfin, utilisent le cadre de référence européen [236] ou définissent leurs propres niveaux « Tranches de 50 heures environ pour niveau 1 » [131].

Nous nous sommes posé la question de rapporter aussi les réponses traitant de « niveau de difficulté » à cette catégorie [6,38]. En effet, dans les autres cas, les enseignants parlent plus du niveau de leurs apprenants que d'un niveau du texte. Mais l'enseignant [6] précise le niveau de difficulté à travers le lien que cela entretient pour lui avec le « degré d'apprentissage ». Son témoignage a donc été inclus dans cette catégorie. L'interprétation du témoignage de l'enseignant 38 se révèle hasardeuse dans la mesure où ses réponses sont les suivantes :

QUESTION N°11	par dates
QUESTION N°12	l'organisation est faite par niveau de difficulté

TABLE 5.3: Témoignage [38], réponse aux questions 11 et 12

Il est difficile de savoir si le « par dates » correspond à la date d'écriture des textes et l'organisation par niveau de difficulté (du texte) ou s'il signifie que le recueil est trié par date d'utilisation du texte. Dans le second cas, il s'agirait finalement d'un classement par progression, la date correspondant à la période de l'année où le texte est supposé être utilisé en classe. La progression de l'année reflétant une augmentation du niveau des étudiants, le niveau de difficulté du texte

11. Nous *soulignons*.

(ou de l'activité dans laquelle il est utilisé) serait alors accru en accord avec cette dernière. Ne sachant pas si nous devons classer ce témoignage ici ou dans la catégorie « séquence », nous excluons ce témoignage de nos statistiques. En effet, il ne s'agit pas ici d'un témoignage qui satisferait d'une certaine manière deux catégories différentes, mais de deux interprétations d'un même énoncé s'excluant mutuellement. Enfin le témoignage [247] n'est pas comptabilisé ici non plus, nous expliquerons pourquoi dans la section d).

d) Public

Il nous a semblé nécessaire de créer une catégorie « public » dont le « niveau » ne soit pas une sous-catégorie (bien que nous n'excluons pas que dans l'absolu ce soit le cas). Dans « public » tel que nous l'entendons ici, le niveau d'apprentissage des apprenants dans la langue enseignée ne sera pas pris en compte. Tout critère de tri qui pourra être considéré comme une caractéristique d'un groupe d'apprenants donné sans être son niveau sera regroupé sous l'étiquette « public ». Prenons un exemple fictif ; un enseignant trie son recueil de la manière suivante :

- sujet
 - technologie
 - arts
 - environnement

S'il enseigne à des étudiants issus de filières scientifiques et de filières littéraires, il est possible de considérer que les textes traitant de technologie seront plus particulièrement destinés à ses étudiants de filières scientifiques et que les textes traitant d'arts iront plus particulièrement à ses étudiants de lettres et en conclure qu'il prend en compte le public (indépendamment de son niveau) dans son classement. À travers cet exemple, nous pouvons à nouveau remarquer le lien existant entre 2 catégories, ici « thème » et « public ». En effet dans cet exemple, le thème des articles est directement influencé par le public.

Dans le cadre de notre questionnaire, seuls trois témoignages font état d'une telle organisation [14, 41, 247]. L'enseignant [14] prend en compte la filière d'enseignement dans son classement. L'enseignant [41] consacre une sous-partie de son recueil pour les cours de traduction qu'il dispense à l'université. Nous avons considéré que ces textes étaient destinés à un public de spécialistes de l'allemand (langue enseignée par l'enseignant [41]) dont le centre d'intérêt du point de vue de l'apprentissage de cette langue est, en ce qui concerne l'enseignant, la traduction¹². Enfin, alors que l'enseignant [247] mentionne le niveau dans les critères d'organisation de son recueil, son témoignage complet (d'enseignant à l'université) fait état d'un recueil par cours, « 5 ou 6 recueils différents pour le présentiel, 2 en EAD (enseignement à distance) ». En conservant la même logique que pour l'enseignant [41], c'est le « cours », c'est-à-dire un groupe d'étudiants intéressés dans leur apprentissage des langues à une problématique particulière (la langue de spécialité ici), qui détermine l'organisation du recueil. De la même manière que pour l'enseignant [41], la distinction faite entre les cours en présentiel et à distance, pourra aussi être comptabilisée comme la prise en compte du public à travers les contraintes matérielles de la situation d'enseignement.

12. Peut-être aurions nous pu considérer qu'il s'agissait d'un critère selon le type d'activité. Cependant, dans ce cas-ci, l'enseignant n'utilise ces textes que pour une filière donnée.

e) **Thème**

Le critère de tri qui guidait notre exemple fictif est celui du thème, le plus revendiqué, qui a été évoqué par 39 enseignants sur 62 (cf. tableau F.11 et tableau F.12). Parmi ces enseignants, 15¹³ précisent ce critère dans le cadre de la question 11 ou 12, les autres l'utilisent sans le définir.

Si nous nous référons au Cadre Européen Commun de Référence pour les Langues (CECR), un *thème de communication* est défini de la manière suivante :

Définition 5.1 (thème (de communication)) « un thème privilégié pour des actes de communication particuliers; c'est autour d'eux que s'articulent le discours, la conversation, la réflexion ou la rédaction. » [Conseil de l'Europe, 2000, p. 45]

Le CECR reprend le travail effectué dans van Ek & Trim [1991], auquel nous nous référons dans la section d) page 172. Il fait état de quatorze catégories thématiques. Elles sont elles-mêmes divisées en sous-thèmes, pour chacun desquels des notions spécifiques sont identifiées. Par exemple, la catégorie « congés et loisirs » contiendra le sous-thème « sports » qui fera intervenir parmi les notions spécifiques qui lui sont associées, des « objets : cartes, ballon ».

L'objet de notre étude n'est pas de privilégier une approche par rapport à une autre ou d'évaluer l'intégration du CECR dans les pratiques des enseignants, mais bien de nous fonder sur les réponses de ces derniers. Il semble que dans la majeure partie des témoignages, ce qui est désigné par thème correspond à la définition qui en est donnée par le CECR. Cependant, pour notre échantillon, le recours à la hiérarchisation en thèmes et sous-thèmes et aux notions spécifiques n'intervient pas dans l'organisation des recueils. Les catégories mentionnées semblent d'ailleurs plus relever des sous-catégories du CECR. D'après les réponses des 15 enseignants qui ont explicité leur classement, il s'agit en général de sujets culturels ou de civilisation. Toutefois, parmi ces enseignants, certains ajoutent une dimension de focalisation sur la forme : les enseignants [198] et [209] parlent de « thèmes de grammaire ».

Il est plus difficile d'apprécier les réponses des enseignants [24] et [114]. L'enseignant [24] ne propose que le thème comme critère de tri de son recueil. Pourtant, à la question suivante, il commence par préciser les différentes composantes de la catégorie « grammaire » avant de mentionner « Thèmes par textes ou articles, documents audio/video, exercices ». Difficile de savoir si la grammaire est l'unique thème du recueil, ou n'en est pas un. Viennent ensuite des catégories « langue générale et de spécialité ». Pour l'enseignant [114], la difficulté d'interprétation est relativement analogue. Seul le thème est proposé comme critère de tri. Dans la question suivante, des exemples sont donnés : ils correspondent à la définition que nous avons donnée plus haut. Interviennent ensuite les catégories « Grammaire », « Vocabulaire », « Chanson », « Textes pour le bac », puis « exercices de grammaire selon les thèmes » et enfin « Miscellaneous (par thèmes[...]) ». La question est donc de savoir si l'enseignant reparle de thèmes parce que dans la description précédente certaines catégories ne s'y rapportaient pas ou si « Textes pour le bac » est considéré par l'enseignant comme un thème au même titre que « Grands hommes ».

Ces deux témoignages utilisent donc très probablement, à l'instar des témoignages [198] et [209], une acception plus large de *thème* tout en n'excluant pas la part « civilisationnelle et culturelle » inhérente au thème de communication. En effet, la notion même de thème grammatical ou le fait de considérer la « grammaire » comme un thème semble aller à l'encontre de la définition 5.1, selon laquelle, la langue est au service d'un besoin communicationnel. Dans le CECR, la grammaire

13. [22, 24, 32, 41, 53, 66, 73, 77, 114, 131, 172, 198, 208, 209, 236]

serait un moyen et non une fin dans le traitement du thème. Nous nous proposons donc dans la suite de ce travail de ne désigner par thème que la notion correspondant à la définition du CECR et de regrouper les informations relatives au « thème grammatical » dans une autre catégorie.

f) Contenu ou objectif linguistique

La forme des textes est une autre catégorie de réponses très représentée dans les témoignages des enseignants. Nous appelons cette catégorie « contenu ou objectif » linguistique, car les témoignages ne permettent pas d'établir de manière précise si les textes ainsi regroupés le sont parce qu'ils contiennent un certain fait de langue ou parce que l'enseignant les considère propices au travail sur ce fait de langue. Par exemple, si l'objectif linguistique est « l'introduction du présent simple en anglais », rien n'indique la stratégie employée. L'activité en question peut s'appuyer sur un texte présentant en contraste des occurrences de présent simple avec des occurrences de présent progressif déjà connu. Si tel est le cas, le texte devra présenter des phrases contenant du présent simple et du présent progressif. Le « contenu linguistique » du texte attendu pour l'activité dépendrait donc de la forme de l'activité et non uniquement de son objectif. Les réponses des enseignants n'entrant pas dans ce type d'explication, nous considérerons qu'ils peuvent parler aussi bien du contenu du texte que de l'objectif de l'activité dans laquelle il sera utilisé.

Quand nous parlons de contenu linguistique, il s'agit d'un choix explicite comme critère de tri. En effet, il est à l'origine de la plupart des critères de tri (si ce n'est tous) : le thème, tout comme le niveau, la langue, les compétences reflètent certains aspects du contenu du texte. La focalisation sur la forme n'est cependant explicite que dans certains cas, avec en plus l'incertitude de ce sur quoi porte l'intitulé, contenu ou objectifs.

En revanche, la plupart des enseignants indique clairement si le tri en fonction de phénomènes explicitement linguistique des textes se fonde sur les structures grammaticales ou sur le vocabulaire présents dans le texte. Nous avons donc décidé de distinguer ces deux champs¹⁴. Dans la suite de ce paragraphe, nous parlerons de contenu syntaxique, quand il s'agit des structures employées et du contenu lexical quand il s'agit des mots employés. Dans ce contexte, seule « par objectif linguistique » [79] ne permet pas de faire la différence entre ces deux aspects.

Le critère grammatical est nettement plus représenté que le critère purement lexical. Cette représentation pourrait être attribuée au tri par thème, qui recoupe partiellement l'accès aux textes via le lexique utilisé. En effet, deux des quatre enseignants qui trient leurs recueils en fonction des unités lexicales présentes dans les textes n'ont pas mentionné le thème comme critère de tri [19,125]. L'enseignant [103] met le « vocabulaire » au même niveau que le thème. Seul l'enseignant [114] différencie explicitement thème et contenu lexical du thème, cependant nous avons constaté dans la section e) (p. 225) que sa classification était ambiguë quant à la valeur de la notion de thème¹⁵.

Tout en l'exprimant de manière différente, 19 enseignants déclarent trier leurs recueils en fonction du contenu syntaxique des textes. Il déclarent organiser leurs textes en fonction :

- de la « progression grammaticale » [19] ;
- des « moyens » [125] ;

14. Même si nous sommes bien conscient que ces deux aspects ne peuvent être strictement séparés, comme le signale Johns, quand il décrit l'utilisation du *Data Driven Learning* (DDL) comme particulièrement efficace à la « frontière collocationnelle entre syntaxe et lexique » [Johns, 2002, p. 109] (cf. note ⁹³ p. 71).

15. Les deux témoignages sont inclus dans les tableaux F.11 p. a-49 et F.12 p. a-55.

- des « points de grammaire » [22, 42, 130, 194] ;
- des « points de langue » [43] ;
- des « points grammaticaux » [71] ;
- des « faits linguistiques présents dans les textes » [32] ;
- ou enfin des « notions grammaticales » [111].

D'autres formulations existent. Plusieurs enseignants parlent d'une catégorie « grammaire » dans leurs recueils [24, 103, 109, 114, 201], couplée avec une catégorie « verbes » pour l'enseignant [109], alors que l'enseignant [47] adresse le problème en terme d'« unités syntaxiques ». Conformément aux remarques faites dans le paragraphe e) (p. 225), nous regroupons ici les témoignages utilisant un quelconque « thème grammatical » comme critère de tri [117, 198, 209].

Le témoignage [27] est difficile à classer : l'enseignant déclare proposer après quelques cours des textes plus longs dans lesquels se retrouvent « les principales règles de grammaire déjà vues ». Les « règles de grammaire » exemplifiées par les textes en question constituent donc l'une des raisons qui placent les textes dans telle ou telle catégorie. Cependant, le critère de tri qu'il déclare utiliser est la « progression de la classe ». Il y a donc ici deux composantes à son organisation : c'est la progression qui est au cœur du mode de tri, l'enseignant semble néanmoins organiser cette dernière en termes de « contenu ou objectif linguistique ». L'indiscutable lien fait par l'enseignant entre la progression de la classe et des questions de forme nous mène à inclure ce témoignage à cette catégorie (cf. § b) p. 222). Nous l'avons toutefois également associé au critère qui nous concerne ici, ce qui est plus discutable, puisque nous pourrions considérer que c'est la manière dont l'enseignant perçoit son tri qui doit primer. Malgré tout, dans l'optique de constater les critères employés par les enseignants, il nous paraît pertinent de savoir combien d'enseignants font des mentions explicites à la forme des textes.

L'enseignant [164] organise lui aussi ses textes en fonction de leur contenu linguistique (ou des objectifs syntaxiques ou lexicaux qu'il peut poursuivre en les réutilisant), il donne comme exemple de catégories :

- + Le style indirect
- + La voix passive
- + Les comparatifs et superlatifs
- + Les temps
 - Les présents
 - Les temps du passé

Ces catégories correspondent clairement à ce que nous avons désigné dans ce paragraphe comme « contenu ou objectif linguistique ». Pourtant l'enseignant parle de « Compétence linguistique », ce qui encore une fois pose un problème de point de vue. La dimension primordiale est elle la notion de compétence ou la dimension linguistique ?

g) Compétence

L'une des notions au centre de l'approche communicative est celle de « compétence de communication ». Comme le souligne Hymes, la définition de la notion n'est pas consensuelle :

« Le fait que la notion de “compétence de communication” ait connu une certaine audience ne suffit pas à garantir une conception commune de la signification qu'on lui donne. »

[Hymes, 1984, p. 177]

Pour Canale & Swain, la notion de « compétence de communication » se décompose en trois secteurs de compétences : la compétence grammaticale, la compétence sociolinguistique et les stratégies de communication ou compétence stratégique [Canale & Swain, 1980, p. 27]. Palmer *et al.* [1981], quant à eux, distinguent d'abord entre parler et lire et reconnaissent, dans un cas comme dans l'autre, une compétence linguistique, sociolinguistique et pragmatique. [...] Une quatrième dimension, la fluence, fut laissée de côté parce qu'elle paraissait impossible à apprécier à l'aide de techniques majeures d'évaluation [Hymes, 1984, p. 183].

Moirand propose elle aussi une vision différente de la compétence de communication, qu'elle décompose en quatre composantes :

«

- une composante linguistique, c'est-à-dire la connaissance et l'appropriation (la capacité de les utiliser) des modèles phonétiques, lexicaux, grammaticaux et textuels du système de la langue ;
- une composante discursive, c'est-à-dire la connaissance et l'appropriation des différents types de discours et de leur organisation en fonction des paramètres de la situation de communication dans laquelle ils sont produits et interprétés ;
- une composante référentielle, c'est-à-dire la connaissance des domaines d'expérience et des objets du monde et de leurs relations ;
- une composante socioculturelle, c'est-à-dire la connaissance et l'appropriation des règles sociales et des normes d'interaction entre les individus et les institutions, la connaissance de l'histoire culturelle et des relations entre les objets sociaux.

»

[Moirand, 1982, p. 20]

La diversité des descriptions des notions et la finesse des différences nous interdit de retenir une acception plutôt qu'une autre pour désigner les témoignages des enseignants, qui la plupart du temps se contentent de faire allusion à la notion [72, 122, 233]. Comme nous l'avons vu, l'enseignant [164] se centre particulièrement sur la composante discursive de Moirand. L'enseignant [190] détaille sa perception de la notion à travers quelques exemples (cf. tableau F.12 p. a-55) et semble considérer un plus grand nombre de composantes de la « compétence de communication ».

Le but de cette catégorisation n'est pas d'évaluer une acception de « compétence de communication » par rapport à une autre, mais bien de regrouper les témoignages qui utilisent les compétences comme critère de tri de leurs recueils.

h) Type de ressource

Plusieurs enseignants font intervenir, plus ou moins explicitement, le type de ressource dans leur classement. Il ne s'agit pas ici d'un « type de texte » mais bien du type de ressource, puisque cela décrit plus souvent la source du texte que réellement le type de texte. En effet, les recueils des enseignants ne contiennent pas nécessairement que du texte. Par exemple, l'enseignant [62] distingue les textes en fonction de « catégories (presse, littérature, iconographie...) ». Dans le témoignage de l'enseignant [109], le type de document intervient à différents niveaux du classement. Il possède un recueil consacré à l'oralité qui est subdivisé entre chansons, jeux et exercices de phonétique. Il dispose aussi d'un recueil de dessins dans lequel sont séparées les bandes dessinées, des publicités. Nous pouvons aussi citer les témoignages des enseignants [24] (Classement par thème puis en type de ressource : « audio/vidéo ») ou [236] qui fait état d'une sous-partie de son recueil consacré à la bande dessinée.

Les enseignants qui cumulent dans leurs recueils plusieurs types de ressources ne l'indiquent cependant pas toujours comme mode de tri. Nous pouvons citer l'exemple des enseignants [53], [190] et [201]. L'enseignant [190] indique bien que son recueil est plus « un recueil de ressources diverses que de textes », mais ne fait pourtant allusion à aucun moment au type de ressource comme un mode de classement. L'enseignant [201] parle de « documents pour la compréhension écrite, classés par niveau [...] » et les sépare des documents « pour la compréhension orale, l'expression orale et écrite ». Un tel classement suggère la présence de ressources audio, notamment pour la compréhension orale, mais le support n'entre pas en ligne de compte pour le classement, c'est le type d'activité dans laquelle la ressource sera utilisée qui est discriminant. En effet, un texte écrit peut être support d'une activité de compréhension orale (ex : *vire-langue*). Enfin l'enseignant [53] explique que pour chaque thème dans son recueil un classement par séquence possible est opéré. Il donne ensuite un exemple de « séquence possible » : « un document iconographique, un texte, un document audio... ». Une fois de plus le type de ressource n'est pas un critère d'organisation du recueil, il suggère cependant que son recueil comporte des ressources non textuelles.

i) Autres catégories

Les 3 dernières catégories ne concernent à chaque fois qu'un enseignant. Elles ne peuvent cependant pas être ramenées à d'autres déjà citées.

Date d'écriture du texte : Outre l'enseignant [38], dont nous évoquons le cas dans le paragraphe Niveau page 223, un seul enseignant [47] déclare trier ses textes en fonction de l'époque à laquelle ils ont été écrits. L'enseignant déclare néanmoins que ce tri ne concerne pas tous les textes (« par siècle, de temps en temps »).

Langue : Alors que parmi les enseignants ayant donné l'organisation de leur recueil de textes, 10¹⁶ enseignent plusieurs langues, seul l'enseignant [52] signale un tri selon ce critère. Nous ne pouvons donc savoir si leurs recueils mélangent au sein des différentes catégories des textes écrits dans différentes langues ou si le tri par langue est tellement évident que les enseignants n'ont pas jugé bon de le signaler, s'ils possèdent un recueil par langue.

Objectif académique : Par objectif académique nous désignons toute classification de textes directement liée à un examen, diplôme ou certification donné. Seul l'enseignant [114] rentre dans cette catégorie, il dédie une partie de son recueil au baccalauréat.

5.2.3 Non pris en compte

Nous avons dû prendre de nombreuses décisions quant à la classification des témoignages, qui demandaient parfois beaucoup d'inférences ; malgré tout, nous avons manqué d'information pour classer 4 témoignages.

- l'exclusion du témoignage [38] est expliquée dans le paragraphe c) p. 223 ;
- la réponse « fonctions » [17] n'est pas suffisamment claire pour être prise en compte ;

16. [5, 6, 8, 24, 52, 131, 134, 194, 198, 267].

- l'enseignant [58] trie par « objectifs visés », mais les informations contextuelles nous manquent pour savoir si cet enseignant exprime ses objectifs en termes de notion grammaticale, de compétence de communication ou de niveau assorti à la programmation annuelle du cours ;
- enfin « en rapport avec la méthode employée » [133] n'était pas interprétable. Veut-il dire par là qu'à chaque unité du manuel employé correspond une partie de son recueil ? que le recueil est trié selon l'ordre de progression du manuel ? ou encore que dans une perspective éclectique, cela dépend de la méthode d'enseignement employée pour l'activité faisant intervenir le texte ?

5.2.4 Des catégories à dominante liée à la situation d'apprentissage

Dans cette énumération, nous récapitulons les décisions prises dans le tableau 5.4 de la présente page.

CATÉGORIE	IDENTIFIANTS DES ENSEIGNANTS CONCERNÉS
Thème	[8], [14], [20], [22], [24], [32], [41], [42], [47], [52], [53], [62], [66], [69], [71], [73], [77], [79], [82], [83], [90], [92], [103], [104], [114], [117], [122], [131], [134], [137], [172], [173], [198], [202], [203], [208], [209], [236], [267]
Contenu ou objectif linguistique : <ul style="list-style-type: none"> • Syntaxe • Lexique • Indéfini 	[19], [22], [24], [27], [32], [42], [47], [71], [103], [109], [111], [114], [117], [125], [130], [164], [194], [198], [201], [209] [19], [103], [114], [125] [43], [79], [122], [127]
Compétence	[72], [122], [164], [190], [233]
Activité	[109], [114], [131], [164], [192], [201], [236]
Type de ressource	[24], [62], [109], [114], [131], [164], [236]
Séquence	[19], [27], [43], [53], [75], [109], [131], [267]
Niveau	[2], [5], [6], [8], [66], [72], [73], [82], [103], [104], [109], [111], [131], [137], [172], [194], [201], [208], [209], [233], [236]
Public	[14], [41], [247]
Langue	[52]
Date d'écriture du texte	[47]
Objectif académique	[114]
Non pris en compte	[17], [38], [58], [138]

TABLE 5.4: Catégorisation des témoignages des enseignants répondant aux questions n°11 et 12 du second questionnaire

À travers cette étude, nous avons pu constater les principales catégories qui permettent aux enseignants d'organiser leurs textes. Elles semblent donc correspondre aux propriétés assignées aux textes pour la phase de sélection (cf. figure 4.7 p. 206).

Il est frappant de constater qu'à l'exception de la langue et de la date d'écriture du texte¹⁷, tous

17. Dans un cas comme dans l'autre, il serait possible d'ajouter ces traits au CP, puisqu'un cours de français

les critères invoqués désignent la situation d'enseignement dans laquelle le texte sera utilisé. Bien évidemment il existe des liens entre la situation d'enseignement et les propriétés du textes, comme nous le faisons remarquer au moment de notre indécision quant au titre de catégorie « contenu ou objectif linguistique ». Cependant pour un témoignage « selon les faits linguistiques présents dans les textes » [32], la plupart des enseignants parlent de « points grammaticaux » [42] ou d'objectifs linguistiques [79].

Cette étude appuie donc notre point de vue selon lequel les enseignants articulent leurs recherches autour des besoins de la situation d'apprentissage. En effet, ces derniers regroupent souvent leurs textes selon un ou plusieurs aspects de ce que nous avons appelé le contexte pédagogique. L'attribution de ces propriétés aux textes s'appuie nécessairement sur leur contenu mais c'est rarement la dimension qui en est retenue. Les degrés de précision de l'organisation des recueils varient d'un enseignant à l'autre, et la répartition en classes d'organisation est complexe et sujette à approximations étant donné les liens qu'entretiennent les critères de tri entre eux.

Ces liens peuvent relever de problèmes de dénomination. La notion de « thème » par exemple, qui d'après les prescriptions du CECR [Conseil de l'Europe, 2000] est articulée autour d'axes de communication (cf. définition 5.1 p. 225). Comme nous le faisons remarquer, la définition du CECR considère les outils grammaticaux comme des moyens et non comme des fins en soi, excluant ainsi les « thèmes grammaticaux » de la catégorie thème. Toutefois, dans les exemples de thèmes développés dans [van Ek & Trim, 1991] (cf. § d) p. 172), ces derniers servent à regrouper des notions spécifiques, lesquelles sont principalement lexicales et syntaxiques et permettent finalement de caractériser les intitulés des thèmes et sous-thèmes. Tout dépend ensuite de la définition que donnerait à « thème grammatical », mais il pourrait ne s'agir que d'une différence terminologique. Ce type de problème serait à traiter par le biais d'une base de connaissance (cf. note ¹⁵ p. 14).

Toutefois, notre propos ici est avant tout de nous intéresser au modèle de document, au modèle de requête et aux fonctions de correspondance (cf. § 1.3.2 p. 12), c'est-à-dire aux liens qui existent entre le contexte pédagogique (requête, par exemple selon les critères de tri du recueil) et les propriétés des textes.

5.3 Tentative d'établissement de relations entre des propriétés de texte et des contextes pédagogiques

Dans le chapitre précédent, nous avons introduit la notion de contexte pédagogique, que nous définissons comme « l'ensemble des traits décrivant la situation d'enseignement » (cf. définition 4.1a p. 211). Dans le début de ce chapitre, nous avons pu constater que les enseignants organisaient leurs recueils de texte personnels en grande majorité à partir de catégories qui se focalisent sur certaines composantes du CP, comme le thème développé (cf. § e) p. 225), le niveau des apprenants (cf. § c) p. 223) ou les objectifs linguistiques de l'activité à planifier (cf. § f) p. 226). Cependant, nous n'avons pas avancé dans notre compréhension de comment le contexte pédagogique s'intègre à la recherche de texte. L'objectif serait donc de mettre en relation le CP avec des propriétés du texte. Toutefois, nous avons souligné dans le chapitre précédent l'extrême

n'est pas un cours d'espagnol ou parce que dans le cadre de cours de littérature, la date de publication des textes fait partie intégrante de la problématique du cours. Peut-on exclure du CP un critère qui influence le choix du texte ? Dans ce cas précis, nous voulons en particulier signaler que ce sont les deux seuls critères qui se rapportent explicitement à des propriétés intrinsèques des textes.

complexité des interactions au sein du flux du faire (cf. § 4.6 p. 195). Pour progresser en direction d'un modèle, nous allons donc nous focaliser sur deux aspects centraux de la situation d'enseignement, l'activité et le public. Nous tenterons de mettre en évidence leur influence sur des propriétés du texte. Notre hypothèse de travail ici est la suivante :

H_j. Le contexte pédagogique intervient dans la recherche comme contraintes portant sur les propriétés des textes.

Avant d'entrer dans le détail des réponses nous allons expliciter les questions posées et le choix des variables. Nous détaillerons d'abord la partie ayant trait aux types d'activités, avant de nous préoccuper de quelques questions se rapportant au public. Avec ces questions, nous essayons de valider l'hypothèse H_j.

5.3.1 Les différents types d'activité considérés

Nous nous intéressons à plusieurs types d'activités, que nous détaillerons dans la suite de cette section. Pour chaque activité, nous posons une série de questions identiques, ce qui doit nous permettre de mettre en évidence l'influence de l'activité sur les différents critères proposés. Conscient de l'aspect fastidieux de ces répétitions de questions et des doutes qu'elles peuvent susciter, nous avons averti les enseignants au début de cette section : « Nous sommes bien conscients qu'un même texte peut être utilisé de plusieurs manières différentes et avec plusieurs buts pédagogiques, cela sera bien évidemment pris en compte au niveau du dépouillement des réponses. » Suite à la remarque d'une enseignante, lassée par la monotonie de cette section, nous avons ajouté sur le site une explication concernant la forme du questionnaire¹⁸.

a) Exercices lacunaires

Nous avons choisi de nous attarder sur les exercices lacunaires, qui sont souvent utilisés comme exercices structuraux. Même si cette forme d'activité est parfois décriée (cf. NB 5.1 p. 233), elle nous est apparue comme largement utilisée. Les réponses nous ont donné raison, 86% des enseignants ayant répondu à la question y ont recours (cf. tableau F.30 p. a-68). Or comme notre point de vue est centré sur les pratiques, cette proportion est un argument suffisant pour l'inclure dans notre réflexion.

Il s'agit d'un type d'exercice qui est aisément modélisable informatiquement et sous certaines de ses formes, générable automatiquement. En plus de correspondre aux pratiques d'une grande proportion des enseignants, c'est donc un type d'activité à prendre en compte dans l'optique de l'intégration d'un système d'indexation pédagogique de textes à Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO)¹⁹. Enfin, nous allons effectuer à travers ces activités un classement selon deux points de vue : type d'activités / objectifs de l'activité pour le type « exercices lacunaires ». Nous avons choisi 3 types d'exercices lacunaires en nous appuyant sur le travail théorique de notre étude préliminaire [Loiseau, 2003, § 3.2.2.].

Inférence du sens d'un mot d'après le contexte : Cette activité est tirée de [Gómez Casañ & Martín Viaño, 1990, p. 50]. Elle a pour but de faire réaliser aux apprenants qu'il n'est pas nécessaire de connaître chacun des mots d'un texte pour en comprendre le sens. D'après les

18. disponible à l'adresse <http://mathieu.loiseau.free.fr/bdtpip/faq.php> .

19. MIRTO constitue le point de départ de notre projet, comme nous l'expliquions dans le chapitre 1.

Nota Bene 5.1 Exercices structuraux

Les exercices structuraux sont un type d'activité de l'enseignement des langues qui émanent des théories béhavioristes et en particulier de la méthodologie audio-orale. Du point de vue de celle-ci « [l'exercice] structural [devait] être fait par les apprenants sans référence au sens et à la règle, il y aurait bien (d'après la théorie audio-oraliste) simple "répétition" immédiate des modèles structuraux, et non application d'une règle » [Puren, 1994, p. 158]. Puren dans une perspective historique de la didactique des langues revient sur cette méthodologie, qui pour lui représente « la version la plus dure d'une subordination de la pratique à la théorie » dont il déplore l'influence sur « l'éclectisme pragmatique » qui s'efface devant l'application de la méthodologie constituée.

« À la pointe extrême de cette logique réductrice ont été imaginées les "machines à enseigner"^a, où l'exercice structural était devenu le mode unique d'enseignement, et dont le programme fixait à l'identique la gradation des contenus pour tous les apprenants et toutes les situations d'enseignement/apprentissage. »
[Puren, 1994, p. 31]

Il évoque cette logique en des termes peu élogieux, parlant de « délire de maîtrise théorique » [Puren, 1994, p. 31]. Il signale ensuite que « les exercices structuraux ont été reconnus "faux" par les praticiens de leur point de vue (ils ne fonctionnaient pas) bien avant de l'être par les linguistes appliqués » [Puren, 1994, p. 198].

a. Cf. § a) p. 30.

auteurs, l'activité s'adresse plus particulièrement à des débutants; elle prend la forme suivante (appliquée ici à un texte de littérature enfantine²⁰) :

Il était une fois une paire de _____ qui étaient mariées ensemble. La _____ droite, qui était le monsieur, s'appelait Nicolas, et la _____ gauche qui était la dame s'appelait Tina. Elles habitaient dans une belle boîte en carton, où elles étaient roulées dans du papier de soie. Elles s'y trouvaient parfaitement heureuses et elles espéraient bien que cela durerait toujours.

Mais voilà qu'un beau matin une vendeuse les sorti de leur boîte afin de les montrer à une dame.

FIGURE 5.1: Exemple d'exercice lacunaire d'inférence du sens d'un mot d'après le contexte [Gómez Casañ & Martín Viaño, 1990, p. 50]

Exercices à objectif « linguistique » : Cette classe d'activités en regroupe plusieurs dont l'objectif est plus porté sur la forme. Elle doit permettre d'adresser des questions de vocabulaire, de phonétique et de syntaxe. Les exercices lacunaires de ce type peuvent prendre plusieurs formes, entre autres :

- Questionnaire à Choix Multiples (QCM) : les mots escamotés sont proposés parmi plusieurs mots inadéquats pour le contexte [Tagliante, 2006, p. 151], [Alonso, 1994, p. 67];
- remplacement d'un mot-forme du texte par un autre qui lui est lié²¹, la transformation à

20. Pierre GRIPARI (1992). *La sorcière de la rue Mouffetard et autres contes de la rue Broca*. Folio Junior, Paris. ISBN : 2-07-033440-6. (p. 42)

21. La relation peut porter sur la syntaxe (une forme fléchie étant remplacée dans l'exercice par son lemme), sur le lexique (une lexie remplacée par son contraire), etc..

effectuer pouvant éventuellement être donnée dans l'énoncé [Tagliante, 2006, p. 160] ;

- suppression complète de mots, en proposant ou non la liste des éléments (sous forme de lemmes mot-forme) escamotés.

La forme de l'exercice, ainsi que les mots escamotés²² dépendent de l'objectif de l'activité. Les combinaisons sont trop nombreuses pour être toutes détaillées ici²³.

Autres exercices de compréhension : Nous proposons aux enseignants, parmi les choix disponibles, un autre type d'exercice lacunaire de compréhension dans lequel l'enseignant ôte des « mots-clés », sans enlever systématiquement le même [Alonso, 1994, p. 67], contrairement au cas traité dans le paragraphe « Inférence du sens d'un mot d'après le contexte : » p. 232. Comme les mots escamotés n'ont pas nécessairement de traits communs, l'exercice est moins systématique et donne une part plus importante à la compréhension qu'un exercice où seuls les verbes au présent ont été enlevés, à plus forte raison quand le verbe est indiqué à l'infinitif.

Importance des types d'exercices lacunaires : Nous nous trouvons donc avec 3 classes d'exercices lacunaires proposés aux enseignants :

- exercice à « objectif linguistique » (ex : suppression de tous les adverbes d'un texte, de tous les adjectifs dits « appréciatifs » d'un texte, de tous les verbes d'un texte en demandant à l'apprenant de les conjuguer au temps approprié, etc...);
- travail sur l'inférence du sens d'un mot suppression de toutes les occurrences d'un même mot ;
- autre exercice de compréhension (Suppression de plusieurs mots-clés différents).

Nous nous centrons donc sur ces trois catégories d'exercices lacunaires pour nos tests quantitatifs tout en permettant à l'utilisateur d'ajouter des réponses. Nous avons obtenu les résultats du tableau F.31 (p. a-69), dans lequel apparaissent les types d'exercices ajoutés par les enseignants. Nous avons pu ramener la plupart de ces différentes propositions aux trois propositions originales, auxquelles nous avons ajouté une proposition supplémentaire (cf. tableau F.32 p. a-70).

Nous avons donc reporté les occurrences de :

- « dans une recette, les verbes » à notre proposition « exercice à objectif linguistique », puisqu'il s'agit d'escamoter des mots selon leur nature dans un texte de type « recette » ;
- « suppression des mots grammaticaux » à notre proposition « exercice à objectif linguistique », puisqu'il s'agit d'un cas particulier de cette proposition.
- « suppression d'un mot sur 7 par exemple sans prise en compte de sa nature ou de sa fonction » à notre proposition à « autre exercice de compréhension (suppression de plusieurs mots-clés différents) » puisqu'elle en est une précision²⁴ ;
- « Reconstitution du texte après écoute », « de compréhension des mots après un travail phonologique » et « dans une chanson, suppression des mots connus, des champs lexicaux connus et à compléter après écoute(s) » à la proposition « Exercice lacunaire de compréhension des mots après un travail phonologique », qui en constituerait une description plus générique.

Enfin nous avons exclu les propositions « dans un texte de civilisation » et « classification des champs d'un texte ». Dans le premier cas, le type d'exercice lacunaire est inconnu, seul le type du

22. Alonso suggère que l'élément supprimé peut être plus petit que le mot, elle propose par exemple de supprimer quelques lettres des mots [Alonso, 1994, p. 68].

23. Les références citées en proposent un large panel, que nous avons tenté de synthétiser dans [Loiseau, 2003, § 3.2.2.4.]

24. Cette proposition remet quelque peu en question l'intitulé de cette classe, puisqu'à supprimer un mot sur 7, les mots peuvent être des « mots-outils ».

texte est signalé et dans l'autre il n'est pas possible de dire s'il s'agit effectivement d'un exercice lacunaire.

La plupart des enseignants ont recours à des exercices lacunaires (86,6% de ceux qui ont répondu à la question). Le type d'exercice le moins utilisé est le travail sur l'inférence du sens d'un mot (24,6% des réponses). Cependant, ce n'est pas surprenant puisque ce type d'exercice était avant tout destiné aux débutants, alors que les autres peuvent en général être utilisés avec des niveaux plus variés. Les exercices lacunaires à objectif « linguistique » sont les plus fréquemment employés (80,7% des enseignants qui ont recours aux exercices lacunaires), ce qui n'est pas non plus surprenant, puisqu'historiquement, ils étaient dédiés à créer des automatismes structuraux chez les apprenants. Ceux que nous avons appelés les « autres exercices de compréhension », même s'ils ne sont pas les plus utilisés, représentent 3 enseignants sur 4. La surprise provient du fait qu'une solution qui n'était pas proposée, la « reconstitution du texte après écoute », s'avère n'être employée que par 2 enseignants de moins que les exercices à objectifs linguistiques. Et ce, alors qu'elle n'a pas été proposée à tous les enseignants²⁵, ce pourrait donc être le type d'exercice lacunaire employé par le plus grand nombre d'enseignants.

b) Introduction de nouvelles notions

La première classe d'activités évoquées était orientée vers la forme de l'activité : des exercices lacunaires, que nous avons ensuite répartis en 3 sous-classes en fonction des objectifs et du mode opératoire. Dans la seconde séquence de questions posées aux enseignants, nous nous intéressons plus à l'objectif de l'activité qu'à la manière dont elle a lieu. L'objectif concerné est l'introduction de nouvelles notions. Cependant, la formulation de notre question impose que cette introduction s'effectue par le biais d'un texte ou plusieurs²⁶. L'introduction de nouvelles notions correspond à une phase récurrente dans l'enseignement des langues, qui s'appuie souvent sur un texte référence. Aucun enseignant n'a d'ailleurs semblé surpris de cette appellation.

Nous avons séparé l'introduction de nouvelles structures grammaticales (syntaxe) et de nouveau vocabulaire (lexique), en nous demandant dans quelle mesure les deux n'étaient pas combinées. D'après les réponses recueillies, ce n'est absolument pas systématique ($\approx 63\%$ des enseignants combinent parfois ou jamais ces objectifs avec le même texte comme support, cf. tableau F.54 p.a-89). Nous avons donc décidé de considérer séparément l'introduction de nouvelles notions syntaxiques et lexicales.

c) Activité de compréhension

La dernière classe d'activités traitées avec cette séquence de questions concerne les activités de compréhension. À nouveau, nous nous focalisons sur l'objectif de l'activité. Nous donnons un exemple d'une telle activité dans la formulation de la question : une activité classique de lecture de texte et réponse à des questions. Toutefois, selon l'objectif de la compréhension ce type d'activité peut prendre des formes différentes, par exemple selon la précision des questions qui peuvent porter sur des faits très globaux ou sur des faits très pointus. Dans ce cas, pour faire intervenir une stratégie de lecture différente (écrémage), elles peuvent être données en amont de la lecture du texte [Tagliante, 2006, p. 136]. Selon les objectifs de l'activité de compréhension, cette

25. Dans le cadre de cette question à visées avant tout exploratoires, nous proposons les réponses « autres » des enseignants à ceux qui remplissaient le questionnaire ensuite.

26. Cf. DDL / conceptualisation grammaticale [Tagliante, 2006, p. 158].

dernière peut également s'assortir d'autres formes de questions. Marcus propose par exemple, dans le cadre de travail sur des ouvrages littéraires, de créer des fiches d'identité des personnages d'un texte [Marcus, 1999, p. 88].

Nous n'allons pas passer en revue la totalité des possibilités, le travail serait d'autant plus fastidieux qu'elles sont très nombreuses. Nous donnons ici des exemples pour expliciter les approximations opérées. Nous voulions recenser des classes d'activités très différentes, et reconnaissons le fait que celles qui sont proposées aux enseignants ont une portée très large. À ce stade, il nous a paru inévitable de passer par ce type de simplifications de toutes les composantes des activités. En effet, pour prendre en compte la complexité du problème, il nous aurait fallu nous focaliser sur des activités très précises, ce qui nous est paru prématuré étant donné l'incertitude qu'il y ait une influence en prenant en compte des classes d'activités très différentes. Les résultats que nous allons obtenir ne sont donc qu'une première étape qui nécessitera à long terme d'intégrer des variables beaucoup plus fines.

5.3.2 Les traits considérés

Les séquences pour chaque classe d'activité comprennent 6 à 8 questions. Nous ne nous préoccupons pour cette analyse que de 4 questions. Deux concernent des propriétés intrinsèques des textes²⁷ : la longueur et le « nombre d'éléments ».

Les deux autres questions considèrent l'influence que peut avoir une activité sur les attentes des enseignants concernant les propriétés du textes. En y répondant les enseignants s'appuient sur leur expérience qui dépend du cadre dans lequel les activités se déroulent, du public avec lequel elles ont lieu. Nous leur demandons si la présence de vocabulaire, puis de structures inconnues est recherchée, tolérée ou proscrite pour chaque activité. La question est de savoir si les attentes des enseignants concernant le contenu du texte (propriétés intrinsèques), qui est exprimée en terme de « tolérance à la nouveauté linguistique » (propriétés extrinsèques), sont influencées par l'activité.

a) Taille du texte

Dans chaque séquence, il est demandé aux enseignants d'indiquer la taille des textes qu'ils utilisent pour le type d'activité en question. Comme il s'agit d'un ordre de grandeur, il ne nous a pas paru nécessaire d'imposer l'unité. En revanche, dans un souci d'exploitation des réponses, il est nécessaire de toutes les rapporter à la même unité. Nous avons donc converti toutes les réponses en « mots ». Meier estime qu'une « ligne ne doit pas contenir plus de huit à douze mots » [Meier, 1992, p. 19]. Nous avons considéré pour la conversion qu'une ligne correspondait à 10 mots. En effet, il n'aurait pas été forcément plus précis de ne considérer que des réponses en mots, puisque c'est une unité qui est plus difficile à maîtriser que la ligne. Si l'enseignant qui répond n'a aucune idée de combien de mots peuvent contenir les textes qu'il utilise (précision à 50% près par exemple), mieux vaut effectuer une estimation du nombre de mots par ligne moins imprécise. Malgré la consigne de donner une réponse en « lignes » ou en « mots », certains enseignants nous ont donné des réponses en « pages ». Le nombre de lignes et, a fortiori, de mots contenus dans une page est trop dépendant de la source du texte pour que l'on puisse faire une

27. Ces propriétés sont intrinsèques dans le sens où à la donnée de consignes d'évaluations les réponses découlent de l'analyse du texte. En effet une fois que la notion de « mot » est définie, la longueur d'un texte en nombre de mots devrait être identique pour tout évaluateur.

approximation acceptable. Les réponses en « pages » et à fortiori en « temps²⁸ » ou en nombre de phrases²⁹ n'ont pas été prises en compte.

Certains enseignants donnaient des fourchettes de valeurs, pour les exploiter nous avons considéré la valeur moyenne entre la valeur minimum et la valeur maximum.

Dans ces questions, il était possible de « ne pas se prononcer », les moyennes produites sont les moyennes des personnes interrogées se prononçant et donnant une réponse interprétable.

b) Nombre d'éléments représentatifs

À l'exception des activités de compréhension (non-lacunaires), chacune des classes d'activités proposées se centrent sur certains éléments des textes en particulier. Les activités d'introduction de notions s'appuient sur des occurrences dans les textes des concepts à présenter. Nous leur demandons donc le nombre d'éléments représentatifs de la notion dont ils ont besoin pour traiter le problème. Ce même type de raisonnement peut être mené avec les exercices lacunaires, les « trous » étant les entités à dénombrer. Ces problématiques sont proches puisque, comme nous l'avons vu, selon le type d'exercice, chaque mot escamoté peut correspondre à une forme fléchée d'un lemme donné ou à un ensemble de traits morphologiques.

Pour chacune de ces classes d'activités nous leur posons donc la question du « nombre d'éléments représentatifs »³⁰. Comme pour les tailles de texte, certains enseignants proposent une fourchette de valeurs. Tous les calculs statistiques impliquant ces réponses utilisent la moyenne entre la valeur minimum et la valeur maximum des fourchettes de valeurs proposées par les enseignants.

c) Tolérance vis-à-vis de la nouveauté

Nous avons vu dans les réponses à notre premier questionnaire que l'une des problématiques du choix de textes pour l'enseignement des langues est la maîtrise du contenu. Nous avons pu le constater lors de notre travail sur les textes fabriqués (cf. § d) p. 172), mais aussi dans une certaine mesure dans le cadre du travail sur les provenances et sources privilégiées des enseignants, à travers la description de tendances dans le processus de recherche de documents (cf. § c) p. 181). À travers les différents témoignages des enseignants, il est apparu que le recours aux textes fabriqués est souvent décidé pour être certain que le niveau des textes est adapté aux apprenants. Il existe des indices de lisibilité³¹, mais pour notre problème, il nous faut des critères qui soient « parlants » pour les enseignants. Nous nous appuyons donc sur la notion de progression et sur la formulation des réponses des enseignants [27] et [117], qui expriment la difficulté de la manière suivante :

- « [...] il ne faut pas que le texte étudié comporte trop de problèmes grammaticaux ou de termes inconnus » [27] ;
- « [...] ne pas [embrouiller les apprenants] avec trop de lexique ou de grammaire inconnus » [117].

28. « 2 minutes » [239].

29. « 12 à 15 phrases » [241].

30. Cf. tableaux F.56 p. a-91, F.60 p. a-95, F.35 p. a-72, F.42 p. a-78 et F.49 p. a-85

31. Sans compter qu'un indice de lisibilité calculé pour des lecteurs natifs ne sera pas forcément utilisable pour des apprenants : la lisibilité dépendant entre autre de la xénité entre la L1 et la L2.

Dans chaque cas, nous demandons donc aux enseignants si, pour un type d'activité donné, ils recherchent ou proscrivent toute nouveauté en termes de vocabulaire (resp. structures grammaticales), tout en leur laissant la possibilité de la tolérer / ne pas trouver le critère pertinent ³².

Pour certaines activités, particulièrement l'introduction de nouvelles notions syntaxiques et lexicales, la présence de structures et de vocabulaire inconnus est une condition nécessaire, nous indiquions donc dans l'énoncé de la question que cette dernière portait sur les nouveautés « autre[s] que l'objectif de la leçon ».

5.3.3 Influences de la classe d'activité

Pour évaluer l'influence potentielle de la classe d'activité, nous allons traiter chaque facteur et le croiser avec la classe d'activité. Nous allons procéder dans le même ordre que pour la présentation des facteurs.

a) Taille du texte

Après avoir codé les données comme nous l'expliquions en a) p. 236, nous avons effectué une ANOVA sur la longueur des textes en fonction de l'activité dans laquelle ils sont utilisés. Les données sont fournies en annexe ³³.

Pour ce test l'hypothèse nulle est la suivante :

H₀. Quel que soit le type d'activité les textes font sensiblement la même longueur.

Les résultats de l'ANOVA sont récapitulés dans le tableau 5.5 de la présente page.

		$n =$	\bar{X}	F	Valeur de p	$<, 05 = \star$
ACTIVITÉ	Notion syntaxe	22	195	3,362	,007	★
	Notion lexicque	24	209			
	Lac. inférence	9	127			
	Lac. obj. linguistique	37	147			
	Lac. compréhension	28	154			
	Compréhension	29	268			

TABLE 5.5: Comparaison entre les types d'activité sur le plan de la longueur des textes

Avant d'entrer dans le détail du test de l'ANOVA nous pouvons remarquer qu'au niveau des moyennes, les exercices lacunaires sont les activités qui font appel aux textes les plus courts. En particulier l'exercice lacunaire d'inférence. Pour ce dernier cas, cette différence pourrait néanmoins tenir au fait qu'il s'agit d'une activité plus particulièrement destinée aux débutants. Ensuite, moins l'activité se focalise sur la forme plus les textes employés sont longs. L'activité de compréhension faisant appel aux textes les plus longs. L'ANOVA doit nous permettre de constater si ces différences sont significatives.

32. Cf. tableaux F.36 p. a-72, F.37 p. a-72, F.43 p. a-78, F.44 p. a-79, F.50 p. a-85, F.51 p. a-86, F.57 p. a-92, F.58 p. a-93, F.61 p. a-96 et F.62 p. a-97.

33. Cf. tableaux F.33 p. a-71, F.40 p. a-75, F.47 p. a-83, F.55 p. a-90, F.59 p. a-94 et F.68 p. a-107.

D'après le test³⁴, H_0 peut être rejetée. Il existe donc des différences significatives entre les longueurs des textes pour les classes d'activité décrites. Ce résultat suffirait à conclure que la classe d'activité a une influence sur la longueur du texte; cela dit, il se pourrait que seul un type d'activité sorte du lot. S'il s'agit des exercices lacunaires d'inférence pour lesquels l'effectif est inférieur aux autres, le résultat perdrait de son importance. Selon le test post-hoc de Tukey, la longueur des textes est significativement différente entre les activités de compréhension et les différents exercices lacunaires. Les valeurs sont récapitulées dans le tableau 5.6 de la présente page³⁵.

ACTIVITÉS		Valeur de p
Compréhension	Exercice lacunaire de compréhension	,019
	Exercice lacunaire d'inférence	,064
	Exercice lacunaire à objectif linguistique	,004

TABLE 5.6: Valeurs significatives pour le test post-hoc de Tukey (Classes d'activités / longueur des textes)

La différence n'est pas significative avec les exercices d'inférence du sens d'un mot du texte, cependant cela peut provenir de notre manque de données. Par contre, les différences respectives entre exercices lacunaires de compréhension et à objectifs linguistiques, d'une part et les activités de compréhension est significative. Les exercices lacunaires sont globalement plus courts que ceux utilisés pour la compréhension. Entre ces extrêmes se situent les textes supports pour l'introduction de nouvelles notions (qui ne s'écartent pas des autres groupes de manière significative).

b) Nombre d'éléments représentatifs

Nous avons déjà expliqué que cette catégorie ne s'appliquait pas aux activités de compréhension (cf. § b) p. 237). La caractérisation des mots à escamoter dans ce type d'activité est hasardeuse puisque si elle est lié à un travail linguistique spécifique (champ lexical, temps, etc.), l'activité n'appartient pas à cette classe. Nous avons laissé les exercices lacunaires de compréhension car il n'est pas exclu qu'il soit possible de caractériser les mots supprimés dans ce type d'exercice. Cependant, n'ayant pas eu vent de telles études, si la seule différence significative entre deux groupes fait intervenir celui-ci, nous ne pourrions considérer que notre hypothèse nulle est invalidée :

H_0 . Quel que soit le type d'activité le nombre d'éléments représentatifs de la notion travaillée est sensiblement identique.

Le test employé est à nouveau une ANOVA, dont nous indiquons les résultats dans le tableau 5.7 page suivante.

En consultant les moyennes, nous pouvons remarquer que l'exercice lacunaire d'inférence ainsi que l'introduction d'une notion syntaxique sont les deux activités pour lesquelles le nombre d'éléments représentatifs est le plus faible. Du point de vue des moyennes, l'introduction de notions grammaticales et de notions lexicales ne nécessitent pas le même nombre d'occurrences. Les notions grammaticales introduites font vraisemblablement³⁶ référence à une construction en particulier répétée 4 ou 5 fois dans le texte alors que les notions lexicales sont en général (au moins dans l'approche communicative) introduites par thème, multipliant ainsi les occurrences. Enfin

34. $F(143) = 3,362 ; p < ,01$.

35. La totalité des valeurs est disponible en ligne : <http://tr.im/anovaltxt> .

36. La tolérance au nouveau vocabulaire est supérieure à la tolérance aux nouvelles structures.

		$n =$	\bar{X}	F	Valeur de p	$<, 05 = *$
ACTIVITÉ	Notion syntaxe	32	4,69	4,739	,001	*
	Notion lexicque	27	8,15			
	Lac. inférence	9	3,68			
	Lac. obj. linguistique	37	11,11			
	Lac. compréhension	27	9,70			

TABLE 5.7: Comparaison entre les types d'activité sur le plan du nombre d'éléments représentatifs d'une notion

les exercices lacunaires linguistiques sont ceux qui font intervenir le plus de mots à escamoter. S'agirait-il d'un reste béhavioriste ?

D'après ces données, l'hypothèse nulle peut être rejetée³⁷. Cette fois-ci les groupes affichant des différences significatives sont plus nombreux, mais l'exercice lacunaire de compréhension est le protagoniste le plus fréquent. Il existe également une différence significative entre l'introduction d'une notion syntaxique et les exercices à objectif linguistique (cf. tableau 5.8 de la présente page³⁸).

ACTIVITÉS		Valeur de p
Notion syntaxique	Exercice lacunaire à objectif linguistique	<,001
	Exercice lacunaire de compréhension	<,001
Notion lexicale	Exercice lacunaire de compréhension	,011

TABLE 5.8: Valeurs significatives pour le test post-hoc de Tukey (Classes d'activités / nombre d'éléments représentatifs)

Cette fois-ci c'est l'introduction d'une notion syntaxique qui sort le plus du lot. La différence en nombre d'éléments représentatifs de la notion est significative avec le nombre de mots escamotés dans un exercice lacunaire de compréhension (qui est également significative avec le nombre d'éléments représentatif pour l'introduction de nouveau vocabulaire). Ne sachant pas caractériser ces mots (sauf quand il s'agit d'un choix systématique comme d'activités où 1 mot sur n est escamoté³⁹), nous nous intéresserons plus particulièrement à la différence significative entre le nombre d'éléments représentatifs quand il s'agit de l'introduction d'une notion syntaxique et d'un exercice à objectif linguistique, qui *a fortiori* peut porter sur cette même notion.

Comme pour la longueur du texte, nous pouvons rejeter l'hypothèse nulle et conclure que la classe d'activité influence le « nombre d'éléments représentatifs ».

c) Une tolérance différente au vocabulaire et aux structures

Avant d'entrer dans le détail des calculs statistiques destinés à mettre en évidence l'influence de la classe d'activité sur les deux dernières variables dépendantes, nous allons les comparer rapidement. Les réponses pour chacun des deux facteurs sont indiquées dans la figure 5.2 (p. 241).

37. $F(127) = 4,739 ; p <, 005$.

38. La totalité des valeurs est disponible en ligne : <http://tr.im/anovaerep>.

39. Cf. note²⁴ p. 234, § a).0.

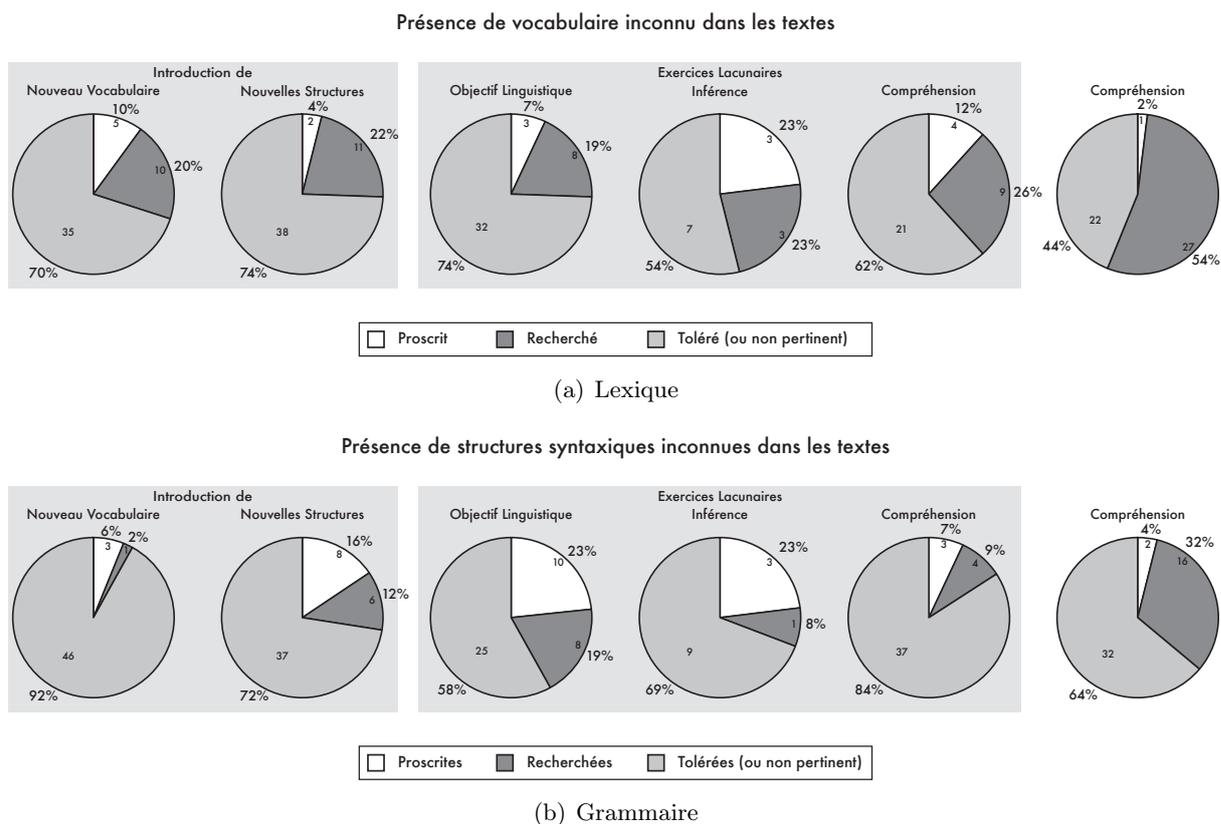


FIGURE 5.2: Tolérance à la présence d'éléments inconnus pour chaque classe d'activité

Dans chaque disque sont représentées les proportions d'enseignants :

- qui n'accordent pas d'importance à la présence ou l'absence de nouveauté dans les textes support de l'activité concernée (gris clair) ;
- qui la recherchent (blanc) ;
- ou la proscrivent (gris foncé).

Chaque « part de livarot » contient le nombre d'enseignants auxquels elle correspond, la proportion est ensuite exprimée en pourcentage.

Globalement, les enseignants semblent plus précautionneux avec les éléments syntaxiques qu'avec les éléments lexicaux inconnus⁴⁰. Le vocabulaire inconnu est systématiquement plus fréquemment recherché. Il est également régulièrement moins fréquemment proscriit. Les enseignants proscrivant le vocabulaire inconnu ne sont plus nombreux que ceux qui proscrivent les éléments grammaticaux inconnus, que dans le cadre des exercices lacunaires et des activités d'introduction de nouveau vocabulaire. Dans tous les autres cas, prohibition et permission sont tout à fait cohérents d'une question à l'autre.

40. Dans ce paragraphe, par inconnu nous désignons les éléments inconnus autres que l'objet de la leçon.

d) Vocabulaire

S'il est intéressant de constater que les enseignants accordent plus d'importance à la maîtrise des faits syntaxiques qu'à celle des faits purement lexicaux, l'objet de notre étude est d'évaluer l'influence de l'activité sur la « tolérance à la nouveauté ». Pour ce faire, nous allons avoir recours à un test de χ^2 , qui va nous permettre de savoir si la différence de répartitions qui semble évidente au regard des diagrammes de la figure 5.2(a) est significative ou non. En effet, à part pour les activités de compréhension, la proportion d'enseignants qui recherchent des éléments lexicaux inconnus (ne faisant pas l'objet de la leçon) est relativement stable. Elle oscille entre 19 et 26%. Par contre, la proportion d'enseignants pour lesquels le vocabulaire inconnu est proscrit varie beaucoup plus d'une activité à l'autre. Elle est pratiquement nulle pour les activités de compréhension et pour l'introduction de nouvelles structures, oscille entre 7 et 12% pour les exercices lacunaires à objectif linguistique, l'introduction de nouveau vocabulaire⁴¹ et les exercices lacunaires de compréhension, pour lesquelles elle est la plus élevée des 3. Enfin pour les exercices d'inférence qui visent à montrer aux apprenants qu'il est possible de comprendre le sens d'un mot sans le connaître, la forte proportion d'enseignants qui proscrit le nouveau vocabulaire (23%) peut vouloir s'assurer que les mots qui doivent servir de contexte pour retrouver le mot supprimé soient effectivement compris par les apprenants.

Pour savoir s'il y a un lien entre le type d'activité et la tolérance au vocabulaire inconnu, le calcul s'appuie sur la répartition des effectifs présentée dans le tableau 5.9 de la présente page.

		VOCABULAIRE INCONNU			TOTAL
		Proscrit	Recherché	Toléré	
ACTIVITÉ	Compréhension	2	16	32	50
	Lacunaire compréhension	3	4	37	44
	Lacunaire inférence	3	1	9	13
	Lacunaire obj. linguistique	10	8	25	43
	Notion grammaticale	8	6	37	51
	Vocabulaire	3	1	46	50
TOTAL		29	36	186	251

TABLE 5.9: Type d'activité croisé avec la tolérance vis-à-vis du vocabulaire inconnu

D'après le test de χ^2 , les différences entre les groupes sont significatives⁴². Cependant ce résultat est à tempérer par le fait qu'il ne satisfait pas les critères de Cochran [1952] (cf. § b) p. 174). Toutefois, d'après Sheskin, beaucoup de sources suggèrent que ces critères sont par trop conservateurs⁴³ [Sheskin, 2004, p. 515].

e) Structures grammaticales

Nous procédons de manière analogue avec les réponses portant sur les structures grammaticales inconnues et qui ne constituent pas l'objet de la leçon, et qui malgré tout sont présentes dans les textes.

41. Il pourrait s'agir d'une gestion de la « concurrence » entre le vocabulaire inconnu qui est l'objet de la leçon et celui qui ne l'est pas.

42. $\chi^2(10) = 28,9 ; p <, 005$.

43. « [Many] sources suggest the latter criteria may be overly conservative ».

En comparant les diagrammes de la figure 5.2(b) et de la figure 5.2(a) (p. 241), nous avons déjà remarqué que les enseignants étaient moins permissifs avec les structures inconnues qu'avec le vocabulaire inconnu (cf. § c) p. 240). C'est particulièrement vrai pour les activités tournées vers des éléments grammaticaux, comme l'introduction de nouvelles structures et les exercices lacunaires à objectif linguistiques.

		STRUCTURES INCONNUES			TOTAL
		Proscrites	Recherchées	Tolérées	
ACTIVITÉ	Compréhension	2	16	32	50
	Lacunaire compréhension	3	4	37	44
	Lacunaire inférence	3	1	9	13
	Lacunaire obj. linguistique	10	8	25	43
	Notion grammaticale	8	6	37	51
	Vocabulaire	3	1	46	50
TOTAL		29	36	186	251

TABLE 5.10: Type d'activité croisé avec la tolérance vis-à-vis des structures grammaticales inconnues

D'après les effectifs présentés dans le tableau 5.10 de la présente page, qui reprennent les informations disponibles dans la figure 5.2(b) (p. 241), les différences de répartition entre les classes d'activité sont à nouveau significatives⁴⁴, à ceci prêt que cette fois-ci les effectifs respectent les critères de Cochran [1952], puisque moins de 20% des cellules ont un effectif théorique inférieur à 5 (16,7%) et qu'aucune n'a un effectif théorique inférieur à 1 (l'effectif théorique minimum est de 1,56).

5.3.4 Influence du public

Pour le traitement (de certains aspects) de l'influence du public sur les textes, nous avons adopté une stratégie différente. Nous demandons aux enseignants comment ils prennent en compte les deux aspects choisis : la L1 des apprenants et leur niveau.

La première série de questions, concernant la L1 des apprenants, a pour vocation de savoir si les enseignants la prennent en compte : nous n'avions pas pu nous y intéresser dans le premier questionnaire (cf. § d) p. 158). Dans la suite logique de ces questions de tolérance à l'inconnu, nous leur demandons comment ils prennent en compte la L1. La dernière question procède de manière analogue mais avec le niveau des apprenants.

a) Influence de la L1 des apprenants

Sans rentrer dans le détail de classes de xénité comme Robert [2004], nous avons posé, aux enseignants qui déclaraient prendre en compte la L1 des apprenants s'ils pensaient que :

- « plus la langue maternelle des apprenants est *proche* de la langue enseignée *plus on acceptera* de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte » ;

44. $\chi^2(10) = 32,2 ; p <, 001$.

- « plus la langue maternelle des apprenants est *proche* de la langue enseignée *moins on acceptera* de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte ».

Les enseignants avaient aussi la possibilité de ne pas se prononcer ou de nous indiquer que dans un sens ou dans l'autre ces considérations ne leur paraissaient pas pertinentes. Les résultats sont indiqués dans les tableaux F.76 et F.77 p. a-115.

Il s'avère que sur les 49 enseignants qui ont répondu à cette question, seulement 16 (32,7%) déclaraient que la langue maternelle de leurs apprenants avait une influence sur le choix des textes qu'ils leur présentaient. Sur ces 16 enseignants :

- 1 (6,3%) estime que les choix qui lui sont proposés ne sont pas pertinents ;
- 2 (12,5%) ne se prononcent pas ;
- 13 (81,3%) choisissent la première solution.

Aux yeux des enseignants qui déclarent prendre en compte la L1 des apprenants, la quantité de structures et de mots inconnus est corrélée positivement avec la proximité entre la L1 et la L2.

b) Influence du niveau des apprenants

De manière tout à fait analogue, nous demandions l'influence qu'avait le niveau des apprenants sur le choix des textes. Sur les 49 enseignants ayant répondu à la question (cf. tableau F.78 p. a-116) :

- 35 (71,4%) considèrent que « *meilleur* est le niveau des apprenants, *plus on acceptera* de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte » ;
- 1 (2%) considère que « *meilleur* est le niveau des apprenants, *moins on acceptera* de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte » ;
- 4 (8,2%) ne considèrent pas le critère comme pertinent ;
- 9 (18,4%) ne se prononcent pas.

Parmi les enseignants qui estiment que le niveau peut être pris en compte selon les propositions précédentes et qui se prononcent sur la question, ils sont 97,2%⁴⁵ à déclarer que plus le niveau des apprenants est élevé et plus ils peuvent être confrontés à du vocabulaire ou des structures inconnus, même dans le cadre d'activités dont l'objectif n'est pas de présenter de nouvelles notions (ou *ces* nouvelles notions).

5.3.5 Synthèse des influences mises en évidence

D'après les tests statistiques que nous avons effectués, chacune des variables dépendantes que nous avons traitées s'est avérée liée de manière significative au trait du contexte pédagogique avec lequel nous l'avons croisée. Nous avons mis en évidence des liens entre :

- l'activité et le nombre de mots ;
- l'activité et le nombre d'« éléments représentatifs » ;
- l'activité et la tolérance en termes de vocabulaire inconnu ;
- l'activité et la tolérance en termes de structures grammaticales inconnues ;
- la L1 et la quantité de notions inconnues ;
- le niveau des apprenants et la quantité de notions inconnues.

45. $35/36 = 0,972$.

Les différences entre longueurs des textes employés sont en particulier significatives lorsque la comparaison effectuée confronte les différents types d'exercices lacunaires aux activités de compréhension. Les premiers utilisant les textes en moyenne les plus courts, alors que les dernières ont recours aux textes les plus longs. Les activités d'introduction de nouvelles notions sont intercalées entre ces extrêmes. Pour tirer plus de conclusions, d'autres études et calculs doivent être menés, mais il semblerait que, si l'on considère que les exercices lacunaires sont les plus spécifiquement centrés sur la forme⁴⁶ et l'activité de compréhension la plus centrée sur le fond en intercalant les activités d'introduction de notions sur ce continuum, alors moins l'activité est focalisée sur la forme et plus les textes sont longs.

Du point de vue du nombre d'éléments représentatifs, nous ne pouvons pas établir ou plutôt suggérer de comportement global. Nous avons cependant pu mettre en évidence des différences significatives principalement à partir des activités d'introduction de notion. La différence emblématique des comparaisons met en évidence le fait que pour travailler une notion syntaxique il faudra un texte comportant beaucoup plus d'« éléments représentatifs » que pour introduire cette même notion.

Pour la tolérance aux notions inconnues (qui ne sont pas l'objet des activités décrites), le test employé ne permet pas de préciser les liens existants, mais uniquement de mettre en évidence une relation entre le type d'activité et cette tolérance. Globalement, les enseignants semblent de ce point de vue plus permissifs avec le lexique qu'avec la grammaire. Et au sein des activités, c'est toujours la compréhension qui est la plus demandeuse d'éléments nouveaux mais sortants du contexte strict des objectifs de l'activité. Au contraire les exercices lacunaires, sont ceux pour lesquels les enseignants sont les plus divisés et les plus restrictifs.

Le calcul employé utilise les réponses des enseignants comme méthode de groupement. Pour certaines activités, principalement au sein des exercices lacunaires, mais aussi dans une moindre mesure dans le cadre de l'introduction de nouvelles notions (et particulièrement pour le type de notion dont fait l'objet la leçon⁴⁷). Cette « discorde » pourrait provenir d'autres facteurs issus du contexte pédagogique comme la proximité de la L1 et de la L2 des apprenants, comme le suggèrent les résultats suivants. Mais pour confirmer ce type d'influence, plus de recherches sont nécessaires.

Les tests que nous avons effectués constituent une première étape vers une compréhension en profondeur de l'influence de chacun de ces traits du contexte pédagogique sur les propriétés des textes. Cependant, nous pouvons nous servir des conclusions que nous avons pu tirer comme support de réflexion pour tenter de modéliser l'influence du CP sur les propriétés du texte.

5.4 Conséquences sur le contexte pédagogique et sa modélisation

Nous avons pu constater dans la première partie de ce chapitre que lorsque les enseignants trient des textes afin d'y avoir accès par la suite dans le cadre d'une recherche pour une activité, ils les organisent avant tout en fonction de critères qui relèvent de ce que nous avons appelé le contexte pédagogique. Les textes sont associés par chaque enseignant à une acception d'une ou plusieurs composantes du CP. Comme l'utilisation du recueil intervient à la phase que nous avons appelée

46. Il s'agit, comme nous l'avons vu, d'un raccourci, cf. § *Importance des types d'exercices lacunaires* : p. 234.

47. Nous voulons parler ici du vocabulaire inconnu ne faisant pas l'objet de l'activité dans le cas de l'introduction de nouveaux éléments lexicaux (respectivement structures grammaticales inconnues / introduction de nouveaux éléments grammaticaux).

« sélection » (cf. figure 4.7 p. 206), qui se trouve correspondre au premier cas d'utilisation que nous avons défini (cf. § b) p. 212). La prise en compte du contexte pédagogique semble donc primordiale dans cette phase. Nous avons donc étudié l'impact de certaines de ses composantes sur des propriétés du textes. Les éléments mis en évidence sont synthétisés dans la figure 5.3 de la présente page. Ils sont présentés vis-à-vis d'un système hypothétique d'indexation des textes.

5.4.1 Exemple de lien entre des aspects du contexte pédagogique et des propriétés des textes

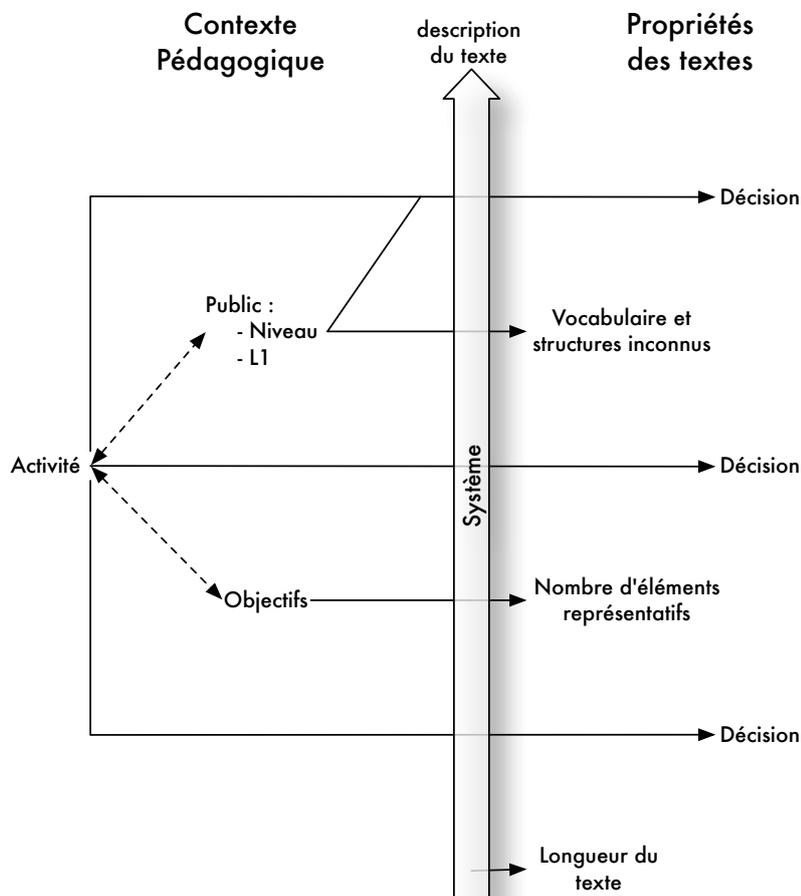


FIGURE 5.3: Influence du contexte pédagogique sur l'assignation de propriétés au texte

Dans la figure 5.3, la partie de gauche concerne différents aspects du contexte pédagogique alors que la partie de droite se rapporte aux propriétés du texte en cours de description. Les décisions que nous indiquons sont considérées comme des propriétés, dans la mesure où il s'agit de décider si le texte est adapté au contexte pédagogique en question.

a) Propriété indépendante du CP et son utilisation en rapport avec ce dernier

Nous pouvons remarquer que certaines propriétés ne sont absolument pas liées au contexte pédagogique. Comme par exemple la longueur, une propriété intrinsèque du texte qui, pour un document donné, sera la même quel que soit le CP. Cependant, nous avons constaté que classe d'activité et longueur du texte étaient liés. À la donnée de la propriété du texte (longueur) et d'un aspect du contexte pédagogique (le type d'activité), une décision peut être prise quant au texte.

Nous ne sommes pas en train de dire que les enseignants procèdent dans l'ordre de la figure 5.3, qui est plus un ordre de complexité qu'un ordre chronologique, ni de définir la décision à prendre. Cependant, d'après les déclarations des enseignants, il s'avère que la longueur des textes utilisés varie avec le type d'activité (cf. § a) p. 238). De plus, en fonction des informations disponibles sur les pratiques des enseignants, un système pourrait prendre des décisions en rapport avec ces éléments. Dans notre cas, nous ne pourrions pas aller plus loin que la définition d'une relation d'ordre entre les textes proposés en fonction de leur longueur. Mais à partir d'études plus poussées, faisant intervenir des observations en situation, un système pourrait exclure des textes en fonction de leur longueur et de l'activité dans laquelle le texte doit être utilisé. Dans un cas comme dans l'autre, le contexte pédagogique agit ici comme contrainte sur la propriété « longueur du texte » : le type d'activité contraint les valeurs que peut prendre cette propriété (afin d'ordonner les textes ou d'en exclure).

b) Propriété dépendante du CP

Le statut de la propriété « nombre d'éléments représentatifs » est beaucoup moins clair que celui de la longueur du texte. En effet, le nombre d'éléments représentatifs est difficile à catégoriser selon la définition de Recker & Wiley [2001]. Nous rappelons que pour eux, les propriétés intrinsèques relèvent de l'objet lui-même et sont dérivables du simple fait d'avoir la ressource en sa possession, alors que les propriétés extrinsèques décrivent le contexte dans lequel la ressource est utilisée (cf. § 3.5.3 p. 120). Or, le nombre d'éléments représentatifs, à partir du moment où la description des éléments est formalisée de manière non ambiguë, relève de l'objet lui-même (son contenu) et est dérivable du simple fait d'avoir la ressource en sa possession. Toutefois, ce calcul dépend directement du contexte dans lequel la ressource doit être utilisée : les objectifs poursuivis. Il s'agirait donc d'une propriété intrinsèque dans la mesure où quel que soit l'agent qui l'évalue elle devrait avoir la même valeur, mais paramétrique, puisque le calcul de cette dernière dépend d'éléments extrinsèques au texte correspondant à son contexte d'usage.

D'après nos données, cette propriété peut elle aussi être prise en compte pour prendre des décisions quant à l'utilisabilité d'un texte dans un contexte pédagogique donné (cf. § b) p. 239). Les objectifs ayant permis de calculer le nombre d'éléments représentatifs en fonction des objectifs de l'activité, cette propriété peut être interprétée à la lumière du type de l'activité dans laquelle le texte doit être utilisé. Encore une fois, nous n'avons pas nécessairement les informations pour tirer des conclusions définitives sur l'adéquation d'un texte avec une situation d'enseignement donnée, mais en attendant d'avoir des informations plus précises, les données recueillies peuvent permettre d'ordonner les textes.

La figure 5.3 (p. 246) montre donc l'influence des objectifs sur le calcul du nombre d'éléments représentatifs et celle de l'activité sur sa prise en compte dans le cadre de la recherche de textes. Il s'agit du jeu de contraintes du point de vue d'un système. La situation réelle est plus complexe,

puisque les actions et prises de décision de l'enseignant s'incluent dans un flux global au sein duquel les *faïres* dépendent les uns des autres (cf. § 4.6.2 p. 198 & § 4.6.4 p. 205). Dans la figure 5.3 nous indiquons cet état de fait en désignant d'autres jeux de contraintes, extérieurs au système, mais agissant sur des traits que ce dernier prend en compte. Un exemple de telles contraintes sont les influences mutuelles exercées par les objectifs sur le type d'activité (et réciproquement), qui sont prises en charge (consciemment ou non) à un autre niveau par l'enseignant (cf. le point de vue centré sur le système de la figure 5.3).

c) Propriété faisant intervenir différents niveaux d'informations

Enfin le dernier type de variable que nous avons testé : la « tolérance à la nouveauté » exploite le même type de fonctionnement que les autres propriétés traitées, mais pousse les traitements hypothétiques à un niveau de complexité largement supérieur.

En effet, pour le calcul de la longueur du texte, il suffisait de définir ce qu'était un mot et dans la perspective d'un traitement automatique, la définition d'une fonction capable de prendre en compte cette définition. La propriété était ensuite contrainte par le trait « activité » du contexte pédagogique, ce qui revient à interpréter la propriété longueur différemment en fonction du contexte. La gestion du nombre d'éléments représentatif est plus complexe. D'un point de vue conceptuel tout d'abord, puisque la valeur de la propriété elle-même est différente selon le CP. La contrainte de la propriété en vue de son interprétation fonctionne ensuite exactement sur le même mode que pour la longueur des textes. Dans la perspective de l'intégration d'une telle propriété à un système informatique, la quantité de traitement et de problèmes est néanmoins nettement supérieure. Elle devra faire intervenir au minimum une analyse morphologique dont le formalisme et l'implémentation permette de décrire les structures recherchées de manière générique et de les repérer dans le texte. Mais il faut aussi définir un formalisme « côté utilisateur⁴⁸ » qui permette aux enseignants de formuler leur problème. C'est à travers ce type de requête que serait pris en charge le point de vue lié aux objectifs.

La « tolérance à la nouveauté » est à notre avis l'un des éléments nécessaire à la prise en compte de la difficulté du texte. D'un point de vue conceptuel, elle fonctionne à quelque chose près de la même manière que le « nombre d'éléments représentatifs ». Le recensement du vocabulaire et des structures inconnus dépend du contexte pédagogique, et plus précisément de sa composante « public » (le niveau ici). La prise en compte de la quantité d'éléments recensés pour savoir si le texte est adéquat ou non pour la situation d'enseignement pourra dépendre, d'après nos informations :

- du type d'activité ;
- de la L1 des apprenants ;
- du niveau des apprenants.

Chacun de ces traits du contexte pédagogique influence la tolérance aux éléments inconnus différemment (nous n'avons d'ailleurs aucune information sur la manière dont ces facteurs interagissent entre eux). Pour arriver à la prise en compte de ce type de phénomènes, les ressources et connaissances du monde nécessaires sont très nombreuses. Il faudrait, par exemple :

- un modèle des apprenants qui recense leurs connaissances ;
- que ce modèle puisse être mis en rapport avec les éléments d'un texte (reconnaissance de structures connues des apprenants dans les textes) ;

48. Il peut s'agir du même formalisme ou d'un formalisme adapté accessible à l'utilisateur en lignes de commandes ou avec une interface graphique.

- un modèle des connaissances de la matière (ici l'enseignement des langues) qui permette de repérer les éléments pertinents du texte pour ce problème (Qu'est-ce qu'une structure ?) ;
- un modèle de description du contenu du texte qui soit adéquat pour décrire les aspects pertinents de la matière contenu dans le texte.

Une fois le repérage possible, il faudra contraindre les valeurs en fonction du niveau et de la L1 des apprenants ainsi que du type d'activités, ce qui demande pour être précis, plus d'informations que celles que nous avons à notre disposition⁴⁹. Nous traitons ici ce problème très rapidement sans vraiment nous poser la question de l'implémenter, ce qui ne manquerait pas de faire surgir de nouvelles difficultés.

5.4.2 Le contexte pédagogique comme paradigme

Les tests que nous avons effectués et les variables choisies constituent une preuve de l'influence du CP sur les propriétés du texte. Ils mettent effectivement en évidence un lien entre l'activité/le public et chacune des variables dépendantes choisies. Cependant, les informations que nous avons ne sont pas toujours suffisamment précises pour caractériser ces relations : l'ANOVA donne certains outils d'évaluation des relations potentielles alors que le test de χ^2 ne permet que de mettre en évidence l'existence d'un lien. Toutefois, ces connaissances parcellaires nous ont permis de donner des exemples de pistes de prise en compte de ces différentes influences dans le cadre d'un système de recherche de textes pour l'enseignement des langues. Il est intéressant de remarquer que quelle que soit la complexité de la question le mécanisme est analogue. Les propriétés qui dépendent du contexte pédagogique constituent en quelque sorte un point de vue du texte reflétant la problématique de l'enseignant dans sa recherche de textes. C'est le cas aussi bien pour des propriétés « quasiment intrinsèques » aux textes comme le nombre d'éléments représentatifs d'une notion, que dans le cas de propriétés de niveau didactique beaucoup plus élevé, comme l'adéquation d'un texte pour être utilisé dans le cadre d'une certaine activité avec un public donné.

Dans la figure 5.3 (p. 246), nous avons de part et d'autre du système deux ensembles de propriétés :

- l'un décrit la situation d'enseignement, le contexte pédagogique ;
- l'autre décrit le texte dans la perspective de son utilisation dans l'enseignement des langues.

Selon la définition de Recker & Wiley [2001], le second ensemble ne contient pas que des données strictement intrinsèques au texte, puisqu'elles varient en fonction du contexte pédagogique⁵⁰. Elles ne sont cependant pas strictement extrinsèques à la ressource non plus, puisque bien que dépendantes d'éléments extérieurs au système, elles ne dépendent pas de l'appréciation de l'enseignant, mais sont censées prendre toujours la même valeur pour un texte et un CP donnés. Au lieu de décrire le contexte dans lequel la ressource est utilisée, elles sont censées décrire un contexte dans lequel la ressource peut être utilisée, ou plutôt l'adéquation de la ressource avec un CP donné. L'appréciation de l'enseignant intervient ultérieurement dans le processus de recherche de texte : dans la phase d'évaluation. Le rôle du système est de permettre de prendre en compte le premier ensemble de caractéristiques (CP) en conjonction avec certains éléments du second (propriétés du texte⁵¹) pour augmenter ce dernier de nouveaux traits contextuels, qui sortent, comme nous l'avons expliqué, du cadre strict des notions de propriétés intrinsèque et

49. La tolérance augmente avec le niveau des apprenants, mais augmente-t-elle toujours selon le même rythme ?

50. Nous rappelons que pour Recker & Wiley, une propriété intrinsèque relève de l'objet lui-même et se dérive du simple fait d'avoir la ressource en sa possession (cf. § 3.5.3 p. 120).

51. Le contenu du texte est en soi l'une de ses propriétés et intervient dans le calcul des autres.

extrinsèque au texte (au sens de Recker & Wiley [2001]).

Dans un cas réel, une base de textes indexée pédagogiquement pour l'enseignement des langues dépendra de la définition, d'une part, de formalismes pour décrire les deux ensembles de propriétés et d'autre part, d'un système de contraintes sur et entre les propriétés des deux ensembles. Comme l'enseignant qui traite sa tâche de planification sous-problème par sous-problème, chacun affectant le problème global [Yinger, 1978] (cf. § 4.6.4 p. 205), les propriétés prises en charge par le système se focaliseront sur certains aspects du contexte pédagogique (et en ignoreront d'autres). Les propriétés ainsi obtenues précisent la description globale du texte pour une situation donnée. Toutes ces propriétés interviennent dans la phase de sélection et sont contraintes plus ou moins explicitement par le contexte pédagogique⁵². Le système devient alors un prolongement du CP, transférant ses contraintes aux propriétés des textes. Nous pouvons adapter la définition de ce dernier au point de vue du système :

Définition 4.1b (Contexte pédagogique) *paradigme à la lumière duquel les propriétés des textes prennent leurs valeurs.*

La définition d'un ensemble de propriétés et de mécanismes permettant de décrire les textes en vue de leur utilisation dans l'enseignement des langues est une activité à long terme. Chaque propriété doit faire l'objet d'observation des besoins, pratiques et conceptions des enseignants, de la création de formalismes pour les décrire et l'implantation d'outils permettant d'utiliser ces formalismes pour construire les propriétés. La complexité de la tolérance aux éléments linguistiques nouveaux qui n'est qu'un élément de la gestion du niveau donne une idée de la difficulté d'une telle tâche si elle veut décrire des propriétés d'un niveau didactique élevé. Une approche moins-disante (cf. § a) p. 59) du problème passe par la définition de propriétés plus élémentaires comme la longueur ou le nombre d'éléments représentatifs d'une notion. Selon cette approche, le traitement de propriétés plus élaborées ne doit se faire qu'après une réflexion suffisante. Cela comprend non seulement l'évaluation de la capacité d'une propriété de texte à focaliser les contraintes exprimées par la composante considérée du contexte pédagogique. Mais il faut aussi évaluer la possibilité d'intégrer cette propriété aux textes de manière fiable et cohérente, que ce soit par des outils Traitement Automatique des Langues (TAL) ou des annotateurs humains.

La suite de notre travail consiste donc à proposer un modèle permettant de prendre en compte le contexte pédagogique comme paradigme d'évaluation et d'interprétation des propriétés. Ce modèle doit pouvoir traiter des aspects du CP tout à fait minimaux en termes de formalisation⁵³, mais aussi pouvoir être augmenté au fur et à mesure que les connaissances sur la problématique évoluent.

52. En définissant un type d'activité pour un public, un enseignant peut limiter indirectement la longueur d'un texte, mais rien ne l'empêche de le faire plus explicitement, puisque pour les exercices lacunaires, plus de 40% des enseignants répondant à la question considèrent que la longueur du texte peut les influencer dans le choix d'un texte pour un exercice lacunaire à objectif linguistique (cf. tableau F.45 p. a-79). Un contexte pédagogique peut donc être du type « Exercice lacunaire avec comme support un texte de moins de 200 mots ».

53. Par exemple : CP = texte de moins de 300 mots / Propriétés du texte : nombre de mots, adéquation avec le CP.

Troisième partie

Proposition d'un modèle

Chapitre 6

Représentation des informations sur les textes, le modèle Prisme-Facette

Nota Bene 6.1 Modèle, version RC

À l'heure où nous écrivons ces lignes, notre modèle a été publié à trois reprises. Tout d'abord, [Loiseau *et al.*, 2008b] le présente selon le point de vue d'un enseignant de langues. Ensuite, Loiseau *et al.* [2008a] en donne une vision nettement plus technique. Loiseau [2008] constitue enfin une version française qui se situerait à mi-chemin entre les deux.

Le contenu de ce chapitre va différer sensiblement de celui de ces articles. Le modèle a évolué depuis. Pour parler en termes de gestion des versions en informatique, ces articles concernaient la version β du modèle, c'est-à-dire une version en période de tests [Wikipedia, 2009h] (via ces conférences). Mais cette première version nous a semblé améliorable à court-terme, notamment à travers le changement de la terminologie employée, ou plutôt de la sémantique associée à la terminologie employée, ainsi qu'une restructuration des statuts des concepts. Pour poursuivre notre métaphore informatique, le modèle présenté ici en est à sa version *Release Candidate* (RC).

Les résultats des questionnaires précédents nous ont permis, en nous appuyant sur les travaux concernant la planification dans l'enseignement de Yinger [1978] et en les rapportant à la notion de flux du *faire* de Holloway [2007], d'introduire la notion de contexte pédagogique, qui est à notre avis une notion primordiale en vue de définir une forme d'indexation pédagogique pour l'enseignement des langues. En effet, il est apparu que la pratique centrale de recherche de textes est la recherche pour une activité (cf. hypothèse H_c , § c) p. 181 et § 4.6.4 p. 205), qui est une recherche pour un contexte pédagogique donné. En détaillant cette pratique, nous avons ensuite suggéré qu'à travers le choix d'un texte, l'enseignant précisait le Contexte Pédagogique (CP), c'est-à-dire les caractéristiques de la situation d'enseignement dans laquelle ce dernier allait être utilisé. Nous avons ensuite convenu que les pratiques de planification menant à l'utilisation de textes dans l'enseignement des langues étaient intimement liées à la définition du contexte pédagogique (au moins pour celles qui sont présentes dans notre premier questionnaire). Toutefois, à ce stade de nos recherches, la notion de contexte pédagogique restait floue du point de vue de son lien avec les textes.

Les éléments au centre de notre problématique proviennent directement de la validation de deux hypothèses. En effet, la recherche de textes se fait en premier lieu dans l'optique de son usage dans une activité (cf. H_c , § a) p. 175) mais chaque texte peut être utilisé dans des contextes pé-

dagogiques variés (cf. H_b, § 4.4.5 p. 187). La conjonction de ces deux résultats rend le traitement de notre problème difficile puisque d'un côté les éléments proposés dans les standards de description de ressources pédagogiques semblent cohérents, mais de l'autre, ils ne sont pas utilisables en l'état, puisque chaque texte peut être utilisé dans plusieurs contextes et l'utilisation d'un texte n'en est pas une propriété intrinsèque. En effet, la description des textes en terme d'intégration à une activité est alors dépendante de la perception de l'annotateur, qui est influencée par son point de vue (enseignant travaillant avec des enfants ou des spécialistes ou des professionnels, dans une école ou un organisme privé, selon un éclectisme personnel, etc.). Une seule personne peut-elle proposer une annotation exhaustive d'une ressource textuelle brute ?

Cette difficulté a été adressée par exemple par Recker & Wiley [2001], qui introduisent une distinction entre propriétés intrinsèque et extrinsèque. Cependant, si leur approche est à long terme complémentaire de la nôtre, elle n'est pas directement réutilisable pour nous. Il ne s'agit pas d'une opposition entre deux approches concurrentes, mais de problématiques de départ différentes : Recker & Wiley revendiquent des statuts différents pour les propriétés selon qu'elles sont extrinsèques ou intrinsèques et proposent d'intégrer les premières à la description des objets par le biais de l'identification de l'annotateur et de la création de groupes d'utilisateurs partageant des « goûts » similaires (cf. § b) p. 122). Cette approche permet de prendre en compte le statut de chaque forme de propriété, remplaçant dans son contexte la définition de propriétés extrinsèques. Ces dernières sont, dans le cadre défini par Recker & Wiley, considérées justement comme l'expression du point de vue d'un annotateur. Mais cette approche comporte des désavantages pour notre problématique, limitant la plupart des descriptions pédagogiques aux usages avérés des textes, qui ne constituent pas une définition exhaustive des usages potentiels. Notre approche se fixe comme objectif de pouvoir proposer aux enseignants des textes qui seraient, par leur action, utilisés pour la première fois dans un contexte donné. Elle doit donc dépendre au minimum des contributions des annotateurs humains pour l'évaluation de propriétés qui dépendent de leur perception de la problématique de l'enseignement des langues. En effet, ce type de propriétés, par essence beaucoup plus précises que celles que nous pourrions employer¹, ne peuvent satisfaire aux exigences d'exhaustivité et de cohérence de la description pour notre problématique. Il pourrait cependant venir compléter un système d'indexation pédagogique de textes, mais pas en être le moteur.

Le modèle qui va être décrit dans ce chapitre doit proposer une manière de considérer les textes et leurs propriétés en rapport avec le contexte pédagogique. Il n'est pas neutre et suit l'approche que nous avons tenté de caractériser dans les lignes qui précèdent. Nous allons dans un premier temps préciser cette approche, avant d'introduire notre modèle, concept par concept.

6.1 Propriétés et stratégie

Le contexte pédagogique semble correspondre à des propriétés extrinsèques au texte. Nous nous sommes appliqué, via le second questionnaire, à mettre en relation ses propriétés avec celles des textes (propriétés intrinsèques). Ce faisant, nous avons constaté que la définition de Recker & Wiley [2001] reflète leur approche du problème, qui n'est pas la nôtre (cf. § 4.6.1 p. 196). En effet, nous constatons en 5.4.2 (p. 249) que les propriétés des textes telles que nous les décrivions

1. Des informations comme la durée de l'activité utilisant le texte, l'énoncé fourni, les difficultés rencontrées par les apprenants, un retour d'expérience comme les modifications à apporter au mode opératoire de la tâche vont bien au delà de notre problématique.

dans la figure 5.3 (p. 246) ne s'intègrent pas parfaitement au cadre défini par Recker & Wiley. Le point de vue sur lequel nous nous focalisons pour ce travail doit, à l'instar de Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO), s'adapter autant que possible à l'ajout de textes nouveaux et, dès lors que c'est possible², être applicable à des textes tout-venant (cf. § 1.1.1 p. 5).

Dans la figure 5.3 (p. 246), nous avons séparé les propriétés entre, d'une part, celles du contexte pédagogique et, d'autre part, celles des textes. Dans ce schéma, une propriété du texte doit être calculable selon une procédure reproductible à partir des données disponibles. Ces dernières sont elles-mêmes des propriétés du texte, de son contenu ou des composantes du contexte pédagogique. Ce dernier n'est donc pas une propriété du texte, mais en offrant un paradigme au sein duquel les considérer, il les influence. Comme le texte en tant qu'objet et, par extension, son contenu participent à calculer ses propriétés, le Traitement Automatique des Langues (TAL) semble particulièrement indiqué pour automatiser une partie des traitements d'indexation. En effet, il offre une porte d'entrée aux propriétés de la langue dans un tel système (cf. [Antoniadis *et al.*, 2005b, p. 67]), qui semblent entrer en compte dans la plupart des catégorisations (cf. § f) p. 226).

Pour illustrer ce point de vue, appuyons-nous sur le cadre théorique défini par Rüschoff dans [Rüschoff, 1993]. Ce dernier se revendique, à travers la figure 6.1 de la présente page, d'une approche orientée processus issue de travaux de Candlin et Widdowson.

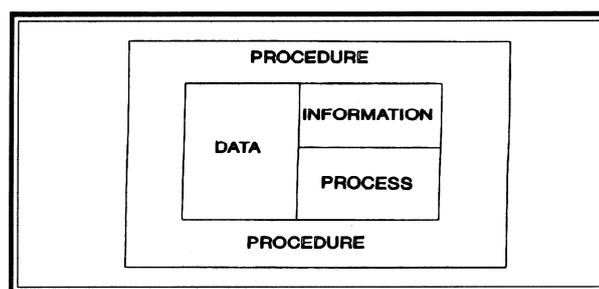


FIGURE 6.1: L'apprentissage des langues [Rüschoff, 1993, p. 7]

Dans la figure 6.1, les « données » (*Data*) représentent le contenu du cours, c'est à dire « la langue comme le sujet de l'apprentissage dans sa forme non didactisée³ » [Rüschoff, 1993, p. 7]. Les « informations » représentent le système utilisé pour « déverrouiller, comprendre et accéder aux données » ; il s'agit des ressources et matériels utilisés dans l'apprentissage des langues (authentiques ou didactisés)⁴. L'information sert de support aux apprenants au cours du « processus d'apprentissage » (*Process*) : les tâches d'apprentissages par lesquelles il leur est demandé de travailler avec les données et les informations. La « procédure » englobe toutes ces entités à travers la manière dont elles sont arrangées et organisées au cours de la situation d'apprentissage elle-même.

2. Nous verrons plus loin (cf. § 6.4.2 p. 268) qu'il ne s'agit pas d'une condition *sine qua non* à la prise en compte d'une propriété. Cependant, l'ajout d'un texte au système devant être aussi simple que possible, les textes annotés ne peuvent constituer qu'une sous-partie de la collection (cf. § 1.2.1 p. 8), sans quoi l'ajout d'un texte nécessiterait d'un enseignant qu'il soit spécialiste de toute problématique traitée par le système et disposé à effectuer une analyse en profondeur de chaque texte. Quoi qu'il en soit, le modèle doit pouvoir prendre en compte ces informations supplémentaires, notamment par l'accès à ces propriétés spécifiques.

3. « *language as the subject of learning in its non-didacticized form.* »

4. « *Information, then, can be referred to as the system that is used to unlock, understand and access the data, i.e. the resources and materials that are used in language learning. This can be both authentic discourse and didacticized learning materials.* » [Rüschoff, 1993, p. 7]

Dans ce cadre théorique, le rôle des textes est prépondérant (sous-partie de l'information) et sert de lien entre l'apprenant et les données. L'enseignant guide cette mise en relation à travers les processus et la manière de les mettre en place. Ce modèle propose un autre point de vue sur la situation d'enseignement décrite dans la figure 1.3 p. 16 (triangle de Houssaye revu par Rézeau [2001]). Les « données » de Rüschoff correspondent au savoir de Houssaye. La définition par l'enseignant de la procédure pourrait s'apparenter à la pédagogie dans le triangle, même si elle comporte une composante didactique. Enfin, la tâche elle-même (effectuée par l'apprenant) relèverait de « l'art de l'apprentissage ».

L'un des intérêts de cette modélisation est, de notre point, de vue d'assimiler le texte à un ensemble d'informations et de le mettre en relation avec les autres composantes de la situation d'apprentissage. En déconstruisant le matériel pédagogique et en l'incorporant, dénué de forme, sous la bannière de l'information, ce cadre rappelle l'objet de son utilisation (dévrouiller, comprendre et accéder aux données) et sa fonction de véhicule de l'information qu'il contient.

Nous ne pouvons nous défaire de la notion de texte, qui est l'unité élémentaire de notre problématique, cependant nous pouvons reprendre à notre compte la position pivot de l'information dans le cadre défini par Rüschoff. Nous devons garder à l'esprit que les objets que nous manipulons sont des vecteurs d'informations liées aux connaissances à acquérir (*data*), au processus selon lequel les acquérir (*process*) et à leur organisation globale (*procedure*). Chacun de ces éléments s'incorpore au contexte pédagogique : l'activité se rapportant au processus, la progression à la procédure, le thème et les structures aux données, pour ne citer que quelques exemples.

Comme dans le cadre du projet MIRTO, le fait de rendre disponible aux enseignants des propriétés des textes relevant de leur problématique passe par l'interprétation du contenu de ces derniers, notamment grâce à des fonctions TAL (cf. citation de [Antoniadis *et al.*, 2005b, p. 65] p. 5 ou § c) p. 55). Dans une optique de cohérence avec notre approche moins-disante, les exemples que nous donnerons dans ce chapitre seront pour la plupart très simples et pour certains implémentables dans notre maquette, mais le modèle se veut générique et capable de prendre en charge, à plus long terme, des propriétés plus élaborées.

6.2 Le couple *facette-prisme*

Ayant rappelé, en nous référant à [Rüschoff, 1993], notre stratégie pour mettre en relation les caractéristiques des textes et du contexte pédagogique, nous pouvons nous atteler à la tâche de formalisation des concepts et à la description de notre modèle.

6.2.1 Prisme

La *facette* est la notion centrale du modèle, mais elle est indissociable de la notion de *prisme*, qui va nous servir à la préciser. C'est autour de ces deux entités que le modèle s'articule⁵. L'objectif est, rappelons-le, de pouvoir prendre en compte l'influence du contexte pédagogique sur les caractéristiques du texte.

5. À travers l'usage de ces termes qui filent la métaphore introduite dans la définition 4.1b (p. 250) du contexte pédagogique – « paradigme à la lumière duquel les propriétés des textes prennent leur valeurs » –, nous tentons de rendre notre modèle visualisable, et plus facile à appréhender.

Définition 6.1 (Prisme) *Un prisme est un mécanisme, informatisable ou non, associé à une propriété définie en vue de l'exploitation pédagogique des textes, permettant d'attribuer une valeur à cette propriété pour tout texte, en fonction d'un contexte pédagogique donné.*

a) Conséquences

D'après cette définition, l'exploitation pédagogique ultérieure des textes considérés préside à la création des prismes. Toute caractéristique de texte pouvant être utile à un enseignant de langues pour la recherche ou le choix d'un texte pourra donc être un prisme applicable à ce texte. Pour calculer la valeur d'une propriété selon un prisme, nous aurons parfois recours à des propriétés sous-jacentes des textes ou de leur contenu, qui en tant que telles n'ont pas nécessairement d'intérêt didactique. C'est donc au niveau du prisme que se crée la valeur ajoutée pédagogique.

Dans cette définition, nous différencions la propriété, qui se situe à un niveau conceptuel (classe de propriétés), de sa valeur (après instanciation). L'essence même du prisme est contenue dans cette relation. Il est la procédure qui permet de considérer un objet (un texte) à l'aune d'un concept.

À titre d'exemple, considérons le prisme associé à la propriété « nombre de mots d'un texte », que nous appellerons dorénavant P_{nbmots} . Le nombre de mots d'un texte est effectivement un prisme, puisque la longueur du texte peut influencer sur la décision d'un enseignant d'utiliser ou non un texte pour son cours (ce critère a plus ou moins d'importance selon le type d'activité que l'enseignant désire préparer avec le texte). P_{nbmots} est un outil qui permet, à la donnée d'un texte, de savoir combien de mots il contient. Ce prisme s'appuie sur le contenu du texte et sur ses propriétés (quelle est l'acception de « mot » considérée?) pour assigner une propriété à l'objet.

6.2.2 Facette et valeur de facette d'un texte selon un prisme

Le prisme est donc l'expression, sur le plan des traitements, de la propriété à laquelle il est associé.

Définition 6.2 (Facette) *Une facette de texte est une propriété de ce dernier définie en vue de son exploitation pédagogique et pour laquelle une procédure d'évaluation peut être définie et appliquée à tout couple (texte, CP).*

Nota Bene 6.2 Flou terminologique

La facette associée à chaque prisme est donc une classe de propriété associant une valeur à chaque couple (texte, CP). Le terme « facette », comme le terme « propriété », porte de ce fait une certaine ambiguïté. En effet, le même mot peut désigner le concept et son application. Le « parallélisme » est par exemple une propriété applicable à certains types d'objets (concept), alors que deux plans (par exemple une étagère et le sol) peuvent avoir la propriété d'être parallèles (attribut). Pour désambiguïser autant que possible les potentiels abus de langage métonymiques de notre terminologie, nous précisons nos usages du terme. Quand nous parlerons de facette F_i , de facette de texte ou *facette* sans autre précision, il s'agira de l'acception conceptuelle du terme. En revanche quand le terme « facette » sera associé à un objet en particulier (facette du texte, $F_i(T)$) ou que nous parlerons de valeur d'une facette, nous nous référerons à l'attribut.

Il existe donc une relation essentielle récursive entre prisme et facette :

- le prisme est associé à une propriété des textes définie en vue de leur exploitation pédagogique. Sans cette facette, nous avons affaire à un outil, une fonction, un mécanisme, mais pas à un prisme.
- les facettes n'existent que par le prisme : sans un mécanisme permettant de les calculer pour tout texte et pour tout CP, il peut s'agir de propriétés mais pas d'une facette⁶.

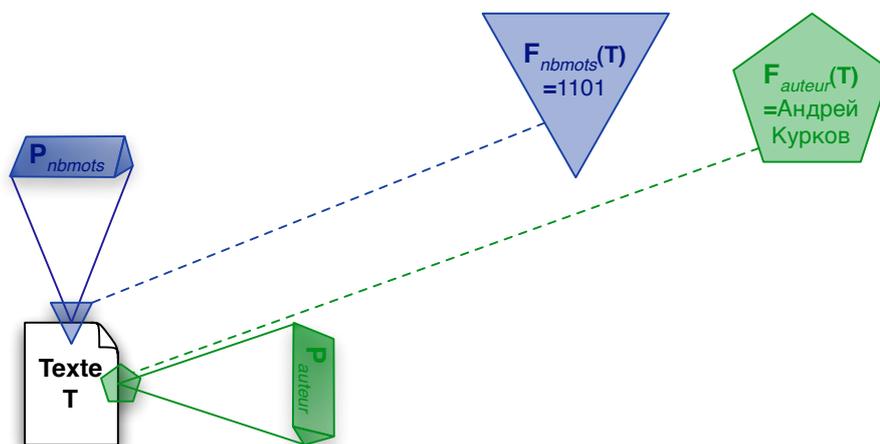


FIGURE 6.2: Exemples de prismes et de valeurs pour les facettes correspondantes (ici, constantes quel que soit le contexte pédagogique)

Bien que, strictement, l'*application* qui permet le calcul des valeurs soit définie au sein du prisme, la notation fonctionnelle que nous adoptons (voir NB 6.2 p. 257) s'appuie sur les facettes (cf. figure 6.2 de la présente page). Outre le fait de permettre une lecture plus claire⁷, elle précise le statut de chaque entité :

- le prisme est un outil, matérialisant un processus ;
- la facette est un concept, une propriété de texte qui a une valeur pour chaque couple (*texte*, *CP*).

La facette est l'entité centrale du point de vue conceptuel : elle représente les notions sur lesquelles portent les raisonnements des enseignants. En effet, si nous rapportons ces concepts au processus de planification, l'enseignant manipule des facettes ; c'est notamment le cas dans la tâche de sélection (cf. § a) p. 206). Cependant, dans un cas comme dans l'autre la propriété est indissociable du processus pour l'évaluer. Par exemple, les propriétés assignées aux textes (thème, activité) sont les critères de classement des recueils de textes, qui constituent un outil pour accéder à des textes à évaluer (cf. § c) p. 212). Mais d'après la vision de Yinger [1979], l'assignation de ces propriétés est liée à des routines développées par l'enseignant (cf. les deux méthodes pour la phase « d'élaboration » [Yinger, 1978, pp. 37–38]). Ces routines d'assignation reflètent la relation de l'enseignant à la sémantique des propriétés et le fait que leurs acceptations varient d'un enseignant à l'autre (cf. thème, § e) p. 225) met en évidence le lien qui existe entre la propriété et le processus pour l'évaluer.

6. Le fait d'avoir été considéré comme propice à l'usage dans un exercice lacunaire par un enseignant tel jour est une propriété du texte, mais ce n'est pas une facette. Elle n'est pas systématisable à d'autres textes ou contextes pédagogiques.

7. Elle remplace à notre avis avantageusement une notation du type « La valeur de la facette $P(t, C)$ ».

Notre modèle s'appuie sur cette constatation, mais avec une approche spécifique, puisque le système ne pourra prendre en compte les points de vue de tous les enseignants⁸. Une réponse de l'indexation à ce problème est la recherche d'une cohérence suffisante dans la description des ressources pour permettre à l'utilisateur d'adapter ses usages à l'acception retenue par le système. Cette cohérence est l'un des objectifs de la prise en charge des propriétés « pédagogiques » des textes sous forme de facettes, nécessairement liées à des prismes. Elle n'en est pas le seul puisque c'est la forte dimension contextuelle des propriétés et la prise en charge du contexte pédagogique qui en sont à l'origine.

6.2.3 Facettes et contexte pédagogique

Pour préciser la notion de facette et la relation avec les prismes, nous allons nous appuyer sur un exemple. Prenons un prisme qui permettrait de compter, dans un texte, le nombre d'éléments représentatifs d'une notion donnée (cf. § b) p. 237). Soient P_{etRep} , ce prisme et F_{etRep} , la facette qu'il permet d'instancier.

Dans le contexte de l'enseignement de l'espagnol, une telle notion pourrait être l'expression du devoir via l'utilisation d'une structure utilisant le verbe **haber** conjugué, suivi de **que** et d'un verbe à l'infinitif. La mise en oeuvre d'un tel prisme demande l'utilisation d'un analyseur morphologique⁹. Une fois l'analyse morphologique d'un texte effectuée, il est en général possible de rechercher des mots dans un texte en fonction de leur lemme (cf. **haber**), de leur forme (cf. **que**), de leur catégorie morphologique (cf. **verbe**) et enfin de leur flexion (cf. **infinitif**), qui dépend de la catégorie. Il faut combiner ces informations avec un programme de *pattern matching*, qui permette de contextualiser ces mots et de définir un patron (cf. recherche d'une forme fléchie du verbe **haber**, suivie de la forme **que**, elle même suivie d'un verbe à l'infinitif), avant d'utiliser un compteur, qui dénombre les structures suivant ce patron.

Considérons l'extrait de la figure 6.3 de la présente page comme un texte entier :

« 19. ¿ Qué opinión tiene usted de la revolución bolivariana de la republica bolivariana de Venezuela ?

Me parece un proyecto político de gran interés que merece ser *apoyado* y más aún *debido* al grado de desinformación y manipulación que flota sobre él aquí en España.

34. Lo siento señora, pero como resumen de su entrevista, lo único que veo es que le molesta que sean otros y no los suyos, los que tienen poder. Su comentario sobre las elecciones la retrata...

Si hay que jugar, a lo que sea, habrá que jugar en igualdad de condiciones. Pero siempre hay quien está *dispuesto* a hacer lo que sea y a costa de quien sea para jugar con ventaja. »

<http://www.elmundo.es/encuentros/invitados/2004/10/1241/>

FIGURE 6.3: Exemple de texte.

Dans cette figure, les occurrences de **haber* que + Inf.** sont encadrées. La facette F_{etRep} pour le texte de la figure 6.3, selon le contexte pédagogique **haber* que + Inf.** sera donc 2. Par la suite, nous adopterons la notation fonctionnelle suivante :

8. Il s'agit d'un problème inhérent à la tâche de classification, voir à ce sujet [Ranganathan, 1965, pp. 204–205], pour le cas d'une bibliothèque.

9. Cf. note ⁵⁶ p. 56.

$$F_{e^t Rep}^{[haber* que + Inf]}(figure\ 6.3) = 2.$$

qui se lira : « la facette du texte de la figure 6.3 selon le prisme $P_{e^t Rep}$ dans le contexte pédagogique **haber* que + Inf**. vaut 2 ». Il est évidemment possible de tirer de ce prisme $P_{e^t Rep}$ d'autres facettes d'un même texte, pour des contextes pédagogiques différents. Nous pourrions par exemple remplacer le contexte pédagogique par **para + Inf**.¹⁰ (souligné dans la figure 6.3) :

$$F_{e^t Rep}^{[Para + Inf]}(figure\ 6.3) = 1.$$

ou par **Participe Passé** (en *italiques* dans la figure 6.3) :

$$F_{e^t Rep}^{[P.Pass.]}(figure\ 6.3) = 3.$$

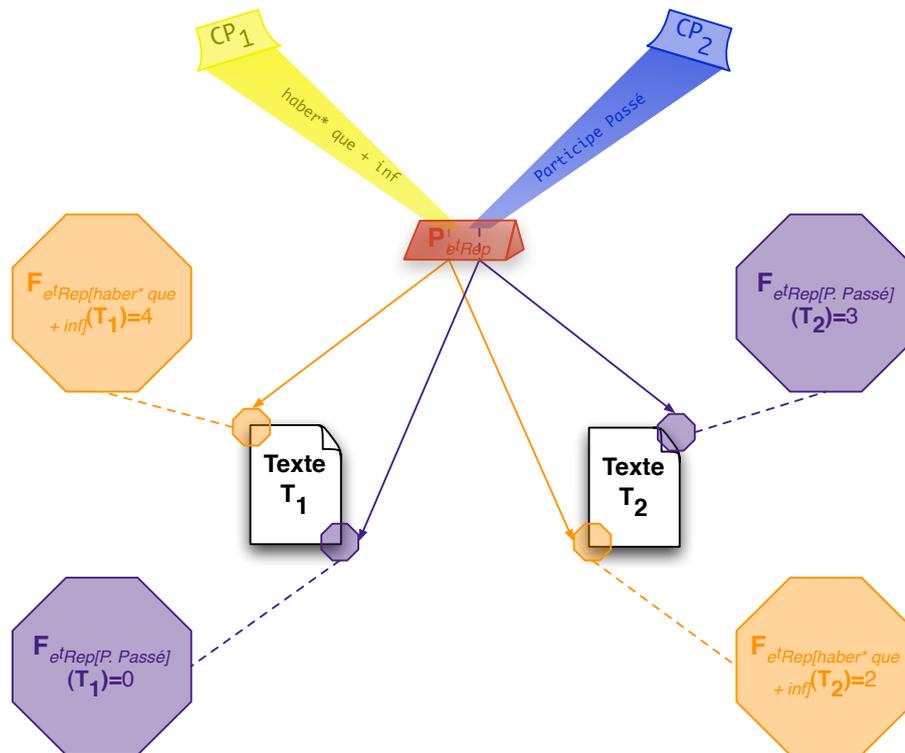


FIGURE 6.4: Exemple de prisme et facettes correspondantes

Dès la notion de prisme, le lien fort avec le contexte pédagogique est défini. La figure 6.4 de la présente page ne présente que le prisme de notre exemple ($P_{e^t Rep}$). Dans ce diagramme, le contexte pédagogique est représenté comme une source de lumière qui, projetée sur un texte par l'intermédiaire d'un prisme, en révèle une facette. Conformément à cette métaphore optique, le prisme divise le faisceau du CP pour n'en conserver que les fréquences (composantes) nécessaires au calcul des facettes.

Dans la figure 6.4, pour chaque texte, le même prisme est à l'origine de deux instances de la facette considérée (cf. couleur dans le schéma). La différence entre ces deux instances pour un même texte tient à la source lumineuse passée par le prisme, le contexte pédagogique.

La notation fonctionnelle, adoptée ci-dessus, souligne le fait qu'une facette, issue d'un prisme, dépend de deux variables : le texte et le contexte pédagogique. Plutôt que de passer par une notation faisant intervenir crochets et parenthèses, nous aurions pu indiquer le couple (*texte*, *CP*).

10. cf. [40] dans le tableau F.25 p. a-64.

Cependant notre notation est plus proche de la manière dont les facettes seront utilisées pour le repérage des textes. En effet, au niveau conceptuel, c'est la facette qui est le critère de recherche (dans la figure 6.4, recherche en fonction du nombre d'occurrences d'une structure donnée). Au niveau du texte, ce que nous avons appelé la décision dans la figure 5.3 (p. 246) s'effectue selon la facette de texte mais pour un certain contexte pédagogique (ex : recherche en fonction du nombre d'occurrences de **présent simple**)¹¹. La distinction d'un texte par rapport à un autre se fera pour un contexte pédagogique défini par l'utilisateur et donc selon l'instance de facette particulière correspondant ; d'où la notation introduite ci-dessus qui ne place pas au même niveau le contexte pédagogique et le texte.

Le couple prisme-facette doit donc nous permettre d'effectuer une indexation cohérente prenant en compte le contexte pédagogique. Le prisme est l'outil qui permet, au niveau de la conception de la base de textes, l'ajout de critères potentiels. Chaque propriété prise en charge par un système utilisant ce modèle l'est par l'ajout d'un prisme, dont la qualité de « prisme » tient à son association à une facette. La recherche étant centrée autour de la définition d'un contexte pédagogique (cf. § c) p. 212), chaque facette peut être instanciée pour un CP donné, pour lequel le prisme pourra calculer une valeur pour chaque texte (facette du texte dans un contexte pédagogique donné).

6.2.4 Prismes et contexte pédagogique

Un système d'indexation pédagogique de textes pour l'enseignement des langues permet donc, suivant ce modèle, d'observer chaque texte à la lumière d'un contexte pédagogique donné à travers un ensemble de prismes. Pour un unique CP, chacun des différents prismes ne considère que les éléments qui lui sont nécessaires. Il convient de s'interroger sur la manière dont il décompose ce CP pour n'en retenir que les traits qu'il prend en charge.

a) Décomposition du CP

Dans la figure 6.5, un faisceau est transmis à 3 prismes, chacun desquels permet d'extraire la composante permettant de révéler la facette correspondante du texte *T*. Mais, si nous sortons un instant de notre métaphore pour exprimer ce fait en termes de traitements et d'interactions homme-machine, il y a une inconnue concernant le moyen par lequel le contexte pédagogique est décomposé. En effet, une possibilité serait de prendre au pied de la lettre la métaphore et de considérer qu'un contexte pédagogique global est défini, un ensemble de prismes choisis et que ces derniers analysent le CP et en tirent les informations nécessaires au bon déroulement de la procédure. Outre l'extrême difficulté technique¹², une telle solution présente peu d'intérêt dans la mesure où elle demanderait à l'enseignant d'ajouter des données sans lui préciser lesquelles sont utiles et lesquelles sont inutiles. Ceci irait directement à l'encontre de la 4^e loi des Sciences de l'Information et des Bibliothèques (SIB) de Ranganathan à savoir « économiser le temps du lecteur¹³ » [Ranganathan, 1965, p. 61]. La stratégie que nous adoptons consiste à ne demander

11. Nous traitons de la recherche de document via les facettes dans le paragraphe 6.3 p. 263.

12. Comment permettre à l'enseignant de décrire des composantes du contexte pédagogique qui ne sont prises en charge par aucun prisme ? Quel formalisme permettra de donner vraiment toutes les informations nécessaires ? Si un tel formalisme peut être créé, comment permettre à chaque prisme d'analyser l'intégralité des données disponibles pour en choisir précisément les données qui le concernent ?

13. « *Save the time of the reader* ».

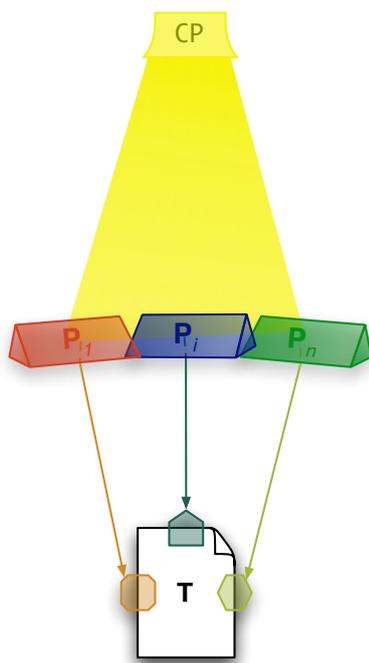


FIGURE 6.5: Facettes d'un texte

à l'enseignant que les composantes du contexte pédagogique qui seront traitées par les prismes sélectionnés.

b) Formalisation d'un prisme

Cette constatation rappelle un aspect essentiel des prismes que nous avons mis en évidence avec la notion de contexte pédagogique. Dans la section 5.4.2 (p. 249), nous faisons remarquer qu'une base de textes indexée pédagogiquement pour l'enseignement des langues dépendrait de la définition de formalismes pour décrire les propriétés du CP et des textes, ainsi que d'un système de mise en relation des propriétés des deux ensembles. Comme nous le faisons remarquer, chaque prisme permet de prendre en compte un sous-ensemble des propriétés du CP (cf. figure 6.4 p. 260), lesquelles doivent être interprétables selon le formalisme qu'il emploie. C'est en fonction de la modélisation de la facette associée, que le prisme fera appel à telle ou telle propriété du contexte pédagogique. La conception d'un prisme nécessite donc la définition de plusieurs éléments :

- facette (que représente la propriété évaluée ?) ;
- données d'entrée :
 - texte (quelles sont les informations contenues dans le texte qui seront nécessaires au calcul de ses facettes ? Comment les formaliser ? Comment les obtenir ?) ;
 - CP (quelles sont les composantes du contexte pédagogique qui seront nécessaires au calcul de ses facettes ? Comment les formaliser pour la formulation par l'utilisateur et pour le traitement associé ?) ;
- données de sortie : valeur de la facette (les valeurs possibles, ensemble de définition, sémantique) ;
- traitement (la procédure permettant d'obtenir les données de sortie à partir des données d'entrée).

Comme l'indique la définition 6.1 (p. 256), la spécification d'un prisme dépend toujours de la définition d'une propriété du monde réel en vue de l'exploitation du texte dans l'enseignement des langues. La conception d'un prisme pour cette facette nécessite l'étude de la perception de la propriété par les enseignants et sa modélisation en vue de la conception d'un traitement. Ce dernier fera donc appel à une formalisation du texte et du contexte pédagogique sous forme de deux ensembles propriétés (sous-ensembles respectifs des univers des propriétés du texte et du CP). Cette analyse de la facette dans le monde réel et sa modélisation dans un système d'indexation pédagogique doivent permettre d'associer des valeurs aux couples (*texte, CP*) formalisés et de définir un mécanisme réalisant cette association.

c) Outil descriptif

Par le mode de décomposition du CP que nous avons adopté et la spécification des principaux aspects de la définition d'un prisme, nous lui conférons une fonctionnalité expressive. En effet, puisque chaque prisme est associé à un (ou plusieurs) aspects du contexte pédagogique, le choix du prisme et la fourniture des informations nécessaires à son utilisation vont permettre à l'enseignant d'exprimer les caractéristiques de la situation d'enseignement (et leurs conséquences sur les textes) qui entreront en jeu dans la phase de sélection. Selon les formalismes proposés, l'action d'expression du CP peut demander à l'enseignant d'en affiner ou reformuler sa conception. Ceci est tout à fait cohérent avec la conception du processus de recherche de texte comme concomitant avec la définition du contexte pédagogique (cf. § d) p. 210).

6.3 Sélection et prismes

Les prismes que nous avons décrits jusqu'à maintenant – P_{nbmots} (§ a) p. 257), P_{etRep} (§ 6.2.3 p. 259) et P_{auteur} (figure 6.2 p. 258) – permettent, quand c'est nécessaire, de prendre en compte un contexte pédagogique dans l'expression et le calcul de propriétés des textes (valeur des facettes). Nous avons montré que ces propriétés intervenaient dans le choix des textes par les enseignants ; le fait d'y avoir accès serait donc une aide. Mais l'unique fait d'avoir accès à ces propriétés n'est pas une fin en soi. Avec ces seules informations, un utilisateur enseignant devrait passer en revue tous les textes du système confrontant les facettes des textes avec les propriétés qu'il recherche. Il effectuerait ainsi lui-même une tâche systématique de discrimination, qui pourrait faire partie de la tâche de sélection effectuée par le système. Les facettes proposées permettent, par exemple, de savoir qu'un texte contient n occurrences d'une structure donnée, que ce n soit 0 ou 15. Bien évidemment, un texte ne contenant aucune occurrence de la structure recherchée par l'enseignant ne lui sera d'aucune utilité (si cette structure est censée faire l'objet de la leçon¹⁴). Un système d'indexation pédagogique de textes doit donc être capable non seulement de prendre en compte le contexte pédagogique pour la description des ressources contenues, mais de l'utiliser pour prendre des décisions.

Dans la figure 5.3 (p. 246), nous replaçons les résultats du second questionnaire dans une perspective de sélection¹⁵ effectuée par le système. Nous avons suggéré que les propriétés des textes étaient influencées par le contexte pédagogique, ce qui correspond aux prismes et facettes tels

14. Même si un enseignant ne spécifiera pas sa recherche en termes de tout ce qu'un texte ne doit pas contenir, il est imaginable, qu'il exclue certaines structures, par exemple pour prendre en compte le niveau des apprenants ou la progression définie. À ce titre l'absence d'une structure pourrait s'avérer être une information pertinente.

15. Cf. § 4.6.4 p. 205 et plus particulièrement § a) p. 206.

que nous les avons définis ici. Toutefois, nous faisons la proposition que la prise de décision dans une optique de sélection pouvait être traitée comme une propriété contextuelle du texte. Nous allons voir ici les conséquences de ce postulat sur la notion de prisme.

6.3.1 La contrainte comme fonctionnalité du prisme

Nous pouvons expliquer le fait de traiter la « décision » comme une propriété du texte à partir des définitions que nous avons données. Pour modéliser cette discrimination entre les textes satisfaisant la requête de l'utilisateur et les autres, deux options s'offraient à nous : laisser à une entité à part le soin de la contrainte des facettes ou inclure cette contrainte dans le prisme chargé de les calculer. D'après sa définition (6.2), la facette correspond à une valeur pour un texte selon un prisme *dans un certain contexte pédagogique*. Or, il semblerait que le fait de contraindre les facettes fasse précisément partie de ce que nous avons appelé le contexte pédagogique. Si un enseignant recherche un texte pour un exercice structural sur le *prétérit* en anglais, un texte n'ayant que trois ou quatre occurrences de *prétérit* aura peu de chance de l'intéresser, alors que cela pourrait être suffisant pour introduire la notion (cf. tableau 5.7 p. 240). Pour un enseignant conscient de ces ordres de grandeur, il s'agit d'une propriété du contexte pédagogique. Les prismes doivent donc pouvoir prendre en compte les contraintes portant sur les valeurs des facettes calculées.

6.3.2 Prismes pour la sélection

L'intégration d'une forme de prise de décision à un prisme revient à en créer un nouveau, qui non seulement calculera une valeur pour une propriété, mais qui en évaluera ensuite la conformité par rapport à des contraintes. Comme un prisme calcule une valeur à la donnée d'un CP et d'un texte, à partir du moment où des relations d'équivalence¹⁶ ou d'ordre¹⁷ sont définies sur l'ensemble des valeurs de sortie (cf. § b) p. 262), il est possible d'augmenter le contexte pédagogique de tests. Ces derniers peuvent être extrêmement simples, un test d'égalité basique (La facette du texte est-elle celle demandée par l'utilisateur dans sa requête ?) ou selon la complexité de la facette beaucoup plus sophistiqués. Par exemple en pondérant les informations du contexte pédagogique et en proposant des seuils de tolérance. Mais dans un cas comme dans l'autre, la prise en compte de la contrainte par le prisme peut se réduire un algorithme simple permettant l'ajout d'un test.

16. Une relation d'équivalence R sur un ensemble E est une relation binaire réflexive (tout élément est en relation avec lui-même : $\forall x \in E, xRx$), symétrique (si x est en relation avec y alors y est en relation avec x : $\forall(x, y) \in E, xRy \Rightarrow yRx$) et transitive (si deux éléments sont en relation avec un même troisième alors ils sont en relation entre eux : $\forall(x, y, z) \in E^3, (xRy \wedge yRz) \Rightarrow (xRz)$) [Wikipedia, 2009e]. Ainsi l'égalité est une relation d'équivalence. Dans le contexte de notre travail, une relation d'équivalence permet de prendre une décision quant à l'adéquation d'un texte pour un CP, mais elle ne permet pas d'ordonner les textes en fonction de cette adéquation.

17. Une relation d'ordre R sur un ensemble E est une relation binaire réflexive (tout élément est en relation avec lui-même : $\forall x \in E, xRx$), antisymétrique (si x est en relation avec y , y ne peut être en relation avec x que s'il est égal à ce dernier : $\forall(x, y) \in E, (xRy \wedge yRx) \Rightarrow x = y$) et transitive (si deux éléments sont en relation avec un même troisième alors ils sont en relation entre eux : $\forall(x, y, z) \in E^3, (xRy \wedge yRz) \Rightarrow (xRz)$) [Wikipedia, 2009f]. Ainsi la relation « inférieur ou égal à » est une relation d'ordre alors que la relation « strictement inférieur » n'en est pas une. Mais elle peut être associée à une relation d'ordre en restreignant celle-ci aux couples d'éléments distincts (relation d'ordre strict). Dans le contexte de notre travail, une relation d'ordre permet de prendre une décision quant à l'adéquation d'un texte pour un CP, mais en plus elle permet de caractériser et ordonner cette adéquation. Elle permettra de dire qu'un texte T_1 est « plus adéquat » pour le contexte pédagogique considéré que le texte T_2 .

6.3.3 Modification du traitement d'un prisme pour permettre la prise de décision

Prenons l'exemple du prisme permettant d'accéder au nombre d'éléments représentatifs d'une notion ($P_{e^t Rep}$), nous pourrions créer un prisme $P_{e^t RepC}$, qui calculera le nombre d'éléments représentatifs de la même manière que $P_{e^t Rep}$, mais qui ne renverra une valeur que pour les textes suivants les contraintes définies par l'utilisateur. La différence, du point de vue du traitement, entre $P_{e^t Rep}$ et $P_{e^t RepC}$ tient à l'utilisation d'une unique fonction supplémentaire (cf. ligne 1 de l'algorithme 1).

Algorithme 1 : Exemple de fonction permettant de créer une facette contrainte à partir d'une facette non contrainte

Data : $F_{e^t Rep[CP]}(T)$, la facette de T selon $P_{e^t Rep}$ pour le contexte pédagogique CP

Input : les contraintes Ct propres au contexte pédagogique utilisé dans $F_{e^t RepC}$

Result : $F_{e^t RepC[CP|Ct]}(T)$, la facette de T selon $P_{e^t RepC}$ pour le contexte pédagogique CP augmenté des contraintes Ct

```

1 if  $F_{e^t Rep[CP]}(T)$  satisfait Ct then
2   | Résultat  $\leftarrow F_{e^t Rep[CP]}(T)$ ;
3 else
4   | Résultat  $\leftarrow$  NULL;
5 end
6 return Résultat ;

```

a) Conséquences conceptuelles

Un prisme ainsi défini, n'a pas de propriétés particulières par rapport à la définition que nous en avons donnée. Il s'agit toujours d'un traitement associé à une propriété (qui prend en compte des composantes de contraintes). Cette propriété est toujours calculée pour chaque texte en fonction d'un contexte pédagogique. Le travail pour arriver à des prismes capables de « prises de décision » réside essentiellement dans le sens associé à la facette traitée, et les décisions de modélisation des données d'entrée.

Dans l'exemple de F_{auteur} , le test pourrait être un test d'égalité, il nécessite comme unique information supplémentaire au contexte pédagogique de la facette de base, le nom de l'auteur recherché. Par contre dans le cas de F_{nbmots} , un tel test n'aurait pas de sens. L'utilisateur ne va vraisemblablement pas rechercher un texte selon sa longueur au mot près. Le contexte pédagogique ne sera donc pas augmenté que de la valeur attendue (le nombre de mots recherché), mais aussi du type de test : au moins un certain nombre de mots (\geq), au plus un certain nombre de mots (\leq). Pour ces deux tests simples, la nature de la contrainte doit être intégrée au contexte pédagogique. Mais le test pourrait tout aussi bien prendre la forme d'une tolérance en mots ou en %, le contexte pédagogique devra alors inclure cette tolérance et son unité, si elle n'est pas fixée. Une autre alternative serait d'exprimer cette contrainte sous forme d'intervalle de confiance du type $[nbmotsMin; nbmotsMax]$. Dans ce cas là, les valeurs extrêmes de l'intervalle constitueront le contexte pédagogique.

b) Facette vide

Pour compléter cette « faculté de décision », un travail doit également être fait sur les données de sortie. Le choix que nous faisons ici est de concentrer dans les prismes toutes les informations du contexte pédagogique que le système sera à même de traiter. La contrainte des valeurs des facettes en fait partie. Dans un souci de concision, nous parlerons de *facette vide* d'un texte lorsque la facette associée prend en charge la contrainte et que la vue du texte ne la satisfait pas (cf. ligne 4 de l'algorithme 1). Nous réutiliserons ce concept dans la partie liée à l'implantation du modèle (cf. § 1.4 p. 15).

6.4 Facette vs métadonnées

Suite à la définition de la notion de contexte pédagogique (cf. définition 4.1a p. 211) et son adaptation à la problématique de la description de ressources (cf. définition 4.1b p. 250), nous proposons dans les pages qui précèdent une modélisation des propriétés des textes qui permet la prise en compte d'informations contextuelles. Pour y parvenir nous utilisons deux entités, chacune nécessaire à l'existence de l'autre, la facette et le prisme. Ces entités permettent :

- d'une part, d'associer le concept (facette) et sa modélisation, rendant ainsi explicite l'acceptation du concept prise en charge par l'outil (prisme)¹⁸ ;
- d'autre part, de modéliser l'influence du contexte pédagogique sur les propriétés des ressources.

6.4.1 Informations sur des ressources

Ce sont ces deux caractéristiques qui différencient en premier lieu métadonnées et facettes. En revenant à la définition de métadonnées de Bourda [2002] (cf. définition 3.5 p. 93), selon laquelle elles sont des « informations sur des [ressources,] compréhensibles par les êtres humains et traitables par des logiciels », les facettes sont proches des métadonnées par la similitude de l'objet de ces propriétés : des informations *sur des ressources*. Il ne s'agit ni dans un cas ni dans l'autre de décrire les informations *des* ressources, mais de pouvoir porter un regard général sur un objet par des informations qui le décrivent dans sa globalité.

Cette dimension est flagrante dans les facettes que nous présentions dans la figure 6.2 (p. 258), qui désignent des propriétés qui pourraient être utilisées telles quelles dans un standard de métadonnées : F_{auteur} correspond par exemple à l'élément 2.3, **Contribute** du *Learning Object Metadata* (LOM) [IEEE, 2002, pp. 16–17] alors que la taille pourrait être associée au raffinement **Extent** du descripteur **Format** dans le profil d'application *Educational Network of Australia* (edna) du *Dublin Core Element Set* (DCES) [edna, 2006, p. 4]. Cependant, cette équivalence n'est possible que pour les facettes de texte constantes pour tout CP.

De la même manière que $f(x) \rightarrow 0$ est une fonction de x , un prisme qui, à l'instar de P_{nbmots} , attribue une unique facette à chaque texte quel que soit le contexte pédagogique, continue de satisfaire la définition que nous en avons donné (cf. définition 6.1 p. 256).

Au contraire, une facette comme F_{etRep} n'est pas représentable sous forme de métadonnées, bien

18. Il s'agit en quelque sorte des fonctions constituantes de l'acceptation de la propriété, que se réapproprient les utilisateurs (cf. [Contamines *et al.*, 2003] § c) p. 61).

que le nombre d'occurrence d'une structure soit une information *sur le texte*. Pour être prise en charge sous forme de métadonnées, celle-ci nécessiterait un descripteur par contexte pédagogique. Un élément de métadonnées correspondrait donc à chaque instance de facette (cf. § 6.2.3 p. 259) et donc à chaque CP possible. Pour fonctionner comme critère d'indexation sous forme de métadonnées, l'annotation devrait pouvoir anticiper toutes les structures utiles à au moins un utilisateur. L'objectif d'intégration à un système d'indexation pédagogique exclut cette solution¹⁹.

a) Multiplication des descripteurs

La détermination *a priori* de toutes les structures pertinentes pour l'enseignement devrait, pour que l'outil ne réduisent pas l'univers des possibles pour les enseignants, considérer toutes les combinaisons imaginables. Le cas de la facette $F_{e^t Rep}$ est un exemple frappant. Imaginons vouloir permettre à l'enseignant de chercher par séquence de deux mots.

Imaginons un texte constitué de l'unique phrase « les lapins courent ». Pour que le texte puisse être correctement indexé selon $P_{e^t Rep}$, les métadonnées le concernant devra contenir un descripteur par requête possible. Cela dépendra du formalisme utilisé, imaginons que l'utilisateur puisse spécifier la catégorie morphologique, le lemme et le mot-forme. Le texte devra être repérable selon le nombre d'occurrences de chaque forme, soient 3 descripteurs²⁰. Il faudrait également pouvoir chercher les occurrences de « les lapins » et de « lapins courent » (2 descripteurs). Il faut aussi permettre de rechercher le texte selon les lemmes²¹, tout en permettant les combinaisons comme pour les mots-formes. Si l'on veut pouvoir trouver tous les textes concernant des lapins qui courent, quel que soit le moment de leur course, il faudra pour trouver le texte, le décrire comme possédant une occurrence de `<nb struct="lapins courir*" valeur="1"/>`. De même pour le ou les lapins. Et ainsi de suite, en prenant en compte la possibilité de chercher des noms, au pluriel, des verbes, au présent, à la 3^e personne du pluriel, etc. Considérant la quantité de métadonnées pour un texte de 3 mots, il est difficile d'imaginer la description d'un texte qui en contiendrait 500.

Dès lors qu'il est pertinent que la propriété concernant le texte dans sa globalité soit affectée par le contexte pédagogique, il devient hasardeux de s'en remettre à une description sous forme de métadonnées. Si les questions des enseignants peuvent être un minimum complexes, ce mode de description n'est pas utilisable du fait de l'explosion combinatoire survient.

b) Recherche plein texte

Au contraire, le pendant procédural de chaque propriété permet à la facette de systématiser un lien entre le contenu du texte et les données destinées à le décrire dans sa globalité. L'indexation

19. Encore une fois, il nous faut rappeler nos objectifs dans le cadre de ce travail. Nous désirons ici pouvoir proposer aux enseignants des textes qui satisfont leurs requêtes sans pour autant que les usages de ces derniers aient été attestés. Ceci limite les facettes possibles puisque celles-ci ne peuvent pas contenir des informations qui sont l'expression d'un point de vue d'un acteur. Un détournement du concept pour prendre en charge des propriétés subjectives consisterait à intégrer l'annotateur au contexte pédagogique. Cependant ce type de réflexion nous paraît prématuré dans le cadre de ce travail. Il pourrait cependant en constituer une extension à long terme.

20. `<nb struct="les" valeur="1"/>`, `<nb struct="lapins" valeur="1"/>`, `<nb struct="courent" valeur="1"/>`.

21. `<nb struct="le*" valeur="1"/>`, `<nb struct="lapin*" valeur="1"/>`, `<nb struct="courir*" valeur="1"/>`.

centrée sur des facettes permettrait alors de rechercher les textes en fonction de propriétés globales protéiformes.

Strictement parlant, il ne s'agit pas de recherche plein texte, même si cette dernière pourrait permettre de résoudre le problème du prisme $P_{e^t Rep}$, par exemple pour des textes dont l'analyse morphologique serait disponible au format *eXtensive Mark-up Language* (XML) [Amer-Yahia *et al.*, 2006]. En effet, la recherche plein texte est un outil (intégrable au sein d'un prisme par exemple) pour la mise en place d'un système utilisant des facettes. Celles-ci sont associées à une sémantique issue de la didactique des langues. Chaque prisme fournira un point de vue sur le texte se voulant pertinent pour l'enseignant. Dans la notion de facette, le traitement est indissociable de l'objectif poursuivi, du sens associé aux propriétés évaluées.

Pour fournir ces différents points de vue, l'une des difficultés est de faire coexister ces différents prismes au sein d'un même système. La facette constitue donc un concept proche dans son objet des métadonnées, mais qui offre la possibilité d'une indexation selon des propriétés dépendant du contexte, s'appuyant pour y parvenir sur des traitements, utilisable par ailleurs à d'autres fins d'indexation. Il s'agit donc d'une indexation moins figée que si elle était effectuée à partir de métadonnées, tout en conservant la sémantique de ces dernières.

6.4.2 Données sous-jacentes

Pour calculer la valeur de certaines facettes (informations *sur le texte*), il est nécessaire d'avoir accès à des informations *du* texte. Revenons par exemple aux facettes F_{nbmots} et $F_{e^t Rep}$. Dans un cas comme dans l'autre en amont du calcul de la valeur de la facette, un certain nombre de traitements est nécessaire : le comptage du nombre de mots s'appuiera sur le découpage du texte en mots ; le calcul du nombre d'éléments représentatifs d'une notion sur une analyse morphologique. Nous appellerons ces informations intermédiaires (portant sur le texte ou son contenu) nécessaires au calcul des valeurs des facettes, « données sous-jacentes ».

Dans le cas des facettes « constantes », la valeur même de la facette peut être une donnée sous-jacente, pour laquelle le prisme, outre ses fonctions conceptuelles, ne servira qu'à récupérer la valeur et éventuellement la contraindre. Cette valeur peut donc être le fruit du travail d'un annotateur humain. Dans ce cas le prisme spécifiera les conventions d'annotation (formalisme, données d'entrées). L'auteur du texte²² (F_{auteur}) est un exemple d'une telle facette. Sa qualité de facette est attestée par la valeur pédagogique évidente lorsqu'il s'agit d'un cours de littérature. Mais cette caractéristique peut aussi avoir une influence dans le cadre de la sélection, dans un contexte plus généraliste²³.

Un autre type de données sous-jacentes concerne les propriétés des éléments du textes (et non du texte lui-même). Dans ce cas, les données sous-jacentes ne sont pas des métadonnées pour le texte en tant qu'objet. C'est par exemple, le cas du résultat d'un découpage en mots où d'une analyse morphologique.

Conformément à nos remarques pour F_{auteur} ces données ne sont pas nécessairement ajoutées

22. Le lien fait avec le texte n'est donc pas uniquement un reflet de son contenu. S'il existe des travaux sur la reconnaissance automatique de l'auteur d'un texte (cf. [Prince *et al.*, 2007]), ce type d'information reste en premier lieu une information qui concerne le texte dans sa globalité et qui ne découle pas nécessairement directement de son contenu au sens le plus strict du terme. Nous mettons l'accent sur les autres possibilités de facettes car elles se différencient plus radicalement des métadonnées.

23. Voir la citation invoquant Roald Dahl, § a) p. 206.

par un outil informatique. Si la procédure « d'annotation » du texte est suffisamment précise pour être utilisable dans des facettes, il n'y a pas, comme le signale la définition 6.1 (p. 256) d'obligation à ce que toutes les étapes du processus soient effectuées automatiquement. Le projet Plan Pluri-Formation (PPF), *Développement et exploitation de ressources linguistiques pour la didactique du français à l'aide d'outils TAL. Étude des marqueurs linguistiques des émotions* (auquel nous nous référerons dorénavant par l'acronyme PPF), est un bon exemple de cet état de fait.

a) Exemple du projet PPF

En cas d'intégration des données du projet PPF [Augustyn *et al.*, 2008], les données sous-jacentes seraient également entrées manuellement. Ce projet visait à la constitution et l'exploitation de ressources linguistiques pour la didactique du français, à destination d'un public de Français Langue Étrangère (FLE) (niveau intermédiaire à avancé) et d'un public de Français Langue Maternelle (FLM) (collège et lycée). Une partie de ce projet concernait le discours rapporté. Après avoir défini une typologie, qui paraissait cohérente avec l'objectif d'utilisation des ressources en cours de français, les discours rapportés du corpus ont tous été annotés manuellement [Rinck & Lebarbé, 2005]. Ces annotations manuelles pourraient tout à fait constituer les données sous-jacentes nécessaires à la création d'une facette de comptage des discours rapportés d'un type donné (F_{DR}). Un programme peut analyser les données du corpus²⁴ pour calculer par exemple F_{DR} pour le contexte pédagogique `discours indirect libre`.

6.5 L'aide à l'évaluation comme complément de la sélection

Les facettes, comme le suggère leur définition (cf. définition 6.2 p. 257), concernent des propriétés sur les textes. En effet, ce point de vue global est inhérent à la tâche de sélection même, qui applique aux textes des propriétés projetées (ou préalablement affectées) en vue de sélectionner les textes qui auront le plus de chance de satisfaire un besoin. À partir du moment où l'enseignant consulte le contenu d'un texte dans le cadre de sa recherche, il sort de la phase de sélection pour passer à celle d'évaluation (cf. § b) p. 208).

L'action des prismes est tournée vers la mise en place d'un système qui permette d'augmenter la portée de la phase de sélection en considérant un nombre de textes plus élevé et en appliquant des critères de sélections relevant plus directement de la problématique de l'enseignement des langues. Elle permet d'associer des propriétés contextuelles qui font le lien entre le contenu du texte, un contexte pédagogique et une propriété du texte en tant qu'objet. Toutefois, la donnée obtenue est toujours une valeur modélisant le texte, sans donner accès à ce dernier. Le couple facette/prisme fournit un cadre propice à proposer des informations sur le texte cohérentes et nombreuses, mais nous avons mis en exergue dès le début de notre travail sur la planification le fait que les enseignants n'ont pas toujours une manière rationnelle d'expliquer leur action (cf. § 4.1.8 p. 145), que certains critères ou aspects du processus de recherche seront difficilement formalisables (cf. « plaisir, intérêt culturel, intérêt des élèves »^[86], § d) p. 185). Les facettes proposées n'ont pas la prétention de proposer à l'enseignant « le » texte répondant à ses besoins. Elles sont censées permettre à l'utilisateur d'effectuer son choix parmi des textes, les plus proches possibles de sa question, mais étant donnée la quantité de facteurs intervenant dans le choix, leur

24. Il s'agit d'annotations XML.

introduction dans un système ne change pas le paradigme de la recherche de textes et demeure un préambule à la phase d'évaluation.

6.5.1 Approche moins-disante et prolongement de la problématique

La stratégie moins-disante impose la question subsidiaire de l'aide au choix de textes (cf. § 1.2.3 p. 9) à la problématique de l'indexation (dont elle n'est que le prolongement). Les études de la manière dont la présentation des résultats de Google influence leur consultation par les utilisateurs²⁵ et les commentaires sur les propositions alternatives (comme celle de Cuil²⁶) rappellent la complémentarité entre indexation des ressources et présentation des résultats à l'utilisateur. En effet, il y a des aspects de la tâche d'évaluation qu'un système sera difficilement à même de modéliser. Toutefois, une aide au choix peut s'articuler autour de la proposition, en plus de points de vue génériques, de représentations des textes permettant à l'utilisateur d'accéder immédiatement à des informations affectant sa décision d'avoir recours ou non à un texte en particulier pour une situation d'enseignement/apprentissage.

Les éléments que nous allons présenter ici, ne sont à notre sens pas des éléments du modèle pour l'indexation pédagogique de textes, mais bien un prolongement dont l'objectif est de rendre un système intégrant prismes et facettes le plus adapté possible à la problématique des enseignants. Pour traiter cet aspect du problème, d'autres recherches et expérimentations seraient nécessaires, en Interface Homme-Machine (IHM) notamment, que ce soit pour sa mise en oeuvre informatique ou pour l'étude des interactions effectives des enseignants avec le système. À ce stade, il s'agit d'hypothèses de travail qui demanderaient à être confirmées par l'usage (ce qui n'est qu'une perspective à long terme).

6.5.2 Évaluation et *visualisations*

Alors que la tâche de sélection doit permettre de considérer les textes à travers leurs propriétés globales, la tâche d'évaluation dépend de leur consultation par l'enseignant. Ce dernier est, à ce moment là, l'acteur qui assigne les propriétés aux textes à partir de leur contenu. Il doit pouvoir porter un regard « qualitatif » et personnel sur les facettes qui ont permis la sélection, en plus d'évaluer d'autres propriétés selon son point de vue.

Sans nous lancer dans une discussion sur les facettes que nous pourrions modéliser ou non (ex : intérêt que présente le texte du point de vue personnel des apprenants), la difficulté de prise en charge des informations qualitatives est claire. Par exemple, repérer les occurrences d'une structure donnée peut s'appuyer sur des outils existants²⁷. Évaluer l'adéquation des occurrences des structures avec les besoins précis de l'enseignant²⁸ est beaucoup plus difficile. Ne pouvant

25. <http://www.miratech.fr/newsletter/eye-tracking-google1.html> .

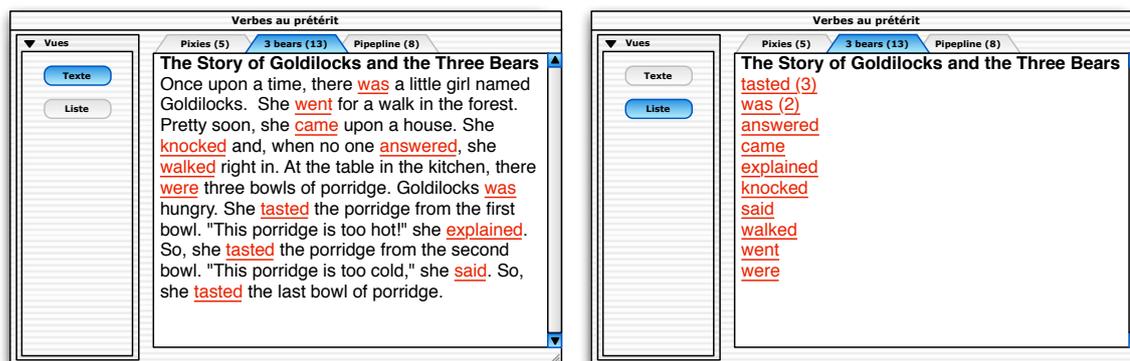
26. <http://www.techcrunch.com/2008/07/27/cuill-launches-a-massive-search-engine/> .

27. Nous pouvons nous contenter d'utiliser des outils TAL comme TreeTagger.

<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

28. Repérer les occurrences de prétérit et de past progressive (prétérit progressif) est beaucoup plus aisé que de repérer les usages en contraste de l'un et de l'autre visant à exprimer une action ponctuelle en contraste avec une action longue (cf. « Occurrence des deux présents ou des deux prétérêts dans un texte en anglais » [216], tableau F.25 p. a-64).

Par exemple : « I was working on the proof of one of my poems all the morning, and took out a comma. In the afternoon I put it back again ». Oscar Wilde (<http://www.quotationspage.com/quote/26788.html>).



(a) En contexte

(b) Index des fréquences

FIGURE 6.6: Exemple de visualisations d'un texte pour un même contexte pédagogique

(ni ne voulant) faire le choix du texte automatiquement, le système peut proposer à l'utilisateur-enseignant une aide : des représentations des textes liées au contexte pédagogique. Ces représentations graphiques sont appelées visualisations du texte.

a) Étude de cas

Revenons à la facette F_{etRepC} (cf. algorithme 1 p. 265). Nous allons considérer un contexte pédagogique centré sur **verbes au prétérit** en anglais.

Imaginons que l'enseignant recherche des textes contenant au moins 5 occurrences du prétérit et que la sélection effectuée par le système lui propose 3 textes à évaluer. Dans figure 6.6, nous donnons un exemple de ce que pourraient être deux visualisations différentes liées à ce contexte pédagogique. Ces éléments ont été repérés pour pouvoir être comptés (valeur de F_{etRepC}). Une visualisation (réutilisant les données sous-jacentes nécessaires au calcul de la facette) pourra donc consister à les mettre en évidence dans le texte ou juste à dresser une liste des occurrences. La première vue permet à l'enseignant de repérer facilement les emplois en contexte. La seconde, un index des fréquences²⁹ regroupant les verbes au prétérit du texte offre la possibilité d'accéder immédiatement à la diversité des formes présentes dans le texte³⁰

6.5.3 Dimension fonctionnelle de la visualisation

Dans les premières versions du modèle (cf. NB 6.1 p. 253), la visualisation était une entité du modèle³¹ à part entière. Il nous semble cependant qu'elle n'intervient pas sur le même plan que prismes et facettes. En effet, elle tient à la conception d'un système et à la nécessité de présenter

En effet, ce dernier cas nécessite des informations supplémentaires comme la notion de phrase. Et pour pouvoir en faire un exemple nous avons utilisé un critère conceptualisable, or nous ne savons pas comment les enseignants prennent leur décision d'un point de vue plus qualitatif, pour une propriété donnée.

29. « index [...] où, pour chaque mot, [on trouve] le nombre des occurrences » [Guilbaud, 1980].

30. Il est bien sûr possible d'imaginer d'autres exemples de visualisations pour ce contexte pédagogique. Celles que nous présentons ici sont des pistes de réflexion.

31. Elle était appelée « visualisation ».

des résultats à un utilisateur. Même sans la moindre prise en compte de la phase d'évaluation, sans préoccupations pédagogiques, l'affichage des résultats nécessite l'existence d'au moins une visualisation. Les prismes et facettes, pour leur part, sont liées à une problématique précise au sein de laquelle la prise en compte du contexte pédagogique dans les propriétés des textes est apparue comme une nécessité. Elle s'inscrit dans un processus de recherche et peuvent être abstraites de la conception d'un système informatique pour appréhender ce processus d'un point de vue descriptif. Nous ne postulons pas ici la validité cognitive d'une telle affirmation, nous signifions ici la nature conceptuelle des notions en l'opposant à la nature fonctionnelle de la visualisation.

Bien évidemment, la conception d'un système est indissociable du problème qu'il doit traiter, et notre « point de vue sur les visualisations » reflète notre objectif d'aide au choix d'un texte dans le cadre de l'enseignement des langues, mais nous avons décidé de considérer la notion de visualisation comme un choix d'implémentation applicable à notre problématique.

6.5.4 Vue et contexte pédagogique

La visualisation peut donc avoir une fonction généraliste liée à des propriétés « pédagogiques » ou non³² (dépendant du contexte pédagogique ou non³³). Mais la notion de visualisation peut s'adapter à notre problème et prendre en charge le contexte pédagogique, à l'instar des deux exemples de visualisation présentés dans la figure 6.6 (p. 271). Dans ces exemples, les éléments au centre des visualisations émanent directement du contexte pédagogique d'une requête imaginaire. Notre système devrait, dans l'hypothèse de telles visualisations, pouvoir en générer d'autres en fonction du contexte pédagogique (par exemple *verbes au présent* ou selon les outils intégrés au système pour les *superlatifs*, etc.).

Si elles ne constituent pas à notre sens une entité du modèle, la conception de certaines visualisations devront prendre en charge des notions issues du contexte pédagogique pour offrir aux enseignants une aide au choix, ou plus précisément une aide à l'évaluation. En effet, il ne s'agit pas seulement d'indiquer aux enseignants quels textes le système considère les plus adaptés pour la demande de l'enseignant (ce problème serait plutôt traité selon les valeurs des facettes bénéficiant d'une relation d'ordre), mais plutôt de lui proposer des informations qui lui soient utiles pour l'évaluation (qui mène à la transformation du texte, le choix d'en faire une ressource pédagogique pour un CP donné ou non). De ce fait, la conception d'un couple facette-prisme peut mener à la création de visualisations.

6.5.5 Vues et compréhension du système

Selon les prismes concernés, il peut être intéressant de donner accès, au cours de la phase d'évaluation, à des propriétés sous-jacentes nécessaires au calcul des valeurs des facettes. Comme nous l'avons déjà expliqué, ceci peut faciliter l'évaluation des textes par l'enseignant. À notre avis, il y a, dans une perspective de conception d'un système lié à l'Apprentissage des Langues Assisté par Ordinateur (ALAO), un effet de bord salubre à la présentation d'informations sous-jacentes à l'utilisateur. Même si ce n'est pas là leur objectif premier, il nous semble que l'accès qualitatif à ces notions donne à l'utilisateur non seulement des informations sur les textes, mais aussi sur le

32. Mise en forme générale des textes.

33. La présentation du texte sous forme d'un index des fréquences pourrait avoir un intérêt pour évaluer la diversité du vocabulaire employé mais ne serait pas dépendante du CP.

système. Nous avons souligné le fait que les réactions des enseignants confrontés à des systèmes d'ALAO pouvait être épidermiques et nous avons également expliqué que celles-ci provenaient souvent d'une incompréhension des possibilités du système (cf. § b) p. 59). Nous concluons ensuite à la nécessité que les enseignants comprennent le système qu'ils utilisent afin qu'ils puissent se l'approprier (cf. § 2.2.5 p. 63). Kübler, avec des utilisateurs-apprenants, évoque un point de vue analogue : « *They need to know how [NLP] works to become expert users of the tools that they will have to become familiar with.*³⁴ » [Kübler, 2002, p. 196]. Or, il nous semble qu'en proposant aux enseignants l'accès à des données sous-jacentes (ayant présidé au calcul des facettes), nous lui offrons une fenêtre vers l'outil qu'il utilise. Il n'est pas évident que tous les enseignants sauront ou seront intéressés à en tirer partie. Cependant, dans l'exemple de la facette $F_{e^t RepC}$, la conscience de ce que le système considère comme **prétérit** ou telle autre structure recherchée par l'enseignant, peut lui permettre de s'approprier le fonctionnement du système en vue de l'instrumenter (et pourquoi pas de l'instrumentaliser).

Cependant, avant de nous poser la question de l'instrumentation d'un système d'indexation pédagogique de textes pour l'enseignement des langues, il nous faut aborder le modèle sous un angle plus informatique, plus tourné vers l'implémentation. Globalement, nous sommes contenté d'introduire les concepts qui nous semblent nécessaires à la mise en oeuvre d'une base de textes indexée pédagogiquement pour l'enseignement des langues. Ceux-ci nous paraissent adéquats à la prise en charge de descripteurs pédagogiques, hautement contextuels dès lors qu'il s'agit de ressources brutes. Pour y parvenir nous formalisons le lien entre la propriété et la procédure permettant de la calculer.

La place faite aux traitements reflète nos préoccupations d'implantation. Malgré quelques décisions relevant de l'implantation (cf. § c) p. 263 et § b) p. 266), nous n'avons pas intégré ici les concepts de prisme et de facette à l'architecture d'un système informatique. Nous allons donc, dans la suite de notre travail proposer une mise en oeuvre de ces concepts qui reflète notre approche de la problématique de l'indexation pédagogique de textes pour l'enseignement des langues.

34. Ils doivent savoir comment fonctionne le TAL, pour devenir des utilisateurs-experts des outils avec lesquels ils vont devoir se familiariser (dans le cadre des activités présentées dans [Kübler, 2002]).

Chapitre 7

Mise en œuvre du modèle prisme-facette

Après avoir introduit la notion de contexte pédagogique, nous nous sommes interrogé sur ses conséquences dans la manière de considérer les propriétés des textes, ce qui a débouché sur les concepts de prisme et de facette. Ces concepts étaient avant tout tournés vers la tâche de sélection, du fait de l'aspect globalisant des informations proposées. Or les limites quant à la modélisation pédagogique des textes et notre décision de travailler selon la stratégie moins-disante, nous ont poussé à nous interroger sur la phase d'évaluation et à sa gestion par un système d'indexation pédagogique pour l'enseignement des langues. Cette gestion repose en partie sur la possibilité de générer pour chaque texte des visualisations différentes, dont certaines soient en lien direct avec la problématique de l'enseignement des langues.

Avant de proposer *une* conception d'un système réutilisant ces notions, il nous faudra nous interroger sur l'ensemble de textes manipulés par le système. Nous aborderons la conception par le biais de séquences d'utilisation et finirons par éprouver la faisabilité des spécifications ainsi produites en nous focalisant sur un sous-ensemble de fonctionnalités à implanter sous forme de maquette.

7.1 Collection de textes

En tant que base de textes, notre système doit pouvoir manipuler des objets « textes ». En effet, les entités définies prennent le texte comme référentiel : les prismes sont destinés à proposer des facettes d'un texte, pour un contexte pédagogique. De même, les visualisations, telles que nous les avons définies, doivent permettre de représenter les textes de différentes manières. Pour agir en tant que *base de textes*, un système doit pouvoir traiter plusieurs textes à la fois, comparer leurs propriétés et d'opérer une sélection d'après une requête.

Pour y parvenir nous proposons d'avoir recours à une entité que nous appellerons « collection », qui s'inspire des lignes directrices définies par Sinclair, que nous citions précédemment (cf. § b) p. 69) :

« Words such as collection and archive refer to sets of texts that do not need to be selected or do not need to be ordered or the selection and/or ordering do not need to

*be on linguistic criteria. They are therefore quite unlike corpora.*¹ »

[Sinclair, 1996, p. 5]

Pour justifier cette appellation, nous allons définir l'usage que nous voulons faire de ces objets et les confronter aux préceptes énoncés par Sinclair.

7.1.1 Collections et contexte pédagogique

Dans la perspective de l'indexation pédagogique pour l'enseignement, c'est-à-dire de pouvoir accéder à un sous-ensemble de textes répondant à une question de l'utilisateur (cf. définition 1.3 p. 12 et définition 1.4 p. 15), nous voulons pouvoir examiner la même instance de facette (cf. § 6.2.3 p. 259) d'un ensemble de textes. Ceci revient à considérer plusieurs textes sous le(s) même(s) prisme(s) avec le même contexte pédagogique.

La conséquence directe de l'application d'un prisme donné, selon un même Contexte Pédagogique (CP), à plusieurs textes tient aux relations que peuvent entretenir entre elles les valeurs des facettes. Cette comparaison tire son sens de la constance du CP. Or nous avons évoqué précédemment le type de relations que pouvaient entretenir les valeurs des facettes, du fait de la définition d'une acception de la propriété concernée. Selon la propriété, il peut s'agir de relations :

- d'équivalence : appliquées à plusieurs textes, elles permettent de les grouper ;
- d'ordre : appliquées à plusieurs textes, elle permettent de les ordonner.

En appliquant ces relations, qui ne s'excluent pas mutuellement, nous avons donc la possibilité de regrouper et d'ordonner les textes. En ajoutant à cela la notion de contrainte sur les valeurs qui peut être jointe au contexte pédagogique, nous avons la possibilité de sélectionner des textes.

Revenons à la notion de facette vide (cf. § b) p. 266), afin d'en constater l'utilité dans ce contexte. La notion de facette vide provenait de l'apposition d'une contrainte à une valeur. Selon qu'elle satisfaisait la contrainte ou non, nous considérions que la facette avait la valeur qui a été testée ou une valeur nulle. Appliquée à une collection de textes C , nous pouvons considérer qu'une image $C_{P_{[CP]}}$ de celle-ci selon un prisme P dans un contexte pédagogique CP est l'ensemble des textes T pour lesquels la facette obtenue par $P_{[CP]}$ n'est pas vide. Ce que nous pouvons résumer selon l'expression suivante :

$$C_{P_{[CP]}} = \{T \in C, F_{[CP]}(T) \neq \emptyset\}$$

Cette utilisation d'un prisme pour traiter un groupe de textes a des conséquences pour la mise en place d'une base de textes indexée pédagogiquement pour l'enseignement des langues (que nous indiquerons plus loin), mais cela soulève aussi une question sur la manière de considérer les prismes dans le système.

a) Retour sur la fonction expressive des prismes

Tout d'abord, il nous faut rappeler que les exigences d'une modélisation informatique de propriétés complexes nécessite le recours à une approximation inhérente à la tâche de modélisation (cf. § a) p. 14). Ensuite, en comparant la notion de facette à celle de métadonnée, nous avons remarqué que par le lien essentiel que nous faisons entre celle-ci et le prisme, ce dernier permettait de définir l'acception de la propriété qui était utilisée (cf. § 6.4 p. 266).

1. Cf. note ⁹⁰ p. 69.

Un enseignant impliqué dans une tâche de planification définit un contexte pédagogique, dans l'acception la plus générale du terme. Il s'agit d'un CP « du monde réel », intériorisé, mais qui correspond malgré tout à la situation d'enseignement qu'il est en train de planifier. Cette formulation non-standardisée de l'ensemble des traits d'une situation d'enseignement (cf. définition 4.1a p. 211), peut influencer plusieurs approximations d'une même propriété (plusieurs couples prisme-facette). Par conséquent un seul tel CP peut être traité de plusieurs manières différentes. Est-ce à dire qu'une même composante d'un CP peut être utilisée par plusieurs prismes ?

Si deux prismes différents considèrent la même propriété du monde réel, alors il y a une différence dans les traitements ou les données d'entrées et de sortie et donc l'acception de la propriété est différente elle aussi d'un prisme à l'autre. En effet, pour prendre un exemple simpliste, le même concept de longueur peut être désigné selon une liste de valeurs (long, moyen, court) ou un nombre de mots, qui peuvent refléter des approches différentes dans deux prismes. Il ne s'agira donc pas de la même composante du CP selon que la propriété est associée à l'acception d'un prisme P_1 ou P_2 . En d'autres termes, un contexte pédagogique dans sa forme la plus générale n'est pas formulé en termes de prismes. C'est donc la fonction expressive (cf. § c) p. 263) du prisme qui permet de prendre en charge le contexte pédagogique au niveau du système et, par conséquent, sa définition progressive (cf. § d) p. 210).

Pour obtenir une collection C_2 issue de la collection C_1 nous pouvons donc considérer qu'il suffit de lui apposer un contexte pédagogique, puisque ce dernier est exprimé par le biais des prismes. Au sein du système, le prisme devra donc être lié à un module permettant à l'enseignant de formuler sa requête et de faire le lien entre le CP ainsi défini et les facettes correspondantes y compris la conséquence de leur nullité.

b) Collections ordonnées

La première conséquence de l'action des prismes concerne les transformations qui peuvent être appliquées à une collection. En ayant recours à des prismes n'offrant pas de possibilité de contrainte des facettes, une collection est malgré tout différente de celle dont elle est issue. En effet, à partir du moment où des valeurs de facettes sont prises en compte, les relations qui peuvent être définies sur leurs valeurs influencent la collection obtenue. Ce qui signifie que les possibilités d'ordonnement et de groupement des textes doivent avoir un effet sur la collection. Par conséquent, la collection résultant de la conjonction d'un CP et d'une autre collection sera ordonnée ou organisée.

Au regard de la définition de corpus de Sinclair (cf. définition 2.3² p. 68), les collections ainsi obtenues possèdent une propriété proche de celles des corpus : les données langagières (les textes) sont « organisées selon des critères pédagogiques explicites ». La seule différence entre la collection de texte et un corpus, de ce point de vue là, est donc le fait que les critères sont pédagogiques et non linguistiques. Cependant, de tels prismes ne nous donnent pas d'information sur la sélection des ressources.

2. « Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage. » [Sinclair, 1996, p. 4].

c) Collection homogène de textes

L'usage de prismes permettant de contraindre les valeurs des facettes, rapprochera la collection de la propriété d'homogénéité du composant :

« *A corpus can be divided into subcorpora. A subcorpus has all the properties of a corpus but happens to be part of a larger corpus. Corpora and subcorpora are divided into components. A component is not necessarily an adequate sample of a language and in that way it is distinct from a corpus and a subcorpus. It is a collection of a language that are selected and ordered according to a set of linguistic criteria that serve to characterise its linguistic homogeneity. Whereas a corpus may illustrate heterogeneity, and also a subcorpus to some extent, the component illustrates a particular type of language.*³ »

[Sinclair, 1996, p. 5]

Encore une fois, l'homogénéité telle que décrite par Sinclair est uniquement linguistique : tous les faits de langues de la collection vérifient les mêmes propriétés linguistiques. Dans notre cas les critères sont pédagogiques (pour l'enseignement des langues) et définis via l'expression du contexte pédagogique, par l'action de prismes et le calcul de facettes.

Une collection, obtenue à partir d'une autre par le biais d'un CP contraint, contiendra donc des textes « sélectionnés selon des critères pédagogiques explicites ». À nouveau, il s'agit d'une propriété directement adaptable à partir de la définition des corpus de Sinclair². Cependant, même adaptée à la pédagogie⁴, la collection n'a pas tous les attributs du corpus, puisqu'elle n'a pas de velléité de servir d'échantillon du langage ou quelque analogie appliquée à la problématique de l'enseignement des langues. Elle a pour but de permettre aux enseignants de considérer un ensemble de textes candidats parmi ceux qui répondent à leur question (formulation d'un CP).

Cela confirme l'appellation de « collection » et non de corpus (même pédagogique). Cependant, une collection à laquelle un contexte pédagogique aura été appliquée satisfera les conditions d'homogénéité définie pour les composants. La différence résidant dans le fait que l'homogénéité est pédagogique et non linguistique.

7.2 La recherche de texte du point de vue de l'utilisateur-enseignant

La fonction première d'une base de textes indexée pour l'enseignement des langues est de permettre la recherche. D'après les entités que nous avons définies et les informations obtenues sur le processus, nous pouvons définir les modalités d'interaction entre un tel système et un utilisateur-enseignant.

La figure 7.1 page ci-contre est une adaptation libre des diagrammes de séquences *Unified Modeling Language* (UML), qui visent à valider un scénario d'usage et explorer un peu plus la

3. Un corpus peut être divisé en sous-corpus. Un sous-corpus a toutes les propriétés d'un corpus mais se trouve faire partie d'un corpus plus grand. Corpus et sous-corpus sont divisés en composants. Un composant n'est pas nécessairement un échantillon adéquat d'une langue et de ce fait se distingue d'un corpus ou d'un sous-corpus. C'est une collection de faits de langue qui sont choisis et ordonnés en fonction d'un critère linguistique qui caractérise son homogénéité linguistique. Alors qu'un corpus peut illustrer des formes de l'hétérogénéité - de même qu'un sous-corpus peut le faire jusqu'à un certain point - le composant illustre un type particulier de langue.

4. Dans notre cas la sélection / l'ordonnement des textes ne dépend pas nécessairement de critères explicitement linguistiques (ex : auteur, date d'écriture du texte, source).

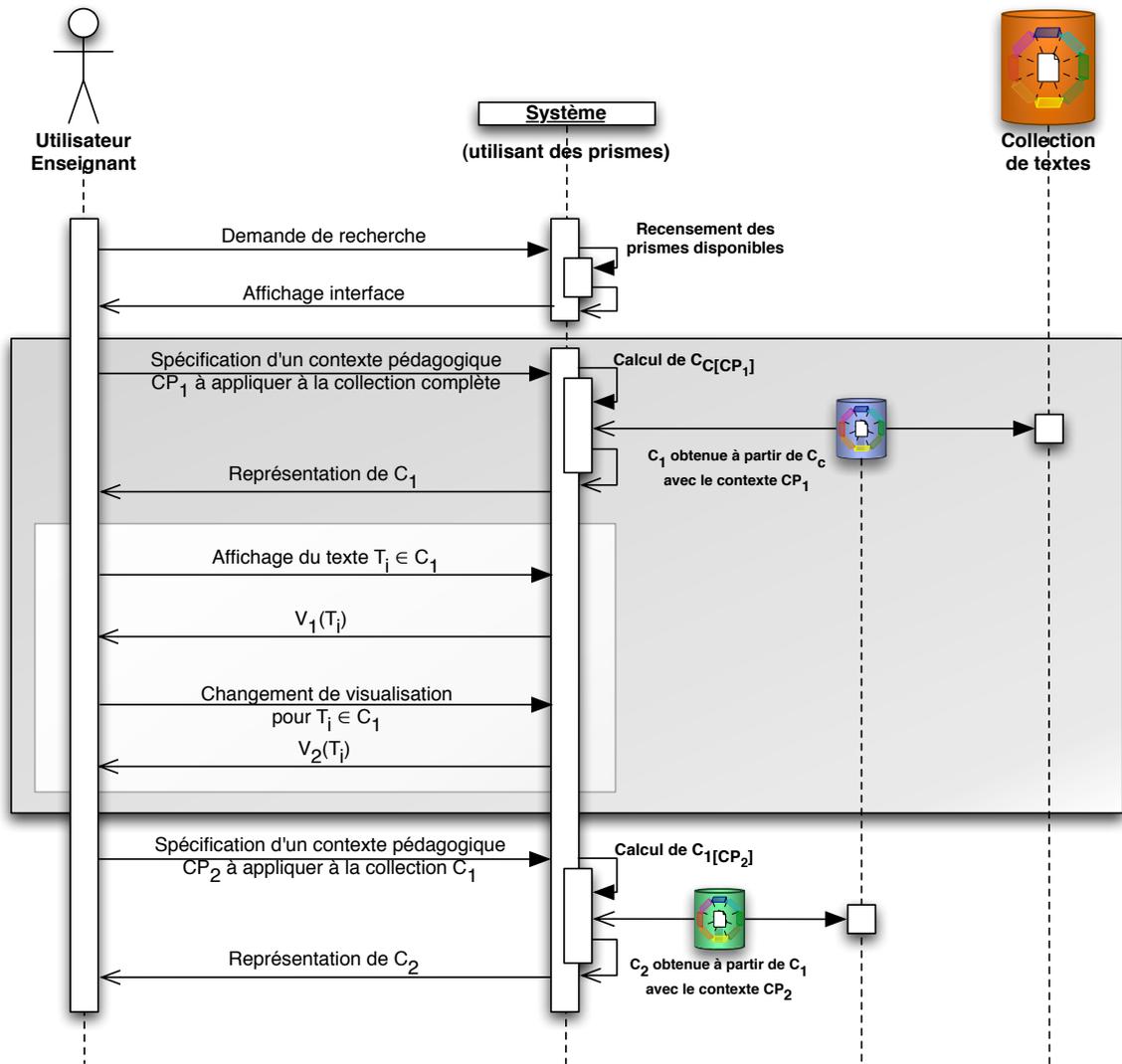


FIGURE 7.1: Exemple de séquence de recherche de textes

conception d'un système, que nous détaillerons ultérieurement [Ambler, 2005, p. 80]. Nous y faisons intervenir l'utilisateur-enseignant dans son interaction avec le système. La troisième entité présente est la collection de textes globale que la recherche de texte vise à restreindre afin de permettre à l'enseignant de choisir un texte parmi une sous-collection de textes candidats. Le système tel qu'il est décrit ici est considéré comme une entité globale permettant l'accès à des collections de textes. Nous ne faisons pas le détail de ses différents modules et de la manière dont ils communiquent.

La séquence du diagramme se décompose en trois parties : l'accès à l'interface de recherche, la spécification d'un contexte pédagogique par l'utilisateur (cadre grisé dans figure 7.1), puis l'affinement de ce CP.

7.2.1 Affichage de l'interface

La première action de la séquence consiste à la demande par l'enseignant d'effectuer une recherche (si le système ne permet que la recherche, cette action peut être apparentée à l'initialisation du système). Le système, pour proposer une interface de requête, doit recenser⁵ les prismes disponibles. Ces derniers sont ici considérés selon leur fonction expressive, que nous avons rapprochée de leur fonction métaphorique de décomposition du CP (cf. § c) p. 263 et § a) p. 276). Selon donc les facettes prises en compte et les données nécessaires à leur calcul pour chaque texte, une interface de requête est proposée à l'utilisateur.

7.2.2 Recherche de textes

a) Obtention d'une sous-collection organisée/ordonnée

L'enseignant utilise donc l'interface pour définir sa recherche. Il compose un contexte pédagogique en spécifiant les données d'entrée contextuelles d'un ensemble de prismes (cf. § b) p. 262)⁶. Le système applique le CP ainsi spécifié à la collection complète des textes du système C_C (toujours par l'entremise des prismes) et génère ainsi une collection homogène C_1 qui selon les prismes employés est ordonnée, organisée ou un sous-ensemble des textes de C_C (ou une combinaison des trois). La phase de sélection s'achève avec la présentation à l'utilisateur des textes candidats, augmentés des informations calculées (cf. ordonnancement). La phase d'évaluation peut alors commencer (cadre clair dans la figure 7.1).

b) Évaluation des textes candidats

L'enseignant peut demander l'affichage de chaque texte T_i de la collection obtenue (C_1), afin de l'évaluer. Pour l'aider dans cette entreprise il peut demander de se voir présenter plusieurs visualisations différentes de chaque texte. À la fin de l'évaluation de chaque texte, il lui associe la propriété d'être adéquat pour son problème ou non. Mais il ne le fait pas nécessairement dans le cadre du système (phase de transformation dans la figure 4.7 p. 206).

7.2.3 Précision du CP

La recherche peut s'avérer infructueuse pour plusieurs raisons. La collection obtenue ne peut contenir aucun texte correspondant réellement à ce que l'enseignant recherche, auquel cas il peut redéfinir le CP en reprenant la séquence à la spécification d'un contexte pédagogique (reformulation). Mais il est également possible que le CP qu'il a défini ne soit pas suffisamment discriminant, auquel cas la création de sous-collection peut s'avérer utile. Au lieu de recommencer le traitement à la première étape, l'enseignant pourra décider de raffiner le contexte pédagogique qu'il a défini et donc de créer une collection C_2 à partir de la collection C_1 , d'où sa présentation comme entité du système dans la figure 7.1.

Après cette nouvelle sélection, la phase d'évaluation s'effectue sur les textes de C_2 de la même manière qu'elle avait lieu pour C_1 . Le processus est itératif et peut continuer en générant C_3 à

5. Ou juste invoquer, selon que l'implémentation est complètement générique ou non.

6. Procédant par là à leur sélection.

partir de C_2 .

7.3 Indexation et propriétés sous-jacentes

Un système d'indexation pédagogique pour l'enseignement des langues doit, pour proposer la séquence présentée précédemment, d'ores et déjà disposer de textes codés de telle manière qu'ils soient utilisables par les prismes. Nous allons ici nous intéresser à comment ces derniers peuvent être utilisés pour permettre la sélection des textes sous forme de collections homogènes. Nous avons déjà expliqué les concepts manipulés⁷, nous allons les considérer ici sous l'angle des traitements qu'ils contiennent, dans le cadre de l'implémentation que nous décrivons ici.

7.3.1 Séparation des données et des traitements

Du fait de l'expérience du projet Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO), il nous semble primordial de séparer les données des traitements notamment pour permettre une meilleure évolutivité d'un système, mais aussi pour éviter les écueils qui touchent la plupart des systèmes d'Apprentissage des Langues Assisté par Ordinateur (ALAO) (cf. § 2.2 p. 47). Dans MIRTO cette séparation des données et des traitements, couplée à une hiérarchisation des entités permet d'attribuer à chaque niveau un rôle (cf. figure 7.2 de la présente page).

a) Le modèle MIRTO

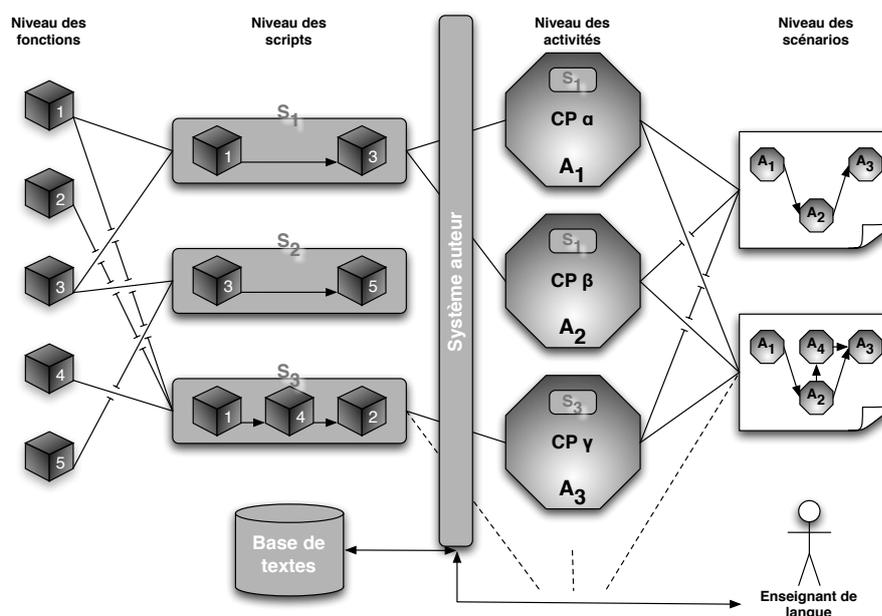


FIGURE 7.2: Modèle hiérarchisé de la plate-forme MIRTO [Antoniadis *et al.*, 2005b, p. 68]

7. Voir. § 6.2.1 p. 256, § b) p. 262 et § 6.3.2 p. 264.

Dans MIRTO, les « fonctions » représentent le niveau le plus bas du modèle ; ce sont des fonctionnalités basiques (Traitement Automatique des Langues (TAL), le plus souvent) qui ont été ou non créés pour la plate-forme. Dans un outil TAL complexe (multi-fonctionnalités), on récupère, si possible⁸, les différentes fonctions qui le composent. Elles ne sont pas accessibles aux utilisateurs-enseignants « du fait de leur nature technique et de leur indépendance d’une application didactique⁹ » [Antoniadis *et al.*, 2004a]. Ces fonctions sont compilées par une équipe composée d’informaticiens et de didacticiens au niveau des scripts. Les scripts permettent à la donnée d’un texte tout venant et d’une série de paramètres constituant les variables qui correspondent à ce que nous avons appelé le contexte pédagogique, de générer une activité. C’est l’utilisateur-enseignant de langue qui, via le système auteur de la plate-forme, fournit le texte et entre les paramètres du CP. Pour choisir ces paramètres, il n’est pas nécessaire d’avoir des connaissances informatiques. Ces choix relèvent de la didactique des langues. L’utilisateur-enseignant n’intervient donc qu’après le niveau des scripts. Il lui est ensuite possible d’organiser ces activités en scénarios.

Cette hiérarchisation permet non seulement d’encapsuler les traitements, mais aussi de les rendre transparents aux utilisateurs qui ne sont confrontés qu’aux niveaux qui leurs sont destinés.

7.3.2 Fonctions et pré-traitements

Sans entrer dans le détail de l’architecture globale du système, le principe de séparation des traitements et des données est inhérent à sa problématique, comme en atteste le schéma d’un modèle de recherche d’information tel que décrit par Géry (cf. § 1.3.2 p. 12).

Dans la figure 1.2 (p. 13), le Système de Recherche d’Information (SRI) (traitements) fait la jonction entre les documents (données) et l’utilisateur. Les composants d’un SRI ont deux fonctions de base : l’indexation et l’interprétation. Dans la figure 7.1 (p. 279), nous nous focalisons sur la fonctionnalité d’interprétation. Ces deux fonctions sont indissociables, comme l’indique la figure 1.2, la pertinence du système dépend directement de la cohérence entre interprétation et indexation. Selon la définition que nous avons donné des prismes, ceux-ci ont également un rôle à jouer au niveau de l’indexation, puisqu’ils sont censés proposer un traitement pour calculer les valeurs des facettes qui nous permettent d’accéder aux textes (cf. définition 6.1 p. 256). Ce calcul passe par ce que nous avons appelé les données sous-jacentes.

a) « Traitement » du prisme et fonctions

Malgré l’introduction de notions destinées à constituer une solution à notre problème, malgré la description des interactions de l’utilisateur-enseignant et du système dans le cadre de la recherche, nous n’avons pas encore défini ce qui constituerait pour nous une architecture adaptée d’un système d’indexation pédagogique de textes pour l’enseignement des langues. Nous allons donc entrer dans le détail d’une architecture possible, ce qui va nous permettre de préciser les concepts introduits et leur mode d’action.

Selon la même démarche que MIRTO, nous concevons un modèle de système de recherche de textes indépendamment des outils qu’il utilise. Nous avons évoqué cette réutilisation d’outils notamment TAL (mais pas uniquement) pour accéder aux concepts contenus dans les textes et

8. Par exemple par le biais d’une *Application Programming Interface* (API), si l’outil en propose une.

9. « *considering its technical nature and its independence from a didactic application.* »

leur associer une sémantique s'approchant de la problématique des enseignants¹⁰. Nous reprenons donc le concept de fonctions tel qu'il est employé dans MIRTO et l'adaptions à la conception de prismes.

Les prismes, rappelons-le, doivent proposer un traitement pour le calcul des facettes. Un prisme sera élaboré à partir de fonctions. Par exemple, $P_{NbEtRep}$, s'appuie sur une « fonction analyseur-morphologique ». Bien sûr, cette fonction à elle seule ne permet pas de modéliser tout le traitement nécessaire à l'obtention de la valeur d'une facette. Il sera nécessaire d'adapter le traitement par la conception de fonctions *ad hoc* ou la prise en charge d'autres outils (cf. § 6.2.3 p. 259).

Nous proposons de représenter les traitements associés aux prismes selon la figure 7.3 de la présente page, dans laquelle nous décrivons deux prismes P_A et P_B . Ces prismes sont mis en relations avec un ensemble de fonctions. Chaque élément Fn est une fonction au sens de MIRTO. Les flèches entre les fonctions représentent des flux de données.

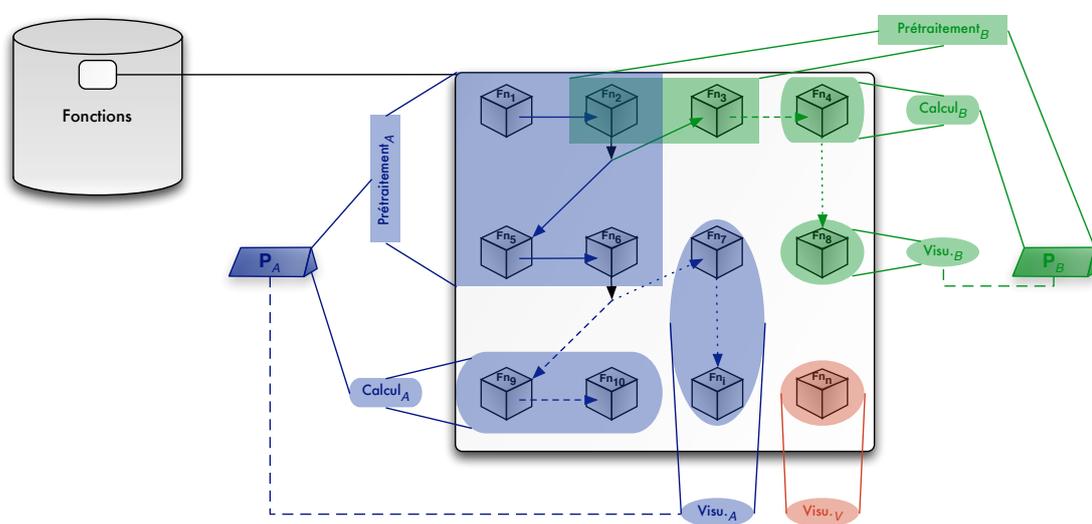


FIGURE 7.3: Prismes et fonctions

Avant d'aller plus loin dans la description de la figure, nous pouvons déjà remarquer que les données de sortie d'une fonction peuvent être réutilisées par plusieurs fonctions différentes. C'est le cas notamment pour les données issues de Fn_2 et Fn_6 . Ce qui ne fait que souligner un problème informatique évident : comme les fonctions ne sont pas nécessairement développées pour le système ou pour fonctionner ensemble, les sorties d'une fonction peuvent ne pas correspondre aux données attendues en entrée pour la fonction suivante dans le traitement d'un prisme. Afin qu'elle puisse prendre en charge les données qu'elle devra traiter ou pour rendre possible l'exploitation des données qu'elle fournira à d'autres fonctions, l'intégration d'une fonction particulière nécessitera parfois la création d'une moulINETTE *ad hoc*. Du fait des formats et des modèles linguistiques sous-jacents aux fonctions, la réutilisation des données d'une fonction par une autre peut soit entraîner la création de programmes complexes, soit s'avérer impossible¹¹.

Dans notre exemple fictif, en suivant les flèches des séquences de chaque prisme, nous obtenons :

- P_A : $Fn_1 \rightarrow Fn_2 \rightarrow Fn_5 \rightarrow Fn_6 \rightarrow Fn_9 \rightarrow Fn_{10}$;

10. Voir par exemple § c) p. 55 et § d) p. 57.

11. si le développement de la « moulINETTE » prend autant de temps que de redévelopper un outil, nous considérerons ce problème insoluble

- $P_B : \boxed{Fn_2 \rightarrow Fn_3} \rightarrow \boxed{Fn_4}$.

b) Réutilisation des fonctions

La figure 7.3 montre aussi que plusieurs prismes peuvent faire appel à la même fonction : F_A et F_B font toutes deux appel à la même fonction Fn_2 . Un *tokenizer* est un exemple d'une telle fonction. Il pourra être utilisé dans P_{nbmots} , ses données de sortie seront envoyées à un compteur qui déterminera le nombre de mots à partir de la représentation du texte sous forme de *tokens*. La tokenisation est en outre un traitement préliminaire de l'analyse morphologique. Dans l'éventualité où l'analyseur morphologique ne l'intègre pas, cette fonction pourrait également être incluse dans le traitement associé à P_{E^tRep} .

Cela suggère non seulement que les fonctions Fn_3 et Fn_5 acceptent au moins un type de données commun en entrée, mais aussi que la séquence à laquelle appartient une fonction n'en est pas une donnée. Le prisme joue donc un rôle de regroupement et de coordination des fonctions.

c) Pré-traitement

Nous avons séparé chacune de ces séquences en deux regroupements. Cette séparation nous permet de rendre compte du fait que le traitement s'effectue en deux phases : le stockage de données sous-jacentes non-contextuelles (flèches pleines) et le calcul de la valeur pour un contexte pédagogique donné (flèches « tirets »).

Dans le souci de ne pas multiplier les traitements, mieux vaudra effectuer ceux qui mènent à des informations non contextuelles une fois pour toutes à l'ajout de chaque texte¹² (pré-traitements). Au contraire, les calculs des valeurs qui nécessitent l'injection d'un CP, ne seront exécutés que lorsque c'est utile, en fonction des interactions avec l'utilisateur-enseignant. Cette manière de procéder reprend nos conclusions sur les différences entre métadonnées et facettes, selon lesquelles les principales limites des premières résident dans le fait qu'elles doivent être calculées / assignées *a priori*.

Cette distinction entre pré-traitement et calcul de valeur permet de spécifier les portions du traitement qui doivent être informatisables pour être intégrés à un système. Tout pré-traitement pourra relever d'un travail d'annotation manuelle, alors que la partie calcul ne peut relever de la responsabilité d'un agent humain¹³.

d) Visualisation

Enfin, la figure contient des visualisations qui sont elles aussi associées à des séquences de traitements. Elles bénéficient également de la séparation entre les traitements utilisés pour l'ajout des informations non contextuelles et ceux qui réutilisent ces données pour calculer les valeurs des facettes. Les fonctions associées aux visualisations peuvent ainsi s'appuyer sur les valeurs calculées pour les facettes (cf. Fn_8 dans figure 7.3), mais aussi traiter directement les données sous-jacentes ajoutées (cf. séquence $Fn_7 \rightarrow Fn_i$ dans figure 7.3). Dans un dernier cas de figure,

12. Si un nouveau prisme est ajouté au système, pour qu'il soit utilisable, il faudra exécuter son pré-traitement sur la totalité des textes recensés.

13. Le pré-traitement peut également être automatisé ou non, alors que le calcul est exclusivement un traitement informatique.

une visualisation peut n'être associée à aucun prisme et concerner un affichage d'ordre tout à fait général (cf. § 6.5.4 p. 272). C'est le cas de Fn_n ¹⁴

Le cas de figure de la séquence $Fn_7 \rightarrow Fn_i$ correspond à ce dont nous parlions en 6.5.5 (p. 272). Tout d'abord, elle indique que les visualisations peuvent également être composées de plusieurs outils. Ensuite, il s'agit d'une visualisation qui donne accès à un niveau d'information neutralisé par l'effet globalisant de la facette. À donner des propriétés *sur* le texte, certaines informations qualitatives ne peuvent pas être transmises aux utilisateurs par le biais des facettes (cf. § 6.5.2 p. 270)¹⁵. Le lien entre prisme et visualisation correspondante peut donc s'appuyer sur les pré-traitements.

7.3.3 Prisme et SRI

Nous avons pu définir la notion d'indexation en combinant la définition issue de [AFNOR, 1993, p. 512] (cf. définition 1.1 p. 10), qui s'appuyait sur la notion de langage documentaire de Lefèvre [2000] (cf. définition 1.3 p. 12). Nous avons remarqué en confrontant la définition de Lefèvre avec l'introduction à la Recherche d'Information (RI) de [Géry, 2002], qu'il regroupait au sein de la notion de langage documentaire aussi bien la notion de modèle de document que celle de modèle de requête (cf. § 1.3.2 p. 12). Du point de vue de l'indexation, c'est le modèle de document qui prime, puisqu'il s'agit de « transcrire en langage documentaire les concepts après les avoir extraits du document par une analyse » [AFNOR, 1993, p. 512]. Le pré-traitement correspond à cette partie du travail, puisqu'il ajoute des informations concernant les concepts des textes, afin que ceux-ci soit réutilisés dans le calcul de la valeur d'une facette servant par le biais de contraintes à la sélection des textes qui « répondent aux questions des enseignants ».

7.3.4 Ajout d'un texte au système

De notre point de vue l'indexation s'apparente donc à l'ajout des propriétés sous-jacentes non contextuelles nécessaires au fonctionnement des prismes. Elle peut donc prendre deux formes (cf. figure 7.4 page suivante) :

- soit des outils informatiques peuvent effectuer en autonomie l'intégralité du pré-traitement ;
- soit une intervention humaine est nécessaire à une étape du pré-traitement (cf. § c) p. 284).

Dans le cadre d'une séquence d'utilisation, un utilisateur voulant ajouter un texte à un système suivant l'approche développée ici commencera par indiquer au système son intention. Le système en réponse devra recenser les propriétés sous-jacentes qui doivent être ajoutées manuellement et générer une interface permettant à l'utilisateur de saisir les informations désirées. Son échange avec le système sera donc composé d'un texte à ajouter et d'un ensemble d'informations parmi lesquelles certaines seront utilisables telles quelles, des valeurs de facettes « constantes », comme par exemple l'auteur du texte ou l'année de publication (cf. figure 6.2 p. 258 ou § 6.4.1 p. 266). Dans la figure 7.4, il s'agit du triangle bleu : par convention, la facette en tant que propriété donne sa forme à la boîte, alors que la projection du contexte pédagogique à travers le prisme lui donne sa couleur (cf. figure 6.4 p. 260 / § 6.2.3 p. 259). L'utilisateur peut aussi être amené à indiquer des informations à partir desquelles le système effectuera des calculs pour obtenir une valeur de facette en fonction du contexte pédagogique (comme par exemple l'auteur du texte

14. Par exemple, l'affichage des 5 premières lignes du texte.

15. Le nombre d'éléments représentatifs de la notion « prétérit » n'indique pas de quels verbes il s'agit ou le contexte dans lequel ils sont employés.

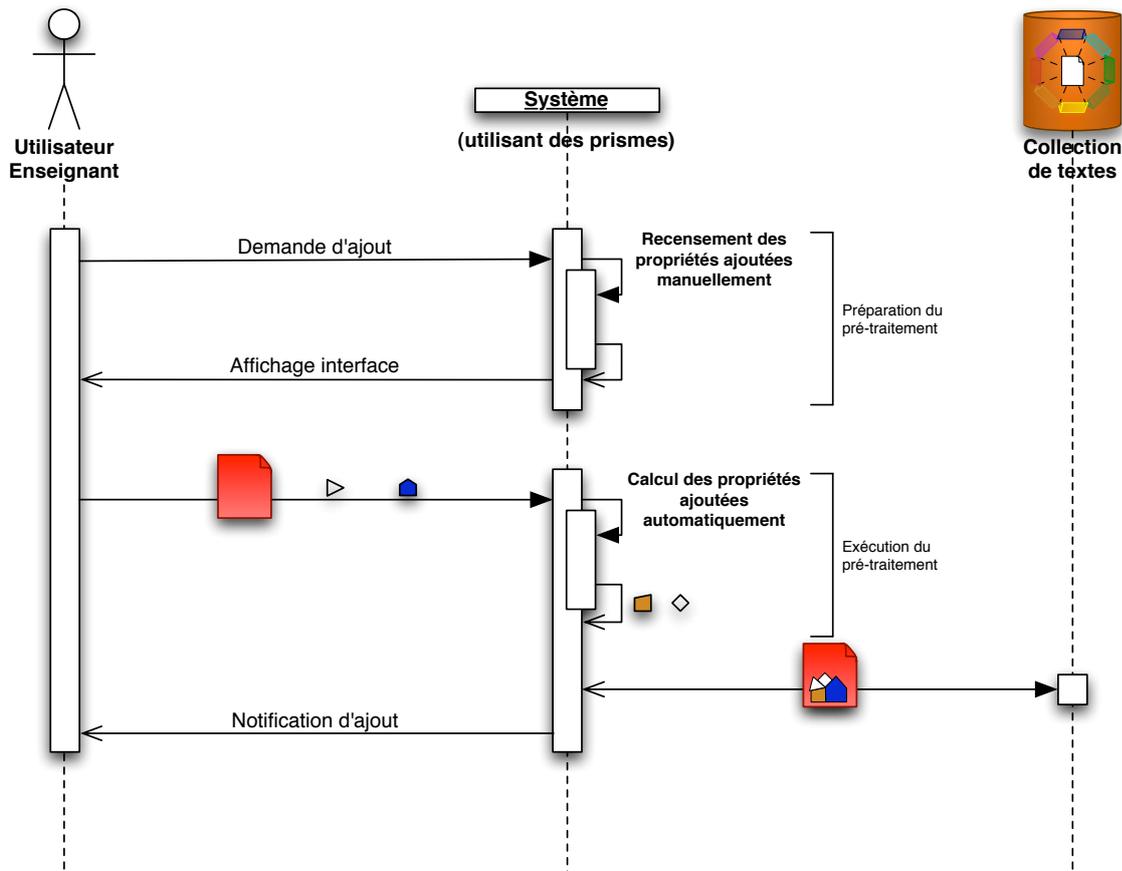


FIGURE 7.4: Séquence d'ajout d'un texte au système

ou son titre¹⁶). Dans notre schéma, cela se traduit par une forme non-colorée (informations fournissant des concepts pour le calcul par le prisme d'une valeur en fonction d'un CP). Enfin, le système effectue la fin des pré-traitements pour ajouter lui aussi des informations qui peuvent constituer une valeur de facette constante (cf. F_{nbMots}), représentée en orange dans la figure 7.4 ou bien des informations sous-jacentes non contextuelles (cf. analyse morphologique dans F_{etRep}) représentées sous forme de losange blanc dans la figure.

7.3.5 Trois types de scripts

À travers la description de l'ajout d'un texte à un système d'indexation pédagogique de textes, nous avons pu décrire l'encapsulation des fonctions sous forme de 3 types de traitements, que nous appellerons scripts (cf. § 7.4.1 p. 287). Nous séparons deux formes de séquences de ces briques élémentaires comme composantes du traitement du prisme. La première, le pré-traitement concerne en premier lieu la tâche d'indexation, le stockage des textes et de leur description sous une forme que les prismes pourront ensuite réutiliser pour calculer la valeur de la facette qui leur est associée. Alors que le pré-traitement concerne exclusivement des propriétés non contextuelles

16. L'annotation des éléments de discours rapporté selon le schéma du projet Plan Pluri-Formation (PPF) (cf. § a) p. 269) pourrait relever de ce type d'information, s'il était proposé une interface d'annotation du texte à l'utilisateur. Cependant, la procédure d'ajout d'un texte en deviendrait très fastidieuse.

du texte ou de son contenu, le calcul analyse ces données à la lumière d'un contexte pédagogique pour permettre à l'enseignant de rechercher des textes (sélection). Enfin les visualisations, le dernier type de traitement, sont destinées à permettre à l'utilisateur-enseignant d'accéder aux textes, dans le cadre de la phase d'évaluation. Celles-ci peuvent notamment lui donner un point de vue plus qualitatif des phénomènes concernés par sa recherche.

7.4 Architecture d'un système d'indexation pédagogique de textes

Avant d'entrer dans le détail des objets nécessaires à une implémentation d'un tel système¹⁷, il nous paraît pertinent de proposer ici une synthèse visuelle des informations proposées jusqu'ici. Nous proposons donc dans cette partie bilan un diagramme reprenant la majorité des éléments développés dans ce chapitre.

Nous avons exposé les séquences d'utilisation d'un système d'indexation pédagogique de textes à base de prismes. Cet examen des interactions avec les utilisateurs nous a mené à nous approcher de l'architecture même d'un système à travers la combinaison des fonctions disponibles. En les combinant nous arrivions à trois types de traitements dont la combinaison assurait l'obtention des valeurs des facettes et l'indexation des textes grâce aux prismes ainsi constitués et la présentation des textes aux utilisateurs sous des formes favorables à l'évaluation.

La figure 7.5 page suivante propose une architecture globale cohérente avec les préceptes énoncés. Pour ne pas la surcharger plus, nous ne tentons pas d'y intégrer les informations de séquences déjà exposées dans les figures 7.1 (p. 279) et 7.4 (p. 286).

7.4.1 Niveau des scripts

Alors que nous parlions de « traitement » pour parler des prismes, nous avons opté pour l'appellation de script dans le cadre de la figure 7.5. Cette décision terminologique nous permet de séparer les couches de l'architecture du système.

a) Réutilisation de la terminologie MIRTO

Tout d'abord, cette appellation inscrit cette architecture dans l'approche présentée dans MIRTO. En effet, nous séparons ici la conception de prisme de l'architecture du système. Nous avons d'ailleurs évoqué la réutilisation de fonctions déjà existantes pour créer de nouveaux points de vues sur les textes. Nous reprenons par là-même les éléments constitutifs de MIRTO, au niveau le plus bas de modélisation didactique : il s'agit d'outils de bas niveau (monofonctionnels) qui sont ensuite combinés de manière incrémentale pour parvenir à la création d'outils pédagogiques. Nous procédons de manière analogue et à ce titre l'appellation de « script » est cohérente.

Nous distinguons donc trois types de scripts (cf. § 7.3.2 p. 282), tous composés à partir de fonctions (cf. traits pleins dans la figure 7.5). Les scripts pré-traitements qui permettent l'indexation sont considérés comme la première partie du traitement des prismes. Les données qu'ils produisent doivent pouvoir être interprétées par les scripts de « calcul » de la valeur de la facette

17. La modélisation UML d'un système sous forme de diagramme de classe peut s'avérer fastidieuse d'accès pour le lecteur non-informaticien. Et ce d'autant plus que nous n'allons pas replacer UML dans la perspective de la programmation objet ou des enjeux de cette dernière (cf. figure 7.6 p. 293).

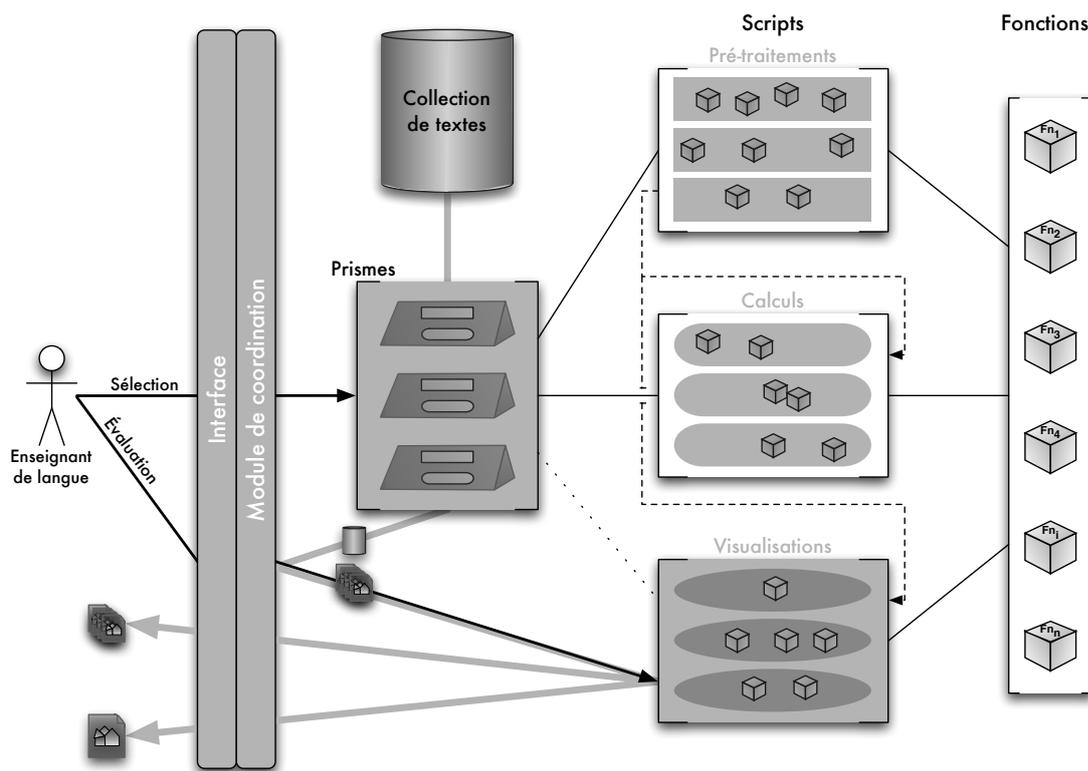


FIGURE 7.5: Architecture modulaire à base de prismes

(flèches avec tirets). Ces deux entités sont donc regroupées au sein du prisme (traits pleins). Enfin, les données issus des pré-traitements peuvent être utilisées dans le troisième type de script, les visualisations. Celles-ci peuvent, notamment quand elles partagent un pré-traitement avec, être liées à des prismes (cf. § d) p. 284). Nous représentons le lien qui unit prismes et visualisations différemment des autres. En effet, la visualisation n'est pas incluse dans le prisme. Elle peut exister indépendamment de ce dernier, qui lui-même ne nécessite pas de se voir associer une visualisation pour fonctionner. Toutefois, l'existence de liens entre prismes et visualisations peut permettre d'associer la phase de sélection avec celle d'évaluation.

b) Scripts et autres traitements

Ensuite, le terme « traitement » est satisfaisant lorsque nous décrivons un traitement en particulier. Mais intégrée à l'architecture globale du système, la notion est trop générale et nous ne pouvons parler *des* traitements en ne considérant que les prismes et les visualisations. En effet, le système est destiné à gérer ces entités en interaction avec un utilisateur. Il existera donc également des traitements exécutés au niveau du module de gestion du système et de l'interface (cf. figure 7.5). Pour ne pas confondre les traitements du prisme ou des visualisations et ceux de gestion du système, l'utilisation du terme « scripts » pour parler des premiers réduit les ambiguïtés potentielles.

Une différence importante entre les scripts et les autres traitements effectués par le système réside dans le fait que ce sont les scripts qui mènent à la modélisation des textes comme objets pédagogiques. Ils ne créent pas la valeur ajoutée pédagogique *per se*, mais permettent de s'en

approcher.

c) Différence conceptuelle avec MIRTO

La similarité entre notre niveau de script et celui de MIRTO est avant tout située au niveau structurel et sur le plan du cycle de développement. Mais d'un point de vue conceptuel, il existe une différence entre les deux. Dans MIRTO, le niveau des scripts est celui qui crée la valeur ajoutée pédagogique : « un script est une suite de fonctions liées par un objectif pédagogique » [Antoniadis *et al.*, 2005b, p. 69].

Dans notre cas, sans le lien avec la propriété à modéliser, les scripts perdent une partie de leur sémantique. Les pré-traitements prennent leur sens couplés avec le calcul d'une facette ou la visualisation d'un texte. Or, c'est au niveau du prisme que s'effectue cette mise en relation, c'est lui qui associe ces traitements à la propriété pédagogique qu'ils doivent permettre de modéliser. Même si c'est l'idée d'un prisme qui gouvernera leur création, pris indépendamment, les scripts ne sont pas porteurs de la valeur ajoutée pédagogique. Ils constituent des moyens de la créer (scripts de pré-traitement et de calcul) mais elle n'apparaîtra réellement qu'au niveau du prisme.

7.4.2 Coordination de prismes et de visualisations

Nous avons d'ores et déjà expliqué l'essentiel de la notion de prisme : sa sémantique et sa dépendance de la notion de facette (cf. § 6.2.1 p. 256), sa gestion du contexte pédagogique (cf. § 6.3 p. 263), sa fonction expressive (cf. § c) p. 263), son lien avec certaines visualisations (cf. § 6.5.5 p. 272) et la notion de collection pour la création d'une sous-collection homogène (cf. § 7.1.1 p. 276). Nous l'avons également abordée d'un point de vue structurel (cf. figure 7.3 p. 283 et § c) p. 289). Nous n'allons donc pas nous appesantir plus sur la notion elle-même.

Dans la figure 7.5 (p. 288), nous ajoutons toutefois quelques éléments nouveaux qui tiennent à l'articulation entre prismes, collections de textes et visualisations au sein d'un système.

a) Modules

Tout d'abord, nous identifions les principales entités/rôles du système (en gris dans la figure 7.5) :

- un module regroupant les différents prismes ;
- un module de gestion des visualisations ;
- un module de gestion des interfaces ;
- un module de coordination.

Nous pouvons ajouter à ces modules des éléments structurels : les scripts permettant de constituer les prismes ainsi que les fonctions permettant de créer les scripts¹⁸. Ne manquent plus alors que les données sur lesquelles l'ensemble agit : une collection de textes.

Les deux éléments que nous n'avions pas mentionnés précédemment sont donc les modules de gestion des interfaces et de coordination. Ils ont été rapprochés dans le diagramme en raison du fait qu'il est difficile de les dissocier complètement, la coordination des entités du système dépend des actions de l'utilisateur (recueillies par l'interface). Le choix des éléments à présenter

18. La portion correspondante de la figure 7.5 (architecture globale) intègre finalement les informations représentées dans la figure 7.3 (fonctions ; p. 283) à un schéma plus global.

à l'utilisateur sont directement issus du module de coordination. Plutôt que d'entrer dans des détails d'implémentation qui relèvent de choix des développeurs, nous avons joint les deux et nous y référerons en tant que « gestionnaire de recherche », pour nous abstraire de ce dilemme peu pertinent en dehors du codage d'un système.

b) Architecture générale

Outre l'ajout d'un texte que nous n'indiquons pas dans la figure 7.5, l'utilisateur interagit avec le système pour demander au système d'opérer une tâche de sélection ou pour évaluer les résultats de cette opération. L'interface lui permettra donc d'agir sur – proposition d'un CP – et en fonction – selon les contraintes – des prismes pour définir un contexte pédagogique à appliquer à la collection de textes. Pour utiliser les éléments d'un modèle de recherche d'information de la figure 1.2 (p. 13), il s'agirait de la formulation d'une requête. Celle-ci est ensuite interprétée par le prisme de manière à la faire correspondre avec le mode de description des textes (modèle de document dans la figure 1.2).

Cependant le prisme permet de calculer une valeur pour un texte, mais n'est pas un outil d'interrogation de base de données. Il faudra donc pouvoir réutiliser ce calcul dans une requête. La fonction de correspondance est donc prise en charge d'une part par le prisme (prise en charge du contexte pédagogique de manière à pouvoir le rendre calculable selon les informations présentes dans les textes), et d'autre part, c'est le « gestionnaire de recherche » qui inclue ces informations sous une forme applicable à un ensemble de textes. Pour formuler cela en « pidgin-SQL », si nous représentons le calcul de la valeur d'une facette par sa représentation dans la figure 7.5, la génération d'une requête pourrait être simplifiée de la manière suivante :

```
SELECT textes, F[CP] FROM collection WHERE F[CP](Txt)
```

Le script de calcul sera donc formulé en fonction des choix d'implémentation du système (base de données, expressions régulières Perl). Quoi qu'il en soit, ces requêtes, une fois interprétables selon le formalisme utilisé pour indexer les documents, seront appliquées à une collection de textes pour permettre d'obtenir une sous-collection homogène (début du trajet issu de la collection de texte dans la figure 7.5 p. 288).

Une collection homogène contient donc un ensemble de textes assortis d'informations supplémentaires (par exemple un ordonnancement). Cependant, pour les présenter à l'utilisateur les textes doivent être transmis un par un aux visualisations (premier « coude » au parcours des données dans la figure 7.5). Bien que des visualisations par défaut soient à prévoir, l'utilisateur spécifie les visualisations qui l'intéressent dans le cadre de l'évaluation. Un tel système devrait pouvoir permettre la visualisation de plusieurs textes simultanément (pages de résultats) ou d'un seul texte à l'écran (analyse du document). À nouveau, c'est le gestionnaire de recherche qui permet à l'utilisateur enseignant de spécifier la visualisation qui l'intéresse et de sélectionner les visualisations à faire intervenir en conséquence.

Nous avons donné une idée des séquences d'interactions possibles dans la figure 7.1 (p. 279). Celle que nous décrivons ici se contente de pointer les actions de l'utilisateur, des modules par lesquels elles transitent et ceux sur lesquelles elles agissent. La figure ne prétend pas représenter un scénario potentiel de recherche de textes.

c) Gestionnaire de recherche

L'entité que nous avons appelée pour les besoins de la description de la figure 7.1 « gestionnaire de recherche » correspond à la partie du système qui va intégrer facettes, prismes et visualisations au processus de recherche. C'est également la partie que nous avons la moins décrite. Il faut toutefois rappeler que le premier objectif de ce travail est de proposer un modèle utilisable pour l'indexation pédagogique de textes pour l'enseignement des langues. Notre modèle permet, indépendamment de la manière dont il est implémenté, de concevoir différemment les propriétés des textes en prenant en compte le CP. Le second objectif est de proposer une maquette qui permette d'attester qu'une implémentation du modèle est possible. Ce sera l'objet de la fin de ce chapitre.

7.5 Conception pour une implémentation informatique du modèle

Le modèle, tel que nous l'avons décrit semble propice à une approche orientée objet. La décomposition des traitements selon différents niveaux, empruntée à MIRTO ainsi que les notions de prisme, facette et collection peuvent chacun constituer une classe d'objet. De plus, la programmation objet permet une forte encapsulation des traitements, ce qui favorise la modularité du système (déjà mise en exergue dans la figure 7.3 p. 283) et son évolutivité. L'objectif de notre travail n'était pas de proposer de nombreuses facettes. Même si notre objectif avait été la création d'un produit « fini », nos données ne peuvent nous permettre que de mettre en place des prismes relativement basiques. Un tel système se devrait de pouvoir intégrer de nouveaux prismes, si ceux-ci sont proposés. Les spécifications ci-dessous sont proposées dans l'optique d'offrir un modèle de données suffisamment générique pour que l'ajout de prismes soit possible dans le cadre du système, sans en redéfinir un nouveau.

Nous allons donc nous appuyer sur un diagramme de classe UML, la figure 7.6 page 293 (recherche) et la figure 7.7 page 297 (indexation) pour décrire les entités qui nous semblent nécessaires à la mise en œuvre d'un système d'indexation pédagogique de textes pour l'enseignement des langues suivant le modèle que nous avons défini. Nous expliquerons ensuite les aspects sur lesquels nous nous focaliserons pour l'implémentation de notre maquette avant de conclure sur le processus de création d'un prisme.

7.5.1 Point de départ : la facette

L'objectif d'une tâche d'implémentation est de concevoir un système qui permette de traiter des textes selon des facettes. Nous allons donc commencer par le package regroupant les objets destinés à en calculer la valeur pour des textes dans un contexte pédagogique donné.

Idéalement, dans l'optique de la création d'un système partant de la problématique des enseignants l'ajout d'un nouveau prisme doit partir des besoins de ces derniers, synthétisés à partir d'une facette¹⁹.

19. Il est tout à fait possible de partir d'un outil proposant de nouvelles informations sur les textes et de tenter de voir dans quelle mesure il peut être utilisable pour des enseignants. Mais même en utilisant cette démarche, l'outil sera le support d'un questionnement menant à conceptualiser les besoins des enseignants.

Nota Bene 7.1 Conventions de notation (programmation objet)

Quand nous nous référons à des entités de l'implémentation que nous proposons, nous l'indiquons en changeant la **police de caractères**. Nous suivons les conventions de la programmation objets :

- les noms de classe commencent par une majuscule ;
- les noms d'objets commencent par une minuscule ;
- les noms d'attributs et de méthodes commencent également par une minuscule ;

Par exemple, « un objet **exemple** » pourrait être glosé par « un objet que nous appellerons **exemple** », alors que « un objet **Exemple** » signifie « un objet de classe **Exemple** ».

Pour rendre la lecture moins difficile, nous tentons d'utiliser des noms d'objets qui renvoient directement à leur classe (et de ce fait l'objet **exemple** est très vraisemblablement de classe **Exemple**) et qui soient lisibles dans une phrase.

Cette volonté d'intégrer des éléments du diagrammes dans le texte nous mène parfois à des formulations douteuses du point de vue informatique du type : « Nos **exemples** sont-ils clairs? », qui signifierait « Nos objets, que nous appelons **exemples**, sont-ils clairs? ». Or, s'il s'agit d'objets, chacun d'entre eux devrait porter un nom différent. Quand nous avons recours à ce type d'artifice, visant à rendre le texte plus lisible, il faut bien réaliser qu'il ne s'agit pas d'un unique objet, mais de plusieurs.

Bien que la facette désigne une propriété du monde réel, nous avons décidé de l'inclure sous forme de classe. En effet, le prisme tire sa qualité de son association avec une facette. Exclure la facette du modèle informatisé nous semblait amputer le prisme d'une partie de son essence. En outre, la classe **Facette** tel que nous l'avons présentée est le support d'informations susceptibles d'aider les concepteurs dans les tâches de maintenance du système : le prisme est indissociable de la propriété à laquelle il permet d'accéder. En effet, les attributs d'un objet **Facette** en font une source de documentation en précisant (cf. figure 7.6 p. 293) :

- ce que représente la propriété (**definition**) ;
- l'ensemble des valeurs possibles ;
- l'interprétation de ses valeurs (**relationValeurs**).

En outre, une partie des informations fournies peuvent être proposées à l'utilisateur-enseignant sous forme d'aide du système. En effet, chaque objet **Prisme** « modélise » un seul objet **Facette** (et inversement) et les objets **Prisme** constitue la pierre angulaire du système par laquelle beaucoup d'éléments (dont la facette correspondante, donc) peuvent être accédés.

7.5.2 *Package* « Fonctions et scripts » et prisme

La propriété modélisée sert donc la création d'un objet **Prisme**²⁰. Nous proposons de regrouper la modélisation informatique des informations de la figure 7.3 (p. 283) au sein d'un *package* « Fonctions et scripts ».

a) La classe **Script**

L'entité centrale de ce *package* est la classe **Script** qui est définie comme la composition d'objets **Fonction**. Chaque **script** peut être composé de une à plusieurs **fonctions** et chaque **fonction** peut être liée à aucun **Script** ou plus. En effet, il est imaginable que le système contienne des

²⁰. Le prisme précisera en retour l'acception de la propriété par les limites inhérentes à la modélisation informatique (cf. § a) p. 14).

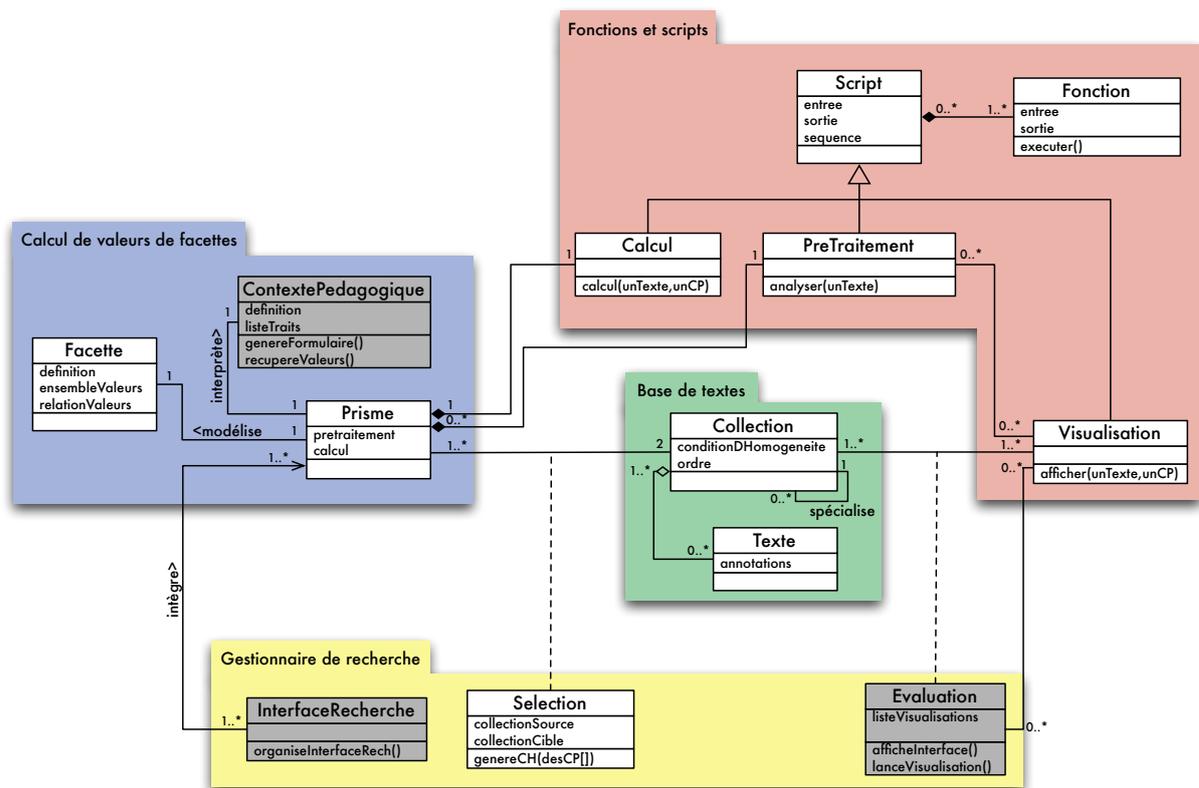


FIGURE 7.6: Diagramme de classe d'une implémentation (recherche de textes)

fonctions qui n'ont pas encore été intégrées au moindre script (cf. relation de composition entre les deux classes dans figure 7.6).

Aucune méthode n'est proposée pour la classe `Script`, il ne s'agit pas d'un choix définitif : des besoins peuvent surgir pendant l'implémentation du système. Cependant, il nous a paru important qu'aucune méthode d'exécution ne soit proposée. L'exécution sera gérée par les sous-classes.

b) Sous-classes de `Script`

Chacune des sous-classes de `Script` a un sens différent. Le traitement n'a pas la même vocation selon qu'il est destiné à proposer un affichage pour un texte, qu'il doit le préparer à un calcul (cf. indexation) ou effectuer le calcul en question. Nous avons donc transféré les méthodes d'exécution des scripts aux sous-classes. À l'exception des objets de classe `Fonction` qui les constituent, aucun objet du système n'a de relation avec un `script`. Le système ne contiendra donc que des instances des sous-classes de `Script`.

Un `calcul` pourra proposer une séquence de `fonctions` permettant d'obtenir une valeur pour un couple (`texte`, `CP`). Un `preTraitement` sera à même de fournir les informations pour l'indexation d'un texte (cf. § 7.5.5 p. 296). Enfin une `visualisation` permettra d'afficher un texte en fonction d'un contexte pédagogique donné.

c) Relations entre la classe Prisme et les sous-classes de Script

La mise en place d'une visualisation peut nécessiter l'ajout d'informations aux textes ou leur conversion d'un format à un autre. Chaque **visualisation** peut donc être liée à 0 ou plusieurs **preTraitements**. Inversement, un **preTraitement** ne donne pas nécessairement lieu à une **visualisation**, mais peut servir de support à plusieurs **visualisations** différentes (cf. figure 6.6 p. 271).

De la même manière un **preTraitement** peut n'être lié à aucun **prisme**, mais aussi à un ou plusieurs. Au contraire un **prisme** n'est composé que d'un et un seul **preTraitement**. Grâce à ces relations successives, à partir d'un **prisme**, il est possible de savoir quel **preTraitement** le compose et par transitivité, quelles visualisations peuvent lui être associées.

Enfin, un **prisme** est composé, d'un seul et unique objet de classe **Calcul** et réciproquement. Du fait de la définition de la notion de facette, il nous semble qu'un calcul ne doit être lié qu'à un seul **prisme**. En effet, un prisme doit permettre le calcul des valeurs d'une facette, qui définit les valeurs potentielles et les relations que ces dernières entretiennent. Si deux **prismes** peuvent recourir exactement aux mêmes objets de classe **Fonction** pour le calcul de la valeur, ce dernier porte en quelque sorte sa sémantique, ne serait-ce que par le processus qui a permis sa définition et deux instances de **Calcul** devront être définis.

7.5.3 La sélection de textes

Ayant défini les composants des objets **Prisme**, nous allons nous intéresser au rôle de cette classe dans la sélection des textes.

a) Interface de recherche

L'utilisateur confronté au système devra pouvoir spécifier un contexte pédagogique correspondant à son besoin. Pour ce faire le *package* « Gestionnaire de recherche » possède une classe **InterfaceRecherche**. Une telle interface recensera les prismes pris en charge par le système. Une **interfaceRecherche** intégrera donc un ou plusieurs **prismes**. À long terme, on peut imaginer qu'un système de recherche de textes selon des facettes classe les **prismes** en fonction de leur degré de difficulté de prise en main et propose une interface différente selon que l'enseignant est novice dans sa maîtrise de l'outil informatique et donc que le système dispose de plusieurs **interfacesRecherche**. Chaque prisme pourra être adapté pour différents degrés d'expertise, d'où la cardinalité de la relation.

Chaque **prisme** recensé par l'interface est capable d'interpréter un type de contexte pédagogique. Nous proposons donc l'usage d'une classe **ContextePedagogique** qui cumule plusieurs rôles :

- de la même manière que la classe **Facette** permettait d'inclure des informations destinées à un agent humain, portant sur sa sémantique, la classe **ContextePedagogique** propose un attribut **definition**;
- un **contextePedagogique** liste les traits que le prisme devra interpréter (cf. § 6.2.4 p. 261 et § 6.3.1 p. 264);
- la classe a des fonctionnalités d'interface²¹, elle permet donc de générer un formulaire permettant à l'utilisateur de formuler sa requête;

21. D'où le fait qu'elle soit grisée dans la figure 7.6 (p. 293).

- le rôle de cette interface est également de récupérer les valeurs entrée par l'utilisateur pour les fournir au **prisme** associé.

La fonctionnalité expressive du prisme (cf. § c) p. 263) est transcrite dans le système par le lien qui est fait entre le rôle d'interface des objets **ContextePedagogique** et une **interfaceRecherche** par un objet **Prisme**.

b) Les objets **Collection**

Avant de d'expliquer comment les **prismes** participent à la phase de sélection nous allons revenir sur la notion de collection de textes dans notre modèle et aux objets que nous lui associons.

Nous définissons un objet **Collection** comme un sous-ensemble homogène (cf. § c) p. 278) et éventuellement ordonné (cf. § b) p. 277) d'une collection de textes. Par homogène, nous entendons ici qu'ils ont comme caractéristique commune de répondre à un besoin donné d'un utilisateur. Ainsi, considérons l'ensemble des textes indexés par la base, ce sera un sous-ensemble de « l'univers des textes ». Les textes de ce sous-ensemble ont tous été considérés par l'utilisateur qui les a ajoutés à la base comme utilisables pour l'enseignement des langues et ils sont tous indexés par notre prototype. Ces caractéristiques satisfont la définition ci-dessus en ce sens que l'indexation par notre prototype répond (ou vise à répondre) aux besoins de ses utilisateurs, qui s'en servent pour rechercher parmi un ensemble de textes, des textes présentant certains traits. À partir de cet ensemble initial, tout sous-ensemble de textes issu d'une requête sera donc un objet **Collection**, sur lequel il est possible ensuite d'exécuter d'autres requêtes pour préciser ses besoins.

Cette propriété est directement appliquée à l'implémentation du modèle via la variable de classe **conditionDHomogeneite**. Cette condition d'homogénéité est la (ou l'ensemble de) propriété(s) se retrouvant dans chacun des textes de la collection considérée. D'un point de vue conceptuel, les valeurs de cette variable devraient être des ensembles d'instances de facettes²² associées chacune à un ensemble de valeurs acceptables.

Une sous-collection possèdera donc une **conditionDHomogeneite** plus restrictive que celle dont elle est issue, nous avons donc appelé la relation entre la **collection** d'origine et ses avatars potentiels **specialise**. On lira : « une **sousCollection specialise** une **collection** ».

D'un point de vue informatique, une implantation possible est d'inclure un Système de Gestion de Base de Données (SGBD) à la classe **collection**, auquel cas la **conditionDHomogeneite** pourra être une requête. La **conditionDHomogeneite** de la collection globale, qui contient tous les textes indexés dans la base²³ serait donc (en **mysql**) une requête du type :

```
SELECT * FROM 'collection' ; (qui permettrait de sélectionner tous les textes de la base).
```

Nous retrouvons donc la condition énoncée dans la définition : tous les textes de la base respectent cette **conditionDHomogeneite** et constituent ainsi la collection globale.

Mais il ne s'agit là que d'une manière d'implémenter cette condition d'homogénéité, qui est en définitive une liste de commandes permettant de générer une sous-collection. En effet, on peut imaginer une implémentation de la collection qui ferait intervenir une base de données pour

22. Au sens de la section 6.2.3 p. 259, pas de la programmation objet.

23. Voir la relation entre la classe **Texte** et la classe **Collection** dans la figure 7.6 (p. 293) qui rend compte des sous-collections et de la possibilité d'avoir des collections vides.

les facettes constantes et des documents *eXtensive Mark-up Language* (XML) parsés par un programme *P* répétant le calcul issu d'un prisme et qui renverrait les identifiants des textes dont la facette pour le CP ne serait pas nulle. La condition d'homogénéité pourrait récupérer des identifiants par une requête (SQL par exemple), en obtenir d'autres par l'exécution du programme *P*, avant d'opérer une intersection des deux ensembles.

c) La classe associative Selection

Nous avons d'une part des objets capables d'interpréter un `contextePedagogique` et de le repercuter dans le calcul de la valeur d'une facette et d'autre part des objets `Collection` capables de générer une sous-collection à la donnée d'une `conditionDHomogeneire`. Le *package* « gestionnaire de recherche » doit donc proposer des objets capables de faire le lien entre les deux. La nécessité de passer par ce module découle d'un problème que nous n'avons pas encore abordé : l'utilisation simultanée de plusieurs objets `prismes`. Si un seul `prisme` était utilisé dans chaque opération de sélection, la classe correspondante pourrait se charger de ce traitement, mais la relation ensembliste d'union nécessite l'intervention d'une classe supplémentaire²⁴. Le rôle de la classe `Selection` est de mettre en relation deux `collections` (source et cible) avec l'ensemble des prismes utilisés pour formuler la `conditionDHomogeneite` (cf. § b) p. 290).

7.5.4 Évaluation

Une fois une `collection` sélectionnée, il faudra permettre à l'utilisateur d'effectuer la tâche d'évaluation. Le système proposera donc une interface du choix des `visualisations` utilisables. L'interface s'appuiera sur une liste établie d'une part directement avec certaines `visualisations` (il s'agit des visualisations les plus génériques, indépendantes du CP), mais aussi en remontant la chaîne des objets en passant par `Collection` et `Prisme` (`visualisations` liées à un `prisme`). Une fois la `visualisation` désirée choisie par l'enseignant un objet `Evaluation` doit pouvoir l'invoquer pour un texte donné.

7.5.5 Indexation

Représenter les relations nécessaires à l'indexation dans le même diagramme que celui de la recherche (cf. figure 7.6 p. 293) aurait rendu ce dernier inutilisable. La figure 7.7 page ci-contre est donc à prendre comme une extension de ce dernier. Un diagramme global du système devrait combiner les éléments des deux figures, les objets recevant attributs, méthodes ou relations des deux diagrammes.

L'indexation fonctionne sur un mode très proche de celui de la sélection : le gestionnaire de recherche²⁵ dispose d'une classe `InterfaceInsertion` qui permet d'intégrer un certain nombre de `prismes` (cf. § a) p. 294).

Ceux-ci sont liés, via le `preTraitement` associé, à un objet de classe `TraitManuel` qui est capable de générer une interface pour la saisie des informations nécessaires au fonctionnement du `prisme` (cf. données envoyées par l'utilisateur dans la figure 7.4 p. 286). Ces données correspondent aux

24. L'intersection pourrait être modélisée par la génération successive de sous-collections.

25. Nous avons décidé de conserver ce nom, puisqu'un tel système est tourné vers la recherche de textes : l'indexation est en faite en vue de pouvoir ensuite rechercher des textes.

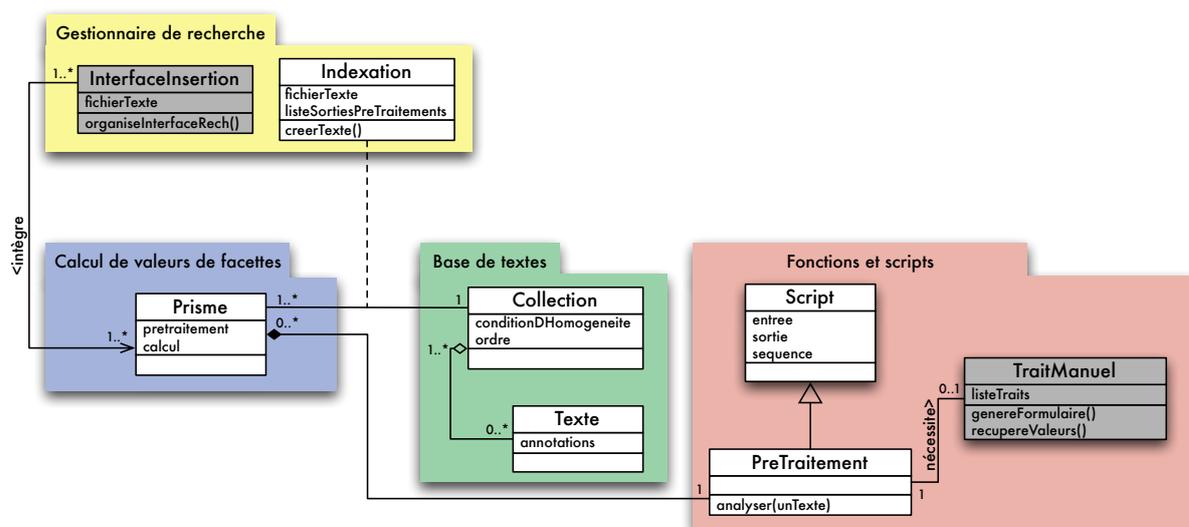


FIGURE 7.7: Diagramme de classe d'une implémentation (indexation)

données d'entrée du `preTraitement` associé au `prisme` (cf. héritage dans la figure 7.7) ou plutôt à une partie de celles-ci. En effet, le texte qui est une donnée transversale à tous les `prismes` du système pour cette opération est pris en charge par la classe `InterfaceInsertion`.

Dans la recherche, l'équivalent de `TraitManuel` était en quelque sorte « ContextePedagogique ». Nous avons décidé de ne pas organiser les deux éléments de la même manière. En effet, du fait du sens du contexte pédagogique, en tant que notion, il nous a paru plus pertinent de l'associer au `Prisme`. Ceci d'autant plus que, comme la relation entre `prisme` et `calcul` est de cardinalité [1,1], il est équivalent de lier un objet `ContextePedagogique` à l'un ou à l'autre. Au contraire d'un `calcul`, un `preTraitement` peut être lié à un `prisme` ou une `visualisation`. Nous avons considéré qu'étant donnée la nature de la tâche d'évaluation (concernant en règle général un texte à la fois), une `visualisation` pourrait effectuer son `preTraitement` au moment de l'affichage. Cependant, il n'est pas impossible que ce soit une erreur. Dans cette éventualité, il paraît plus pertinent de lier la demande d'information au `preTraitement` accessible via la classe `Prisme` et la classe `Visualisation` que de l'associer à la classe `Prisme`. En effet, certaines `visualisations` ne sont liées à aucun `prisme` et si nous avons lié `TraitManuel` à la classe `Prisme`, les `preTraitements` de ces `visualisations` n'auraient aucune possibilité de prendre en compte des `TraitManuel`. Nous considérons ici, la nécessité de permettre ce type de traitements²⁶ très improbable. Nous ne le présentons donc pas dans notre diagramme. Cependant, il est possible de revenir sur cette décision à moindre frais.

Une fois la saisie de l'utilisateur terminée chaque `prisme` dispose de toutes les données d'entrées nécessaires à l'exécution du `preTraitement` qui lui est associé. Le `prisme` pourrait alors effectuer son analyse (via le `preTraitement`) ; toutefois, la collection globale doit pouvoir prendre en charge tous les points de vue proposés par chacun des `prismes`, d'où la création d'une classe associative `Indexation` qui met en relation une `collection` et un ensemble de `prismes`. Sa fonction est de combiner toutes les informations liées aux textes au sein de la `collection`. Ceci nous permet de faire la différence entre le fichier texte envoyé par l'enseignant (pour lequel nous ne définissons

26. `preTraitement` associé à une `visualisation` non contextuelle, indépendante de tout `Prisme` s'appuyant sur des données manuelles.

pas de classe) et un objet de classe `Texte`. Le fichier est le texte dans sa forme la plus brute, alors que l'objet de classe `Texte` est augmenté des informations issues des prismes.

7.5.6 Maquette

La structure précédemment définie nous semble cohérente pour l'implantation d'un système de recherche de textes indexés pédagogiquement pour l'enseignement des langues. Cependant, la création d'un tel système ne se limite pas à la définition de chacune des classes. Le développement d'un tel outil passerait à notre sens par la définition précise de facettes (cf. § 7.6 p. 302), par une phase extensive de test sur l'adéquation entre les facettes définies et l'usage qui en est fait par les utilisateurs. Tout ceci est en soi un projet de recherche nécessitant la collaboration de multiples acteurs. Notre objectif est uniquement de tester la possibilité d'implanter une version d'un tel système. Nous allons donc nous contenter de l'implantation d'une maquette qui proposera une version réduite du modèle ci-dessus.

Nous tenterons malgré tout d'implémenter le système de façon relativement générique, de manière à ce qu'il soit possible d'en réutiliser certains éléments pour en faire un véritable prototype, qui puisse être soumis à des utilisateurs-enseignants. Le code généré est donc commenté en conséquence et compilé avec « `phpDocumentor`²⁷ »²⁸.

a) Langage(s)

Un tel système se doit de supporter les notions de client et de serveur, puisqu'il s'agit de proposer à plusieurs utilisateurs d'accéder au même ensemble de textes. Nous nous sommes à nouveau tourné vers *PHP: Hypertext Preprocessor* (PHP) lié avec des bases de données MySQL (et JavaScript pour certaines interfaces). Outre l'avantage d'être une configuration avec laquelle nous avons travaillé régulièrement (confort de programmation), elle est gratuite, portable, supporte la programmation objet et surtout ne demande pas d'autre logiciel côté client, qu'un navigateur Web. Une telle implantation ne demande donc *a priori* aucune d'installation côté client (enseignant).

b) Transformation de textes et formats

À travers les différents traitements appliqués aux textes, des informations leur sont ajoutées, leur format altéré. Une système devra jongler avec plusieurs formats.

Tout d'abord, quel format pour les textes ajoutés à la base ? Les enseignants utilisent vraisemblablement divers formats pour le stockage de leurs fichiers, au sein desquels coexisteront formats propriétaires et formats ouverts. L'une des premières étapes d'un pré-traitement sera le transcodage vers un format utilisable par les fonctions. Ce problème n'est absolument pas basique dans la mesure où un prisme peut avoir recours à des éléments de structuration des documents, qui nécessiteront une prise en charge fine. Dans le cadre de notre maquette, nous nous contenterons de considérer des textes bruts et au format *HyperText Mark-up Language* (HTML), dont nous n'exploiterons pas toutes les informations. Les textes seront stockés dans leur forme origi-

27. <http://www.phpdoc.org/>

28. Voir l'exemple en annexe G p. a-127

nale pour pouvoir être restitués non modifiés aux utilisateurs, mais le calcul s'effectuera sur une version transformée du texte.

Le problème de la structure interne du système est complexe, puisqu'il faudra que toutes les informations nécessaires à l'exécution des calculs soient présentes. Or, décider de centraliser toutes les informations pour un texte donnée au sein d'un document XML, ce qui pourrait être une solution, ne sera pas forcément viable selon les prismes présents dans le système. Une solution à court terme (celle que nous avons choisie) est de multiplier les versions des textes en fonction des prismes présents, chaque prisme nécessitant une représentation spécifique créera au cours de son pré-traitement une nouvelle occurrence du texte répondant à ses besoins. Cependant, dans l'éventualité de la création d'un système contenant de nombreux textes, cette solution particulièrement coûteuse en ressource sera à reconsidérer.

c) Classes

N'étant pas destinée à être testée en l'état par des utilisateurs, la maquette aura une interface utilisateurs extrêmement spartiate. Nous n'avons implémenté aucune interface pour l'indexation, tout devra s'effectuer en lignes de commandes. Pour la recherche l'accent est mis avant tout sur la sélection : les classes `InterfaceRecherche` et `ContextePedagogique` bénéficieront d'interfaces minimalistes, alors que nous n'allons pas développer d'autres visualisations qu'une représentation tout à fait basique du texte, limitant ainsi au maximum la mise en œuvre de la classe `Evaluation`.

Les autres classes seront toutes implantées (même si à nouveau il s'agit d'une implémentation minimaliste). La maquette tentera malgré tout d'être générique et de permettre l'ajout de prismes en affectant le moins possible la structure globale du système. Nous avons donc ajouté une classe `BD`²⁹ pour permettre de stocker et récupérer des objets dans une base de données (par exemple des `prismes` ou des `scripts`).

d) F_{nbmots}

Comme première facette à réaliser, nous nous sommes tourné vers une facette élémentaire non contextuelle. Il s'agit donc de la longueur du texte, exprimée en mots. Le `prisme` associé est relativement basique. Le `preTraitement` consiste à découper un texte en mots, mais également à les compter. Comme le nombre de mots d'un texte ne dépend pas du contexte pédagogique, l'assignation d'une valeur pour un texte donné a lieu au niveau du pré-traitement, le « calcul » de la valeur de la facette ne consistera donc qu'à récupérer une valeur préalablement stockée.

Le pré-traitement prendra en entrée un texte brut (ou XML) et renverra un entier.

Ce type de calcul serait applicable à toute facette constante (cf. § 6.4.1 p. 266). Cependant, dans certains cas, comme celui de F_{nbmots} la sémantique de la facette influencera le `prisme` associé, notamment par la prise en compte de la contrainte (cf. § 6.3.1 p. 264). Pour le `prisme` du modèle la contrainte pourra être exprimée de plusieurs manières : $valeur \pm tolérance$ ou sous forme d'intervalle $min \leq valeur \leq max$.

29. Cf. Annexe G p. a-127.

e) $F_{e^t Rep}$

L'autre facette à mettre en œuvre dans la maquette est le nombre d'éléments représentatifs d'une notion, que nous avons déjà évoqué à de nombreuses reprises (cf. § b) p. 237). La première étape de la prise en charge de $F_{e^t Rep}$ est la définition d'une acception de la propriété. Qu'est-ce qu'une notion dans notre maquette ?

Dans leurs recueils les enseignants évoquant la « grammaire » comme critère de tri font souvent référence aux temps et aux modes ([32, 109, 164, 209], cf. tableau F.1 p. a-56). Cependant une notion pourrait également être « la voix passive » [164] (cf. tableau F.1 p. a-56), « le devoir » [47] ou « le conseil » [103] (cf. tableau F.25 p. a-64). Ces notions décrites en des termes hétéroclites et concernant des notions d'ordre tout à fait différent, nous ne pourrions toutes les modéliser au sein du prisme en question. C'est ici qu'interviennent les outils informatiques disponibles qui doivent nous permettre de modéliser la propriété. Un analyseur morphologique (comme TreeTagger³⁰, qui comporte notamment l'avantage d'être gratuit et multilingue) peut nous permettre d'arriver à une acception de « notion » pour le prisme en question. Nous réutiliserons les sorties de l'outil pour permettre d'effectuer une recherche. Le formalisme proposé pour la question 24 du second questionnaire (cf. tableau F.24 p. a-61) semblait aisément pris en main par les enseignants qui la plupart du temps s'y conforment. Nous proposerons donc aux utilisateurs de rechercher des triplets de token consécutifs. Chaque élément du triplet sera décrit par un trait pris en charge par l'outil. Le choix de procéder par triplet est une limitation que nous imposons pour limiter les temps de calcul. Selon les fonctions utilisées l'utilisateur pourra combiner différents types de propriété.

TreeTagger propose par exemple pour chaque *token* un triplet (*word, pos, lemma*)³¹ (cf. tableau 7.1 de la présente page).

WORD	POS	LEMMA
Cinnamon	NN	cinnamon
rolls	NNS	roll
are	VBP	be
delicious	JJ	delicious
.	SENT	.

TABLE 7.1: Exemple de sorties de TreeTagger

En cas d'usage de TreeTagger, il sera possible de faire le choix de décrire le mot, le *pos* ou le lemme. Permettant ainsi des triplets présent dans le tableau F.25 (p. a-64) comme, par exemple, (*lemma = want, word = to, pos = VB*) [22].

Nous pourrions développer cet exemple en montrant comment le calcul peut assumer « l'interprétation de la requête » (cf. figure 1.2 p. 13). En effet, l'utilisation des traits TreeTagger peut être difficile d'accès et permettre à l'enseignant de proposer des structures émanant de son vocabulaire (ex : *-ing* pour *pos = VBG*) serait primordial dans l'optique de confronter le système aux utilisateurs.

Pour ce prisme la contrainte sera exprimée sous forme d'intervalle $min \leq valeur \leq max$.

30. <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>

31. (mot-forme, catégorie morphologique (*pos* signifie *part of speech*), lemme).

f) Indexation

La **collection** sera représentée par une table de tous les textes qu'elle contient avec leurs caractéristiques (identifiant, lien vers le fichier original, $F_{nbmots}(T)$) et un index des *token* existants dans les **textes** (lemme, mot, cat³², listePos³³).

Pour rendre les textes utilisables au niveau de la recherche, il faut pouvoir les transformer de manière à ce qu'ils soient pris en charge par la **collection**. Dans le cas où cette dernière cumule plusieurs tables SQL au sein d'une même base, l'objet **indexation** recevra :

- P_{nbmots} ->preTraitement() : le nom d'un champ dans une table et une valeur ;
- P_{etRep} ->preTraitement() : un ensemble de requêtes pour la table créée pour la facette correspondante ;
- un fichier ;

L'objet créera alors une requête pour ajouter le texte à la base (le fichier) récupérera un identifiant qu'il utilisera dans P_{nbmots} ->preTraitement() pour renseigner le champ concerné et dans P_{etRep} ->preTraitement() pour toutes les requêtes.

g) Selection

Les deux **prismes** pourront être combinés et l'utilisateur spécifier la longueur et une fourchette du nombre d'éléments représentatifs d'une structure définie par une séquence de trois éléments. C'est un objet **selection** qui s'en chargera. Il recevra des deux prismes des portions de requêtes permettant d'obtenir la valeur de la facette³⁴. La fonction de l'objet **selection** est donc de créer une condition d'homogénéité interprétable à partir des données des prismes.

h) Extensions

Dans la maquette, peu de temps a été accordé aux interfaces qui fonctionnent donc principalement directement à partir des formalismes des fonctions utilisées. Dans le cadre d'une version « admissible³⁵ » du système beaucoup d'étapes supplémentaires seraient nécessaires, ne serait-ce que pour permettre aux utilisateurs de formuler les CP selon des termes proches de leur pratiques. Il faudrait savoir comment les enseignants appellent les traits morphologiques correspondants aux valeurs utilisées en sortie des fonctions. Ceci demanderait la consultation de didacticiens et d'enseignants. Il faudrait ensuite travailler sur l'interface : utiliser des menus contextuels ? laisser la possibilité de formuler en ligne de commande ? etc. Pour cela, il serait opportun de travailler avec des spécialistes d'Interface Homme-Machine (IHM), avant de soumettre le produit à des β -testeurs et d'adapter ensuite le **contextePedagogique** et le calcul du **prisme** en conséquence.

32. Catégorie morphologique.

33. Pour simplifier : liste des couples (*identifiant de texte, liste des positions du token dans le texte*). Nous passons outre certains détails de l'implémentation, puisque notre objectif ici est de montrer un exemple d'**Indexation**.

34. Par exemple, P_{nbmots} ->calcul(CP) :

```
IF('collection'.'nbmots'> 200 AND 'collection'.'nbmots'<350, 'nbmots', NULL) ;
```

35. Voir Wikipedia [2009h].

7.5.7 Conception d'un système générique

Nous avons proposé ici, une modélisation objet permettant de rechercher des textes en fonctions de facettes. Il s'articule autour de 4 *packages* principaux dont l'objectif est de permettre une encapsulation suffisante des différents niveaux de traitement et de données, pour permettre d'ajouter quasi-indépendamment des prismes. Dans notre maquette, la collection recense les textes, dans une table et à partir du moment où les identifiants des textes sont cohérents il est possible d'associer aux nouveaux prismes de nouvelles tables.

Il serait d'ailleurs imaginable de proposer d'autres implémentations de P_{etRep} , par exemple en créant des fichiers XML qui seraient ensuite passé par des moulinettes Perl³⁶. Les modifications toucheraient :

- le **preTraitement** du prisme en question : il n'est plus utile de stocker les informations dans une base de données, il suffirait de créer un document XML ;
- l'**indexation** : il ne s'agirait donc plus d'ajouter des éléments à une table, mais de stocker des documents XML (lien stocké dans la table principale de la collection) ;
- le **calcul** du prisme : génération d'une expression régulière Perl en fonction du CP qui permette pour un texte de connaître la valeur de la facette ;
- la **selection** : au lieu de créer une condition d'homogénéité sous forme de requêtes SQL, il faudra combiner les résultats de la requête de F_{nbmots} avec une liste d'identifiant de textes issu d'un programme Perl répétant le calcul pour tous les éléments de la collection.

Le système est donc minimal et pourrait être amélioré d'un point de vue fonctionnel, mais l'architecture devrait être suffisamment générique pour que les modifications liées à la mise à jour ou l'ajout d'un prisme ne demandent pas une refonte complète du système.

7.6 Ajout d'un prisme au système

Si l'architecture proposée est censée permettre de faire évoluer le système, le travail nécessaire pour faire d'un prisme une fonctionnalité « admissible » est un processus potentiellement long (cf. § h) p. 301). Qu'en est-il de la création d'une nouvelle facette ? Après avoir donné des exemples de facettes très basiques, nous allons conclure ce chapitre par une description de la problématique de la création d'une facette à travers l'exemple du projet PPF que nous avons déjà évoqué (cf. § a) p. 269). Nous n'allons pas discuter ici les résultats du projet, mais plutôt l'aborder comme un projet pluridisciplinaire à l'image du domaine de l'ALAO (cf. § 2.1.2 p. 35) à travers son déroulement et les interactions entre les spécialistes des différents domaines (cf. § c) p. 50).

Bien que les objectifs du projet PPF soient différents des nôtres, nous allons les décrire selon les concepts que nous avons introduits.

7.6.1 Définition de la facette

Nous allons ici nous focaliser sur la partie du projet dédiée à la description de la polyphonie dans les textes. L'idée d'aborder cette question émanait en particulier d'un échange entre des enseignants et didacticiens du projet. À partir du moment où le phénomène à étudier était défini, l'étape suivante était d'arriver à une acception de la notion qui corresponde à la problématique

36. <http://www.perl.org/>

du projet, c'est-à-dire une utilisation avec les apprenants. Cette partie du projet était piloté en particulier par les linguistes du projet qui ont proposé une série de descriptions des phénomènes aux autres participants afin d'arriver à un consensus. Le choix final du paradigme employé s'est fait après avoir sacrifié la puissance descriptive des premières propositions pour parvenir à des propriétés jugées adaptables au public d'apprenants³⁷. Cette phase correspondrait dans notre problématique à la définition de la facette (la propriété, ses valeurs et sa sémantique) ou plutôt la définition de la facette potentielle, puisque sans prisme associé, la propriété n'est pas une facette.

7.6.2 Définition du pré-traitement

Ces concepts devaient être employés pour annoter manuellement des corpus. Cette annotation étant destinée à être utilisée dans un système informatique, il fallait adapter la description des passages entre guillemets à un traitement informatique. Les informaticiens du projet sont donc intervenus pour s'assurer que l'acceptation de la propriété définie soit interprétable avec des outils informatiques : l'un des problèmes était la possibilité d'annotation en XML. Une fois un schéma d'annotation défini, un petit outil a été développé et une première phase de pré-traitement (manuel) lancée. Cette phase était un test du schéma produit. Malgré des résultats de comparaison interannotateurs encourageants, des désaccords subsistaient dans certains cas [Rinck & Tutin, 2007, p. 83].

Une nouvelle phase de réflexion sur les valeurs de la propriété s'est engagée à partir de la catégorisation des situations posant problème aux annotateurs (modification de la facette du fait de problèmes soulevés pendant la définition du prisme). Finalement, un schéma d'annotation stable est défini en accord avec les différents pôles de compétences représentés : il décrit adéquatement les phénomènes du point de vue des linguistes et des didacticiens, permet une annotation cohérente et réutilisable dans les outils informatiques assurant l'équivalent de ce que nous appelons dans notre modèle le calcul de la valeur de la facette.

7.6.3 Définition du contexte pédagogique (et du calcul)

L'objectif du projet n'est pas de rechercher des textes, mais notamment de produire un corpus utilisable en classe dans des activités de type *Data Driven Learning* (DDL) (cf. [Rinck & Lebarbé, 2005, p. 93]). Malgré des problématiques différentes, nous continuons d'utiliser notre terminologie : nous ne rentrons pas dans le détail des notions et les structures recherchées (type de discours rapporté, croisé avec le segment introducteur) dans le corpus pourraient être réutilisées dans le contexte de notre système pour rechercher des textes entiers. La différence de problématique ici, concerne en particulier l'objet de la recherche.

La définition d'un outil qui permette de repérer une structure en fonction d'une requête (équivalent du CP) pourrait constituer un équivalent de la définition du calcul dans notre implémentation. Cette phase fait encore intervenir plusieurs types de compétences. Il faut s'interroger sur la manière dont l'outil peut être utilisé en classe (didactique) [Antoniadis *et al.*, 2007a] et en fonction de ces besoins permettre des recherches adaptées (informatique).

37. Pour la typologie employée, voir [Rinck & Tutin, 2007, pp. 80–81].

7.6.4 Cadre pour la conception de facettes

À travers un exemple sortant du cadre strict de notre problématique nous avons pu donner des exemples des difficultés liées à ce qui pourrait être une facette pour notre système. En effet, plutôt que de rechercher des portions de texte (corpus), le travail du projet PPF pourrait être réutilisé pour chercher des textes contenant ces portions. Les trois phases du projet que nous avons décrites sont synthétisées dans la figure 7.8 de la présente page à travers leurs objectifs.

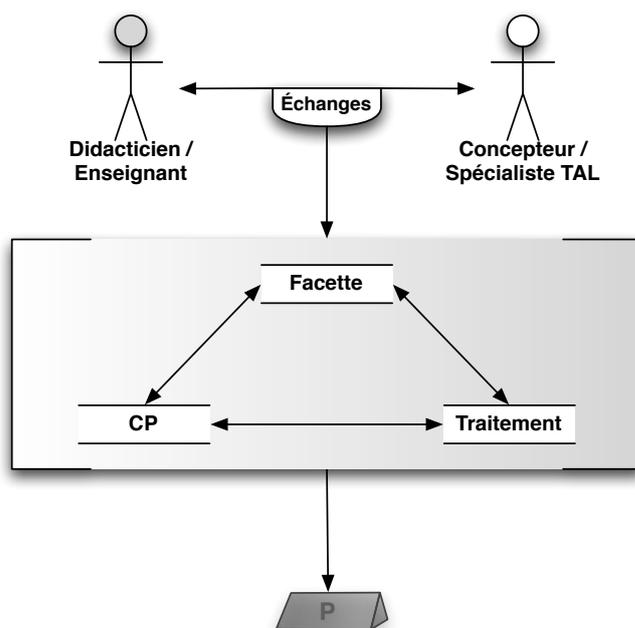


FIGURE 7.8: Problèmes clés dans la définition d'un prisme

Dans le traitement de chacun des trois problèmes clés de la figure 7.8, différentes compétences interviennent : si chaque résolution de sous-problème est plus ou moins pilotée par un pôle, l'échange s'est avéré nécessaire pour chacune. Enfin, chaque sous-problème peut influencer les autres et ne peut être traité indépendamment.

Une prise en charge dans notre système reprendrait les problèmes énoncés :

- définition de la facette ;
- adéquation avec le prisme ;
- mise en place d'un pré-traitement ;
- adéquation avec le calcul ;
- modélisation du CP ;
- création d'une interface de recherche adaptée.

Mais à ceux-ci s'ajouteraient les difficultés liées à l'évaluation, à savoir la définition de visualisations adaptées aux besoins des utilisateurs. La proposition de notre modèle, comme l'illustre cet exemple, ne pouvait déboucher sur une base de textes indexée pédagogiquement pour l'enseignement des langues complètement opérationnelle. En revanche, il fournit un cadre au sein duquel analyser les besoins et les pratiques des enseignants en vue de définir de nouvelles facettes, proposant de plus en plus de valeur ajoutée pédagogique.

Conclusion

Chapitre 8

Perspectives et points de vues

Comme l'indique le titre de ce manuscrit, l'objectif de notre travail n'était pas de développer un produit « fini ». Notre approche se centrait sur la nécessité de fournir un cadre pour la conception de bases de textes indexées pédagogiquement pour l'enseignement des langues. Dans ce qui constitue finalement un travail de défrichage nous n'avons pu prendre en compte l'intégralité des problèmes qui se présentaient à nous.

8.1 Quels textes indexer ?

La première question que nous avons décidé d'éviter concerne l'ajout de textes à un système. Nous avons décidé de laisser les utilisateurs se charger de peupler la base, orientant donc le système vers un partage de ressources. Nous considérons toujours qu'il s'agit d'une solution viable (à quelques détails légers près). De plus, les textes proposés seront des textes éprouvés dans l'enseignement des langues, que les utilisateurs pourront réutiliser.

8.1.1 Textes tout-venants et RSS

Cependant cette conception de la population de la base pose un problème par rapport à notre approche globale : nous voulons proposer aux enseignants des textes qui n'ont jamais été utilisés dans le contexte qui les concerne (dans la mesure où ils sont jugés adaptés). Or dans cette optique, le fait de ne dépendre que des enseignants pour l'ajout de ressources limite le fond documentaire à des textes dont l'usage dans l'enseignement des langues est attesté (ou projeté). Certes, ils pourront être inclus dans une activité relevant d'un contexte pédagogique différent de celui qui a mené à leur première activité, mais il y a là une forme de paradoxe.

Tout d'abord, pour pouvoir ajouter un texte à une base de textes indexée pédagogiquement pour l'enseignement des langues, il faut préalablement le rechercher par un autre biais. Ensuite, comme le signalait une enseignante : « Les textes allemands perdent vite leur intérêt, vu le contexte socio-économique et vu l'évolution de nos élèves. » [128] (cf. § b) p. 168). C'est d'ailleurs une remarque qui nous avait été faite par Liam Murray à UNTELE 2004 : il avait noté que les enseignants cherchaient régulièrement des textes en fonction de l'actualité et que dans ce contexte une telle base serait inutile. Même si les textes en rapport avec l'actualité ne concernent pas toute recherche, cette série de remarques soulève un point qui semble pouvoir être résolu à moyen terme,

pour les prismes d'un système capables de prendre en charge des textes tout-venants (et même certains autres).

Une piste est offerte par Fairon [2006] ou Kurella *et al.* [2008] qui proposent de s'appuyer sur des flux *Really Simple Syndication* (RSS). En effet, ces derniers proposent des en-têtes permettant d'accéder à des informations telles que l'auteur, la date de publication, le titre de l'article ou la source. En outre un lecteur de flux pourra accéder à des textes récemment publiés. Enfin, pour les structures proposant des liens permanents, le système pourrait renvoyer vers les textes dans leur contexte original et ainsi éviter certains problèmes légaux.

8.1.2 Textes « didactisés »

Nos données suggéraient que les textes fabriqués pouvaient eux aussi faire l'objet de recherches (cf. tableau 4.8 p. 170) et donc être inclus dans une base de textes indexée pédagogiquement pour l'enseignement des langues. Dans la figure 4.6 (p. 199), nous regroupons les textes fabriqués et les textes que les enseignants appelaient « didactisés » sous la même étiquette « écrire ». Selon le point de vue qui était le nôtre à ce moment là, cette acception est correcte, d'autant plus que, pour certains, l'authenticité n'est pas à prendre comme une propriété binaire, mais un continuum (cf. § b) p. 67). Néanmoins, du point de vue des ressources elles-mêmes il y a des informations qui sont perdues à considérer les textes « didactisés » comme des textes fabriqués.

Un texte « didactisé » est créé à partir d'un autre texte, par abrègement ou adaptation (cf. § a) p. 65) ; il est donc fortement lié à ce dernier. Il nous semble que conserver une trace de ce type de relation est un exemple de données sous-jacentes principalement destiné aux visualisations qui mériterait que l'on s'y intéresse. D'une part, ceci permettrait la mise en place d'une facette qui ferait une différence entre texte fabriqué et texte « didactisé ». Ensuite, un enseignant pourrait dans le cadre de sa tâche d'évaluation consulter non seulement les textes, mais aussi les textes dont ils proviennent ou ceux dont ils découlent, ce qui lui permettrait pour un objet qui lui paraît pertinent de choisir la version la plus adaptée à son propos.

Mais l'intérêt principal de suggérer aux enseignants de réinjecter dans le système leurs textes didactisés est un objectif de compréhension du processus de recherche. En effet, à travers le cycle de vie des documents, on pourra tenter de comprendre pourquoi tels éléments ont été reformulés, pourquoi d'autres ont été abrégés (des explications fournies par les enseignants ajoutant le texte ainsi modifié permettraient d'en savoir plus sur leurs intentions). Ceci constituerait donc une instrumentation¹, afin d'obtenir de nouvelles informations pour la compréhension de la planification à partir de textes dans l'enseignement des langues (et éventuellement d'arriver à la création ou l'adaptation de prismes).

8.2 Élaboration de prismes

S'il nous semble que la possibilité de rechercher des textes en fonction du nombre d'éléments représentatifs d'une notion répond à un besoin des enseignants, sa mise en place ne nécessite pas la production d'un modèle pour décrire les propriétés. Notre travail est avant tout destiné à proposer une manière de considérer les propriétés des textes pour une indexation pédagogique. Sa

1. Instrumentalisation si le système est modifié pour demander des informations supplémentaires aux enseignants.

généricité doit permettre l'ajout de nouvelles facettes et la réutilisation de traitements existants. L'idée globale serait de créer des prismes de portée didactique croissante.

8.2.1 Réutilisation de pré-traitements et combinaison avec de nouvelles fonctions

Reprenons les résultats des questions posées dans notre second questionnaire (cf. chapitre 5 p. 215). Il est apparu que la longueur des textes et le nombre d'éléments représentatifs d'une notion sont liés à l'activité dans laquelle le texte est utilisé. Il paraît donc pensable de réutiliser les pré-traitements des prismes correspondants pour en faire un nouveau qui permettrait à l'utilisateur de choisir les textes en fonction d'un type d'activité et d'une notion. Il paraît évident qu'un prisme ainsi constitué serait réducteur des problèmes des enseignants et probablement peu performant. Une manière d'améliorer un tel prisme serait de prendre en compte le niveau du public posant ainsi la question de savoir comment le modéliser². Le travail de Zampa est une piste dans cette direction.

Dans le contexte du développement de la compréhension écrite en anglais langue de spécialité à travers la lecture de textes, elle propose d'utiliser *Latent Semantic Analysis* (LSA)³, pour décrire le niveau des apprenants. Celui-ci est alors symbolisé par l'espace regroupant tous les textes qu'il est censé avoir lus. En comparant cet espace avec celui généré pour chaque texte, Zampa définit une proximité optimale d'acquisition [Zampa, 2005, p. 138]. En intégrant son prototype, Recueil Automatique Favorisant l'Acquisition d'une Langue Étrangère de Spécialité (Rafales)⁴, en tant que fonction dans un système d'indexation pédagogique pour l'enseignement des langues, il serait possible de définir de nouvelles zones auxquelles appartiennent les textes, selon leur proximité avec le modèle de l'apprenant. Elles pourraient être définies en fonction du type d'activité, afin de prendre en compte la relation qui existe entre la recherche de nouveau vocabulaire / nouvelles structures et le type d'activité. Enfin, il serait possible de prendre en compte le « volume » du modèle de l'apprenant (son niveau) pour modéliser l'influence de ce dernier sur la quantité de notions inconnues dans les textes.

Il s'agit ici uniquement de pistes de facettes, les réponses des enseignants à l'ultime question de notre second questionnaire peuvent en fournir d'autres (cf. tableau F.83 p. a-125), sans pour autant couvrir la liste des possibles, qui dépend avant tout de nos connaissances des pratiques et attentes des enseignants.

Le thème est, comme nous pouvions nous y attendre, au centre de leurs préoccupations⁵, mais ils proposent plusieurs autres propriétés. Parmi elles, certaines paraissent plus aisément convertibles en facettes. Nous pouvons citer par exemple :

- « actualité » [109] (cf. § 8.1.1 p. 307) ;
- « type de texte » [201, 208, 240, 241, 247, 254] (dont la complexité est palpable à la consultation des tableaux F.79 p. a-116, F.80 p. a-118 et F.81 p. a-121) ;
- « champ lexical » [40, 208, 230, 236, 241] ;
- « rareté/fréquence des mots (critères que l'on trouve dans les dictionnaires unilingues pour apprenants type Longman Dictionary of Contemporary English) » [231], propriété appuyée par

2. Même indépendamment d'une telle facette la question risque de se poser.

3. Un système qui infère les relations sémantiques entre les graphèmes (transformés en vecteurs) à partir d'analyses uniquement statistiques.

4. Ou certains de ses composants.

5. [22, 23, 24, 26, 32, 39, 40, 57, 68, 85, 94, 101, 164, 166, 172, 196, 200, 201, 206, 208, 215, 237, 240, 242, 247].

le témoignage [26], qui propose d'indiquer si le vocabulaire du texte appartient à l'ensemble des « 300/3000/30000 » mots les plus employés ;

- « régularité de prononciation (même sur des textes écrits) » [231] ;
- « époque » [237] ;
- « variété de langue (pour anglais, ang GB et US par ex)... » [247] ;
- « y a-t-il beaucoup de mots transparents (selon la langue des élèves) » [26] ;
- les principales facettes que nous avons évoquées se trouvent également dans les propositions (structures⁶, longueur⁷, auteur⁸).

D'autres réponses pourraient également être analysées puisque nous demandions aux enseignants les critères qui pouvaient les influencer dans leurs choix pour chaque type d'activité dans la séquence utilisée dans le chapitre 5 (p. 215)⁹. La littérature peut également fournir des pistes de facettes. Tagliante [2006] propose par exemple des critères de choix de textes (tout à fait cohérents avec les propositions des enseignants)¹⁰ :

- « que le texte dose convenablement les éléments lexicaux et grammaticaux connus et inconnus » ;
- « que le contenu socioculturel permette une comparaison interculturelle » ;
- « que les différents textes proposés couvrent les domaines répertoriés dans le *Cadre européen commun de référence* (public, privé, éducationnel, professionnel) » ;
- « que les textes soient représentatifs des différents types dominants de textes (narratif, explicatifs, infonctifs, informatifs, poétiques, argumentatifs, descriptifs) » ;
- « qu'ils s'adressent à la bonne classe d'âge, surtout au début de l'apprentissage » ;
- « que le texte apporte des informations originales qui permettent d'éveiller la curiosité ».

8.2.2 Le système comme source d'informations pour les chercheurs

Toutefois, il ne suffit pas d'ajouter de nouvelles fonctions pour obtenir de nouvelles prismes : chacune des propositions ci-dessus doit être confrontée aux pratiques constatées des enseignants, afin d'évaluer la pertinence et l'utilisabilité de chaque modélisation. Les prismes doivent être dessinés pour servir un besoin des enseignants et les facettes testées afin d'évaluer leur cohérence avec les conceptions de ces derniers.

Dans ce contexte, à la manière de ce que nous suggérions dans la section 8.1.2 (p. 308), l'implémentation d'un prototype permettrait d'avoir accès à de nouvelles informations sur les pratiques des enseignants. Tout d'abord, elle serait un support pour interroger les utilisateurs sur le système (cf. processus incrémentiel et itératif, § d) p. 52). Elle pourrait également servir de support à la mise au point d'expérimentations ponctuelles sur les conceptions des enseignants et leur compatibilité avec les acceptions des propriétés utilisées dans les facettes développées.

Mais d'autres expérimentations concernant le processus de recherche et son imbrication avec des facettes élémentaires pourraient être menées. Une étude utilisant un prototype muni d'un

6. [18, 22, 23, 32, 40, 39, 65, 68, 101, 201, 206, 208, 215, 230, 236, 240, 241].

Pour en savoir plus sur les structures que les enseignants voudraient pouvoir rechercher et sur la manière dont ils s'y réfèrent, certaines questions de la section F.2 (p. a-57) peuvent être consultées.

7. [68, 85, 201, 208].

8. [66, 208, 237].

9. Voir § 5.3 p. 231 et les réponses dans les tableaux F.39 p. a-74, F.46 p. a-82, F.53 p. a-88, F.64 p. a-100 et F.72 p. a-111.

10. Nous reproduisons la liste dans son intégralité [Tagliante, 2006, pp. 136–137], même si certains critères nous semble « difficiles » à formaliser, comme par exemple le fait que le texte éveille la curiosité.

système de trace, permettrait d'étudier le comportement des enseignants une fois confrontés aux résultats d'une requête donnée :

- quels textes seront considérés comme adéquats ;
- lesquels seront rejetés ;
- au bout de combien de temps d'évaluation ;
- après avoir consulté quelle(s) visualisation(s).

Utilisé dans le cadre d'une étude longitudinale, le corpus de traces permettrait d'assister au processus d'instrumentation du prototype. Il serait éventuellement possible de constater comment les enseignants adaptent leurs requêtes pour trouver plus rapidement des textes, quels sont les prismes qu'ils combinent, quelles sont les visualisations qu'ils privilégient. Il est imaginable de demander des informations supplémentaires pour associer les actions avec une intention de l'enseignant, en demandant à l'enseignant d'exprimer sa requête librement avant de l'effectuer. Toutes ces informations pourraient être réunies dans l'optique d'améliorer le système (et les prismes), mais aussi la compréhension du processus de recherche et des conceptions des propriétés par les enseignants.

Ce type d'informations pourrait également soulever de nouveaux problèmes que nous n'avons qu'effleuré dans ce travail. Nous avons tenté d'extraire des invariants dans les pratiques d'enseignants aux parcours et problématiques très divers. En scrutant les comportements comme nous proposons de le faire, des profils d'enseignants pourraient surgir, qu'il s'agisse de leur point de vue sur les propriétés (lesquelles correspondent à leur manière de concevoir la langue qu'ils enseignent, travail sur l'acceptation des propriétés), mais aussi la manière dont ils s'adaptent au système (travail sur les interactions avec le système).

8.3 L'architecture du système

Toutes les questions posées sur les facettes potentielles et la manière de les concevoir, mettent l'accent sur la nécessité de généralité de la modélisation informatique du concept de facettes. Cela soulève le problème de la capacité de l'architecture proposée à répondre aux problèmes plus complexes auxquels elle pourrait être confrontée.

8.3.1 Modèles de l'enseignant et de l'apprenant

Les expérimentations potentielles mentionnées précédemment sont souvent perspectives à long terme, puisque la réalisation d'un prototype assez performant pour être « β -testé », le développement de suffisamment de facettes, la mise en place d'un tel protocole et son exécution demanderaient à eux seuls plusieurs dizaines « d'hommes-mois¹¹ » de travail. Mais surtout ils ne manqueraient pas de soulever des interrogations plus spécifiques dont nous ne soupçonnons pas le contenu, comme la nécessité de développer de nouveaux outils, par exemple issus du Traitement Automatique des Langues (TAL)¹² pour modéliser le point de vue des enseignants.

Toutefois, de la même manière que nous mentionnions une piste de modèle de l'apprenant à travers LSA (cf. § 8.2.1 p. 309), il paraît peu probable que la problématique de l'indexation pédagogique de textes pour l'enseignement des langues puisse affranchir complètement le contexte

11. Unité familière aux porteurs et autres rédacteurs de projets ANR.

12. Apportant ainsi des éléments de réponse à la 4^e question posée par Zock [1996] (cf. § d) p. 57)

pédagogique d’une prise en compte minimale du profil de l’enseignant. Ces deux modules constitueraient des extensions au *package* « calcul de valeurs de facettes ¹³ ». Ils semblent à première vue correspondre à des instances de notre classe `ContextePedagogique`.

Dans le cas du modèle de l’apprenant, une modélisation sous forme de divers objets `ContextePedagogique` semble adéquat. Dans la proposition d’une modélisation à partir de LSA, même si le profil d’un groupe d’étudiants doit être stocké d’une manière ou d’une autre, le fait d’avoir recours à ce profil est dépendant du prisme utilisé et fonctionnerait de manière cohérente en tant que `contextePedagogique` (cf. relation entre les deux classes d’objet dans la figure 7.6 p. 293). Ce point de vue serait cumulable avec d’autres facettes potentielles, par exemple la quantité de mots transparents, indiquée par l’enseignant [26]. Celle-ci pourrait être approchée en utilisant un dictionnaire bilingue pour généré automatiquement une liste des mots transparents en fonction de la distance d’édition qui existe entre chaque mot et sa traduction ¹⁴. La spécification du profil de l’apprenant se ferait alors, pour ce prisme, en spécifiant sa langue maternelle (et éventuellement des contraintes sur la quantité de mots transparents ¹⁵). Les deux aspects du modèle de l’étudiant pourraient coexister tout en n’étant invoqués que si l’enseignant n’a recours qu’au prisme concerné.

8.3.2 Interfaces

La question du modèle de l’enseignant implique directement les interfaces que nous n’avons pas traitées dans le cadre de notre implémentation. En effet, la prise en compte de l’enseignant correspondrait idéalement à tenter d’intégrer sa perception des propriétés dans les résultats des recherches : la même requête pour deux enseignants différents ne renverrait pas les mêmes résultats. Mais une solution plus réaliste (au moins à court terme) est de tenter de lui proposer des interfaces plus adaptées. Il s’agit là d’un problème à part entière. D’un point de vue didactique d’abord, puisqu’il faudrait pouvoir caractériser des phénomènes selon l’acceptation de l’enseignant. Un exemple simpliste et caricatural serait de proposer à l’enseignant la terminologie qu’il emploie (prétérit ou *simple past*). Mais il peut s’agir d’aspects beaucoup plus directement liés à l’interaction de l’enseignant avec le système (mode expert en ligne de commande ou mode débutant avec des menus déroulants ¹⁶).

Dans un cas comme dans l’autre, la gestion du problème passerait par la conception d’interface en ayant à cœur leur « plasticité » : « La plasticité d’une matière est sa capacité à se déformer suivant des contraintes, sans se casser. De la même manière, nous définissons la plasticité d’une interface, comme sa capacité à s’adapter (se déformer) aux contraintes du matériel ou de l’environnement, en restant utilisable (sans se casser) » [Thevenin, 1999, p. 152]. Il s’agira d’adapter l’Interface Homme-Machine (IHM) à son contexte d’usage : un triplet <utilisateur, plate-forme, environnement> Serna *et al.* [2009].

13. Cf. figure 7.6 p. 293.

14. En utilisant l’algorithme de Levenshtein [Algo-Lev, 2008] ou l’un de ses avatars (par exemple l’algorithme de Myers [1986]).

15. La quantité de mots transparents pourrait être ramenée à une proportion parmi les mots du textes ou parmi les noms (par exemple) du textes en utilisant une analyse morphologique.

16. Dans le cas par exemple des structures à rechercher, sera-t-il possible de proposer un formalisme qui permette aux enseignants d’exprimer simplement leurs besoins ? La formulation des requêtes à partir d’exemples de structures recherchées présenterait-elle un avantage suffisant pour les enseignants pour justifier le travail qu’elle nécessiterait ?

8.3.3 Combinaison de prismes

Outre l'adaptation à l'utilisateur, la plasticité des interfaces concerne la plate-forme. Il pourra s'agir de préoccupations standards concernant le dispositif¹⁷ utilisé pour accéder au système, mais aussi de la version de ce dernier. Le modèle se veut évolutif et permettant d'ajouter des prismes à un système existant. L'interface devra donc pouvoir assumer cet aspect du modèle tout « en restant utilisable (sans se casser) ». Cela pose une nouvelle question. Dans le cadre de notre modèle, la solution pour combiner des prismes consiste à en créer un nouveau. Celui-ci réutiliserait des fonctions issues des traitements des autres.

Reprenons l'exemple simpliste d'une modélisation de l'activité par la conjonction d'un nombre d'éléments représentatifs d'une notion et d'une longueur de texte¹⁸. Le pré-traitement d'un tel prisme ne demande pas d'autre information que la séquence de traitement des deux autres. Par contre, son calcul ajoute une couche répercutant l'expression d'une contrainte du Contexte Pédagogique (CP) vers les éléments utilisés par ailleurs par les deux prismes sur lesquels il s'appuie (en orange¹⁹ dans la figure 8.1 de la présente page).

FIGURE 8.1: Exemple de portion d'interface associée à un prisme

Dans l'état actuel du modèle, l'interface (objet `ContextePedagogique` qui lui est associé) est censée expliciter à l'utilisateur sa reformulation (dans la figure 8.1), le cadre « critères de recherche » pourrait donc être mis à jour automatiquement quand l'utilisateur modifie la partie « recherche assistée », explicitant ainsi le lien existant entre les deux et lui permettant d'affiner les critères qui lui sont proposés. Dans cette solution, il est considéré que c'est l'enseignant qui, à travers les informations qui lui sont transmises, devra faire le lien entre ce prisme (et son interface) et les

17. Caractéristiques matérielles et logicielles.

18. Nous utilisons cet exemple pour sa simplicité, sans postuler son efficacité ou sa cohérence.

19. Dans cette exemple, l'activité ($P_{activite}$) se fonde sur une longueur (P_{nbmots}) et un nombre d'éléments représentatifs (P_{etRep}) qui constitueraient par ailleurs deux autres prismes associés à deux autres interfaces. Pourtant l'interface de saisie du CP de $P_{activite}$ reprend (dans un sous-cadre) le type d'information qui pourrait être associée aux deux autres prismes.

autres prismes qui désignent les mêmes propriétés mais avec un point de vue différent. On peut imaginer que l'enseignant ne combinera pas la recherche de la figure 8.1 avec une recherche du type *nbmots* > 300.

Cependant nous pouvons nous demander s'il serait pertinent d'indiquer les liens entretenus par les différents prismes du système, et permettre leur composition sans repasser par le niveau des fonctions. Il suffirait peut être d'utiliser les liens entre prisme et fonctions (via les scripts) pour permettre la gestion, par les modules d'interface, des liens qui peuvent exister entre les propriétés (cf. plasticité selon la plate-forme). Enfin, faudrait-il permettre à l'utilisateur de proposer ses propres « prismes composés » ? Nous entendons par là l'ajout d'une couche supplémentaire (du type « utilisation » dans la figure 8.1) qui mettrait à jour des relations existantes entre les deux prismes. La simple conjonction de deux prismes est bien sûr possible dans l'état actuel des choses.

8.3.4 Collections

La conjonction de prismes évoquée dans la section précédente n'est, malgré tout, en rien triviale. Dans le cas de notre maquette, le problème est facilité par l'homogénéité des structures de données. Nous avons utilisé pour tous les prismes à mettre en œuvre deux tables de la même base de données MySQL. Les jointures entre les différentes tables ne sortent donc pas du cadre des pratiques les plus standards dans le domaine.

Cependant, l'objectif est de permettre l'ajout de fonctions diverses, quitte à cumuler des représentations très différentes des textes. C'est notamment l'exemple que nous voulions donner avec la recherche dans une collection à partir d'expressions régulières perl (cf. § 7.5.7 p. 302). Le concept de collection et de sous-collection est destiné à faciliter la communication entre des systèmes de stockage de données différents, réduisant les requêtes complexes à des requêtes successives. Cependant, ce type de solution maximisant la quantité de résultats intermédiaires peut s'avérer coûteux en termes de stockage et/ou d'efficacité, en particulier pour des requêtes complexes.

Afin d'améliorer la réutilisabilité de notre maquette et de faciliter la programmation objet, nous avons créé une classe *BD*²⁰ qui encapsule les bases de données employées et permet de récupérer directement des objets. Cependant, la conception de cette classe ne s'est pas articulée autour d'autres structures de stockage de données en particulier. Elle est à concevoir comme une boîte vide à construire en fonction des solutions choisies. Il peut s'agir d'opter pour une technologie existante. Celle-ci peut être propriétaire comme IBM DB2, qui permet le stockage hybride de données dans une base de données relationnelle et sous forme de documents *eXtensive Mark-up Language* (XML) [Nicola & van der Linden, 2005, figure 1], ce qui est particulièrement adapté dans le cas du traitement de ressources textuelles. Une solution libre comme PostgreSQL²¹ semble offrir le même genre de possibilités.

Ce problème de l'hybridation des bases de données, n'est pas l'apanage de notre seule problématique. Nous pouvons par exemple citer Panchenko [2009] qui désire proposer à des développeurs de rechercher des portions de code et leur offrir plusieurs points de vue sur les programmes analysés (recherche plein texte ou structurelle). Sa solution passe alors par le développement d'une structure de stockage *ad hoc*, ce qui pose la question de savoir si l'utilisation d'un SGBD existant est viable dans l'optique d'obtenir un système performant ou s'il faudra développer une structure pour le système. Le problème d'une hyperspécialisation de la structure de données, plus efficace,

20. Cf. Annexe G p. a-127.

21. <http://www.postgresql.org/docs/current/static/datatype-xml.html>.

est qu'elle irait à l'encontre de nos objectifs de modularité et d'évolutivité du système. La piste d'un SGBD est donc à poursuivre en priorité.

8.4 Adaptation à d'autres utilisations

De la même manière que des solutions prévues pour d'autres problématiques (par exemple celle de Panchenko [2009]) pourraient être utilisées pour l'indexation pédagogique pour l'enseignement des langues, les concepts développés ici peuvent être réutilisés pour d'autres problématiques.

8.4.1 Indexation pédagogique de problèmes pour l'enseignement de l'algorithmique

Dans l'optique de proposer des exercices dans le contexte de l'enseignement de l'algorithmique, une base pourrait être créée pour recenser les problèmes proposés aux apprenants et leurs solutions. Il semblerait que les concepts développés pour l'indexation pédagogique de textes pour l'enseignement des langues puissent être adaptés à cette application²².

Par la nature non-ambigüe des langages informatiques, leur analyse est nettement plus simple que celle de la langue naturelle. En considérant les concepts de base de l'algorithmique²³ et en les augmentant de quelques considérations, nous pouvons établir une liste de facettes aisément implémentables permettant à l'enseignant de formuler des requêtes élaborées :

- le nombre de structures de contrôles ;
- le degré d'imbrication des structures ;
- la présence de récursions ;
- le nombre de variables ;
- le nombre de types de variables utilisés ;
- le nombre de types personnels créés pour l'algorithme ;
- le nombre d'instructions.

Comme dans notre cas, ces facettes relativement élémentaires peuvent donner naissance à des facettes plus élaborées, comme par exemple un indice de difficulté du problème. Celui-ci pour avoir un sens demanderait la prise en compte du niveau des apprenants et du type d'activité dans le contexte pédagogique. D'un point de vue instinctif qui serait à éclaircir, nous pouvons proposer quelques ébauches de relations entre ces différents éléments :

- plus le niveau de l'apprenant est élevé plus le nombre d'instructions, de structures de contrôles (etc.) pourra être élevé ;
- plus le niveau de l'apprenant est élevé, moins le nombre d'instructions augmente la difficulté du problème ;
- si l'apprenant doit écrire l'algorithme il faudra que les métriques définies ci-dessus aient des valeurs plus faibles que s'il doit comprendre un algorithme qui lui est fourni.

Enfin, on peut imaginer des facettes intégrant le langage dans lequel l'exercice sera effectué. Quoi qu'il en soit, les couples facettes-prisme semblent une solution viable pour la modélisation de ressources dans d'autres domaines, par exemple des problèmes dans le contexte de l'enseignement de l'algorithmique.

22. Du fait de notre expérience de l'enseignement de l'informatique, cela nous semble une application utile, mais il faudrait effectuer un état de l'art ou une enquête pour étoffer ce point de vue.

23. Voir par exemple [Dijkstra, 1971].

8.4.2 Plateforme, par exemple MIRTO

Bien évidemment une base de texte indexée pédagogiquement pour l'enseignement des langues pourrait être utilisée dans le domaine de l'Apprentissage des Langues Assisté par Ordinateur (ALAO) : l'intégration à une plateforme d'enseignement des langues, en l'occurrence Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral (MIRTO) est d'ailleurs à l'origine de notre projet. Une utilisation dans MIRTO pourrait être de considérer la sélection comme une fonction d'un script pour une activité (cf. § 7.2 p. 281). Un script pourrait y faire appel pour procéder à la constitution d'une collection de textes en fonction d'un contenu linguistique. L'enseignant fournirait donc au sein du contexte « didactique²⁴ » la requête à effectuer, mais ce serait l'apprenant qui déciderait du texte sur lequel travailler « [rompant ainsi] avec le système qui veut que seul l'enseignant a le droit de choisir les textes qui vont être travaillés en classe²⁵ » [Tagliante, 2006, p. 137].

Dans ce contexte, même pour un usage en présentiel, il serait pertinent de permettre aux enseignants de stocker leurs recherches afin de pouvoir utiliser le résultat comme corpus. La notion de collection se prête bien à ce genre de problématique puisque la donnée de la `conditionDHomogeneite` permet à elle seule de régénérer la collection concernée.

Enfin, à plus long terme, dans une perspective d'hybridation et de communication entre différents systèmes, un système de diagnostic comme ExoGen [Kraif & Ponton, 2007] pourrait, en fonction des erreurs détectées, proposer une requête permettant d'obtenir des textes adéquats pour générer une activité portant sur la notion incriminée.

8.5 Hybridation et ALAO intégratif

Ces exemples prospectifs sont à notre avis représentatifs de l'hybridation, à laquelle nous nous référerions dans notre description de la phase intégrative de l'ALAO (cf. § 2.1.6 p. 44). Ces applications consisteraient, en quelque sorte, à combiner des outils qui analysent, chacun selon un point de vue, la problématique de l'ALAO :

« What is now needed is not so much a single solution to guide Computer Assisted Language Learning (CALL), but rather a careful weighing of the options so that strengths and limitations become evident. This viewpoint grows from a belief that postgraduates, researchers, and teachers increasingly have to relate to a number of parallel perspectives in their work. This requires an appreciation and understanding not of one single viewpoint alone, but instead of a number of different perspectives concerning the nature of language and language learning.²⁶ »

[Levy & Stockwell, 2006, p. 5]

24. Pour plus d'information sur la création d'activité avec MIRTO, voir par exemple [Antoniadis *et al.*, 2005a].

25. « [Les] apprenants peuvent, eux aussi, apporter des textes qui les intéressent ou choisir un texte parmi l'éventail proposé par l'enseignant. » [Tagliante, 2006, p. 137]

26. Ce dont nous avons besoin actuellement, ce n'est pas tant une solution unique pour guider l'ALAO, mais plutôt un examen attentif des options de manière à ce que les points forts et limitations deviennent évidents. Ce point de vue naît de la conviction que, de plus en plus, doctorants, chercheurs et enseignants doivent considérer des perspectives parallèles dans le cadre de leur travail. Cela demande non pas l'appréciation et la compréhension d'un unique point de vue, mais au contraire d'une quantité de perspectives différentes concernant la nature de la langue et de son apprentissage.

Le concept de facette (associée à un prisme) s'inscrit tout à fait dans cette direction générale, en ce sens que chaque prisme permet de considérer le texte selon un point de vue différent. Cette mouvance semble s'accompagner de la prise en compte de plus en plus généralisée des limites des outils. Tout ceci poursuit une certaine logique : conscients de la difficulté de résoudre un problème complexe, plutôt que de prétendre proposer une solution globale, les équipes abordent des problèmes spécifiques (cf. § a) p. 59). En parallèle, d'autres outils sont développés en se focalisant sur la capacité à intégrer le fruit de ces travaux spécifiques (type Système de Gestion de l'Apprentissage (SGA)). Ils ne proposent pas forcément une conception unifiée de la langue ou même son enseignement et visent à offrir ces différents points de vue à travers l'intégration d'outils. S'agit-il d'un avatar de l'ALAO intégratif provenant de la généricité d'outils comme les SGA ou du début de la transition vers une phase différente ? Quoi qu'il en soit, il s'agit, comme souvent dans l'histoire des Environnements Informatiques pour l'Apprentissage Humain (EIAH), d'une direction globale qui n'est pas cantonnée au domaine, comme en témoignent :

- l'importance de l'éclectisme dans les pratiques ou sa reconnaissance dans la réflexion théorique [Puren, 1994] ;
- la position centrale de la possibilité de lier et partager des données (et donc la possibilité d'offrir différents points de vue sur celle-ci) dans la problématique de l'avenir du Web [Berners-Lee, 2009]²⁷.

27. Ou pour un exposé plus formel [Bizer *et al.*, 2009].

Acronymes

ADL	<i>Advanced Distributed Learning</i>
AGLS	<i>Australian Government Locator Service</i>
AICC	<i>Aviation Industry Computer-Based Training (CBT) Committee</i>
AFNOR	Agence Française de NORmalisation
ALAO	Apprentissage des Langues Assisté par Ordinateur
ANOVA	<i>ANalysis Of VAriance</i>
ANR	Agence Nationale de la recherche
API	<i>Application Programming Interface</i>
ARIADNE	<i>Alliance of Remote Instructional Authoring & Distribution Networks for Europe</i>
ATALA	Association pour le Traitement Automatique de la Langue
ATIEF	Association des Technologies de l'Information pour l'Éducation et la Formation
BEP	Brevet d'Études Professionnelles
BNC	<i>British National Corpus</i>
CALL	<i>Computer Assisted Language Learning</i>
CAM	<i>Content Aggregation Model</i>
CAPES	Certificat d'Aptitude au Professorat de l'Enseignement du Second degré
CASLA	<i>Computer Applications in Second Language Acquisition</i>
CBT	<i>Computer-Based Training</i>
CECR	Cadre Européen Commun de Référence pour les Langues
CEN	Comité Européen de Normalisation
CLELIA	Corpus Littéraire et Linguistique assisté par des outils d'Intelligence Artificielle
CMO	Communication Médiatisée par Ordinateur
CNTRL	Centre National de Ressources Textuelles et Lexicales
CP	Contexte Pédagogique
CREA	<i>Corpus de Referencia del Español Actual</i>
CSCL	<i>Computer-Supported Collaborative Learning</i>
CLCS	<i>Centre for Language and Communication Studies</i>
CSS	<i>Cascading Style Sheet</i>
DCES	<i>Dublin Core Element Set</i>
DCMI	<i>Dublin Core Metadata Initiative</i>

DDL	<i>Data Driven Learning</i>
DEA	Diplôme d'Études Approfondies
DoD	<i>Department of Defense</i>
edna	<i>Educational Network of Australia</i>
EIAH	Environnements Informatiques pour l'Apprentissage Humain
EAO	Enseignement Assisté par Ordinateur
EIAO	Enseignement Intelligemment Assisté par Ordinateur
EIAO₂	Environnement Interactif d'Apprentissage avec Ordinateur
ELISA	<i>English Language Interview Corpus as a Second-Language Application</i>
EML	<i>Educational Modelling Language</i>
EPI	Enseignement Public et Informatique
ESP	<i>English for Specific Purposes</i>
EXXELANT	<i>EXample eXtractor Engine for LANguage Teaching</i>
FLE	Français Langue Étrangère
FLM	Français Langue Maternelle
FLS	Français Langue Seconde
FOAD	Formation Ouverte À Distance
GEM	<i>the Getaway to Educational Material</i>
GRACE	Grammaire et Ressources pour les Analyseurs de Corpus et leur Évaluation
HTML	<i>HyperText Mark-up Language</i>
HTTP	<i>HyperText Transfer Protocol</i>
IA	Intelligence Artificielle
ICALI	<i>Intelligent Computer Assisted Language Instruction</i>
IE	<i>Internet Explorer</i>
IEEE	<i>Institute of Electrical and Electronics Engineers, Inc</i>
IHM	Interface Homme-Machine
IMS	<i>Instructional Management Systems</i>
IMS-LD	<i>IMS-Learning Design</i>
INSEE	Institut National de la Statistique et des Études Économiques
ISO	<i>International Organization for Standardization</i>
ISSN	<i>International Standard Serial Number</i>
KWIC	<i>KeyWord In Context</i>
LIDILEM	laboratoire LInguistique et DIidactique des Langues Étrangères et Maternelles
LMS	<i>Learning Management System</i>
LOM	<i>Learning Object Metadata</i>
LSA	<i>Latent Semantic Analysis</i>
LTSC	<i>Learning Technology Standards Committee</i>
LTSO	<i>Learning Technologies Standards Observatory</i>

LV	Langue Vivante
MANOVA	analyse de variance multivariée
MIRTO	Multi-apprentissages Interactifs par des Recherches sur des Textes et l'Oral
NLP	<i>Natural Language Processing</i>
NSF	<i>National Science Foundation</i>
ODF	<i>Open Document Format</i>
ODF₂	Observatoire Du Français et des langues nationales
OOXML	<i>Office Open XML</i>
OUN	<i>Open Universiteit Nederland</i>
PHP	<i>PHP: Hypertext Preprocessor</i>
PIB	Produit Intérieur Brut
PLATO	<i>Programmed Logic for Automated Teaching Operations</i>
PO	Pouvoirs Organismes
PPF	Plan Pluri-Formation
QCM	Questionnaire à Choix Multiples
Rafales	Recueil Automatique Favorisant l'Acquisition d'une Langue Étrangère de Spécialité
RC	<i>Release Candidate</i>
RI	Recherche d'Information
RSS	<i>Really Simple Syndication</i>
RTE	<i>Run Time Environment</i>
RTF	<i>Rich Text Format</i>
SACODEYL	<i>System Aided Compilation and Open Distribution of European Youth Language</i>
SCTIC	Sous-comité sur les technologies de l'information et de la communication
SCO	<i>Sharable Content Object</i>
SCORM	<i>Sharable Content Object Reference Model</i>
SGA	Système de Gestion de l'Apprentissage
SGBD	Système de Gestion de Base de Données
SIB	Sciences de l'Information et des Bibliothèques
SN	<i>Sequencing and Navigation</i>
SRI	Système de Recherche d'Information
TAL	Traitement Automatique des Langues
TEI	<i>Text Encoding Initiative</i>
TICCIT	<i>Time-shared Interactive Computer Controlled Information Television</i>
URL	<i>Unified Resource Locator</i>
UML	<i>Unified Modeling Language</i>
W3C	<i>World Wide Web Consortium</i>
XAIRA	<i>XML Aware Indexing and Retrieval Architecture</i>
XML	<i>eXtensive Mark-up Language</i>

Bibliographie

- [Adamczewski & Gabilan, 1993] Henri ADAMCZEWSKI & Jean-Pierre GABILAN (1993). *Les clés de la grammaire anglaise*. Armand Colin. ISBN : 2-200-01174-1.
- [AFNOR, 1993] AFNOR (1993). Information et documentation - principes généraux pour l'indexation des documents. Norme homologuée Z47-102.
url : <http://tr.im/Z47102>. Cité dans [Bertrand *et al.*, 1996; U-Montréal, 2002].
- [AGLS, 2002] AGLS. AGLS metadata element set - part 1: Reference description v1.3. (2002). ISBN : 0-642-34491-4.
url : http://www.naa.gov.au/Images/AGLS_reference_description_v1-3_tcm2-880.pdf.
- [Ahmad *et al.*, 1985] Khurshid AHMAD, Greville CORBETT, Margaret ROGERS & Roland SUSSEX (1985). *Computers, language learning, and language teaching*. New directions in language teaching. Cambridge University Press. ISBN : 052126569X. Cité dans [Bannert *et al.*, 2003].
- [AICC, 2002] AICC (2002). Digital electronic library system. Document technique ELS001 - Draft 7, AICC.
url : <http://www.aicc.org/docs/working/els001d7.pdf>.
- [AICC, 2006] AICC (2006). Aviation industry metadata description. Document technique DELS002 - 1.0, AICC.
url : <http://www.aicc.org/docs/working/els001d7.pdf>.
- [Alcaraz & Pérez Paredes, 2008] José Maria ALCARAZ & Pascual PÉREZ PAREDES (2008). What do annotators annotate? An analysis of language teachers' corpus pedagogical annotation. *In TALC [2008]*, pages 27–37.
- [Algo-Lev, 2008] The Levenshtein-algorithm – how Levenshtein works... (2008).
url : <http://www.levenshtein.net/>.
- [Allum, 2001] Paul ALLUM (2001). Principles applicable to the production of CALL-ware: learning from the field of Human Computer Interaction (HCI). *ReCALL*, 13(2):146–166. ISSN : 0958-3440.
url : http://journals.cambridge.org/article_S0958344001000222.
- [Alonso, 1994] Encina ALONSO (1994). *¿ cómo ser profesor/a y querer seguir siéndolo ?*. Investigación didáctica. Edelsa. ISBN : 84-7711071-9.
- [Álvarez, 2007] Alfredo ÁLVAREZ ÁLVAREZ (2007). *Estudio de los recursos internet aplicados a la enseñanza y a la traducción del francés*. Thèse de doctorat, Universidad Autónoma de

Madrid.

url : <http://digitool-uam.greendata.es/webclient/DeliveryManager?pid=1785>.

[Ambler, 2005] Scott W. AMBLER (2005). *The elements of uml 2.0 style*. Cambridge University Press. ISBN : 0-521-61678-6.

url : <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521616782>.

[Amer-Yahia *et al.*, 2006] Sihem AMER-YAHIA, Emiran CURTMOLA & Alin DEUTSCH (2006). Flexible and efficient XML search with complex full-text predicates. *In Proceedings of the 2006 ACM SIGMOD international conference on Management of data, Actes de ACM SIGMOD/PODS 2006* (Chicago, IL, 26-29 juin 2006), pages 575–586. ISBN : 1-59593-434-0.

url : <http://portal.acm.org/citation.cfm?id=1142537>.

[Angers, 2000] Maurice ANGERS (2000). *Initiation pratique à la méthodologie des sciences humaines*. Les éditions CEC, 3e édition. ISBN : 2-7617-1673-6.

[Antoniadis, 2004] Georges ANTONIADIS (2004). Les logiciels d'apprentissage des langues peuvent-ils ignorer le TAL? *Les cahiers de l'APLIUT*, 23(2):82–97. ISSN : 0248-9430.

[Antoniadis, 2008] Georges ANTONIADIS (2008). *Du TAL et son apport aux systèmes d'apprentissage des langues : Contributions*. Habilitation à diriger des recherches, Université Stendhal Grenoble 3.

[Antoniadis *et al.*, 2007a] Georges ANTONIADIS, Françoise BOCH, Francis GROSSMANN, Claude PONTON & Thomas LEBARBÉ (2007a). L'apport des corpus annotés à la didactique des langues : l'exemple des guillemets en français. Exposé à : *14e Colloque International de l'Association Grecque de Linguistique Appliquée* (Thessalonique, 14-16 décembre 2007).

[Antoniadis *et al.*, 2004a] Georges ANTONIADIS, Sandra ÉCHINARD, Olivier KRAIF, Thomas LEBARBÉ, Mathieu LOISEAU & Claude PONTON (2004a). NLP-based scripting for CALL activities. *In Erhard Hinrichs Lothar LEMNITZER, éditeur, COLING 2004 eLearning for Computational Linguistics and Computational Linguistics for eLearning*, pages 18–25. COLING. url : <http://hal.archives-ouvertes.fr/hal-00190373/fr/>.

[Antoniadis *et al.*, 2005a] Georges ANTONIADIS, Sandra ÉCHINARD, Olivier KRAIF, Thomas LEBARBÉ, Mathieu LOISEAU & Claude PONTON (2005a). MIRTO: a new approach to CALL systems. *In Jean-Pierre COURTIAT, Costas DAVARAKIS & Thierry VILLEMUR, éditeurs, Technology Enhanced Learning: IFIP TC3 Technology Enhanced Learning Workshop (Tel'04), World Computer Congress, August 22-27, 2004, Toulouse, France*, volume 171 de *IFIP International Federation for Information Processing*, pages 147–156. Springer. ISBN : 0-387-24046-2.

[Antoniadis *et al.*, 2005b] Georges ANTONIADIS, Sandra ÉCHINARD, Olivier KRAIF, Thomas LEBARBÉ & Claude PONTON (2005b). Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO. *ALSIC*, 8(Numéro spécial TALAL):65–79. ISSN : 1286-4986.

url : http://alsic.u-strasbg.fr/v08/antoniadis/alsic_v08_04-rec4.htm.

[Antoniadis *et al.*, 2004b] Georges ANTONIADIS, Sandra ÉCHINARD, Olivier KRAIF, Thomas LEBARBÉ, Claude PONTON & Mathieu LOISEAU (2004b). CALL: from current problems to NLP solutions. *In Proceedings of EUROCALL*.

- [Antoniadis & Ponton, 2002] Georges ANTONIADIS & Claude PONTON (2002). Le TAL : une nouvelle voie pour l'apprentissage des langues. Exposé à : *UNTELE* (Compiègne, 28-30 mars 2002). UTC.
url : http://www.utc.fr/~untele/abst_2002/antoniad.html.
- [Antoniadis & Ponton, 2004] Georges ANTONIADIS & Claude PONTON (2004). MIRTO : un système au service de l'enseignement des langues. Actes de *UNTELE* (Compiègne, 16-20 mars 2004). UTC.
url : <http://www.utc.fr/~untele/abst04/anton04.htm>.
- [Antoniadis *et al.*, 2007b] Georges ANTONIADIS, Claude PONTON & Virginie ZAMPA (2007b). De la nécessité du TAL dans les EIAH de langues : les cas EXXELANT et MIRTO. In *Actes de la conférence EIAH 2007*, pages 251–256.
url : <http://hal.archives-ouvertes.fr/docs/00/16/15/56/PDF/28.pdf>.
- [Aston, 1997] Guy ASTON (1997). Small and large corpora in language learning. In Barbara LEWANDOWSKA-TOMASZCZYK & Patrick James MELIA, éditeurs, *PALC 97: Practical applications in language corpora*, pages 51–62. Łódź University Press.
url : <http://home2.sslmit.unibo.it/~guy/wudj1.htm>.
- [Aston, 2002] Guy ASTON (2002). The learner as corpus designer. In Kettemann & Marko [2002], pages 9–25.
url : <http://www.sslmit.unibo.it/~guy/graz.htm>.
- [Augustyn *et al.*, 2008] Magdalena AUGUSTYN, Sabrina Ben HAMOU, Gwendoline BLOQUET, Vannina GOOSSENS, Mathieu LOISEAU & Fanny RINCK (2008). Constitution de ressources pédagogiques numériques : le lexique des affects. In Mathieu LOISEAU, Myriam ABOUZAÏD, Laurence BUSON, Cristelle CAVALLA, Ali DJAROUN, Céline DUGUA, Anna GHIMENTON, Vannina GOOSSENS, Thomas LEBARBÉ, Aurélie NARDY, Fanny RINCK & Christian SURCOUF, éditeurs, *Autour des langues et du langage – perspective pluridisciplinaire*, pages 407–414. Presses Universitaires de Grenoble. ISBN : 978-2-7061-1427-4.
- [Balatsoukas *et al.*, 2008] Panos BALATSOUKAS, Anne MORRIS & Ann O'BRIEN (2008). Learning objects update: Review and critical approach to content aggregation. *Journal of Educational Technology & Society*, 11(2):119–130.
url : http://www.ifets.info/journals/11_2/11.pdf.
- [Bangs, 2002] Paul BANGS (2002). Authoring, pedagogy and the web: Expectations versus reality. In Pascual PÉREZ PAREDES & Pascual CANTOS GÓMEZ, éditeurs, *New Trends in Computer Assisted Language Learning and Teaching, International Journal of English Studies*, 2(1):19–30. ISSN : 1578-7044.
url : <http://www.um.es/ijes/vol2n1/03-PaulBangs.pdf>.
- [Bangs, 2003] Paul BANGS (2003). *Engaging the learner—how to author for best feedback*. Language learning and language technology. Swets & Zweitlinger. ISBN : 9026519486.
url : <http://books.google.fr/books?id=F111uAtL7HkC>.
- [Bangs & Cantos Gómez, 2004] Paul BANGS & Pascual CANTOS GÓMEZ (2004). What can Computer Assisted Language Learning contribute to foreign language pedagogy? *International Journal of English Studies*, 4(1):221–239. ISSN : 1578-7044.
url : <http://www.um.es/ijes/vol4n1/11-Bangs%20%20Cantos.pdf>.

- [Bannert *et al.*, 2003] Robert BANNERT, Antonis BOTINIS, Barbara GAWRONSKA & Gerd HOLLENSTEIN (2003). Requirements on infrastructure for CALL (Computer Assisted Language Learning). In *Fifth International Conference on Advances in Infrastructure for e-Business, e-Education, e-Science, e-Medicine on the Internet*.
url : <http://kondor.etf.bg.ac.yu/~vm/cd1/papers/107.pdf>.
- [Bar-Hillel, 1964] Yehoshua BAR-HILLEL (1964). The future of machine translation. In Hartley ROGERS, éditeur, *Language and information – selected essays on their theory and application*, Logic, pages 180–184. Addison-Wesley. Cité dans [Kraif *et al.*, 2004; Macklovitch, 1995].
- [Bax, 2003] Stephen BAX (2003). CALL—past, present and future. *System*, 31(1):13–28. ISSN : 0346-251X.
url : [http://dx.doi.org/10.1016/S0346-251X\(02\)00071-4](http://dx.doi.org/10.1016/S0346-251X(02)00071-4).
- [Beacco, 1995] Jean-Claude BEACCO (1995). La méthode circulante et les méthodologies constituées. *Le français dans le monde*, Recherche et applications (numéro spécial):42–49. ISSN : 0015-9395.
- [Benjamin & Blunt, 1992] Robert I. BENJAMIN & Jon BLUNT (1992). Critical IT issues: The next ten years. *Sloan Management Review*, 33(4):7–19. ISSN : 0019-848X.
- [Bérard, 1991] Évelyne BÉRARD (1991). *L’approche communicative – théorie et pratiques*. Didactique des langues étrangères. CLE International. ISBN : 2-19-033352-0.
- [Bernard, 1995] Michel BERNARD (1995). Lire l’hypertexte. In Alain VUILLEMIN & Michel LENOBLE, éditeurs, *Littérature et informatique : la littérature générée par ordinateur*, Actes de *Journées d’études internationales* (Paris, 20-22 avril 1994), pages 313–325. Artois Presses Université. ISBN : 2-910663-02-7.
- [Berners-Lee, 1989] Tim BERNERS-LEE (1989). Information management: a proposal. Proposition de projet, CERN.
url : <http://www.w3.org/History/1989/proposal.html>.
- [Berners-Lee, 1997] Tim BERNERS-LEE (1997). Metadata architecture.
url : <http://www.w3.org/DesignIssues/Metadata.html>.
- [Berners-Lee, 2009] Tim BERNERS-LEE (2009). On the next Web. Exposé à : *Technology Entertainment Design* (Long Beach, 3-7 février 2009).
url : http://www.ted.com/talks/lang/fre_fr/tim_berniers_lee_on_the_next_web.html.
- [Bertrand *et al.*, 1996] Annick BERTRAND, Jean-Marie CELLIER & Luc GIROUX (1996). Expertise and strategies for the identification of the main ideas in document indexing. *Applied Cognitive Psychology*, 10(5):419–433. ISSN : 0888-4080.
url : <http://www3.interscience.wiley.com/journal/21437/abstract>.
- [Besse, 1987] Henri BESSE (1987). Documents authentiques et enseignement / apprentissage de la grammaire d’une langue étrangère. In Jacques CORTÈS, éditeur, *Une Introduction à la recherche scientifique en didactique des langues*, pages 181–214. Didier. ISBN : 2-278-03747-1.
- [Biber, 1986] Douglas BIBER (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62(2):384–414. ISSN : 0097-8507.
url : <http://www.jstor.org/stable/414678>.

- [Binon *et al.*, 1998] Jean BINON, Catherine GEYSKENS & Karen LAHOUSSE (1998). Les stratégies de lecture dans quelques manuels de Français Langue Étrangère ou Seconde (FLES) en flamand. *Romanesque*, (984):17–31. ISSN : 0772-6406.
url : <http://www.vlrom.be/pdf/984leesvaa1.pdf>.
- [Bizer *et al.*, 2009] Christian BIZER, Tom HEATH & Tim BERNERS-LEE (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*. À paraître. ISSN : 1552-6283.
url : <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.
- [Blanchard *et al.*, 2009] Alexia BLANCHARD, Olivier KRAIF & Claude PONTON (2009). Mastering noise and silence in learner answers processing: simple techniques for analysis and diagnosis. *CALICO Journal*. ISSN : 0742-7778.
url : <http://tr.im/calicoabokcp>.
- [Boch & Grossmann, 2001] Françoise BOCH & Francis GROSSMANN (2001). De l'usage des citations dans le discours théorique : des constats aux propositions didactiques. In Françoise BOCH & Francis GROSSMANN, éditeurs, *Apprendre à citer le discours d'autrui, Lidil*, (24):91–111. ISSN : 1146-6480.
- [Bouchard, 1990] Robert BOUCHARD (1990). *Interaction et discursivité*. Thèse de doctorat, Université Stendhal Grenoble 3.
- [Bouillon *et al.*, 1998] Pierrette BOUILLON, Françoise VANDOOREN, Lyne DA SILVA, Laurence JACQMIN, Sabine LEHMANN, Graham RUSSELL & Evelyne VIEGAS (1998). *Traitement automatique des langues naturelles*. Champs linguistiques. Duculot. ISBN : 2-8011-1181-3.
- [Boukottaya *et al.*, 2009] Aida BOUKOTTAYA, Bernadette CHARLIER & Christine VANOIRBEEK (2009). A document reuse tool for communities of practice.
url : <http://en.scientificcommons.org/48838438>.
- [Boulidoires, 1997] Alain BOULDOIRES (1997). Multimédia et enseignement : la problématique de l'usage. *Revue de l'EPI*, (88):125–131. ISSN : 1254-3985.
url : <http://archive-edutice.ccsd.cnrs.fr/docs/00/03/09/10/PDF/ba4p125.pdf>.
- [Boulidoires, 1998] Alain BOULDOIRES (1998). L'usage du multimedia dans l'enseignement. *Revue de l'EPI*, (91):101–108. ISSN : 1254-3985.
url : <http://archive-edutice.ccsd.cnrs.fr/edutice-00001167/en/>.
- [Boulton, 2008] Alex BOULTON (2008). DDL: reaching the parts other teaching can't reach? In TALC [2008], pages 38–44.
- [Boulton, 2007] Alex BOULTON (2007). Esprit de corpus : Promouvoir l'exploitation de corpus en apprentissage des langues. In Geoffrey WILLIAMS, éditeur, *Actes des cinquièmes Journées de la linguistique de Corpus, Texte et corpus*, (3):37–46. ISSN : 1958-5306.
url : <http://hal.archives-ouvertes.fr/hal-00326988/fr/>.
- [Bourda, 2002] Yolaine BOURDA (2002). Des objets pédagogiques aux dossiers pédagogiques (via l'indexation). *Document numérique*, 6(1-2):115–128. ISSN : 1279-5127.
url : <http://www.cairn.info/revue-document-numerique-2002-1-page-115.htm>.

- [Bourdieu, 2002] Pierre BOURDIEU (2002). L'opinion publique n'existe pas. Exposé à Noroît (Arras) (janvier 1972). Cf. *Questions de sociologie*, pages 222–235. Les éditions de minuit. ISBN : 2-7073-1825-6.
- [Bourguignon, 2007] Claire BOURGUIGNON (2007). Apprendre et enseigner les langues dans la perspective actionnelle : le scénario d'apprentissage-action. *Les langues modernes*.
url : <http://www.aplv-languesmodernes.org/spip.php?article865>.
- [Braun, 2005] Sabine BRAUN (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1):47–64. ISSN : 0958-3440.
url : <http://dx.doi.org/10.1017/S0958344005000510>.
- [Braun, 2007] Sabine BRAUN (2007). Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora. *ReCALL*, 19(03):307–328. ISSN : 0958-3440.
url : http://journals.cambridge.org/article_S0958344007000535.
- [Braun et al., 2004] Sabine BRAUN, Kurt KOHN, Caroline LENZ, Gregor SIEBER & Hannes WIDMANN. Demo de ELISA. (2004).
url : http://www.uni-tuebingen.de/elisa/html/elisa_index.html.
- [Breen, 1985] Michael P. BREEN (1985). Authenticity in the language classroom. *Applied Linguistics*, 6(1):60–70. ISSN : 0142-6001.
url : <http://applij.oxfordjournals.org/cgi/reprint/6/1/60.pdf>. Cité dans [Taylor, 1994; Tatsuki, 2006].
- [Brett, 1994] Paul BRETT (1994). Using text reconstruction software. *ELT J*, 48(4):329–336. ISSN : 0951-0893.
url : <http://eltj.oxfordjournals.org/cgi/content/abstract/48/4/329>. Cité dans [Levy, 1997].
- [Brodine, 2001] Ruey BRODINE (2001). Integrating corpus work into an academic reading course. In Guy ASTON, éditeur, *Learning with corpora*, pages 138–176. Athelstan. ISBN : 0940753162.
- [Bronckart & Plazaola Giger, 1996] Jean-Paul BRONCKART & Itziar PLAZAOLA GIGER (1996). Théories des actes de langage et enseignement de la L2. Un exemple de transposition didactique. *Diálogos hispánicos de Amsterdam*, (18):13–36. ISBN : 90-5183-945-6, ISSN : 0929-2217.
- [Broudoux & Chartron, 2008] Évelyne BROUDOUX & Ghislaine CHARTRON, éditeurs (2008). *Traitements et pratiques documentaires – vers un changement de paradigme ?*, actes de la 2^e conférence “Document numérique et société” (Paris, 17-18 novembre 2008). ADBS. ISBN : 978-2-84365-116-8, ISSN : 1762-8288.
- [Bru, 2002] Marc BRU (2002). Pratiques enseignantes : des recherches à conforter et à développer. *Revue Française de Pédagogie*, (138):63–73. ISSN : 0556-7807.
url : http://www.inrp.fr/edition-electronique/archives/revue-francaise-de-pedagogie/web/fascicule.php?num_fas=393.
- [Bru, 2004] Marc BRU (2004). *Les pratiques enseignantes hors de la classe*, chapitre Les pratiques enseignantes comme objet de recherche, pages 281–299. Savoir et formation. L'Harmattan. ISBN : 2-7475-6120-8.

- [Bruillard, 1997] Éric BRUILLARD (1997). *Les machines à enseigner*. Hermès. ISBN : 2-86601-610-6.
url : http://www.stef.ens-cachan.fr/annur/bruillard/mae_somr.htm.
- [Bruillard & de La Passadière, 2003] Éric BRUILLARD & Brigitte DE LA PASSADIÈRE, éditeurs (2003). *Ressources numériques, xml et éducation*. Numéro Hors série in Sciences et Techniques Educatives. Lavoisier. ISBN : 2-7462-0682-X.
- [Burden, 1981] Paul R. BURDEN (1981). Teachers' perceptions of their personal and professional development. Exposé à : *Annual meeting of the Midwestern Educational Research Association* (Des Moines, Iowa, 20 novembre 1981).
url : <http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?accno=ED210258>.
- [Burnard, 2001] Lou BURNARD (2001). XML+TEI, un mariage fait aux cieux. Exposé à MSH de Lyon.
url : <http://www.tei-c.org/Talks/lyon-3.ppt>.
- [Bush, 1945] Vannevar BUSH (1945). As we may think. *The Atlantic Monthly*.
url : <http://www.theatlantic.com/doc/194507/bush>.
- [Calzada Pérez, 2008] María CALZADA PÉREZ (2008). The ECPC archive: a gateway to the merging of corpus based translation studies and critical discourse analysis. Exposé à : *8th Teaching and Language Corpora Conference (TaLC)* (Lisbonne, 4-6 juillet 2008). Cf. TALC [2008]. ISBN : 978-989-95523-1-9.
- [Canale & Swain, 1980] Michael CANALE & Merrill SWAIN (1980). Communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1):1-47. ISSN : 0142-6001.
url : <https://segue.atlas.uiuc.edu/uploads/nppm/CanaleSwain.80.pdf>.
- [Carol *et al.*, 2004] Rita CAROL, Peter GRIGGS & Pierre BANGE (2004). Les exercices dans l'apprentissage de langue étrangère. Critique de l'approche communicative. Exposé à : *La didactique des langues face aux cultures linguistiques et éducatives* (Paris, 12-14 décembre 2002). Marges Linguistiques.
url : http://www.revue-texto.net/1996-2007/marges/marges/Documents%20Site%206/doc0199_colloque_paris/58_griggs_al.pdf.
- [Carroll, 1965] John B. CARROLL (1965). The contributions of psychological theory and educational research to the teaching of foreign languages. *The Modern Language Journal*, 49(5):273-281. ISSN : 0026-7902.
url : <http://www.jstor.org/stable/322133>.
- [CEN, 2008] Comité Européen de Normalisation / Information Society Standardization System. Learning technologies standards observatory - observatory contents. (2008).
url : <http://www.cen-ltso.net/Descargas/book/CEN-LTSO-eng.pdf>.
- [CGal66, 2007] CONSEIL GÉNÉRAL DES PYRÉNÉES ORIENTALES (2007). Charte en faveur du catalan. Charte.
url : http://www.cg66.fr/culture/patrimoine_catalanite/catalanite/charte.html.
- [Chambers, 2007a] Angela CHAMBERS (2007a). Integrating corpora in language learning and teaching. *ReCALL*, 19(03):249-251. ISSN : 0958-3440.
url : http://journals.cambridge.org/article_S0958344007000134.

- [Chambers, 2007b] Angela CHAMBERS (2007b). Popularising corpus consultation by language learners and teachers. In Encarnación HIDALGO TENORIO, Luis QUEREDA RODRÍGUEZ-NAVARRO & Juan SANTANA, éditeurs, *Corpora in the Foreign Language Classroom*, Actes de *Sixth International Conference on Teaching and Language Corpora (TaLC 6)* (Université de Grenade (Espagne), 4-7 juillet 2004), volume 61 de *Language and Computers – Studies in Practical Linguistics*, pages 3–16. Rodopi. ISBN : 90-420-2142-X.
- [Chanier, 1995] Thierry CHANIER (1995). *Acquisition des langues assistée par ordinateur (ALAO)*. Habilitation à diriger des recherches, Université Blaise Pascal - Clermont-Ferrand II.
url : <http://tel.archives-ouvertes.fr/tel-00180340/fr/>.
- [Chappelle, 2001] Carol A. CHAPPELLE (2001). *Computer applications in Second Language Acquisition: Foundations for teaching, testing and research*. Cambridge University Press. ISBN : 0-521-62646-3.
- [Chappelle, 2004] Carol A. CHAPPELLE (2004). Technology and second language learning: expanding methods and agendas. *System*, 32(4):593–601. Incorporating Multimedia Capability in the Reporting of Applied Linguistics Research. ISSN : 0346-251X.
url : <http://www.sciencedirect.com/science/article/B6VCH-4DTKX01-1/2/3484fed0afde3071c694b19fcbf0a74a>.
- [Charlier, 1989] Évelyne CHARLIER (1989). *Planifier un cours, c'est prendre des décisions*. Pédagogies en développement. Série 5, Nouvelles pratiques de formation. De Boeck Université. ISBN : 2-8041-1271-3.
- [Charlier & Charlier, 1998] Évelyne CHARLIER & Bernadette CHARLIER (1998). *La formation au cœur de la pratique : analyse d'une formation continuée d'enseignants*. Pratiques pédagogiques. Série 6, Pratiques méthodologiques. De Boeck Université. ISBN : 2-8041-2816-4, ISSN : 0778-0451.
- [Chastain & Woerdehoff, 1968] Kenneth D. CHASTAIN & Frank J. WOERDEHOFF (1968). A methodological study comparing the audio-lingual habit theory and the cognitive code-learning theory. *The Modern Language Journal*, 52(5):268–279. ISSN : 0026-7902.
url : <http://www.jstor.org/stable/322883>.
- [Chatzissavidis, 2003] Sofronis CHATZISSAVIDIS (2003). Enseignement précoce des langues vivantes : à quelles conditions ; quels résultats? Actes de *Promouvoir l'apprentissage des langues européennes moins enseignées : Politiques et Méthodologies* (Dijon, 7-9 avril 2003).
- [Chevallard, 1982] Yves CHEVALLARD (1982). Pourquoi la transposition didactique? In *Actes de l'année 1981-1982*, Actes de *Séminaire de didactique et de pédagogie des mathématiques de l'IMAG*, pages 167–194. IMAG.
url : http://yves.chevallard.free.fr/spip/spip/IMG/pdf/Pourquoi_la_transposition_didactique.pdf.
- [Chomsky, 1959] Noam CHOMSKY (1959). A review of B.F. Skinner's verbal behavior. *Language*, 35(1):26–58. ISSN : 0097-8507.
url : <http://cogprints.org/1148/0/chomsky.htm>.

- [Clanet, 2001] Joël CLANET (2001). Étude des organisateurs des pratiques enseignantes à l'université. *Revue des sciences de l'éducation*, 27(2):327–352. ISSN : 0318-479X.
url : <http://id.erudit.org/iderudit/009936ar>.
- [Clanet, 2007] Joël CLANET (2007). Éléments organisateurs de séances de lecture en cours préparatoire. *Repères*, (64):211–230. ISBN : 978-2-7342-1093-1, ISSN : 0755-7817.
url : http://visa.inrp.fr/visa/presentation/Seminaires/Journees_inaugurales/Clanet.pdf.
- [Clark & Yinger, 1979] Christopher M. CLARK & Robert J. YINGER (1979). The hidden world of teaching: implications of research on teacher planning. Exposé à : *Annual meeting of the American Educational Research Association* (San Francisco, CA, 8-12 avril 1979). Numéro 55 in Research Series. Michigan State University, Institute for Research on Teaching.
url : <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED175855>.
- [Clark & Yinger, 1980] Christopher M. CLARK & Robert J. YINGER (1980). The hidden world of teaching: implications of research on teacher planning. Exposé à : *Annual meeting of the American Educational Research Association* (Boston, MA, mai). Numéro 77 in Research Series. Michigan State University, Institute for Research on Teaching.
url : <http://eric.ed.gov/ERICWebPortal/detail?accno=ED191844>.
- [CNTRL, 2008] CNTRL (2008). Morphalou 2.0. Documentation, CNTRL.
url : <http://www.cnrtl.fr/lexiques/morphalou/LMF-Morphalou.php>.
- [Cochran, 1952] William G. COCHRAN (1952). The χ^2 test of goodness of fit. *The Annals of Mathematical Statistics*, 23(3):315–345. ISSN : 0003-4851.
url : <http://www.jstor.org/stable/2236678>.
- [Cockburn, 1997a] Alistair A.R. COCKBURN (1997a). Goals and use cases. *Journal of Object Oriented Programming*, 10(5):35–40. ISSN : 0896-8438.
- [Cockburn, 1997b] Alistair A.R. COCKBURN. Structuring use cases with goals. (1997).
url : <http://alistair.cockburn.us/Structuring+use+cases+with+goals>.
- [Cockburn, 1997c] Alistair A.R. COCKBURN (1997c). Using goal-based use cases. *Journal of Object Oriented Programming*, 10(7):56–62. ISSN : 0896-8438.
- [Cole & Foster, 2008] Jason COLE & Helen FOSTER (2008). *Using moodle - teaching with the popular open source course management system*. Community Press. O'Reilly, 2e édition. ISBN : 0-596-52918-X.
url : http://docs.moodle.org/en/Using_Moodle_book.
- [Conseil de l'Europe, 2000] Conseil de la coopération culturelle / Comité de l'éducation / Division des langues vivantes, éditeur (2000). *Cadre européen commun de référence pour les langues*. Conseil de l'Europe ; Didier. ISBN : 227805075-3.
url : <http://medias.didierfle.com/media/contenuNumerique/007/4140016745.pdf>.
- [Contamines et al., 2003] Julien CONTAMINES, Sebastien GEORGE & Richard HOTTE (2003). Approche instrumentale des banques de ressources éducatives. In Bruillard & de La Passadière [2003], pages 157–178. ISBN : 2-7462-0682-X.
url : http://hal.archives-ouvertes.fr/hal-00298189_v1/.

- [Cros, 1965] Louis CROS (1965). Intervention lors de la réunion du 19 mai 1965 au comité éducation de la commission française pour l'Unesco. Document 208.882, INRDP. Cité dans [Bruillard, 1997].
- [Cuq, 1991] Jean-Pierre CUQ (1991). *Le français langue seconde - origines d'une notion et implications didactiques*. références. Hachette.
- [Cuq, 2003] Jean-Pierre CUQ, éditeur (2003). *Dictionnaire de didactique du français langue étrangère et seconde*. CLE International. ISBN : 209-033972-1.
- [Dabène, 1975] Louise DABÈNE (1975). L'enseignement de l'espagnol aux francophones (pour une didactique des langues « voisines »). *Langages*, 9(39):51–64. ISSN : 1958-9549.
url : http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726x_1975_num_9_39_2291.
- [Dabène, 1989] Louise DABÈNE (1989). Exposé introductif. *In Actes du premier colloque de l'A.C.E.D.L.E.*, pages 4–8. École Normale Supérieure de Saint-Cloud. Cité dans [Rézeau, 2001].
- [Dabène, 1994] Louise DABÈNE (1994). Le projet européen GALATEA : pour une didactique de l'intercompréhension en langues romanes. *In* Jeanine STOLIDI, éditeur, *Recherches en linguistique hispanique, Études Hispaniques*, (22):41–45. ISSN : 0246-5523.
url : http://www.galanet.be/publication/fichiers/L_Dabene1994_EH22.pdf.
- [Dabène, 1995] Louise DABÈNE (1995). Apprendre à comprendre une langue voisine, quelles conceptions curriculaires? *Études de Linguistique Appliquée*, (98):103–112. ISSN : 0071-190X.
url : http://www.galanet.be/publication/fichiers/L_Dabene1995_ELA98.pdf.
- [Dargirolle, 1999] Françoise DARGIROLLE (1999). L'évolution de la conception de l'observation de classes en didactique des langues étrangères. *Études de Linguistique Appliquée*, 4(114):141–152. ISSN : 0071-190X.
- [DCMI, 2000] DCMI. Memorandum of understanding between the Dublin Core Metadata Initiative and the IEEE Learning Technology Standards Committee. (2000).
url : <http://dublincore.org/documents/2000/12/06/dcmi-ieee-mou/>.
- [DCMI, 2003] DCMI Usage Board. DCMI grammatical principles. (2003).
url : <http://dublincore.org/usage/documents/principles/>.
- [DCMI, 2005] DCMI (2005). Marc relator terms and dublin core. Document technique, Dublin Core Metadata Initiative.
url : <http://dublincore.org/usage/documents/relators/>.
- [DCMI, 2008] DCMI Usage Board. Dublin core element set. (2008).
url : <http://dublincore.org/documents/dces/>.
- [de Montmollin, 1975] Maurice DE MONTMOLLIN (1975). *L'enseignement programmé*. Numéro 1171 *in* Que sais-je? Presses Universitaires de France, 4e édition mise à jour.
- [de Nuchèze, 1998] Violaine DE NUCHÈZE (1998). Approche pragmatico-énonciative du discours de recherche (à l'usage des apprenants-chercheurs). *In* Michel DABÈNE & Yves REUTER, éditeurs, *Pratique de l'écrit et modes d'accès au savoir dans l'enseignement supérieur, Lidil*, (17):25–40. ISSN : 1146-6480.

- [de Singly, 1992] François DE SINGLY (1992). *L'enquête et ses méthodes : le questionnaire*. sociologie 128. Nathan Université. ISBN : 209-190-567-4.
- [de Vries, 2001] Erica DE VRIES (2001). Les logiciels d'apprentissage : panoplie ou éventail? *Revue Française de Pédagogie*, (137):105–116. ISSN : 0556-7807.
url : http://www.inrp.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP_RF137_10.pdf.
- [de Vries & Baillé, 2006] Erica DE VRIES & Jacques BAILLÉ (2006). Apprentissage : référents théoriques pour les EIAH. In Monique GRANDBASTIEN & Jean-Marc LABAT, éditeurs, *Environnements informatiques pour l'apprentissage humain*, pages 27–45. Hermès et Lavoisier. ISBN : 2-7462-1171-8.
- [Decoo, 1984] Wilfried DECOO (1984). An application of didactic criteria to courseware evaluation. *CALICO Journal*, 2(2):42–46. ISSN : 0742-7778.
url : <https://www.calico.org/a-246-An%20Application%20of%20Didactic%20Criteria%20to%20Courseware%20Evaluation.html>.
- [Delcloque, 2000] Philippe DELCLOQUE, éditeur. History of CALL. (2000).
url : http://www.eurocall-languages.org/resources/history_of_call.pdf.
- [Delcloque & Bramoullé, 2001] Philippe DELCLOQUE & Alexandre BRAMOULLÉ (2001). DISSEMINATE, an initial implementation proposal: a new point of departure in CALL for the 'year 01' ? *ReCALL*, 13(2):277–292. ISSN : 0958-3440.
url : http://journals.cambridge.org/article_S0958344001001124.
- [Demaizière & Dubuisson, 1992] Françoise DEMAIZIÈRE & Colette DUBUISSON (1992). *De l'EAO aux NTF : Utiliser l'ordinateur pour la formation*. Numéro 8 in Autoformation et Enseignement Multimedia. Ophrys. ISBN : 2-7080-0674-6.
- [Depover *et al.*, 2000] Christian DEPOVER, Jean-Jacques QUINTIN & Bruno De LIÈVRE (2000). La conception des environnements d'apprentissage : de la théorie à la pratique/de la pratique à la théorie. *ALSIC*, 3(1):3–18. ISSN : 1286-4986.
url : http://alsic.u-strasbg.fr/Num5/depover/alsic_n05-rec4.htm.
- [Dijkstra, 1971] Edsger W. DIJKSTRA (1971). *A short introduction to the art of programming*. EWD316.
url : <http://www.cs.utexas.edu/~EWD/transcriptions/EWD03xx/EWD316.html>.
- [Draelants & Maroy, 2002] Hugues DRAELANTS & Christian MAROY (2002). Analyse de l'évolution des modes de régulation institutionnalisée dans le système éducatif de la communauté française de Belgique. Rapport de l'étape 1 de : *European project « changes in regulation modes and social production of inequalities in educational systems : a european comparison »*, GIRSEF – Université catholique de Louvain.
url : <http://www.girsef.ucl.ac.be/Belgique2.pdf>.
- [Duval, 2004] Erik DUVAL (2004). Learning technology standardization: making sense of it all. *ComSIS*, 1(1):33–43. ISSN : 1820-0214.
url : <http://www.comsis.org/ComSIS/Volume01/InvitedPapers/ErikDuval.htm>.
- [Duval *et al.*, 2001] Erik DUVAL, Eddy FORTE, Kris CARDINAELS, Bart VERHOEVEN, Rafael Van DURM, Koen HENDRIKX, Maria Wentland FORTE, Norbert EBEL, Maciej MACOWICZ, Ken

- WARKENTYNE & Florence HAENNI (2001). The ariadne knowledge pool system. *Commun. ACM*, 44(5):72–78. ISSN : 0001-0782.
url : <http://doi.acm.org/10.1145/374308.374346>.
- [Duval *et al.*, 2002] Erik DUVAL, Wayne HODGINS, Stuart SUTTON & Stuart L. WEIBEL (2002). Metadata principles and practicalities. *D-Lib Magazine*, 8(4). ISSN : 1082-9873.
url : <http://www.dlib.org/dlib/april02/weibel/04weibel.html>.
- [Duval *et al.*, 2000] Erik DUVAL, Erwin VERVAET, Bart VERHOEVEN, Koen HENDRIKX, Kris CARDINAELS, Henk OLIVIE, Eddy FORTE, Florence HAENNI, Ken WARKENTYNE, Maria Wentland FORTE & Fabian SIMILLION (2000). Managing digital educational resources with the ariadne metadata system. In Jane GREENBERG, éditeur, *Metadata and Organizing Educational Resources on the Internet*, pages 145–171. Haworth Press. ISBN : 0789011786.
- [Ebbrell, 1998] Dawn EBBRELL. Hot potatoes (review). (1998).
url : <http://www.fredriley.org.uk/call/reviews/hotpot.htm>.
- [edna, 2002] EdNA. Edna metadata standard v1.1. (2002).
url : http://www.edna.edu.au/edna/webdav/site/myjahiasite/shared/edna_metadata.pdf.
- [edna, 2006] edna. edna resources - metadata application profile. (2006).
url : http://www.edna.edu.au/edna/webdav/site/myjahiasite/shared/edna_resources_metadata_1.0.pdf.
- [Ellis *et al.*, 1991] Clarence A. ELLIS, Simon J. GIBBS & Gail REIN (1991). Groupware: some issues and experiences. *Commun. ACM*, 34(1):39–58. ISSN : 0001-0782.
- [Englebart, 1970] Douglas ENGLEBART (1970). X-Y position indicator for a display system. Brevet 3541541, United States Patent Office.
url : <http://www.pat2pdf.org/pat2pdf/foo.pl?number=3541541>.
- [Evans & Lefferts, 1995] David A. EVANS & Robert G. LEFFERTS (1995). Clarit-trec experiments. *Information Processing & Management*, 31(3):385–395. The Second Text Retrieval Conference (TREC-2). ISSN : 0306-4573.
url : <http://www.sciencedirect.com/science/article/B6VC8-3YMFVB3-18/2/e60d87d54606ce4e40b05957b9ecce42>.
- [Everett, 2007] Dan EVERETT (2007). Cultural constraints on grammar in PIRAHÃ: a reply to Nevins, Pesetsky, and Rodrigues (2007).
url : <http://ling.auf.net/lingBuzz/000427>.
- [Fairon, 2006] Cédric FAIRON (2006). Corporator: a tool for creating RSS-based specialized corpora. In *Proceedings of the Workshop Web as corpus. EACL 2006*.
url : <http://www.uclouvain.be/9002.html?publication=393186&format=none&Envoi=1>.
- [Feynman, 2006] Richard Phillips FEYNMAN (2006). *Leçons sur l'informatique*. Odile Jacob. ISBN : 2-7381-1415-6.
- [Flowerdew, 1993] John FLOWERDEW (1993). Concordancing in language learning. *Perspectives*, 5(Autumn):87–101.
url : <http://sunzi1.lib.hku.hk/hkjo/view/10/1000064.pdf>.

- [Francis, 1967] Winthrop Nelson FRANCIS (1967). The Brown University standard corpus of English: some implications for TESOL. In Betty Wallace ROBINETT, éditeur, *On Teaching English to Speakers of other Languages*, Actes de *TESOL Conference, New York City, March 17-19, 1966*, Series III, pages 131-137. Georgetown University.
url : <http://eric.ed.gov/ERICWebPortal/detail?accno=ED032525>.
- [Friesen *et al.*, 2003] Norm FRIESEN, Sue FISHER, Susan HESEMEIER & Anthony ROBERTS. Lignes directrices de cancore version 2.0 : Introduction. (2003).
url : http://www.cancore.ca/lignes_directrices/00a_Introduction_2.0.pdf.
- [Fuchs *et al.*, 1993] Catherine FUCHS, Laurence DANLOS, Anne LACHERET-DUJOUR, Daniel LUZATTI & Bernard VICTORI (1993). *Linguistique et traitements automatiques des langues*. HU Linguistique. Hachette. ISBN : 2-01-016908-5.
- [Fuller, 1969] Frances F. FULLER (1969). Concerns of teachers : a developmental conceptualization. *American Educational Research Journal*, 6(2):207-226. ISSN : 0002-8312.
url : <http://www.jstor.org/stable/1161894>. Cité dans [Lamarre, 2003; Huberman, 1989a].
- [Gamper & Knapp, 2002] Johann GAMPER & Judith KNAPP (2002). A review of intelligent CALL systems. *Computer Assisted Language Learning*, 15(4):329 - 342. ISSN : 0958-8221.
url : <http://www.eurac.edu/NR/rdonlyres/3CA3C4BB-664D-4B41-BF55-98A5ED0BEEF8/0/icallExtended.pdf>.
- [GEM, 2001] GEM. Listing of GEM 1.0 elements and namespaces. (2001).
url : <http://www.thegateway.org/about/documentation/elementsArchive/gem-1-elements/>.
- [GEM, 2004a] GEM. Application profiles for GEM 2.0 including GEM (in general), gateway lite and gateway full. (2004).
url : <http://www.thegateway.org/about/documentation/ApplicationProfiles-9-8-04/>.
- [GEM, 2004b] GEM. instructionalmethod. (2004).
url : <http://www.thegateway.org/about/documentation/metadataElements/documentation/metadataElements/instructionalMethodRefinements/instructionalMethod-11-11-04>.
- [GEM, 2004c] GEM. Listing of GEM 2.0 top-level elements. (2004).
url : <http://www.thegateway.org/about/documentation/metadataElements/index.html>.
- [Germain, 1993] Claude GERMAIN (1993). *Évolution de l'enseignement des langues : 5000 ans d'histoire*. Didactique des Langues Étrangères. CLE International. ISBN : 2-19-033353-9.
- [Géry, 2002] Mathias GÉRY (2002). *Indexation et interrogation de chemins de lecture en contexte pour la recherche d'information structurée sur le web*. Thèse de doctorat, Université Joseph Fourier - Grenoble I.
url : <http://tel.archives-ouvertes.fr/documents/archives0/00/00/44/53/>.
- [Ghiglione & Matalon, 1978] Rodolphe GHIGLIONE & Benjamin MATALON (1978). *Les enquêtes sociologiques : théories et pratiques*. Armand Colin. ISBN : 2-200-31046-3.
- [Godwin-Jones, 2000] Bob GODWIN-JONES (2000). Web browser trends and technologies. *Language Learning & Technology*, 4(1):6-11. ISSN : 1094-3501.
url : <http://llt.msu.edu/vol4num1/emerging/>.

- [Gómez Casañ & Martín Viaño, 1990] Pilar GÓMEZ CASAÑ & María del Mar MARTÍN VIAÑO (1990). *Didáctica de las segundas lenguas : estrategias y recursos básicos*, chapitre La expresión escrita : de la frase al texto, pages 43–63. Santillana. ISBN : 84-294-3152-7.
- [Goodfellow *et al.*, 1999] Robin GOODFELLOW, Patricia MANNING & Marie-Noëlle LAMY (1999). Building an online open and distance language learning environment. In Robert DEBSKI & Michael LEVY, éditeurs, *WorldCALL: Themes for the New Millennium*, Actes de *WorldCALL 1998 : CALL to Creativity* (Melbourne, Australie), pages 267–286. Swets and Zeitlinger.
url : <http://iet-staff.open.ac.uk/r.goodfellow/worldcall/worldcall.html>.
- [Goodwin & Stevens, 1993] Laura D. GOODWIN & Ellen A. STEVENS (1993). The influence of gender on university faculty members' perceptions of "good" teaching. *Journal of Higher Education*, 64(2):166–185. ISSN : 0022-1546.
url : <http://www.jstor.org/stable/2960028>.
- [Grandbastien & Labat, 2006] Monique GRANDBASTIEN & Jean-Marc LABAT (2006). Introduction. In Monique GRANDBASTIEN & Jean-Marc LABAT, éditeurs, *Environnements informatiques pour l'apprentissage humain*, pages 17–24. Hermès et Lavoisier. ISBN : 2-7462-1171-8.
- [Granger, 2002] Sylviane GRANGER (2002). A bird's-eye view of learner corpus research. In Granger *et al.* [2002], pages 3–33. ISBN : 90-272-1701-7.
- [Granger *et al.*, 2002] Sylviane GRANGER, Joseph HUNG & Stephanie PETCH-TYSON, éditeurs (2002). *Computer learner corpora, second language acquisition and foreign language teaching*, volume 6 de *Language Learning & Language Teaching*. John Benjamins Publishing Company. ISBN : 90-272-1701-7.
- [Granger *et al.*, 2007] Sylviane GRANGER, Olivier KRAIF, Claude PONTON, Georges ANTONIADIS & Virginie ZAMPA (2007). Integrating learner corpora and natural language processing: a crucial step towards reconciling technological sophistication and pedagogical effectiveness. *ReCALL*, 19(03):252–268. ISSN : 0958-3440.
url : http://journals.cambridge.org/article_S0958344007000237.
- [Grasha, 2002] Anthony F. GRASHA (2002). *Teaching with style*. Alliance Publisher. ISBN : 0-6745071-1-0.
url : http://ilte.ius.edu/pdf/teaching_with_style.pdf.
- [Guidon, 2001] Jacques GUIDON (2001). Note sur les normes et standards pour la formation.
url : <http://www.educnet.education.fr/superieur/normes.htm>.
- [Guilbaud, 1980] Georges-Théodule GUILBAUD (1980). Zipf et les fréquences. *Mots*, 1(1):97–126. ISSN : 0243-6450.
url : http://www.persee.fr/web/revues/home/prescript/article/mots_0243-6450_1980_num_1_1_1007.
- [Habert *et al.*, 1998] Benoît HABERT, Cécile FABRE & Fabrice ISAAC (1998). *De l'écrit au numérique - constituer, normaliser et exploiter les corpus électroniques*. Masson / InterEditions. ISBN : 2-225-82953-5.
- [Habert *et al.*, 1997] Benoît HABERT, Adeline NAZARENKO & André SALEM (1997). *Les linguistiques de corpus*. Collection U. Série Linguistique. A. Colin. ISBN : 2-200-01775-8.

- [Hakstian *et al.*, 1979] A. Ralph HAKSTIAN, J. Christian ROED & John C. LIND (1979). Two-sample t-2 procedure and the assumption of homogeneous covariance matrices. *Psychological Bulletin*, 86(6):1255–1263. ISSN : 0033-2909.
url : <http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=bul-86-6-1255&site=ehost-live>.
- [Harris, 2000] Ruth HARRIS (2000). Are CALL packages disregarding the research on dealing with authentic materials? *ITB Journal*, (1):4–13.
url : <http://www.itb.ie/site/researchinnovation/itbjournal/ITB-Journal-May-2000.pdf>.
- [Hart, 1995] Robert S. HART (1995). The illinois PLATO foreign languages project. *CALICO Journal*, 12(4):15–37. ISSN : 0742-7778.
url : <https://www.calico.org/a-586-The%20Illinois%20PLATO%20Foreign%20Languages%20Project.html>.
- [Hashweh, 1987] Maher Z. HASHWEH (1987). Effects of subject-matter knowledge in the teaching of biology and physics. *Teaching and Teacher Education*, 3(2):109–120. ISSN : 0742-051X. Cité dans [Tochon, 1989].
- [Heift & Schulze, 2007] Trude HEIFT & Mathias SCHULZE (2007). *Errors and intelligence in computer-assisted language learning - parsers and pedagogues*. Numéro 2 in Routledge Studies in Computer-Assisted Language Learning. Routledge. ISBN : 0-415-36191-5.
- [Herreras, 1998] José Carlos HERRERAS, éditeur (1998). *L'enseignement des langues étrangères dans les pays de l'union européenne*. Numéro 92 in Bibliothèque des Cahiers de l'Institut de Linguistique de Louvain. Peeters. ISBN : 2-87723-376-6, ISSN : 0779-1666.
- [Herreras, 2001] José Carlos HERRERAS (2001). De l'apprentissage à la maîtrise des langues étrangères en Europe. In *De la méthodologie à la didactologie, Études de Linguistique Appliquée*, 3-4(123):313–323. ISSN : 0071-190X.
url : <http://www.cairn.info/revue-ela-2001-3-page-313.htm>.
- [Higgins, 1986] John HIGGINS (1986). Introduction: smart learners and dumb machines. *System*, 14(2):147–150. ISSN : 0346-251X.
- [Higgins & Johns, 1984] John HIGGINS & Tim JOHNS (1984). *Computers in language learning*. Collins ELT / Addison-Wesley, World Language Division. ISBN : 0-00-370623-0.
- [Hillmann, 2005] Diane HILLMANN (2005). Using dublin core. Usage Guide (première version en jan 1998), DCMI.
url : <http://dublincore.org/documents/usageguide/>.
- [Hodgins & Duval, 2004] Wayne HODGINS & Erik DUVAL (2004). Learning object metadata (présentation). In *CEN/ISSS + LTSC Meetings*.
url : http://ltsc.ieee.org/wg12/files/CEN-ISS_LOM_2004_Jan27_Madrid.ppt.
- [Holloway, 2002a] John HOLLOWAY (2002a). *Cambiar el mundo sin tomar el poder. El significado de la revolución hoy*. Herramienta / Universidad de Puebla, édition mexicaine de [Holloway, 2002b] (traduction : Marcela ZANGARO). ISBN : 987-9306-10-4. Cité dans [Holloway, 2007].

- [Holloway, 2002b] John HOLLOWAY (2002b). *Change the world without taking power: the meaning of revolution today*. Pluto, 1st edition. ISBN : 0745318649.
url : <http://libcom.org/library/change-world-without-taking-power-john-holloway>.
Cité dans [Holloway, 2007, 2002a].
- [Holloway, 2007] John HOLLOWAY (2007). *Changer le monde sans prendre le pouvoir : le sens de la révolution aujourd'hui*. Syllepse / Lux, édition française de [Holloway, 2002a] (traduction : Sylvie BOSSERELLE). ISBN : 978-2-84950-14-29.
- [Houssaye, 1988] Jean HOUSSAYE (1988). *Le triangle pédagogique*, volume 1 de *Théorie et pratiques de l'éducation scolaire*. Peter Lang. ISBN : 3-261-03827-6.
- [Hubbard, 1987] Philip HUBBARD (1987). *Modern media in foreign language education: theory and implementation*, chapitre Language Teaching Approaches, the Evaluation of CALL Software, and Design Implications, pages 227–254. ACTFL Foreign Language Education Series. National Textbook Company. ISBN : 0-8442-9386-5.
url : <http://tr.im/Hubbard>.
- [Hubbard, 1996] Philip HUBBARD (1996). Elements of CALL methodology: development, evaluation, and implementation. In Martha PENNINGTON, éditeur, *The Power of CALL*. Athelstan. ISBN : 0940753049. Cité dans [Rézeau, 2001; Levy & Stockwell, 2006].
- [Huberman, 1989a] Michaël HUBERMAN (1989a). *La vie des enseignants : évolution et bilan d'une profession*, chapitre Le cycle de vie professionnelle des enseignants, pages 9–35. In Huberman et al. [1989]. ISBN : 2-603-00690-8, ISSN : 1019-3391.
- [Huberman, 1989b] Michaël HUBERMAN (1989b). *La vie des enseignants : évolution et bilan d'une profession*, chapitre Les phases de la carrière de l'enseignant, pages 47–144. In Huberman et al. [1989]. ISBN : 2-603-00690-8, ISSN : 1019-3391.
- [Huberman et al., 1989] Michaël HUBERMAN, Marie-Madeleine GROUNAUER & Jürg MARTI (1989). *La vie des enseignants : évolution et bilan d'une profession*. Actualités pédagogiques et psychologiques. Delachaux et Niestlé. ISBN : 2-603-00690-8, ISSN : 1019-3391.
- [Huberty & Olejnik, 2006] Carl J. HUBERTY & Stephen OLEJNIK (2006). *Applied manova and discriminant analysis*. Wiley series in probability and statistics. John Wiley & Sons, 2nd edition. ISBN : 0-471-46815-0.
- [Hutchinson & Waters, 1987] Tom HUTCHINSON & Alan WATERS (1987). *English for specific purposes: a learning-centred approach*. New directions in language teaching. Cambridge University Press. ISBN : 0-521-31837-8.
- [Hymes, 1984] Dell H. HYMES (1984). *Vers la compétence de communication*. Langues et apprentissage des langues. Hatier-Didier. ISBN : 2-218-05877-4.
- [IEEE, 2002] IEEE (2002). Final 1484.12.1 LOM draft standard document. Document technique, IEEE LTSC WG12.
url : http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.
- [IMS, 2006] IMS meta-data best practice guide for IEEE 1484.12.1-2002 standard for Learning Object Metadata. (2006).
url : http://www.imsglobal.org/metadata/mdv1p3/imsmd_bestv1p3.html.

- [INSEE, 2008a] INSEE (2008a). Dépense intérieure d'éducation, *In* INSEE [2008c]. Document technique, INSEE.
url : http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=NATTEF07304.
- [INSEE, 2008b] INSEE (2008b). Dépense pour la formation professionnelle continue et l'apprentissage, *In* INSEE [2008c]. Document technique, INSEE.
url : http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=NATTEF07304.
- [INSEE, 2008c] INSEE (2008c). Tableaux de l'économie française. Document technique, INSEE.
url : http://www.insee.fr/fr/publications-et-services/sommaire.asp?id=185®_id=0.
- [ISO, 2004] ISO (2004). ISO/IEC guide 2. Document technique, ISO.
url : http://www.iso.org/iso/fr/how_ISO_develops_standards.
- [ISO, 2007] ISO/IEC JTC1 SC36 (2007). SC36 business plan for period september 2007 to september 2008. Business plan, ISO.
url : <http://isotc.iso.org/livelink/livelink?func=ll&objId=1056984&objAction=browse&sort=name>.
- [Jackson, 1968] Philip W. JACKSON (1968). *Life in classrooms*. Holt, Rinehart and Winston. Cité dans [Tochon, 1989].
- [Johns, 1986] Tim JOHNS (1986). Micro-concord: a language learner's research tool. *System*, 14(2):151–162. ISSN : 0346-251X.
url : <http://www.sciencedirect.com/science/article/B6VCH-465D5H9-V/2/cbbb6d3cbd43dee0377e6c1676888ab1>.
- [Johns, 1991] Tim JOHNS (1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *In* Tim JOHNS & Philip KING, éditeurs, *Classroom Concordancing, English Language Research Journal*, 4:27–45. Cité dans [Boulton, 2007].
- [Johns, 2002] Tim JOHNS (2002). Data-driven learning: the perpetual challenge. *In* Kettemann & Marko [2002], pages 107–117.
- [Jones, 1986] Christopher JONES (1986). It's not so much the program, more what you do with it: the importance of methodology in CALL. *System*, 14(2):171–178. ISSN : 0346-251X.
url : <http://www.sciencedirect.com/science/article/B6VCH-465D5H9-X/2/2e22378443427ea670660219c906535d>.
- [Jones, 1995] Randall L. JONES (1995). TICCIT and CLIPS: the early years. *CALICO Journal*, 12(4):84–96. ISSN : 0742-7778.
url : <https://www.calico.org/a-587-TICCIT%20and%20CLIPS%20The%20Early%20Years.html>.
- [Jonnaert & Laurin, 2001a] Philippe JONNAERT & Suzanne LAURIN (2001a). *Les didactiques des disciplines : un débat contemporain*, chapitre Des orientations didactiques, pages 1–7. *In* Jonnaert & Laurin [2001b]. ISBN : 2-7605-1153-7.
- [Jonnaert & Laurin, 2001b] Philippe JONNAERT & Suzanne LAURIN, éditeurs (2001b). *Les didactiques des disciplines : un débat contemporain*. Éducation Recherche. Presses de l'Université du Québec. ISBN : 2-7605-1153-7.

- [Jung, 2005] Udo JUNG (2005). CALL: past, present and future - a bibliometric approach. *ReCALL*, 17(1):4–17. ISSN : 0958-3440.
url : <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=305544&fulltextType=RA&fileId=S0958344005000212>.
- [Kaltenböck & Mehlmauer-Larcher, 2005] Gunther KALTENBÖCK & Barbara MEHLMAUER-LARCHER (2005). Computer corpora and the language classroom: on the potential and limitations of computer corpora in language teaching. *ReCALL*, 17(1):65–84. ISSN : 0958-3440.
url : http://journals.cambridge.org/article_S0958344005000613.
- [Katz, 1972] Lilian G. KATZ (1972). Developmental stages of preschool teachers. *Elementary School Journal*, 73(1):123–127. ISSN : 0013-5984.
url : <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED057922>.
- [Kern *et al.*, 2004] Richard KERN, Paige WARE & Mark WARSCHAUER (2004). Crossing frontiers: new directions in online pedagogy and research. *Annual Review of Applied Linguistics*, 24:243–260. ISSN : 0267-1905.
url : http://journals.cambridge.org/article_S0267190504000091.
- [Kettemann & Marko, 2002] Bernhard KETTEMANN & Georg MARKO, éditeurs (2002). *The learner as corpus designer*, actes de *Fourth International conference on Teaching and Language Corpora (TaLC)* (Graz, 19-24 juillet 2000), numéro 42 in *Language and Computers – Studies in Practical Linguistics*. Rodopi. ISBN : 90-420-1450-4.
- [Kirby *et al.*, 2004] Simon KIRBY, Kenny SMITH & Henry BRIGHTON (2004). From UG to universals: linguistic adaptation through iterated learning. *Studies in Language*, 28(3):587–607.
url : <http://www.ling.ed.ac.uk/~simon/Papers/Kirby/From%20UG%20to%20Universals%20linguistic%20adaptation%20through.pdf>.
- [Koper, 2000] Rob KOPER (2000). From change to renewal: educational technology foundations of electronic learning environments. Exposé. Cf. *OUN inaugural address*, LN : Publications and Preprints. Dspace.
url : <http://dspace.learningnetworks.org/handle/1820/38>.
- [Koteyko, 2006] Nelya KOTEYKO (2006). Corpus linguistics and the study of meaning in discourse. *The Linguistics Journal*, 1(2):132–157. ISSN : 1718-2298.
url : http://www.linguistics-journal.com/June2006_nk.php.
- [Kraif *et al.*, 2004] Olivier KRAIF, Georges ANTONIADIS, Sandra ÉCHINARD, Mathieu LOISEAU, Thomas LEBARBÉ & Claude PONTON (2004). NLP tools for CALL: the simpler, the better. *In Proceedings of InSTIL/ICALL 2004 Symposium on Computer Assisted Learning*.
url : <http://w3.u-grenoble3.fr/lidilem/labo/file/Instill2004.PDF>.
- [Kraif & Ponton, 2007] Olivier KRAIF & Claude PONTON (2007). Du bruit, du silence et des ambiguïtés : que faire du TAL pour l'apprentissage des langues ? Actes de *TALN* (Toulouse).
url : <http://www.u-grenoble3.fr/ponton/perso/docs/TALN07.pdf>.
- [Krashen, 1982] Stephen D. KRASHEN (1982). *Principles and practice in second language acquisition*. Language teaching methodology series. Pergamon Press. ISBN : 0-08-028628-3.
- [Krüger, 1981] Michael KRÜGER (1981). *Übungstypologie zum kommunikativen deutschunterricht*, chapitre Unterrichtsplanung : Übungs- und Handlungssequenzen. Langenscheidt. ISBN : 3468494300. Cité dans [Carol *et al.*, 2004].

- [Kübler, 2002] Natalie KÜBLER (2002). Linguistic concerns in teaching with language corpora. In Kettemann & Marko [2002], pages 193–202.
url : http://www.eila.univ-paris-diderot.fr/_media/recherche/kublertalc002.pdf.
- [Kurella *et al.*, 2008] Svitlana KURELLA, Serge SHAROFF & Antony HARTLEY (2008). Rhetorical text structure in acquiring reading skills in L3. In TALC [2008], pages 193–197.
- [Labov, 1972] William LABOV (1972). *Sociolinguistic patterns*. Numéro 4 in Conduct and communication. University of Pennsylvania press. ISBN : 0-8122-1052-2. Cité dans [Labov, 1976].
- [Labov, 1976] William LABOV (1976). *Sociolinguistique*. Les éditions de minuit, édition française de [Labov, 1972] (traduction : Alain KIHM). ISBN : 2-7073-0120-5.
- [Lamarre, 2003] Anne-Marie LAMARRE (2003). *Étude de l'expérience de la première année d'enseignement au primaire dans une perspective phénoménologico-herméneutique*. Thèse de doctorat, Université du Québec à Chicoutimi.
url : <http://dx.doi.org/doi:10.1522/18194167>.
- [Lamy & Klarskov Mortensen, 2007] Marie-Noëlle LAMY & Hans Jørgen KLARSKOV MORTENSEN (2007). Using concordance programs in the modern foreign languages classroom.
url : http://www.ict4lt.org/en/en_mod2-4.htm.
- [Larose *et al.*, 2002] François LAROSE, Yves LENOIR, Thierry KARSENTI & Vincent GRENON (2002). Les facteurs sous-jacents au transfert des compétences informatiques construites par les futurs maîtres du primaire sur le plan de l'intervention éducative. *Revue des sciences de l'éducation*, 28(2):265–287. ISSN : 0318-479X.
url : <http://id.erudit.org/iderudit/007354ar>.
- [Laurin & Gaudreau, 2001] Suzanne LAURIN & Louise GAUDREAU (2001). *Les didactiques des disciplines : un débat contemporain*, chapitre De la didactique aux didactiques : dialogue sur des enjeux éducatifs, pages 9–27. In Jonnaert & Laurin [2001b]. ISBN : 2-7605-1153-7.
- [Lebarbé, 2002] Thomas LEBARBÉ (2002). *Hiérarchie inclusive en analyse syntaxique coopérative*. Thèse de doctorat, Université de Caen.
url : <http://w3.u-grenoble3.fr/lebarbe/index.php/Main/These>.
- [Lebarbé & Meynard, 2008] Thomas LEBARBÉ & Cécile MEYNARD (2008). CLELIA : un modèle documentaire et une plateforme de gestion de manuscrits. In Broudoux & Chartron [2008], pages 317–334.
- [LeBlanc, 1989] Raymond LE BLANC (1989). Le curriculum multidimensionnel : une approche intégrée pour l'enseignement de la langue seconde. *Études de Linguistique Appliquée*, (75):78–94. ISSN : 0071-190X.
url : http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&res_dat=xri:pao-us:&rft_dat=xri:pao:article:t611-1989-075-00-000007:1.
- [Lecocq, 2008] Katia LECOCQ (2008). *Acquisition d'une langue seconde en milieu scolaire – évaluation longitudinale réalisée auprès d'enfants francophones immergés en néerlandais*. Thèse de doctorat, Université Libre de Bruxelles.
- [Lefèvre, 2000] Philippe LEFÈVRE (2000). *La recherche d'informations : du texte intégral au thésaurus*. Hermès Science. ISBN : 2-7462-0173-9.

- [Lemieux, 1964] Claude P. LEMIEUX (1964). Harold E. Palmer's contribution to the oral method of teaching foreign languages. *The Slavic and East European Journal*, 8(3):320–326. ISSN : 0037-6752.
url : <http://www.jstor.org/stable/304223>.
- [Levin *et al.*, 1991] Lori S. LEVIN, David A. EVANS & Donna M. GATES (1991). The ALICE system: a workbench for learning and using language. *CALICO Journal*, 9(1):27–56. ISSN : 0742-7778.
url : <https://www.calico.org/a-522-The%20Alice%20System%20A%20Workbench%20for%20Learning%20and%20Using%20Language.html>.
- [Levy, 1997] Michael LEVY (1997). *Computer-assisted language learning - context and conceptualization*. Oxford University Press. ISBN : 0-19-823631-X.
- [Levy & Kennedy, 2004] Michael LEVY & Claire KENNEDY (2004). A task-cycling pedagogy using stimulated reflection and audio-conferencing in foreign language learning. *Language Learning & Technology*, 8(2):50–68. ISSN : 1094-3501.
url : <http://llt.msu.edu/vol8num2/levy/>.
- [Levy & Stockwell, 2006] Michael LEVY & Glenn STOCKWELL (2006). *CALL dimensions – options and issues in computer assisted language learning*. ESL & applied linguistics professional series. Lawrence Erlbaum Associates. ISBN : 0-8058-5633-1.
- [Lévy, 1997] Pierre LÉVY (1997). *Cyberculture*. Odile Jacob. ISBN : 2-7381-0512-2.
- [Li, 2003] Yiping LI (2003). Un système de segmentation du chinois basé sur des triplets. In *RÉCITAL*.
url : http://www-list.cea.fr/gb/publications/docs/si/ingenierie_connaissance/fr/taln2003_li_recital_li_segchinois.pdf.
- [Lin Hsiao, 1996] Jy Wana Daphne LIN HSIAO (1996). CACL theories.
url : <http://www.edb.utexas.edu/csclstudent/dhsiao/theories.html>.
- [Little, 1997] David LITTLE (1997). Language awareness and the autonomous language learner. *Language Awareness*, 6(2&3):93–104. ISSN : 0965-8416. Cité dans [Goodfellow *et al.*, 1999].
- [Little & Singleton, 1988] David LITTLE & David SINGLETON (1988). Authentic materials and the role of fixed support in language teaching: towards a manual for language learners. CLCS Occasional paper 20, Trinity College.
url : <http://eric.ed.gov/ERICWebPortal/detail?accno=ED316033>.
- [Loiseau, 2003] Mathieu LOISEAU (2003). Vers la création d'une base de données de ressources textuelles indexée pédagogiquement pour l'enseignement des langues. Mémoire de DEA, Université Stendhal Grenoble 3.
url : http://mathieu.loiseau.free.fr/bdtp/fichiers/DEAScL_BDTIP-2003.pdf.
- [Loiseau, 2004] Mathieu LOISEAU (2004). La description de ressources pédagogiques : état de l'art et application aux ressources textuelles pour l'enseignement des langues. In *TAL et apprentissage des langues*, Actes de *TAL & AL* (Grenoble, 22 octobre 2004). LIDILEM et ATALA, XRCE.
url : <http://w3.u-grenoble3.fr/lidilem/talal/actes/JourneeTALAL-041022-loiseau.pdf>.

- [Loiseau, 2008] Mathieu LOISEAU (2008). Le « contexte pédagogique » comme pré-requis à la notion de facette de texte. Actes de *1st International Conference on Linguistic and Intercultural Education (CLIE 2008)* (Alba Iulia, Roumanie, 27-29 Novembre 2008).
url : <http://mathieu.loiseau.free.fr/bdtpip/fichiers/articles/AlbaIulia.pdf>.
- [Loiseau *et al.*, 2005] Mathieu LOISEAU, Georges ANTONIADIS & Claude PONTON (2005). Pedagogical text indexation and exploitation for language teaching. In A. Méndez VILAS, B. Gonzalez PEREIRA, J. Mesa GONZÁLEZ & J.A. Mesa GONZÁLEZ, éditeurs, *Recent research developments in learning technologies*, volume III, pages 984–994. FORMATEX. ISBN : 609-5997-X.
url : <http://www.formatex.org/micte2005/165.pdf>.
- [Loiseau *et al.*, 2008a] Mathieu LOISEAU, Georges ANTONIADIS & Claude PONTON (2008a). Model for pedagogical indexation of texts for language teaching. In José CORDEIRO, Boris SHISHKOV, Alpesh RANCHORDAS & Markus HELFERT, éditeurs, *ICSOFT (ISDM/ABF)*, Actes de *ICSOFT 3* (Porto, 5-8 juillet 2008), volume ISDM/ABF, pages 212–217. INSTICC Press. ISBN : 978-989-8111-53-1.
url : <http://mathieu.loiseau.free.fr/bdtpip/fichiers/articles/icsoft-2008.pdf>.
- [Loiseau *et al.*, 2008b] Mathieu LOISEAU, Georges ANTONIADIS & Claude PONTON (2008b). The concept of “text facet” as a means to achieve pedagogical indexation of a text base dedicated to language teaching. In TALC [2008], pages 421–425.
url : <http://mathieu.loiseau.free.fr/bdtpip/fichiers/articles/Talc8.pdf>.
- [Love & Nader, 1997] James LOVE & Ralph NADER (1997). Microsoft, monopole du prochain siècle. *Le Monde diplomatique*.
url : <http://www.monde-diplomatique.fr/1997/11/NADER/9458>.
- [Lowe, 1999] Carrie A. LOWE (1999). Meta-data resource discovery and educational information on the internet: the gateway to educational materials GEM project. In *Museums and the Web*.
url : <http://www.archimuse.com/mw99/papers/lowe/lowe.html>.
- [Luhn, 1960] Hans Peter LUHN (1960). Keyword-in-context index for technical literature. *American Documentation*²⁸, 11(4):288–295. ISSN : 0002-8231.
url : http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&res_dat=xri:pao-us:&rft_dat=xri:pao:article:5002-1960-011-04-000003:1.
- [Lusalusa *et al.*, 2000] Simon K.T. LUSALUSA, Alain LAMMÉ & Éric UYTTEBROUCK (2000). Quelles compétences en nouvelles technologies de l’information et de la communication pour les enseignants du supérieur. In *Apprendre et enseigner dans l’enseignement supérieur*, Actes de *Congrès international Francophone ADMES-AIPU* (Paris, 10-13 avril 2000).
url : http://www.ulb.ac.be/ulb/cte/docs/Article_Compétences_TICE.pdf.
- [Lynch, 1997] Clifford A. LYNCH (1997). The Z39.50 information retrieval standard - Part I: a strategic view of its past, present and future. *D-Lib Magazine*.
url : <http://www.dlib.org/dlib/april97/04lynch.html>.
- [Mackey, 1997] William Francis MACKKEY (1997). Langue maternelle, langue première, langue seconde, langue étrangère. In Marie-Louise MOREAU, éditeur, *Sociolinguistique – concepts de base*, numéro 218 in *Psychologies et Sciences Humaines*, pages 183–185. Mardaga. ISBN : 2-87009-664-X.

28. Maintenant *Journal of the American Society for Information Science and Technology*, ISSN : 1532-2882

- [Macklovitch, 1995] Elliott MACKLOVITCH (1995). The future of MT is now and Bar-Hillel was (almost entirely) right. In *Proceedings of the Fourth Bar-Ilan Symposium on the Foundations of Artificial Intelligence*.
url : <http://rali.iro.umontreal.ca/Publications/urls/bisfai95.ps>.
- [Mangenot, 2002] François MANGENOT (2002). Produits multimédias : médiation ou médiatisation ? *Le français dans le monde*, (322):34–35. ISSN : 0015-9395.
url : http://w3.u-grenoble3.fr/espace_pedagogique/fdlm.doc.
- [Mangenot, 2003] François MANGENOT (2003). Présentation d'EUROCALL 2003 "New literacies in language learning & teaching". *ALSIC*, 6(2):117–125. ISSN : 1286-4986.
url : http://alsic.u-strasbg.fr/Num11/mangenot/alsic_n11-poi1.htm.
- [Mangeot-Lerebours *et al.*, 2003] Mathieu MANGEOT-LEREBOURS, Gilles SÉRASSET & Mathieu LAFOURCADE (2003). Construction collaborative d'une base lexicale multilingue – le projet papillon. *TAL*, 44(2):151–176. ISSN : 1965-0906.
url : http://jibiki.univ-savoie.fr/~mangeot/Publications/TAL03_MM-GS-ML.pdf.
- [Manning & Schütze, 1999] Christopher D. MANNING & Hinrich SCHÜTZE (1999). *Foundations of statistical natural language processing*. MIT Press. ISBN : 0-262-13360-1.
- [Marcus, 1999] Catherine MARCUS (1999). *Français seconde langue, 36 lectures pour les collèges*. CRDP de l'académie de Grenoble ; Delagrave. ISBN : 2-206-08161-X, ISSN : 1264-3483.
- [Martinez, 1996] Pierre MARTINEZ (1996). *La didactique des langues étrangères*. Numéro 3199 in *Que sais-je ?* Presses Universitaires de France. ISBN : 2130565530. Cité dans [Rézeau, 2001].
- [Matthey, 2003] Marinette MATTHEY (2003). *Apprentissage d'une langue et interaction verbale - sollicitation, transmission et construction de connaissances linguistiques en situation exolingue*. Exploration - Recherches en sciences de l'éducation. Peter Lang, 2e édition. ISBN : 3-906766-62-4.
- [McEnery & Wilson, 1997] Tony MCENERY & Andrew WILSON (1997). Teaching and language corpora (TALC). *ReCALL*, 9(01):5–14. ISSN : 0958-3440. Cité dans [Chambers, 2007a].
- [McEnery & Wilson, 2001] Tony MCENERY & Andrew WILSON (2001). *Corpus linguistics*. Edinburgh textbooks in empirical linguistics. Edinburgh University Press, 2e édition. ISBN : 0748611657.
- [MEDEF & CCIP, 2004] MOUVEMENT DES ENTREPRISES DE FRANCE & CHAMBRE DE COMMERCE ET D'INDUSTRIE DE PARIS (2004). Consultation nationale sur l'avenir de la recherche : premiers axes de réflexion. Avenir de la recherche : contribution commune MEDEF - CCIP.
url : http://www.medef.fr/main/core.php?pag_id=20704.
- [Meier, 1992] Hans Ed. MEIER (1992). Règles fondamentales de mise en page. *Cahiers GUTenberg*, (13):5–38.
url : <http://www.gutenberg.eu.org/pub/GUTenberg/publicationsPDF/13-meier.pdf>.
- [MEN, 2003] DIRECTION DE L'ÉVALUATION DE LA PROSPECTIVE ET DE LA PERFORMANCE (2003). Repères et références statistiques sur les enseignements, la formation et la recherche. Document technique, Ministère de la Jeunesse, de l'Éducation nationale et de la Recherche. ISBN : 2-11-093455-7, ISSN : 1635-9089.
url : http://media.education.gouv.fr/file/2003/73/8/rers2003_27738.pdf.

- [MEN, 2004] DIRECTION DE L'ÉVALUATION DE LA PROSPECTIVE ET DE LA PERFORMANCE (2004). Repères et références statistiques sur les enseignements, la formation et la recherche. Document technique, Ministère de l'Éducation nationale, de l'Enseignement Supérieur et de la Recherche. ISBN : 2-11-094345-9, ISSN : 1635-9089.
url : http://media.education.gouv.fr/file/2004/73/1/rers2004_27731.pdf.
- [MEN, 2005] DIRECTION DE L'ÉVALUATION DE LA PROSPECTIVE ET DE LA PERFORMANCE (2005). Repères et références statistiques sur les enseignements, la formation et la recherche. Document technique, Ministère de l'Éducation nationale, de l'Enseignement Supérieur et de la Recherche. ISBN : 2-11-095390-X, ISSN : 1635-9089.
url : <http://media.education.gouv.fr/file/58/2/2582.pdf>.
- [MEN, 2006] DIRECTION DE L'ÉVALUATION DE LA PROSPECTIVE ET DE LA PERFORMANCE (2006). Repères et références statistiques sur les enseignements, la formation et la recherche. Document technique, Ministère de l'Éducation nationale, de l'Enseignement Supérieur et de la Recherche. ISBN : 2-11-095401-9, ISSN : 1635-9089.
url : <http://media.education.gouv.fr/file/40/0/6400.pdf>.
- [MEN, 2007] DIRECTION DE L'ÉVALUATION DE LA PROSPECTIVE ET DE LA PERFORMANCE (2007). Repères et références statistiques sur les enseignements, la formation et la recherche. Document technique, Ministère de l'Éducation nationale et ministère de l'Enseignement supérieur et de la Recherche. ISBN : 978-11-095421-3, ISSN : 1635-9089.
url : <http://media.education.gouv.fr/file/61/2/6612.pdf>.
- [Moirand, 1982] Sophie MOIRAND (1982). *Enseigner à communiquer en langue étrangère*. Recherches / applications. Hachette. ISBN : 2-01-008787-9.
- [Mossel, 1960] Rudolf MOSSEL (1960). Pédagogie des langues étrangères et linguistique. *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Éducation*, 6(3):309–317. ISSN : 0020-8566.
url : <http://www.jstor.org/stable/3441582>.
- [Mucchielli, 1990] Roger MUCCHIELLI (1990). *Le questionnaire dans l'enquête psycho-sociale*. Formation permanente en sciences humaines. ESF, 9e édition. ISBN : 2-7101-0761-9.
- [Murray & Barnes, 1998] Liam MURRAY & Ann BARNES (1998). Beyond the “wow” factor—evaluating multimedia language learning software from a pedagogical viewpoint. *System*, 26(2):249–259. ISSN : 0346-251X.
url : <http://www.sciencedirect.com/science/article/B6VCH-3TN9MNX-8/2/a08cda3e81819a6d441c357f4eb5b60c>.
- [Myers, 1986] Eugene MYERS (1986). An O(ND) difference algorithm and its variations. *Algorithmica*, 1(1-4):251–256. ISSN : 0178-4617.
url : <http://dx.doi.org/10.1007/BF01840446>.
- [Netter, 2008] Carole NETTER. ClicNet. (2008).
url : <http://clicnet.swarthmore.edu/>.
- [Nevins et al., 2007] Andrew Ira NEVINS, David PESETSKY & Cilene RODRIGUES (2007). Pirahã exceptionality: a reassessment.
url : <http://ling.auf.net/lingBuzz/000411>.

- [Nicola & van der Linden, 2005] Matthias NICOLA & Bert VAN DER LINDEN (2005). Native XML support in DB2 universal database. In *VLDB '05, Actes de 31st international conference on Very large data bases* (Trondheim, Norway, 30 août - 2 septembre 2005), pages 1164–1174. VLDB Endowment. ISBN : 1-59593-154-6.
url : <http://www.vldb2005.org/program/paper/thu/p1164-nicola.pdf>.
- [Nunan, 1989] David NUNAN (1989). *Designing tasks for the communicative classroom*. Cambridge language teaching library. Cambridge University Press. ISBN : 0-521-67915-6.
- [Nunan, 2006] David NUNAN (2006). Task-based language teaching in the asia context: defining 'task'. *Asian EFL Journal*, 8(3).
url : http://www.asian-efl-journal.com/Sept_06_dn.php.
- [Oxford, 1993] Rebecca L. OXFORD (1993). Intelligent computers for learning languages: The view for language acquisition and instructional methodology. *Computer Assisted Language Learning*, 6(2):173–188. ISSN : 0958-8221.
url : <http://www.informaworld.com/10.1080/0958822930060206>. Cité dans [Chanier, 1995].
- [Oxford, 2006] Rebecca L. OXFORD (2006). Task-based language teaching and learning: an overview. *Asian EFL Journal*, 8(3).
url : http://www.asian-efl-journal.com/Sept_06_ro.php.
- [Pallotti, 2002] Gabriele PALLOTTI (2002). La classe dans une perspective écologique de l'acquisition. In *L'Acquisition en classe de langue, Acquisition et Interaction en Langue Étrangère [En ligne]*, (16). ISSN : 1778-7432.
url : <http://aile.revues.org/document1395.html>.
- [Palmer et al., 1981] Adrian PALMER, Peter GROOT & George TROSPER, éditeurs (1981). *The construct validation of tests of communicative competence*, actes de *TESOL Conference* (Boston, 27-28 février 1979). TESOL. ISBN : 0-939791-09-9.
url : <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED223103>. Cité dans [Hymes, 1984].
- [Palmer, 1922] Harold E. PALMER (1922). *The oral method of teaching languages – a monograph on conversational methods together with a full description and abundant examples of fifty appropriate forms of work*. W. Heffer & Sons Ltd., 2e impression.
url : <http://www.archive.org/details/oralmethodofteac00palmuoft>.
- [Panchenko, 2009] Oleksandr PANCHENKO (2009). Hybrid storage for enabling fully-featured text search and fine-grained structural search over source code. In *SUITE '09 : Proceedings of the 2009 ICSE Workshop on Search-Driven Development-Users, Infrastructure, Tools and Evaluation*, Actes de *31st International Conference on Software Engineering (ICSE)* (Vancouver, 16-24 mai 2009), pages 37–40. IEEE Computer Society. ISBN : 978-1-4244-3740-5.
- [Papert, 1980] Seymour PAPERT (1980). *Mindstorms: children, computers, and powerful ideas*. Basic Books. ISBN : 0-465-04627-4.
- [Papert, 1981] Seymour PAPERT (1981). *Jaillissement de l'esprit : ordinateurs et apprentissage*. Flammarion, édition française de [Papert, 1980] (traduction : Rose-Marie VASSALLO-VILLANEAU). ISBN : 2080812106. Cité dans [Bruillard, 1997; de Vries, 2001].

- [Paroubek *et al.*, 1999] Patrick PAROUBEK, Gilles ADDA, Joseph-Jean MARIANI, Josette LECOMTE & Martin RAJMAN (1999). La campagne d'évaluation GRACE (grammaire et ressources pour les analyseurs de corpus et leur évaluation) des analyseurs morpho-syntaxiques pour le français. Rapport d'activité LIMSI.
url : <http://www.limsi.fr/RS99FF/CHM99FF/TLP99FF/tlp10/>.
- [Partington, 1998] Alan PARTINGTON (1998). *Patterns and meanings: using corpora for english language research and teaching*. Studies in Corpus Linguistics. John Benjamins Publishing Company. ISBN : 1-55619-396-3.
- [Patesson, 2007] René PATESSON (2007). Aspects de l'usage des questionnaires et l'auto-observation de l'activité en ergonomie. *Psychologie du Travail et des Organisations*, 13(4). ISSN : 1420-2530.
url : <http://tr.im/mzqb>.
- [Pernin, 2003] Jean-Philippe PERNIN (2003). Objets pédagogiques, activités ou ressources ? In Bruillard & de La Passadière [2003], pages 179–210. ISBN : 2-7462-0682-X.
- [Pernin, 2004] Jean-Philippe PERNIN (2004). À propos d'objets pédagogiques... Exposé à : *colloque IRDP* (2003). Cf. *Entre technique et pédagogie : la création de contenus multimédia pour l'enseignement et la formation*. ISBN : 2-88198-010-4.
url : http://www-clips.imag.fr/arcade/User/jean-philippe.pernin/recherche/download/Article_Pernin_Neuchatel07Nov03.pdf.
- [Pernin, 2006] Jean-Philippe PERNIN (2006). Normes et standards pour la conception, la production et l'exploitation des EIAH. In Monique GRANDBASTIEN & Jean-Marc LABAT, éditeurs, *Environnements informatiques pour l'apprentissage humain*, pages 201–222. Hermès et Lavoisier. ISBN : 2-7462-1171-8.
- [PHP, 2009] PHP (2009). PHP licensing. Licence, The PHP Group.
url : <http://www.php.net/license/>.
- [Polguère, 2002] Alain POLGUÈRE (2002). Notions de base en lexicologie. Notes de cours photocopiées, Observatoire de linguistique Sens-Texte. Cité dans [Mangeot-Lerebours *et al.*, 2003].
- [Polguère, 2003] Alain POLGUÈRE (2003). *Lexicologie et sémantique lexicale – notions fondamentales*. Paramètres. Presses de l'Université de Montréal. ISBN : 2-7606-1860-9.
url : <http://books.google.fr/books?id=rI74jhDqso4C>.
- [Poussard, 2000] Cécile POUSSARD (2000). *La compréhension de l'anglais oral et les technologies éducatives*. Thèse de doctorat, Université Paris 7 - Denis Diderot.
url : http://didatic.net/doc_pdf/thentier.pdf.
- [Pravec, 2002] Norma A. PRAVEC (2002). Survey of learner corpora. *ICAME Journal*, (26):81–114. ISSN : 0801-5775.
url : <http://icame.uib.no/ij26/pravec.pdf>.
- [Prince *et al.*, 2007] Violaine PRINCE, Yves KODRATOFF, Jérôme AZÉ & Mathieu ROCHE, éditeurs (2007). *Défi fouille de textes : reconnaissance automatique des auteurs de discours*. Numéro RNTI-E-10 in *Revue des Nouvelles Technologies de l'Information*. Cépaduès Éditions. ISBN : 978-2-85428-7.
url : <http://www.cepades.com/docs/776/RNTI-E-10.pdf>.

- [Puren, 1991] Christian PUREN (1991). Innovation et variation en didactique des langues. *Le français dans le monde*, (244):39–47. ISSN : 0015-9395.
- [Puren, 1994] Christian PUREN (1994). *La didactique des langues étrangères à la croisée des méthodes – essai sur l’éclectisme*. Essais. Crédif; Didier. ISBN : 2-278-04267-X.
url : http://www.ppp-esn.ch/tools/73056/files/PUREN_Essai_eclectisme.pdf.
- [Puren, 2002] Christian PUREN (2002). Perspectives actionnelles et perspectives culturelles en didactique des langues-cultures : vers une perspective co-actionnelle co-culturelle. *Les langues modernes*, (3):55–71.
url : <http://www.aplv-languesmodernes.org/spip.php?article844>.
- [Puren, 2006] Christian PUREN (2006). De l’approche communicative à la perspective actionnelle. Exposé à : *14^e congrès RANACLES* (Poitiers, 23-25 novembre 2006).
url : <http://uptv.univ-poitiers.fr/web/canal/61/theme/28/manif/125/video/1180/>.
- [Ramsden, 1992] Paul RAMSDEN (1992). *Learning to teach in higher education*. Routledge. ISBN : 0-415-06415-5.
- [Ranganathan, 1965] Shiyali Ramamrita RANGANATHAN (1965). *The colon classification*, volume 4 de *Rutgers series on systems for the Intellectual Organization of Information*. Rutgers University Graduate School of Library Service.
- [Ranguin, 1995] Josiane RANGUIN (1995). L’évaluation des application interactives multimédias : une approche résolument pragmatique. *Revue de l’EPI*, (78):151–157. ISSN : 1254-3985.
url : http://www.epi.asso.fr/fic_pdf/b78p151.pdf.
- [Rastier, 1999] François RASTIER (1999). *Eloquio del senso. Dialoghi semiotici per Paolo Fabbrì*, chapitre Dalla significazione al senso : per una semiotica senza ontologia, pages 213–240. Riscontri. Costa & Nolan. ISBN : 8876483748. Cité dans [Rastier, 2003].
- [Rastier, 2003] François RASTIER (2003). De la signification au sens. Pour une sémiotique sans ontologie (édition française de [Rastier, 1999]). *Texto*. ISSN : 1773-0120.
url : http://www.revue-texto.net/Inedits/Rastier/Rastier_Semiotique-ontologie.html.
- [Recker & Wiley, 2001] Mimi M. RECKER & David A. WILEY (2001). A non-authoritative educational metadata ontology for filtering and recommending learning objects. *Interactive Learning Environments*, 9(3):255–271. ISSN : 1049-4820.
url : <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=5848430&site=ehost-live>.
- [Rediker, 2008] Marcus REDIKER (2008). *Pirates de tous les pays – l’âge d’or de la piraterie atlantique (1716-1726)*. Libertalia, édition française de *Villains of All Nations – Atlantic pirates in the golden age* (traduction : Fred ALPI). ISBN : 978-2-952-8292-7-4.
- [Reeves & Marlino, 1994] Tom REEVES & Mary MARLINO (1994). Atelier consacré à l’évaluation des multimédias interactifs. Exposé à : *ED-MEDIA 94–World Conference on Educational Multimedia and Hypermedia* (Vancouver, 25-30 juin 1994). Cité dans [Ranguin, 1995].
- [Reffay & Chanier, 2005] Christophe REFFAY & Thierry CHANIER (2005). Approche minimaliste de la formation d’enseignants du supérieur au rôle de concepteur-tuteur de cours en ligne.

- url:<http://archive-edutice.ccsd.cnrs.fr/docs/00/03/94/49/PDF/ReffayChanierprepublie2005.pdf>.
- [Reffay *et al.*, 2008] Christophe REFFAY, Thierry CHANIER, Muriel NORAS & Marie-Laure BETBEDER (2008). Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche. *Sticef*, 15(EPAL). ISSN : 1764-7223.
url:http://sticef.univ-lemans.fr/num/vol2008/01-reffay/sticef_2008_reffay_01.htm.
- [Rézeau, 1999] Joseph RÉZEAU (1999). Analyse de computer-assisted language learning context and conceptualization. *ALSIC*, 2(1):61–68. ISSN : 1286-4986.
url:http://alsic.u-strasbg.fr/Num3/rezeau/alsic_n03-liv1.htm.
- [Rézeau, 2001] Joseph RÉZEAU (2001). *Médiatisation et médiation pédagogique dans un environnement multimédia*. Thèse de doctorat, Université Bordeaux 2.
url:<http://perso.wanadoo.fr/joseph.rezeau/recherche/thesePDF/TheseJosephRezeau.pdf>.
- [Riba, 2006] Patrick RIBA (2006). De l'approche communicative à l'approche actionnelle : apports du cadre européen commun de référence pour les langues. In *VIIe Séminaire national des professeurs de français - L'Évaluation en classe de FLE : quels modèles pour la Colombie ?*
url:<http://seminairebogota2006.googlepages.com/del%E2%80%9999approchecommunicative%C3%A0l%E2%80%9999approcheactionnelle>..
- [Rinck & Lebarbé, 2005] Fanny RINCK & Thomas LEBARBÉ (2005). Constitution et exploitation pédagogique de ressources linguistiques pour un enseignement / apprentissage du discours rapporté. In Geoffrey WILLIAMS, éditeur, *Actes des quatrièmes Journées de la Linguistique de Corpus, Texte et corpus*, (2):83–94. ISSN : 1958-5306.
url:http://web.univ-ubs.fr/corpus/jlc4/acteJLC2005_9_rinck.pdf.
- [Rinck & Tutin, 2007] Fanny RINCK & Agnès TUTIN (2007). Annoter la polyphonie dans les textes : le cas des passages entre guillemets. *Corpus*, (6):79–100. ISSN : 1765-3126.
url:<http://corpus.revues.org/index1102.html>.
- [Robert, 2004] Jean-Michel ROBERT (2004). Proximité linguistique et pédagogie des langues non maternelles. In *Accès aux langues proches et aux langues voisines, Études de Linguistique Appliquée*, 4(136):499–511. ISSN : 0071-190X.
url:<http://www.cairn.info/revue-ela-2004-4-p-499.htm>.
- [Robert, 2008] Jean-Pierre ROBERT (2008). *Dictionnaire pratique de didactique du FLE*. L'essentiel français. Ophrys. ISBN : 2708011979.
- [Roberts & Bosse-Andrieu, 2006] Roda P. ROBERTS & Jacqueline BOSSE-ANDRIEU (2006). Corpora and translation. In Lynne BOWKER, éditeur, *Lexicography, Terminology, and Translation - Text-based Studies in Honour of Ingrid Meyer*, Translation/Traduction, pages 201–214. Presses de l'Université d'Ottawa. ISBN : 9780776606279. Cité dans [Calzada Pérez, 2008].
- [Rodríguez, 2001] Ana RODRÍGUEZ SEARA (2001). L'évolution des méthodologies dans l'enseignement du français langue étrangère depuis la méthodologie traditionnelle jusqu'à nos jours. *Cuadernos del Marqués de San Adrián*, (1). ISSN : 1578-5696.
url:http://www.uned.es/ca-tudela/revista/n001/art_8.htm.

- [Rogerson-Revell, 2007] Pamela ROGERSON-REVELL (2007). Directions in e-learning tools and technologies and their relevance to online distance language education. *Open Learning*, 22(1):57–74. ISSN : 0268-0513.
url : <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=24078576&site=ehost-live>.
- [Rüschhoff, 1993] Bernd RÜSCHOFF (1993). Language learning and information technology: State of the art. *CALICO Journal*, 10(3):5–17. ISSN : 0742-7778.
url : <https://calico.org/a-552-Language%20Learning%20and%20Information%20Technology%20State%20of%20the%20Art%20.html>.
- [Sanz, 2003] Mercedes SANZ GIL (2003). *Las tecnologías de la información y de la comunicación y la autonomía de aprendizaje de lenguas. Análisis crítico y estudio de casos en el aprendizaje del FLE*. Thèse de doctorat, Universitat Jaume I. ISBN : 84-688-7842-1.
url : http://www.tdr.cesca.es/TDX-0628104-113234/index_cs.html.
- [Schmied, 1993] Josef SCHMIED (1993). Qualitative and quantitative research approaches to English relative constructions. In Clive SOUTER & Eric ATWELL, éditeurs, *Corpus-based Computational Linguistics: Papers Presented at the 12th Conference of the International Computer Archive of Modern English (ICAME)*, numéro 9 in Language and computers – Studies in practical linguistics, pages 85–97. Rodopi. ISBN : 90-5183-485-3.
- [Schuman & Presser, 1979] Howard SCHUMAN & Stanley PRESSER (1979). The open and closed question. *American Sociological Review*, 44(5):692–712. ISSN : 0003-1224.
url : <http://www.jstor.org/stable/2094521>.
- [SCORM, 2006a] SCORM (2006a). SCORM content aggregation model. Spécification SCORM 2004 3rd Edition Content Aggregation Model Version 1.0, Advance Distributed Learning.
url : http://tr.im/scorm2004_3.
- [SCORM, 2006b] SCORM (2006b). SCORM overview. Spécification SCORM 2004 3rd Edition Content Aggregation Model Version 1.0, Advance Distributed Learning.
url : http://tr.im/scorm2004_3.
- [SCTIC, 2002] Groupe de travail sur les normes et standards de la formation en ligne du SCTIC (2002). Les normes et standards de la formation en ligne - état des lieux et enjeux.
url : <http://www.profetic.org/file/norm-0210-d-RAPPORT.pdf>.
- [Seidlhofer, 2002] Barbara SEIDLHOFER (2002). Pedagogy and local learner corpora: working with learning-driven data. In Granger *et al.* [2002], pages 213–234. ISBN : 90-272-1701-7.
- [Selva & Chanier, 2000] Thierry SELVA & Thierry CHANIER (2000). Génération automatique d'activités lexicales dans le système alexia. *Sciences et Techniques Éducatives*, 7(2):385–412. ISSN : 1265-1338.
url : <http://edutice.archives-ouvertes.fr/edutice-00000282/fr/>.
- [Serna *et al.*, 2009] Audrey SERNA, Sébastien PINEL & Gaëlle CALVARY (2009). La plasticité des IHM en action : un exemple de téléprocédure plastique. In *proceedings of IHM'09, Actes de IHM'09, 21ème Conférence Francophone sur l'Interaction Homme-Machine* (Grenoble, 13-16 octobre 2009).

- [Sharoff *et al.*, 2008] Serge SHAROFF, Svitlana KURELLA & Anthony HARTLEY (2008). Seeking needles in the Web haystack: finding texts suitable for language learners. In TALC [2008], pages 268–273.
- [Shavelson & Stern, 1981] Richard J. SHAVELSON & Paula STERN (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51(4):455–498. ISSN : 0034-6543.
url : <http://www.jstor.org/stable/1170362>. Cité dans [Tochon, 1989].
- [Shepherd, 1987] John SHEPHERD (1987). *Computer assisted language learning and innovative efl methodology*. Ausburger I & I-Schriften Band 38, Augsburg. Universität Augsburg. ISBN : 3-923549-20-2. Cité dans [Rézeau, 2001].
- [Sheskin, 2004] David J. SHESKIN (2004). *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, 3rd edition. ISBN : 1-58488-440-1.
- [Simon, 1969] Herbert Alexander SIMON (1969). *The sciences of the artificial*. MIT Press. ISBN : 0262691914 (3e édition, 1996). Cité dans [Clark & Yinger, 1980].
- [Simon, 1997] Herbert Alexander SIMON (1997). *Administrative behavior, 4th edition*. Free Press. ISBN : 0-684-83582-7. Cité dans [Wikipedia, 2009c].
- [Sinclair, 1996] John SINCLAIR (1996). Preliminary recommendations on corpus typology. Document technique, EAGLES (Expert Advisory on Language Engineering standards).
url : <http://www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/corpus typ.ps.gz>.
- [Skinner, 1963] Burrhus Frederic SKINNER (1963). L'avenir des machines à enseigner. *Psychologie Française*, 8(3):170–180. ISSN : 0033-2984.
- [Spielmann, 1992] Guy SPIELMANN (1992). À la recherche de l'enseignement communicatif. *The French Review*, 65(6):908–918. ISSN : 0016-111X.
url : <http://www.jstor.org/stable/394641>.
- [Springer, 1996] Claude SPRINGER (1996). *La didactique des langues face aux défis de la formation des adultes*. Autoformation et Enseignement Multimedia. Ophrys. ISBN : 2-7080-0819-6.
- [St. John, 2001] Elke ST. JOHN (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning & Technology*, 5(3):185–203. ISSN : 1094-3501.
url : <http://llt.msu.edu/vol5num3/stjohn/>.
- [Starbuck, 2003] Gene H. STARBUCK (2003). College teaching styles by gender. Exposé à : *Western Social Science Association Annual Meeting* (Las Vegas, 9-12 avril 2003).
url : <http://www.mesastate.edu/~starbuck/CTSbG.pdf>.
- [Stevens, 2002] James STEVENS (2002). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, 4th edition. ISBN : 0-8058-3776-0.
- [Tagliante, 2006] Christine TAGLIANTE (2006). *La classe de langue*. Techniques de classe. CLE International, nouvelle édition. ISBN : 209-033068-6.
- [TALC, 2008] ASSOCIAÇÃO DE ESTUDOS E DE INVESTIGAÇÃO CIENTÍFICA DO ISLA-LISBOA, éditeur (2008). *Proceedings of the 8th teaching and language corpora conference (talc)*. Offset-mais Artes Gráficas, S.A. ISBN : 978-989-95523-1-9.

- [Tatsuki, 2006] Donna TATSUKI (2006). What is authenticity ? In Tim NEWFIELDS, Ian GLEDALL, Megumi KAWATE-MIERZEJEWSKA, Yvonne ISHIDA, Mark CHAPMAN & Peter ROSS, éditeurs, *Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference*, pages 1–12.
url : <http://www.jalt.org/pansig/2006/HTML/Tatsuki.htm>.
- [Taylor, 1994] David TAYLOR (1994). Inauthentic authenticity or authentic inauthenticity? *Teaching English as a Second or Foreign Language*, 1(2):A–1[1–12].
url : <http://www-writing.berkeley.edu/TESL-EJ/ej02/a.1.html>.
- [Terrell, 1977] Tracy D. TERRELL (1977). A natural approach to second language acquisition and learning. *The Modern Language Journal*, 61(7):325–337. ISSN : 0026-7902.
url : <http://www.jstor.org/stable/324551>.
- [Teubert, 2005] Wolfgang TEUBERT (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10(1):1–14. ISSN : 1384-6655.
- [Thevenin, 1999] David THEVENIN (1999). La plasticité en interaction homme-machine. In *Actes IHM99, Actes de 11ème Conférence Francophone Interaction Homme-Machine* (Montpellier, 22-26 novembre), pages 152–155.
- [Tochon, 1989] François TOCHON (1989). À quoi pensent les enseignants quand ils planifient leurs cours? *Revue Française de Pédagogie*, (86):23–33. ISSN : 0556-7807.
url : http://www.inrp.fr/publications/edition-electronique/revue-francaise-de-pedagogie/INRP_RF086_3.pdf.
- [Trim, 2002] John TRIM (2002). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer – guide pour les utilisateurs*, chapitre De la didactique aux didactiques : dialogue sur des enjeux éducatifs, pages 3–52. Division des Politiques Linguistiques.
url : <http://tr.im/CECRuser>.
- [Tsui, 2004] Amy Bik May TSUI (2004). What teachers have always wanted to know - and how corpora can help. In John McHardy SINCLAIR, éditeur, *How to Use Corpora in Language Teaching*, volume 12 de *Studies in Corpus Linguistics*, pages 39–61. John Benjamins Publishing Company.
url : <http://site.ebrary.com/pub/benjamins/Top?docID=10056604>.
- [Tyler, 1949] Ralph W. TYLER (1949). *Basic principles of curriculum and instruction*. University of Chicago Press, édition de 2003. ISBN : 0226820319. Cité dans [Tochon, 1989].
- [U-Montréal, 2002] Terminologie de base en sciences de l'information : volets 1 et 2 - indexation. (2002).
url : <http://www.ebsi.umontreal.ca/termino/00000133.htm>.
- [Underwood, 1984] John H. UNDERWOOD (1984). *Linguistics, computers, and the language teacher: A communicative approach*. Newbury House. ISBN : 0-88377-470-4.
- [Valcke, 1974] Louis VALCKE (1974). Le monisme épistémologique de la science contemporaine. *Philosophiques*, 1(2):3–13. ISSN : 0316-2923.
url : <http://id.erudit.org/iderudit/203010ar>.
- [van Ek & Trim, 1991] Jan Ate VAN EK & John TRIM (1991). *Threshold level 1990*. Council of Europe Press. ISBN : 92-871-2003-x.

- [Vidal *et al.*, 2004] Philippe VIDAL, Julien BROISIN, Erik DUVAL & Stefaan TERNIER (2004). Normalisation et standardisation des objets d'apprentissage : l'expérience ARIADNE. Actes de *miage et e-mi@ge* (Marrakech, 15-17 mars 2004), pages 48–64. ESG.
- [Viéville & de La Passadière, 2003] Claude VIÉVILLE & Brigitte DE LA PASSADIÈRE (2003). Méta-données pour les ressources éducatives et parcours de formation personnalisés. *In* Bruillard & de La Passadière [2003], pages 135–155. ISBN : 2-7462-0682-X.
- [Vigner, 2001] Gérard VIGNER (2001). *Enseigner le français comme langue seconde*. Didactique des langues étrangères. CLE International. ISBN : 2-09-033342-1.
- [Vitalis, 1993] André VITALIS (1993). La part de citoyenneté dans les usages. *In* André VITALIS, éditeur, *Médias et nouvelles technologies : pour une socio-politique des usages*, Médias et nouvelles technologies, pages 35–43. Apogée. ISBN : 2-909275-41-8.
- [Vivet, 1997] Martial VIVET (1997). *Préface*, pages 11–13. *In* Bruillard [1997]. ISBN : 2-86601-610-6.
url : http://www.stef.ens-cachan.fr/annur/bruillard/mae_somr.htm.
- [Warschauer, 1996] Mark WARSCHAUER (1996). Computer assisted language learning: an introduction. *In* Sandra FOTOS, éditeur, *Multimedia language teaching*, pages 3–20. Logos International. ISBN : 4947561041.
url : <http://www.ict4lt.org/en/warschauer.htm>.
- [Warschauer, 2000] Mark WARSCHAUER (2000). The death of cyberspace and the rebirth of CALL. *English Teachers' Journal*, (53):61–67. ISSN : 0333-533X.
url : http://www.gse.uci.edu/person/warschauer_m/cyberspace.html.
- [Warschauer & Healey, 1998] Mark WARSCHAUER & Deborah HEALEY (1998). Computers and language learning: an overview. *Language teaching*, 31:57–71. ISSN : 0261-4448.
url : http://www.gse.uci.edu/person/warschauer_m/overview.html.
- [Wauthy, 2008] Xavier WAUTHY (2008). No free lunch sur le web 2.0! ce que cache la gratuité apparente des réseaux sociaux numériques. *Regards économiques*, (59).
url : <http://www2.econ.ucl.ac.be/pphlogger/dlcount.php?id=regards&url=/Archives/RE059.pdf>.
- [Weinrich, 1989] Harald WEINRICH (1989). Les langues, les différences. *Le français dans le monde*, (289):49–56. ISSN : 0015-9395.
- [Wenger, 1987] Étienne WENGER (1987). *Artificial intelligence and tutoring systems: computational and cognitive approaches to the communication of knowledge*. Morgan Kaufmann Publishers Inc. ISBN : 0-934613-26-5. Cité dans [Zampa, 2003].
- [Widdowson, 1981] Henry George WIDDOWSON (1981). *Une approche communicative de l'enseignement des langues*. Langues et apprentissage des langues. Hatier-Didier, édition française de *Teaching language as communication* (traduction : Katsy BLAMONT & Gérard BLAMONT). ISBN : 2-278-05893-2.
- [Widdowson, 2003] Henry George WIDDOWSON (2003). *Defining issues in English language teaching*. Oxford applied linguistics. Oxford University Press. ISBN : 0-19-437445-9.

- [Widmann *et al.*, 2008] Johannes WIDMANN, Kurt KOHN & Ramon ZIAI (2008). The SACODEYL search tool - exploiting corpora for language learning purposes. *In* TALC [2008], pages 321–327.
- [Wikipedia, 2008a] Wikipédia (fr) (2008a). Accès direct.
url : http://fr.wikipedia.org/w/index.php?title=Acc%C3%A8s_direct&oldid=32099476.
- [Wikipedia, 2008b] Wikipédia (fr) (2008b). Filtrage par motif.
url : http://fr.wikipedia.org/w/index.php?title=Filtrage_par_motif&oldid=36574415.
- [Wikipedia, 2008c] Wikipedia (en) (2008c). Georgetown–IBM experiment.
url : http://en.wikipedia.org/w/index.php?title=Georgetown-IBM_experiment&oldid=221249154.
- [Wikipedia, 2008d] Wikipédia (fr) (2008d). International Standard Serial Number.
url : <http://fr.wikipedia.org/w/index.php?title=ISSN&oldid=36175777>.
- [Wikipedia, 2008e] Wikipedia (en) (2008e). Linguistic universal.
url : http://en.wikipedia.org/w/index.php?title=Linguistic_universal&oldid=235768666.
- [Wikipedia, 2008f] Wikipédia (fr) (2008f). National Science Foundation.
url : http://fr.wikipedia.org/w/index.php?title=National_Science_Foundation&oldid=34388855.
- [Wikipedia, 2008g] Wikipedia (en) (2008g). PILOT.
url : <http://en.wikipedia.org/w/index.php?title=PILOT&oldid=223528928>.
- [Wikipedia, 2008h] Wikipédia (fr) (2008h). Précision et rappel.
url : http://fr.wikipedia.org/w/index.php?title=Pr%C3%A9cision_et_rappel&oldid=34537117.
- [Wikipedia, 2009a] Wikipédia (fr) (2009a). Athénée (homonymie).
url : [http://fr.wikipedia.org/w/index.php?title=Ath%C3%A9n%C3%A9e_\(homonymie\)&oldid=40473156](http://fr.wikipedia.org/w/index.php?title=Ath%C3%A9n%C3%A9e_(homonymie)&oldid=40473156).
- [Wikipedia, 2009b] Wikipédia (fr) (2009b). Espace de noms.
url : http://fr.wikipedia.org/w/index.php?title=Espace_de_noms&oldid=40856346.
- [Wikipedia, 2009c] Wikipédia (fr) (2009c). Herbert Simon.
url : http://fr.wikipedia.org/w/index.php?title=Herbert_Simon&oldid=44146772.
- [Wikipedia, 2009d] Wikipédia (fr) (2009d). John Holloway.
url : http://fr.wikipedia.org/w/index.php?title=John_Holloway&oldid=43305787.
- [Wikipedia, 2009e] Wikipédia (fr) (2009e). Relation d'équivalence.
url : http://fr.wikipedia.org/w/index.php?title=Relation_d%27%C3%A9quivalence&oldid=41915397.
- [Wikipedia, 2009f] Wikipédia (fr) (2009f). Relation d'ordre.
url : http://fr.wikipedia.org/w/index.php?title=Relation_d%27ordre&oldid=45069334.

- [Wikipedia, 2009g] Wikipédia (fr) (2009g). Rich Text Format.
url : http://fr.wikipedia.org/w/index.php?title=Rich_Text_Format&oldid=39843344.
- [Wikipedia, 2009h] Wikipédia (fr) (2009h). Version d'un logiciel.
url : http://fr.wikipedia.org/w/index.php?title=Version_d%27un_logiciel&oldid=44393125#Version_admissible_ou_pre-release.
- [Wikiversity, 2008] James NEILL (2008). Advanced ANOVA/MANOVA.
url : http://en.wikiversity.org/w/index.php?title=Advanced_ANOVA/MANOVA&oldid=385133.
- [Wilkins, 1976] David WILKINS (1976). *Notional syllabuses*. Oxford University Press. Cité dans [Germain, 1993; Springer, 1996].
- [Willis, 2000] Jane WILLIS (2000). A holistic approach to task-based course design. *The Language Teacher Online*, 24(02). ISSN : 0289-7938.
url : <http://www.jalt-publications.org/tlt/articles/2000/02/willis>.
- [Wyatt, 1983] David H. WYATT (1983). Three major approaches to developing computer-assisted language learning materials for microcomputers. *CALICO Journal*, 1(2):34–38. ISSN : 0742-7778.
url : <https://www.calico.org/a-159-Three%20Major%20Approaches%20To%20Developing%20ComputerAssisted%20Language%20Learning%20Materials%20For%20Microcomputers.html>.
- [Yinger, 1978] Robert J. YINGER (1978). A study of teacher planning: Description and a model of preactive decision making. Exposé à : *Annual meeting of the American Educational Research Association* (Toronto, Canada, 27-31 mars 1978). Michigan State University, Institute for Research on Teaching.
url : <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED152747>.
- [Yinger, 1979] Robert J. YINGER (1979). Routines in teacher planning. *Theory Into Practice*, 18(3):163–169. ISSN : 0040-5841.
url : <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=5204010&site=ehost-live>.
- [Zahorik, 1975] John A. ZAHORIK (1975). Teachers' planning models. *Educational Leadership*, 33(2):134–139. ISSN : 0013-1784.
url : <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=7720946&site=ehost-live>.
- [Zampa, 2003] Virginie ZAMPA (2003). *Les outils dans l'enseignement : conception et expérimentation d'un prototype pour l'acquisition par expositions à des textes*. Thèse de doctorat, Université Pierre-Mendès-France Grenoble II.
url : <http://w3.u-grenoble3.fr/zampa/pdf/these.pdf>.
- [Zampa, 2005] Virginie ZAMPA (2005). Utilisation de l'analyse sémantique latente pour tenter d'optimiser l'acquisition par exposition à une langue étrangère de spécialité. *ALSIC*, 8(Numéro spécial TALAL):135–146. ISSN : 1286-4986.
url : <http://alsic.revues.org/index339.html>.

[Zeilinger-Trier, 2002] Manuela ZEILINGER-TRIER (2002). Analyse de *Fremdsprachenlernen in der Wissensgesellschaft – Zum Einsatz der Neuen Technologien in Schule und Unterricht*. *ALSIC*, 5(1):81–91. ISSN : 1286-4986.

url : http://alsic.u-strasbg.fr/Num8/zeilinger/alsic_n08-liv2.htm.

[Zock, 1996] Michael ZOCK (1996). Computational linguistics and its use in real world: the case of computer assisted-language learning. In *Proceedings of the 16th conference on Computational linguistics*, pages 1002–1004. Association for Computational Linguistics.

url : <http://www ldc.upenn.edu/acl/C/C96/C96-2171.pdf>.

Index

- AFNOR [1993], 10, 285
AGLS [2002], 119
AICC [2002], 95
AICC [2006], 95
Adamczewski & Gabilan [1993], 114
Ahmad *et al.* [1985], 53
Alcaraz & Pérez Paredes [2008], 79, 81, 85, 214
Algo-Lev [2008], 312
Allum [2001], 50, 51
Alonso [1994], 233, 234
Ambler [2005], 7, 9, 279
Amer-Yahia *et al.* [2006], 268
Angers [2000], 134
Antoniadis & Ponton [2002], 5, 53, 55, 68
Antoniadis & Ponton [2004], 55, 68
Antoniadis *et al.* [2004a], 55, 282
Antoniadis *et al.* [2004b], 55
Antoniadis *et al.* [2005a], 54, 55, 316
Antoniadis *et al.* [2005b], 5, 53–55, 57, 126, 255, 256, 281, 289
Antoniadis *et al.* [2007a], 303
Antoniadis *et al.* [2007b], 55
Antoniadis [2004], 11, 53–55
Antoniadis [2008], 31
Aston [1997], 76
Aston [2002], 71, 72, 76
Augustyn *et al.* [2008], 269
Balatsoukas *et al.* [2008], 104, 121, 128
Bangs & Cantos Gómez [2004], 49, 111
Bangs [2002], 37, 49, 55, 125
Bangs [2003], 51, 54
Bannert *et al.* [2003], 53
Bar-Hillel [1964], 1, 59
Bax [2003], 37, 38, 40, 41, 43, 44, 60, 64, 85
Beacco [1995], 25
Benjamin & Blunt [1992], 82
Bernard [1995], 32
Berners-Lee [1989], 39
Berners-Lee [1997], 93
Berners-Lee [2009], 317
Bertrand *et al.* [1996], 10
Besse [1987], 64, 66
Biber [1986], 69, 70
Binon *et al.* [1998], 201
Bizer *et al.* [2009], 317
Blanchard *et al.* [2009], 54, 59
Boch & Grossmann [2001], 29
Bouchard [1990], 66
Bouillon *et al.* [1998], 57
Boukottaya *et al.* [2009], 203
Boulidoires [1997], 49, 92
Boulidoires [1998], 61
Boulton [2007], 71
Boulton [2008], 74
Bourda [2002], 93, 128, 266
Bourdieu [2002], 217
Bourguignon [2007], 20
Braun *et al.* [2004], 77
Braun [2005], 75–78, 81, 127
Braun [2007], 76, 78, 85
Breen [1985], 66
Brett [1994], 41
Brodine [2001], 72
Bronckart & Plazaola Giger [1996], 167
Bruillard [1997], 30–33, 48
Bru [2002], 136
Bru [2004], 136
Burden [1981], 162, 193
Burnard [2001], 92
Bush [1945], 33
Bérard [1991], 19
CEN [2008], 91, 94, 95, 97, 98
CGal66 [2007], 159
CNTRL [2008], 56
Canale & Swain [1980], 228
Carol *et al.* [2004], 167
Carroll [1965], 21
Chambers [2007a], 71
Chambers [2007b], 79
Chanier [1995], 36, 41, 42

- Chappelle [2001], 36, 48
 Chappelle [2004], 214
 Charlier & Charlier [1998], 137
 Charlier [1989], 205, 206, 211
 Chastain & Woerdehoff [1968], 21
 Chatzissavidis [2003], 19
 Chevallard [1982], 16
 Chomsky [1959], 40
 Clanet [2001], 144
 Clanet [2007], 136
 Clark & Yinger [1979], 205, 221
 Clark & Yinger [1980], 144, 197, 203
 Cochran [1952], 175, 242, 243
 Cockburn [1997a], 7
 Cockburn [1997b], 7
 Cockburn [1997c], 7
 Cole & Foster [2008], 46
 Conseil de l'Europe [2000], 143, 160, 172, 225, 231
 Contamines *et al.* [2003], 61, 63, 92, 117, 121, 196, 197, 200, 212, 266
 Cros [1965], 48
 Cuq [1991], 7
 Cuq [2003], 7, 21, 22
 DCFI [2000], 96
 DCFI [2003], 96
 DCFI [2005], 120
 DCFI [2008], 95, 119
 Dabène [1975], 36
 Dabène [1989], 17
 Dabène [1994], 37
 Dabène [1995], 117
 Dargirolle [1999], 36, 137, 201, 203
 Decoo [1984], 49
 Delcloque & Bramoullé [2001], 55, 85
 Delcloque [2000], 38
 Demaizière & Dubuisson [1992], 31, 43, 48, 50, 51
 Depover *et al.* [2000], 46, 52, 57, 85, 125, 196
 Dijkstra [1971], 315
 Draelants & Maroy [2002], 160
 Duval *et al.* [2000], 97
 Duval *et al.* [2001], 97
 Duval *et al.* [2002], 93, 96
 Duval [2004], 88, 91
 Ebbrell [1998], 45
 Ellis *et al.* [1991], 33
 Englebart [1970], 39
 Evans & Lefferts [1995], 84
 Everett [2007], 37
 Fairon [2006], 308
 Feynman [2006], 14
 Flowerdew [1993], 62, 71–73
 Francis [1967], 70
 Friesen *et al.* [2003], 120
 Fuchs *et al.* [1993], 11
 Fuller [1969], 161, 162
 GEM [2001], 96
 GEM [2004a], 97, 120
 GEM [2004b], 120
 GEM [2004c], 142
 Gamper & Knapp [2002], 11, 64
 Germain [1993], 18, 19, 21–23, 25, 65, 143, 163
 Ghiglione & Matalon [1978], 134, 139, 149–152, 219
 Godwin-Jones [2000], 50
 Goodfellow *et al.* [1999], 43, 44
 Goodwin & Stevens [1993], 141
 Grandbastien & Labat [2006], 34, 124
 Granger *et al.* [2007], 70, 75
 Granger [2002], 75
 Grasha [2002], 141
 Guidon [2001], 90
 Guilbaud [1980], 271
 Géry [2002], 12–14, 282, 285
 Habert *et al.* [1997], 68–70, 72
 Habert *et al.* [1998], 68, 69
 Hakstian *et al.* [1979], 191
 Harris [2000], 67
 Hart [1995], 39, 40
 Hashweh [1987], 203
 Heift & Schulze [2007], 36, 55, 58
 Herreras [1998], 141
 Herreras [2001], 141
 Higgins & Johns [1984], 41
 Higgins [1986], 63
 Hillmann [2005], 96
 Hodgins & Duval [2004], 100
 Holloway [2002b], 196, 198
 Holloway [2007], 196–198, 200, 202, 213, 214, 253
 Houssaye [1988], 15, 32
 Hubbard [1987], 37, 48, 49
 Hubbard [1996], 48, 59
 Huberman *et al.* [1989], 192
 Huberman [1989a], 161, 162, 194

- Huberman [1989b], 162
 Huberty & Olejnik [2006], 191
 Hutchinson & Waters [1987], 66, 214
 Hymes [1984], 18, 227, 228
 IEEE [2002], 98, 99, 106, 110, 111, 113, 116, 118, 128, 142, 266
 IMS [2006], 97
 INSEE [2008a], 92
 INSEE [2008b], 92
 ISO [2004], 89
 ISO [2007], 98
 Jackson [1968], 137
 Johns [1986], 62, 71–73, 75, 81, 202
 Johns [1991], 71
 Johns [2002], 71–74, 127, 226
 Jones [1986], 40, 63
 Jones [1995], 38, 40
 Jonnaert & Laurin [2001a], 17
 Jung [2005], 40, 41, 67
 Kaltenböck & Mehlmauer-Larcher [2005], 73, 74
 Katz [1972], 161, 162, 193
 Kern *et al.* [2004], 214
 Kirby *et al.* [2004], 37
 Koper [2000], 103
 Koteyko [2006], 70
 Kraif & Ponton [2007], 11, 59, 316
 Kraif *et al.* [2004], 55, 59
 Krashen [1982], 22
 Krüger [1981], 167
 Kurella *et al.* [2008], 308
 Kübler [2002], 273
 Labov [1976], 136
 Lamarre [2003], 161, 162, 194
 Lamy & Klarskov Mortensen [2007], 72
 Larose *et al.* [2002], 134
 Laurin & Gaudreau [2001], 17
 LeBlanc [1989], 43
 Lebarbé & Meynard [2008], 92
 Lebarbé [2002], 12
 Lecocq [2008], 155
 Lefèvre [2000], 12, 87, 133, 187, 197, 285
 Lemieux [1964], 22, 29
 Levin *et al.* [1991], 83, 84
 Levy & Kennedy [2004], 43
 Levy & Stockwell [2006], 44–47, 316
 Levy [1997], 18–20, 23, 24, 26, 35, 36, 39, 41, 84, 124, 142, 163, 166
 Lin Hsiao [1996], 33
 Little & Singleton [1988], 64, 65, 83
 Little [1997], 44
 Li [2003], 11
 Loiseau *et al.* [2005], 15
 Loiseau *et al.* [2008a], 253
 Loiseau *et al.* [2008b], 253
 Loiseau [2003], 6, 52, 110, 133, 134, 139, 143, 146, 147, 207, 232, 234
 Loiseau [2004], 96
 Loiseau [2008], 253
 Love & Nader [1997], 90, 91
 Lowe [1999], 97
 Luhn [1960], 61, 62, 80
 Lusalisa *et al.* [2000], 134
 Lynch [1997], 93
 Lévy [1997], 14
 MEDEF & CCIP [2004], 138
 MEN [2003], 157
 MEN [2004], 157
 MEN [2005], 157
 MEN [2006], 157
 MEN [2007], 157
 Mackey [1997], 7
 Mangenot [2002], 46
 Mangenot [2003], 134
 Mangeot-Lerebours *et al.* [2003], 56
 Manning & Schütze [1999], 61
 Marcus [1999], 236
 Martinez [1996], 25
 Matthey [2003], 7
 McEnery & Wilson [1997], 71
 McEnery & Wilson [2001], 70
 Meier [1992], 236
 Moirand [1982], 19, 113, 228
 Mossel [1960], 24
 Mucchielli [1990], 151
 Murray & Barnes [1998], 52, 55, 59, 60, 67, 125
 Myers [1986], 312
 Netter [2008], 83
 Nevins *et al.* [2007], 37
 Nicola & van der Linden [2005], 314
 Nunan [1989], 65, 67
 Nunan [2006], 19, 20
 Oxford [1993], 42
 Oxford [2006], 19
 PHP [2009], 149
 Pallotti [2002], 36

- Palmer *et al.* [1981], 228
 Palmer [1922], 21
 Panchenko [2009], 314, 315
 Papert [1981], 32
 Paroubek *et al.* [1999], 11
 Partington [1998], 74
 Patesson [2007], 135
 Pernin [2003], 92, 98, 101, 102, 104, 118, 128
 Pernin [2004], 98, 117, 118, 128
 Pernin [2006], 88–90, 95
 Polguère [2002], 56
 Polguère [2003], 56
 Poussard [2000], 43
 Pravec [2002], 75
 Prince *et al.* [2007], 268
 Puren [1991], 25
 Puren [1994], 233, 317
 Puren [2002], 19
 Puren [2006], 19, 20
 Ramsden [1992], 17, 136
 Ranganathan [1965], 259, 261
 Ranguin [1995], 48
 Recker & Wiley [2001], 121, 129, 196, 197, 212, 247, 249, 254, 255
 Rediker [2008], 80
 Reeves & Marlino [1994], 48
 Reffay & Chanier [2005], 134
 Reffay *et al.* [2008], 214
 Riba [2006], 19
 Rinck & Lebarbé [2005], 269, 303
 Rinck & Tutin [2007], 303
 Roberts & Bosse-Andrieu [2006], 82
 Robert [2004], 158, 192, 243
 Robert [2008], 221
 Rodríguez [2001], 23
 Rogerson-Revell [2007], 43–46, 51, 85
 Rüschoff [1993], 12, 255, 256
 Rézeau [1999], 38
 Rézeau [2001], 15–17, 25, 31, 36, 53, 59, 60, 123, 124, 256
 SCORM [2006a], 98, 101–103
 SCORM [2006b], 101
 SCTIC [2002], 88, 89, 101
 Sanz [2003], 65
 Schmied [1993], 70
 Schuman & Presser [1979], 140
 Seidlhofer [2002], 75
 Selva & Chanier [2000], 55–57, 59
 Serna *et al.* [2009], 312
 Sharoff *et al.* [2008], 83, 127
 Shavelson & Stern [1981], 187, 205, 221
 Shepherd [1987], 60
 Sheskin [2004], 165, 174, 242
 Simon [1969], 197
 Simon [1997], 197
 Sinclair [1996], 68, 69, 276–278
 Skinner [1963], 49
 Spielmann [1992], 26
 Springer [1996], 25
 St. John [2001], 72
 Starbuck [2003], 141, 159
 Stevens [2002], 189–191
 Tagliante [2006], 202, 233–235, 310, 316
 Tatsuki [2006], 66
 Taylor [1994], 66, 73, 74, 214
 Terrell [1977], 23
 Teubert [2005], 70
 Thevenin [1999], 312
 Tochon [1989], 137, 145, 187, 203, 205, 221
 Trim [2002], 163
 Tsui [2004], 16, 73
 Tyler [1949], 205
 U-Montréal [2002], 10
 Underwood [1984], 40–42
 Valcke [1974], 110
 Vidal *et al.* [2004], 88, 90
 Vigner [2001], 65, 67, 110, 134, 171, 210
 Vitalis [1993], 61, 84
 Vivet [1997], 34
 Viéville & de La Passadière [2003], 113
 Warschauer & Healey [1998], 37, 38, 40, 41, 43, 85, 124
 Warschauer [1996], 37–43, 46, 85
 Warschauer [2000], 37, 38, 47, 85
 Wauthy [2008], 8, 9, 122
 Weinrich [1989], 36
 Wenger [1987], 32
 Widdowson [1981], 67, 112, 210
 Widdowson [2003], 73
 Widmann *et al.* [2008], 77, 78, 85
 Wikipedia [2008a], 39
 Wikipedia [2008b], 83
 Wikipedia [2008c], 59
 Wikipedia [2008d], 10
 Wikipedia [2008e], 37
 Wikipedia [2008f], 38

- Wikipedia [2008g], 41
 Wikipedia [2008h], 11
 Wikipedia [2009a], 160
 Wikipedia [2009b], 119
 Wikipedia [2009c], 197
 Wikipedia [2009d], 197
 Wikipedia [2009e], 264
 Wikipedia [2009f], 264
 Wikipedia [2009g], 149
 Wikipedia [2009h], 253, 301
 Wikiversity [2008], 193
 Wilkins [1976], 25
 Willis [2000], 20
 Wyatt [1983], 41, 45
 Yinger [1978], 137, 205, 206, 208–210, 214, 215, 250, 253, 258
 Yinger [1979], 194, 200, 203, 258
 Zahorik [1975], 162, 193
 Zampa [2003], 32
 Zampa [2005], 309
 Zeilinger-Trier [2002], 41
 Zock [1996], 57, 64, 311
 de Montmollin [1975], 31
 de Nuchèze [1998], 29
 de Singly [1992], 139, 140, 153
 de Vries & Baillé [2006], 31
 de Vries [2001], 32–34
 edna [2002], 96
 edna [2006], 96, 119, 142, 266
 Gómez Casañ & Martín Viaño [1990], 232, 233
 Álvarez [2007], 82
 van Ek & Trim [1991], 172, 225, 231
- Activité pédagogique, 6, 57, 71–74, 77, 104, 114, 116, 118, 133, 282
 Apport (*input*), 22, 24, 42, 43, 66, 67, 73, 82
 Approche instrumentale, 61, 63, 84, 126–128, 196, 203
 Artefact, 61, 71, 76, 196
 Champ instrumental, 62, 75, 84, 128, 133
 Fonctions constituantes, 35, 61, 73, 84, 118, 203, 266
 Fonctions constituées, 61, 64, 84
 Genèse instrumentale, 61–63, 71, 127, 196, 200
 Instrumentalisation, 61, 62, 64, 71, 73, 75, 76, 84, 127, 273, 308
 Instrumentation, 61, 62, 64, 70, 71, 75, 84, 127, 273, 308, 311
 Instrument, 61, 64, 196
 Aptitude, 21, 22, 24, 43
 Authenticité (*Texte authentique*, voir index pour *Texte*), 64–67, 73–75, 81, 126, 172, 201, 214, 308
- Collection de textes, 69, 70, 82, 126, 275, 277, 278
 (Entité d'un système), 275–278, 280, 281, 289–291, 295, 314
 Concordancier, 61, 62, 64, 71–73, 78, 81
 Concordances, 61, 62, 72–74, 76, 77, 81, 82
 KWIC, 61, 62, 76
 Contexte pédagogique, a-140, 137, 195, 210–215, 217–219, 230–232, 245–250, 253–267, 271, 272, 275–280, 282, 284–287, 289–291, 293, 294, 296, 297, 299, 301–304, 307, 312, 313, 315, 358
 Corpus, 5, 12, 13, 15, 16, 26, 30, 68–77, 80–82, 84–86, 126, 127, 277, 278, 304
 alignés, 72
 annotés, 69
 d'apprenants, 75
 dans l'enseignement des langues, 68
 pédagogique, 45, 77–79, 81, 82, 85, 86, 127
 annotation pédagogique, 77–79, 85, 86
 sous-corpus, 76
- Distribution normale (ou Gaussienne), 165
 EIAH, 29, 30, 34, 35, 37, 38, 46–49, 63, 84–86, 124–127, 317
 ALAO, 5, 6, 18, a-140, 29, 30, 35–44, 46–50, 52, 53, 55, 57–61, 63, 64, 67, 69, 84, 85, 111, 124–127, 272, 273, 302, 316, 358
 Point de vue communicatif, 40–44, 46, 47, 124
 Point de vue intégratif, 41–47, 77, 78, 85, 124, 125, 316, 317
 Point de vue structural, 38, 41, 46, 47, 55, 124
 Reconstruction de textes, 41, 43
 Travers, 109
 Enseignement programmé, 30, 31, 38
 Programme, 31
 Environnements d'apprentissage collaboratifs, 33

- Environnements d'apprentissage ouverts, 32, 43, 45
- Feed-back, 38, 55, 59, 75, 126
- Hybridation, 46, 125, 316
- Hypermédia, 32, 33, 38, 43, 46
- Hypertexte, 32, 33
- L'ordinateur comme outil, 32, 43, 45, 46, 63
- Langage auteur, 39, 41
- Machines à enseigner, 30, 31, 233
- Micromondes, 32
- SGA, 6, 45, 46, 85, 86, 103, 105, 125
- Simulations, 32, 43
- Système auteur, 41, 45, 55, 83
- Tuteurs intelligents, 32, 41, 43, 83
- Exercice structural, 114, 233
- Faire*, 196–198, 200–204, 208, 210, 213, 214, 247, 253
- Flux du *Faire*, 196–198, 200–204, 206, 208, 213, 214, 247
- Flexion, 56, 83, 259
- Affixe flexionnel, 56
- Catégorie flexionnelle, 56
- Forme fléchie, 56, 57, 83, 233, 237
- Radical, 56
- Forme canonique, 56
- Langue
- 1 (L1), 7, 22, 23, 36, 37, 65, 73, 75, 117, 134, 141, 158, 159, 161, 166, 218, 237, 243–245, 248
- 2 (L2), 6, 7, 16, 21, 22, 24, 36, 37, 41, 42, 44, 49, 65–67, 70, 74, 75, 78, 138, 141, 144, 154, 155, 157, 158, 166, 188, 189, 193, 195, 197, 213, 218, 237, 244, 245
- De spécialité, 66, 71, 207
- Étrangère, 7, 75, 155, 166
- Maternelle, 7
- Officielle, 75, 166
- Seconde, 7, 75, 155, 166, 207
- Xénité, 36, 117, 158, 237, 243
- Lemme, 56, 57, 233, 234, 259, 267, 300
- Lexème, 56
- Lexie, 56, 233
- Littérature informatique, 133, 134, 138, 139, 142, 148, 150, 188
- Manuel, a-119–a-121, 167, 168, 170, 171, 178–182, 188, 189, 202, 203
- Métadonnées, 93, 98, 102, 105, 118, 128, 201, 266–268, 276, 284
- Méthode scientifique, 110
- Méthodes et approches didactiques
- Approche communicative, 18–20, 24, 25, 39–41, 43, 44, 46, 47, 64, 65, 74, 81, 112, 113, 124, 143, 145, 163, 164, 172, 227, 239
- Compétence de communication, 18, 19, 43
- Approche naturelle, 22, 24, 164
- Approche notionnelle-fonctionnelle, 21, 24, 25, 164
- Cognitive code-learning theory*, 21, 163
- Data Driven Learning* (DDL) (apprentissage à partir des données), 71, 73, 74, 127, 235, 303
- Éclectisme, 21, 25, 124, 138, 163, 164, 188, 195, 213, 230, 233, 254, 317
- Grammaire Formelle*, 20, 22, 24, 163
- Grammaire-Traduction, 24, 163
- Méthode audio-orale, 23, 24, 46, 47, 163, 164, 233
- Méthode directe, 21–24, 164
- Méthode grammaire-traduction, 21, 22, 164
- Méthode orale, 21, 163, 168
- Méthode par le mouvement, 24, 164
- Méthode situationnelle, 21, 24, 164
- Task-based approach*, 19, 20, 163
- Perspective actionnelle, 19, 20, 24, 43, 46, 47, 81, 143, 164
- Tâche, 19, 20, 43, 77, 163, 169
- Modèle
- Facette, 253, 256–273, 275–278, 283–287, 289, 291, 292, 294–296, 299, 300, 302–304, 309–311, 315, 317
- constante, 258, 266, 268, 285, 286, 296, 299
- instance, 260, 261, 267, 276, 295
- vide, 266, 276
- Prisme, 253, 256–266, 268, 269, 271–273, 275–278, 280–292, 294, 295, 299, 302–304, 308, 310–315, 317
- Visualisation, 270–272, 275, 280, 284, 285, 287–289, 291, 311
- Mot-forme, 56, 233, 234, 267, 300

- Niveau d'agrégation, 99, 102, 104, 105, 121, 128, 222
- Normalisation, 86–93, 95, 98, 101, 105, 111, 127
 Norme, 10, 90, 91, 105, 107, 127, 128
 Profil d'application, 93–95, 97, 102, 105, 119, 120, 122, 128, 133, 266
 Recommandation, 89–91
 Standard, 90, 92
 Standard de fait, 90, 91
- Objet pédagogique, 98, 99, 101–108, 110–113, 115–117, 120–122, 128, 169, 198, 207, 288
- Pratiques, 134–136
 constatées, 136, 310
 déclarées, 133–137, 143
 effectives, 135, 136
- Propriété
 extrinsèque, 121, 122, 129, 187, 196, 211, 214, 236, 247, 249, 254
 intrinsèque, 118, 121, 122, 129, 172, 187, 211, 214, 231, 236, 246, 247, 249, 254
- Recherche d'Information, 285
 Indexation, 10, 12–15, 17, 61, 69, 80, 83, 84, 87, 99, 104, 123, 246, 255, 267, 268, 270, 281, 282, 285, 287, 293
 pédagogique, 5, 9, 12, 15, 17, 18, 26, a-140, 29, 30, 57, 63, 64, 67, 69, 77, 79–81, 84, 86, 87, 108, 121, 123, 124, 127, 133, 138, 140, 146, 187, 188, 196, 212, 249, 253, 254, 261–263, 267, 270, 273, 275, 276, 281, 285–287, 291, 298, 304, 307–309, 311, 315, 358
 Langage Documentaire, 10, 12–15, 26, 87, 108, 124, 133, 285
 Modèle de recherche d'information, 12–14, 26, 282, 290
 Fonction de correspondance, 13, 14
 Modèle de documents, 13
 Modèle de requête, 13, 14
 Précision, 11, 59
 Rappel, 11, 59
 Système de recherche d'information, 12–14, 282
- Ressource brute, 82, 99, 109, 112, 114, 117, 118, 121, 128, 133, 254
- Standard de description de ressources pédagogiques, 94, 98, 104, 120, 121, 127, 128, 133, 135, 172, 187, 205, 254
 AICC, 94, 101
 ARIADNE, 97, 105
 CEN-LTSO, 94
 Dublin Core
 Element Set (DCES), 96, 97, 105, 118–120
 Metadata Initiative (DCMI), 95–98, 118, 120, 128
 edna, 96, 118–120, 128, 142
 Éléments, 119, 142, 266
 EML, 98, 103, 128
 GEM, 96, 118, 120, 128, 142
 Éléments, 120, 142
 IMS, 97, 98, 101–103, 105, 128
 LOM, 92, 95–99, 101–110, 112, 113, 115–121, 128, 142, 146
 Attributs, 106
 Éléments, 100, 106–110, 112–118, 128, 142, 266
Smallest permitted maximum, 106, 110, 115
 Types, 107, 118
 SCORM, 46, 98, 101–103, 105, 128
- Stratégie moins-disante, 9, 11, 12, 59, 60, 64, 126, 211, 250, 256, 270, 275
- Systèmes ALAO
 ALEXIA, 56, 57, 59
 ALICE, 83, 84
Learn Welsh, 45
 ClicNet, 83
 ELISA, 77, 78
 Hot Potatoes, 45, 46
 MIRTO, 5, 6, 17, 35, 47, 57, 68, 84, 114, 115, 123, 126, 281, 283, 287, 289, 291
 Moodle, 46
 PLATO, 38–40
 SACODEYL, 77–82
storyboard, 41
 TexToys, 46
 TICCIT, 38–40
- TAL, 5, 53, 55–59, 61, 63, 64, 68–70, 85, 122, 126, 250, 255, 256, 270, 273, 282
 Analyse Morphologique, 11, 56, 57, 83, 248, 259, 286, 312

Texte

- authentique, 19, 20, 23, 24, a-13, 64–68, 109, 112, 126, 127, 134, 143, 151, 167–171, 173, 174, 181, 188, 201, 255
- fabriqué, a-13, 65, 67, 113, 134, 143, 146, 148, 151, 155, 167–174, 180, 181, 186, 188, 202–204, 207–209, 211, 237, 308
- pédagogique, 167
- recherche passive, 134, 135, 144–146, 173–175, 177, 181–184, 188, 189, 212, 213, 215
- recueil, 6
- tout-venant, 5, 255, 307, 308

Théorie sens-texte, 56

UML, 7–9, 278, 287, 291

Diagramme

- Cas d'utilisation, 7–9
- Classes, 291, 293, 297
- Séquence, 278

Usage, 114, 117

Vocabulaire informatique, 14, 110, 113, 115, 116, 118

Résumé

Cette thèse traite la notion d'indexation pédagogique et l'aborde sous l'angle de la recherche et du choix de textes pour l'enseignement des langues. Ce problème est replacé dans le champ disciplinaire de l'Apprentissage des Langues Assisté par Ordinateur (ALAO) et des apports potentiels du Traitement Automatique des Langues (TAL) à cette discipline, avant d'être confronté à des éléments provenant plus directement de la didactique des langues, pour proposer une approche empirique. Celle-ci est ensuite justifiée par l'incapacité des standards de description de ressources pédagogiques actuels à modéliser des ressources brutes de manière cohérente, en particulier des textes pour l'enseignement des langues. La thèse s'articule ensuite autour de deux questionnaires visant à connaître les pratiques déclarées des enseignants quant à la recherche et au choix de textes dans le cadre de la tâche de planification des cours. Le premier questionnaire permet la formalisation de la notion de contexte pédagogique, qui est ultérieurement appréhendée à travers certaines de ses composantes grâce au second questionnaire. Enfin, ces premières ébauches de formalisation servent de fondement à la définition d'un modèle dont l'objectif est de rendre compte de la contextualité des propriétés dites pédagogiques quand elles sont appliquées à des ressources brutes. Des pistes d'implantation du modèle sont finalement proposées dans le cadre de la description d'un système informatique.

Mots-clés: Indexation pédagogique, ALAO, EIAH, Didactique des langues, Métadonnées, TAL, Modélisation, Corpus

Abstract

This PhD thesis deals with the notion of pedagogical indexation and tackles it from the point of view of searching for and selecting texts for language teaching. This particular problem is set in the field of Computer Assisted Language Learning (CALL) and of the potential contribution of Natural Language Processing (NLP) to this discipline, before being considered within the scope of elements more directly relevant to language didactics, in order to propose an empirical approach. The latter is then justified by the inadequacy of current description standards for pedagogical resources where modeling of raw objects in a consistent fashion is concerned. This is particularly true for texts in the context of language learning. The thesis subsequently revolves around two questionnaires the aim of which is to provide insight into language teachers' declared practices regarding searching for and selecting texts in the context of class planning. The first questionnaire provides data to formalize the notion of pedagogical context, which is later considered through some of its components thanks to the second questionnaire. Finally, these first formalization drafts provide foundations for the definition of a model aiming at taking into account the contextuality of the properties said to be pedagogical, which is inherent to raw resources. Finally, possible leads for implementing this model are suggested through the description of a computerized system.

Keywords: Pedagogical indexation, CALL, Language didactics, Metadata, NLP, Modeling, Corpus

Élaboration d'un modèle pour une base de textes indexée pédagogiquement pour l'enseignement des langues

Volume 2 - Annexes

THÈSE

présentée et soutenue publiquement le 9 décembre 2009

pour l'obtention du

Doctorat de l'Université Stendhal Grenoble 3

(Spécialité Industries de la langue)

par

Mathieu Loiseau

Composition du jury

<i>Directeurs :</i>	Georges ANTONIADIS	LIDILEM (Université Stendhal Grenoble 3)
	Claude PONTON	LIDILEM (Université Stendhal Grenoble 3)
<i>Rapporteurs :</i>	Éric BRUILLARD	UMR STEF (ENS Cachan) – INRP
	Natalie KÜBLER	CLILLAC (Université Paris Diderot - Paris 7)
<i>Examineur :</i>	Cédrick FAIRON	CENTAL (Université Catholique de Louvain)

Mis en page avec la classe thloria.

Annexes

Annexe A

Questionnaire 1 : Choix et utilisation de ressources textuelles en cours de langues

A.1 Votre enseignement

A.1.1 La ou les langue(s) que vous enseignez :

A.1.2 La ou les langue(s) maternelle(s) de vos apprenants :

A.1.3 Le nombre d'années durant lequel vous avez enseigné les langues :

A.1.4 Le cadre dans lequel vous l'enseignez (collège, lycée, université, organisme privé) :

A.1.5 Le pays dans lequel vous l'enseignez

A.1.6 Le public auquel vous l'enseignez (âge, niveau) :

A.1.7 Parmi les méthodes, philosophies et approches d'enseignement suivantes, cochez les cases correspondant à celles qui décrivent le mieux votre manière de travailler

- | | |
|---|---|
| <input type="checkbox"/> Approche axée sur la compréhension | <input type="checkbox"/> Méthode directe |
| <input type="checkbox"/> Approche communicative | <input type="checkbox"/> Méthode par le mouvement |
| <input type="checkbox"/> Approche fonctionnelle - notionnelle | <input type="checkbox"/> Méthode par le silence |
| <input type="checkbox"/> Approche intégrative | <input type="checkbox"/> Méthode situationnelle |
| <input type="checkbox"/> Approche naturelle | <input type="checkbox"/> Méthode suggestopédique |
| <input type="checkbox"/> Audio-orale | <input type="checkbox"/> SGAV |
| <input type="checkbox"/> Curriculum multidimensionnel | <input type="checkbox"/> Autre : _____ |
| <input type="checkbox"/> Grammaire traduction | _____ |
| <input type="checkbox"/> Méthode communautaire | _____ |
| <input type="checkbox"/> Méthode des séries | |

A.2 Utilisation des textes dans le cadre de vos cours :

Par texte on entend toute ressource textuelle : aussi bien un texte dans sa forme originale qu'un texte modifié pour les besoins d'une activité, ou même un court extrait de texte.

A.2.1 Utilisez-vous des textes authentiques ?

- Toujours
 Le plus souvent
 Parfois
 Jamais

A.2.2 Quand vous utilisez un texte pédagogique (non authentique) vous :

- l'écrivez
 le recherchez
 l'un ou l'autre, selon le contexte

A.2.3 Dans quels cas privilégiez-vous le texte pédagogique sur le texte authentique ?

A.2.4 Avez-vous un recueil de textes personnel ?

- Oui
- Non

a) Si oui, ce recueil est-il organisé ?

- Oui
- Non

b) Si oui, comment ?

A.2.5 Les textes dont vous disposez pour vos cours proviennent de (trier les propositions par ordre de fréquence) :

- Recherches pour une activité précise
 - Textes repérés au cours de lectures personnelles
 - Imposés par le programme
 - Autre : _____
-
-

A.2.6 Lorsque vous recherchez un texte pour votre cours vous le cherchez (trier les propositions par ordre de fréquence) :

- Dans des revues / journaux
 - Dans des ouvrages littéraires
 - Dans des méthodes
 - Sur Internet
 - Autre : _____
-
-

A.2.7 Vous arrive-t-il de trouver « par hasard » un texte que vous décidez de garder pour vos cours ?

- Oui
- Non

a) **Si oui, vous le gardez :**

- Pour une activité précise
- Pour un public précis
- En attendant de trouver une manière adéquate de l'utiliser
- Si le texte est bon il sera utilisable dans de nombreux contextes radicalement différents

A.2.8 Pensez-vous qu'un texte donné puisse être utilisé dans plusieurs contextes différents ? (ex : avec un public de niveau différent / pour un autre type d'activité ou pour une activité portant sur une notion différente, etc...)

- Oui
- Non

a) **Cela vous est-il déjà arrivé ? (d'utiliser un même texte dans plusieurs contextes différents)**

- Oui
- Non

A.2.9 Quand vous recherchez un texte (même au sein de votre recueil personnel), vous effectuez votre recherche selon quels critères (plusieurs réponses possibles) :

- Le but pédagogique de l'activité dans laquelle le texte sera utilisé
- Le *modus operandi* de l'activité dans laquelle le texte sera utilisé
- Le public auquel est destiné le texte
- L'intérêt du texte indépendamment du contexte scolaire
- Vous recherchez un moyen d'utiliser un texte donné plus qu'un texte pour une activité donnée
- Autre : _____

A.3 Remarques

Une version électronique (rtf) de ce questionnaire est disponible en ligne :

<http://mathieu.loiseau.free.fr/QBaseDeTextes1.rtf>¹

Merci de remplir ce questionnaire (électroniquement ou sur papier) et de le retourner avant le 20 décembre à : mathieu.loiseau@u-grenoble3.fr

ou

Mathieu LOISEAU

Département d'Informatique Pédagogique

DIP – BP 25 – 38040 Grenoble Cedex 9

France

1. Depuis le document a été déplacé vers <http://mathieu.loiseau.free.fr/bdtp/q1/QBaseDeTextes1.rtf>.

En cas de questions ou remarques, notamment sur la clarté des questions ou les choix de réponses proposés, n'hésitez pas à les joindre au questionnaire, me les envoyer directement par e-mail ou me contacter au 04 76 82 77 63 (+33 4 76 82 77 63).

Si vous souhaitez recevoir des informations sur l'avancement du projet il suffit de me le spécifier. Enfin si vous êtes prêt(e) à nous aider dans nos recherches concernant l'apprentissage des langues assisté par ordinateur en nous faisant bénéficier de votre expérience de l'enseignement, merci de ne surtout pas hésiter à me le faire savoir (et à me transmettre vos coordonnées : une adresse e-mail par exemple).

A.4 Suivi du projet

Les questions suivantes remplacent la partie Remarques dans la version électronique du questionnaire :

Votre adresse e-mail : _____
(à remplir si vous cochez l'une des cases ci-dessous)

A.4.1 Cochez les cases suivantes si :

- Vous souhaitez recevoir des informations sur l'avancement du projet
- Vous êtes prêt(e) à nous aider dans nos recherches concernant l'apprentissage des langues assisté par ordinateur en répondant à d'éventuels questionnaires.

A.4.2 Si vous avez des remarques sur le questionnaire :

(questions dont l'énoncé n'est pas clair, auxquelles vous n'avez pas pu répondre comme vous l'auriez aimé, ou autres commentaires)

A.4.3 Remarques

Une version utilisable "*hors-ligne*" (rtf) de ce questionnaire est disponible à :

<http://mathieu.loiseau.free.fr/bdtp/QBaseDeTextes1.rtf>¹.

Si vous préférez remplir (électroniquement ou sur papier) la version hors-ligne, merci de la retourner à :

mathieu.loiseau@u-grenoble3.fr

ou

Mathieu LOISEAU

Département d'Informatique Pédagogique

DIP – BP 25 – 38040 Grenoble Cedex 9
France

En cas de questions ou remarques, n'hésitez pas à les joindre au questionnaire, me les envoyer directement par e-mail ou me contacter au 04 76 82 43 17 (+33 4 76 82 43 17)

Annexe B

Statistiques officielles

Concours Année	Agrégation		CAPES		Total	
	Effectifs	%	Effectifs	%	Effectifs	%
2006						
Anglais	201	52,76%	1176	62,69%	1377	61,01%
Espagnol	85	22,31%	463	24,68%	548	24,28%
Allemand	65	17,06%	125	6,66%	190	8,42%
Autre	30	7,87%	112	5,97%	142	6,29%
Total	381	100,00%	1876	100,00%	2257	100,00%
2005						
Anglais	213	50,00%	1319	62,51%	1532	60,41%
Espagnol	92	21,60%	525	24,88%	617	24,33%
Allemand	74	17,37%	130	6,16%	204	8,04%
Autre	47	11,03%	136	6,45%	183	7,22%
Total	426	100,00%	2110	100,00%	2536	100,00%
2004						
Anglais	194	49,62%	1106	58,18%	1300	56,72%
Espagnol	91	23,27%	536	28,20%	627	27,36%
Allemand	70	17,90%	137	7,21%	207	9,03%
Autre	36	9,21%	122	6,42%	158	6,89%
Total	391	100,00%	1901	100,00%	2292	100,00%
2003						
Anglais	222	56,78%	1687	88,74%	1909	56,31%
Espagnol	119	30,43%	856	45,03%	975	28,76%
Allemand	84	21,48%	210	11,05%	294	8,67%
Autre	38	9,72%	174	9,15%	212	6,25%
Total	463	118,41%	2927	153,97%	3390	100,00%
2002						
Anglais	222	56,78%	1695	89,16%	1917	53,09%
Espagnol	119	30,43%	1005	52,87%	1124	31,13%
Allemand	84	21,48%	225	11,84%	309	8,56%
Autre	51	13,04%	210	11,05%	261	7,23%
Total	476	121,74%	3135	164,91%	3611	100,00%
TOTAL 2002-2006						
Anglais	8035	57,04%				
Espagnol	3891	27,62%				
Allemand	1204	8,55%				
Autre	956	6,79%				
Total	14086	100,00%				

FIGURE B.1: Places ouvertes aux concours de l'agrégation et du CAPES de 2002 à 2006. Données compilées à partir de [MEN, 2003, 2004, 2005, 2006, 2007].

Annexe C

Recours au texte fabriqué (questionnaire 1)

ID & PARCOURS		TEXTES AUTHENTIQUES	TEXTES FABRIQUÉS	CIRCONSTANCES OÙ LES TEXTES FABRIQUÉS SONT PRIVILÉGIÉS
1	C	le plus souvent	rechercher	pas compris cette question et la précédente
2	C	le plus souvent	rechercher	Si je travaille dans un livre scolaire (unité de leçon)
3	C	le plus souvent	contextuel	Quand je n'en trouve pas du niveau des élèves
4	C	le plus souvent	rechercher	pour des débutants
5	C	le plus souvent	rechercher	Pour illustrer un point de grammaire ou un thème particulier. Quand le niveau des élèves ne permet pas la compréhension d'un texte authentique
6	C	le plus souvent	rechercher	En fonction des besoins, des thèmes, des fonctions langagières, des points de grammaire à intégrer
7	C	parfois	contextuel	avec des débutants
8	C	le plus souvent	contextuel	en 5 et 6 toujours authentique
9	C	parfois	rechercher	-
10	C	le plus souvent	rechercher	Lorsque le vocabulaire n'est pas approprié au niveau des élèves ou au sujet que je veux aborder
11	C	toujours	contextuel	dans le cadre de l'enseignement d'un point de grammaire
12	C	le plus souvent	rechercher	-
13	C	le plus souvent	rechercher	Dans les niveaux débutants car il est difficile de trouver des texte adaptés à ce niveau. Lorsque les textes authentiques ne répondent pas au besoins pédagogiques.
14	B	le plus souvent	rechercher	didactiser un texte permet parfois de le rendre plus accessible, de mieux contrôler ce qu'il va apporter aux étudiants
16	B	toujours	contextuel	Grammaire
17	B	le plus souvent	contextuel	groupes de bas niveau ; dans le cadre d'une explication grammaticale

ID & PARCOURS		A.2.1	A.2.2	A.2.3
15	B	le plus souvent	contextuel	Lorsque le texte authentique est abordable et suffisamment riche pour faire exemple.
18	A	parfois	contextuel	petites classes,
19	C	le plus souvent	rechercher	Pour du vocabulaire de la vie quotidienne, des situations quotidiennes
20	A	le plus souvent	contextuel	pour étudier la vie de tous les jours, pour créer les situations quotidiennes et apprendre le langage parlé
21	A	le plus souvent	rechercher	lorsque le texte pédagogique me paraît plus accessible et qu'il est bien écrit
22	C	le plus souvent	rechercher	Compréhension de l'oral
23	A	parfois	rechercher	avec des débutants, faux débutants et élémentaires pour exploiter un point de grammaire traité auparavant
24	A	le plus souvent	rechercher	niveau faible : méthodes structurées
25	A	le plus souvent	rechercher	acquisition de notions spécifiques
26	A	le plus souvent	contextuel	Quand je ne trouve pas de texte authentique correspondant à ce que je veux travailler ; si le texte que j'ai trouvé est trop difficile, je le simplifie
27	A	parfois	rechercher	Vu le niveau des étudiants (débutant) il ne faut pas que le texte étudié comporte trop de problèmes grammaticaux ou de termes inconnus.
28	A	le plus souvent	rechercher	par facilité et quand je ne trouve pas de texte authentique
29	A	toujours	rechercher	aucun!!!!!!!!!!
30	A	le plus souvent	rechercher	presque jamais! Disons que j'utilise parfois des textes didactisées pour faciliter la compréhension, les textes de base étant parfois un peu compliqués pour mes classes
31	A	toujours	rechercher	cas rarissimes d'enseignement à des Lansad mes cours doivent forcément porter sur des textes (ou documents) authentiques (civilisation, littérature etc)
32	A	toujours	rechercher	Jamais
33	A	toujours	rechercher	point de langue
35	A	le plus souvent	rechercher	Pour des remédiations à des notions non acquises
36	A	toujours	contextuel	dans le cas de révisions , de bilan de séquence afin de m'assurer de la bonne et totale compréhension de tous.
37	A	le plus souvent	rechercher	pour la mise en situation de communication
38	A	le plus souvent	rechercher	pour aborder un point de grammaire
39	A	toujours	contextuel	Exercices de grammaire
40	A	le plus souvent	contextuel	lorsque le deuxième nécessite un niveau de compréhension trop élevé pour les élèves

ID & PARCOURS		A.2.1	A.2.2	A.2.3
41	A	le plus souvent	rechercher	Je ne le préfère pas vraiment, mais il m'arrive d'utiliser un texte pédagogique si je tombe sur un qui convient bien à l'activité ou au thème que je souhaite traiter avec mes élèves.
42	A	toujours	rechercher	points grammatical précis
43	A	toujours	rechercher	Je n'ai pas encore utilisé de texte pédagogique
44	A	toujours	contextuel	Lorsque le le texte authentique comporte trop de difficultés et que je veux le mettre à la portée des élèves
45	A	le plus souvent	rechercher	pour les débutants (donc en l'occurrence en quatrième) ou pour revoir une notion censée avoir déjà été étudiée : donc en réemploi et fixation dans la mesure du possible. en quatrième je l'utilise à des fins strictement communicatives. Ex : des dessins pour des dessins pour utiliser la forme progressive (deux dessins quasi-identiques ou des images montrant des actions en cours)
47	A	toujours	contextuel	pour éclaircir une notion culturelle ou syntaxique
48	A	parfois	contextuel	si la demarche pour l'approche du fait de langue paraît la plus judicieuse
49	A	parfois	rechercher	faciliter la compréhension
50	A	parfois	contextuel	au début de l'apprentissage d'une langue et lorsque le niveau de la classe est faible.
51	A	le plus souvent	rechercher	POUR RECR2ER UNE SITUATION VUE PAR UN DOC. AUTHENTIQUE afin d'insister et "d'enfoncer le clou", par rapport à un objectif linguistique en particulier...
52	A	le plus souvent	contextuel	situation concrète
53	A	le plus souvent	contextuel	-Situations de communication très concrètes : lettres, CV... -Intérêt du document pédagogique (vidéos faits par des éditeurs espagnols de ELE, par exemple)
54	A	toujours	ecrire	exercices tous premiers cours
55	A	le plus souvent	rechercher	lorque les objectifs linguistiques à travailler sont une priorité et qu'ils sont plus présents dans le support didactisé
56	A	toujours	contextuel	xxxxxxx
57	A	toujours	rechercher	TOUJOURS
58	A	toujours	ecrire	objectif précis
59	A	le plus souvent	contextuel	pour insister sur un ou des fait(s) de langue
61	A	le plus souvent	ecrire	pour approfondir le vocabulaire et la grammaire
62	A	le plus souvent	rechercher	pour des points ciblés de grammaire
63	A	le plus souvent	contextuel	évaluation ou mise au point de difficultés récurrentes

ID & PARCOURS	A.2.1	A.2.2	A.2.3	
64	A	le plus souvent	contextuel	Compréhension de l'écoute
65	A	parfois	rechercher	Débutants
66	A	le plus souvent	rechercher	Pour aborder un problème complexe (sur le fond) de manière simplifiée, surtout du point de vue lexical de manière à ce que les difficultés de vocabulaire ne gâchent pas le plaisir et l'intérêt de traiter des problématiques en rapport avec l'âge des élèves.
67	A	parfois	contextuel	Dans le cadre de la création de dialogues ou saynètes correspondants au réinvestissement des séquences d'enseignement précédentes
68	A	parfois	contextuel	lorsque je veux cibler la notion, ou pour la simplicité du support, particulièrement en 6e 5e
69	A	toujours	contextuel	dans le cadre de la trace écrite sur le cahier qui émane de la production élève parfois améliorée par moi-même
71	A	le plus souvent	rechercher	dans la cas d'un exercice dont l'objectif est linguistique (grammaire, conjugaison).
72	A	toujours	rechercher	Apprentissage de la grammaire
73	A	parfois	contextuel	En l'absence d'un texte authentique - pour la compréhension de l'oral destinée aux débutants, par exemple
74	A	le plus souvent	rechercher	surtout pour les débutants
75	A	toujours	rechercher	si trop difficile
76	A	parfois	contextuel	Les textes utilisés doivent être très simples, les élèves ne pratiquant l'italien que 45 minutes par semaine. Souvent, je simplifie la syntaxe des textes trouvés.
77	A	parfois	contextuel	Dans les situations d'apprentissage.
78	A	le plus souvent	contextuel	grammaire
79	A	le plus souvent	rechercher	Tout texte authentique est remanié, découpé, pour permettre à mes élèves de se l'approprier, c'est à dire didactisé. Je ne modifie rien dans le cas de recherche et/ou travail sur internet.
80	A	le plus souvent	contextuel	pour la civilisation
81	A	parfois	rechercher	Plutôt pour une évaluation
82	A	le plus souvent	rechercher	Pour l'apprentissage plus ciblé d'un point particulier de grammaire mis en situation
83	A	le plus souvent	contextuel	bilan de plusieurs séances
84	A	toujours	rechercher	Pratiquement jamais
85	A	toujours	contextuel	exercices à compléter
86	A	parfois	contextuel	pour faire plaisir a l'institution et aux parents, il faut "faire" le livre

ID & PARCOURS		A.2.1	A.2.2	A.2.3
87	A	toujours	ecrire	Lorsqu'il s'agit de créer les exercices pratiques. Exemple création de poèmes avec modèle où l'on réemploie les structures vues en cours
88	A	le plus souvent	contextuel	Lorsqu'il y a un point de grammaire abstrait dans la compréhension. Mais aussi, lorsque je désire travailler sur les structures à réutiliser en contexte inconnu.
89	A	le plus souvent	rechercher	Lorsque le texte pédagogique me semble plus adapté au niveau de mes élèves
90	A	toujours	ecrire	ex de grammaire, conjugaison
91	A	parfois	contextuel	Quand on fait une révision sur ce qu'on vient d'étudier.
92	A	parfois	rechercher	Dans tous les cas où j'ai un texte pédagogique immédiatement disponible
93	A	toujours	contextuel	aucun
94	A	parfois	contextuel	Quand le texte authentique est trop difficile à comprendre.
95	A	toujours	rechercher	POUR SERVIR DE BASE A DES DEBATS (entraînement à l'expression orale)
97	A	toujours	contextuel	aucun cas
98	A	le plus souvent	rechercher	travail sur du lexique, des faits de langue précis
99	A	parfois	contextuel	en raison des contraintes du programme
100	A	parfois	contextuel	en fonction du niveau des élèves
101	A	le plus souvent	contextuel	pour des notions linguistiques qui demandent un entraînement rigoureux
102	A	le plus souvent	contextuel	Quand je ne trouve aucun texte authentique correspondant à la fois au thème central de la séquence et au point de langue à aborder
103	A	le plus souvent	contextuel	fixation des structures
104	A	le plus souvent	contextuel	Afin de l'adapter au niveau des élèves et pour éviter la dispersion dans des difficultés qui risquent de nous éloigner des objectifs. (problème de l'arabe littéral et l'arabe dialectal)
105	A	le plus souvent	rechercher	Lorsque les textes authentiques trouvés sont trop complexes pour mes élèves
106	A	le plus souvent	ecrire	exercices
107	A	toujours	rechercher	
108	A	le plus souvent	contextuel	pour expliquer un point grammatical
109	A	le plus souvent	contextuel	pour commencer les leçons et pour les devoirs
110	A	toujours	rechercher	jamais
111	A	le plus souvent	contextuel	pédagogie différenciée
112	A	toujours	contextuel	évaluation
113	A	toujours	rechercher	Jamais, je ne travaille qu'avec des textes authentiques.

ID & PARCOURS	A.2.1	A.2.2	A.2.3	
114	A	toujours	contextuel	D'abord pour les étudiants, toujours. Pour le collègue, le plus souvent, afin que la langue soit vivante
115	A	le plus souvent	contextuel	Analyse grammaticale.
116	A	toujours	contextuel	Applications grammaticales
117	A	toujours	ecrire	Pour les débutants pour ne pas les embrouiller avec trop de lexique ou de grammaire inconnus
119	A	le plus souvent	contextuel	pour l'approche de la grammaire et de la conjugaison, dans le cadre d'exercices d'application
120	A	le plus souvent	rechercher	quand il s'agit de travailler une notion spécifique ou un point de grammaire déterminé.
121	A	le plus souvent	ecrire	situations que je crée moi même : ex apprentissage du lexique de l'alimentation
122	A	toujours	contextuel	Pour une activité d'applications notionnelles s'il y a lieu...
123	A	toujours	contextuel	dans les exercices
137	A	parfois	rechercher	THEMES CULTURELS? EXPERIENCES INTERCULTURELLES
125	A	toujours	contextuel	cas ponctuels (activités ludiques, exploitation points précis grammaire ou lexique)
126	A	le plus souvent	contextuel	Dans les petites classes
127	A	toujours	contextuel	modèle grammatical
128	A	toujours	rechercher	Activités grammaticales
129	A	parfois	contextuel	afin de faire acquérir une notion bien précise, ou un lexique particulier
130	A	parfois	contextuel	élèves en difficulté
131	A	le plus souvent	rechercher	importance régressive sur les premières 150 heures
132	A	toujours	contextuel	je sais pas
133	A	le plus souvent	contextuel	pour évaluer
134	A	parfois	rechercher	surtout au niveau débutant
135	A	parfois	rechercher	La plupart des textes authentiques ne sont pas adaptés au niveau de mes élèves
136	A	le plus souvent	rechercher	en cas d'objectif grammatical. Les occurrences sont généralement en nombre suffisant. Mais il est vrai que je pourrais les trouver dans des textes authentiques.

TABLE C.1: Réponses aux questions A.2.1, A.2.2 et A.2.3 p. a-

Annexe D

Croisement « sources » et « provenance » (questionnaire 1)

Descriptives

Remarques		
Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif Filtrer Poids Scinder fichier N de lignes dans le fichier de travail	DataSet1 recherche passive = 0 (FILTER) <aucune> <aucune> 7
Gestion des valeurs manquantes	Définition des valeurs manquantes Observations prises en compte	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes. Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur Temps écoulé	0:00:00.005 0:00:00.000

"recherche passive" = 0

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	7	0	4	2,14	1,107
ouvrages littéraires	7	0	3	,64	1,107
méthodes/manuels	7	3	4	3,43	,607
Internet	7	0	4	2,64	1,249
N valide (listwise)	7				

. Pour une explication du calcul des valeurs associées à chaque groupe, voir § a) p. 175.

Descriptives**Remarques**

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	recherchepassive = 1 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	28
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.002
	Temps écoulé	0:00:00.000

"recherche passive" = 1

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	28	0	4	2,34	1,000
ouvrages littéraires	28	0	3	1,18	,819
méthodes/manuels	28	0	4	3,32	1,219
Internet	28	0	4	2,70	1,066
N valide (listwise)	28				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	recherchepassive = 2 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	65
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.004
	Temps écoulé	0:00:00.000

"recherche passive" = 2

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	65	0	4	2,59	1,176
ouvrages littéraires	65	0	4	1,54	1,200
méthodes/manuels	65	0	4	2,77	1,344
Internet	65	1	4	2,65	1,057
N valide (listwise)	65				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	recherche passive > 2 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	30
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.003
	Temps écoulé	0:00:00.000

"recherche passive" > 2 (ou = 3)

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	30	0	4	2,68	1,256
ouvrages littéraires	30	0	4	1,45	1,392
méthodes/manuels	30	0	4	2,47	1,332
Internet	30	0	4	2,37	1,408
N valide (listwise)	30				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	recherche passive > 1 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	95
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.009
	Temps écoulé	0:00:00.000

"recherche passive" > 1

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	95	0	4	2,62	1,196
ouvrages littéraires	95	0	4	1,51	1,257
méthodes/manuels	95	0	4	2,67	1,340
Internet	95	0	4	2,56	1,179
N valide (listwise)	95				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	recherche passive > 0 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	123
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.003
	Temps écoulé	0:00:00.000

"recherche passive" > 0

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	123	0	4	2,56	1,156
ouvrages littéraires	123	0	4	1,43	1,177
méthodes/manuels	123	0	4	2,82	1,337
Internet	123	0	4	2,59	1,151
N valide (listwise)	123				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	imposéparleprogramme = 0 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	34
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.004
	Temps écoulé	0:00:00.000

Chez les "imposés par le programme" = 0

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	34	0	4	2,51	1,329
ouvrages littéraires	34	0	4	1,09	1,328
méthodes/manuels	34	0	4	2,65	1,374
Internet	34	0	4	2,54	1,311
N valide (listwise)	34				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	imposéparleprogramme = 1 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	56
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.002
	Temps écoulé	0:00:00.000

"imposés par le programme" = 1

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	56	0	4	2,58	1,163
ouvrages littéraires	56	0	4	1,68	1,162
méthodes/manuels	56	0	4	2,68	1,295
Internet	56	1	4	2,67	1,105
N valide (listwise)	56				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	imposéparleprogramme = 2 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	13
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.004
	Temps écoulé	0:00:00.000

"imposés par le programme" = 2

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	13	2	4	2,85	,801
ouvrages littéraires	13	0	4	1,23	1,166
méthodes/manuels	13	2	4	3,54	,776
Internet	13	0	4	2,08	1,115
N valide (listwise)	13				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	imposéparleprogramme > 2 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	27
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.004
	Temps écoulé	0:00:00.000

"imposés par le programme" > 2 (ou = 3)

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	27	0	4	2,31	1,048
ouvrages littéraires	27	0	3	1,26	,944
méthodes/manuels	27	0	4	3,15	1,379
Internet	27	0	4	2,76	1,032
N valide (listwise)	27				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	imposéparleprogramme > 1 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	40
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.004
	Temps écoulé	0:00:00.000

"imposés par le programme" > 1

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	40	0	4	2,49	,997
ouvrages littéraires	40	0	4	1,25	1,006
méthodes/manuels	40	0	4	3,28	1,219
Internet	40	0	4	2,54	1,094
N valide (listwise)	40				

Descriptives

Remarques

Résultat obtenu		
Commentaires		
Entrée	Ensemble de données actif	DataSet1
	Filtrer	imposéparleprogramme > 0 (FILTER)
	Poids	<aucune>
	Scinder fichier	<aucune>
	N de lignes dans le fichier de travail	96
Gestion des valeurs manquantes	Définition des valeurs manquantes	Les valeurs manquantes définies par l'utilisateur sont traitées comme manquantes.
	Observations prises en compte	Toutes les données non manquantes sont utilisées.
Syntaxe		DESCRIPTIVES VARIABLES=revuesjournaux ouvrageslittéraires méthodesmanuels Internet /STATISTICS=MEAN STDDEV MIN MAX.
Ressources	Temps de processeur	0:00:00.004
	Temps écoulé	0:00:00.000

"imposés par le programme" > 0

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
revues/journaux	96	0	4	2,54	1,092
ouvrages littéraires	96	0	4	1,50	1,114
méthodes/manuels	96	0	4	2,93	1,292
Internet	96	0	4	2,61	1,097
N valide (listwise)	96				

Annexe E

Documentation du gestionnaire de questionnaires

E.1 Documentation

Documentation gestionnaire de questionnaires

Documentation du gestionnaire de questionnaires

1. Installation et stockage des questions du questionnaire :

1.1. *Tout copier dans un dossier d'un serveur php*

1.2. *Modifier :*

- **connectBD.php** : avec les codes d'accès de la bonne base SQL
- Choisir un nom de variable de session pour l'identifiant de la personne qui est en train de remplir le questionnaire et le reporter dans **login.php**, **exploitation.php** à la place de `$_SESSION['s_idProf']`. (si `s_idProf` convient, ça diminue le nombre de modifications)
- **exploitation.php** : changer si nécessaire le nom de `$_SESSION['s_idProf']` et les appels aux fonctions de menu de manière à modifier les bonnes tables.
- **fonctionsFormulaires.php** : changer si nécessaire le nom de `$_SESSION['s_idProf']`
- **login.php** : un truc de login sans mot de passe. Login possible en réutilisant le même login (adresse e-mail chez moi) ou en suivant un lien généré automatiquement (cf. fonctionsFormulaires) et toujours `$_SESSION['s_idProf']` à modifier via les variables déclarées en début de fichier. Modifier les appels des fonctions de menu pour bien prendre en compte les noms des tables associées.
- **logout.php** : modifier en accord avec **fonctionsFormulaires.php** et **login.php**
- **menu.php** : modifier les fonctions entete et endHTML pour qu'elles prennent en compte votre feuille de style. À consulter pour les appels des fonctions des menus (beaucoup de paramètres appelés par défaut, cf. modification de **exploitation.php**).

1.3. Les Tables :

1.3.1. La table regroupant toutes les questions des questionnaires :

```
CREATE TABLE `questionnaire` (  
  `IdQuest` tinyint(3) unsigned NOT NULL auto_increment,  
  `question` text collate latin1_general_ci NOT NULL,  
  `type` enum('_text', '_textArea', '_orderList', '_radio',  
  '_checkBox', '_listMultiple', '_list') collate  
latin1_general_ci NOT NULL default '_text',  
  `largeur` int(4) NOT NULL default '-1',  
  `hauteur` int(4) NOT NULL default '-1',  
  `autre` enum('_none', '_singleText', '_multipleText') collate  
latin1_general_ci NOT NULL default '_none',  
  `nomTabRep` varchar(24) collate latin1_general_ci NOT NULL  
default '',  
  `nomClef` varchar(24) collate latin1_general_ci NOT NULL  
default '',  
  `nomClefJointure` varchar(24) collate latin1_general_ci NOT  
NULL default '_ibid',  
  `nomTexte` varchar(24) collate latin1_general_ci NOT NULL  
default '',  
  `nomIdUsr` varchar(16) collate latin1_general_ci NOT NULL
```

Documentation gestionnaire de questionnaires

```

default '',
`nomTabJointure` varchar(24) collate latin1_general_ci NOT
NULL default '',
`nomOrdre` varchar(16) collate latin1_general_ci NOT NULL
default '',
`update` enum('true','false') collate latin1_general_ci NOT
NULL default 'false',
UNIQUE KEY `IdQuest` (`IdQuest`)
)ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci;

```

`IdQuest` : l'identifiant g r  automatiquement (auto_increment)

`question` : la question pos e au format HTML

`type` : le type de r ponse :

- `_text` : zone de texte sur une ligne
- `_textArea` : zone de texte sur plusieurs lignes
- `_orderList` : liste   ordonner (peut- tre un peu bugg )
- `_radio` : choix multiple, un seul choix
- `_checkBox` : choix multiple, plusieurs choix
- `_list` : choix multiple, un seul choix
- `_listMultiple` : choix multiple, plusieurs choix

`largeur` : -1 choix par d faut

pour `_radio` et `_checkBox` : le nombre de colonnes affich es

pour `_textArea` : Le nombre de caract res par ligne

pas d'influence pour les autres types de questions

`hauteur` : -1 choix par d faut

pour `_listMultiple` et `_list` : le nombre de lignes

pour `_textArea` : le nombre de lignes

pour les autres, pas d'influence pour les autres types de question

`autre` :

ne fonctionne que pour `_radio`, `_checkBox`, `_list` et `_listMultiple`

`_none` : choix par d faut, pas de case "autre" apr s les propositions

`_singleText` : possibilit  de rajouter une proposition

`_multipleText` : possibilit  de rajouter plusieurs proposition (ne marche pas pour `_radio` ou `_list`)

NB : les propositions ajout es sont ensuite propos es aux autres sond s.

`nomTabRep` : le nom de la table avec les r ponses (qui servira de nom   la liste de valeurs)

`nomClef` : le nom de la cl  correspondant   l'id des r ponses (dans le cas d'un texte ou d'une r ponse   choix unique, le `nomClef` peut  tre le m me que `nomIdUsr`)

`nomClefJointure` : nom de `nomClef` dans la table de jointure (questions choix multiple, ou question ou une table de jointure est utilis e pour stocker les r ponses) si "`_ibid`" reprend la valeur de "`nomClef`"

`nomTexte` : le nom du champ dans la table de r ponse contenant le texte associ    la r ponse (les r ponses aux `_text` et `_textArea` seront stock es dedans)

`nomIdUsr` : le nom du champ contenant l'identifiant de l'utilisateur dans la table de jointure

`nomTabJointure` : le nom de la table o  seront stock es les r ponses des utilisateurs

`nomOrdre` : Seulement pour les `_orderList`. Le nom du champ contenant la relation d'ordre

Documentation gestionnaire de questionnaires

`update` : seulement pour les `_list` et `_radio`. Mettre à vrai si la même table contient les réponses à plusieurs questions. Comme on ne traite qu'un champ à la fois cela demandera un traitement spécial. Par défaut, vaut faux.

1.3.2. Exemple d'enregistrement et de table de réponses et de texte, correspondant.

```
INSERT INTO `questionnaire` VALUES (6, 'Le(s) pays dans
le(s)quel(s) vous enseignez', '_listMultiple', -1, -1,
'_multipleText', 'q2_pays', 'idPays', '_ibid', 'nomPays',
'idProf', 'q2_prof-pays', '-', 'false');
-- '_multipleText' : si le pays où l'enseignant travaille n'est pas
répertorié dans la liste, il peut l'ajouter (et s'il travaille dans
plusieurs pays, il peut en ajouter plusieurs).
-- La table contenant les choix proposés
```

`IdQuest` : 6

`question` : 'Le pays dans lequel vous enseignez'

`type` : `_listMultiple` (choix multiple dans une liste)

`largeur` : -1 choix par défaut

`hauteur` : -1 choix par défaut

`autre` : `_multipleText` : si le pays où l'enseignant travaille n'est pas répertorié dans la liste, il peut l'ajouter (et s'il travaille dans plusieurs pays, il peut en ajouter plusieurs).

`nomTabRep` : `q2_pays`

`nomClef` : `idPays`

`nomClefJointure` : `_ibid` donc `idPays`

`nomTexte` : `nomPays`

`nomIdUsr` : `idProf`

`nomTabJointure` : `q2_prof-pays`

`nomOrdre` : Seulement pour les `_orderList`. Aucun ici

`update` : `false` car liste multiple

```
-- La table contenant les réponses possibles à la question :
--
-- Structure de la table `q2_prof-pays`
--
CREATE TABLE `q2_pays` (
  `idPays` smallint(6) unsigned NOT NULL auto_increment,
  `nomPays` varchar(32) collate latin1_general_ci NOT NULL,
  PRIMARY KEY (`idPays`)
) ENGINE=MyISAM AUTO_INCREMENT=26 DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci COMMENT='La liste des q2_pays où
exercer les enseignants' AUTO_INCREMENT=26 ;
--La table contenant les réponses des enseignants interrogés :
--
-- Structure de la table `q2_prof-pays`
--
CREATE TABLE `q2_prof-pays` (
  `idProf` int(11) unsigned NOT NULL,
  `idPays` smallint(6) unsigned NOT NULL,
  PRIMARY KEY (`idProf`,`idPays`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci COMMENT='Là où les enseignants ont
travaillé';
--
-- Contenu de la table `q2_pays`
-- à remplir manuellement
INSERT INTO `q2_pays` VALUES (1, 'Allemagne');
INSERT INTO `q2_pays` VALUES (2, 'Angleterre');
```

Documentation gestionnaire de questionnaires

```

INSERT INTO `q2_pays` VALUES (3, 'Argentine')
INSERT INTO `q2_pays` VALUES (4, 'Autriche');
INSERT INTO `q2_pays` VALUES (5, 'Bulgarie');
INSERT INTO `q2_pays` VALUES (6, 'Canada');
INSERT INTO `q2_pays` VALUES (7, 'Colombie');
INSERT INTO `q2_pays` VALUES (8, 'Congo');
INSERT INTO `q2_pays` VALUES (9, 'Espagne');
--
-- Exemple de réponses (insérées automatiquement par le système)
--
INSERT INTO `q2_prof-pays` VALUES (1, 8);
INSERT INTO `q2_prof-pays` VALUES (2, 4);
INSERT INTO `q2_prof-pays` VALUES (2, 7);

```

Le pays dans lequel vous enseignez

Pour sélectionner plusieurs éléments, maintenir la touche Ctrl (Windows) ou pomme (Mac) enfoncée en même temps que vous cliquez sur les propositions de votre choix

Autre(s):

Séparer les propositions avec un caractère '|' (Alt Gr + 6 sous windows, alt + shift + L sous Mac)

Figure 1 - La portion de formulaire générée par les enregistrements ci-dessus

Pour les questions où l'utilisateur ne donne qu'une réponse (_list, _radio, _text ou _textArea), plusieurs réponses peuvent être stockées dans une même table (table de jointure) si le paramètre « update » est true :

```

INSERT INTO `questionnaire2` VALUES (2, 'Souhaitez-vous
&ecirc;tre contact&eacute;(e) pour avoir des informations sur
l'avancement du projet ?', '_radio', 2, -1, '_none',
'q2_bool', 'id', 'feedback', 'bool', 'idProf',
'q2_enseignant', '-', 'true');

```

Souhaitez-vous être contacté(e) pour avoir des informations sur l'avancement du projet ?

oui non

```

INSERT INTO `questionnaire2` VALUES (7, 'Le nombre
d'années durant lequel vous avez enseigné;
<br><span class="consigne">entrer uniquement le
nombre</span>', '_text', -1, -1, '_none', 'q2_enseignant',
'idProf', '_ibid', 'experience', 'idProf', 'q2_enseignant',
'-', 'true');
-- Pas de table contenant les réponses pour une question de type texte
TabRep et TabJointure sont identiques

```

Le nombre d'années durant lequel vous avez enseigné :

entrer uniquement le nombre

```

INSERT INTO `questionnaire2` VALUES (9, 'Avez-vous un recueil
de texte personnel ?', '_radio', 1, -1, '_none', 'q2_bool',
'id', 'rec', 'bool', 'idProf', 'q2_enseignant', '-', 'true');

```

Avez-vous un recueil de texte personnel ?

oui

non

Documentation gestionnaire de questionnaires

```

--
-- Ces 3 questions nécessiteront l'utilisation des tables suivantes
--
-- Réponses proposées
CREATE TABLE `q2_bool` (
  `id` tinyint(1) NOT NULL default '0',
  `bool` enum('oui','non') collate latin1_general_ci NOT NULL
default 'non',
  PRIMARY KEY (`id`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci;
INSERT INTO `q2_bool` VALUES (1, 'oui');
INSERT INTO `q2_bool` VALUES (0, 'non');
-- Stockage des réponses aux trois questions (2, 7, 9)
CREATE TABLE `q2_enseignant` (
  `idProf` int(11) unsigned NOT NULL auto_increment,
  `feedback` tinyint(1) default NULL,
  `experience` tinyint(4) default NULL,
  `rec` tinyint(1) NOT NULL default '0',
  UNIQUE KEY `idProf` (`idProf`)
);

```

2. Affichage conditionnel

2.1. Création de la base

Pour les questions du type « si oui, ... ».

Tout est géré via une base de donnée. Pour faire un formulaire avec affichage conditionnel, utiliser le constructeur à deux paramètres :

```

$formulaire = new Form("questionSuivante.php",
                      "NomTableAffichageCond");

```

Pour que cela fonctionne il faut créer une table avec le nom indiqué dans le constructeur :

```

CREATE TABLE `NomTableAffichageCond` (
  `nomTable` varchar(24) collate latin1_general_ci NOT NULL
default '',
  `nomClef` varchar(24) collate latin1_general_ci NOT NULL
default '',
  `valClef` varchar(24) collate latin1_general_ci NOT NULL
default '',
  `evenement` varchar(11) collate latin1_general_ci NOT NULL
default '',
  `init` varchar(32) collate latin1_general_ci NOT NULL
default '',
  `listeFonctions` varchar(192) collate latin1_general_ci NOT
NULL default '',
  `listeItems` varchar(192) collate latin1_general_ci NOT
NULL default '',
  PRIMARY KEY (`nomTable`,`nomClef`,`valClef`,`evenement`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci;

```

nomTable : Le nom de la table qui contient les réponses proposées à la question qui appelle un affichage conditionnel

nomClef : La clef dans la table **nomTable**

valClef : La valeur de la clef pour laquelle une action doit être effectuée

Documentation gestionnaire de questionnaires

`evenement` : L'événement javascript qui déclenche l'action (ex : `onClick`, `onChange`)
`init` : la liste des numéros des items à afficher si l'élément est (`nomTable`.`nomClef`=valClef`) séparés par des ',' (en comptant à partir de l'élément courant). Ce champ est utile si l'utilisateur veut revenir sur une réponse donnée (n'est pas utilisé quand il n'a pas encore répondu à la question).`

`listeFonctions` : La liste des fonctions javascript à appeler. Les fonctions définies dans 'affichageConditionnel.js' sont :

`hide` : cacher

`hideAllFromTo` : cacher tous les éléments dont les identifiants sont compris entre

a et b

`show` : montrer

`affCondId` : cacher si visible et montrer si caché

`affCondRadio` : la même chose que `affCondId` mais pour les boutons radio

`init` : fonction créée dynamiquement par la classe php et qui affiche tous les éléments qui sont censés être affichés quand un élément est coché, si cet élément est coché.

`initFull` : fonction créée dynamiquement par la classe php et qui affiche tous les éléments qui sont censés être affichés quand un élément est coché, si cet élément est coché et dissimule ceux qui sont censés être dissimulés. N'utiliser cette fonction que si l'affichage d'un élément ne peut être déclenché que par UN SEUL autre élément. Cette fonction générera des erreurs d'affichage sinon.

`listeItems` : la liste des paramètres des fonctions, doit faire la même longueur que la liste des Fonctions :

' ' pour séparer deux paramètres d'une même fonction

'|' pour séparer les paramètres de deux fonctions

'#' précède un numéro d'élément mobile relatif (#n affecte le n^{ème} élément en partant de l'élément concerné. NB: pas très propre, demande de gérer l'enchaînement des questions à la main en dehors de la BD)

NB1 : Il est possible d'appeler une séquence de fonctions en les séparant par un |

NB2 : Il est possible d'appeler n'importe quelle fonction javascript incluse dans la page.

2.2. Exemples

```
INSERT INTO `NomTableAffichageCond` VALUES ('q2_ex_reponses',
'clefRep', '1', 'onchange', '1,2', 'affCondID|affCondID',
'#1|#2');
```

Avec cet enregistrement : quand une question prend ses réponses dans la table `q2_ex_reponses`, à chaque **changement** de valeur de l'item correspondant l'enregistrement désigné par l'identifiant **1** dans la colonne `clefRep` (item0), le système affichera ou escamotera selon qu'ils étaient affichés avant ou non les deux items suivant l'item0. A l'initialisation les deux items suivant l'item0 sont affichés si l'item0 est choisi par l'utilisateur.

Si l'item0 est un bouton radio il faudra utiliser une formule du style :

```
INSERT INTO `NomTableAffichageCond` VALUES
('q2_ex_reponses2', 'clefRep', '1', 'onclick', '1',
'affCondRadio|affCondRadio', '#1,this,1|#2,this,2');
```

`affCondRadio` prend trois paramètres : le numéro de l'élément à afficher / escamoter, le bouton radio qui déclenche l'affichage et la valeur pour laquelle l'élément doit être afficher.

Documentation gestionnaire de questionnaires

L'enregistrement ci-dessus signifie donc qu'à chaque clic sur le bouton radio 1 on va retester les valeurs du bouton 1 et du bouton 2 et afficher l'item suivant si la valeur de l'ensemble des éléments radio est 1 ou celui d'après si la valeur est 2. Pour fonctionner (dissimuler l'élément suivant quand on clique sur le bouton 2, il faudra ajouter l'enregistrement suivant :

```
INSERT INTO `NomTableAffichageCond` VALUES
('q2_ex_reponses2', 'clefRep', '2', 'onclick', '2',
'affCondRadio|affCondRadio', '#1,this,1|#2,this,2');
```

Un dernier exemple :

```
INSERT INTO `q_affcond` VALUES
--il s'agit de la table contenant les résultats à la question 9 ci-dessus
--si la valeur stockée est 0 on cache toutes les question suivantes
('q2_bool_rec', 'id', '0', 'onclick', '', 'hide|hide|hide',
'#1|#2|#3'),
--sinon on affiche la suivante et on réinitialise la suite (en fonction des
valeurs cochées)
('q2_bool_rec', 'id', '1', 'onclick', '1', 'show|init',
'#1|#1');
```

Avez-vous un recueil de texte personnel ?

oui
 non

Avez-vous un recueil de texte personnel ? Ce recueil est-il organisé ?

oui
 non

Figure 2 - exemple d'affichage conditionnel

3. Ecriture d'un formulaire d'un questionnaire

Voir Exemple de page de question.php

4. Gestion du menu

Le module de gestion du menu, ne gère complètement qu'un site pour lequel tous les fichiers sont dans le même dossier. Pour gérer un site utilisant plusieurs dossiers, dans l'état actuel de la programmation, il faudrait une table par dossier utilisé, de manière à mettre à jour les liens en fonction du dossier dans lequel on se trouve.

4.1. Le menu

Toutes les données du menu sont stockées dans une table de la forme suivante :

```
CREATE TABLE `q2_menu` (
  `idPos` tinyint(3) unsigned NOT NULL auto_increment,
  `texte` varchar(64) collate latin1_general_ci NOT NULL
  default '',
  `lien` varchar(64) collate latin1_general_ci NOT NULL
  default '',
  `sousMenu` enum('true','false') collate latin1_general_ci
  NOT NULL default 'false',
  `fromQuestionnaire` enum('true','false') collate
  latin1_general_ci NOT NULL default 'false',
  UNIQUE KEY `idPos` (`idPos`)
```

Documentation gestionnaire de questionnaires

```
) ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci COMMENT='Menu (nécessite menu.php
et menuDynamique.js)';
```

idPos : identifiant des items. Le menu est créé dans l'ordre de idPos croissants. Les premiers éléments seront en haut du menu

texte : le texte qui sera affiché

lien : la page vers laquelle pointe l'objet

sousMenu : Si vrai, alors l'élément considéré ne s'affichera que lorsque la souris pointera sur le dernier élément non sous-menu le précédent

fromQuestionnaire : est-ce un lien vers une page du questionnaire. Auquel cas s'il existe un historique et si une feuille de style est définie l'utilisateur pourra voir dans le menu les pages sur lesquelles il a répondu à toutes les questions.

4.1.1. Exemple de menu

```
INSERT INTO `q2_menu` VALUES (1, 'Accueil', 'index.php',
'false', 'false');
INSERT INTO `q2_menu` VALUES (2, 'Projet', 'bdtip.php',
'false', 'false');
INSERT INTO `q2_menu` VALUES (3, 'Questionnaire n°1',
'q1/index.php', 'false', 'false');
INSERT INTO `q2_menu` VALUES (4, 'Résultats du
précédent questionnaire', 'q1/stats.php',
'true', 'false');
--sous menu de Questionnaire n°1
INSERT INTO `q2_menu` VALUES (5, 'Questionnaire n°2',
'login.php', 'false', 'false');
INSERT INTO `q2_menu` VALUES (6, 'Précisions quest.
1', 'q2.01_prcisq1.php', 'true', 'true');
--sous menu de Questionnaire n°2 faite partie du questionnaire (gestion de
l'historique)
INSERT INTO `q2_menu` VALUES (7, 'Critères
potentiels', 'q2.02_outilsTAL.php', 'true', 'true');
--sous-menu de Questionnaire n°2 faite partie du questionnaire (gestion de
l'historique) (sera situé en dessous de "Précisions quest. 1")
```

4.2. Gestion de l'historique

La gestion de l'historique nécessite une table de la forme :

```
CREATE TABLE `q2_historique` (
  `idProf` int(11) NOT NULL default '0',
  `idItem` tinyint(4) NOT NULL default '0',
  PRIMARY KEY (`idProf`,`idItem`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1
COLLATE=latin1_general_ci COMMENT='Pour stocker le parcours
des enseignants';
```

Une fois la table définie le programme gère le reste.

5. Feuilles de style

Le menu, comme les formulaires demandent la définition d'un certain nombre de styles pour bien fonctionner :

Documentation gestionnaire de questionnaires

La feuille de style Formulaire.css récapitule les classes nécessaires. Mais pour n'utiliser que ces classes là, il faudra modifier les fonctions entete et endHtml de menu.php qui sont faites pour mes feuilles de style.

Les feuilles de style utilisées sont disponibles.

E.2 Fichier d'exemple

Exemple de page de questions.php

Page 1 of 1

```

<? session_start(); ?>
<!-- Toujours commencer par le session_start pour ne pas perdre les variables de session -->
<?
//pour pouvoir utiliser la fonction q2_login et writeMenu
include_once ("fonctionsFormulaires.php");
//Vérification du "logage" de l'utilisateur
//et écriture de l'en-tête le cas échéant
//on imagine avoir 3 feuilles de styles pour les pages une pour la partie menu, une pour
//le questionnaire et à charger dans une balise <!--IF IE--> pour gérer internet explorer
//on veut de plus ajouter une balise meta "langage", comme q2_login fait appel à entete()
//on aura l'appel suivant :
$fds = array("feuillesDeStyle/menu.css", "feuillesDeStyle/questionnaire.css");
$fdsIE = array("fdsHackIE/questionnaireIE.css");
//NB : d'autres paramètres sont possibles, cf. fonctionsFormulaires.php et doc
if (q2_login("Titre de la page", $fds, $fdsIE, "<meta name=\"Language\" content=\"fr\">")){
//Si l'utilisateur est loggé on peut afficher le menu :
writeMenu("Table de menu", "Titre de la page", $_SESSION['s_idProf']/*identifiant de l'utilisateur*/); //Cf. menu.php
pour paramètres additionnels
include_once("classForm.php"); //pour pouvoir créer des questionnaires
$formulaire = new Form("PageSuivante.php", "tableAffCond"); //pas d'affichage conditionnel sans le second paramètre
//à partir de la création du formulaire tout se fait grâce à l'objet créé (surtout utile pour l'affichage conditionnel,
chaque commande texte ou question ajoute un item) :
//ajout de texte
$formulaire->texte("<h2>Informations personnelles</h2>");
$formulaire->texte("<h3>Contact</h3>");
//ajout de question
$formulaire->question("questionnaire2", 1, $_SESSION['s_idProf']);
//3 paramètres : de quelle table vient la question, quel est son identifiant et quel est l'identifiant du "sondé"
$formulaire->question("questionnaire2", 2, $_SESSION['s_idProf']);
$formulaire->question("questionnaire2", 3, $_SESSION['s_idProf']);
//Bouton d'envoi du questionnaire
$formulaire->texte("_bouton");
//Si affichage conditionnel il y a
$formulaire->lancerAffCond();
}
//fermer le HTML du document
endHTML();
?>

```

Annexe F

Réponses au second questionnaire

Les questions sont proposées dans l'ordre dans lequel elles ont été posées, entre-coupées des différents titres et remarques figurant dans le questionnaire. Chaque section (numérotées de 1 à 8) correspond à une page du questionnaire.

F.1 Précisions / mise à jour des réponses au questionnaire n°1

QUESTION	Votre adresse e-mail Cette adresse e-mail va juste vous servir d'identifiant, elle ne sera pas utilisée pour vous contacter à moins que vous ne m'en donniez l'autorisation ci-dessous. Elle ne sera en aucun cas transmise à qui que ce soit (personne physique ou organisme)
IDPROF	RÉPONSES
-	Dans un souci d'anonymat des témoignages, ces réponses ne sont pas divulguées.
203 personnes ont répondu à cette question	

TABLE F.1: Réponses à la question n° 1 du second questionnaire

QUESTION	Souhaitez-vous être contacté(e) pour avoir des informations sur l'avancement du projet ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non	 29.4%	57 réponses parmi 194
oui	 70.6%	137 réponses parmi 194

TABLE F.2: Réponses à la question n° 2 du second questionnaire

QUESTION	Êtes-vous prêt(e) à nous aider dans nos recherches concernant l'apprentissage des langues assisté par ordinateur en répondant à d'éventuels questionnaires ou en testant des prototypes ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non	 31.4%	61 réponses parmi 194
oui	 68.6%	133 réponses parmi 194

TABLE F.3: Réponses à la question n° 3 du second questionnaire

QUESTION	La (ou les) langue(s) que vous enseignez :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
anglais	 41.8%	81 réponses parmi 194
castillan espagnol	 27.3%	53 réponses parmi 194
français	 21.6%	42 réponses parmi 194
allemand	 9.8%	19 réponses parmi 194
néerlandais flamand	 6.2%	12 réponses parmi 194
italien	 3.6%	7 réponses parmi 194
arabe	1%	2 réponses parmi 194
occitan (après 1500) provençal	1%	2 réponses parmi 194
hongrois	0.5%	1 réponses parmi 194
latin	0.5%	1 réponses parmi 194
lingala	0.5%	1 réponses parmi 194
malais	0.5%	1 réponses parmi 194
portugais	0.5%	1 réponses parmi 194
russe	0.5%	1 réponses parmi 194

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
afar, abkhaze, avestique, afrikaans, akan, amharique, aragonais, assamais, avar, aymara, azéri, bachkir, biélorusse, bulgare, bihari, bichlamar, bambara, bengali, tibétain, breton, bosniaque, catalan valencien, tchétyhène, chamorro, corse, cree, tchèque, tchouvache, gallois, danois, maldivien, dzongkha, éwé, grec moderne (après 1453), espéranto, estonien, basque, persan, peul, finnois, fidjien, féroïen, frison, irlandais, gaélique gaélique écossais, galicien, guarani, goudjrati, manx mannois, haoussa, hébreu, hindi, hiri motu, croate, haïtien créole haïtien, arménien, herero, interlingua (langue auxiliaire internationale), indonésien, interlingue, igbo, yi de Sichuan, inupiaq, ido, islandais, inuktitut, japonais, javanais, géorgien, kongo, kikuyu, kuan-yama kwanyama, kazakh, groenlandais, khmer, kannada, coréen, kanouri, kashmiri, kurde, komi, cornique, kirghize, luxembourgeois, ganda, limbourgeois, lao, lituanien, luba-katanga, letton, malgache, marshall, maori, macédonien, malayalam, mongol, moldave, marathe, maltais, birman, nauruan, bokmål, norvégien bokmål norvégien, ndébélé du Nord, népalais, ndonga, norvégien nynorsk nynorsk, norvégien, norvégien, ndébélé du Sud, navaho, chewa, chichewa, nyanja, ojobwa, galla, oriya, ossète, pendjabi, pali, polonais, pachto, quechua, rhéto-roman, rundi, roumain, rwanda, sanskrit, sarde, sindhi, sami du Nord, sango, singhalais, slovaque, slovène, samoan, shona, somali, albanais, serbe, swati, sotho du Sud, soundanais, suédois, swahili, tamoul, télougou, tadjik, tha, tigrigna, turkmène, tagalog, tswana, tongan (îles Tonga), turc, tsonga, tatar, twi, tahitien, ouïgour, ukrainien, ourdou, ouzbek, venda, vietnamien, volapük, wallon, wolof, xhosa, yiddish, yoruba, chuang zhuang, chinois, zoulou		Aucun témoignage
194 personnes ont répondu à cette question et ont choisi en moyenne 1.2 solutions par personne		

TABLE F.4: Réponses à la question n° 4 du second questionnaire

QUESTION	Les langues maternelles les plus représentées parmi vos apprenants :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
français	 71.4%	135 réponses parmi 189
arabe	 14.8%	28 réponses parmi 189
anglais	 13.8%	26 réponses parmi 189

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
castillan espagnol	<input type="checkbox"/>	13.8% 26 réponses parmi 189
chinois	<input type="checkbox"/>	6.9% 13 réponses parmi 189
portugais	<input type="checkbox"/>	5.8% 11 réponses parmi 189
turc	<input type="checkbox"/>	5.8% 11 réponses parmi 189
allemand	<input type="checkbox"/>	4.8% 9 réponses parmi 189
japonais	<input type="checkbox"/>	3.7% 7 réponses parmi 189
russe	<input type="checkbox"/>	3.7% 7 réponses parmi 189
italien	<input type="checkbox"/>	3.2% 6 réponses parmi 189
polonais	<input type="checkbox"/>	2.1% 4 réponses parmi 189
coréen	<input type="checkbox"/>	1.6% 3 réponses parmi 189
vietnamien	<input type="checkbox"/>	1.6% 3 réponses parmi 189
catalan valencien	<input type="checkbox"/>	1.1% 2 réponses parmi 189
persan	<input type="checkbox"/>	1.1% 2 réponses parmi 189
indonésien	<input type="checkbox"/>	1.1% 2 réponses parmi 189
géorgien	<input type="checkbox"/>	1.1% 2 réponses parmi 189
néerlandais flamand	<input type="checkbox"/>	1.1% 2 réponses parmi 189
roumain	<input type="checkbox"/>	1.1% 2 réponses parmi 189
singhalais	<input type="checkbox"/>	1.1% 2 réponses parmi 189
suédois	<input type="checkbox"/>	1.1% 2 réponses parmi 189
afrikaans	<input type="checkbox"/>	0.5% 1 réponses parmi 189
biélorusse	<input type="checkbox"/>	0.5% 1 réponses parmi 189
bulgare	<input type="checkbox"/>	0.5% 1 réponses parmi 189
bambara	<input type="checkbox"/>	0.5% 1 réponses parmi 189
bengali	<input type="checkbox"/>	0.5% 1 réponses parmi 189
bosniaque	<input type="checkbox"/>	0.5% 1 réponses parmi 189
danois	<input type="checkbox"/>	0.5% 1 réponses parmi 189
croate	<input type="checkbox"/>	0.5% 1 réponses parmi 189
khmer	<input type="checkbox"/>	0.5% 1 réponses parmi 189
lingala	<input type="checkbox"/>	0.5% 1 réponses parmi 189
malgache	<input type="checkbox"/>	0.5% 1 réponses parmi 189
malais	<input type="checkbox"/>	0.5% 1 réponses parmi 189
norvégien	<input type="checkbox"/>	0.5% 1 réponses parmi 189
slovaque	<input type="checkbox"/>	0.5% 1 réponses parmi 189
tagalog	<input type="checkbox"/>	0.5% 1 réponses parmi 189
tongan (îles Tonga)	<input type="checkbox"/>	0.5% 1 réponses parmi 189
wolof	<input type="checkbox"/>	0.5% 1 réponses parmi 189

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
afar, abkhaze, avestique, akan, amharique, aragonais, assamais, avar, aymara, azéri, bachkir, bihari, bichlamar, tibétain, breton, tchéchène, chamorro, corse, cree, tchèque, tchouvache, gallois, maldivien, dzongkha, éwé, grec moderne (après 1453), espéranto, estonien, basque, peul, finnois, fidjien, féroïen, frison, irlandais, gaélique gaélique écossais, galicien, guarani, goudjrati, manx mannois, haoussa, hébreu, hindi, hiri motu, haïtien créole haïtien, hongrois, arménien, herero, interlingua (langue auxiliaire internationale), interlingue, igbo, yi de Sichuan, inupiaq, ido, islandais, inuktitut, javanais, kongo, kikuyu, kuanyama kwanyama, kazakh, groenlandais, kannada, kanouri, kashmiri, kurde, komi, cornique, kirghize, latin, luxembourgeois, ganda, limbourgeois, lao, lituanien, luba-katanga, letton, marshall, maori, macédonien, malayalam, mongol, moldave, marathe, maltais, birman, nauruan, bokmål, norvégien bokmål norvégien, ndébélé du Nord, népalais, ndonga, norvégien nynorsk nynorsk, norvégien, ndébélé du Sud, navaho, chewa, chichewa, nyanja, occitan (après 1500) provençal, ojibwa, galla, oriya, osète, pendjabi, pali, pachto, quechua, rhéto-roman, rundi, rwanda, sanskrit, sarde, sindhi, sami du Nord, sango, slovène, samoan, shona, somali, albanais, serbe, swati, sotho du Sud, soundanais, swahili, tamoul, télougou, tadjik, tha, tigrigna, turkmène, tswana, tsonga, tatar, twi, tahitien, ouïgour, ukrainien, ourdou, ouzbek, venda, volapük, wallon, xhosa, yiddish, yoruba, chuang zhuang, zoulou		Aucun témoignage
189 personnes ont répondu à cette question et ont choisi en moyenne 1.7 solutions par personne		

TABLE F.5: Réponses à la question n° 5 du second questionnaire

QUESTION	Le pays dans lequel vous enseignez	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
France	 74.9%	146 réponses parmi 195
Belgique	 8.2%	16 réponses parmi 195
Canada	 3.1%	6 réponses parmi 195
U.S.A.	 1.5%	3 réponses parmi 195
Congo	 1%	2 réponses parmi 195
Espagne	 1%	2 réponses parmi 195
Mexique	 1%	2 réponses parmi 195
Portugal	 1%	2 réponses parmi 195

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
Turquie	0	1% 2 réponses parmi 195
Allemagne	0	0.5% 1 réponses parmi 195
Angleterre	0	0.5% 1 réponses parmi 195
Argentine	0	0.5% 1 réponses parmi 195
Autriche	0	0.5% 1 réponses parmi 195
Bulgarie	0	0.5% 1 réponses parmi 195
Colombie	0	0.5% 1 réponses parmi 195
T.O.M.	0	0.5% 1 réponses parmi 195
Italie	0	0.5% 1 réponses parmi 195
Malaisie	0	0.5% 1 réponses parmi 195
Maroc	0	0.5% 1 réponses parmi 195
Martinique	0	0.5% 1 réponses parmi 195
Provence - occitanie	0	0.5% 1 réponses parmi 195
Suisse	0	0.5% 1 réponses parmi 195
Syrie	0	0.5% 1 réponses parmi 195
Togo	0	0.5% 1 réponses parmi 195
Pologne	0	0.5% 1 réponses parmi 195
195 personnes ont répondu à cette question et ont choisi en moyenne 1 solutions par personne		

TABLE F.6: Réponses à la question n° 6 du second questionnaire

QUESTION : Le nombre d'années durant lequel vous avez enseigné :

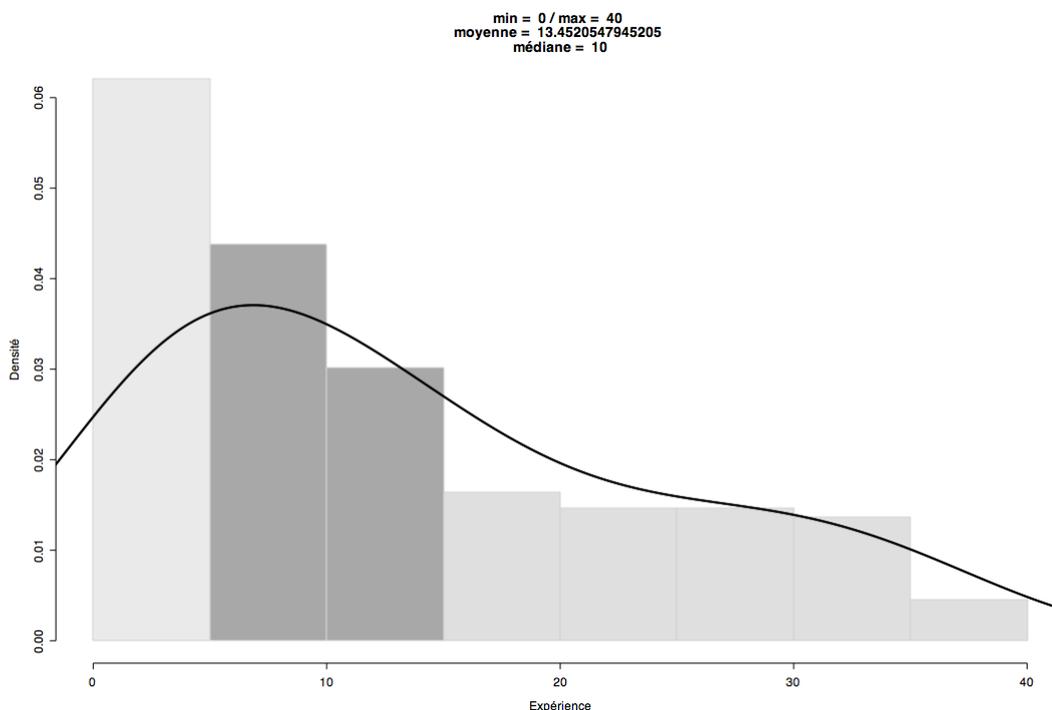


FIGURE F.1: Expérience des enseignants ayant répondu au questionnaire

QUESTION	Le (ou les) cadre(s) dans le(s)quel(s) vous enseignez : J'ai restreint volontairement les choix pour faciliter les calculs statistiques. Cependant, si vous avez besoin d'ajouter des catégories, utilisez la case prévue à cet effet. Vos propositions feront partie des choix laissés aux enseignants qui rempliront le questionnaire après vous. Essayez donc de rester aussi général que possible		
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL	
collège	<input type="checkbox"/>	39.8%	43 réponses parmi 108
lycée général	<input type="checkbox"/>	38.9%	42 réponses parmi 108
université (non-spécialistes)	<input type="checkbox"/>	20.4%	22 réponses parmi 108
université (spécialistes)	<input type="checkbox"/>	15.7%	17 réponses parmi 108
primaire	<input type="checkbox"/>	12%	13 réponses parmi 108
organisme de formation privé	<input type="checkbox"/>	7.4%	8 réponses parmi 108
Alliance française, école de langue	<input type="checkbox"/>	5.6%	6 réponses parmi 108
lycée professionnel	<input type="checkbox"/>	4.6%	5 réponses parmi 108
Centre de Formation par Apprentissage	<input type="checkbox"/>	3.7%	4 réponses parmi 108

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX		DÉTAIL
IUFM	<input type="checkbox"/>	3.7%	4 réponses parmi 108
maternelle	<input type="checkbox"/>	2.8%	3 réponses parmi 108
entreprise	<input type="checkbox"/>	1.9%	2 réponses parmi 108
Secondaire public Québec	<input type="checkbox"/>	1.9%	2 réponses parmi 108
ministère de l'immigration et des communautés culturelles	<input type="checkbox"/>	1.9%	2 réponses parmi 108
GRETA	<input type="checkbox"/>	0.9%	1 réponses parmi 108
Classe Relais	<input type="checkbox"/>	0.9%	1 réponses parmi 108
association d'aide à l'insertion	<input type="checkbox"/>	0.9%	1 réponses parmi 108
lycee bilingue	<input type="checkbox"/>	0.9%	1 réponses parmi 108
CDP	<input type="checkbox"/>	0.9%	1 réponses parmi 108
alphabetisation	<input type="checkbox"/>	0.9%	1 réponses parmi 108
bénévolat ds association	<input type="checkbox"/>	0.9%	1 réponses parmi 108
Didactique des LE	<input type="checkbox"/>	0.9%	1 réponses parmi 108
néant	<input type="checkbox"/>	0.9%	1 réponses parmi 108
<i>108 personnes ont répondu à cette question et ont choisi en moyenne 1.7 solutions par personne</i>			

TABLE F.7: Réponses à la question n° 8 du second questionnaire

QUESTION	Le (ou les) public(s) au(x)quel(s) vous enseignez :		
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX		
	DÉTAIL		
non-spécialistes	<input type="checkbox"/>	94.3%	100 réponses parmi 106
spécialistes	<input type="checkbox"/>	16%	17 réponses parmi 106
professionnels	<input type="checkbox"/>	7.5%	8 réponses parmi 106
<i>106 personnes ont répondu à cette question et ont choisi en moyenne 1.2 solutions par personne</i>			

TABLE F.8: Catégorisation des réponses à la question n° 8 du second questionnaire

QUESTION	Avez-vous un recueil de texte personnel ?		
CHOIX	% DE RÉPONSES		
	DÉTAIL		
oui	<input type="checkbox"/>	50,8%	100 réponses parmi 197
non	<input type="checkbox"/>	49,2%	97 réponses parmi 197

CHOIX	% DE RÉPONSES	DÉTAIL
-------	---------------	--------

TABLE F.9: Réponses à la question n° 9 du second questionnaire

QUESTION	Ce recueil est-il organisé ?	
CHOIX	% DE RÉPONSES	DÉTAIL
oui	 63%	63 réponses parmi 100
non	 37%	37 réponses parmi 100

TABLE F.10: Réponses à la question n° 10 du second questionnaire

QUESTION	Comment ce recueil est-il organisé ?
IDPROF	RÉPONSES
2	par niveau
5	par niveau, par classe, par niveau d'apprentissage
6	en fonction des niveaux de difficulté, des degrés d'apprentissage
8	selon les thèmes et le niveau
14	par thèmes et par filières d'enseignement, selon les cas
17	fonction ;
19	Progression grammaticale + réactivation du vocabulaire
20	par thème par difficulté
22	Par thèmes et selon les points de grammaire à revoir.
24	par thème
27	Selon les difficultés d'apprentissage ou plutôt selon la progression de la classe. Un texte comportant de nombreux impératifs illustrera l'étude de ce mode.
32	Par thème (d'après le contenu). Selon les faits linguistiques présents dans les textes.
38	par dates
41	Surtout selon des thèmes (fêtes, la vie des jeunes en Allemagne, l'immigration/les étrangers/la société multiculturelle, les villes ...) Mais j'ai aussi des textes que j'utilise uniquement dans les cours de traduction à la fac.
42	en fonction des thèmes et des points grammaticaux
43	Selon les points de langue qui seront abordés avec les élèves, dans l'ordre de difficulté croissante

IDPROF	RÉPONSES
47	par thèmes, principalement. par siècle, de temps en temps. par unités syntaxiques, plus rarement
52	par langue et par theme
53	Par thèmes : La ville, relations homme-femme, les dictatures... Dans chaque thème : par séquence possible : ex : un document iconographique, un texte, un document audio...
58	par objectifs visés
62	par thèmes, par catégories (presse, littérature, iconographie...)
66	par thèmes (le sport, la ville...) subdivisés par niveaux
69	organisation thématique
71	par thème et points grammaticaux et ou points de conjugaison.
72	Par compétences et par niveaux
73	Par niveau d'enseignement et par thème (culturel / civilisationnel)
75	par unités
77	Par thèmes ex sci-fi /teenage problems /The environment
78	par séquence
79	par thème par objectif linguistique
82	Par thème et par niveau d'études
83	thèmes...
90	par thèmes
92	par dossiers thématiques
103	thèmes/niveaux
104	Thèmes et niveaux de langues
109	J'ai 5 classeurs classés par niveaux...I pour Primaire-initiation , II pour Collège - Lycée, et toujours du plus "basique" au plus complexe...
111	par thème par niveau de classe par notion grammaticale
114	Par thèmes
117	par thèmes, soit grammaticaux, soit de civilisation
122	Par thématiques et selon des compétences d'ordre général, méthodologique, linguistique dont l'objet est le "cadre commun européen de référence pour les langues pages 26/27 entre autre"
125	activites sujets
130	selon le point de grammaire à traiter
131	Par niveaux. Tranches de 50 heures environ pour niveau 1.
133	en rapport avec la méthode employée
134	thèmes
137	PAR NIVEAU ET PAR THEMES
164	* competence linguistique * civilisation * activites ludiques
172	Par niveau d'abord, par thèmes culturels ensuite.

IDPROF	RÉPONSES
173	par thèmes du programme
190	C'est plus un recueil de ressources diverses que de textes. Il était jusque là hiérarchisé comme ci-dessous et j'essaye maintenant de l'organiser par compétence. Ce qui me conduit à sauvegarder la même ressource dans plusieurs dossiers. A terme j'avais l'intention de créer une base de données me permettant de pointer vers une même ressource à partir de différents critères, l'organisation hiérarchique n'étant pas satisfaisante.
192	Par activités d'apprentissage Par activités d'évaluation Par exercices Par la lecture de textes ciblés
194	Niveaux Points de grammaire
198	Par thèmes de civilisation ou grammaticaux
201	*grammaire *fêtes et "special days" *games *documents pour la compréhension écrite, classés par niveau (6è, 5è, ...) *idem pour la compréhension orale, l'expression orale et écrite *... et un classeur "miscellaneous"
202	par thèmes
203	par thème
208	par niveau et par thèmes
209	par niveau de classe, par thème de grammaire ou civilisation
233	Par compétence et par niveau
236	Thématique générale répondant à un objectif particulier
247	ça dépend du niveau, un recueil par cours, organisation différente
267	Par compétences et par thèmes et par programmations (prise en compte de la progression sur l'année)
63 personnes ont répondu à cette question	

TABLE F.11: Réponses à la question n° 11 du second questionnaire

QUESTION	<p>Afin de préciser cette organisation et de donner une idée de la manière dont vous les formulez, pouvez-vous donner un exemple des catégories employées pour le classement de ce recueil ?</p> <p>Si vous désirez hiérarchiser vos exemples vous pouvez le faire de la manière suivante :</p> <p>Dans cet exemple, les présents simple et progressif sont regroupés sous la rubrique présent faisant elle-même partie de la rubrique grammaire.</p> <ul style="list-style-type: none"> * Civilisation <ul style="list-style-type: none"> + Civil rights movement + Halloween + Sundance film festival * Grammaire <ul style="list-style-type: none"> + Présent <ul style="list-style-type: none"> - Présent progressif - Présent simple + Possessifs + Pronoms personnels
IDPROF	RÉPONSES
22	Exemple :
24	<p>* Groupement de textes "Creepy stories" avec comme point de grammaire le present perfect simple et progressif, en classe de première.</p> <p>Grammaire</p> <p>temps</p> <p>prépositions</p> <p>articles</p> <p>idioms</p>
27	<p>Thèmes par textes ou articles,doc. audio/video,exercices</p> <p>Langue générale : actualité</p> <p>Langue spécifique : CV,lettres,doc marketing,RH,CEI</p> <p>En fait, j'utilise d'une part une méthode où l'étude de la grammaire ou de la civilisation est proposée en fonction des textes "imposés".</p> <p>D'autre part, après quelques cours, je propose la lecture d'une histoire qui ne fait pas plus de 5 à 6 pages où nous retrouvons les principales règles grammaticales déjà vues et pouvons en "effleurer" de nouvelles.</p>

IDPROF	RÉPONSES
32	Grammaire : - temps : présent, parfait, prétérit - modes : subjonctif 1 -2 - possessifs Civilisation : - système scolaire - histoire de l'Allemagne - Autriche - Suisse Thèmes : - argent / société de consommation - vie en ville / à la campagne - éducation des enfants - intégration des étrangers dans la société - voyages - amour - famille
38	l'organisation est faite par niveau de difficulté
47	cela correspondant au schéma proposé
103	Themes/ vocabulaire + Grammaire + notions

IDPROF	RÉPONSES
109	<p>* Classeur "oralidad"</p> <ul style="list-style-type: none"> + Canciones <ul style="list-style-type: none"> - Infantiles y retahilas - Populares - "a la moda" + Juegos con palabras <ul style="list-style-type: none"> - Scrabble - bingos - Situaciones(ex. je vais chez le médecin, je cherche un travail et je me présent, ...) + Ejercicios de fonética <p>* "Gramàtica"</p> <ul style="list-style-type: none"> + la pronunciacion, alfabeto, acentuacion + los numeros + formacion de palabras + pronombres + preposiciones + otros + Verbos <ul style="list-style-type: none"> -presentacion tiempos verbales -conjugaciones, infinitivos, ser y estar, ... -Presente -Pasado -Futuro -Condicional -Subjuntivo,... + resúmenes, evaluaciones, test <p>* Dibujos</p> <ul style="list-style-type: none"> + por temas : la ropa, la casa, la hora, alimentacion, ... + Comics, B.D., ... : (nivel I)saludos, pensamientos, descripciones, tengo frio/calor/sed,....(nivel II) la naturaleza, la familia, las fiestas, la imàgen(famosos, politicos,...) la educacion, accidentes, viajes a paises latinoamericanos... + anuncios <p>* Fichas</p> <ul style="list-style-type: none"> +Primaria (actividades ya realizadas o preparadas para presentarlas en clase, sobre todo a nivel de iniciacion) : <ul style="list-style-type: none"> -los colores, los meses del año, las descripciones, la casa,el cuerpo humano, la ropa, los animales, hablando por teléfono, ... +Textos <ul style="list-style-type: none"> - recopilatorio por temas, por ejemplo, "San Valentin", Las fiestas, La gastronomia, las supersticiones(ex, Galicia y la mitologia, la suerte y sus simbolos, ... - cuentos, leyendas, <p>* Varios</p> <ul style="list-style-type: none"> + juegos de logica, sopas de letras, otros + unidades didacticas + Ideas de actividades (sin leer o sin aplicar) + Sin clasificar (fotos, mapas, recortes,

IDPROF	RÉPONSES
114	<p>Fêtes (Hallowe'en, Christmas, les fêtes telles que Diwali, Hannukah etc etc)</p> <p>Pays (histoire, géographie, faune, flore, grands hommes, arts, cinéma, etc etc)</p> <p>Grands hommes (JFK, MLK, Mandela, présidents des USA, rois et reines de Grande Bretagne etc etc)</p> <p>Grammaire (groupe nominal, groupe verbal, syntaxe)</p> <p>Vocabulaire (par champs lexicaux)</p> <p>Chansons</p> <p>Textes pour le bac (lycée)</p> <p>exercices de grammaire selon les thèmes</p> <p>Miscellaneous (par thèmes : addiction : to video games, drugs, cigarettes, computers, ...ou encore : supernatural : superstition, etc etc)</p> <p>films : scripts, vocabulary, etc etc</p>
125	<p>Tâche (exemple) :C'est Noel. Il faut penser aux cadeaux ...</p> <p>Activités : faire des courses, décrire des objets, connaître les traditions de Noel en France, Québec... Comparer avec les traditions du Portugal</p> <p>Moyens linguistiques : Avoir besoin de ; couleurs et matières, chiffres et prix, nom de boutiques (chez, à la, au)</p>
131	<p>*Documents authentiques</p> <p>+ élémentaire (documents simples + act. progression vocabulaire, grammaire, syntaxe)</p> <p>+ Niveaux 1 - 4</p> <p>+ Intermédiaire (articles, structure discours + thèmes civ.)</p> <p>+ Niveaux 1 - 4</p> <p>+ Avancés (sans subdivisions)</p> <p>+ chansons poésie</p>
164	<p>* Compétence linguistique</p> <p>+ Le style indirect</p> <p>+ La voix passive</p> <p>+ les comparatifs et superlatifs</p> <p>+ les temps</p> <p>- les présents</p> <p>- les temps du passé</p> <p>*Civilisation</p> <p>+ Immigration waves</p> <p>+ Native Americans</p> <p>+ The civil rights movement</p> <p>+ Artetc / etc</p>
172	je ne vais pas au-delà du niveau et du thème

IDPROF	RÉPONSES
190	<p>Ancienne organisation (très semblable à l'exemple) ex collègue</p> <p>*At home</p> <ul style="list-style-type: none"> + my house + my family + me - body - clothes + meals <p>* Countries</p> <p>.....</p> <p>Nouvelle organisation (pas encore au point)</p> <p>* Je suis capable de parler de moi</p> <ul style="list-style-type: none"> + my family + me - body <p>* Je suis capable de parler de mes goûts</p> <ul style="list-style-type: none"> + food + colours + sports
208	<p>les thèmes en terminale</p> <ul style="list-style-type: none"> - apprendre une langue étrangère dès la maternelle - le bonheur - la gastronomie - les conduites à risques - l'Europe - esclavage et colonialisme - la guerre - ...
209	<p>*niveau -present perfect</p>

IDPROF	RÉPONSES
236	<ul style="list-style-type: none"> * Ecrit <ul style="list-style-type: none"> + compréhension <ul style="list-style-type: none"> - niveau A1 - niveau A2... + expression <ul style="list-style-type: none"> - niveau A1 - niveau A2... * Interculturel <ul style="list-style-type: none"> + stéréotypes + bd + poèmes * Civilisation <ul style="list-style-type: none"> + Actualités + géographie + histoire + enseignement + fêtes + littérature + politique + peinture, sculpture + cuisine * Francophonie <ul style="list-style-type: none"> + géographie + histoire + fêtes + littérature + politique + art + musique + cuisine
247	<p>Impossible de répondre à cette question en détail, j'utilise actuellement 5 ou 6 recueils différents pour le présentiel, 2 en EAD (enseignement à distance).</p> <p>J'enseigne la langue de spécialité, d'une façon générale, mes recueils intègrent : du lexique spé, des textes spé, des cadrages méthodologiques. Pas de grammaire ni de civi</p>
18 personnes ont répondu à cette question	

TABLE F.12: Réponses à la question n° 12 du second questionnaire

Les réponses aux deux questions précédentes (11 et 12) ont été catégorisées dans le tableau 5.2.

F.2 Apports potentiels du TAL dans l'organisation de la base de textes

F.2.1 Présentation de la rubrique

Toutes les questions situées sur cette page sont prospectives. Elles concernent les fonctionnalités potentielles de la base de texte, aussi bien au niveau de la recherche de textes, que de leur exploitation en classe. Il vous faut imaginer que vous disposez d'un outil permettant d'avoir accès à un ensemble de textes satisfaisant un ou plusieurs des critères ci-dessous ou d'afficher toutes ces caractéristiques pour chaque texte sous formes de statistiques ou de **mise en évidence** de ces phénomènes.

Enfin, pour chacun de ces critères potentiels, il faut imaginer qu'ils peuvent être combinés les uns avec les autres. Par exemple rechercher des textes qui contiennent un maximum d'occurrences d'adjectifs au pluriel et dont le pluriel se forme comme celui de "travail".

Dans cette optique, évaluez la pertinence des critères suivants pour la sélection d'un document pour votre cours de langue

F.2.2 Au niveau du mot

QUESTION	La mise en évidence des mots grammaticaux d'un texte.	
CHOIX	% DE RÉPONSES	DÉTAIL
1 - pas du tout pertinent	<input type="checkbox"/> 2.6%	2 réponses parmi 78
2 - peu pertinent	<input type="checkbox"/> 17.9%	14 réponses parmi 78
3 - pertinent	<input type="checkbox"/> 48.7%	38 réponses parmi 78
4 - très pertinent	<input type="checkbox"/> 26.9%	21 réponses parmi 78
5 - primordial	<input type="checkbox"/> 3.8%	3 réponses parmi 78

TABLE F.13: Réponses à la question n° 16 du second questionnaire

QUESTION	Le nombre de ou la liste des mots grammaticaux / non-grammaticaux différents dans un texte.	
CHOIX	% DE RÉPONSES	DÉTAIL
1 - pas du tout pertinent	<input type="checkbox"/> 9.9%	8 réponses parmi 81
2 - peu pertinent	<input type="checkbox"/> 39.5%	32 réponses parmi 81
3 - pertinent	<input type="checkbox"/> 38.3%	31 réponses parmi 81
4 - très pertinent	<input type="checkbox"/> 12.3%	10 réponses parmi 81
5 - primordial		Aucun témoignage

TABLE F.14: Réponses à la question n° 15 du second questionnaire

QUESTION	Le nombre de ou la liste des mots différents dans un texte (cheval, chevaux = 1 mot)		
CHOIX	% DE RÉPONSES		DÉTAIL
1 - pas du tout pertinent		12.3%	10 réponses parmi 81
2 - peu pertinent		43.2%	35 réponses parmi 81
3 - pertinent		33.3%	27 réponses parmi 81
4 - très pertinent		9.9%	8 réponses parmi 81
5 - primordial			Aucun témoignage

TABLE F.15: Réponses à la question n° 14 du second questionnaire

QUESTION	Le nombre / la liste des verbes différents présents dans un texte.		
CHOIX	% DE RÉPONSES		DÉTAIL
1 - pas du tout pertinent		5.3%	4 réponses parmi 76
2 - peu pertinent		22.4%	17 réponses parmi 76
3 - pertinent		42.1%	32 réponses parmi 76
4 - très pertinent		28.9%	22 réponses parmi 76
5 - primordial		1.3%	1 réponses parmi 76

TABLE F.16: Réponses à la question n° 17 du second questionnaire

QUESTION	Le nombre / la liste des noms et des adjectifs différents présent dans un texte.		
CHOIX	% DE RÉPONSES		DÉTAIL
1 - pas du tout pertinent		5.1%	4 réponses parmi 78
2 - peu pertinent		29.5%	23 réponses parmi 78
3 - pertinent		43.6%	34 réponses parmi 78
4 - très pertinent		20.5%	16 réponses parmi 78
5 - primordial		1.3%	1 réponses parmi 78

TABLE F.17: Réponses à la question n° 18 du second questionnaire

QUESTION	La mise en évidence des mots se prononçant de la même manière dans le texte ex : pin / pain	
CHOIX	% DE RÉPONSES	DÉTAIL
1 - pas du tout pertinent		16% 12 réponses parmi 75
2 - peu pertinent		28% 21 réponses parmi 75
3 - pertinent		26.7% 20 réponses parmi 75
4 - très pertinent		21.3% 16 réponses parmi 75
5 - primordial		8% 6 réponses parmi 75

TABLE F.18: Réponses à la question n° 19 du second questionnaire

F.2.3 Formes fléchies

QUESTION	Le nombre de ou la liste des temps / personnes différents employés	
CHOIX	% DE RÉPONSES	DÉTAIL
1 - pas du tout pertinent		7.7% 6 réponses parmi 78
2 - peu pertinent		15.4% 12 réponses parmi 78
3 - pertinent		43.6% 34 réponses parmi 78
4 - très pertinent		25.6% 20 réponses parmi 78
5 - primordial		7.7% 6 réponses parmi 78

TABLE F.19: Réponses à la question n° 20 du second questionnaire

QUESTION	La mise en évidence des temps / personnes différents dans tout le texte.	
CHOIX	% DE RÉPONSES	DÉTAIL
1 - pas du tout pertinent		3.8% 3 réponses parmi 78
2 - peu pertinent		16.7% 13 réponses parmi 78
3 - pertinent		46.2% 36 réponses parmi 78
4 - très pertinent		19.2% 15 réponses parmi 78
5 - primordial		14.1% 11 réponses parmi 78

TABLE F.20: Réponses à la question n° 21 du second questionnaire

QUESTION	Le nombre de / la liste des "modèles de conjugaison" différents employés dans le texte pour un temps donné On conviendra que aller et travailler suivent un même modèle de conjugaison à l'imparfait, mais qu'ils suivent des modèles différents au présent		
CHOIX	% DE RÉPONSES	DÉTAIL	
1 - pas du tout pertinent		10%	8 réponses parmi 80
2 - peu pertinent		31.3%	25 réponses parmi 80
3 - pertinent		28.7%	23 réponses parmi 80
4 - très pertinent		26.3%	21 réponses parmi 80
5 - primordial		2.5%	2 réponses parmi 80

TABLE F.21: Réponses à la question n° 22 du second questionnaire

QUESTION	Le nombre de / la liste des "modèles de flexion" différents de noms ou d'adjectifs employés dans le texte pour une flexion donnée On conviendra que "spacial" et "illégal" suivent un même modèle pour le passage de l'adjectif au pluriel		
CHOIX	% DE RÉPONSES	DÉTAIL	
1 - pas du tout pertinent		11.4%	9 réponses parmi 79
2 - peu pertinent		43%	34 réponses parmi 79
3 - pertinent		31.6%	25 réponses parmi 79
4 - très pertinent		13.9%	11 réponses parmi 79
5 - primordial			Aucun témoignage

TABLE F.22: Réponses à la question n° 23 du second questionnaire

F.2.4 Au niveau de la phrase

QUESTION	Le nombre de ou la mise en évidence des cooccurrences de deux mots Ex : tous les emplois de "jouer" et de "musique" au sein d'une même phrase.		
CHOIX	% DE RÉPONSES	DÉTAIL	
1 - pas du tout pertinent		15.4%	12 réponses parmi 78
2 - peu pertinent		34.6%	27 réponses parmi 78
3 - pertinent		38.5%	30 réponses parmi 78
4 - très pertinent		6.4%	5 réponses parmi 78
5 - primordial		5.1%	4 réponses parmi 78

CHOIX	% DE RÉPONSES	DÉTAIL
-------	---------------	--------

TABLE F.23: Réponses à la question n° 30 du second questionnaire

QUESTION	La présence / le comptage / la mise en évidence d'occurrences d'une structure donnée ex : haber que + infinitif en espagnol, looking forward to + ing en anglais, il faut que + GN + subjonctif en français, etc	
CHOIX	% DE RÉPONSES	DÉTAIL
1 - pas du tout pertinent	<input type="checkbox"/> 3.9%	3 réponses parmi 77
2 - peu pertinent	<input type="checkbox"/> 11.7%	9 réponses parmi 77
3 - pertinent	<input type="checkbox"/> 35.1%	27 réponses parmi 77
4 - très pertinent	<input type="checkbox"/> 32.5%	25 réponses parmi 77
5 - primordial	<input type="checkbox"/> 16.9%	13 réponses parmi 77

TABLE F.24: Réponses à la question n° 24 du second questionnaire

QUESTION	Pouvez-vous donner un ou plusieurs exemple(s) de structure que vous pourriez vouloir rechercher ou mettre en évidence ?
IDPROF	RÉPONSES
39	tener que+infinitif, haber de+ infinitif, ser necesario que + subjonctif, querer que + subjonctif...
85	seguir+ gérondif
65	se me ocurre una cosa, llevar + durée + gérondif, estar+ gérondif, etc
32	verbe de modalité + infinitif / subordonnées infinitives.
103	le conseil : should / had better / If I were you
54	es posible que + subjonctif
109	l'application des prépositions en espagnol
57	GUSTAR
21	GN+ devoir+infinif il faut que+GN+subjonctif il faut+infinitif si+GN+imparfait+GN+conditionnel
166	TO LIKE + ING
165	to want to do somethinh to want something to want somebody to do something to make somebody do something Present perfect/past simple+for since ago If, les trois "formes"

IDPROF	RÉPONSES
164	Les verbes ou expressions suivies d'un gérondif / d'une base verbale /d'un infinitif (to)
167	verbes d'opinion et de sentiment + subjonctif (en italien) verbes impersonnels (en italien) andare a + infinitif (en italien) cercare di + infinitif (en italien)
172	pedir que+ subjonctif, no ... sino,
68	proposition infinitive if+présent, will+V
184	Structures répétitives des programmes
185	little did he know ... that / no sooner had he.. than.. / the more.. the more../ to have sbdy do sthg.../ etc...
47	Le devoir : tener que + inf. mais deber (ensemble vide) + inf
108	used to /would
22	want to + V want something want somebody to do something have somebody do something make somebody do something it's the first time + present perfect just + present perfect if only... I wish...
31	' '
194	anglais : have to, I wish I were,I had..., I'd rather... espagnol : tener que français : vouloir +ind., devoir + ind,
101	comparatif; subjonctif 2
196	les expressions impersonnelles (ex : il faut ...)
198	Formes verbales
200	he learns English to go to America
201	used to + BV; would + BV (répétition); look + for, look + at, look + after; ...
24	style indirect et concordance des temps : il a dit qu'il avait été malade Subjonctif/indicatif : je doute qu'il puisse.../je sais qu'il peut...
206	prétérit vs present perfect
208	emploi des temps du passé : imparfait et passé composé emploi du verbe impersonnel
50	les structures fonctionnant en V someone to V, an V someone V, V suivi de V_ing
213	although unless
215	make someone do something
216	Occurrence des deux présents ou des deux prétérits dans un texte en anglais par exemple.
225	INSISTANCE (c'est lui qui...)

IDPROF	RÉPONSES
229	structure du subjonctif Si + imparfait, conditionnel présent Si +.....
230	je suis content/triste/heureux/ravi que + subjonctif accords à la forme pronominale : ils se sont promenés, ils se sont lavé les mains les doubles pronoms : il le lui a dit, on leur en a parlé, il l'y place expression de la comparaison : plus ... que, moins ... que, autant ... que, aussi ... que, meilleur, pire, mieux etc
231	-adverbes en -ly (badly) versus adjectifs en -ly (friendly) -verbes de deux syllabes (avant flexion) (pour un travail sur leur accentuation) -mots en -tion, -ic, -ate (pour leur accentuation) -phrasal verbs -for + V-ING
233	donner quelque chose à quelqu'un espérer + infinitif avoir envie de + infinitif etre intéresser par quelque chose
236	verbes + de verbes + à xxx que + subjonctif
237	C'est un,une / pallier qqch
238	les participes passés les articles les formes verbales comme le passé surcomposé
232	J'aimerais que + subj / J'espère que + indicatif
241	subjonctif, hypothèse, cause, conséquence, gérondif, mise en valeur...
242	Les verbes et locutions verbales introducteurs du subjonctif (craindre que, avoir peur que, avoir envie que, vouloir que...)
243	la négation , Si condition, hypothèse, regret la concordance des temps, passé composé -imparfait-plus-que parfait etc
251	faire faire qqch à qqun, il faut que,
66	les compléments circonstanciels
252	je monte en haut chercher un livre :la semantique du verbe monter ne demande pasl'emploi de "en haut".le sens impose la structure de la phrase .
26	go to (niveau élémentaire) adj-N-ed (intermédiaire)N's N ou NN
27	Je n'ai guère d'exemples mais je voulais préciser que mon manque d'enthousiasme (3 en général) s'explique par le fait, qu'utilisant des textes courts, je fais rapidement ce travail de présence, comptage... et que ce travail peut bien sûr être demandé aux étudiants.
119	l'inversion du sujet

IDPROF	RÉPONSES
40	cuando + subj; querer que + subj, decir que +subj et decir que +indicatif; no... sino (que)(il n'est pas blanc mais noir : no es blanco sino negro) ; ir a + infinitifcomo si +subj imparfait ; para +infinitf; para que + subjonctif; ...
88	par exemple dans un texte, rechercher la structure de divers auxiliaires modaux et leur rôle.
253	Expressions de sentiments + subjobctif. Exemple : "Ça me ferait plaisir que tu viennes avec nous au Biodôme"
131	que + subjonctif
56 personnes ont répondu à cette question	

TABLE F.25: Réponses à la question n° 25 du second questionnaire

QUESTION	La présence / le comptage / la mise en évidence d'occurrences appartenant à une même catégorie syntaxique ex : subordonnée relative		
CHOIX	% DE RÉPONSES		DÉTAIL
1 - pas du tout pertinent	<input type="checkbox"/>	6.6%	5 réponses parmi 76
2 - peu pertinent	<input type="checkbox"/>	18.4%	14 réponses parmi 76
3 - pertinent	<input type="checkbox"/>	40.8%	31 réponses parmi 76
4 - très pertinent	<input type="checkbox"/>	27.6%	21 réponses parmi 76
5 - primordial	<input type="checkbox"/>	6.6%	5 réponses parmi 76

TABLE F.26: Réponses à la question n° 26 du second questionnaire

QUESTION	Pouvez-vous donner un ou plusieurs exemple(s) de "catégorie(s) syntaxique(s)" que vous pourriez vouloir rechercher ou mettre en évidence ?
IDPROF	RÉPONSES
32	subordonnées conjonctives. subordonnées infinitives. interrogations globales directes et indirectes. compléments de lieux (locatifs / directifs) discours direct / indirect
57	PARA + INFINITIF PARA QUE + SUBJONCTIF
21	subodonnee relative
38	adjectifs, pronom
166	LES CONNECTEURS

IDPROF	RÉPONSES
165	link words, les subordonnées if when to/in order to etc.
164	Les comparatifs ou superlatifs
167	phrase hypothétique phrase passive paricipe absolu
172	subordonnée temporelle avec subjonctif, es + adjectif (difficile, normal, ...) suivi d'infinitif ou de que+ subjonctif,...
68	proposition infinitive
185	relatives, forme verbale dans les complétives, etc..
47	Verbe de volonté + subj => querer/mandar...+ subj
108	sentiment
45	querer que + subjonctif soler + infinitif
31	'-'
101	chronologie ; expression de la cause
23	conjonctives/relatives/subjonctives
198	Formes verbales
201	subordonnées relatives, justement ! subordonnées de temps introduites pas WHEN, AS SOON AS conditionnelles (IF) ...
24	le livre dont tu m'as parlé Lequel d'entre vous.. la chaise sur laquelle je suis assise
206	subordonnée infinitive
208	subordonnée relative expression de la condition ou de l'hypothèse
50	expressions de temps par exemple
213	infinitives
215	proposition infinitive
216	les adjectifs au comparatif en anglais
225	proposition infinitive / tags /
230	pronoms ou adjectifs possessifs ; pronoms et adjectifs indéfinis ; articles ; pronoms ou adjectifs démonstratifs ; subordonnées complétives, relatives
231	-les subordonnées de temps (coordonnance des temps délicate en anglais) -les subordonnées nominales en fonction sujet (ex : Killing a lamb is a horrible thing.)
233	les coordinaisons, les subordinaisons, les compléments relatifs
236	Relatives Complétives

IDPROF	RÉPONSES
238	phrase nominale
241	apposition, subordonnées (cause, conséquence, but, opposition, condition, relatives...)
242	les expressions du temps, du lieu (complément circonstanciels), de but, etc.. pendant .. ; dans .. ; depuis.. ..
243	qualifier par une relative usage des relatifs après préposition
66	phrase relative, phrase conditionnelle
252	le cheval est mort (est un verbe "auxiliaire" au présent qui désigne un événement passé "mort" il y a une nuance ici.
26	relatives, complétives, adverbes, V-ing (pour les spécialistes)
27	nuances dues à l'emploi de l'indicatif ou du subjonctif en espagnol
88	le discours indirect
253	Place de la négation/ de l'adverbe Exemples : Je ne tiens vraiment pas à aller 'la cabane à sucre. Je n'ai pas encore goûté à la tige d'érable.
131	articulateurs logiques
42 personnes ont répondu à cette question	

TABLE F.27: Réponses à la question n° 27 du second questionnaire

QUESTION	La présence / le comptage / la mise en évidence d'occurrences de structures par rapport à un contexte donné ex : avoir + participe passé dans une subordonnée introduite par "que"		
CHOIX	% DE RÉPONSES		DÉTAIL
1 - pas du tout pertinent	<input type="checkbox"/>	2.6%	2 réponses parmi 77
2 - peu pertinent	<input type="checkbox"/>	35.1%	27 réponses parmi 77
3 - pertinent	<input type="checkbox"/>	44.2%	34 réponses parmi 77
4 - très pertinent	<input type="checkbox"/>	11.7%	9 réponses parmi 77
5 - primordial	<input type="checkbox"/>	6.5%	5 réponses parmi 77

TABLE F.28: Réponses à la question n° 28 du second questionnaire

QUESTION	RÉPONSES
	Pouvez-vous donner un ou plusieurs exemple(s) de structures en contexte que vous pourriez vouloir rechercher ou mettre en évidence ?
IDPROF	RÉPONSES
32	Emploi du subjonctif 1 et 2.

IDPROF	RÉPONSES
54	Acaso+ subjonctif
64	Would+BV ayant valeur d'habitude dans le passé
57	tener que
165	'_'
164	want + proposition infinitive Conjonction de temps + présent et non futur
172	pas d'idée particulière
68	durée avec for et since
184	Structures répétitives des programmes
47	Un seul auxiliaire pour construire les temps composés : Haber+ ppassé. (Le verbe Tener introduit une autre nuance peu utile au niveau des lycéens et encore moins aux collégiens)
108	want to, passif
22	After + BVING
31	'_'
198	temps
24	Phrases introduites par :alors que, bien que = concordance des temps avec la phrase principale :bien qu'il soit fatigué, il a répondu au questionnaire style indirect avec concordance des temps au passé/futur : il m'a dit qu'il avait répondu...
206	have something done
208	emploi du subjonctif dans la relative emploi des articulateurs dans l'argumentation
213	le but, les objectifs
215	when-futur + présent
216	Exclamatives en anglais
230	concordance des temps dans une subordonnée
231	-FOR + sujet + TO + Base Verbale -so that + modal
238	les articulateurs logiques lorsque ainsi mais
242	le discours rapporté au style indirect (au présent puis au passé pour la concordance des temps : ex : j'ai dit qu'il viendrait)
26	un verbe au present perfect/prétérit après/avant 'since' V-ing après before/after
40	le subjonctif employé dans une subordonnée pour exprimer le futur (en espagnol)
88	l'usage du conditionnel, le gérondif....
267	Jeux de la marchande avec des questions
28 personnes ont répondu à cette question	

TABLE F.29: Réponses à la question n° 29 du second questionnaire

F.3 Utilisation des textes

Nous sommes bien conscients qu'un même texte peut être utilisé de plusieurs manières différentes et avec plusieurs buts pédagogiques, cela sera bien évidemment pris en compte au niveau du dépouillement des réponses.

F.3.1 Exercices Lacunaires

QUESTION	Avez-vous recours à des exercices lacunaires dans votre enseignement ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non	13.4%	9 réponses parmi 67
oui	86.6%	58 réponses parmi 67

TABLE F.30: Réponses à la question n° 31 du second questionnaire

NB : La suite des questions ne s'affichait qu'en cas de réponse positive à la question ci-dessus.

QUESTION	Sélectionnez les types d'exercices lacunaires que vous utilisez :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
exercice à "objectif linguistique" (ex : suppression de tous les adverbes d'un texte, de tous les adjectifs dits "appréciatifs" d'un texte, de tous les verbes d'un texte en demandant à l'apprenant de les conjuguer au temps approprié, etc...)	78.9%	45 réponses parmi 57
autre exercice de compréhension (Suppression de plusieurs mots-clés différents)	64.9%	37 réponses parmi 57
Reconstitution du texte après écoute	63.2%	36 réponses parmi 57

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
de compréhension des mots après un travail phonologique.	<input type="checkbox"/>	31.6% 18 réponses parmi 57
suppression des mots grammaticaux	<input type="checkbox"/>	29.8% 17 réponses parmi 57
travail sur l'inférence du sens d'un mot	<input type="checkbox"/>	24.6% 14 réponses parmi 57
suppression de toutes les occurrences d'un même mot	<input type="checkbox"/>	24.6% 14 réponses parmi 57
dans un texte de civilisation	<input type="checkbox"/>	24.6% 14 réponses parmi 57
suppression aléatoire d'un mot/7 par exemple sans prise en compte de sa nature ou fonction	<input type="checkbox"/>	10.5% 6 réponses parmi 57
dans une chanson	<input type="checkbox"/>	7% 4 réponses parmi 57
suppression des mots connus, des champs lexicaux connus et à compléter après écoute(s)	<input type="checkbox"/>	7% 4 réponses parmi 57
classification des champs dans un texte	<input type="checkbox"/>	5.3% 3 réponses parmi 57
dans une recette, les verbes	<input type="checkbox"/>	1.8% 1 réponses parmi 57
57 personnes ont répondu à cette question et ont choisi en moyenne 3.4 solutions par personne		

TABLE F.31: Réponses à la question n° 32 du second questionnaire

Les données du tableau précédent ont été reformatées, comme indiqué en § a) p. 232 :

QUESTION	Sélectionnez les types d'exercices lacunaires que vous utilisez :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
exercice à "objectif linguistique" (ex : suppression de tous les adverbes d'un texte, de tous les adjectifs dits "appréciatifs" d'un texte, de tous les verbes d'un texte en demandant à l'apprenant de les conjuguer au temps approprié, etc...)	 80.7%	46 réponses parmi 57
Reconstitution du texte après écoute	 77.2%	44 réponses parmi 57
autre exercice de compréhension (Suppression de plusieurs mots-clés différents)	 75.4%	43 réponses parmi 57
travail sur l'inférence du sens d'un mot suppression de toutes les occurrences d'un même mot	 24.6%	14 réponses parmi 57
57 personnes ont répondu à cette question et ont choisi en moyenne 2.6 solutions par personne		

TABLE F.32: Catégorisation des réponses à la question n° 32 du second questionnaire

Pour chacun des types d'exercices lacunaires sélectionnés (parmi ceux proposés initialement, à savoir « travail sur l'inférence du sens d'un mot », « exercice à "objectif linguistique" » et « autre exercice de compréhension ») la même séquence de questions était proposée. Nous avons converti en mots les longueurs de textes quand cela était possible :

a) Travail sur l'inférence du sens d'un mot

QUESTION	Longueur moyenne d'un texte pour un exercice lacunaire de travail sur l'inférence du sens d'un mot : en mots ou en lignes, merci de préciser l'unité
IDPROF	RÉPONSES
146	10 lignes
64	à l'intérieur d'un texte étudié
166	10 LIGNES
164	120 mots
108	12 lignes
201	10 lignes
206	10-15 lignes
231	10 à 15 lignes
237	20
232	150
252	éviter l'ambiguïté d'untex
253	cela varie beaucoup selon le ni
12 personnes ont répondu à cette question	

TABLE F.33: Réponses à la question n° 33 du second questionnaire

QUESTION	Longueur moyenne d'un texte pour un exercice lacunaire de travail sur l'inférence du sens d'un mot : en mots
IDPROF	RÉPONSES
146	100
166	100
164	120
108	120
201	100
206	125
231	125
237	200
232	150
9 personnes ont répondu à cette question	

TABLE F.34: Normalisation des réponses à la question n° 33 du second questionnaire

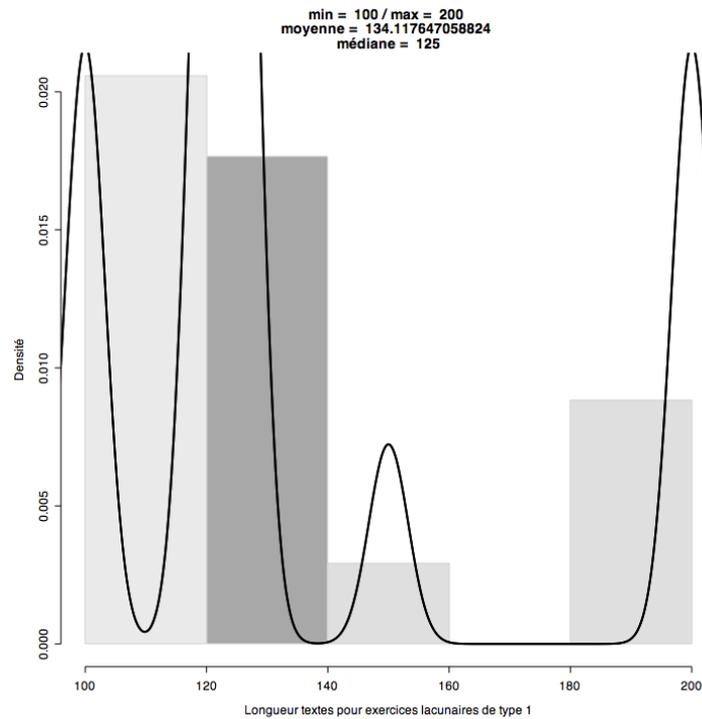


FIGURE F.2: Longueur d'un texte pour un exercice lacunaire de travail sur l'inférence du sens d'un mot

QUESTION	Nombre moyen de mots escamoté par texte (Ordre de grandeur)
IDPROF	RÉPONSES
146	15
64	3
166	10
164	8
108	5
201	10
206	10
231	12
237	12
9 personnes ont répondu à cette question	

TABLE F.35: Réponses à la question n° 34 du second questionnaire

QUESTION	Dans un texte utilisé dans le cadre d'un travail sur l'inférence du sens d'un mot (à part le mot escamoté) la présence de vocabulaire inconnu est :		
CHOIX	% DE RÉPONSES		DÉTAIL
recherchée	<input type="text"/>	23.1%	3 réponses parmi 13
tolérée (ou non pertinent)	<input type="text"/>	53.8%	7 réponses parmi 13
proscrite	<input type="text"/>	23.1%	3 réponses parmi 13

TABLE F.36: Réponses à la question n° 35 du second questionnaire

QUESTION	Sélectionner les critères qui peuvent vous influencer dans le choix d' un texte à utiliser comme support d'un exercice lacunaire d'inférence du sens d'un mot :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
thème	 84.6%	11 réponses parmi 13
vocabulaire	 76.9%	10 réponses parmi 13
longueur du texte	 46.2%	6 réponses parmi 13
objectifs de communication	 46.2%	6 réponses parmi 13
structures	 38.5%	5 réponses parmi 13
type de texte	 38.5%	5 réponses parmi 13
date	 15.4%	2 réponses parmi 13
auteur	 7.7%	1 réponses parmi 13
bac, style		Aucun témoignage
13 personnes ont répondu à cette question et ont choisi en moyenne 3.5 solutions par personne		

TABLE F.38: Réponses à la question n° 37 du second questionnaire

QUESTION	Pouvez-vous nous donner un exemple pour chacune des caractéristiques que vous avez sélectionné(es) ? ex : thème : musique
IDPROF	RÉPONSES
146	thème : musique vocabulaire : la joie, l'amitié structures : les collocations, les expressions figées de la joie type de texte : un dialogue entre amis longueur : pas très long (500 mots maxi)
64	no
166	LE MONDE DU TRAVAIL/LE SPORT/....
164	thème : art / en rapport avec le thème abordé lors des cours précédents longueur : jamais plus de 120 mots
108	amour, études, la révolte
206	Native Americans buffaloes, bows, arrows, tipis 10-15 lignes invitation
231	thème : justice (en lien avec le thème travaillé) vocabulaire : au moins 5 mots a priori inconnus longueur du texte : jamais plus de 15 lignes
237	thème : saint-valentin, les couleurs, le racisme, l'esclavage, l'adolescence type de texte :le biographique, policier, poème

IDPROF	RÉPONSES
232	Tout dépend de l'objectif du cours, exemple : parler de sa santé, acheter des vêtements, chercher un emploi...
251	Thème : informatique Vocabulaire : disque dur, cd... Type de texte : en fait, origine : magazine, manuel, internet, publicité... Longueur de texte : nombre de mots Date : date de publication
252	on sait q'un mot en français peut avoir plusieurs sens, or la structure ou le contexte determine le sens de ce mot : ex : une femme bonne et une bonne femme (ds le 1 ex une servante et ds le 2 ex une femme généreuse altruiste...
253	Je ne suis pas certaine de comprendre ce que vous cherchez ici.
12 personnes ont répondu à cette question	

TABLE F.39: Réponses à la question n° 38 du second questionnaire

b) Exercice à « objectif linguistique »

QUESTION	Longueur moyenne d'un texte pour un exercice lacunaire à "objectif linguistique" : en mots ou en lignes, merci de préciser l'unité
IDPROF	RÉPONSES
146	200 mots (avancés)
39	5 lignes
32	5-8
54	15 lignes
109	10
57	10 LIGNES
21	10 lignes
166	15
164	120 mots
172	12 lignes
68	5 lignes
47	pas plus de 15 lignes
108	12 lignes à 40
22	100 mots
45	7 à 10 lignes :
23	20 lignes
201	10-20 lignes (suivants la classe
24	500 à 1000 mots
206	10-15 lignes
208	200 mots
213	20 lignes

IDPROF	RÉPONSES
215	50 mots
216	10 lignes
229	200 mots
230	20-30 vers (poème ou chanson)
231	10 à 15 lignes
237	10 lignes
239	10 lignes
232	150 mots environ
241	12 à 15 phrases
242	10
66	5 à 10 lignes
252	selon le niveau des apprenants
18	8 lignes
26	15 lignes
120	10
40	150 à 200 mots, une dizaine de l
88	15 lignes
131	10 lignes
39 personnes ont répondu à cette question	

TABLE F.40: Réponses à la question n° 39 du second questionnaire

QUESTION	Longueur moyenne d'un texte pour un exercice lacunaire à "objectif linguistique" : en mots
146	200
39	50
32	65
54	150
109	100
57	100
21	100
166	150
164	120
172	120
68	50
47	150
108	260
22	100
45	85
23	200
201	150

IDPROF	RÉPONSES
24	750
206	125
208	200
213	200
215	50
216	100
229	200
230	250
231	125
237	100
239	100
232	150
242	100
66	75
18	80
26	150
120	100
40	175
88	150
131	100
37 personnes ont répondu à cette question	

TABLE F.41: Normalisation des réponses à la question n° 39 du second questionnaire

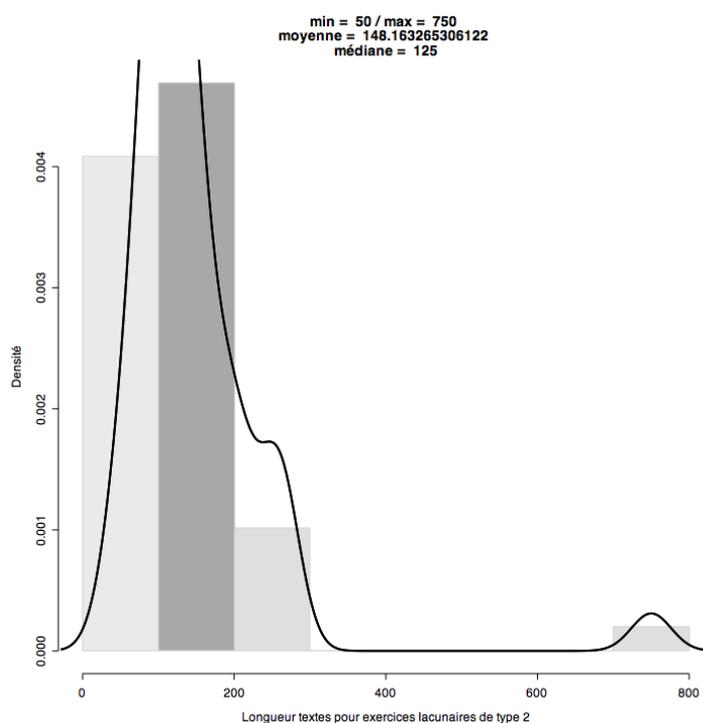


FIGURE F.3: Longueur d'un texte pour un exercice lacunaire à « objectif linguistique »

QUESTION	Nombre moyen de mots escamoté par texte (Ordre de grandeur)
IDPROF	RÉPONSES
146	20
39	5
32	10
54	10
109	6
57	10
21	10
166	10
164	10
172	10
68	5
47	1
108	10
22	20
23	10
201	15
24	20
206	7
208	20
213	20
215	6

IDPROF	RÉPONSES
216	15
229	10
230	5
231	15
237	15
239	10
232	30
241	1
242	10
66	10
18	10
26	10
120	10
40	10
88	5
131	10
37 personnes ont répondu à cette question	

TABLE F.42: Réponses à la question n° 40 du second questionnaire

QUESTION	Dans un texte utilisé dans le cadre d'un travail à "objectif linguistique", la présence de vocabulaire inconnu est :		
CHOIX	% DE RÉPONSES		DÉTAIL
recherchée	<input type="checkbox"/>	18.6%	8 réponses parmi 43
tolérée (ou non pertinente)	<input type="checkbox"/>	74.4%	32 réponses parmi 43
proscrite	<input type="checkbox"/>	7%	3 réponses parmi 43

TABLE F.43: Réponses à la question n° 41 du second questionnaire

QUESTION	Dans un texte utilisé dans le cadre d'un travail à "objectif linguistique", la présence de structures grammaticales incon- nues est :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	 18.6%	8 réponses parmi 43
tolérée (ou non perti- nent)	 58.1%	25 réponses parmi 43
proscrite	 23.3%	10 réponses parmi 43

TABLE F.44: Réponses à la question n° 42 du second ques-
tionnaire

QUESTION	Sélectionner les critères qui peuvent vous influencer dans le choix d'un texte à utiliser comme support d'un exercice la- cunaire à "objectif linguistique" :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
structures	 81%	34 réponses parmi 42
thème	 73.8%	31 réponses parmi 42
vocabulaire	 59.5%	25 réponses parmi 42
objectifs de communica- tion	 45.2%	19 réponses parmi 42
type de texte	 42.9%	18 réponses parmi 42
longueur du texte	 42.9%	18 réponses parmi 42
date	<input type="checkbox"/> 7.1%	3 réponses parmi 42
auteur	<input type="checkbox"/> 7.1%	3 réponses parmi 42
style	<input type="checkbox"/> 2.4%	1 réponses parmi 42
bac		Aucun témoignage
42 personnes ont répondu à cette question et ont choisi en moyenne 3.6 solutions par personne		

TABLE F.45: Réponses à la question n° 43 du second ques-
tionnaire

QUESTION	Pouvez-vous nous donner un exemple pour chacune des caractéristiques que vous avez sélectionné(es) ? ex : thème : musique
IDPROF	RÉPONSES
146	structures : adjectifs détachés type de texte : roman pour enfants longueur : une page maxi pour des avancés

IDPROF	RÉPONSES
39	thème en rapport avec celui de la séquence structures identiques au point de grammaire abordé
32	thème : amour, amitié, école, modes de transport, voyages, famille, vacances, sport. structures : locatif / directif, subordonnées. vocabulaire : emploi des verbes de modalité, verbes forts.
64	no, cela se fait naturellement à partir d'un texte étudié
109	thème : se presenter vocabulaire : la nature communication : un jour dans une ville espagnole
57	THEME : la famille vocabulaire : les goûts structures :estar + gérondif longueur :pas trop long
21	theme : rentree vocabulaire : adjectifs appreciatifs et depreciatifs objectifs : exprimer son opinion
164	Voir réponses précédentes + communication : les tags dans un dialogue
172	Thème : la publicité Vocabulaire de la description Structures de la critique : pienso que + indicatif, no pienso que + subjonctif, etc
68	indiens télévision should
47	thème : le roman de la forêt en Amérique latine, etc...
108	immigration
22	thème : new technologies vocabulaire : champ lexical lié à l'environnement structures : prétérit
45	thème : la famille, le corps, la découverte (de l'Amérique) vocabulaire : vocabulaire thématique et réemploi de mots appartenant à la même famille structures : tener que + infinitif. structures grammaticales vues en classe
23	thème : actualités, vie pratique; vocabulaire :mobilier; longueur du texte :+/-20 lignes; objectifs de com :féliciter
201	thème : pilgrim fathers structures : prétérit type de texte : - longueur : 20 lignes
24	articles de presse : l'élite polonaise fait du baby-sitting à Paris (Le Monde) Livres,CD,films : l'industrie culturelle est dans le rouge (Le Monde)
206	passif

IDPROF	RÉPONSES
208	thème et vocabulaire : la gastronomie structure : l'expression du but, de la cause, de la conséquence type de texte : littéraire, article de presse, ... longueur : une page date : 17e siècle, 20e siècle, ... objectifs de communication : texte polémique qui se prête au débat auteur : Camus
213	le pour et le contre notice explicative, texte informatif
215	thème : voyages structures : grammaticales, eg
216	Retrouver le vocabulaire appris Retrouver les participes passé des verbes au present perfect
230	thème : politique ; famille ; couple ; musique ; alimentation vocabulaire : lié au thème, ce sont les champs lexicaux correspondants, plus ou moins large selon le niveau des apprenants structures : bien sûr, c'est l'essentiel ici, il faut retrouver la structure sur laquelle on travaille dans différents contextes type de texte : poème, chanson, dialogue, description, extrait de roman, de pièce de théâtre, article de journal, etc longueur du texte : ne doit pas être trop long si on veut pouvoir l'utiliser pendant un cours (le cours doit être diversifié), mais on peut toujours faire des coupes nous-mêmes - ceci dit ça dépend bien sûr du niveau et de l'âge des apprenants Objectifs de communication : se présenter, faire une réservation, parler du temps, raconter une anecdote... Auteur : ce n'est pas le plus important, mais les étudiants de niveau avancé sont flattés de se rendre compte qu'ils peuvent lire des textes d'auteurs français ou francophones connus.
231	thème : justice vocabulaire : expressions idiomatiques liées au thème structures : génitifs génériques longueur : pas plus de 15 lignes
237	thème : amour, fêtes, l'étranger, la rencontre,
239	thème : grève structures : exceptions au thème travaillé type de texte : quotidien
232	Le nombre de mots dans le texte dépend du niveau du groupe et du nombre de séances prévues.
241	thème : le travail vocabulaire : la nominalisation structures : le passif type de texte : chanson, article, critique... longueur : une demi-page (en times 12) objectifs de communication : se plaindre style : littéraire, juridique...

IDPROF	RÉPONSES
242	structures : syntaxe du discours rapporté type de texte : narratif longueur du texte : 10 lignes max.
18	objectifs : comment rapporter les paroles de quelqu'un ? vocab : un champ lexical intéressant structures a travailler
26	obj de communication : faire parler les eleves thème : celui qu'on est en train de travailler (les médias :biographie d'un acteur)
40	structure : celle sur laquelle porte l'exercice ! thème : les contes / les émigrés clandestins vocabulaire : champ lexical des animaux / champs lexical du voyage, de l'angoisse, des sentiments
88	structures : les indicateurs temporels + le temps adéquat (ex : tous les jours + imparfiat ; soudain +passé simple) / tener que +infinitif... type de texte : conte, fantastique / article de journal, témoignage... vocabulaire : aliments structures : adverbes de fréquence type de texte : poésie, chanson.. objectifs de communication :au téléphone
131	50 mots
34 personnes ont répondu à cette question	

TABLE F.46: Réponses à la question n° 44 du second questionnaire

c) Exercice lacunaire de compréhension

QUESTION	Longueur moyenne d'un texte pour un exercice lacunaire de compréhension : en mots ou en lignes, merci de préciser l'unité
IDPROF	RÉPONSES
21	60 mots
38	10 lignes
94	200 mots
162	5
166	15 LIGNES
165	8 lignes
164	120mots
68	5 lignes
47	10 lignes
22	300
45	10 à 15 lignes

IDPROF	RÉPONSES
101	6 à 8 lignes
23	25 lignes
201	10 à 20 lignes
24	600 à 1000 mots
210	20 lignes
216	15
225	8 10 lignes
230	de 10 à 20 lignes (plus court po
239	2 minutes
232	150/200
241	une page
240	10 lignes
18	8 lignes
26	15 lignes
120	15
40	150 à 200 mots, une dizaine de l
88	5 à 10 lignes
131	10 - 15 lignes
125	10 lignes
30 personnes ont répondu à cette question	

TABLE F.47: Réponses à la question n° 45 du second questionnaire

QUESTION	Longueur moyenne d'un texte pour un exercice lacunaire de compréhension : en mots
IDPROF	RÉPONSES
21	60
38	100
94	200
162	50
166	150
165	80
164	120
68	50
47	100
22	300
45	125
101	70
23	250
201	150
24	800
210	200

IDPROF	RÉPONSES
216	150
225	90
230	150
232	175
240	100
18	80
26	150
120	150
40	175
88	75
131	125
125	100

28 personnes ont répondu à cette question

TABLE F.48: Normalisation des réponses à la question n° 45 du second questionnaire

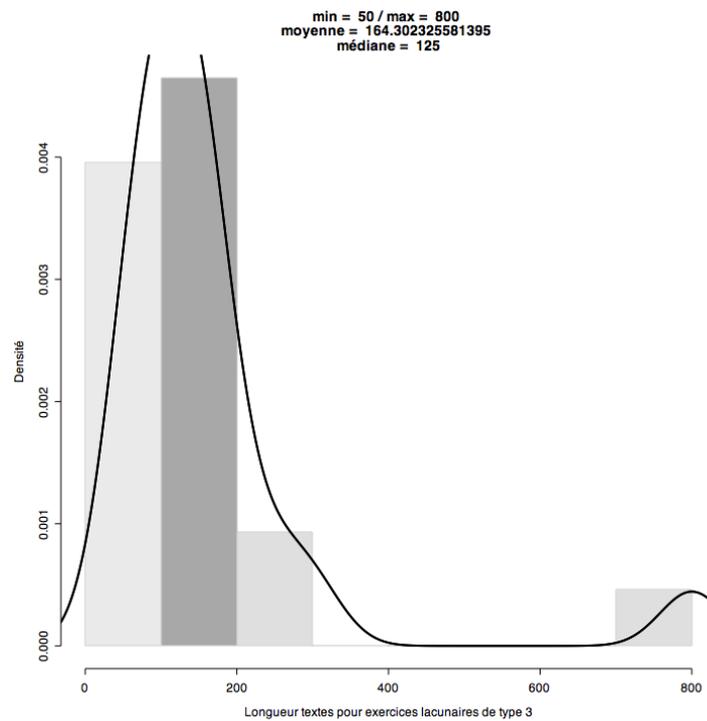


FIGURE F.4: Longueur d'un texte pour un exercice lacunaire de compréhension

QUESTION	Nombre moyen de mots escamoté par texte (Ordre de grandeur)
IDPROF	RÉPONSES
21	10
38	20
94	10
162	8
166	10
165	10
164	8
68	5
22	30
45	10
101	10
23	15
201	10
24	20
210	6
216	15
225	10
239	20
232	50
241	30
240	8
18	12
26	10
120	10
40	10
88	5
131	10
27 personnes ont répondu à cette question	

TABLE F.49: Réponses à la question n° 46 du second questionnaire

QUESTION	Dans un texte utilisé dans le cadre d'un exercice lacunaire de compréhension, la présence de vocabulaire inconnu est :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	26.5%	9 réponses parmi 34
tolérée (ou non pertinente)	 61.8%	21 réponses parmi 34
proscrite	 11.8%	4 réponses parmi 34

TABLE F.50: Réponses à la question n° 47 du second questionnaire

QUESTION	Dans un texte utilisé dans le cadre d'un exercice lacunaire de compréhension, la présence de structures grammaticales inconnues est :		
CHOIX	% DE RÉPONSES		DÉTAIL
recherchée		11.8%	4 réponses parmi 34
tolérée (ou non pertinent)		79.4%	27 réponses parmi 34
proscrite		8.8%	3 réponses parmi 34

TABLE F.51: Réponses à la question n° 48 du second questionnaire

QUESTION	Sélectionner les critères qui peuvent vous influencer dans le choix d'un texte à utiliser comme support d'un exercice lacunaire de compréhension :		
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX		DÉTAIL
thème		91.2%	31 réponses parmi 34
vocabulaire		79.4%	27 réponses parmi 34
structures		44.1%	15 réponses parmi 34
type de texte		41.2%	14 réponses parmi 34
longueur du texte		41.2%	14 réponses parmi 34
objectifs de communication		41.2%	14 réponses parmi 34
auteur	<input type="checkbox"/>	5.9%	2 réponses parmi 34
date	<input type="checkbox"/>	2.9%	1 réponses parmi 34
style	<input type="checkbox"/>	2.9%	1 réponses parmi 34
bac	Aucun témoignage		
34 personnes ont répondu à cette question et ont choisi en moyenne 3.5 solutions par personne			

TABLE F.52: Réponses à la question n° 49 du second questionnaire

QUESTION	Pouvez-vous nous donner un exemple pour chacune des caractéristiques que vous avez sélectionné(es) ? ex : thème : musique
IDPROF	RÉPONSES
21	theme : la majorite en france type de texte : chanson

IDPROF	RÉPONSES
38	cinéma, culture quotidienne
94	C'est très variable
165	non, mais je pense un peu aux attentes de CECR
164	cf réponses précédentes
68	charlie chaplin, ou un autre personnage célèbre voc du cinéma
47	favoriser le réemploi de structures et du vocabulaire au sein d'un même thème, quel qu'il soit
22	thème : musique rock
45	thème : l'altérité, la découverte de l'autre vocabulaire : thématique sur un thème : vocabulaire induit ou déduit du contexte
101	thème : sport vocabulaire : école structures : adjectifs épithètes objectifs de communication : proposer, accepter, refuser
23	thème : actu ; voc : cuisine ; structures : utilisation du quantitatif+de ex : un peu de... ; longueur : +/-25 lignes ; objectifs :se diriger
201	thème : le temps vocabulaire : activités quotidiennes, habituelles structures : subordonnées de temps type de texte : chanson longueur : deux - trois couplets, un refrain objectif de communication : parler de l'avenir, de ce qu'on sera / fera dans plusieurs années
24	Doc.video : "trente ans de pub à la télé
210	chanson
216	Mêmes exemples que ci-dessus
225	tatouages violence uniformes à l'école

IDPROF	RÉPONSES
230	<p>thème : politique ; famille ; couple ; musique ; alimentation</p> <p>vocabulaire : lié au thème, ce sont les champs lexicaux correspondants, plus ou moins large selon le niveau des apprenants</p> <p>structures : l'idéal est de retrouver dans ce texte une ou plusieurs structures vues récemment, cela constitue en même temps une petite révision - les nouvelles structures ne sont pas à proscrire mais il ne peut pas y en avoir en trop grand nombre sinon les apprenants sont découragés, le texte n'est pas à leur niveau</p> <p>type de texte : poème, chanson, dialogue, description, extrait de roman, de pièce de théâtre, article de journal, etc</p> <p>longueur du texte : ne doit pas être trop long si on veut pouvoir l'utiliser pendant un cours (le cours doit être diversifié), mais on peut toujours faire des coupes nous-mêmes - ceci dit ça dépend bien sûr du niveau et de l'âge des apprenants</p> <p>Objectifs de communication : se présenter, faire une réservation, parler du temps, raconter une anecdote...</p> <p>Auteur : ce n'est pas le plus important, mais les étudiants de niveau avancé sont flattés de se rendre compte qu'ils peuvent lire des textes d'auteurs français ou francophones connus.</p>
232	<p>Je voudrais juste ajouter qu'il m'arrive d'adapter les textes aux niveaux débutants et intermédiaires, généralement la structure et le vocabulaire. Evidemment, je signale aux étudiants que j'ai adapté le texte et qu'il ne s'agit pas de l'original.</p>
241	idem ci-dessus
240	<p>thème : ségrégation raciale vocabulaire : les métiers structures : le present perfect type de texte : dialogue longueur du texte : 15 lignes objectifs de communication : expression du conseil</p>
18	thème : ecologie
26	Je les utilise essentiellement en évaluation.
40	idem
88	<p>thème : violence, repas....</p> <p>type de textes : publicité</p>
125	<p>Thème : Vacances</p> <p>Type de texte : cartes postales, dépliants touristiques</p> <p>Objectifs comm. : recherche info, argumentation, description, souhaits</p>
25 personnes ont répondu à cette question	

TABLE F.53: Réponses à la question n° 50 du second questionnaire

F.3.2 Introduction d'une nouvelle notion grammaticale ou de nouveau vocabulaire

QUESTION	Utilisez-vous le même texte pour présenter de nouvelles structures grammaticales et du nouveau vocabulaire ?		
CHOIX	% DE RÉPONSES		DÉTAIL
jamais		13.7%	7 réponses parmi 51
parfois		49%	25 réponses parmi 51
en général		31.4%	16 réponses parmi 51
toujours		5.9%	3 réponses parmi 51

TABLE F.54: Réponses à la question n° 51 du second questionnaire

a) Présentation d'une nouvelle notion grammaticale

QUESTION	Longueur moyenne d'un texte utilisé pour l'introduction de nouvelles structures grammaticales : (en mots ou en lignes, préciser l'unité) Si ce critère ne vous paraît pas pertinent répondez '-'
IDPROF	RÉPONSES
146	tout dépend du niveau des appren
39	"_"
85	30 l
32	25-30
109	6
57	10-15 lignes
21	'_'
38	20 lignes
166	40 LIGNES
165	'_'
164	30 lignes
172	20 lignes
47	je ne regarde pas nécessairement
22	300
31	'_'
101	'_'
196	'_'
23	+/-25 lignes
198	30 lignes
200	100
24	'_'
208	'_'
209	'_'

IDPROF	RÉPONSES
210	peu importe
213	20 lignes
216	20 lignes
230	15 lignes
237	' '
238	10 lignes
232	300 mots
241	une demi-page
240	15 lignes
242	10
66	' '
18	' '
120	8 lignes
40	' '
88	5 lignes
254	' '
131	10 - 20 lignes

40 personnes ont répondu à cette question

TABLE F.55: Réponses à la question n° 52 du second questionnaire

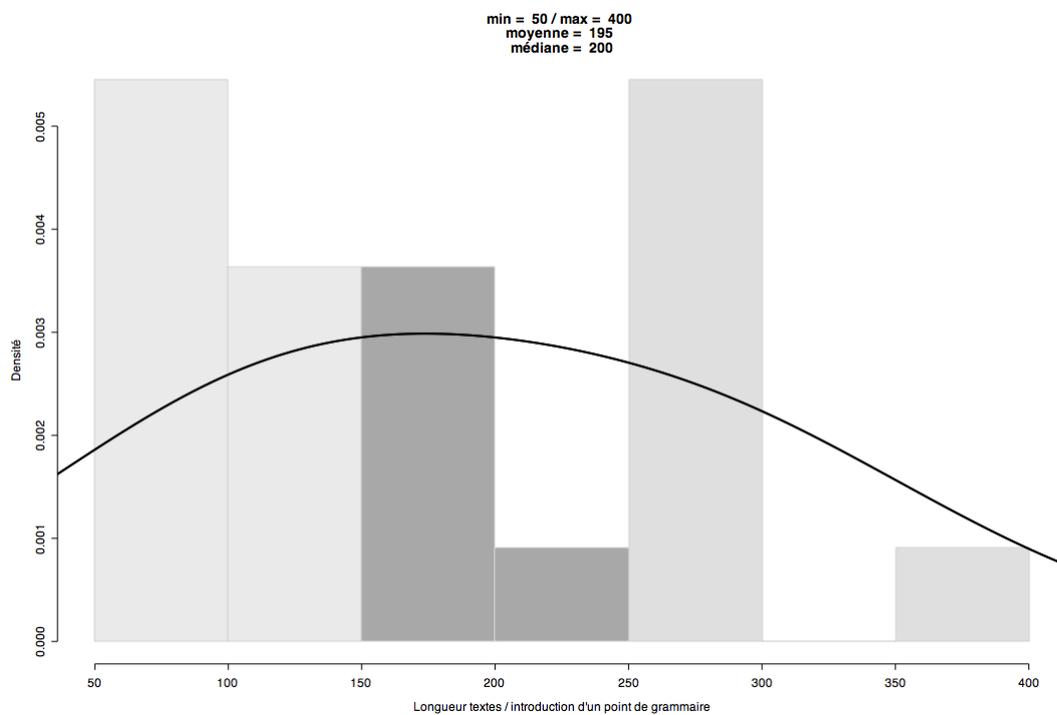


FIGURE F.5: Longueur d'un texte pour l'introduction d'une nouvelle notion grammaticale

QUESTION	Nombre moyen d'éléments représentatifs de la nouvelle notion grammaticale à introduire : (ordre de grandeur) Si ce critère ne vous paraît pas pertinent répondez '-'
IDPROF	RÉPONSES
146	8 à 10 e
39	4
32	5-6
109	10
57	5
21	'-'
38	5
166	'-'
165	'-'
164	5 / 7
172	3
68	3
47	3 max
22	5
31	'-'
101	3
196	entre 5
23	3 à 5
198	'-'
24	'-'
206	un seul
208	'-'
209	5,6
215	3
216	2
231	2-3
230	10
237	'-'
238	3
232	2/3
241	7
240	3
242	5
251	10
66	5
18	6
26	3
40	3
88	5
254	'-'
131	3 - 5
41 personnes ont répondu à cette question	

TABLE F.56: Réponses à la question n° 54 du second questionnaire

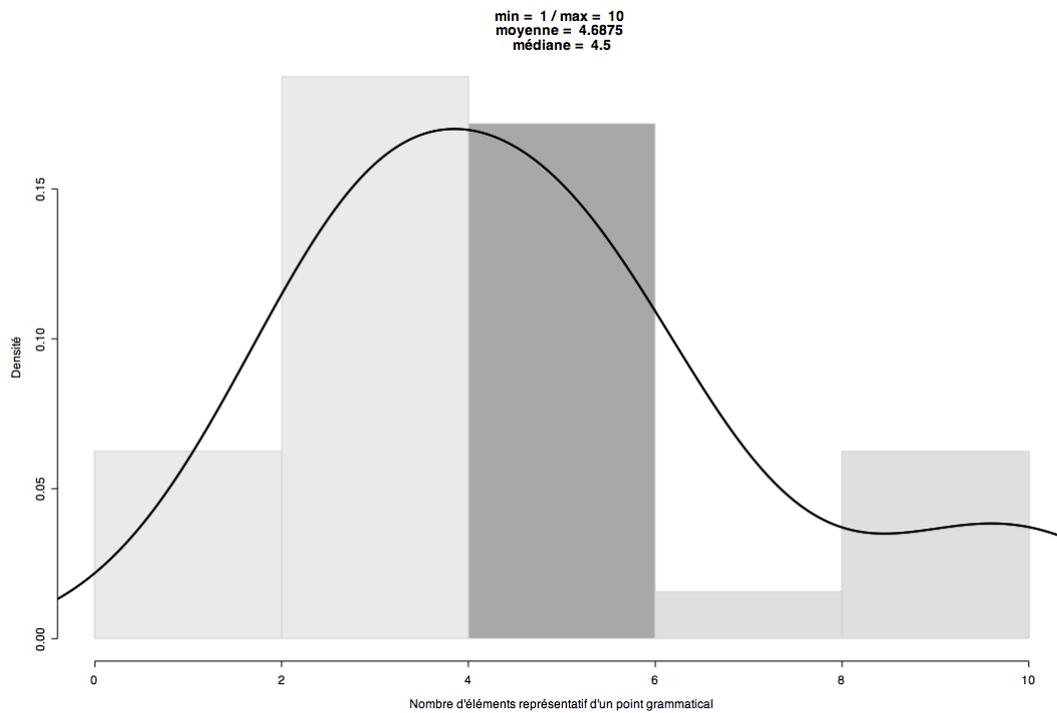


FIGURE F.6: Nombre d'éléments représentatifs de cette notion grammaticale dans le texte

QUESTION	Présence de vocabulaire inconnu dans un texte utilisé pour introduire une nouvelle notion grammaticale :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	<input type="checkbox"/> 21.6%	11 réponses parmi 51
tolérée (ou non pertinent)	<input checked="" type="checkbox"/> 74.5%	38 réponses parmi 51
proscrite	<input type="checkbox"/> 3.9%	2 réponses parmi 51

TABLE F.57: Réponses à la question n° 56 du second questionnaire

QUESTION	Présence de structures grammaticales inconnues dans un texte utilisé pour présenter une nouvelle notion grammaticale (autre que l'objet de la leçon) :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée		11.8% 6 réponses parmi 51
tolérée (ou non pertinente)		72.5% 37 réponses parmi 51
proscrite		15.7% 8 réponses parmi 51

TABLE F.58: Réponses à la question n° 58 du second questionnaire

b) Introduction de nouveau vocabulaire

QUESTION	Longueur moyenne d'un texte utilisé pour l'introduction de nouveau vocabulaire : (en mots ou en lignes, préciser l'unité) Si ce critère ne vous paraît pas pertinent répondez '-'
IDPROF	RÉPONSES
146	niveau apprenants
85	30l
32	25-30
109	5
57	15-20
21	'-'
38	20 lignes
94	Cela dépend du niveau enseigné
166	IDEM
165	'-'
164	30 lignes
172	20 lignes
68	10 lignes
47	je ne regarde pas nécessairement
22	300
31	'-'
101	'-'
196	entre 10 et 20 lignes
23	'-'
198	'-'
200	300 mots
24	'-'
208	'-'
209	'-'
210	1 page

IDPROF	RÉPONSES
213	20 lignes
216	20 lignes
237	60 lignes
238	12 lignes
232	300 mots
241	une demi-page
240	15 lignes
242	10
66	10
18	' '
120	15 lignes
40	' '
88	10 lignes
254	10 lignes
131	10 - 20 lignes
40 personnes ont répondu à cette question	

TABLE F.59: Réponses à la question n° 53 du second questionnaire

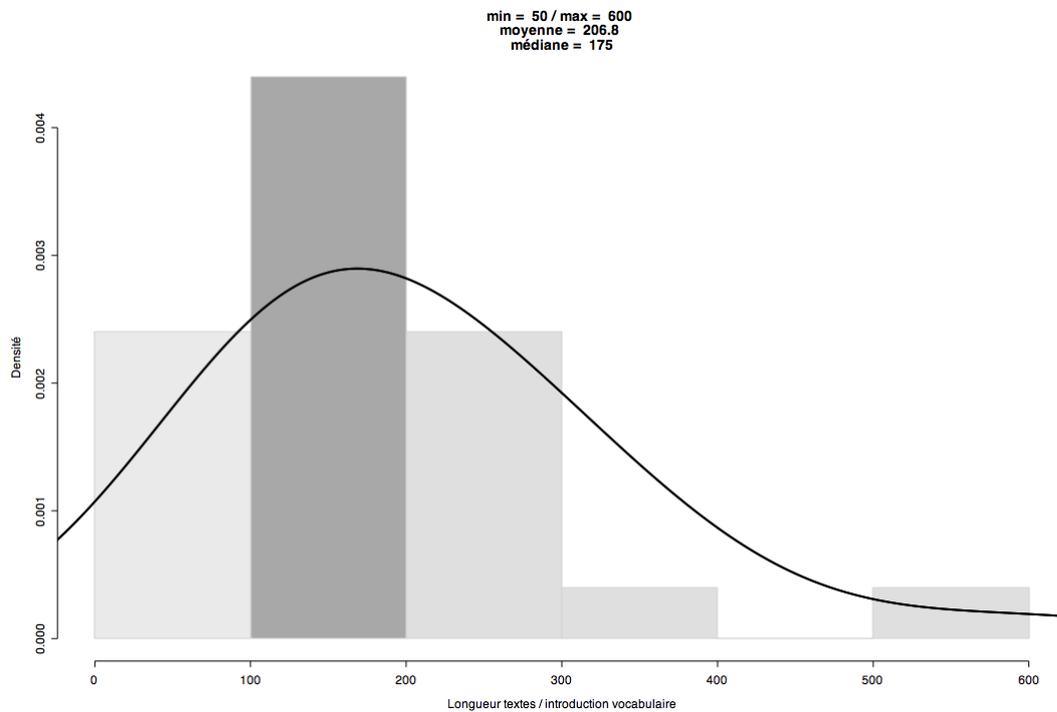


FIGURE F.7: Longueur d'un texte pour l'introduction de nouveau vocabulaire

QUESTION	Nombre moyen d'éléments représentatifs du vocabulaire à introduire (nombre d'occurrences des mots de la liste que vous voulez présenter) : (ordre de grandeur) Si ce critère ne vous paraît pas pertinent répondez '-'
IDPROF	RÉPONSES
146	10-15 (c
39	6
32	6-8
109	7
57	5
21	'-'
38	30
165	'-'
164	10
172	1
47	en fonct
31	'-'
101	'-'
196	entre 5
23	'-'
198	'-'
200	15
24	'-'
206	un seul
208	'-'
209	'-'
210	15
215	10
216	15-20
231	1-2
230	10
237	'-'
238	10
232	2 ou 3
240	5
242	8
66	10
18	10
26	1
40	8
88	5
254	3
131	5
38 personnes ont répondu à cette question	

TABLE F.60: Réponses à la question n° 55 du second questionnaire

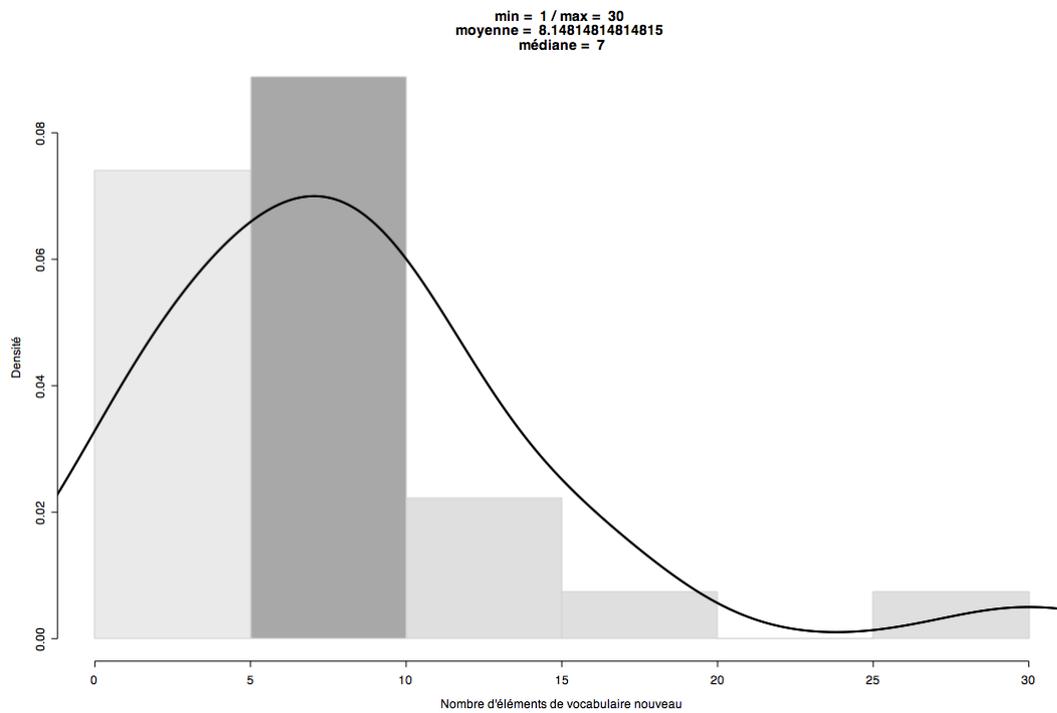


FIGURE F.8: Nombre d'éléments représentatifs de ce nouveau vocabulaire dans le texte

QUESTION	Présence de vocabulaire inconnu dans un texte utilisé pour introduire du nouveau vocabulaire (autre que l'objet de la leçon) :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	<input type="text" value="20%"/>	10 réponses parmi 50
tolérée (ou non pertinent)	<input type="text" value="70%"/>	35 réponses parmi 50
proscrite	<input type="text" value="10%"/>	5 réponses parmi 50

TABLE F.61: Réponses à la question n° 57 du second questionnaire

QUESTION	Présence de structures grammaticales inconnues dans un texte utilisé pour introduire du vocabulaire nouveau :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	<input type="checkbox"/> 2%	1 réponses parmi 50
tolérée (ou non pertinent)	<input checked="" type="checkbox"/> 92%	46 réponses parmi 50
proscrite	<input type="checkbox"/> 6%	3 réponses parmi 50

TABLE F.62: Réponses à la question n° 59 du second questionnaire

c) Critères

QUESTION	Sélectionner les critères qui peuvent vous influencer dans le choix d'un texte à utiliser comme support de la présentation d'une nouvelle notion grammaticale ou de nouveau vocabulaire :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
thème	<input checked="" type="checkbox"/> 86%	43 réponses parmi 50
vocabulaire	<input checked="" type="checkbox"/> 72%	36 réponses parmi 50
structures	<input checked="" type="checkbox"/> 72%	36 réponses parmi 50
type de texte	<input checked="" type="checkbox"/> 56%	28 réponses parmi 50
objectifs de communication	<input checked="" type="checkbox"/> 52%	26 réponses parmi 50
longueur du texte	<input checked="" type="checkbox"/> 34%	17 réponses parmi 50
date	<input type="checkbox"/> 12%	6 réponses parmi 50
auteur	<input type="checkbox"/> 10%	5 réponses parmi 50
bac	<input type="checkbox"/> 4%	2 réponses parmi 50
style	<input type="checkbox"/> 4%	2 réponses parmi 50

50 personnes ont répondu à cette question et ont choisi en moyenne 4 solutions par personne

TABLE F.63: Réponses à la question n° 60 du second questionnaire

QUESTION	Pouvez-vous nous donner un exemple pour chacune des caractéristiques que vous avez sélectionné(es) ? ex : thème : musique
IDPROF	RÉPONSES
146	thème : musique vocabulaire : les types de musique structures : la syntagmatique type de texte : descriptif longueur du texte : dépend du niveau
85	thème : la guerre civile espagnole
32	Thème et vocabulaire sont toujours lié.
109	thème : les vêtements vocabulaire : la classe structures : se presenter
57	theme : famille royale structures : obligation type de texte : description
21	theme : anniversaire vocabulaire : les aliments et les boissons type de texte : dialogue
38	vie quotidienne, cinéma
94	non
165	' '
164	ce sont toujours les mêmes réponses : je n'aime pas la répétition donc la monotonie de votre questionnaire! Je vais aller jusqu'au bout mais vraiment sans enthousiasme.....
172	Thème : la Guerra Civil Vocabulaire : los partidarios de cada bando Structures : Verbes au passé et concordance des temps (imparfaits du subjonctif)
68	science fiction science fiction have+en, have+be+ing dialogue
47	les relations mexique- etats-unis
22	structures : be+ing type de texte : journalistique ou littéraire.
187	theme :it's cheaper to keep her ;longueur :30 à200lignes ;date :pas imp. obj.com . :fonctions
31	thème : cuisine feux d'artifice théâtre structures : concordance des temps type de textes : notices dépliants publicitaires articles de presse essais biographies extraits de roman
101	thème : biographie type de texte : biographie objectifs de communication : faire le portrait de quelqu'un

IDPROF	RÉPONSES
196	thème : le sport Vocabulaire : champ lexical important structures : variées
23	thème : actu ; voc : moyens de transport ; structures : emploi des prépositions à/en+moyen de transport ; objectifs de com : récit de voyage
24	Bibliographie d'un personnage contemporain (Jane Fonda dans Libération)
206	champ lexical de la maison style indirect suggestions
208	thème et vocabulaire : gastronomie structure : l'expression du but, de la cause, de la conséquence type de texte : poème, théâtre, article de presse, ... longueur : une page date : texte récent ou non objectif de communication : prendre position, argumenter, ... auteur : un autre auteur, d'une autre époque ou d'un autre genre littéraire
213	point de vue sur un problème
215	thème : voyages vocabulaire : vêtements structures : le conseil
216	thème : au restaurant vocabulaire : le restaurant structures : le present perfect type de texte : dialogue
231	thème : les Indiens vocabulaire : 5 mots inconnus structures : au moins une nouvelle structure objectifs de communication : comparer à une autre support sur le même thème
230	bon c'est toujours la même chose quoi...
236	thème : cuisine, présentation d'une ville, artiste, profession, vie quotidienne
238	thème : actualité, civilisation, culture Date : éviter les F, donc avant 2003
232	Thème : La santé, rechercher un logement, se vêtir, trouver un emploi ... Longueur : important selon le niveau du groupe (débutant, intermédiaire, avancé) et le temps prévu pour le texte (un cours de 1 heure, plusieurs heures etc) Objectifs de communication : faire une requête, s'informer, informer sur son état de santé etc
241	idem réponses précédentes
240	thème : santé vocabulaire : santé structure : must+base verbale type de texte : article de presse longueur : de 15 à 20 lignes objectifs de communication : expression de l'obligation

IDPROF	RÉPONSES
242	thème : la ville vocabulaire : immeuble, cité, banlieue, centre-ville, citadin, mairie, maire, métro, ... type ; descriptif, narratif longueur ; 10 lignes objectifs : s'orienter, situer
247	impossible de répondre précisément à ces questions et aux précédentes. Je ne fais pas travailler à mes étudiants la grammaire en tant que telle (j'ai surtout des niveaux master). Je travaille à partir de thèmes : les documents sont choisis en fonction des thèmes, le lexique peut être nouveau pour certains, pour la plupart il s'agit de révision finalement, le but étant une pratique de la langue la plus fluide possible, mais il faut bien avouer que le gros du travail a déjà été fait en amont, je n'introduis pas vraiment de point grammatical ou de lexique totalement nouveau.
66	thème : histoire
18	structures : be going to + bv
26	Style : texte avec une structure répétitive, ce qui facilite la compréhension
40	.
88	thème : très diversifié
254	thème : l'automne Type de texte : comptine, poésie, chanson, ronde longueur du texte : 10 lignes maximum
40 personnes ont répondu à cette question	

TABLE F.64: Réponses à la question n° 61 du second questionnaire

F.3.3 Activité de compréhension

QUESTION	Quand vous utilisez un texte comme support d'une activité de compréhension (par exemple : lecture / questions sur le contenu du texte), vous utilisez aussi ce texte comme support d'autres activités ?	
CHOIX	% DE RÉPONSES	DÉTAIL
jamais	<input type="checkbox"/> 6%	3 réponses parmi 50
parfois	<input type="checkbox"/> 44%	22 réponses parmi 50
en général	<input type="checkbox"/> 38%	19 réponses parmi 50
toujours	<input type="checkbox"/> 12%	6 réponses parmi 50

TABLE F.65: Réponses à la question n° 62 du second questionnaire

La question suivante n'est posée que si la réponse n'est pas « jamais ».

QUESTION	Pouvez-vous nous donner un (ou plusieurs) exemple(s) des autres activités que vous pouvez effectuer avec un texte utilisé comme support d'une activité de compréhension.
IDPROF	RÉPONSES
146	travail sur les champs lexicaux (vocabulaire), travail sur les structures grammaticales spécifiques au texte...
39	expression orale autour du thème du texte
85	traduction
65	non, car ça varie beaucoup + j'ai le sentiment que les questions de cette enquête se réfèrent à une pratique beaucoup trop différente de la mienne pour être pertinentes
32	point de départ pour l'expression orale ou écrite.
109	recherche des champs lexicaux déjà vus
57	expresion personnelle
21	element d'inspiration pour un jeu de role
38	reconstitution de sens d'un texte
94	production orale ou écrite
166	OBJECTIF CULTUREL
164	points grammaticaux
172	Transposer certains verbes à une autre personne (pertinent en cas de verbes à diphtongue)
68	imaginer un pair work changer le point de vue de l'écriture
47	Compréhension de l'écrit et de l'oral
22	linguistiques et phonologiques
187	questions
31	enrichissement (culturel/historique) expression approfondie sur le sujet
101	structures linguistiques
196	pour la grammaire, pour les rimes, pour les champs lexicaux, comme base à un projet...
23	réutilisation des structures grammaticales dans un contexte différent, prononciation, expression écrite sur le même thème
201	phonologie grammaire (ex : écrire des questions, ...) écriture (plus ou moins guidée : écrire une suite, écrire sur le même modèle, ...)
206	structures, vocabulaire, phonologie
208	conversation, débat, écriture
209	pairwork, writing
210	reprises grammaticales ou de structures
213	compléter les connaissances techniques
215	support à discussion, controverse

IDPROF	RÉPONSES
230	expression écrite (ex : écrire une chanson sur un thème proche) ; débat oral ; introduction à un thème ou à un champ lexical ; révision de structures grammaticales
231	grammaire expression
236	support de lecture à haute voix (correction phonétique) déclancheur de production orale (discussion, jeux de rôles) déclancheur de production écrite (courrier des lecteurs, critiques, opinions, réponses...)
237	production écrite ou orale
238	Champ lexical Syntaxe stylistique
232	recherche sur le vocabulaire, utilisation du dictionnaire, travail en équipe sur les homonymes, synonymes... Utilisation du vocabulaire dans différents contextes :registres de langue, important en FLE. Oral : Discussion dirigée sur le thème du texte. Comparaison avec le pays d'origine, exposé en classe. Ecrit : Prolongation du texte.
241	compréhension d'une chanson, qui est ensuite à mettre au discours indirect, et qui amène à un débat sur le problème soulevé par la question
240	inférence de mots en contexte analyse de fait de langue dans le cadre d'un extrait de roman : imaginer ce qui s'est passé avant ou ce qui va se produire après
242	réécriture du texte (en changeant le temps, le point de vue ; continuation (imaginer une suite) ; passer du style indirect au style direct...)
247	activités d'expression diverses : à l'écrit, type résumé-discussion, à l'oral comme base d'argumentation dans un débat...
66	grammaire, activité orale
252	texte ,images, corpus ...
18	comprehension orale, vocabulaire nouveau, faits de langues, civilisation...
26	lecture à voix haute, prononciation
40	créer en suivant la structure du texte, poème, un aute texte/ poème. relater au discours direct les propos des personnes écrits au style indirect
88	résumé du texte
254	illustrer ce texte ; le réciter
131	support imitation, réorganiser paragraphes,
46 personnes ont répondu à cette question	

TABLE F.66: Réponses à la question n° 63 du second questionnaire

a) Catégorisation des réponses à la question 63

La question 63 (cf. tableau F.66 page a-102) est un champ libre laissé afin que les enseignants puissent préciser dans quelles autres activités ils réutilisent les textes dédiés à des activités de compréhension. Comme pour l'organisation de leurs recueils, l'aspect qualitatif sera utilisé dans la création du modèle, mais il peut être intéressant de consulter ces réponses d'un point de vue quantitatif pour croiser ces réponses avec d'autres ou établir des tendances. Nous avons donc eu recours à des familles d'activités.

Nous avons, dans un premier temps, tenté de recourir au Cadre Européen Commun de Référence pour les Langues (CECR) pour classer les réponses des enseignants, mais séparer les témoignages en *activité de production* (orale ou écrite), *activité de réception* (orale ou écrite), *activité d'interaction* (orale ou écrite) et *activité de médiation* (orale ou écrite)¹ ne traduisait pas réellement le contenu des réponses. Dans certains cas, les réponses ne sont pas assez précises pour pouvoir être classées selon cette typologie : « expression orale autour du thème d'un texte » [39]. S'agit-il ici de d'une activité de production orale ou d'une activité d'interaction orale dans laquelle les apprenants doivent échanger des points de vue ? D'autre part, alors que certains enseignants se focalisent sur le but de l'activité (« révision de structures grammaticales » [230]), la plupart se concentre sur l'activité elle-même (« relater au discours direct les propos des personnes écrits au style indirect » [40]).

Modes de description : Une classification selon le CECR, ne serait donc pas suffisante. Nous avons décidé de faire découler notre classification directement des réponses données, en articulant nos catégories selon deux sous-ensembles. Un sous-ensemble regroupant les différents témoignages adressant la question en termes d'objectif pédagogique et un autre pour ceux qui l'adressent en termes de déroulement de l'activité.

Certains témoignages font appel aux deux modes de description, pour la même activité (déroulement : « support de lecture à haute voix », objectif : « Correction phonétique », [236]) ou même en changeant de mode d'une activité à l'autre (activité 1, déroulement : « Compréhension orale » ; activité 2, objectif : « civilisation », [18]). Ce type de distinction ne sera pas traduit par notre classification, qui permet uniquement de dire que tel ou tel enseignant utilise les deux modes de description dans sa réponse, sans spécifier si c'est dans le cadre de la même activité ou non.

Le fait que les enseignants emploient un mode de description plutôt qu'un autre ne signifie pas qu'ils occultent les aspects liés à cet autre mode. Nous examinons ici, la description que les enseignants font des activités qu'ils utilisent en classe, sans que cela ne présage de la manière dont ils conçoivent ces activités.

Types d'activités : Quand le déroulement de l'activité est indiqué, on peut en général le lier à un des types d'activités du CECR, sans toutefois toujours avoir la possibilité de spécifier s'il s'agit d'oral ou d'écrit ou encore si l'activité est une activité de production simple ou d'interaction. Cette typologie ne fait pas non plus la distinction entre les activités de production écrite, que l'on pourra désigner comme *créative* et les exercices structuraux et autres exercices lacunaires. Dans les premières, la responsabilité de l'apprenant se situe aussi bien au niveau du *quoi dire* que du *comment le dire*, pour reprendre la terminologie de la génération automatique de texte.

1. Cf. [Conseil de l'Europe, 2000, chapitre 4, pp. 39–80].

Alors que dans les secondes, il n'interviendra qu'au niveau du *comment le dire*. Nous avons donc eu recours au types d'activité suivants :

- Production orale (nous regroupons sous l'étiquette production orale les activités de production orale et d'interaction)
- Production écrite créative
- Production écrite non-créative
- Production « indéfinie » (orale ou écrite, créative ou non)
- Médiation

Objectifs d'activité : Malgré la tendance à décrire les objectifs d'une activité en terme de compétence de communication à développer, nous avons décidé de suivre strictement les réponses des enseignants. C'était dans ce cas d'autant plus approprié que les réponses des enseignants ne permettent, en général, pas d'établir clairement les compétences à développer. Les différents objectifs des activités seront donc les suivants :

- Lexique
- Grammaire
- Phonologie
- Stylistique
- Culture, civilisation

Liste des témoignages : Les témoignages sont regroupés dans le tableau ci-dessous.

Dérroulement de l'activité	Production orale	[39], [32], [21], [94], [68], [208], [209], [215], [230], [236], [237], [232], [241], [66], [26], [254]
	Production écrite créative	[57], [23], [201], [230], [236], [232], [241], [240], [242], [247], [40], [88], [131]
	Production écrite non-créative	[172], [201], [241], [242], [247], [40]
	Production « indéfinie »	« Point de départ pour l'expression orale ou écrite » [32], « Production orale ou écrite » [94], [68] ² , [31], [196], « réutilisation des structures grammaticales dans un contexte différent » [23] (décrit donc l'activité en termes d'objectif et de déroulement, mais ne permet pas de dire s'il s'agit d'oral ou d'écrit), [208], [209], [237], [254]
	Médiation	[85]
Objectif de l'activité	Lexique	[146], [32], [196], [206], [230], [238], [232], [240], [18]
	Grammaire	[146], [164], [172], [22], [101], [196], [23], [201], [206], [210], [230], [231], [238], [240], [66], [18]
	Phonologie	[22], [23], [201], [206], [236], [26]
	Stylistique	« pour les rimes » [196], [238], [232]
	Culture, civilisation	[166], [31], [213], [232], [18]

2. « changer le point de vue de l'écriture », s'agit-il d'écriture créative (par exemple, donner le rôle de narrateur à un protagoniste du texte) ou non (par exemple, passer le texte de la première à la troisième personne) ?

Hors-sujet ou ambigu	[65] (Hors-sujet), « reconstitution du sens d'un texte » [38], « compréhension de l'oral et de l'écrit » [47] (ne constitue pas une nouvelle utilisation du texte), « questions » [187], « texte, image, corpus » [252]
----------------------	---

TABLE F.67: Classement des réponses à la question 63

b) Fin de la série de question pour les activités de compréhension

QUESTION	Longueur moyenne d'un texte pour une activité de compréhension : (en mots ou en lignes, préciser l'unité) Si ce critère ne vous paraît pas pertinent répondez '-'
IDPROF	RÉPONSES
146	niveau apprenants
85	30l
32	25-30
57	10-15 lignes
38	20 lignes
94	selon niveau enseigné
166	40 LIGNES
165	15lignes
164	30 lignes
172	30 lignes
47	15 lignes
22	300
187	40 90
31	'-'
101	6 à 7 lignes
196	entre 10 et 30 lignes
23	'-'
200	300 mots
201	dépend du niveau de la classe
24	600 à1000 mots
206	10-15 lignes
208	une page
209	'-'
210	1 page
215	30 lignes
216	15-20 lignes
230	15 - 20 lignes, ça dépend toujours du niveau des apprenants
231	500 mots
237	60
238	50
232	130/ 150 mots
241	une page
240	15 à 20 lignes
242	10
247	1 ou 2 pages
66	'-'
252	entre 5 et 8 lignes
18	10 lignes
40	'-'
88	10 à 30 lignes
254	10 lignes

IDPROF	RÉPONSES
131	15 - 150
42 personnes ont répondu à cette question	

TABLE F.68: Réponses à la question n° 64 du second questionnaire

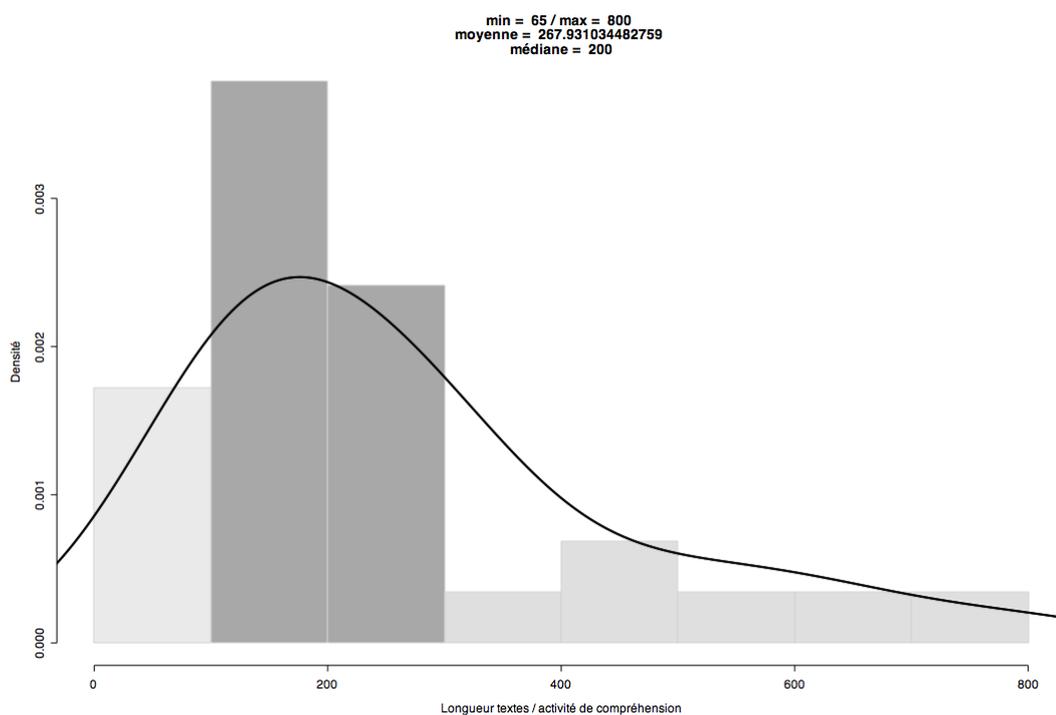


FIGURE F.9: Longueur d'un texte pour une activité de compréhension

QUESTION	Dans un texte utilisé dans le cadre d'une activité de compréhension, la présence de vocabulaire inconnu est :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	<input type="checkbox"/>	54% 27 réponses parmi 50
tolérée (ou non pertinente)	<input type="checkbox"/>	44% 22 réponses parmi 50
proscrite	<input type="checkbox"/>	2% 1 réponses parmi 50

TABLE F.69: Réponses à la question n° 65 du second questionnaire

QUESTION	Dans un texte utilisé dans le cadre d'une activité de compréhension, la présence de structures grammaticales inconnues est :	
CHOIX	% DE RÉPONSES	DÉTAIL
recherchée	 32%	16 réponses parmi 50
tolérée (ou non pertinente)	 64%	32 réponses parmi 50
proscrite	<input type="checkbox"/> 4%	2 réponses parmi 50

TABLE F.70: Réponses à la question n° 66 du second questionnaire

QUESTION	Sélectionner les critères qui peuvent vous influencer dans le choix d'un texte à utiliser comme support d'une activité de compréhension :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
thème	 91.8%	45 réponses parmi 49
vocabulaire	 73.5%	36 réponses parmi 49
structures	 65.3%	32 réponses parmi 49
type de texte	 59.2%	29 réponses parmi 49
objectifs de communication	 55.1%	27 réponses parmi 49
longueur du texte	 40.8%	20 réponses parmi 49
date	 16.3%	8 réponses parmi 49
auteur	<input type="checkbox"/> 8.2%	4 réponses parmi 49
bac	<input type="checkbox"/> 4.1%	2 réponses parmi 49
style	<input type="checkbox"/> 4.1%	2 réponses parmi 49
49 personnes ont répondu à cette question et ont choisi en moyenne 4.2 solutions par personne		

TABLE F.71: Réponses à la question n° 67 du second questionnaire

QUESTION	Pouvez-vous nous donner un exemple pour chacune des caractéristiques que vous avez sélectionné(es) ? ex : thème : musique
IDPROF	RÉPONSES
85	theme : la conquete de l'Amérique
65	/

IDPROF	RÉPONSES
32	thème : école, vacances, famille, voyages, modes de transports, amitié, amour... structures : utilisation de certains temps ou modes (parfait, prétérit, subjonctif). objectifs : expression de l'intention, du regret, de l'opposition.
109	parler au téléphone
57	theme !les loisirs vocabulaire : les goûts structures :encantar preferir gustar type de texte : annonce longueur du texte :selon la difficulté
21	theme : introduction d'une nouvelle personne vocabulaire : description physique objectifs de communication : savoir introduire/reconnaitre une personne
38	vie quotidienne, cinéma, musique
94	non
165	'_'
164	idem
172	Thème : la vie de couple Vocabulaire : activités ménagères Structures : alternance je-tu
68	musique instruments musique, types de musique had+en biographie interview
47	l'Espagne du siècle d'or
22	identité
31	thème : réforme de la justice en Italie vocabulaire : lexique juridique type de texte : article de presse
101	mêmes exemples que dans les questions précédentes
196	thème : sport date : récent
23	thème : actu ; voc :banlieues ; longueur : - de 50 lignes
200	psychologie art géographie
24	Articles de presse (même type que précédemment) Doc.audio/video (enregistrées à la radio ou à la télé)
206	Dans le cadre de l'étude sur les native Americans, texte de compréhension sur l'histoire des Lakotas.

IDPROF	RÉPONSES
208	thème et vocabulaire : le roman policier structures : emploi des temps du passé type de texte : nouvelle à chute longueur : une page, 3 maximum date : 19e siècle ou 21e siècle ? objectifs : raconter, reformuler, (ré)écrire la fin de l'histoire auteur : Simenon, ...
213	point de vue sur le sujet
215	thèmes, voc, structures : idem type de texte : critique de film
216	Mêmes exemples que dans la section précédente
230	En général c'est avant tout le thème qui importe, mais aussi si c'est de la compréhension orale, la compréhensibilité du support (si c'est une chanson par exemple ou une vidéo, il faut voir si le débit de parole n'est pas trop rapide etc); pour la compréhension écrite, il faut limiter les expressions typiques que les apprenants ne connaissent pas. Rien n'est proscrit vraiment, mais il faut que le texte soit adapté au niveau des apprenants tout simplement, et aux objectifs du cours. Tous ces éléments sont interdépendants... Si le texte est assez simple il peut être plus long, s'il est complexe il doit être plus court; s'il y a beaucoup de vocabulaire, il faut éviter les nouvelles structures grammaticales, tout simplement parce que ça fait trop en une fois...
231	thème : droits de l'homme vocabulaire : 5 mots inconnus difficiles à inférer + 10 à 15 mots faciles à inférer (procédés de lexicologie simples) structures : idiomatisme de l'écrit longueur du texte : liée à la longueur de la séquence voulue
238	structures : des paragraphes courts type de texte : extraits de journaux et revues diverses auteur : exclusivement français
232	Date : pour un texte de compréhension, je préfère les textes d'actualité ce qui engendre un meilleur débat en classe. Également important pour connaître le nouveau pays d'accueil dans le cadre du français pour immigrants.
241	idem
240	thème : les nouvelles technologies vocabulaire : informatique structure : will+base verbale type de texte : page web objectif de communication : l'expression du futur longueur du texte 15 à 20 lignes
242	ex : thème : la ville vocabulaire : immeuble, centre-ville, banlieue, cité, citadin, métro, mairie, maire square, rue... type : descriptif ou narratif objectif de communication : situer, s'orienter.. longueur : 10 lignes (niveau élémentaire)

IDPROF	RÉPONSES
247	thème : négociations commerciales internationales, marchés énergétiques, chômage... vocabulaire : lexique spécifique à un thème, un style... type de texte : article de presse, article de recherche publié, working paper, compte-rendu de table-ronde... date : plus récent bien sûr, ou alors ancien avec perspective de lecture diachronique (que disait-on dans la presse lors de la mise en place de l'OMC en 1994-5 par ex) style : travail sur les registres et les styles en fonction du genre/type de texte
66	thème : sujet de société
18	theme : sports vocabualaire : activités, loisirs structures : can, be able to, manage to...
26	Thème, vocabulaire et structures qu'on est en train de travailler ou qu'on a déjà travaillés (exemples : présentation personnelle//d'un personnage de roman)
40	.
88	thème : divers
254	idem page précédente
39 personnes ont répondu à cette question	

TABLE F.72: Réponses à la question n° 68 du second questionnaire

F.3.4 Activité de traduction

QUESTION	Dans les exercices de traduction que vous proposez à vos apprenants, vous proposez de travailler sur :	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
phrases	<input type="text" value="69.2%"/>	27 réponses parmi 39
texte	<input type="text" value="61.5%"/>	24 réponses parmi 39
39 personnes ont répondu à cette question et ont choisi en moyenne 1.3 solutions par personne		

TABLE F.73: Réponses à la question n° 69 du second questionnaire

QUESTION	Classer l'importance des facteurs suivants dans votre choix d'un texte pour un exercice de traduction Le facteur le plus important en premier	
CHOIX	SCORE*	DÉTAIL
*Sur 6 propositions celle classée la première obtient 6 points la seconde 5 et ainsi de suite		
auteur du texte	294	choisi 25 fois comme choix numéro 1 choisi 1 fois comme choix numéro 3 choisi 4 fois comme choix numéro 4 choisi 6 fois comme choix numéro 5 choisi 13 fois comme choix numéro 6
date (époque) d'écriture ou de parution du texte	200	choisi 11 fois comme choix numéro 2 choisi 4 fois comme choix numéro 3 choisi 4 fois comme choix numéro 4 choisi 14 fois comme choix numéro 5 choisi 7 fois comme choix numéro 6
source (ex : journal, manuel, roman, etc...)	160	choisi 2 fois comme choix numéro 2 choisi 16 fois comme choix numéro 3 choisi 14 fois comme choix numéro 4 choisi 6 fois comme choix numéro 5 choisi 2 fois comme choix numéro 6
thème	81	choisi 11 fois comme choix numéro 2 choisi 1 fois comme choix numéro 3 choisi 13 fois comme choix numéro 4 choisi 2 fois comme choix numéro 5
type de texte	76	choisi 9 fois comme choix numéro 2 choisi 13 fois comme choix numéro 3 choisi 4 fois comme choix numéro 4 choisi 11 fois comme choix numéro 5 choisi 1 fois comme choix numéro 6
présence de certaines tournures (ex : "se + 3e personne" en espagnol traduit avec "on" en français)	29	choisi 8 fois comme choix numéro 2 choisi 6 fois comme choix numéro 3 choisi 1 fois comme choix numéro 5 choisi 14 fois comme choix numéro 6

TABLE F.74: Réponses à la question n° 70 du second questionnaire

QUESTION	Si la présence de certaines tournures vous a paru un critère de recherche pertinent, pouvez-vous nous donner quelques exemples de tournures que vous pouvez rechercher dans un texte que vous aller faire traduire à vos apprenants (thème ou version)
IDPROF	RÉPONSES
39	expression de l'obligation personnelle et impersonnelle, expression de on, expression de devenir...
85	traductions de "on", de "devenir"...
65	Je ne fais pas d'exercices de traduction !! C'est complètement dépassé à mon avis (sauf si on forme des...traducteurs)
32	comparatif, verbes de modalité, discours direct / indirect, interrogatives partielles / globales.
57	gustar
21	pronoms possessifs
38	participe passé, pronom
94	verbes de modalité
	pertinence de la différenciation entre un GN à l'accusatif ou au datif
	fonctions langagières (surtout concernant l'aspect perlocutoire des énoncés)
165	paul makes his bed every morning. He does his homework after tea I am 13 I am thirsty It is cold He makes me go to music lessons traduction des 'liking verbs' Subordonnées : if, when, to/on order to vs for etc.
164	passives à traduire par "on" tournures impersonnelles en français à construction très différente en anglais
172	soy yo quien, allí es donde, el chico del que te hablaba, etc
68	must et les modaux present perfect
47	la traduction du devenir, la valeur des passés simples, le déroulement de l'action..
22	vouloir que quelqu'un fasse quelquechose
31	concordance des temps, passé simple, subjonctif, traduction de "on", de "il faut", "il y a"
101	je ne fais pas d'exercices de traduction
23	"je ne peux pas faire de traduction avec autant d'origines différentes!" désolée de mettre ce commentaire ici mais c'est le seul espace d'expression que j'ai trouvé sur cette page!
201	used to phrases au présent simple avec adverbe de fréquence (anglais) ...

IDPROF	RÉPONSES
24	Jamais de traductions
206	Traduction du passif par le pronom 'on' ou traduction du pronom 'on' par le passif en anglais
208	je ne donne pas de thèmes ni de versions
209	langue 'orale', populaire
210	thème date auteur source type tournures
213	tournures idiomatiques
215	passif
216	savoir utiliser soit le prétérit soit le présent perfect pour un passé composé français.
231	thème : participes présents, 'on', futur dans le passé, adjectif en tête de phrase, 'depuis', propositions relatives version : V+ING (hors présent continu), phrasal verbs, prépositions
236	aucune traduction
232	Je ne pratique pas la traduction car j'enseigne à des Chinois, des Bengalais, des Pakistanais ...
241	je ne fais pas de traduction
240	traduction de la structure passive anglaise traduction des "phrasal verbs" ou verbes à particule
242	tournure passive
66	phrase conditionnelle, négation
252	je prends la route. difficile à traduire pour exprimer je voyage je vais à
18	je ne fais pas d'ex de traduction
26	je ne fais pas de traduction avec les non-spécialistes ; avec les spés : verbes de mouvement, structures résultatives, modaux
40	l'exemple déjà donné ; le passé simple en espagnol traduit par le passé composé en français ; les verbes d'ordre en espagnols suivis de "que + verbe au subjonctif" traduits par verbe d'ordre en français suivis de "de+infinitif" ; le verbe "demander" en français qui se traduit par "preguntar ou pedir" suivant le sens ...
88	les mots de liaison qui amène une fluidité de lecture, les dialogues, les expressions idiomatiques ou autres,...
254	pas de traduction
131	Nous faisons très peu de traduction.
40 personnes ont répondu à cette question	

TABLE F.75: Réponses à la question n° 71 du second questionnaire

F.4 Influence de la langue maternelle des apprenants et de leur niveau

QUESTION	La langue maternelle de vos apprenants a-t-elle une influence sur le choix des textes que vous leur présentez ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non	 67.3%	33 réponses parmi 49
oui	 32.7%	16 réponses parmi 49

TABLE F.76: Réponses à la question n° 77 du second questionnaire

NB : La question suivante ne s'affichait que lorsque l'utilisateur répondait par l'affirmative à la question précédente

QUESTION	Quelle influence la langue maternelle de vos apprenants a-t-elle sur la quantité tolérée de vocabulaire nouveau ou de structures grammaticales inconnues (n'étant pas l'objet du cours) dans un texte ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non pertinent	<input type="checkbox"/> 6.3%	1 réponses parmi 16
ne se prononce pas	<input type="checkbox"/> 12.5%	2 réponses parmi 16
plus la langue maternelle des apprenants est proche de la langue enseignée plus on acceptera de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte.	<input type="checkbox"/> 81.3%	13 réponses parmi 16
plus la langue maternelle des apprenants est proche de la langue enseignée moins on acceptera de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte.		Aucun témoignage

TABLE F.77: Réponses à la question n° 72 du second questionnaire

QUESTION	Quelle est l'influence du niveau de vos apprenants sur la quantité tolérée de vocabulaire nouveau ou de structure grammaticales inconnues (n'étant pas l'objet du cours) dans un texte ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non pertinent	<input type="checkbox"/> 8.2%	4 réponses parmi 49
ne se prononce pas	<input type="checkbox"/> 18.4%	9 réponses parmi 49
meilleur est le niveau des apprenants, plus on acceptera de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte.	<input type="checkbox"/> 71.4%	35 réponses parmi 49
meilleur est le niveau des apprenants, moins on acceptera de structures grammaticales inconnues ou de vocabulaire nouveau (ne faisant pas partie de l'objet du cours) dans un texte	<input type="checkbox"/> 2%	1 réponses parmi 49

TABLE F.78: Réponses à la question n° 73 du second questionnaire

F.5 Types de textes

QUESTION	Effectuez-vous avec vos apprenants un travail portant sur les différents types de texte ?	
CHOIX	% DE RÉPONSES	DÉTAIL
non	<input type="checkbox"/> 14%	7 réponses parmi 50
oui	<input type="checkbox"/> 86%	43 réponses parmi 50

TABLE F.79: Réponses à la question n° 74 du second questionnaire

NB : La question suivante ne s'affichait que lorsque l'utilisateur répondait par l'affirmative à la question précédente

QUESTION	Quels types de textes abordez-vous avec vos apprenants ?
IDPROF	RÉPONSES
146	journalistiques, romans, dialogues, descriptif, narratifs, dans tous les cas oral et écrits
39	articles de presse, poésies, chansons, romans, nouvelles...
85	poésie, théâtre, texte de presse, roman
65	En fait je ne comprends pas la question posée ici. Mais il fallait bien choisir oui ou non. Désolée.
32	Autobiographie, romans, théâtre, poésie, articles de journaux.
109	documentation authentique, textes pas très compliqués (car ce sont des élèves entre 8 et 12 ans....)comptes, ...
38	textes journalistiques (libération et vingt minutes essentiellement
94	texte "littéraire" (extrait de roman ou de nouvelle, au Lycée nouvelles entières), poèmes articles de presse texte "documentaire" (expliquant un fait) page web dialogues (souvent fabriqués dans un butr d'enseignement au Collège, extraits de film ou interview au Lycée - à partir du niveau A2)
166	NARRATIFS/DESCRITPTIFS/INJONCTIFS
165	monologues, dialogues, biographies,narrations, contes
164	extraits de romans (classiques ou non) extraits de journaux ou magazines dialogues en langage très courant chez les jeunes
172	Extraits de romans, articles de presse, poèmes, contes, dialogues (théâtre ou non), publicités.
68	biographies, dialogues, extraits de romans, articles
47	roman, conte, journal, chanson, théâtre, script d'un film mais il existe d'autres supports comme la peinture, un extrait du journal télévisé ou d'une émission..
22	journalistiques et littéraires
187	tous
31	textes littéraires en prose (roman, nouvelles), essais, bandes dessinées, articles de presse (quotidiens, périodiques, presse spécialisée), notices techniques, dépliants, manuels d'utilisation
101	dialogue texte relevant de l'écrit
196	informatif, descriptif, argumentatif
201	articles chansons dialogues (films, ceux des manuels) extraits de romans, de pièces de théâtre BD ...
24	presse écrite, radio, télé

IDPROF	RÉPONSES
206	articles de journaux, lettres
208	autant que possible tous les types de textes : tous les genres littéraires et toutes sortes de textes ou documents non littéraires (pub, mode d'emploi, recette, article, chanson, BD, forum, ...)
209	textes enregistrés, textes écrits venant de manuels, presse, littérature...
210	tous : presse théâtre poésie littérature chanson synopsis
213	annonces pour l'emploi courriels CV
215	résumés de livres ; règlements ; critiques de films
216	Dialogues Texte / récit
230	articles de journaux, chansons, poèmes, dialogues, scénarios, publicités, extraits de romans, tracts, affiches, recettes de cuisine, notices, ...
236	dialogues, articles, récits, lettres, publicité...
	informatif, narratif, descriptif, explicatif, argumentatif...
237	tous types
238	littéraires, journalistiques, informatifs (genre "Francoscopie")
232	Informatifs, narratifs, descriptifs
241	articles de journaux, chansons, extraits de romans, transcriptions d'un reportage, tableaux statistiques, chapeaux d'articles divers
240	textes littéraires, presse, dialogues, chansons
242	tous (descriptif, narratif, argumentatif, explicatif, informatif)
247	déjà écrit : articles de presse, articles de recherche publiés, working papers, scripts de reportages ou d'entretiens...
66	tout type de texte
18	littéraires, dialogues, articles de journaux, news sur le net...
40	contes, articles, poèmes, chansons, extraits de romans...
88	cela va de la poésie, au tongue twister, à une publicité, une chanson, un extrait d'œuvre, un article de journal....
254	poésie, chansons, histoires, notice de montage, recette de cuisine, lettre, programmes
131	articles, occ. statistiques, schémas, documents quotidiens pratiques(menus, mode d'emploi etc...)
43 personnes ont répondu à cette question	

TABLE F.80: Réponses à la question n° 75 du second questionnaire

F.5.1 Catégorisation des résultats

Nous avons posé aux enseignants la question : « Quels types de textes abordez-vous avec vos apprenants ». L'objectif de cette question était double : savoir effectivement les types de textes

employés mais aussi et surtout en quels termes ils désignent ces différents types de textes. L'intégralité des réponses se trouve dans le tableau F.80. Dans le tableau ci-dessous (tableau F.81), nous avons isolé tous les termes employés, indiqué lesquels nous avons regroupés et quels enseignants les avaient employés.

TERME REPRÉSENTATIF	TERMES EMPLOYÉS	ENSEIGNANTS
Journalistique	journalistique article articles de presse articles de presse (quotidiens, périodiques, presse spécialisée) texte de presse articles de journaux article de journal textes journalistiques extraits de journaux ou magazines journal presse écrite presse	[146], [22], [238] [68], [201], [208], [236], [40], [131] [39], [94], [172], [247] [31] [85] [32], [206], [230], [241], [18] [88] [38] [164], [206] [47] [24] [209], [210], [240]
Littéraire	littéraire texte "littéraire" textes littéraires textes littéraires en prose tous les genres littéraires littérature extrait d'œuvre	[22], [238], [18] [94] [240] [31] [208] [209], [210] [88]
Roman	roman extrait de roman extrait de roman (classique ou non)	[146], [39], [85], [32], [47], [31] [94], [172], [68], [201], [230], [241], [40] [164]
Nouvelle	nouvelle extrait de nouvelle	[39], [31] [94]
Récit	récit	[216], [236]
Conte	histoire conte	[254] [109], [165], [47], [40]
Théâtre	théâtre extraits de pièces de théâtre	[85], [32], [47], [210] [201]
Script	script d'un film scenarior dialogues (films, ceux des manuels) transcription d'un reportage script de reportages ou d'entretiens	[47] [230] [201] [241] [247]
Synopsis	synopsis	[210]
Critique	critique de films	[215]

TERME REPRÉSENTATIF	TERMES EMPLOYÉS	ENSEIGNANTS
Résumé	résumés de livres	[215]
Poésie	poésie poèmes	[39], [85], [32], [210], [88], [254] [94], [172], [230], [40]
Chanson	chanson	[39], [47], [201], [208], [210], [230], [241], [240], [40], [88], [254]
Articles de recherche	article de recherche publiés	[247]
Essai	essai	[31]
Rapport	working papers	[247]
Dialogue	dialogue dialogues en langage très courant chez les jeunes dialogues (théâtre ou non) dialogues (films, ceux des manuels)	[146], [94], [165], [68], [101], [216], [230], [236], [240], [18] [164] [172] [201]
Monologue	monologue	[165]
Biographie	biographie	[165], [68]
Autobiographie	autobiographie	[32]
Informatif	informatif informatif (genre "Francoscopie")	[196], [236], [232], [242] [238]
Explicatif	texte "documentaire" (expliquant un fait) explicatif	[94] [236], [242]
Descriptif	descriptif	[146], [166], [196], [236], [232], [242]
Argumentatif	argumentatif	[196], [236], [242]
Narratif	narratif narration	[146], [166], [236], [232], [242] [165]
Injonctif	injonctif	[166]
Page Web	page web news sur le net	[94] [18]
Forum Correspondance	forum lettre courriel	[208] [206], [236], [254] [210]
Textes authentiques	documentation authentique	[109]
Tongue twister ³	tongue twister	[88]
Textes pas très compli- qués	textes pas très compliqués	[109]
Manuel	textes écrits venant de manuels	[209]
Recette	recette recette de cuisine	[208] [230], [254]

3. Phrase difficile à prononcer. Ex : Les chaussettes de l'archiduchesse sont elles sèches, archi-sèches ?

TERME REPRÉSENTATIF	TERMES EMPLOYÉS	ENSEIGNANTS
Règlement	règlements	[215]
Documents quotidiens pratiques	Documents quotidiens pratiques (menu, mode d'emploi etc...)	[131]
Menu	menu	[131]
Mode d'emploi	notice technique manuel d'utilisation mode d'emploi notice notice de montage	[31] [31] [208], [131] [230] [254]
Annonces	annonces pour l'emploi	[213]
Curriculum Vitae	CV	[213]
Publicité	publicités pub	[172], [230], [236], [88] [208]
Tract	tracts	[230]
Dépliant	dépliant	[31]
Types de textes a priori non utilisables dans le cadre de notre base de textes	oral et écrits - peinture, extrait du journal télévisé ou émission tous Bande dessinée texte relevant de l'écrit radio, télé textes enregistrés texte affiches tableaux statistiques, chapeaux d'articles divers programmes occ. statistiques, schémas	[146] [65] [47] [187], [237], [66] [31], [201], [208] [101] [24] [209] [216] [230] [241] [254] [131]

TABLE F.81: Classement des réponses à la question 75

QUESTION	Quels types de textes abordez-vous avec vos apprenants ?	
CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
Journalistique	 65.1%	28 réponses parmi 43
Types non utilisables ici	 39.5%	17 réponses parmi 43
Roman	 32.6%	14 réponses parmi 43
Dialogue	 30.2%	13 réponses parmi 43
Chanson	 25.6%	11 réponses parmi 43

CHOIX	% DE RÉPONSES CONTENANT CE CHOIX	DÉTAIL
Littéraire	<input type="checkbox"/>	23.3% 10 réponses parmi 43
Poésie	<input type="checkbox"/>	23.3% 10 réponses parmi 43
Descriptif	<input type="checkbox"/>	14% 6 réponses parmi 43
Narratif	<input type="checkbox"/>	14% 6 réponses parmi 43
Théâtre	<input type="checkbox"/>	11.6% 5 réponses parmi 43
Script	<input type="checkbox"/>	11.6% 5 réponses parmi 43
Informatif	<input type="checkbox"/>	11.6% 5 réponses parmi 43
Mode d'emploi	<input type="checkbox"/>	11.6% 5 réponses parmi 43
Publicité	<input type="checkbox"/>	11.6% 5 réponses parmi 43
Conte	<input type="checkbox"/>	9.3% 4 réponses parmi 43
Correspondance	<input type="checkbox"/>	9.3% 4 réponses parmi 43
Nouvelle	<input type="checkbox"/>	7% 3 réponses parmi 43
Récit	<input type="checkbox"/>	7% 3 réponses parmi 43
Explicatif	<input type="checkbox"/>	7% 3 réponses parmi 43
Argumentatif	<input type="checkbox"/>	7% 3 réponses parmi 43
Recette	<input type="checkbox"/>	7% 3 réponses parmi 43
Biographie	<input type="checkbox"/>	4.7% 2 réponses parmi 43
Page Web	<input type="checkbox"/>	4.7% 2 réponses parmi 43
Synopsis	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Critique	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Résumé	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Articles de recherche	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Essai	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Rapport	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Monologue	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Autobiographie	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Injonctif	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Forum	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Textes authentiques	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Tongue twister	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Textes pas très compliqués	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Manuel	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Règlement	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Documents quotidiens pratiques	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Menu	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Annonces	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Curriculum Vitae	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Tract	<input type="checkbox"/>	2.3% 1 réponses parmi 43
Dépliant	<input type="checkbox"/>	2.3% 1 réponses parmi 43
43 personnes ont répondu à cette question et ont choisi en moyenne 4.4 solutions par personne		

TABLE F.82: Catégorisation des réponses à la question n° 75 du second questionnaire

F.6 Suggestions

QUESTION	En imaginant que TOUT soit possible, quel(s) critère(s) de recherche de textes vous paraissent primordiaux ? Même si le critère vous semble proche de critères suggérés par les questions précédentes, n'hésitez pas à le proposer malgré tout.
IDPROF	RÉPONSES
146	les critères morphologiques, les recherches sur le lexique (mots isolés et expressions) ; en nombre et en liste
39	Recherche de textes en fonctions du thème et des objectifs grammaticaux
85	thème, longueur
65	les objectifs de communication avant tout certains objectifs grammaticaux ou lexicaux éventuellement
32	Thème (et vocabulaire s'y rapportant). Structures grammaticales.
109	actualité
57	la facilité de compréhension et le choix du thème
21	presence de notions culturelles éveillant la curiosité thème createur de débat
114	Désolée, votre questionnaire est particulièrement incompréhensible, et très mal expliqué aussi ne puis-je continuer de l'instruire. Si toutefois, vous reformulez ce questionnaire de façon compréhensible, et non avec des termes "IUFMesques" réservés aux initiés, grands ou petits, je serai heureuse de contribuer à votre travail. Véronique *****, enseignante d'anglais en Fac d'anglais et *****
38	date
94	a) Tout d'abord : le thème (adéquation avec les programmes en vigueur, notamment en ce qui concerne le contenu civilisationnel) b) le niveau pour lequel on peut l'utiliser (en lien avec la grille de compétences du cadre européen de référence) c) le potentiel didactique (à l'entraînement de quelle compétence le texte sert-il ?) d) "le degré d'implicite" j'ai du mal à le formuler, c'est en lien avec le potentiel didactique. Il faudrait un critère (une échelle) qui indique dans quel mesure le texte peut susciter la curiosité des élèves. e) le média (sous quel forme le texte existe-t-il ?
166	THEME DU TEXTE LIE AUX CENTRES D'INTERET DS JEUNES/ CIVILISATION/CULTURE

IDPROF	RÉPONSES
165	un rapprochement avec les tâches à proposer dans le CECR! C'est déjà compliqué pour moi, alors ça faciliterait les choses.
164	Oui, plus d'items proches des descripteurs et tâches préconisés par les CECR Un thème susceptible de les intéresser et qui leur permette de "polémiquer" (ex : le rôle de la femme dans la société / l'enseignement idéal / le cinéma / la conquête spatiale / le racisme)
172	Le thème abordé avant tout.
68	thème ; structure qui apparaît dans le texte, longueur
47	Tout est important en réalité mais étudier un texte ne doit pas être un prétexte pour étudier la grammaire
22	thème et grammaire
187	qui intéresse les élèves
101	objectifs de communication ; thème ; structures linguistiques
196	thème
23	documents authentiques ; thème ; voc ; temps grammaticaux ; ton (sérieux, statistique, humoristique, ironique...)
200	thème
201	thème vocabulaire et structures grammaticales type de texte longueur
24	Thèmes d'actualité ou en relation directe avec les disciplines enseignées (Marketing, RH, communication, etc...)
206	thèmes, structures, vocabulaire, phonologie.
208	type de texte thème auteur longueur notion grammaticale présente dans le texte champ lexical présent dans le texte
209	il faut que le sujet intéresse les jeunes du groupe d'âge visé.
210	l'intérêt qu'ils suscitent chez mes ELEVES (je déteste ce terme ringard d'apprenant) la culture que cela leur apporte, la cohérence avec mes choix de séquences.
213	l'actualité
215	temps de conjugaison ; tournures grammaticales ; thème
216	pour des textes audio, l'origine de la ou des personnes qui parlent (accent)
230	les champs lexicaux et les structures grammaticales

IDPROF	RÉPONSES
231	-rareté/fréquence des mots (critères que l'on trouve dans les dictionnaires unilingues pour apprenants type Longman Dictionary of Contemporary English) -dérivation des mots (pour le travail sur l'inférence) -régularité de prononciation (même sur des textes écrits)
236	champs lexicaux dérivations lexicales flexions catégories grammaticales structures grammaticales
237	thème, tournure, type de texte, époque, auteur
238	Les textes littéraires
232	L'intérêt des étudiants. Que ces textes puissent les aider dans leur adaptation à leur nouveau pays.
241	occurrences grammaticales, vocabulaire par champ lexical, vocabulaire par type de mots (noms, adjectifs...), style, type de texte, niveau dans le cadre européen commun de référence, objectif communicatif
240	le thème les structures grammaticales le type de texte
242	thème ; vocabulaire, style...
247	thèmes, type de document, registre, variété de langue (pour anglais, ang GB et US par ex)...
66	sujet du texte, auteur, niveau de complexité, richesse culturelle, richesse de l'implicite, richesse des idées
252	le plus facile qui relate des événements de leurs environnements
18	motivation des élèves, actualité, structures grammaticales,
26	est-ce que le vocabulaire du texte appartient aux mots les plus fréquents / rares de la langue enseignée ? (300/3000/30000 mots) y a-t-il beaucoup de mots transparents (selon langue des élèves) thème
40	recherche par champs lexicaux, par points grammaticaux (ser y estar ; gustar...) par structure (mas....que ; no ... sino) par temps (impératif négatif, imparfait, présent progressif...), par thème (la famille, l'environnement, les aztèques...) . Quand je recherche un texte j'essaie de croiser 2 de ces critères entre eux, en sachant qu'à partir de trois, c'est quasi-impossible car très fastidieux (on y passe beaucoup de temps).
88	celui qui intéresse les élèves, les informe sur un fait de société ou culturel, les amenant à discuter.
254	type de texte
131	complexité syntaxique. Nota : je travaille avec des hispanophones
50 personnes ont répondu à cette question	

TABLE F.83: Réponses à la question n° 76 du second questionnaire

Annexe G

Documentation de la classe BD de la maquette



Utilitaires
BasesDeDonnées

[class tree: Utilitaires] [index: Utilitaires] [all elements]

Utilitaires

Packages:

phpDocumentor
Configuration
Exemples
GénérationFacettes
GénérationRequetes
GroupementTextes
Utilitaires

Files:

BasesDeDonnées

Champs.sql.php
classBD.php
classTableObjet.php
configBD.inc.php
Outils.sql.php
Prismes.sql.php
Recherche.sql.php

GestionDesErreurs

gestionDesErreursUtilisateur.php

GestionDesFichiers

gestionFichiers.php

Classes:

BasesDeDonnées
BD
TableObjet

Class: BD

Source Location: /BD/classBD.php

Class BD

Class Overview

[Class Overview](#) | [Property Summary](#) | [Properties Detail](#) | [Method Summary](#) | [Methods Detail](#) | [Constants Summary](#) | [Constants Detail](#)

Une classe pour l'encapsulation de la gestion des bases de données.

Cette classe permettant (conjointement avec TableObjet) de récupérer des objets, on obtient ainsi une approximation d'interface de OODBMS sans respecter ou s'occuper des problèmes d'accès concurrentiel aux données. Pour ce prototype l'utilisateur ne modifiera pas les objets qu'il récupère. Les modifications toucheront soit des enregistrements non issus d'une classe décrite par un objet TableObjet (accès à une BD relationnelle standard), soit des objets qui ne sont pas encore "persistants" (pas d'accès concurrentiel, puisqu'avant sauvegarde dans la BD, ils n'existeront que pour l'utilisateur qui les crée et qu'après ils ne seront accessibles qu'en lecture).

Les variables de classe sont toutes privées et accessibles uniquement via les sélecteurs et modificateurs (sauf pour l'utilisateur et le mot de passe auxquels aucune méthode ne permet d'accéder). Les constantes de classes permettent de définir les types de bases de données prises en charge, pour le moment seulement mysql.

Located in /BD/classBD.php [line 43]

Properties

- \$base
- \$host
- \$ID
- \$nbUnreadLines
- \$password
- \$req
- \$res
- \$type
- \$user

Methods

- __construct
- connectee
- connexion
- erreurRequete
- getBase
- getHost
- getID
- getNbUnreadLines
- getNextObject
- getRequete
- getResultat
- getTypeBD
- insertID
- nbRes
- requete
- select1ouPlus
- selectionBD
- setID
- setNbUnreadLines
- setRequete
- setResultat

Author(s):

- Mathieu Loiseau <loiseaum@gmail.com>

API Tags:

See: `BD::__mysql`

Usedby: `configBD.inc.php` - Pour pouvoir créer des instances de BD, avec toutes les bases de données auxquelles on sera amené à se connecter.

Usedby: `classPrisme.php` - Pour stocker et récupérer des Prismes

Usedby: `outils.php` - Pour exemple

Usedby: `classCollection.php` - Pour stocker et récupérer des Textes

Usedby: `Recherche::__construct()` - pour récupérer la Recherche dans la BD.

Docs For Class BD

Usedby: Collection - Pour stocker et récupérer des Textes

Usedby: Outil::__construct() - Dans les versions 2 et 3, quand on veut récupérer un outil depuis une BD

Usedby: Outil::store() - pour stocker l'outil dans une BD

Usedby: Prisme::__construct() - pour récupérer le Prisme dans la BD.

Usedby: Prisme::store() - pour stocker le prisme dans une BD

Usedby: Collection::__construct() - pour récupérer le Prisme dans la BD.

Information Tags:

[Top]

Constant Summary

[Class Overview](#) | [Constants Summary](#) | [Constants Detail](#) | [Property Summary](#) | [Properties Detail](#) | [Method Summary](#) | [Methods Detail](#)

`_mysql` Constante de classe, pour la gestion du langage MySQL.

[Top]

Property Summary

[Class Overview](#) | [Property Summary](#) | [Properties Detail](#) | [Method Summary](#) | [Methods Detail](#) | [Constants Summary](#) | [Constants Detail](#)

<i>string</i>	\$base	le nom de la base
<i>string</i>	\$host	l'hôte
<i>resource</i>	\$ID	link_identifier de la base, pour la gestion des requêtes
<i>integer</i>	\$nbUnreadLines	le nombre de lignes à lire pour la requête courante (select)
<i>string</i>	\$password	le mot de passe
<i>string</i>	\$req	la dernière requête effectuée
<i>resource</i>	\$res	le handler du résultat de la dernière requête
<i>integer</i>	\$type	type de BD, cf. constantes de classe
<i>string</i>	\$user	le nom d'utilisateur

[Top]

Method Summary

[Class Overview](#) | [Property Summary](#) | [Properties Detail](#) | [Constants Summary](#) | [Constants Detail](#) | [Method Summary](#) | [Methods Detail](#)

<i>BD</i>	__construct()	Seul constructeur pour la classe
<i>boolean</i>	connectee()	Méthode disant si la base est connectée
<i>resource/boolean</i>	connexion()	Méthode qui connecte, puis sélectionne la base et renvoie le résultat de la connexion
<i>boolean</i>	erreurRequete()	Méthode qui dit si la dernière requête a renvoyé une erreur.
<i>string</i>	getBase()	Sélecteur
<i>string</i>	getHost()	Sélecteur
<i>resource/boolean</i>	getID()	Sélecteur
<i>int</i>	getNbUnreadLines()	Sélecteur
<i>integer/boolean</i>	getNextObject()	Méthode qui renvoie l'objet suivant dans la liste des réponses (lignes) à la dernière requête effectuée
<i>string</i>	getRequete()	Sélecteur
<i>resource</i>	getResultat()	Sélecteur
<i>int</i>	getTypeBD()	Sélecteur
<i>integer/boolean</i>	insertID()	Renvoie l'identifiant (auto_increment) de la dernière opération d'insertion
<i>integer/boolean</i>	nbRes()	Méthode qui dit si la dernière requête effectuée a renvoyé une erreur.
<i>boolean</i>	requete()	Méthode qui exécute une requete
<i>boolean</i>	select1ouPlus()	Méthode qui exécute une requete select dans le but d'obtenir au moins une ligne satisfaisant la requête.
<i>resource/boolean</i>	selectionBD()	Méthode qui sélectionne la base à utiliser
<i>resource/boolean</i>	setID()	Modificateur
<i>integer</i>	setNbUnreadLines()	Modificateur

Docs For Class *BD*

<i>string</i>	setRequete()	Modificateur
<i>resource</i>	setResultat()	Modificateur

[Top]

Properties

[Class Overview](#) | [Property Summary](#) | [Properties Detail](#) | [Method Summary](#) | [Methods Detail](#) | [Constants Summary](#) | [Constants Detail](#)

string **\$base** [line 54]

le nom de la base

API Tags:

Access: private

[Top]

string **\$host** [line 49]

l'hôte

API Tags:

Access: private

[Top]

resource **\$ID** [line 76]

link_identifieur de la base, pour la gestion des requêtes

API Tags:

Access: private

[Top]

integer **\$nbUnreadLines** [line 91]

le nombre de lignes à lire pour la requête courante (select)

API Tags:

Access: private

[Top]

string **\$password** [line 64]

le mot de passe

API Tags:

Access: private

[Top]

string **\$req** [line 81]

la dernière requête effectuée

API Tags:

Access: private

Docs For Class BD

[\[Top \]](#)*resource* **\$res** [line 86]**le handler du résultat de la dernière requête****API Tags:****Access:** private[\[Top \]](#)*integer* **\$type** [line 71]**type de BD, cf. constantes de classe****API Tags:****See:** BD::*mysql***Internal:** Pour le moment, ici on aura que du *mysql***Access:** private[\[Top \]](#)*string* **\$user** [line 59]**le nom d'utilisateur****API Tags:****Access:** private[\[Top \]](#)

Methods

[Class Overview](#) | [Property Summary](#) | [Properties Detail](#) | [Constants Summary](#) | [Constants Detail](#) | [Method Summary](#) | [Methods Detail](#)**Constructor `__construct`** [line 110]

```
BD __construct( string $aHost, string $aBase, string $aUser,
string $aPassword, integer $aType )
```

Seul constructeur pour la classe**Parameters:**

string **\$aHost:** L'hôte hébergeant la base
string **\$aBase:** Le nom de la base
string **\$aUser:** Le nom d'utilisateur pour la connexion
string **\$aPassword:** Le mot de passe pour la connexion
integer **\$aType:** Le type de base de données utilisé. Le seul type pris en charge est
BD::*mysql*

API Tags:**Return:** Un objet base de données utilisable[\[Top \]](#)**connectee** [line 228]

Docs For Class BD

```
boolean connectee( )
```

Méthode disant si la base est connectée

création : 08/08/2007

API Tags:**Return:** vrai si la base est connectée, faux sinon

[Top]

connexion [line 240]

```
resource | boolean connexion( )
```

Méthode qui connecte, puis sélectionne la base et renvoie le résultat de la connexion

création : 08/08/2007

API Tags:**Return:** Le résultat de la sélection de la base (resource identifier), faux si la connexion ou la sélection échoue.**Internal:** Erreurs possibles à cause de problèmes de connexion, de sélection ou de type de base

[Top]

erreurRequete [line 364]

```
boolean erreurRequete( )
```

Méthode qui dit si la dernière requête a renvoyé une erreur.

Création : 08/08/2007

API Tags:**Return:** Vrai si la requête s'est exécutée sans erreur.

[Top]

getBase [line 143]

```
string getBase( )
```

Sélecteur**API Tags:****Return:** Le nom de la base de données

[Top]

getHost [line 125]

```
string getHost( )
```

Sélecteur

API Tags:

Return: L'hôte hébergeant la base

[Top]

getID [line 135]

```
resource|boolean getID( )
```

Sélecteur

API Tags:

Return: L'identifiant de la base si la connexion a été faite, faux sinon.

Internal: Peut-être un peu lié à MySQL, mais pas sûr...

[Top]

getNbUnreadLines [line 175]

```
int getNbUnreadLines( )
```

Sélecteur

API Tags:

Return: Le nombre de lignes non lues de la dernière requête effectuée.

[Top]

getNextObject [line 401]

```
integer|boolean getNextObject( [TableObjet|boolean
    $tableObjet = false] )
```

Méthode qui renvoie l'objet suivant dans la liste des réponses (lignes) à la dernière requête effectuée

Création : 09/08/2007

Parameters:

TableObjet|boolean **\$tableObjet:** Si \$tableObjet est un objet TableObjet alors on va se servir d'une classe pour récupérer les lignes de la base TableObjet

API Tags:

Return: Le nombre de résultats à la dernière requête (false s'il y a erreur ou si c'est un insert ou assimilé)

Uses: TableObjet - Pour récupérer directement des objets de la classe TableObjet::getNomClasse() et non des objets indéfinis.

Docs For Class BD

[Top]
<p>getRequete <small>[line 159]</small></p> <pre>string getRequete()</pre> <p style="text-align: center;">Sélecteur</p> <p>API Tags: Return: La dernière requête effectuée.</p>
[Top]
<p>getResultat <small>[line 167]</small></p> <pre>resource getResultat()</pre> <p style="text-align: center;">Sélecteur</p> <p>API Tags: Return: Le handler de la dernière requête effectuée. Access: private</p>
[Top]
<p>getTypeBD <small>[line 151]</small></p> <pre>int getTypeBD()</pre> <p style="text-align: center;">Sélecteur</p> <p>API Tags: Return: Le type de la base de données</p>
[Top]
<p>insertID <small>[line 453]</small></p> <pre>integer boolean insertID()</pre> <p style="text-align: center;">Renvoie l'identifiant (auto_increment) de la dernière opération d'insertion</p> <p>Création : 10/08/2007</p> <p>API Tags: Return: L'ID généré pour une colonne AUTO_INCREMENT par la dernière requête INSERT en cas de succès, 0 si la dernière requête n'a pas généré de valeur AUTO_INCREMENT, ou FALSE si aucune connexion MySQL n'a été établie. cf. <code>mysql_insert_id</code> Internal: Sûrement pas utilisable pour toutes les BD.</p>
[Top]

nbRes [line 381]

```
integer|boolean nbRes( )
```

Méthode qui dit si la dernière requête effectuée a renvoyé une erreur.

Création : 09/08/2007

API Tags:

Return: Le nombre de résultats à la dernière requête (false s'il y a erreur ou si c'est un insert ou assimilé)

[Top]

requete [line 302]

```
boolean requete( string $requete, [boolean $verbose = true] )
```

Méthode qui exécute une requete

création : 08/08/2007

Parameters:

string **\$requete:** La requête à effectuer

boolean **\$verbose:** Vrai pour afficher les messages d'erreur, faux pour ne pas les voir

API Tags:

Return: Vrai si la requête s'exécute sans erreur.

[Top]

select1ouPlus [line 345]

```
boolean select1ouPlus( string $requete, [boolean $verbose = true] )
```

Méthode qui exécute une requete select dans le but d'obtenir au moins une ligne satisfaisant la requête.

Comme BD::requete() sauf qu'elle ne marche que pour les SELECT et qu'elle renvoie faux si la requete ne renvoie pas d'éléments

création : 08/08/2007

Parameters:

string **\$requete:** La requête à effectuer

boolean **\$verbose:** Vrai pour afficher les messages d'erreur, faux pour ne pas les voir

API Tags:

Return: Vrai si la requête s'exécute sans erreur et renvoie au moins un résultat.

Internal: On ne teste pas la requête, on ne fait qu'afficher un warning si verbose est vrai

[Top]

selectionBD [line 277]

Docs For Class BD

```
resource|boolean selectionBD( )
```

Méthode qui sélectionne la base à utiliser

(possibilité de changer la base, pour un objet donné) création : 08/08/2007

API Tags:

Return: Le résultat de la sélection de la base (resource identifiant), faux si la sélection échoue.

Internal: Erreurs possibles à cause de problèmes de sélection, on ne traite pas le cas (improbable) d'un type de base non traité car la connexion sera impossible de toutes façons.

[Top]

setID [line 186]

```
resource|boolean setID( resource|boolean $newID )
```

Modificateur

Parameters:

resource|boolean **\$newID:** Une nouvelle valeur de l'identifiant de la BD.

API Tags:

Return: Le résultat de l'affectation de \$newID à la variable de classe.

[Top]

setNbUnreadLines [line 215]

```
integer setNbUnreadLines( integer $entier )
```

Modificateur

Parameters:

integer **\$entier:** Un nouveau nombre de lignes non lues.

API Tags:

Return: La valeur de l'affectation de \$entier à la variable de classe.

Access: private

[Top]

setRequete [line 195]

```
string setRequete( string $req )
```

Modificateur

Parameters:

string **\$req:** Une requete à effectuer.

API Tags:

Return: La valeur de l'affectation de \$req à la variable de classe.

Access: private

Docs For Class BD

[\[Top \]](#)**setResultat** [line 205]

```
resource setResultat( resource $res )
```

Modificateur**Parameters:**

resource **\$res**: Le nouveau résultat

API Tags:

Return: La valeur de l'affectation de \$req à la variable de classe.

Access: private

[\[Top \]](#)**Constants**[Class Overview](#) | [Constants Summary](#) | [Constants Detail](#) | [Method Summary](#) | [Methods Detail](#)**_mysql = 1** [line 97]

Constante de classe, pour la gestion du langage MySQL.

[\[Top \]](#)

Résumé

Cette thèse traite la notion d'indexation pédagogique et l'aborde sous l'angle de la recherche et du choix de textes pour l'enseignement des langues. Ce problème est replacé dans le champ disciplinaire de l'Apprentissage des Langues Assisté par Ordinateur (ALAO) et des apports potentiels du Traitement Automatique des Langues (TAL) à cette discipline, avant d'être confronté à des éléments provenant plus directement de la didactique des langues, pour proposer une approche empirique. Celle-ci est ensuite justifiée par l'incapacité des standards de description de ressources pédagogiques actuels à modéliser des ressources brutes de manière cohérente, en particulier des textes pour l'enseignement des langues. La thèse s'articule ensuite autour de deux questionnaires visant à connaître les pratiques déclarées des enseignants quant à la recherche et au choix de textes dans le cadre de la tâche de planification des cours. Le premier questionnaire permet la formalisation de la notion de contexte pédagogique, qui est ultérieurement appréhendée à travers certaines de ses composantes grâce au second questionnaire. Enfin, ces premières ébauches de formalisation servent de fondement à la définition d'un modèle dont l'objectif est de rendre compte de la contextualité des propriétés dites pédagogiques quand elles sont appliquées à des ressources brutes. Des pistes d'implantation du modèle sont finalement proposées dans le cadre de la description d'un système informatique.

Mots-clés: Indexation pédagogique, ALAO, EIAH, Didactique des langues, Métadonnées, TAL, Modélisation, Corpus

Abstract

This PhD thesis deals with the notion of pedagogical indexation and tackles it from the point of view of searching for and selecting texts for language teaching. This particular problem is set in the field of Computer Assisted Language Learning (CALL) and of the potential contribution of Natural Language Processing (NLP) to this discipline, before being considered within the scope of elements more directly relevant to language didactics, in order to propose an empirical approach. The latter is then justified by the inadequacy of current description standards for pedagogical resources where modeling of raw objects in a consistent fashion is concerned. This is particularly true for texts in the context of language learning. The thesis subsequently revolves around two questionnaires the aim of which is to provide insight into language teachers' declared practices regarding searching for and selecting texts in the context of class planning. The first questionnaire provides data to formalize the notion of pedagogical context, which is later considered through some of its components thanks to the second questionnaire. Finally, these first formalization drafts provide foundations for the definition of a model aiming at taking into account the contextuality of the properties said to be pedagogical, which is inherent to raw resources. Finally, possible leads for implementing this model are suggested through the description of a computerized system.

Keywords: Pedagogical indexation, CALL, Language didactics, Metadata, NLP, Modeling, Corpus