



HAL
open science

Médiatiser l'annotation pour une herméneutique numérique : AnT&CoW, un collecticiel pour une coopération via l'annotation de documents numériques

Gaëlle Lortal

► To cite this version:

Gaëlle Lortal. Médiatiser l'annotation pour une herméneutique numérique : AnT&CoW, un collecticiel pour une coopération via l'annotation de documents numériques. Autre [cs.OH]. Université de Technologie de Troyes, 2006. Français. NNT : . tel-00136042

HAL Id: tel-00136042

<https://theses.hal.science/tel-00136042>

Submitted on 11 Mar 2007

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.

UNIVERSITÉ DE TECHNOLOGIE DE TROYES

THÈSE

EN VUE DE L'OBTENTION DU GRADE DE
DOCTEUR DE L'UNIVERSITÉ DE TECHNOLOGIE DE TROYES

DISCIPLINE : INFORMATIQUE

**Médiatiser l'annotation pour une herméneutique
numérique :**
**AnT&CoW, un collecticiel pour une coopération via l'annotation
de documents numériques**

présentée par
Gaëlle Lortal

Novembre 2006

Cette thèse est intitulée :

**Médiatiser l'annotation pour une herméneutique
numérique :**
**AnT&CoW, un collecticiel pour une coopération via l'annotation
de documents numériques**

présentée par :

Gaëlle Lortal

a été évaluée par un jury composé des personnes suivantes :

M. Jean Charlet, rapporteur

M. Gérard Sabah, rapporteur

M. Manuel Zacklad, examinateur, président du jury

Mme Sylvie Calabretto, examinateur

M. Michaël Zock, membre invité

Mme Myriam Lewkowicz, co-directrice de recherche

Mme Amalia Todirascu-Courtier, co-directrice de recherche

Résumé

Suite au projet CNRS/ PI-TCAN Mediannote, centré sur les activités coopératives en conception mécanique, le projet Mediannote, dans lequel s'inscrivent nos travaux, se concentre sur l'annotation dans ces activités. Dans un contexte où les échanges médiatisés s'accroissent, le document numérique devient central. Pour soutenir les échanges et la construction d'une interprétation collective autour de ce document, nous proposons d'instrumenter l'annotation - définie comme un fragment de discours à propos d'un texte, un support à l'argumentation -. Nous proposons donc de concevoir un collecticiel pour annoter collectivement des documents numériques et ainsi soutenir l'herméneutique numérique.

L'annotation comme support au travail coopératif est envisagée à la fois comme un objet qui relève de l'étiquette et du commentaire et comme une activité qui relève de la communication, de l'indexation et de l'élaboration de discours. La conception de notre collecticiel se fonde sur un modèle d'activité d'annotation qui souligne la dimension interactionnelle et coopérative de l'annotation. Cette démarche guidée par les modèles est enrichie par l'utilisation de corpus qui permet de conserver l'utilisateur final au centre de nos préoccupations. Nous présentons une maquette du collecticiel, AnT&CoW, utilisant des outils de TAL pour le soutien de l'utilisateur à différents niveaux : soutien à la construction de classification et aide à l'indexation. Une première étape d'évaluation de cette maquette est également présentée, une validation des fonctionnalités.

Mots-clés : Travail Coopératif Assisté par Ordinateur, Ingénierie des Connaissances, Traitement Automatique des Langues, Collecticiel, Document numérique, Herméneutique, Annotation, Logique de communication, Ressources Termino-Ontologiques, Web Sémantique.

Abstract

Our work is related with the CNRS/PI-TCAN Mediannote project, following a previous project called Mediapro which was focused on cooperative activities in mechanical design. While mediated exchanges are increasing everyday, digital document is a central element and tool. We design a groupware to support exchanges and collective sense making about documents via annotations, so to support digital hermeneutics. In this context, an annotation is a discourse fragment talking about a text, which is an argumentation support.

This research aims at defining annotation for cooperative work. The annotation is not only an object used as a tag as well as a comment, but also an activity related to communication, indexation and discourse elaboration. Our groupware design process is based on an activity model emphasizing the interactional and cooperative dimensions of annotation activity. This model-oriented, user-centered framework is designed on the basis of corpora studies.

Hence, we propose the AnT&CoW groupware, supporting annotation activities by means of visualisation, classification and editing. It integrates NLP tools to support classification building and fragments indexing. A first evaluation of this groupware has been leaded.

Keywords : Computer Supported Cooperative Work, Knowledge Engineering, Natural Language Processing, Groupware Design, Digital document, Hermeneutics, Annotation, Communication Rationale, Termino-Ontological resources, Semantic Web.

Table des matières

Introduction	1
1^{ère} partie : Positionnement et méthodologie	9
1. Le Travail Coopératif Assisté par Ordinateur, un cadre adapté	11
1.1. Les objets-frontières	12
1.2. Les schèmes de classification	14
1.3. Les artefacts de négociation de frontières	14
2. Méthodologie de conception d'outil	15
2.1. Les modèles	16
2.2. Le corpus	19
2.2.1. <i>Le corpus : un genre</i>	19
2.2.2. <i>Exhaustivité ou pertinence ?</i>	21
2.2.3. <i>L'interprétation du corpus</i>	22
2.3. Les scénarios	23
2.3.1. <i>Un cadre pour la conception</i>	23
2.3.2. <i>Les atouts du scénario</i>	25
2.3.3. <i>Le scénario pour le prototypage et l'évaluation d'outil</i>	26
3. Conclusion	28
2^e partie : Autour de l'objet annotation	31
Chapitre I Des annotations sémantiques, discursives, élaborantes	33
1. L'annotation sémantique : un index pour mettre la main sur le document	34
1.1. Le Web Sémantique	35
1.2. La Recherche d'Information et de Document	36
1.2.1. <i>Les principes de base de la Recherche d'Information</i>	36
1.2.2. <i>RI et TAL</i>	37
1.3. Les ontologies	39
1.4. La boîte à outils de l'annotation sémantique	42
1.4.1. <i>Échanger des annotations : Annotea, protocole et serveur d'annotation</i>	43
1.4.2. <i>Clients et serveurs d'annotation</i>	45
2. L'annotation discursive	49
2.1. Herméneutique	49
2.1.1. <i>L'herméneutique médiévale</i>	50
2.1.2. <i>Une herméneutique littéraire ?</i>	50
2.1.3. <i>L'herméneutique philosophique</i>	51
2.2. La boîte à outils de l'annotation discursive	52

3. L'annotation élaborante	56
3.1. Web Social	56
3.2. Communication Médiatisée par Ordinateur	57
3.2.1. CMO et analyse du discours	58
3.2.2. CMO et Textualité	59
3.2.3. Des annotations en CMO ?	60
3.3. La boîte à outils de l'annotation élaborante	61
3.3.1. WYSIWYG, blog, Wiki et CMO	62
3.3.2. Des outils de rédaction coopérative	63

4. Conclusion	68
----------------------	-----------

Chapitre II Proposition d'une définition de l'annotation pour la coopération

73

1. Un lien vers le document	74
1.1. Folksonomies, ethno-classifications et classifications vulgaires	76
1.2. Attribution du sociolecte	79
1.2.1. Schibboleth	79
1.2.2. Point de vue	81
2. Un fragment de conversation	84
3. Un fragment de document	86
4. Une définition unifiée : L'annotation coopérative, une trace de la logique de communication	88

3^e partie : Autour de l'activité annotation

93

Chapitre III Quel modèle de description de l'activité ?

95

1. Les modèles cognitifs	95
1.1. La théorie propositionnelle (Kintsch et Van Dijk, 1978, Kintsch, 1988)	96
1.2. Le modèle de processus d'écriture de Hayes et Flower (1980)	97
1.3. Modèle de production verbale orale, (Levelt, 1989)	99
1.4. Van Wijk et Sanders, (1999)	100
2. Les modèles interactionnistes	102
2.1. Le modèle de l'interaction sociale (Nystrand, 1989)	102
2.2. Les modèles rhétoriques	104

Chapitre IV Proposition d'un modèle de l'activité d'annotation

107

1. Un modèle de description de l'activité d'annotation	107
2. Adaptation d'un modèle à la médiatisation	111

2.1. Modèle d'activité instrumentée	111
2.2. Objets du système	115
3. Conclusion	115
4^e partie : AnT&CoW : Description et simulation	119
Chapitre V AnT&CoW, un collecticiel pour la gestion des DoPAs dans un groupe projet	121
1. Documents pour l'action et documentarisation	121
2. Spécifications fonctionnelles	122
3. Une architecture client-serveur	126
3.1. Le serveur	126
3.2. Le client	130
4. Fonctionnalités du client AnT&CoW	131
4.1. Un espace partagé sur le serveur	131
4.2. Des fonctionnalités pour la communication	134
4.2.1. Ancrer une annotation	135
4.2.2. Saisir une annotation	138
4.2.3. Visualiser le fil d'annotation	140
4.3. Des fonctionnalités pour l'élaboration	140
4.3.1. Recueil d'annotations	141
4.3.2. Sélection suivant l'indexation	142
4.3.3. Édition d'un brouillon	143
4.3.4. Enregistrement d'un brouillon	143
4.4. Des fonctionnalités pour l'indexation	144
5. Une classification ajustée	145
5.1. Une Ressource Termino-Ontologique adaptée : la carte de thèmes	146
5.2. Démarche pour une construction sur mesure	147
5.2.1. Présentation du corpus de référence	149
5.2.2. Construction de la carte de thèmes	151
5.2.3. Construction de relations	157
5.2.4. Représentation de la carte de thèmes	162
5.2.5. Création d'une carte	169
5.2.6. Indexation selon plusieurs index et plusieurs granularités	171
6. Conclusion	173
Chapitre VI Scénarios d'utilisation d'AnT&CoW	175
1. Dispositif de simulation	175
1.1. Présentation du corpus de travail	177
1.2. Recueil du corpus	179
1.3. Traitement du corpus	180

	x
1.3.1. <i>Des fragments de conversation</i>	181
1.3.2. <i>Des fragments de document</i>	183
1.3.3. <i>Des fragments pour l'indexation</i>	184
2. Trois scénarios d'annotation dans un projet de conception distribuée	187
2.1. La communication	188
2.1.1. <i>Scénario d'annotation support à la communication</i>	188
2.1.2. <i>Scène 1 : Échange autour de documents</i>	189
2.2. L'élaboration	190
2.2.1. <i>Scénario d'annotation support à la création de documents</i>	190
2.2.2. <i>Scène 2 : Élaboration de document via des annotations</i>	190
2.3. L'indexation	193
2.3.1. <i>Scénario d'annotation support à l'indexation</i>	193
2.3.2. <i>Scène 3 : Construction de la classification et indexation</i>	194
3. Évaluation	196
3.1. Description du protocole	196
3.1.1. <i>Utilisation des fonctionnalités d'annotation</i>	197
3.1.2. <i>Validation des scénarios</i>	198
3.1.3. <i>L'utilité de l'outil ou utilité des fonctionnalités dans un usage d'ingénierie mécanique concourante</i>	201
4. Conclusion	204
Conclusion et perspectives	207
1. Conclusion générale	207
2. Limites et perspectives	209
Bibliographie	213
Annexes	229
Index	II
Liste des acronymes et abréviations	II
Glossaire	III
La pondération en RI	IV
Outils d'annotation	VI
Corpus	XIII
Extrait de Topic Maps du projet	XXXII
AnT&CoW	XXXIII
Questionnaire d'évaluation d'AnT&CoW	XXXIV
Liste des publications au cours de la thèse	XLI

Liste des figures

FIG. 1.	DÉMARCHE DE CONCEPTION D'UN COLLECTICIEL (LEWKOWICZ ET AL., 2006).....	18
FIG. 2.	SPECTRE DE L'ONTOLOGIE.....	39
FIG. 3.	LE PROTOCOLE ANNOTEA.....	43
FIG. 4.	SCHÉMA D'UNE ANNOTATION SELON ANNOTEA-RDF.....	44
FIG. 5.	PAGE D'ACCUEIL DE XWIKI.....	63
FIG. 6.	POPULARITÉ DES ÉTIQUETTES DE DEL.ICIO.US.....	78
FIG. 7.	DÉFINITION DE L'ANNOTATION EN UN CONTINUUM.....	89
FIG. 8.	MODÈLE DE PROCESSUS D'ÉCRITURE DE HAYES ET FLOWER (1980).....	98
FIG. 9.	MODÈLE DE PRODUCTION VERBALE ORALE, (LEVELT, 1989).....	100
FIG. 10.	MODÈLE DE PRODUCTION ÉCRITE SELON VAN WIJK (1999) IN (PIOLAT, 2004).....	101
FIG. 11.	PHASES PRINCIPALES DE L'ACTIVITÉ DE PRODUCTION DE DISCOURS.....	108
FIG. 12.	DIAGRAMME D'ACTIVITÉ DU MODÈLE DE PRODUCTION DE DISCOURS.....	111
FIG. 13.	DIAGRAMME D'ACTIVITÉ DU MODÈLE DE PRODUCTION DE DISCOURS INSTRUMENTÉE.....	114
FIG. 14.	DIAGRAMME DES CLASSES GÉRÉES PAR L'OUTIL.....	115
FIG. 15.	PAQUETAGES DES CAS D'UTILISATION DE L'OUTIL.....	125
FIG. 16.	ARCHITECTURE D'ANT&COW.....	127
FIG. 17.	PROTOCOLE DE COMMUNICATION D'ANT&COW.....	128
FIG. 18.	MODULE ZTAL.....	129
FIG. 19.	LE PLUG-IN D'ANNOTATION.....	131
FIG. 20.	ACCUEIL SUR L'ESPACE PARTAGÉ.....	132
FIG. 21.	BIBLIOTHÈQUE DE L'ESPACE PARTAGÉ.....	132
FIG. 22.	CHOIX DU FICHER A TELECHARGER.....	133
FIG. 23.	BIBLIOTHÈQUE DE DOCUMENTS PARTAGÉS.....	133
FIG. 24.	ÉDITION D'UN DOCUMENT.....	134
FIG. 25.	ANCRAGE D'UNE ANNOTATION.....	136
FIG. 26.	MULTI-ANCRAGE.....	137
FIG. 27.	FONCTIONNALITÉ DU PLUG-IN : AFFICHER, CRÉER, RECHERHER, GÉNÉRER UN BROUILLON.....	138
FIG. 28.	SAISIE D'UNE ANNOTATION.....	139
FIG. 29.	MISE À JOUR D'UNE ANNOTATION.....	139
FIG. 30.	FIL D'ANNOTATIONS.....	140
FIG. 31.	COLLECTION D'ANNOTATION POUR L'ÉLABORATION D'UN BROUILLON.....	142
FIG. 32.	SÉLECTION SELON UN DESCRIPTEUR.....	142

FIG. 33.	ÉDITION D'UN BROUILLON	143
FIG. 34.	SAUVEGARDE D'UN BROUILLON, CRÉATION D'UN DOCUMENT PARTAGÉ.....	144
FIG. 35.	EXEMPLE DE SORTIE DE SYNTAX SUR NOTRE CORPUS	155
FIG. 36.	FORMULE DE FRÉQUENCE	157
FIG. 37.	PATRONS DE RELATIONS CONCEPTUELLES DE (SÉGUÉLA ET AUSSENAC, 1999).....	158
FIG. 38.	PATRONS LEXICO-SYNTAXIQUES POUR L'HYPONYMIE DE (HEARST, 1992).....	159
FIG. 39.	PATRONS DE (FUCHS, 1982)	162
FIG. 40.	DÉCLARATIONS EN TM	164
FIG. 41.	ASSOCIATIONS SELON DES POINTS DE VUE EN TM	166
FIG. 42.	INTERFACE D'ÉDITON DE LA CARTE DE THÈMES EN TM	171
FIG. 43.	PROPOSITION ET ENREGISTREMENT D'INDEX	172
FIG. 44.	LISTE DÉROULANTE DES DESCRIPTEURS ARGUMENTATIFS.....	172
FIG. 45.	FRAGMENTS DE CONVERSATION.....	182
FIG. 46.	DES FRAGMENTS D'ÉLABORATION DE DOCUMENT	184
FIG. 47.	SUJETS DES ÉCHANGES ORGANISÉS PAR POINTS DE VUE SUR LE PROJET.....	186
FIG. 48.	ÉCHANGE ENTRE PÉPITO ET LÉON SUR BRIDE ET RÉDUCTEUR	189
FIG. 49.	LOGIQUE DE COMMUNICATION	191
FIG. 50.	DOCUMENT CRÉÉ SUR LA BASE D'UN RECUEIL D'ANNOTATION	192
FIG. 51.	ANNOTATION D'UN DOCUMENT DU PROJET	192
FIG. 52.	RECUEIL DE SOLUTIONS POUR LA PLANIFICATION.....	193
FIG. 53.	VISUALISATION DU POINT DE VUE ARGUMENTATION ET DE SES CONCEPTS.....	195
FIG. 54.	VISUALISATION DU THÈME RÉGLAGE ET DE SES OCCURRENCES.....	195
FIG. 55.	VISUALISATION D'ASSOCIATIONS DÉCOUVERTES DANS LE CORPUS	196
FIG. 56.	CAPTURE D'ÉCRAN D'AMAYA.....	VIII
FIG. 57.	INTERFACE DE JANNO.....	IX
FIG. 58.	INTERFACE DE COMPENDIUM.....	XIII

Liste des tableaux

TAB. 1.	PROPRIÉTÉS DU SCÉNARIO NÉCESSAIRE À UNE ACTIVITÉ (CARROLL, 1999).....	25
TAB. 2.	GRILLE DE CLASSEMENT DES OUTILS D'ANNOTATION SÉMANTIQUE.....	47
TAB. 3.	GRILLE DE CLASSEMENT DES OUTILS D'ANNOTATION DISCURSIVE	53
TAB. 4.	GRILLE DE CLASSEMENT DES OUTILS DE RÉDACTION COLLABORATIVE.....	65
TAB. 5.	TABLEAU COMPARATIF DES OUTILS D'ANNOTATION EXISTANTS	69
TAB. 6.	TABLEAU COMPARATIF DES OUTILS EXISTANTS POUR LE CSCWRITING.....	69
TAB. 7.	SPÉCIFICATIONS GÉNÉRALES D'UN OUTIL D'ANNOTATION POUR LE TRAVAIL COOPÉRATIF	123
TAB. 8.	ANALYSE DIFFÉRENTIELLE ANNOZILLA-ZANNOT / ANT&COW	124
TAB. 9.	RELATIONS UTILISÉES POUR L'EXTRACTION DE TERMES.....	160
TAB. 10.	PATRONS D'EXTRACTION ET STRUCTURATION EN TM	167

À Pifiou,

Remerciements

Je remercie avant tout mes co-directrices, Myriam Lewkowicz et Amalia Todirascu-Courtier de la régularité et du sérieux de leur encadrement et de leur résistance exceptionnelle au travail de nuit et dans l'urgence.

Je remercie Jean Charlet et Gérard Sabah de me faire l'honneur de leurs annotations et de leur présence au sein du jury. De même, je remercie vivement Sylvie Calabretto d'avoir accepté de faire partie du jury de cette thèse. Merci aussi à Michaël Zock d'avoir accepté mon invitation.

Merci aussi à Nada Matta et Manuel Zacklad de m'avoir permis de travailler dans de bonnes conditions scientifiques et matérielles, aux mécaniciens du LASMIS et de la Halle Industrielle pour leur temps et leur gentillesse ; à Didier Bourigault pour le prêt de l'outil Syntex.

Remerciements chatoyants à Nadia, Michel, Aurélien, Christophe, Hassan, pour, respectivement, le divan, la bonne humeur et les coups de gueule, les coups de pouce, la Science humaine et les vanes XL ; à tous les autres thésards, Anne-Laure, Sacha, Anne-Sophie, Benjamin, Denis, Jean-Sébastien, L'Hédi, Chao, Valérie, Oswaldo, Bertrand, Xavier, pour leur *sympathie* ; à Foued, Simon, Fabien, Maxime, Cyril, Alison, et Fabien, pour leur participation au projet ; Special thanks to Hung !

Et à tous ceux qui ont gravité autour de cette thèse... Merci !

... avant tout à la constance, la patience, la bienveillance... l'héroïsme ? dont tu fais preuve, Rolando ;

... à la longanimité de Véronique, Isabelle, Anne-Marie et Raymond, aux sourires de Luna ;

... au soutien permanent de Michael, Brigitte et Anne ;

... à l'amitié indéfectible de François, Catherine, Léna, Sandra, Sandrine et Coraline ;

... à tous ceux dont la pensée éclaire mes jours, Sandrine, Dominique, Éric, Mylène, Alain, Séwan, Florent, Marc, Sandra, Véronique, Séverine, Guillaume, Xavier, Chantal, ...

Introduction

Trop de lecture peut étouffer le génie
Jean le Rond d'Alembert

Le texte ne peut être séparé d'une chaîne interprétative qui le précède, le constitue et l'actualise. Cette chaîne est réalisée par la communauté, qui élabore, qui lit, qui relie. Cette activité d'interprétation des textes, ou herméneutique, est à repenser dans une nouvelle dimension, celle de l'avènement du monde numérique. En effet, le contexte actuel de numérisation des documents a mis l'accent sur le partage du plus grand nombre de sources possibles. Les mécanismes de compréhension, d'interprétation et de production de textes, pourtant à l'origine du fondement et de l'évolution des communautés de pensée se retrouvent négligés. C'est ce constat du manque de soutien à la lecture critique qui est à l'origine de nos travaux sur ce que nous appellerons ici une herméneutique numérique.

Le travail de recherche que nous présentons s'intègre, suivant (Pédauque, 2003, p. 3), dans une problématique qui considère le document comme « porteur de sens et doté d'une intentionnalité [...] indissociable du sujet en contexte qui le construit ou le reconstruit et lui donne sens ». Dans cette vision du document comme signe, nous portons plus particulièrement notre intérêt sur la création et l'interprétation d'un document, c'est-à-dire sur les signes qui le constituent. Nous abordons ces questions sous l'angle de la lecture critique des documents, que nous opposons à une lecture qui ne serait guidée par aucun principe productif, qui ne viserait ni un savoir, ni la production d'un autre texte. Dans ce cadre, le document représente un support classique de partage de la connaissance au sein d'un groupe, par exemple un groupe projet, constitué afin d'effectuer un travail précis.

Notre propos se fonde sur un constat dans le contexte du travail en conception mécanique : de nombreux projets sont effectués par des groupes distribués sur plusieurs sites, qui partagent un grand nombre de documents numériques au centre des processus de travail. Dans ce cadre, plusieurs études ont montré que les phases de travail asynchrones peuvent être soutenues par l'usage d'annotations sur les documents (Darses, 2001, Martin *et al.*, 2001, ou encore le projet Mediannote - appel d'offre CNRS/ PI-TCAN - : Zacklad *et al.*, 2003). Ces annotations sont des commentaires déposés sur un document par les membres du projet à des fins de communication. Elles permettent de négocier le sens d'un terme ou d'un document, d'élaborer de nouveaux concepts en laissant des commentaires sur un document. Elles ont un rôle important dans le suivi des interactions entre les membres : elles

sont les traces d'une décision, d'une négociation, ou même d'une solution apportée à un problème ; elles jouent un rôle central dans le discours produit par les membres du groupe. Certaines informations contenues dans les annotations pourront être intégrées dans de nouveaux documents. Ces documents et fragments de documents peuvent être qualifiés de DoPAs (DOcuments Pour l'Action)¹ (Zacklad, 2004). Le concept de DoPA permet de considérer le document en évolution, qu'il soit produit ou utilisé par des situations d'activité collective. Dans cette approche, un ensemble de fragments documentaires se constitue en un document au cours d'un processus de documentarisation. Dans une activité collective, un fragment est une production sémiotique qui peut être en élaboration, ou finalisée (quoique tout document s'inscrit potentiellement dans un inachèvement prolongé) et qui intervient par exemple dans un projet au sein d'un groupe de co-auteurs. Les différents auteurs de ces DoPAs ne pourront participer efficacement à leur élaboration que si les fragments de documents sont conservés et organisés dès la première lecture et pour toute version. De plus, l'explicitation offerte par une communication écrite est réalisée en donnant aux messages une structure reconnue publiquement et en les transformant en artefact partagé, ce qui est déjà en soi un moyen de favoriser la conscience mutuelle (Goody, 1987). Le partage ainsi permis participera à la formation d'un référentiel commun (*common ground*, Clark, 1996, Clark et Brennan, 1991) et facilitera donc la conscience des activités réalisées par les autres participants (*activity awareness*, Dourish et Belotti, 1992, Carroll *et al.*, 2003).

Pour illustrer ce besoin de création et de gestion de DoPAs, nous avons observé les pratiques documentaires d'une équipe de concepteurs en mécanique. Dans le cadre de leur activité, les concepteurs sont amenés à échanger nombre de documents (plan, maquette, spécifications de fabrication, etc.) et d'opinions (vérification, opposition d'interprétation de résultats, etc.) afin de déterminer les spécifications d'une pièce (la courroie d'un moteur par exemple) et du produit (le moteur). Pour soutenir ces pratiques, il faut donc rendre possible la gestion des DoPAs par le groupe projet : visualisation des documents de travail, annotation de ces documents, gestion des modifications, discussions entre les membres du groupe par le biais d'annotations, réutilisation des annotations, création de nouveaux documents sur la base des réflexions au travers des annotations, etc.

Dans ce contexte, la documentation doit traiter des documents atypiques. En effet, les documentalistes considèrent en général le document comme un enregistrement textuel ou

¹ Les acronymes et abréviations sont listés en annexe.

pour le moins comme un enregistrement *plat* (Ranganathan cité dans (Buckland, 1998a)). La combinaison d'un médium, d'un message et d'une signification est enregistrée telle que, contenant son contexte via un ensemble de métadonnées ou de mots-clés. Mais les technologies numériques ont remis au goût du jour la sempiternelle question de « Qu'est-ce qu'un document ? ». Dans un contexte numérique, le principe du médium semble beaucoup moins satisfaisant que celle d'une approche fonctionnelle. Dans une visée plus sémiotique de la documentation, il est possible de définir le document comme tout élément consigné en tant que *vecteur de sens* (*vehicle of meaning* selon Buckland, 1998b), c'est-à-dire tout objet archivé en tant qu'il est significatif dans un certain contexte. Il reprend alors des considérations de Otlet et Briet. Briet considère qu'est document « Tout indice concret ou symbolique, conservé ou enregistré, aux fins de représenter, de reconstituer ou de prouver un phénomène ou physique ou intellectuel » (Briet cité dans (Buckland, 1998a)). Elle prend pour exemple un animal en liberté versus un animal dans un zoo. L'animal en liberté n'est pas un document, mais l'animal enfermé dans l'optique d'exemplifier tous les animaux de ce type en devient un. Les documents sont des objets qui acquièrent leur statut documentarisé du moment qu'ils sont mis dans une relation organisée et significative avec d'autres objets (Buckland, 1998b). Un document numérique peut donc être vu comme une représentation documentarisée d'un objet du monde, un texte, un objet physique ou encore un objet audio ou vidéo.

Les travaux au sein du Réseau Thématique Pluridisciplinaire 33 « Documents et contenu : création, indexation, navigation » (RTP-DOC) ont permis une réflexion nouvelle et consensuelle sur le document. Selon ce groupe, le document est utilisé dans trois contextes de médiatisation : privé, collectif ou public (Pédauque, 2006). Dans un contexte collectif, les documents sont utilisés pour structurer un groupe, son identité, son activité, sa cohésion. Toujours dans la définition collective de R.T. Pédauque (Pédauque, 2006), le document est présenté comme un médium qui possède quatre propriétés principales. Le document permet (1) la mémorisation, (2) l'organisation, (3) la création et (4) la transmission. La mémorisation et l'organisation sont les propriétés illustrées dans notre projet par les *proto-documents* utilisés pour organiser les idées dans un document. Une annotation est donc une sorte de proto-document qui soutient non seulement la mémorisation mais aussi la transmission d'idées ou de révisions à opérer. Les annotations peuvent aussi être colligées et former un document puisqu'elles peuvent représenter les parties d'un document en cours d'organisation. Les annotations ainsi regroupées en un nouveau document acquièrent une

signification nouvelle par le renouveau (la remotivation) de leur sens dans un nouveau cotexte².

Ces pratiques annotatives ont été étudiées dans des recherches sur l'élaboration de texte (Weng et Gennari, 2004), la communication dans une communauté (Bringay, 2006, Bringay *et al.*, 2006) et l'indexation de document (Dzbor, 2004). Forte de l'étude de ces travaux sur le soutien aux activités collectives distribuées, nous souhaitons fournir un outil qui permettrait aux utilisateurs d'annoter des documents, de stocker ces annotations ainsi que de les retrouver afin de comprendre la logique de conception du document. Cet outil gèrera ainsi la connaissance en action, ou *knowing*, par opposition au *knowledge*, comme différencié par (Cook, 1999) et (Pfeffer, 1999). Il s'insèrera dans le courant initié par (Bannon et Kuuti, 1996) de mémoire organisationnelle constructive (en complément d'une mémoire organisationnelle passive ou active) où les outils ont pour but d'aider à une construction d'espaces communs d'information en prenant en compte les aspects sociologique et psychologique des organisations.

Nos travaux de recherche menés dans un objectif de médiatisation de l'annotation pour la coopération autour de documents nous ont incitée à tenter de définir l'annotation, après une revue des définitions proposées par différentes disciplines. Ces définitions portent sur l'annotation en tant qu'objet ou sur l'activité d'annotation. Dans des travaux liés à l'herméneutique (De Libera, 2000) ou à la rédaction assistée par ordinateur (CSCWriting en anglais) (Weng et Gennari, 2004), l'annotation est envisagée comme la production de fragments de documents liés à un document d'origine et aidant à expliquer ce document, à lui donner un sens. Nous avons ici affaire à l'annotation telle que définie aussi par S. Bringay : « Une annotation est une note particulière liée à une cible. La cible peut être une collection de documents, un document, un segment du document (un paragraphe, un groupe de mots, une image, une partie d'image, etc.) ou une annotation. Chaque annotation a un contenu, matérialisé par une inscription. C'est la trace de la représentation mentale élaborée par l'annotateur au sujet de la cible. Le contenu de l'annotation peut être interprété par un autre lecteur. L'ancre lie l'annotation à la cible (une flèche, une phrase entourée, etc.). » (Bringay, 2006, p. 67). M. Veron considère que l'annotation possède en plus de son moyen annotatif (une forme – un surlignement, une croix, ...-), un lieu, un auteur, une histoire, un support et surtout un but (mettre en valeur pour un surlignement) (Veron, 1998). En revanche, les

² Ce terme est explicité dans le glossaire en annexe.

travaux sur les langages de représentation, comme ceux du Web Sémantique par exemple (Berners-Lee *et al.*, 2001), définissent l'annotation comme un index montrant le chemin vers l'information et permettant son accès. Les recherches menées sur les annotations sont généralement focalisées sur l'une ou l'autre de ces orientations. Nous adopterons ici un point de vue focalisé sur le document en tant que construction, sur la connaissance élaborée grâce aux argumentations et aux échanges produits par des utilisateurs autour d'un document. Nous considérons ainsi l'annotation non seulement comme un objet, mais aussi comme un processus, une activité qui permet la communication de connaissances et d'information dans un contexte de travail distribué et médiatisé autour de documents. Afin de comprendre cette élaboration de connaissance et de document sur la base d'annotations échangées, il est nécessaire d'indexer finement les fragments produits par les utilisateurs. Cette tâche étant lourde, nous souhaitons la soutenir par l'utilisation d'outils de Traitement Automatique des Langues (TAL). Il s'agit alors d'utiliser ces outils de TAL dans une optique de construction de classification et de soutien de l'utilisateur dans sa tâche d'indexation.

Ce mémoire est structuré de la manière suivante :

Dans la première partie consacrée au positionnement et à la méthodologie, nous établissons dans un premier temps l'annotation en tant qu'élément de négociation et d'organisation de concepts partagés en conformité avec des concepts de TCAO (Travail Coopératif Assisté par Ordinateur). Nous exposons ensuite notre démarche de conception d'outil fondée sur (1) la modélisation de l'activité d'annotation, (2) la construction d'outils d'aide à l'utilisateur selon un corpus représentatif du sociolecte du groupe observé et (3) la validation des fonctionnalités de l'outil selon des scénarios pertinents.

Dans la deuxième partie, nous présentons tout d'abord, dans le chapitre I, un éventail d'emploi du terme annotation selon des perspectives herméneutique, communicationnelle, indexante et coopérative. Ces différentes utilisations relèvent de différentes définitions de l'objet annotation. Ce panel nous permettra de proposer une classification d'outils d'annotation existants obéissant à des objectifs d'indexation, de communication et d'élaboration de documents. Dans le chapitre II, nous présentons une définition unifiée de l'annotation qui permet de tracer la logique de communication d'un document. Le continuum d'annotation présenté part de la balise pour l'indexation et arrive au fragment de document pour la négociation. Nous considérons que nous avons affaire à un continuum dans la mesure où les frontières des différentes annotations sont floues et que leurs propriétés restent partagées.

La troisième partie introduit, en chapitre III, différents modèles issus principalement des sciences cognitives et présentant des activités de lecture ou d'écriture décrivant sous plusieurs angles l'activité d'annotation. Dans le quatrième chapitre, nous développons le modèle évalué comme le plus adéquat aux besoins d'annotation dans notre contexte. C'est un modèle descriptif issu de la rhétorique médiévale et adapté pour permettre la définition de l'activité d'annotation médiatisée.

Dans la quatrième partie, nous décrivons, dans le chapitre V, les fonctionnalités d'indexation, de communication et d'élaboration mises en place pour la médiatisation de l'activité d'annotation. Elles permettent le travail en coopération autour de documents numériques textuels. Dans le cadre du soutien de l'utilisateur dans son activité d'indexation, nous décrivons la mise en place et l'utilisation d'outils de TAL (Traitement Automatique des Langues) pour la construction d'une classification et son utilisation pour de l'indexation supervisée. Le sixième et dernier chapitre est consacré à la simulation de scénarios d'utilisation dans l'outil. Dans un premier temps, nous présentons le corpus utilisé comme ressource linguistique et les scénarios rejoués via l'outil et utilisés pour valider les fonctionnalités du collecticiel. La simulation est ainsi menée dans un contexte médiatisé de projet de conception mécanique.

Nous concluons enfin sur l'annotation et les problèmes soulevés par sa médiatisation puis nous exposons les perspectives de recherches envisagées.

1^{ère} partie : Positionnement et méthodologie

Ce que j'appelle cristallisation, c'est l'opération de l'esprit, qui tire de tout ce qui se présente la découverte que l'objet aimé a de nouvelles perfections.
Henri Beyle

Nous considérons l'annotation comme une activité qui produit des fragments textuels, et, par eux, qui permet d'échanger et de négocier le sens de documents mis en commun au sein d'un groupe. Ces fragments sont nécessaires pour contextualiser la compréhension d'un document, et partager cette représentation du sens d'un document. Par cette contextualisation partagée, les membres d'un groupe négocient des frontières de leur projet, ce qui leur permet de concevoir collectivement un produit ou de nouveaux concepts. Ces fragments, utilisés pour la négociation de sens et de frontières, sont des productions textuelles de différents auteurs selon différents points de vue. Les membres du groupe vont concevoir un document textuel, mais aussi les nouveaux concepts qui vont constituer ce document. Produire du texte, c'est aussi construire des idées en échangeant avec l'autre, comme au cours d'une discussion en face à face. C'est ce principe d'échange qui est mis en œuvre en herméneutique. Celle-ci consiste en l'élaboration de discussions critiques ou explicatives autour de textes pour recréer un sens autour d'un texte qui ne possède plus qu'intrinsèquement son contexte. Pas ou peu d'informations sur ses conditions de production subsistent hors de son corps lui-même. Il s'agit de redécouvrir un sens au texte par une succession d'énoncés expliquant une interprétation, se complétant, se répondant, construisant de nouvelles pistes d'interprétation. Ces échanges sont matérialisés par des annotations qui sont ancrées au texte d'origine, ou relient différents textes ou fragments de texte. Dans un contexte de projet, mémoriser ces annotations et les structurer peut être envisagé comme un moyen non seulement de capter la logique de conception du produit ou des concepts mais aussi d'observer l'organisation commune des connaissances que le groupe a produites au cours de ces échanges. Médiatiser l'activité d'annotation est donc à la fois un moyen de soutenir la discussion autour de textes élaborés ou en cours d'élaboration, et un moyen de concevoir un référentiel commun dans le groupe et de conserver sa logique de conception.

Dans cette partie, nous présentons le cadre théorique et la démarche suivie pour concevoir un collecticiel permettant la coopération via l'annotation de documents numériques.

Dans un premier temps, nous nous positionnons dans le champ du Travail Coopératif Assisté par Ordinateur par l'adoption de concepts tels que les objets-frontières, les schèmes de classification et les artefacts de négociation de frontières. Ces concepts permettent de situer l'annotation dans une perspective coopérative.

Suite à ce positionnement, nous définissons notre méthode de conception d'outil comme inspirée de l'Ingénierie de Connaissances, des Sciences Humaines et Sociales et de l'Interaction Homme-Machine. L'Ingénierie de Connaissances nous offre des principes de modélisation qui permettent le passage d'une vision anthropocentrée de l'outil à une vision technocentrée. Les Sciences Humaines et Sociales nous permettent d'ancrer dans un contexte pertinent les modèles définis par le recours au corpus, un recueil de données réelles. L'Interaction Homme-Machine nous fournit enfin un cadre pour l'évaluation à base de scénarios.

1. Le Travail Coopératif Assisté par Ordinateur, un cadre adapté

Le Travail Coopératif Assisté par Ordinateur (TCAO - traduction du terme anglais CSCW) se propose d'étudier les mécanismes de coopération à l'aide des théories existantes en Sciences Humaines et Sociales, de proposer des outils supports à ces mécanismes et d'étudier leurs impacts. Le TCAO voit le jour en 1984 par l'invention du terme Computer Supported Cooperative Work par I. Greif et P. Cashman (1988, p. 5). Il est défini comme :

« a computer-assisted coordinated activity such as problem solving and communication carried out by a group of collaborating individuals.³ »

Cette définition est reprise par (Schmidt et Bannon, 1992, p. 3) :

« CSCW should be conceived as an endeavor to understand the nature and characteristics of cooperative work with the objective of designing adequate computer-based technologies for cooperative work arrangements.⁴ »

Le collecticiel, traduction de *groupware* est un outil d'aide au travail collectif. Ce terme désigne les logiciels informatiques plus que les processus mis en oeuvre par le travail coopératif. (Ellis *et al.*, 1991, p. 40) définissent le collecticiel ainsi :

« Computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment.⁵ »

Les technologies coopératives en TCAO sont le plus souvent classées selon une matrice espace-temps définie par (Johansen, 1988). Dans un cadre de médiatisation de la coopération Homme-Homme par la machine, nous nous plaçons résolument dans une optique où les collecticiels sont des outils qui « facilitent, améliorent, augmentent la communication Homme-Homme sans perturber l'activité des utilisateurs » (Chibout *et al.*,

³ Une activité coordonnée, assistée par ordinateur telle que la résolution de problème ou la communication lorsqu'elles sont menées par un groupe d'individus collaborant. (Traduction personnelle)

⁴ Le TCAO doit être vu comme une tentative de compréhension de la nature et des caractéristiques du travail coopératif afin de concevoir des technologies informatiques adaptées aux dispositifs de travail coopératif. (Traduction personnelle)

⁵ Systèmes informatiques soutenant des groupes de personnes engagés dans une tâche (ou un objectif) commun et qui fournissent une interface à un environnement partagé. (Traduction personnelle)

2003, p. 2). Nous considérons à la suite de ((Dejours, 1995) cité dans (Chibout *et al.*, 2003, p. 13)) que la « coopération est l'une des activités qui échappe à la procéduralisation complète et qui relève en propre de la contribution des acteurs à l'organisation du travail ». Il est donc primordial de proposer aux utilisateurs de collecticiels des fonctionnalités de communication fondées sur le langage naturel spontané. (Lacoste, 1991) envisage la communication orale comme une dimension essentielle dans les activités de travail. Elle souligne l'aspect réducteur des langages spécialisés pour l'écrit. Dans une de ses études dans le milieu médical, elle constate que l'oral permet aux collaborateurs d'introduire des éléments de négociation et apporte des éléments contextuels en explicitant l'écrit, même si l'écrit prédomine dans l'activité (Dossiers médicaux par exemple). Aujourd'hui nombre d'outils autorisent un écrit oralisé pour permettre la communication (par exemple, les blogs, les Wikis, les chats, (Delacroix, 2005)).

Le collecticiel issu de nos recherches (présenté en 4^e partie) propose d'assister la communication via de l'édition et de la négociation de texte. Cette dernière activité, la négociation, permet au groupe de s'accorder sur les problèmes de coordination et de décision. L'édition permet de partager un objet commun pour la discussion et ainsi de communiquer à propos de cet objet.

Afin de concevoir ce collecticiel, nous nous penchons sur les éléments théoriques pouvant orienter nos choix. En effet, le TCAO s'appuie sur des concepts clés venus de différentes disciplines. Nous intéressent particulièrement, le concept d'*objet-frontière* (Boundary Object), le concept de *schème de classification* (Classification Scheme) ainsi que celui d'*artefacts de négociation de frontières* (Boundary Negotiation Artifact) qui offrent un cadre théorique intéressant dans l'optique de l'analyse de l'annotation en tant qu'entité porteuse d'une argumentation et permettant de négocier une interprétation partagée dans une communauté.

1.1. Les objets-frontières

La notion d'objet frontière est cruciale pour notre propos. Elle a été développée dans le cadre de la théorie des acteurs-réseaux (TAR, en anglais ANT - Actor Network Theory-) de S. Star (1989) et reprise par E. Wenger (1998, Wenger *et al.*, 2002) dans sa réflexion sur les communautés de pratique. Les communautés de pratique sont des groupes de personnes qui partagent une préoccupation, un ensemble de problèmes, ou une passion à propos d'un

sujet, et qui approfondissent leurs connaissances et expertise dans ce domaine en interagissant de manière régulière⁶. Dans un contexte de conception mécanique médiatisée, différentes communautés de pratique cohabitent et effectuent diverses tâches dans un même objectif de finalisation d'un produit.

Dans la TAR, les objets-frontières circulent au travers des réseaux en jouant différents rôles selon différentes situations. Ils ne sont pas forcément tangibles et peuvent être des « trucs et des choses, des outils, des artefacts et des techniques, des idées, des histoires et des mémoires » ((Clarke et Fujimura, 1992) cité par (Bowker et Star, 2000, p. 298)). Pour (Wenger, 1998), les objets-frontières se trouvent à la limite de communautés de pratique et permettent de créer des relations extérieures à la communauté. Ce type d'objet permet la coordination entre deux communautés, sans pour autant faire un pont que celles-ci traverseraient. En effet, selon E. Wenger, ils ne produisent pas de modifications des significations ou des perspectives inhérentes à chacune des communautés. Les objets-frontières sont interprétés différemment selon les communautés de référence des interlocuteurs. La connaissance et les discussions autour de ces objets-frontières permettent d'édifier une compréhension partagée entre ces communautés de pratique et de les instituer en une communauté de partage. Dans le cadre de notre projet, nous allons nous intéresser aux moyens de communication mis en place entre les communautés, principalement les artefacts qui médiatisent la communication entre les membres des communautés.

Suivant M. Ackerman et C. Halverson (1999), les documents et collections de documents dans des contextes de coopération peuvent être considérés comme des objets frontières (au sens de S. Star (1989)) permettant au groupe de structurer un contexte propre à ses membres et à ses objectifs. Les objets-frontières font partie des mécanismes d'articulation du travail dans un projet déjà organisé, et sont considérés comme relativement statiques. La collection de documents numériques est un élément partagé permettant d'ancrer des tâches distribuées dans un référentiel commun, autorisant ainsi une médiation de ce travail. Le groupe de conception observé possède un grand nombre de documents, base de communication entre les différents membres du projet. Ces documents sont peu organisés et ne répondent pas à une classification explicitée.

⁶ « Communities of practice are groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis » (Wenger *et al.*, 2002, p. 4)

1.2. Les schèmes de classification

C. Simone et M. Sarini (2001) considèrent également les schèmes de classification (Classification Schemes), comme des objets frontières définis au cours d'une construction collective et permettant de structurer les connaissances d'un groupe selon une organisation commune. Intégrer un document dans une classification physique (un thésaurus par exemple), c'est aussi l'intégrer à ses concepts individuels selon un schème de classification qui peut être partagé. Cette intégration de documents dans un schème individuel revient à comprendre le document par rapport à son contexte. C. Simone et M. Sarini (2001) soulignent que, malgré l'importance des schèmes de classification dans les mécanismes de coordination, ceux-ci ne sont pas suffisamment observés sur le plan de leurs évolutions et mises à jour. Cette évolution serait appréhensible en parvenant à tracer les négociations menées dans les phases de construction des schèmes de classification d'un groupe. En structurant l'ensemble des échanges du groupe, nous pensons qu'il serait possible de comprendre les négociations et argumentations intervenues dans la construction d'éléments ou produits du projet. Cela reviendrait alors à tracer la *logique de communication* d'un projet ou d'un document (définie chp. II section 4).

1.3. Les artefacts de négociation de frontières

Les activités de conception donnent naissance à des artefacts nouveaux, émergeant dans des contextes organisationnels qui ne sont pas formellement établis. Les artefacts et le groupe qui collabore à leur conception se co-construisent au fil de l'activité. Dans ces contextes de coopérations émergentes, non encore établies, nous nous penchons avec intérêt sur le concept d'artefact de négociation de frontières (*Boundary Negotiation Artifact* (Lee, 2005)) utilisé pour noter, organiser, explorer, partager des idées, introduire de nouveaux concepts et techniques, créer des alliances ou encore créer une compréhension partagée de certains problèmes (Lee, 2005). Ces artefacts de négociation de frontières ont un rôle comparable à celui des objets intermédiaires (Boujut et Blanco, 2003) ou des prototypes définis en ingénierie de la conception (Subrahmanian *et al.*, 2005).

Dans des phases de réflexion dans l'action (*in action* (Schön et Bennet, 1996)), les concepteurs conçoivent un objet par une création continue en s'écoutant et en répondant aux surprises de la conversation. Les artefacts de négociation de frontières sont aussi construits par des échanges entre les participants d'une ou plusieurs communautés au cours d'une

activité collective et vont jusqu'à repousser certaines frontières entre les communautés. En participant à la construction négociée d'un référentiel commun au groupe (Clark, 1996) (une classification commune par exemple), un acteur manifeste son engagement dans la volonté de construire une communauté. Il participe ensuite à la création de documents et à leur classement selon le plan de classement défini. Les échanges permettent l'élaboration collective de schèmes de classification, mais aussi de concepts et de documents. Le document issu des échanges médiatisés devient un élément de communication, à la fois base de discussion et d'élaboration de nouveaux concepts et reflet de la *logique de communication*.

Les documents, comme les schèmes de classification, évoluent et sont élaborés au fur et à mesure de l'apparition de nouveaux objectifs, de nouveaux membres du projet, de nouveaux documents, ... Ce sont des documents en élaboration constante, en action, qui impliquent des mouvements endogènes - les documents donnent naissance à de nouveaux concepts qui s'intègrent dans ce même document au cours de l'élaboration des différentes versions - et exogènes - les documents font émerger de nouvelles classifications et documents possibles, ou de nouvelles actions, reliés au contenu des premiers mais extérieurs à ce contenu tout de même -. Ce sont des DOcuments Pour l'Action (DoPA) (Zacklad, 2006). On peut soutenir cette génération de DoPAs par le biais d'annotations suivant ces deux grands mouvements (endogène et exogène) que l'on retrouve dans le principe du cercle herméneutique (Gadamer, 1960) et qui permettent une interprétation des textes toujours renouvelée.

Afin de concevoir un collectif adapté au travail médiatisé d'un groupe projet en conception, nous souhaitons favoriser l'utilisation d'annotations en tant qu'artefacts de négociation de frontières qui soutiendront habilement la communication. Nous allons maintenant exposer les principes de conception que nous estimons les plus appropriés dans ce contexte.

2. Méthodologie de conception d'outil

Le recueil de données pour comprendre une activité se fait de différentes manières selon la discipline qui se propose d'analyser les informations. L'Ingénierie des Connaissances (IC) est une discipline modélisatrice qui consiste à concevoir des systèmes dont le fonctionnement permet d'opérationnaliser des connaissances portant sur le traitement

ou la résolution d'un problème donné. J. Charlet, C. Reynaud et R. Teulier (2001) définissent l'IC comme proposant des modélisations conceptuelles du monde plus ou moins formelles. Pour élaborer ces représentations, l'IC se fonde sur des observations d'activités avec des méthodes issues de psychologie ou d'ergonomie cognitive, des sciences de gestion, ... au travers, par exemple, du recueil de formulations linguistiques du problème, de transcriptions d'interviews d'experts, de descriptions techniques, de notices de maintenance, etc. (Charlet *et al.*, 2001). Nous pouvons considérer trois instruments primordiaux à la conception : le corpus, le scénario et le modèle. Le modèle est un outil de représentation a priori d'une activité qui nécessite selon nous d'être complété par une approche de recueil de données existantes structurées en corpus. Dans le cadre d'une conception itérative, une première phase d'évaluation peut être menée sur la base de scénarios.

2.1. Les modèles

La situation de recherche dans laquelle nous évoluons nous amène à définir de nouvelles pratiques à assister, comme l'interprétation collective de documents numériques. Cette pratique peut être analysée à l'aide de concepts issus des SHS ou des Sciences et Technologies de l'Information et de la Communication. De ce fait, un processus de conception classique en informatique fondé sur une analyse des besoins, ou sur une analyse de l'activité existante, pour en déduire des primitives de conception, n'est pas adapté. « La recherche en informatique a donc ici un rôle fondamental, celui d'inventer de nouveaux possibles » (Tchounikine, 2002b, p. 207).

Dans ce contexte, nous proposons de recourir aux théories qui permettent de fournir une description, d'expliquer, de donner une idée générale du comportement d'un utilisateur, de fournir un cadre d'analyse (c'est-à-dire de fournir un outil de haut niveau conceptuel pour identifier des problèmes et modéliser certains types d'interaction), de fournir un langage commun (un ensemble de concept simple à utiliser pour discuter d'une conception), ou encore de générer des dimensions de conception et des représentations interactives. B. Shneiderman (2002 cité dans Rogers, 2004) souligne qu'il existe au moins cinq types de théories utilisables dans le cadre de la définition de systèmes d'Interaction Homme-Machine :

- descriptive : au sens où la théorie fournit des concepts, qu'elle clarifie une terminologie et guide des recherches à venir ;
- explicative : au sens où la théorie fournit des relations et des processus explicatifs

- prédictive : c'est à dire qui permet de prévoir des performances des utilisateurs
- prescriptive : qui fournit un guide pour la conception
- générative : au sens où la théorie permet aux domaines appliqués de créer.

Ces théories peuvent permettre de fournir des « intuitions » pour produire des modèles de l'activité qui peuvent être utilisés à différentes fins :

- modèle informatif : fournit des résultats de recherche utiles ;
- modèle prédictif : fournit des outils pour modéliser les comportements d'utilisateurs ;
- modèle prescriptif : fournit des indications et des conseils sur la façon de concevoir ou d'évaluer un système.

Le modèle est un objet interdisciplinaire. Dans cette pluridisciplinarité d'utilisation, J - L. Le Moigne (2004) reconnaît deux options épistémologiques de validation du modèle dues à une conception « ontologique » (donnée) ou une conception « phénoménologique » (construite) du modèle. La première conception est légitimée par « la *réalité* présumée objectivement observable ou expérimentable de ces faits » (Le Moigne, 2004, p. 3) (réalité présumée indépendante de l'observateur qui les décrit) et la seconde par « la modélisation artificielle puisque le modèle n'est plus tenu pour donné par la réalité, mais construit intentionnellement, à partir de l'un des points de vue possibles du phénomène, à l'aide d'artefacts symboliques, par un modélisateur » conscient (Le Moigne, 2004, p. 3). Dans la mesure où l'on peut aussi bien tenir pour construits des modèles de *faits* initialement tenus pour *donnés*, J - L. Le Moigne (Le Moigne, 2004) se penche sur le glissement réducteur du terme *modèle* au terme *modélisation* depuis ces vingt dernières années où *modélisation* représente plutôt un patron taillé pour de la *programmation à fin de simulation* qui permet de rendre pensable l'interdisciplinarité. Dans le champ de la conception des Environnements Informatiques pour l'Apprentissage Humain (EIAH), M. Baker (2000) et P. Tchounikine (2002a) mènent une réflexion où ils distinguent les modèles comme outil scientifique, utilisés pour comprendre ou prédire une situation, des modèles pour la conception de systèmes, qui traduisent les premiers en modèle permettant l'implémentation de systèmes supports à l'activité.

(Lewkowicz et al., 2006) font remarquer que « les théories issues des Sciences Humaines habituellement mobilisées lors de la conception de collecticiels (théorie de l'activité, théorie de l'apprentissage, théorie de l'agir communicationnel, ...) sont difficiles à exploiter telles quelles pour en déduire des primitives de conception » (p. 297). De même, il

est difficile de « transposer leurs définitions dans un cadre médiatisé par un système informatique » (p. 297). Le travail de conception consiste alors à mettre en place de nouveaux modèles, définissant les éléments du système, suivant une théorie qui décrit l'activité existante. La théorie permet ensuite d'analyser les traces ainsi mémorisées.

Nous suivons ce positionnement méthodologique et adoptons la démarche suivante, illustrée en figure 1 (Lewkowicz *et al.*, 2006). Ainsi, dans le cadre d'une théorie en Sciences Humaines adaptée aux phénomènes que l'on souhaite assister/observer, nous proposons un modèle de description de l'annotation en tant que production de discours. Ce modèle permet d'opérationnaliser la théorie et sert de base de réflexion à la définition de fonctionnalités d'un système qui soutiendrait l'activité médiatisée. Cette réflexion conduit à un modèle de *l'activité instrumentée*. Ce modèle de l'activité instrumentée est ensuite matérialisé dans un modèle de conception d'outil. Le collecticiel développé permet d'assister les interactions entre les utilisateurs et peut également être un moyen de recueil de corpus. Le corpus, analysé à l'aide de la théorie mobilisée permet alors de valider ou revoir la théorie mobilisée et ainsi de faire évoluer notre compréhension de l'activité observée.

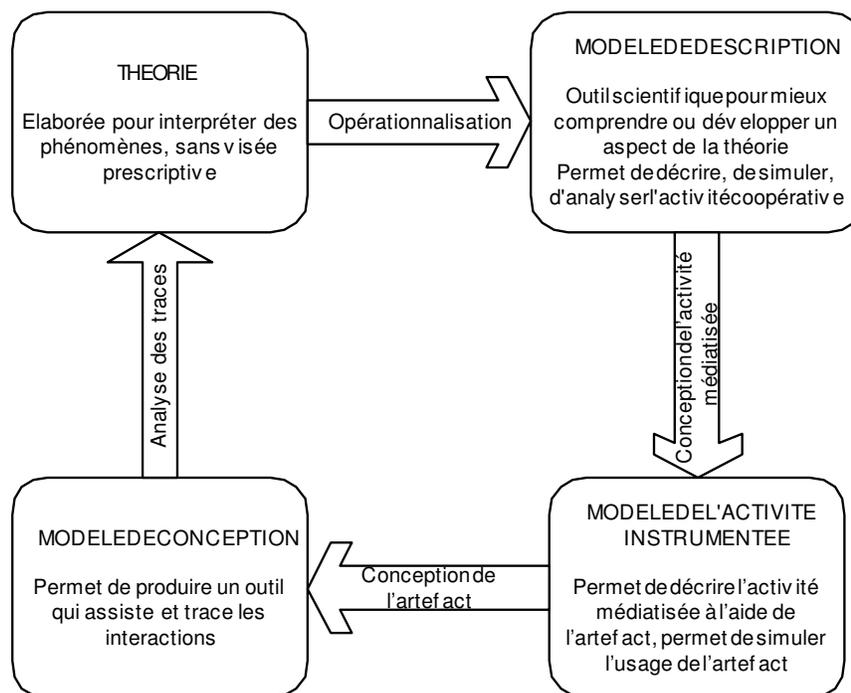


Fig. 1. DÉMARCHE DE CONCEPTION D'UN COLLECTICIEL (LEWKOWICZ ET AL., 2006)

Les modèles de description de l'activité et de l'activité instrumentée proposent une représentation a priori du support possible de l'activité. Afin de relier cette modélisation a priori et technocentrée à une représentation effective d'une activité humaine, nous proposons, dans le cadre de notre terrain, de constituer une collection des productions des collaborateurs au cours du projet. Cette collection de documents sera utilisée comme corpus d'analyse et comme ressource.

2.2. Le corpus

Le corpus est un objet remis à la mode, déjà largement défini et utilisé en Sciences Humaines et Sociales. Son renouveau s'est fondé sur des oppositions voire des contradictions liées à ses utilisations. Les principales différences s'établissent entre des genres de corpus, des analyses quantitative ou qualitative, une linguistique *sur* corpus et *de* corpus menant à des techniques d'interprétations différentes.

2.2.1. Le corpus : un genre

La constitution du corpus est une question centrale en linguistique de la *parole* (par opposition à *langue* et une linguistique hors corpus) depuis de nombreuses années. La construction du corpus, sur la base d'une validation empirique et d'une procédure explicite, était perçue comme une donnée scientifique permettant d'élargir la continuité du social vers les phénomènes langagiers (Guilhaumou, 2002). En analyse de discours par exemple, elle fait ses débuts dans les années 70. On définit ainsi généralement des :

- Corpus de référence, très étendus de manière à pouvoir servir de base représentative des variétés de la langue pour la construction de grammaires, de dictionnaires, etc.
- Corpus spécialisés, limités à une situation de communication, ou à un domaine.

On parle aujourd'hui souvent de corpus électroniques, mais B. Habert, A. Nazarenko et A. Salem (1997- p. 145) soulignent à juste titre que « la simple existence sur support électronique ne fait pas d'un ensemble de textes un corpus électronique. Encore faut-il que ce document obéisse à des conventions de représentation, de codage répandues, voire consensuelles, qui permettent la transmission et la réutilisation des données textuelles en cause. »

D. Mayaffre (2006) raffine ces types de corpus en reconnaissant des genres de corpus plus fins définis selon leurs utilisations suivant des champs d'étude spécifiques. Il définit donc des :

- Corpus phrastiques de grammairiens ou de syntacticiens (recueil d'exemples non pas attestés mais forgés, non pas trouvés mais *controuvés*).
- Corpus lexicographiques ou sacs de mots (recueil de l'ensemble du dictionnaire) dont la grande spécificité et l'incroyable avantage est de pouvoir prétendre à l'exhaustivité. Ce sont donc des corpus clos mais aussi des corpus finis.
- Corpus textuels qui ne peuvent aspirer ni à l'exhaustivité ni même à la représentativité et qui concentrent toujours des données attestées puisqu'on ne saurait fabriquer artificiellement un texte pour prétendre en appréhender le sens.

En effet, c'est au début des années 80 que la notion de *corpus naturel* (Guilhaumou, 2002) ou corpus *attesté* prend son essor, largement représenté par les corpus textuels. Ce naturalisme « révoque le corpus clos préféré dans des corpus phrastiques [...], remet en cause la confrontation entre le corpus et le hors-corpus, [...] refuse enfin la présentation référentielle des conditions de production, au profit d'une description de la réflexivité du discours » (Guilhaumou, 2002, p. 2). La propriété réflexive de la langue est celle qui consiste en l'analyse de la langue par la langue ; la langue s'analyse par ses propres termes.

Cette caractéristique naturelle du corpus est un des critères de qualité chez C. Plantin ((2005) cité dans (Atifi *et al.*, 2006)). Pour ce dernier, le corpus doit manifester les trois dimensions suivantes :

- Technique : il faut réaliser des corpus de bonne qualité technique sonore et visuelle pour faciliter leur conservation, leur transcription et leur exploitation manuelle ou informatique ;
- Juridique : il faut réaliser des corpus qui respectent la vie privée des personnes enregistrées (accord préalable des enquêtés, anonymisation des données), le droit d'auteur (entre collecteurs, transcripateurs et chercheurs) et le recueil et la diffusion de données (préparation et mise en place de l'enregistrement) ;
- Sociolinguistique : il faut réaliser des corpus décrits souvent dans la littérature comme naturels, authentiques et représentatifs (dans le sens évoqué ci-dessus).

Dans le cadre de la représentativité du corpus, J. Sinclair préconise la typicalité du corpus. Ce qui est nouveau, surreprésenté dans un laps de temps court n'est souvent que peu représentatif (Sinclair, 1991). Des textes *atypiques* peuvent donc être inclus dans un corpus, mais le corpus devrait être complété par des textes plus *neutres*. Cependant, la typicalité est difficile à définir. Quel est le degré de typicalité dans un corpus dédié à une analyse de la néologie ? Dans le cadre des annotations, quelle forme, quel style a une annotation stéréo- ou proto- typique? Quel contexte de référence est le plus favorable à une annotation prototypique? Ces questions doivent être posées afin de recueillir un corpus pertinent où le corpus amènera des réponses.

2.2.2. Exhaustivité ou pertinence ?

Jusqu'aux années 70, le corpus défini comme « un ensemble déterminé de textes sur lesquels on applique une méthode définie » ((Dubois, 1969), cité dans (Guilhaumou, 2002)) est adopté. Cette définition implique une neutralisation de l'effet discursif comme vu dans la linguistique de Harris et sa présentation dans *Discours analysis* (Harris, 1952). À la suite de la préférence pour les données naturelles sur des données construites, l'analyse du discours et la constitution de corpus prennent la forme d'une approche sociolinguistique.

La lexicométrie qui quantifie les faits langagiers apparaît au cours des années 1970 et ouvre la voie à la linguistique de corpus. Selon, B. Habert (2000), en linguistique de corpus, un corpus est « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques et extra - linguistiques explicites pour servir d'échantillon d'emplois déterminés d'une langue ». Il reprend en cela la définition désormais classique de J. Sinclair ((1996), cité dans (Habert, 2000)) « 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.⁷ » Cette définition représente à la fois la démarche descriptive et interprétative de la linguistique de corpus et son intérêt quantitatif, gage de sa représentativité par l'exhaustivité. L'explosion de la numérisation des ressources textuelles dans les années 90 a permis la constitution d'une nouvelle étape de la longue histoire entre linguistique et corpus puisqu'elle a permis le recueil de *très grand corpus* attesté. La linguistique *de corpus* a pour objectifs la constitution et la mise à disposition de ressources, et leur confrontation par la communauté des linguistes (Pery-Woodley, 2005). Le temps du

⁷ Un corpus est une collection de fragments de langage sélectionnés et ordonnés selon des critères linguistiques explicites afin d'être utilisé comme échantillon du langage. (Traduction personnelle)

corpus limité, clos, est ici révolu. Le linguiste finit en effet par concentrer son attention sur l'enrichissement des corpus, via les banques de données, l'accroissement de leur taille et l'amélioration des accès aux corpus.

A. Condamines (2005) souligne cependant que le problème de la représentativité n'est pas nécessairement articulé à une ambition de généralisation et que le chercheur dégage seulement ce qui est propre au corpus étudié, ce qui en est un phénomène récurrent. En linguistique interactionnelle, H. Atifi et C. Lejeune (2006) expliquent que l'analyse ne se limite pas au contenu, et le chercheur prend également en considération le niveau formel, à travers la structure et l'organisation des interactions. L'unité d'analyse sera, en fonction du niveau de granularité, la conversation, la séquence, l'échange, l'intervention ou l'acte de langage.

Il est donc important de tenir compte de l'exhaustivité d'un corpus par rapport à un contexte d'analyse bien défini pour permettre une représentation quantitative des données, ainsi que de la pertinence et de la fiabilité de ce corpus pour permettre une représentation qualitative fine. Un corpus spécialisé exhaustif extrait d'un corpus de référence large doit permettre une analyse complexe. Reste à définir le type d'analyse possible à effectuer sur un corpus.

2.2.3. L'interprétation du corpus

La réflexivité du discours a ouvert la voie au corpus réflexif sous la forme du très grand corpus où texte et contexte se retrouvent dans un même espace construit. Le corpus donne alors lieu à une interprétation, « un geste de lecture de facture herméneutique » (Guilhaumou, 2002, p. 7) fondé sur les capacités réflexives, donc interprétatives. D. Mayaffre développe d'ailleurs une herméneutique numérique (Mayaffre, 2002), dans le cadre de l'analyse de textes assistée par ordinateur (ADT, Salem, 1986) principalement fondé sur des analyses lexicométriques statistiques. Dans la veine saussurienne, il considère le texte (et ce qui va avec le texte, le con-texte) comme « l'unité fondamentale d'une linguistique aboutie » d'où naît le sens. Dans un contexte de linguistique de corpus, F. Rastier considère donc que :

« Le texte est pour une linguistique évoluée l'unité minimale, et le corpus l'ensemble dans lequel cette unité prend son sens » (Rastier, 2005, p. 31).

Cette double incidence du texte est celle qui guide l'interprétation en linguistique *de / sur* corpus. G. Williams (2005) distingue ainsi les études *corpus-based* (*basées sur* le corpus) et les études *corpus-driven* (guidée par le corpus). Le positionnement est simple, soit le

corpus est un recueil d'une théorie a priori (un observatoire, (Mayaffre, 2006)), soit il est producteur de sens dans un parcours interprétatif (un observé réflexif (Mayaffre, 2006)). En bref, soit le corpus sert à révéler un sens qui serait préexistant, soit il sert fondamentalement à le construire.

En terme d'interprétation du corpus, il s'agit donc de se positionner du côté de la linguistique *sur* corpus interprétative et inductive, où le corpus a une fonction validante ou illustrative ; ou du côté de la linguistique *de* corpus, où le corpus n'est plus seulement un support, mais un apport (Mayaffre, 2006).

À la lumière de ces considérations sur la constitution et l'interprétation de corpus et sur le domaine spécifique qui nous intéresse, la conception mécanique médiatisée par ordinateur, nous avons recueilli notre corpus directement auprès des membres d'un projet en mécanique afin que le corpus soit pertinent. L'analyse du corpus guidera notre interprétation et permettra de modéliser l'activité d'annotation telle qu'effectuée dans notre groupe projet. *Rejouer* l'activité sur la base d'éléments du corpus permet une réflexion sur l'usage de l'outil. Via une scénarisation des données du corpus recueillies, nous pourrions mener une validation des fonctionnalités du collecticiel développé. Le corpus ne nous permettant de ne conserver que le texte (l'annotation) et son co-texte (le lieu où elle est déposée), nous devons recréer un contexte (le ego, hic et nunc, - les trois coordonnées du discours singulier - et le alii, dans le cas d'une conversation) et prêter un sens à l'annotation par le biais d'un scénario, une confrontation à une réalité.

2.3. Les scénarios

Afin d'aider à la conception d'outils, il est nécessaire de se munir d'Objet Permettant de Penser Avec (OPPA, (Papert, 1993)). Les OPPAs sont des artefacts cognitifs qui fournissent un lien entre les connaissances abstraites et sensorielles, aussi bien qu'entre le monde individuel et social. Le scénario est un objet qui permet aussi de passer d'une réflexion technocentrée à une réflexion anthropocentrée.

2.3.1. Un cadre pour la conception

En conception logicielle, le scénario est un outil courant. Il permet au concepteur de représenter l'activité d'utilisateurs via le logiciel conçu. Un scénario est une description narrée d'un ensemble d'utilisateurs dans un contexte de travail et d'un ensemble de tâches que

l'utilisateur effectue ou souhaite effectuer (Nardi, 1992). Le scénario est un objet utilisé de longue date mais pour lequel les discussions théoriques restent actuelles. K. Go et J.M. Carroll (2004) retracent un rapide historique de la notion de scénario et de son utilisation. Herman Kahn (1962), dans *Thinking About the Unthinkable* (1962) décrit son utilisation de scénario comme une aide pour penser l'incertain. Il utilise le scénario comme un outil d'analyse sous forme de narration comportant des appréciations de types psychologique, social et politique. Selon lui, le scénario est une aide à la réflexion qui permet :

- La saillance des événements à prendre en compte par le concepteur
- Le traitement du détail et de la dynamique
- L'assistance pour articuler l'interaction entre les aspects individuels, psychologiques, sociaux, politiques et militaires
- L'illustration de certains principes ou de certaines questions
- Le raisonnement sur les possibilités de développement alternatives

Dans les années 70, Pierre Wack (cité par Go et Carroll, 2004) souligne l'utilisation de scénarios pour deux grandes situations de conception : la prévision du développement et la planification de l'utilisation future de l'outil conçu.

Dans le cadre de l'interaction Homme-Machine, J. M. Carroll (Carroll et Rosson, 1992), étend cette vision de l'utilisation des scénarios à la compréhension et à la création de systèmes et d'applications informatiques. Il propose d'utiliser les scénarios pour la création d'artefacts, médium d'interaction, comme des systèmes d'apprentissage. Un scénario est en fait une description de l'activité qui permet de mimer le contexte d'utilisation de l'outil au cours de l'activité. Le scénario mime les contraintes, les restructurations de l'activité à travers l'utilisation de l'outil... Il tient son nom des descriptions de l'action d'une histoire (le canevas d'un film par exemple). Le scénario remet l'utilisateur au centre du processus de conception, en focalisant l'attention du concepteur sur l'activité de l'utilisateur et en minimisant la dimension technologique du développement. Le scénario est un prototype d'une activité qui selon J. M. Carroll (1999) peut fournir un cadre adapté pour la conception d'outils. La rencontre du corpus et du scénario, entre les Sciences Humaines et l'informatique, donne naissance à des champs d'étude importants comme le *storytelling* (Soulier, 2006). Le storytelling est une technique de capitalisation des connaissances narrées par un expert. Dans ce cadre, la narration permet de mettre en valeur des actions et de représenter une

activité que l'expert rend sous forme de connaissances considérées stratégiques pour une tâche donnée.

J. Bardram (2000) applique la notion de scénario au développement d'outils coopératifs. Il définit donc des scénarios collaboratifs (collaborative scénarii) qu'il ne considère plus comme focalisés sur une tâche isolée d'un acteur. Un scénario collaboratif a explicitement pour but de capturer l'ensemble de l'activité coopérative alors qu'elle se joue aux frontières de domaines, d'acteurs, de temps et de lieux.

« A collaborative scenario explicitly aims at capturing a whole collaborative activity as it plays out across departmental boundaries, actors, time, and space.⁸» (Bardram, 1998, p. 76)

De plus, un scénario collaboratif porte sur la façon dont les utilisateurs interagissent mais aussi sur le pourquoi ils le font. Ainsi, ce type de scénario essaie de représenter les activités conflictuelles ou *poly-motivées*. Dans notre cadre de conception de collecticiel, les scénarios collaboratifs tels que définis par J. Bardram (2000) semblent appropriés.

2.3.2. Les atouts du scénario

Le scénario permet de relever dans une narration des éléments de contexte modélisables : les scènes de l'action (lieu), les acteurs de l'action, l'objectif de l'action, les objets et concepts manipulés au cours de cette action... Autant d'éléments qui permettent de modéliser une activité dans son contexte. Le scénario permet donc au concepteur de réfléchir et de se projeter sur l'utilisation future de son outil au fil de sa conception.

J.M. Carroll (1999) expose cinq raisons de faire reposer sa conception sur des scénarios, liées à cinq propriétés du scénario. Ainsi, le scénario permet une activité grâce à une ou plusieurs propriétés comme dans le tableau suivant (Tab.1) :

Tab. 1. PROPRIÉTÉS DU SCÉNARIO NÉCESSAIRE À UNE ACTIVITÉ (CARROLL, 1999)

Propriétés	Activité
Prototypique	Réflexion pendant la conception

⁸ Un scénario collaboratif a explicitement pour but de saisir l'ensemble de l'activité collaborative alors qu'elle se déroule à travers des frontières de domaines, d'acteurs, de temps et d'espace. (Traduction personnelle)

Propriétés	Activité
Réaliste et flexible	éviter les inconsistances dues à l'indétermination sans s'enfoncer dans une impasse
Multi - facette Granularité multiple	Montrer différentes facettes d'un même objet et adapter la granularité de la description
Abstrait et catégorisés	guider sa pratique et permettre d'apprendre de sa pratique
Orienté sur la tâche	être concentré sur l'usage et l'intérêt des utilisateurs

Ainsi, le scénario est un langage approprié à la conception de logiciel orientée utilisateurs qui permet une réflexion effective liée à l'action. Le scénario ne permet pas une complétude close pour définir un développement mais au contraire guide le concepteur dans une conception ouverte qui permet d'articuler de nouveaux possibles. Il répond alors aux besoins exposés par D. Schön (Schön et Bennet, 1996) d'une conversation entre concepteurs dans une situation observée (acteur, matériel,...) où le matériel en situation peut *répondre* aux modifications et tests des concepteurs. Cependant, les actions peuvent modifier l'objectif premier défini dans le scénario. Bien que guidées par l'objectif, les actions le modifient. Il est donc crucial, ici comme dans le cas des corpus, de documenter son scénario, d'en expliquer avec précision les tenants et les aboutissants. Ainsi, nous pouvons ajouter une nouvelle raison d'utiliser des scénarios pour la conception logicielle. Le scénario sert à la fois à la contextualisation et à la documentation du corpus recueilli dans l'objectif de conception.

Il nous paraît donc judicieux de lier ces deux objets d'aide à la conception que sont le corpus et le scénario. Ils rendent possible une conception adaptée au contexte et à la tâche grâce au scénario, et pertinente sur le plan de la représentativité grâce au corpus.

2.3.3. Le scénario pour le prototypage et l'évaluation d'outil

J. Bardram (2000) souligne que la limitation principale des scénarios est leur descriptivité puisqu'ils ne font que décrire des activités détachées des conditions réelles de travail. Il préconise donc la confrontation de ces scénarios au monde réel avec ses véritables

conditions de travail. Pour ce faire, J. Bardram décline des scénarios décrivant des pratiques de travail existantes, en scénarios décrivant des pratiques de travail médiatisées par l'outil conçu. De cette utilisation des scénarios, il conclut que la conception fondée sur le scénario est applicable et utile (Bardram, 1998). Il considère que le scénario est bel et bien le pont qui permet de passer de l'existant au futur, ce qui permet un prototypage de l'outil à venir efficace. Le scénario est donc utile au prototypage puisqu'il permet de tracer les besoins en terme de fonctionnalités d'un outil informatique par exemple. J.M. Carroll (2000) présente d'ailleurs diverses méthodes de dérivation de fonctionnalités requises à la suite de l'observation de scénarios.

Une seconde façon de confronter l'outil en cours de conception au monde réel est d'utiliser un échantillon représentatif d'un domaine ou d'une activité. Nous souhaitons effectuer cette confrontation par une série de *scénographies* des scénarios, leur *représentation* au sens théâtral du terme, avec des mots empruntés au contexte, c'est-à-dire appartenant au corpus. Le corpus représenterait donc la partie réelle du monde, du moins ce qui est échangé grâce à la langue, à propos du domaine réel observé. M. Prilla et C. Ritterskamp (2006) utilisent aussi leurs scénarios pour illustrer des considérations sur leur outil. Ils utilisent les scénarios comme des exemples d'utilisation de l'outil conçu. En effet, les scénarios semblent pertinents pour servir de langage commun entre les concepteurs, mais aussi entre les concepteurs et les utilisateurs.

Dans une étude sur l'évaluation d'outil fondée sur des scénarios, S. Haynes, S. Puroo et A. Skattebo déclarent que les scénarios sont efficaces pour aider à concentrer l'évaluation en identifiant des catégories de facteurs (humains, techniques, organisationnels, ...) à analyser (Haynes *et al.*, 2004). De même, la méthode permet d'identifier des actions spécifiques qui ouvrent des possibilités d'amélioration de l'outil. Cependant, ces auteurs concluent que la méthode d'évaluation fondée sur le scénario ne permet pas de mesurer des bénéfices quantifiables dans les applications de Travail Coopératif Assisté par Ordinateur (Haynes *et al.*, 2004). Puisque c'est aussi notre but, nous garderons cette faiblesse à l'esprit et ne tenterons pas d'évaluer quantitativement les bénéfices apportés par notre outil.

L'analyse qualitative sur les outils reste tout de même problématique dans la mesure où elle reste toujours indéterminée. Ceci est dû au fait que la conception d'un outil change la façon dont un utilisateur agit et appréhende son environnement. Il est donc nécessaire de s'appuyer aussi sur des éléments de plus haut niveau que le scénario qui placent l'activité dans un cadre théorique fort. Le modèle est un moyen de traduire, de faire un pont entre

deux activités, l'une naturelle, bien que simulée (le scénario), et l'autre conceptualisée dans le cadre d'une théorie.

3. Conclusion

Par cette démarche nous souhaitons mêler le *donné* et le *construit* d'une façon consciente et pertinente. L'utilisation de corpus (le donné) ou de scénarios (le construit) permet de mener une analyse anthropocentrée inspirée par des concepts théoriques du TCAO comme les artefacts de négociation de frontières. Le modèle de conception permet le passage de l'anthropocentré au technocentré. Il spécifie le collectif sur la base d'une modélisation de l'activité documentaire. Dans cette modélisation, l'accent sera porté sur les mécanismes de négociation.

L'annotation semble être un artefact de négociation de frontière pertinent dans un cadre de partage de documents et de travail médiatisé par ordinateur. En effet, l'annotation permet à la fois la lecture, l'élaboration de texte et l'échange aussi bien dans des interactions en face à face (notes explicatives au fil de l'eau sur un schéma) qu'entre des utilisateurs distants (post-it sur un dossier envoyé). La démarche que nous adoptons permet de rester pertinent et concentré sur les besoins de l'utilisateur tout en traduisant ses besoins en fonctionnalités nécessaires au soutien de son activité. Nous allons donc tout d'abord (1) définir les besoins d'un utilisateur pour annoter sur la base de la littérature grise et des modèles existants. Puis, nous allons (2) construire des outils de soutien à la coopération. Ces outils sont fondés sur (a) la transformation d'un modèle de description de l'activité en un modèle instrumenté permettant de définir des fonctionnalités et (b) sur un corpus reflétant la langue spontanée utilisée dans la communauté de l'utilisateur au cours de son activité de conception. Enfin, nous (3) validerons l'outil conçu par des scénarios construits suite à l'observation des échanges effectivement produits et une représentation de l'activité. La coopération étant fondée sur une communication en Langue Naturelle (LN), il semble naturel aussi de soutenir les activités des utilisateurs à l'aide d'outils de Traitement Automatique des Langues (TAL) traitant les ressources en LN utilisées et portant une orientation plus cognitive au soutien de l'utilisateur.

Nous souhaitons donc soutenir la coopération au sein d'un groupe de concepteur grâce à l'annotation. Les deux parties suivantes de ce mémoire vont être consacrées à l'annotation, sous deux angles, celui de l'activité d'annotation et celui de la production résultant de cette

activité. Le chapitre I est consacré à un état de l'Art sur l'objet annotation. Le Web Sémantique et la Recherche d'Information et de Document, l'Herméneutique, le Web Social et la Communication par Ordinateur sont mobilisés. Leur étude nous permet de dégager des annotations sémantique, discursive et élaborante permettant d'indexer, de commenter et de rédiger des documents. Nous présentons alors un relevé des outils développés pour soutenir ces différentes annotations afin de nous positionner et de réutiliser potentiellement des outils existants.

Nous présentons ensuite en chapitre II notre proposition de définition de l'annotation. Au vu de cette proposition, nous critiquons les outils existants présentés précédemment qui manquent selon nous d'un modèle sous-jacent représentant l'activité d'annotation de l'utilisateur et donc la production de l'annotation polymorphe que nous identifions.

Le chapitre III expose alors divers modèles de description d'activité de lecture et d'écriture en vue de leur utilisation en tant que modèle de description de l'activité d'annotation. Ces modèles sont issus majoritairement de la Psychologie Cognitive qui instaure d'emblée une dichotomie entre lecture et écriture et se focalise sur une activité individuelle. Les modèles interactionniste sont plus pertinents pour notre propos, mais tendent à minimiser les processus de régulation du discours.

Dans le chapitre IV, nous proposons alors un modèle adapté à l'activité d'annotation pour la coopération. Ce modèle de description de l'activité d'annotation est traduit en un modèle de l'activité instrumentée afin de nous permettre de passer d'un modèle anthropocentré à un modèle technocentré permettant la spécification d'un collectifiel.

2^e partie :

Autour de l'objet annotation

Le terme *annotation*⁹ est défini comme une note critique ou explicative qui accompagne un texte. Le terme *note* est lui-même fortement polysémique. Il peut signifier la « marque faite pour garder mention, indication de quelque chose » (proche alors du verbe *adnotare* qui donne *annoter* dans le sens *inventorier*), ou « le mot, la phrase se rapportant à un texte et qui figure à côté du texte », ou encore des éléments brefs du type « bref éclaircissement nécessaire à l'intelligence d'un texte », « brève communication écrite », « brève indication recueillie par écrit », « brève appréciation ». Il existe beaucoup de synonymes à *annotation* et *note* sur un document¹⁰, mais l'annotation est toujours un élément scriptural, tangible, ancré (*adnotare*) au document.

Dans cette partie, nous nous focalisons sur l'annotation en tant qu'objet et non en tant qu'activité (3^e partie). Nous présentons tout d'abord (chapitre II) trois facettes de l'annotation que nous avons identifiées, ainsi que les outils proposés pour leur soutien. Dans un second temps (chapitre III), nous proposons une définition unifiée de l'annotation pour la coopération. Cette définition unifiée nous permet de proposer le soutien d'une activité d'annotation complexe par un outil support au travail coopératif.

⁹ in Le Petit Robert, s.v. « annotation », venu du latin *adnotatio*, apparu au XIV^e siècle

¹⁰ Apostille, glose, remarque, addition, note manuscrite, note marginale, observation, remarque, explication, référence, scolie

Chapitre I

Des annotations sémantiques, discursives, élaborantes

Article 4 : Vols et dégradations [...] il est interdit d'annoter ou de souligner les ouvrages, ou de dégrader intentionnellement les documents. [...] (Cela) sera sanctionnée et fera l'objet d'un signalement à la police et de poursuites judiciaires (pour) acte de malveillance.
Règlement intérieur de la Médiathèque de l'agglomération troyenne

Nous considérons qu'il existe plusieurs types d'annotations¹¹, marqués en anglais par la différence *to tag* (étiqueter) et *to annotate* (annoter) classant d'emblée les outils d'annotation en deux familles. La première se concentre sur l'indexation de pages Web favorisant leur rappel par le biais de *tags* (étiquettes) par des outils issus du Web Sémantique (WS). La seconde se concentre sur la communication humaine via des annotations, des commentaires, des annotations discursives. Pour les documents numériques, l'annotation est généralement considérée comme l'ajout d'information à un fragment de document. Elle est un élément central dans la gestion des documents Web. Cependant, les échanges textuels via des annotations ne consistent pas seulement en l'ancrage de fragments à un document. Ils peuvent aussi donner lieu à l'élaboration d'un document suite à une interprétation. Il s'agit alors d'annotations élaborantes (comme les notes discutées dans les ateliers PRESCOT (Salembier, 2004)).

Dans ce chapitre, nous verrons comment l'annotation est abordée selon différentes disciplines. De même, nous présenterons différents types d'outils qui se proposent de la soutenir. Nous présenterons l'annotation sous l'angle de l'étiquetage de texte (tag) dans le cadre du Web Sémantique, sous celui de la récupération de documents et enfin sous l'angle de la communication avec l'annotation (annotating) de documents.

¹¹ Cette réflexion a donné naissance à bon nombre d'outils d'annotations pour la Linguistique ou les Sciences Humaines et Sociales comme Prospero par exemple (Chateauraynaud, 2003) s'apparentant pour certains à de la transcription pure (WinPitch, (Martin, 2000) ; Transcriber (Barras et al., 2000)) qui ne font pas l'objet de notre propos, et pour d'autres à de la communication (ceux-ci seront détaillés plus bas). Pour un état de l'Art des outils d'analyse de corpus, voir (Lejeune C., 2007).

1. L'annotation sémantique : un index pour mettre la main sur le document

D'après (Leech, 1997) ou (Habert, 1997) dans le cadre de l'annotation de corpus en linguistique informatique, l'annotation est un enrichissement, une *valeur ajoutée*, consistant en un apport d'information de nature interprétative aux données brutes. En effet, il est difficile de séparer d'une façon stricte représentation des données et interprétation. J. Véronis (2000) adopte donc une vue assez radicale, en appelant annotation « tout apport d'information aux données brutes originales ». Cependant, l'annotation est un terme qui est souvent ramené à son sens d'étiquette et son usage indexant. Dans cet usage, les *marginalia* anglaises déposées sur un document et reliées au sens de *to annotate*, s'effacent au profit de l'annotation sémantique *taggée* dans le document pour l'indexer. Ainsi, de l'index¹² qui *indique* un lieu, l'annotation est devenue la ressource elle-même selon certains auteurs, le document étant modifié dans son contenu même par cette annotation informatique. L'annotation sémantique est la descendante de la marque utilisée sur un document pour situer et inventorier. L'annotation sémantique est une métadonnée (une donnée sur une donnée) qui permet d'associer une information exploitable à une ressource (Prié et Garlatti, 2004).

Afin de structurer le nombre grandissant d'informations et de documents déposés sur le Web, l'initiative du Web Sémantique propose une organisation des documents et informations pour les rendre accessibles et partageables pour différents services Web. Cette initiative d'interopérabilité impose le recours à l'annotation sémantique pour donner un sens à l'information en tatouant des métadonnées dans le corps des documents. Le WS développe alors des langages de balise normés qui permettent d'annoter les entités du Web. Il s'inspire en cela de la recherche documentaire qui utilise des descripteurs, des index permettant de retrouver un document. Ces problèmes de typage des informations se posent aussi dans le cadre d'un travail distribué où les membres d'un projet échangent un grand nombre de documents.

¹² Index est un mot latin signifiant « indicateur » dans son acception française en 1503

1.1. Le Web Sémantique

Le Web Sémantique (WS) (Berners-Lee *et al.*, 2001) provient des efforts de normalisation du Web dans le consortium du W3C (World Wide Web), mais également d'autres domaines comme la recherche d'information, le traitement automatique des langues ou la représentation de connaissances. Il est vu comme « un vaste espace d'échange de ressources entre êtres humains et machines permettant une exploitation, qualitativement supérieure, de grands volumes d'informations et de services variés » (Laublet *et al.*, 2004, p. 1). Un espace si vaste que les efforts de normalisation ont pour but non seulement l'interopérabilité des systèmes mais aussi la gestion du nombre grandissant de documents sur le Web. Pour ce faire, le WS s'applique à améliorer la gestion de document par des techniques de Recherche d'Information (RI), en incluant entre autres des techniques d'extraction d'information (de termes et de relations) et de structuration des données (ontologies). Le WS propose différents outils qui utilisent des données normalisées ou qui aident à structurer les données du Web et à leur associer une sémantique. Une couche syntaxique est ajoutée aux informations disponibles sur le Web pour représenter l'apport sémantique. C'est elle qui permet une intercompréhension machine-machine ou homme-machine. Tout élément du WS est étiqueté. Cette étiquette est comprise par des systèmes logiciels ce qui permet leur interopérabilité (services Web). Ces étiquettes normalisées et comprises par des agents humains et logiciels permettent d'attribuer une sémantique aux entités du Web.

Dans cette perspective WS, l'annotation est traitée comme une technique d'indexation de documents ou de parties de documents. L'annotation est vue comme l'enrichissement de données par des métadonnées pour faciliter l'interopérabilité entre systèmes à différents niveaux sémantiques. Une ancre attache alors matériellement l'annotation à un document mais c'est l'annotation, le mot-clé contenu dans la balise, qui lie sémantiquement deux documents. Le contenu de l'annotation est l'élément rattaché sémantiquement aux documents et faisant le lien entre ces deux documents. La caractéristique relationnelle de l'annotation est celle qui autorise le WS à considérer l'annotation comme enrichissant le document. L'annotation enrichit un document en permettant de le relier à d'autres documents contenant la même annotation et donc en améliorant la récupération automatique de documents par des moteurs de recherche. S. Handschuh et S. Staab (2003) distinguent une annotation de surface (shallow annotation) et une annotation profonde (deep annotation) en faisant référence aux documents Web statiques et dynamiques. Ces derniers sont composés d'une structure (deep) et d'un contenu visible par

l'utilisateur (shallow) généré selon des paramètres modifiables. Une annotation profonde va concerner la partie statique d'un document dynamique (sa structure), alors qu'une annotation de surface va concerner un document statique. Pour le W3C, l'annotation est utilisée pour récupérer des données, structurer un document, permettre l'interopérabilité des services et un certain type de coopération (partage de classifications, de signets, ..., (Koivunen et Swick, 2003)).

La tâche principale du Web Sémantique vis-à-vis de ses utilisateurs reste de leur permettre de retrouver des informations et des documents dans ce Web démesuré. Dans cette perspective, le WS propose d'annoter pour permettre à un système d'accéder à l'information. Il suit alors les mêmes objectifs que la Recherche d'information et de document et s'inspire des techniques développées dans ce cadre.

1.2. La Recherche d'Information et de Document

L'annotation sémantique utilisée pour indexer les documents et pour permettre aux moteurs de recherche de retrouver des documents sert aussi à la Gestion Électronique de Documents (GED), au traitement du contenu, à la structuration de document, à l'interopérabilité des services documentaires. Elle est ainsi l'outil de prédilection de la Recherche d'Information (RI) ou de l'ingénierie documentaire.

1.2.1. Les principes de base de la Recherche d'Information

La Recherche d'information (RI) partage une grande partie de ses préoccupations avec la recherche documentaire. La RI traite de la représentation, la conservation, l'organisation et l'accès à des informations (Baeza Yates et Ribeiro Neto, 1999). Née de l'expression de Mooers en 1948, la recherche d'information a rapidement été utilisée dans des applications documentaires (Ihadjadène, 2005). En 1951, Taube propose le concept d'*uniterm* ou mot-clé simple dans une perspective d'application de la logique booléenne à la recherche documentaire (Ihadjadène, 2005). La RI et la recherche documentaire sont bien souvent mêlées puisque les évaluations de SRI (Systèmes de Recherche d'Information) se fondent en fait sur les taux de récupération de documents. Les deux domaines recherchent des descripteurs associés aux documents qui permettent la récupération d'information dans des documents ou les documents eux-mêmes.

Avec l'explosion du document numérique et les milliards de pages Web qui apparaissent, la RI s'est orientée vers une recherche des descripteurs en *plein texte* (c'est-à-dire dans le contenu du document). Le développement des descriptions de documents sous forme de vecteurs de termes en est un exemple (Salton et McGill, 1983, Joachims, 2002). Le recours aux termes ouvre sur deux espaces ; la linguistique automatique statistique (Manning et Schütze, 1999), ainsi que les techniques plus classiques de pondération (Salton et Buckley, 1988). Tous les termes d'un document n'ont pas le même poids, selon leur fréquence d'apparition, leur contexte d'apparition, etc. Plusieurs méthodes de pondération sont en concurrence, principalement la méthode de Jaccard, du χ^2 , ou le *tf.idf* - term frequency.inversed document frequency - (voir en annexe). La pondération des termes dans le traitement du contenu des documents est centrale pour la récupération de documents pertinents.

La pertinence des documents et informations récupérés se fait par rapport à des calculs de rappel et de précision. Le rappel de documents pertinents se doit d'être le plus proche des cent pour cent (dans un groupe défini de documents) par rapport à une requête spécifique, et le bruit (les documents qui ne sont pas pertinents par rapport à la recherche) doit être minimisé afin d'avoir une précision forte. Ces deux paramètres peuvent être améliorés par l'utilisation des techniques avancées d'extraction de descripteurs, fondées sur des critères syntaxiques ou sémantiques.

1.2.2. RI et TAL

Ces techniques d'indexation ou de recherche d'information statistiques, qui sont mises en œuvre par les moteurs de recherche actuels, proposent peu de documents pertinents et ne fournissent donc pas de résultats très satisfaisants. Ces résultats sont dus principalement aux deux phénomènes spécifiques du langage naturel : la polysémie (un mot peut avoir plusieurs sens possibles) et le polymorphisme (le même concept peut avoir plusieurs formulations linguistiques). Ces phénomènes ne sont pas traités dans un moteur de recherche utilisant des mots-clés comme index.

Plusieurs recherches se focalisent donc sur l'amélioration des techniques d'indexation et de recherche d'information par le biais des techniques de traitement automatique des langues (TAL), qui prennent en compte ces phénomènes. Ainsi, une des pistes explorée et développée dans le contexte du Web Sémantique est représentée par l'application d'outils de TAL pour l'identification de termes décrivant le document, à l'aide des critères syntaxiques (Strzalkowski,

1999). Les termes qui indexent les documents sont identifiés suite à une analyse syntaxique (avec éventuellement lemmatisation des mots), parmi les groupes nominaux trouvés dans le document. Les requêtes utilisateur sont également analysées pour identifier les termes pertinents et ces termes vont être recherchés parmi les descripteurs des documents. Malgré les techniques d'analyse syntaxique robustes mises en œuvre, il est difficile d'appliquer ces méthodes sur l'ensemble des documents du Web ; s'agissant de texte libre, l'ensemble des termes utilisés comme descripteurs est pratiquement infini (Voorhees, 1999). Cette famille de méthodes peut s'appliquer sur un ensemble de documents liés à un domaine précis, sur un site dédié à ce domaine, en utilisant comme descripteurs un ensemble de termes spécifiques au domaine extraits automatiquement (à l'aide d'un extracteur de termes). Ces méthodes, qui s'appuient sur des connaissances syntaxiques, permettent d'éviter des erreurs dues à l'ambiguïté lexicale (comme dans le cas des mots clés), mais elles restent toutefois limitées, puisqu'elles ne prennent pas en compte le sens des mots.

Pour dépasser ces limites, une autre famille de méthodes tente d'améliorer les méthodes de Recherche d'Information et s'appuie pour cela sur une analyse sémantique des documents. Ainsi, l'utilisation d'une base lexico-sémantique comme Corelex (Buitelaar, 1998), EuroWordnet (Vossen et *al.*, 1998) ou FrameNet (Baker et *al.*, 1998) est possible. Les descripteurs de chaque document seront donc choisis parmi les synonymes ou les concepts de la base. Ainsi, les mots qui composent les requêtes sont également associés à des concepts identifiés de la base. Ces concepts vont être recherchés dans l'index pour être retrouvés dans la base de documents. Certains systèmes de question-réponse (Narayan et *al.*, 2004) ont prouvé que l'utilisation de ces ressources peut améliorer les résultats, même pour des textes libres. Si l'utilisation de ces méthodes sur des textes libres est possible, pour une application de RI dans un domaine spécifique le problème qui se pose est de trouver les bases lexico-sémantiques adaptées au domaine, puisque celles citées ci-dessus ne couvrent pas la totalité des domaines.

Malgré ces limites, ces méthodes permettraient de retrouver automatiquement des descripteurs dans le corps (le texte) d'une annotation ou de nouveaux documents et donc d'aider l'utilisateur dans son activité d'indexation de documents. Nous envisageons d'adapter des méthodes fondées sur l'identification syntaxiques des descripteurs pour notre application.

Pour éviter les problèmes liés à l'utilisation d'outils de TAL ou de bases lexico-sémantique pour identifier les descripteurs, une autre piste de recherche explorée dans le cadre du Web Sémantique est représentée par l'utilisation d'ontologies pour indexer et annoter des documents.

1.3. Les ontologies

En Intelligence Artificielle, les ontologies, utilisées pour les systèmes mimant les comportements humains (langage par exemple), sont définies par (Gruber, 1993) comme *une spécification explicite d'une conceptualisation*. Dans (Gruber et Lytras, 2004), Thomas Gruber raffine sa définition de ce type de base de connaissances en prenant en compte la coopération nécessaire pour arriver à un accord sur le sens : « Every ontology is a treaty – a social agreement – among people with some common motive in sharing¹³ » (p. 5). Nous arrivons ainsi à une définition où T. Gruber porte l'accent sur le processus de négociation orienté vers un objectif de la conceptualisation plutôt que sur sa structure.

« The ontology is representation artifact (a specification), distinct from the world it models, and that it is a designed artifact, built for a purpose. [...] I would try to emphasize that we design ontologies.¹⁴ » (Gruber et Lytras, 2004, p. 1)

Nous adoptons cette définition suffisamment large et orientée sur l'usage qui permet de prendre en compte de nombreuses acceptions du mot *ontologie* (vocabulaire technique, référentiel métier, terminologie/thesaurus, système de classes d'une représentation par objet, base de connaissances terminologique,...). (Lassila et McGuinness, 2001), répertorient ces ontologies plus ou moins formelles dans un spectre (Fig. 2) que (Uschold et Grüninger, 1996) organisent suivant leurs utilisations :

- pour la communication humaine : ontologie non équivoque mais informelle ;
- pour l'interopérabilité des systèmes informatiques : format d'échange ;
- pour la conception de systèmes : codage formel, utilisée en tant que métadonnées.

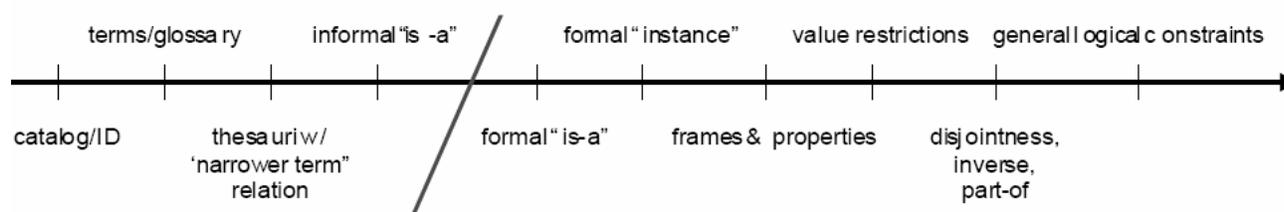


Fig. 2. SPECTRE DE L'ONTOLOGIE

¹³ Chaque ontologie est un traité – un accord social – entre des personnes qui ont des raisons communes de partager. (Traduction personnelle)

¹⁴ L'ontologie est un artefact de représentation (une spécification) distinct du monde qu'elle modélise, et qui est un artefact conçu à dessein et construit pour un propos. [...] J'essaierais de souligner que nous concevons intentionnellement les ontologies. (Traduction personnelle)

Parmi toutes les catégories de ressources ontologiques, nous nous intéressons plus particulièrement aux ressources utiles pour la communication Homme-Homme via la machine. Les ressources peu formelles du type thésaurus, catalogue, ou glossaire ne sont majoritairement pas reconnues dans la grande famille des ontologies et restent qualifiées de terminologies. Nous utiliserons donc le terme de *Ressource Termino-Ontologique* (RTO, Bourigault et Aussenac, 2003) dans un sens générique (ressources formelles, semi-formelles, et informelles) et *ontologies* pour les ressources formelles ou semi-formelles.

Nous pouvons définir plus en détail une ontologie formelle comme une modélisation des connaissances du monde, organisées en un réseau de concepts. Une ontologie consiste en un ensemble de définitions de catégories de base (objet, relations, propriétés) qui permettent de décrire les objets du domaine que l'on traite, leurs propriétés et les relations qu'ils entretiennent les uns avec les autres. Comme le souligne Fabien Gandon¹⁵, « dans une ontologie, les intensions sont organisées, structurées et contraintes pour représenter notre conception du monde et de ses contraintes ». Par exemple, un verre serait (in s.v « verre », Le Robert), « La substance fabriquée, ou un récipient à boire, ou le contenu d'un verre par extension ». Cette formalisation de l'ontologie permet de calculer des inférences guidées par les *intensions* de la conception. L'ontologie permet d'utiliser un verre à *vin* ou *d'eau*, ... Sa formalisation en l'occurrence ne permet pas d'utiliser un verre pour y mettre des fleurs (qui ne se boivent pas), ni les cendres d'un mégot à la fin de la soirée. Les ontologies formelles sont donc utilisées principalement pour la conception de systèmes. Elles proposent une modélisation complexe et rigide des connaissances, utilisable par les agents logiciels. Ainsi, les relations hiérarchiques (de type est-un) ou horizontales entre les concepts ou les instances des concepts sont définies rigoureusement. Les propriétés des concepts peuvent avoir des valeurs dans un intervalle limité et défini d'avance et les axiomes définis imposent des contraintes logiques strictes qui contrôlent les inférences logiques qu'on peut appliquer sur les données (héritage des propriétés, relations transitives ou inverses). Par le biais des inférences logiques, il est possible de déduire de nouvelles connaissances à partir de celles qui sont déjà définies et qui permettent aux agents de comprendre le contenu du document. Malgré ces avantages, la complexité des calculs nécessaires est trop importante et la construction d'une ontologie complète nécessite l'intervention des experts, ingénieurs de la connaissance ou ontologistes, ce qui n'est pas adapté à notre situation, où les utilisateurs doivent mettre à jour l'ontologie au fur et à mesure de leur activité.

¹⁵ http://interstices.inria.fr/display.jsp?qs=id%3Djalios_5127&id=c_17672&part=1

Si l'interopérabilité des systèmes est l'objectif visé par l'utilisation de l'ontologie, alors les inférences logiques ne sont pas absolument nécessaires. Une définition complète des propriétés de chaque concept et des relations qui relient les concepts est suffisante pour pouvoir échanger des données entre les applications, à condition que cette définition soit unique et non-ambiguë. Un fois la définition du concept ou de la relation établie, les diverses applications l'utilisent pour échanger les données. L'intervention d'un expert est cependant toujours nécessaire pour établir les définitions des concepts. Cette intervention rend difficile son utilisation dans notre contexte. On parle alors d'ontologie semi-formelle. T. Gruber considère que les ontologies sont toujours des mélanges de parties informelles et d'autres formelles. Les ontologies dites semi-formelles sont majoritairement informelles et de conclure que « all practical ontologies are semiformal ¹⁶ », toutes ces ontologies qui permettraient de transformer un verre en vase ou en cendrier selon la situation.

Ainsi, pour une communication Homme-Homme, une ressource de type thésaurus ou base terminologique est suffisante, même si les termes du thésaurus ne sont pas toujours explicités par un ensemble de propriétés et que les relations sont en général des relations hiérarchiques. Les définitions des termes sont décrites en langage naturel, donc difficilement utilisables directement par les agents logiciels. En revanche, la mise à jour d'une telle ressource est réalisable facilement par des utilisateurs qui connaissent bien le domaine d'application, comme c'est le cas pour les membres du groupe de travail étudié.

Dans un souci d'améliorer le partage des connaissances sur le Web, l'interopérabilité entre les systèmes, la communication Homme-machine et machine-machine, le Web Sémantique regroupe des recherches sur les ontologies qui se focalisent sur leur construction et leur mise à jour, leur réutilisation, la fusion de plusieurs ontologies et le développement des langages de représentation d'ontologies (RDF (Brickley et *al.*, 2004), OWL (McGuinness et *al.*, 2004), Topic Maps (Biezunski et *al.*, 1999) etc.). L'amélioration des résultats d'un système de recherche d'information peut être obtenue par exemple par l'utilisation des ontologies pour décrire le contenu des documents. Ainsi, les documents sont annotés avec des concepts ou des termes qui apparaissent dans ces documents. Les ontologies jouent alors le rôle d'index, pour retrouver les documents qui contiennent des instances de ces concepts. Les requêtes utilisateurs sont aussi interprétées pour identifier les concepts ou les termes. Les concepts et les relations définies dans les ontologies sont utilisés pour annoter certaines parties de documents

¹⁶ Toutes les ontologies pratiques sont semi-formelles. (Traduction personnelle)

qui contiennent des instances de ces concepts. Les agents logiciels peuvent alors déduire des relations entre les documents, ou retrouver l'information dans un document puisqu'il sera indexé par son contenu et que le descripteur contiendra le concept. Un agent humain peut se servir de l'ontologie pour naviguer dans la base de documents. Si la ressource utilisée est une RTO, les documents sont indexés par les termes de la RTO, et l'agent humain va pouvoir retrouver le document par navigation dans cette RTO. Ce dernier scénario est adapté à notre contexte, et nous allons donc privilégier l'usage d'une RTO par rapport à une ontologie formelle.

Pour mettre en œuvre un système d'indexation sur la base d'une RTO, il est nécessaire de construire une RTO adaptée à notre application. Malgré l'existence d'ontologies génériques (comme SUMO (Nils et Pease, 2001), ou DOLCE (Gangemi *et al.*, 2003)), et de quelques ontologies du domaine, nous ne disposons pas d'ontologie adaptée à notre domaine d'application. Pour créer la RTO nécessaire à l'indexation, il faut donc appliquer une méthode classique de construction de RTO à partir de textes, que nous allons détailler en chapitre 5 section 5. Une fois la RTO construite, le problème qui doit être résolu concerne l'indexation de nouveaux documents. Cette tâche est difficile pour l'utilisateur, il doit identifier les termes pertinents dans le document et les proposer comme index, en les ajoutant dans la RTO. Si le nombre de documents est trop important, l'utilisation d'un outil d'extraction de terme s'avère nécessaire, pour soutenir l'utilisateur dans son activité de classement des documents.

Pour illustrer divers usages d'ontologies pour la RI nous nous focalisons sur la présentation d'outils d'annotation sémantique.

1.4. La boîte à outils de l'annotation sémantique

Les outils ayant pour objectif la récupération d'information dans un document, ou d'un document à part entière, mettent principalement en œuvre une activité de classement, automatique ou supervisée par un utilisateur. Elle consiste donc en l'indexation ou l'étiquetage automatique ou semi-automatique de documents ou de fragments de documents. Cette activité d'annotation sémantique, en raison de sa finalité de communication machine, met en jeu des balises indissociables des données sur le plan de la visualisation de données *logiques* mais qui permettent une visualisation graphique privilégiée. Les annotations sémantiques sont géographiquement dépendantes d'une partie de la page Web. Elles enrichissent la page de concepts issus de l'extérieur pour l'indexation automatique de pages Web, mais n'assistent pas les interactions entre lecteurs d'une même page. Les outils d'annotation répondant à la

philosophie du WS fonctionnent de préférence par des actions de type inférences dans un espace inférentiel assez réduit. En général, cet espace peut être représenté par une taxinomie aux relations *est-un* voire une ontologie complète (formelle). Les inférences se font entre un ensemble de serveurs qui ont besoin d'échanger des informations.

Les outils proposés par le WS permettent une interopérabilité sémantique à différents niveaux de la machine. Le WS propose donc des langages, des formalismes, des protocoles ou encore des ontologies ou des ressources normalisées pour le balisage, sur lesquels se fondent les outils (client et serveur) développés pour permettre un Web Sémantique.

1.4.1. Échanger des annotations : Annotea, protocole et serveur d'annotation

¹⁷Annotea (Kahan *et al.*, 2001, Koivunen *et al.*, 2001) est un LEAD (Live Early Adoption and Demonstration) du W3C qui se focalise sur la normalisation pour faciliter le travail coopératif. À la fois un protocole d'annotation standardisé du W3C et un serveur d'annotation répondant à ce protocole, il fonctionne avec différents clients et serveurs (principalement Amaya, Annotea, Zannot, ...) qui acceptent le protocole (Annotea, 2001) (Voir Fig. 3).

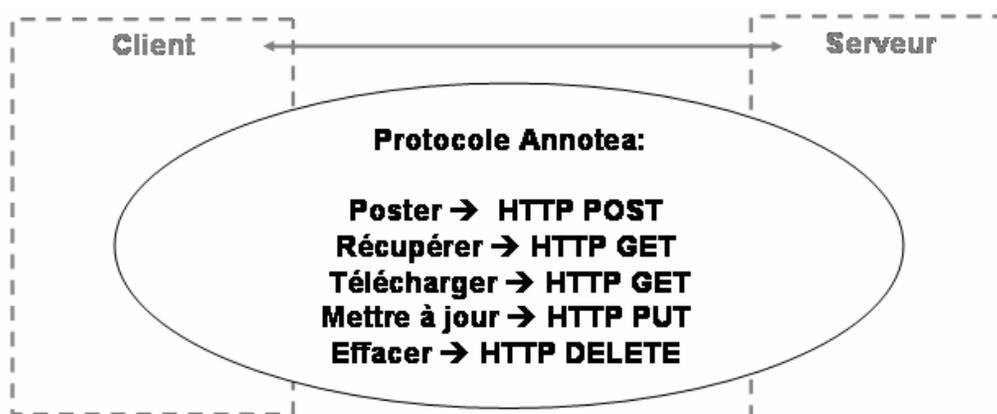


Fig. 3. LE PROTOCOLE ANNOTEA

Ce standard propose à la fois une architecture client/serveur d'un système d'annotation partagé, un protocole de communication des annotations selon le protocole HTTP et une description des annotations selon un schéma RDF (cf. Fig. 4¹⁸), où les annotations sont

¹⁷ Cette partie est issue des travaux de M.F.Sriti (Sriti, 2004)

¹⁸ de <http://www.w3.org/2001/Talks/www10-Annotea-paper/slide7-0.html>

représentées par un ensemble de métadonnées et un corps (le contenu de l'annotation). En effet, les métadonnées contiennent des informations sur l'annotation comme le nom de son auteur, sa date de création, son type, l'URL du document annoté, le XPointer pour spécifier les parties annotées du document. Les annotations sont extérieures au document et elles peuvent être stockées dans un ou plusieurs serveurs d'annotation.

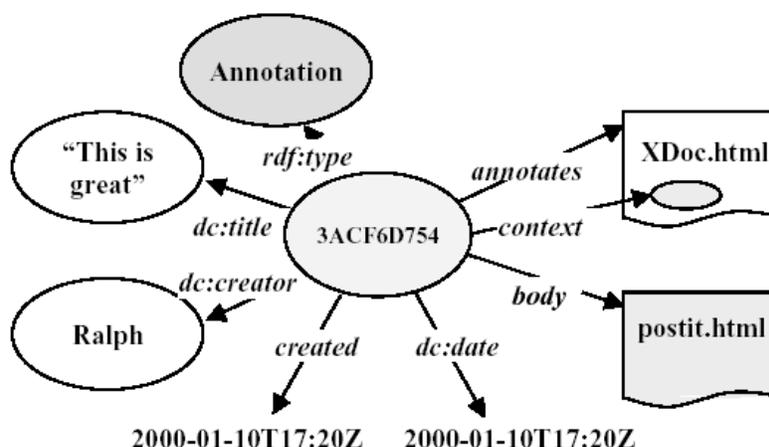


Fig. 4. SCHÉMA D'UNE ANNOTATION SELON ANNOTEA-RDF

Le standard (Annotea, 2001) propose un schéma RDF (Annotea Annotation Schema¹⁹) décrivant un ensemble de propriétés (sous forme de métadonnées) caractérisant l'annotation. Ceci pour standardiser le format d'échanges entre les applications manipulant l'ensemble des annotations d'une part, et pour pouvoir conserver le contexte dans lequel les annotations sont créées, d'autre part. Les propriétés proposées par Annotea sont :

- type : le type d'annotation (*Advice* : conseil pour le lecteur, *Change* : proposition de changement, *Comment* : commentaire, *Example* : exemple, *Explanation* : explication du contenu, *Question* : question sur le contenu, *SeeAlso* : référence à une autre ressource)
- annotates : URL du document annoté
- body : contenu de l'annotation
- context : l'ancre (le XPointer)
- author : l'auteur de l'annotation
- created : date et heure de création
- date : date et heure de la dernière modification.
- relate : relation entre l'annotation et une autre ressource

¹⁹ <http://www.w3.org/2000/10/annotation-ns>

Ces propriétés s'avèrent très utiles surtout pour la phase d'indexation. Il existe aussi un schéma complémentaire à Annotea, Annotation Schema définissant la réponse (reply) à une annotation (Thread Schema²⁰). Ce schéma suppose que la réponse à une annotation est une annotation possédant deux propriétés en plus : *root* et *inReplyTo*. En effet, cette définition est suffisante pour reconstruire l'arborescence des fils de discussion, car *root* pointe vers l'annotation racine d'un fil et *inReplyTo* pointe vers l'annotation parent.

Le premier client implémentant Annotea est Amaya (Amaya, 2003), développé aussi au sein du W3C. Il existe d'autres applications conformes au standard Annotea notamment Annozilla et Janno. Quelques caractéristiques de ces clients sont présentées par la suite.

Annotea est un développement qui marque la volonté du WS à s'ouvrir à l'utilisateur en lui proposant un système pour les interactions Homme-Homme via la machine. Cependant, la volonté d'interopérabilité portée par le WS y est aussi visible.

1.4.2. Clients et serveurs d'annotation

1.4.2.1. Serveur

Le W3C propose donc Annotea protocole et Annotea serveur comme présentés ci-dessus. Le serveur Annotea est un serveur qui reste buggé et n'est plus entretenu aujourd'hui. Un autre serveur d'annotation proposé en Open-Source est le serveur ZAnnot (Zannot, 2003). Zope Annotation Server (ZAnnot) est le serveur Open-Source encapsulable sur la plateforme de développement Zope (Latteier *et al.*, 2003). C'est un outil simple et puissant développé en Python sur des bases standardisées et réutilisables. Il gère les annotations envoyées par un client d'annotation selon le protocole Annotea. Il récupère donc, l'Url de la page annotée, l'ancre de l'annotation dans le document (par des XPointer) et le corps de l'annotation. ZAnnot conserve l'ensemble dans un objet annotation, sur le serveur d'annotation choisi. L'annotation est affichée sur demande d'un utilisateur d'afficher les annotations mises en mémoire sur ce serveur pour une page Web donnée. Le serveur est utilisable avec de nombreux clients et son code ouvert permet de l'adapter à de nouveaux usages.

²⁰ <http://www.w3.org/2001/03/thread>

1.4.2.2. Clients

De nombreux clients d'annotation ont été développés. Nous n'avons pas pour objectif de faire un relevé exhaustif des clients existants, mais cependant, nous souhaitons mettre en valeur le fait que les outils d'annotation, bien que comportant des fonctionnalités proches et se conformant aux recommandations du W3C, ont des utilisations très diverses.

La plupart des outils d'annotations répondant aux recommandations du W3C ont pour objectif d'indexer des pages Web plus ou moins automatiquement. Ils sont utilisés pour la création de métadonnées et certains se fondent sur des ontologies pour soutenir l'annotation sémantique. Ces annotations sont liées géographiquement à une partie d'une page Web et ajoutent des informations supplémentaires, mais elles n'aident pas à la coopération ou aux interactions entre les lecteurs d'une même page. Ce sont des métadonnées qui indexent une page, et permettent aux moteurs de recherche un meilleur rappel d'informations. Les métadonnées sont placées sur la page soit manuellement par l'utilisateur, soit automatiquement par appariement.

Nous classerons les outils d'annotation du WS en trois catégories principales selon leur degré de coopération. Chaque catégorie possède un type d'objet permettant la mise en relation / coopération (ontologie, document, utilisateur). Pour chacune de ces catégories, nous décrivons le prototype selon ses avantages et ses inconvénients et citons un outil en exemple. Les descriptions de ces outils sont disponibles en annexe.

Tab. 2. GRILLE DE CLASSEMENT DES OUTILS D'ANNOTATION SÉMANTIQUE

Catégorie	Point de mise en relation	Description	Point fort	Point faible	exemple
Interopérabilité machine-machine	Lien sur ontologie	Interopérabilité basique et annotation de pages Web manuelle, Interface de choix et de saisie, visualisation d'ontologie	Simplicité	Long et Fastidieux (manuel)	OntoMat-annotizer, SHOE Knowledge Annotator
		Construction de terminologies sur mesure, multilinguisme	Adaptabilité	Fastidieux (manuel)	SyDoM
		Extraction de concepts, Apprentissage sur les documents annotés	Soutien à l'utilisateur	Beaucoup d'erreurs en tout automatique	S-CREAM, Melita
Interfaçage annotateur-machine	Lien sur document	Ontologie a un rôle d'index permettant la mise en relation de documents,	Adaptabilité de la classification	Base de documents personnelle et fermée	MnM
		Navigation dans une base de documents indexés	Partage de documents Indexation et partage larges (pages Web) Lien automatique selon l'index	Concepts figés Relation à une ontologie à la fois	COHSE

Catégorie	Point de mise en relation	Description	Point fort	Point faible	exemple
Aide à l'utilisateur	Lien sur utilisateur	Navigation dans des documents sur la base d'une ontologie partagée, Indexation de documents Compréhension entre utilisateurs	Mise en valeur thématique du document, Visualisation partagée du document	Limitation de la coopération entre utilisateurs à la visualisation (pas de communication)	Magpie

Ces outils sont les principaux dans le champ de l'annotation sémantique. Ils donnent un aperçu intéressant des outils disponibles. Principalement développés dans des contextes de recherche, ils sont peu utilisés et réutilisables (non disponibles, implémentation chaotique). Peu d'outils ont pour objectif de faciliter le travail cognitif sur des pages Web. Nous soulignons toutefois Magpie, une initiative du Knowledge Media Institute (KMI) qui, bien que développé dans une optique WS, propose une manipulation de l'annotation par l'utilisateur.

Ainsi, la définition d'une annotation sémantique, pertinente au niveau *computationnel*, c'est-à-dire pour une utilisation des annotations par des systèmes informatiques, permet la mise en place de nombreux systèmes d'annotation pertinents. Cependant, ceux-ci restent faibles pour une utilisation cognitive de l'annotation. Poursuivant notre objectif de soutien à la communication et à la négociation, nous nous intéressons à d'autres travaux plus orientés vers le discours et la communication. Notre intérêt pour l'annotation sémantique et les techniques d'ancrage ou de récupération des informations liées à une balise est d'avantage lié au besoin de soutenir l'accès aux ressources documentaires du Web par un acteur humain, qu'au besoin de soutenir l'interopérabilité de systèmes. Nous adoptons donc le principe de structuration de plus haut-niveau des descripteurs sous forme de RTO, mais souhaitons moduler les principes de structuration de RTO semi-formelle en fonction de leur objectif d'utilisation. En effet, dans un cadre de soutien à l'utilisateur, les outils proposés pour soutenir l'indexation et fondés sur des ontologies sont intéressants. Pourtant, ils s'appuient sur des classifications trop rigides et définies a priori ce qui gêne leur adaptation à l'activité d'utilisateurs spécifiques.

2. L'annotation discursive

Nous entendons ici annotation discursive dans le sens de l'annotation conçue comme un discours, suivant la dichotomie discours/texte courante dans le champ de la linguistique où le discours est l'« inclusion d'un texte dans son contexte » (Charaudeau et Maingueneau, 2002, s.v. *discours*). Il s'agit donc de considérer l'annotation comme un texte inclus dans son contexte, c'est-à-dire lié à ses conditions de production et de réception. Son contexte est formé notamment par le rôle de son auteur, son contenu sémantique, ou encore sa place dans le document. L'annotation est un discours secondaire (Derrida, 1991, p. 202) qui ne peut que « répondre » et non « parler en premier », mais qui permet entre autres d'élaborer une interprétation partagée du document (comme dans le champ de l'herméneutique).

2.1. Herméneutique

L'herméneutique, du grec *hermêneuein*, *interpréter*, est une discipline qui consiste en l'interprétation des textes. L'herméneutique fournit un ensemble de techniques pour permettre de recréer un sens autour d'un texte qui ne possède plus qu'intrinsèquement son contexte. Ce sens laissé au sein du texte même par l'auteur est redécouvert par une succession d'énoncés expliquant une interprétation, se complétant, se répondant, construisant de nouvelles pistes d'interprétation. L'interprétation des textes est traditionnellement accompagnée de gloses, commentaires et autres annotations ancrées au texte même, ou reliant différents textes ou

fragments de texte. Dans ce contexte, l'annotation est une production discursive de la lecture critique. Elle permet de contextualiser le texte et de le faire entrer ainsi dans un contexte discursif. Les annotations, en tant qu'ensemble de fragments textuels reliés au texte et formant son *co-texte* (contexte textuel) rendent possible cette lecture critique. Une lecture critique peut donc être productive d'un autre texte, commentaire ou critique. Le corps textuel de l'annotation marque l'argumentation portée par un lecteur autour d'un texte, un point de vue d'un lecteur. Elle est productive d'une interprétation qui éclaire non seulement le texte lu, mais également d'autres textes. Dans une optique de coopération, nous nous intéressons plus particulièrement à une lecture critique à plusieurs comme mise en place en herméneutique, permettant la construction d'une interprétation partagée du document initial.

2.1.1. L'herméneutique médiévale

L'herméneutique est souvent reliée aux pratiques religieuses du Moyen-Âge où les manuscrits étaient diffusés et annotés par tous leurs lecteurs dans toute l'Europe. L'herméneutique est alors une technique de lecture qui permet une compréhension véritable d'un message divin en parcourant différents plans de signification : passer du sens littéral du texte à un sens allégorique puis tropologique (ou moral) pour atteindre enfin le sens anagogique (ou mystique). (Carruthers, 1990) souligne que les annotations n'étaient pas déposées uniquement à des fins de commentaires critiques, mais bien souvent aussi à des fins de mémorisation des textes. L'annotation mémorielle est une marque typographique (virgule, paragraphe), symbolique (lettrine, animaux fabuleux) ou verbale, le plus souvent laissée dans la marge. Ces signets graphiques sont à la fois indicatifs (un signal indicateur) et mnémotechniques ; ils servent à repérer des passages importants et à faciliter la mémorisation en créant des liens dans la mémoire visuelle.

La philologie (analyse critique et étude formelle des textes), la philosophie et la littérature intègrent rapidement les techniques scientifiques de l'herméneutique. L'annotation est alors un élément d'interprétation qui permet d'exprimer le contexte d'un document. L'annotation est la pierre angulaire de la re-construction d'une interprétation d'un document.

2.1.2. Une herméneutique littéraire ?

Bien que Ricoeur définisse l'herméneutique littéraire (Titre de HR., Jauss, 1988) comme « un dévoilement du sens caché sous le sens apparent » (cité dans Rastier, 1989, p. 18), une

large réflexion est menée par les théoriciens littéraires. Genette, dans ses réflexions sur la *transtextualité*²¹ prend en compte l'annotation dans le cadre de cinq formes de relations *transtextuelles*²² du texte (Genette, 1982). Ces relations ne sont bien entendu pas exclusives et une annotation comprise en un sens plus large que la note de bas de page prise en compte dans la perspective de Genette, peut appartenir à différentes relations transtextuelles. (Genette, 1987) considère la note de bas de page comme un élément facilement incorporable au reste du texte, ce qui peut être vrai dans un cadre de parole auctoriale unique comme en littérature, mais pas dans notre contexte de conception collective. Pour Genette, l'annotation peut donc faire partie du paratexte, et plus précisément du péri-texte (autour du texte mais à l'intérieur du livre) par opposition à l'épi-texte (à l'extérieur du livre) (Genette, 1982), même s'il n'y a pas toujours lieu de mobiliser la notion de paratexte pour cette unité. Cependant, l'annotation dénote bien une contamination du hors-texte dans le texte. Toujours en littérature, (Derrida, 1991) souligne cette double contrainte (double-bind) : Le texte dit qu'il doit être lu en silence, et à la fois crie qu'il veut une réponse du lecteur (Derrida, 1991, p. 203) :

« 'Be quiet, all has been said, you have nothing to say, obey in silence,' while at the same time it implores, it cries out, it says, 'Read me and respond: if you want to read me and hear me, you must understand me, know me, interpret me, translate me, and hence, in responding to me and speaking to me, you must begin to speak in my place, to enter into a rivalry with me'. »

Plus un texte se pose en *inannotable*, plus il va générer d'annotations de la part du lecteur. C'est le paradoxe de la double contrainte derridienne. Cette double contrainte, révélatrice de la reproduction - concept clé du poststructuralisme - est proche des notions d'intérieur-extérieur de la philosophie heideggérienne et du cercle herméneutique gadamérien.

2.1.3. L'herméneutique philosophique

L'herméneutique se généralise en philosophie avec Schleiermacher. Il considère alors que l'herméneutique n'est pas seulement une technique de lecture pour comprendre le sens d'un oracle. L'herméneutique est surtout un dialogue avec un auteur. L'objectif de cet échange est de s'efforcer de retrouver l'intention d'un auteur, de comprendre un esprit par le biais du texte, de son expression. Trois démarches sont mises en place dans l'herméneutique philosophique : la

²¹ « tout ce qui met le texte en relation, manifeste ou secrète, avec d'autres textes » (Genette, 1982)

²² Intertextualité, métatextualité, architextualité, hypertextualité, paratextualité (voir Glossaire)

compréhension, l'interprétation, et enfin l'application. (Gadamer, 1960) ouvre cette expérience en définissant l'herméneutique moins comme une méthode que comme une interprétation d'une expression située de l'homme. Le texte ouvre des *horizons* dans lesquels l'interprète se situe. L'interprétation est alors une compréhension de l'interprète dans la tradition dont il provient : C'est le cercle herméneutique, concept majeur de l'herméneutique philosophique. L'interprète s'inscrit dans son horizon propre, ce qui lui permet de prendre la mesure de son identité par rapport à l'autre et donc de comprendre le vécu de l'autre, de l'interpréter. Le transfert volontaire ou non des points de vue permet à deux altérités de parvenir à un langage commun et à une énonciation commune.

Le cercle herméneutique décrivant un dialogue avec un auteur, il est d'autant plus facile de mettre en relation l'échange herméneutique avec les échanges numériques de la communication médiatisée par ordinateur.

2.2. La boîte à outils de l'annotation discursive

L'annotation discursive est un échange entre différents lecteurs par le biais de l'écrit. Plusieurs outils informatiques permettent ce type d'échange grâce aux protocoles de communication Web. Nous classerons les outils d'annotation pour l'annotation discursive selon le type de coopération permis. Chaque catégorie possède un type d'objet permettant la mise en relation / coopération (commentaire, document, utilisateur). Pour chacune de ces catégories, nous décrivons les spécificités d'un ou plusieurs prototypes. Les descriptions détaillées des outils sont disponibles en annexe. Les outils décrits répondent tous à un principe d'architecture client – serveur et de visualisation de commentaires d'un utilisateur humain en langue naturelle sur un document disponible sur un serveur. Nous avons choisi de nous limiter aux outils Open-Source ou non-propriétaires. Cependant, dans la littérature, des logiciels propriétaires ou certains plug-in de logiciels spécifiques sont disponibles. Ils proposent parfois de déposer des commentaires qui ne sont ni indexés ni différenciés du document comme dans Windows Word Comments²³ (Windows, 2003). D'autres peuvent adopter une approche distribuée où les annotations peuvent être stockées à part sur des serveurs d'annotations et organisées de façon minimaliste (Acrobat PDF²⁴).

²³ <http://office.microsoft.com/fr-fr/assistance/HA010714941036.aspx>

²⁴ <http://www.adobe.com/support/techdocs/ac76.htm>

Tab. 3. GRILLE DE CLASSEMENT DES OUTILS D'ANNOTATION DISCURSIVE

Catégorie	Point de mise en relation	Description	Point fort	Point faible	exemple
Annotation du document	Lien entre document et commentaire	Annotation individuelle pour le dépôt de commentaire personnel	Respect d'un modèle individuel d'annotation	Activité limitée	XLibris
		Partage d'un commentaire sur un document partagé	Interface agréable	Pas de mise en relation des différents commentaires	Janno
		Analyse et extraction manuelle d'information du document via l'annotation personnelle	Mise en relation et réutilisation des annotations via des inférences dans des résumés	Annotation personnelle	Trellis
Annotation pour l'échange	Lien entre documents	Échanges via des annotations Mise en relation de lecteur de pages Web et de documents sur un serveur via des annotations	Fil d'annotation	Indexation minimaliste (date, auteur, type) Pas d'exploitation des annotations	Amaya Annozilla
		Échanges d'opinion via des annotations sur des documents partagés	Structuration argumentative du fil d'annotation	Pas d'exploitation des annotations	Tafannote

		Échanges d'opinion via des annotations sur des pages Web publiques	Annotation simple sans restrictions propriétaires Mise en relation par lien hypertexte Interopérabilité via le serveur CritLink	Lien entre document et utilisateur public Pas de création de communautés de travail, de pratique ou d'intérêt	CritLink
		Échanges d'opinion et indexation de pages Web	Fil d'annotation Visualisation adaptée à l'annotation personnelle	Indexation minimaliste (date, auteur, type) Pas d'exploitation des annotations	Yawas
Annotation pour l'interprétation collective	Lien entre utilisateurs via les commentaires	Échanges d'opinion et indexation de pages Web pour la navigation dans les documents	Fil d'annotation Choix de la sphère de publication	Indexation minimaliste (date, auteur, type) Pas d'exploitation des annotations au niveau discursif	ComMentor

		Échanges d'opinion et indexation de documents Web	Fil d'annotation Interface de qualité	Indexation minimaliste (date, auteur, type) Pas d'exploitation des annotations au niveau discursif	D3E
--	--	---	--	--	-----

Pourtant, même si ces outils d'annotation choisis pour leur représentativité et leur qualité se focalisent sur les interactions entre utilisateurs contrairement aux outils supports à l'annotation sémantique (index), ils ne permettent que rarement de relier les annotations entre elles, et ne permettent donc pas de constituer un ensemble homogène d'échanges entre des utilisateurs à propos d'un document. On retrouve ici le problème de définition de frontière entre le commentaire et le document, l'annotation et son ancre. Le résultat obtenu est ainsi un document (une page ou un site Web) d'une part, et d'autre part des annotations esseulées. Ces outils traitent majoritairement l'annotation comme un commentaire décontextualisé, sans lien avec les annotations de son entourage. Ils ne prévoient pas plus de fonctionnalités pour la récupération et la réutilisation d'annotations.

Si l'on reprend les idées développées dans un cadre herméneutique, il est possible de comprendre les extensions à apporter à ces catégories d'outils. Il existe plusieurs types d'herméneutique, liés aux différentes époques et différentes théories d'interprétation au fil des siècles. L'herméneutique romantique comprend deux dimensions, la compréhension et l'interprétation, puisqu'on ne comprend qu'en produisant soi-même du sens. Cette compréhension est vue par (Gadamer, 1960) comme un dialogue dans lequel le locuteur chercherait à s'impliquer. Ce dialogue ne s'arrêterait pas à l'interprétation comme dans l'herméneutique romantique, mais implique une dimension d'application. Suivant sa théorie, il est possible de comprendre l'annotation comme un élément du discours créant une interprétation et permettant une actualisation de cette interprétation dans la conception d'un nouvel objet, un autre texte par exemple. Dans notre optique, il est donc nécessaire de créer un lieu d'échanges homogène. Or, dans les outils d'annotation décrits jusqu'ici, c'est justement l'absence de soutien à l'élaboration de texte à plusieurs qui est à déplorer.

3. L'annotation élaborante

L'annotation élaborante est en fait le commentaire déposé par un annotateur à l'attention d'un pair et qui sera utilisé dans l'élaboration d'un nouveau texte. Il s'agit donc de considérer l'annotation comme une partie d'une version d'un texte qui peut être incluse dans une version à venir. Elle permet de communiquer d'une façon générale comme dans une optique de Web Social (WSO) ou de Communication Médiatisée par Ordinateur (CMO), et ainsi d'élaborer des concepts partagés et des connaissances via la production d'un document. Cette production coopérative de documents est soutenue aussi bien par des outils permettant une Communication Médiatisée par Ordinateur que par des outils de Rédaction Coopérative Assistée par Ordinateur (RCOA, CSCWriting en anglais).

3.1. Web Social

Le Web Social (WSO) est né de l'apparition d'une nouvelle génération d'outils destinée au soutien de la communication humaine. Il correspond à un ensemble d'applications du Web qui visent essentiellement à fournir des espaces de rencontre accroissant la conscience mutuelle entre les partenaires (mutual awareness, Dourish et Belotti, 1992) dans les interactions distantes. Sans s'inscrire dans un champ de recherche institutionnel, le terme WSO (Jordan *et al.*, 2003), (Reed *et al.*, 2004) s'est propagé sur le Web par le biais de journaux en ligne, Wikis et blogs. L'appellation Web Social (WSO) est directement liée au terme de *logiciel social* ou *relationnel* de (Shirky, 2003). L'objectif du WSO est la communication pour favoriser les liens

sociaux au travers d'échanges, de discours. L'annotation soutenue dans le cadre du WSo est donc une annotation discursive qui trouve sa place en tant que *post* (nom donnée à un message posté dans un forum) dans un forum ou mél d'accompagnement. Avec le WSo, nous avons affaire à des outils qui prônent une informatique communicante, donc dont la formalisation est peu contraignante et d'une grande richesse sémantique (au sens de la sémantique lexicale). Le réseau du WSo est un réseau social, il n'y a pas ici de problématique de classement pour favoriser une récupération de document, mais une problématique de commentaire, de glose grâce à des annotations postées dont le corps est constitué de termes. L'objet ancré est donc un signe (au sens linguistique) et l'ancre un fragment de document. Par exemple, une annotation peut être un message en langue naturelle relié à un fragment discursif. Grâce à ces annotations, un utilisateur de logiciel social peut effectuer des inférences cognitives sur des objets du monde, dans un cadre pragmatique. Il n'en résulte pas de visualisation en soi comme par un langage de balise, plutôt une représentation mentale créée par l'utilisateur, interprétable par exemple par les traces textuelles, les annotations discursives, des utilisateurs.

La discipline académique qui paraît complémentaire au développement d'applications du Web Social est la Communication Médiatisée par Ordinateur (CMO). La CMO propose des grilles d'analyse pertinentes pour des logiciels sociaux reposant sur l'utilisation de la langue naturelle.

3.2. Communication Médiatisée par Ordinateur

La Communication Médiatisée par Ordinateur (CMO ou CMC - Computer Mediated Communication - en anglais) est une sous-partie de la communication qui a pour étude la communication à travers l'outil informatique. La CMO, par son objet d'étude même, entre naturellement dans les thématiques du Web Social (WSo). (Herring, 2000) la définit comme l'interaction entre deux personnes ou plus, grâce à la transmission de messages entre ordinateurs branchés en réseau. En effet, la communication médiatisée diffère de la communication en face à face puisque entre autres, elle modifie l'homogénéité synchronique de la conversation. Les conversations asynchrones reviennent à un type de *conversations persistantes* comme présentées par (Erickson, 1999, p. 1). Elles ouvrent sur de nouvelles possibilités de gestion de la communication :

« Persistent conversations may be searched, browsed, replayed, annotated, visualized, restructured, and recontextualized, with what are likely to be profound impacts on personal, social, and institutional practices. ²⁵»

La CMO est en plein essor depuis le début des années 90 et touche aussi bien à l'éducation, au commerce et aux médias en général qu'à la coopération. Dans la mesure où l'une des questions qui se posent est : « comment les réseaux informatiques peuvent permettre à des humains de communiquer? », les technologies permettant une communication médiatisée recouvrent une partie des besoins du TCAO. Orientée vers l'analyse des technologies de la communication, la CMO s'appuie principalement sur les théories des Sciences Humaines et Sociales, comme l'analyse du discours par exemple. Internet est devenu pour la CMO un champ d'étude important avec l'apparition des *logiciels sociaux*, ces outils, le plus souvent en ligne, qui permettent de communiquer et de relier des communautés en ligne. Les outils permettant la communication médiatisée sont ceux qui permettent la communication à travers l'ordinateur (courrier, chat, ICQ, Web, etc.).

3.2.1. CMO et analyse du discours

Les approches linguistiques en CMO se cristallisent en 1995 avec l'expression de S. Herring « Computer-Mediated Discourse Analysis²⁶ ». Il s'agit de l'analyse (qualitative ou quantitative) du contenu d'unités linguistiques médiatisées par ordinateur (des lettres, des mots, des phrases, des messages, des tours de parole, des narrations, des échanges, des fils de discussion, des archives). Ces données peuvent être produites dans des conditions naturelles ou expérimentales. L'étude des données porte alors généralement sur différents niveaux de la structuration du message, et selon (Herring, 2000), principalement sur :

- la structure linguistique : la typographie, le choix d'un mot, l'organisation du message, par exemple ;
- le sens : des symboles, de l'échange, par exemple ;
- la cohérence interne : prise de tours de parole, réparations, décomposition thématique, par exemple ;
- la fonction sociale : l'identité, l'appartenance au groupe, la gestion de la face, le conflit et la négociation, par exemple.

²⁵ Les conversations persistantes peuvent être fouillées, visitées, rejouées, annotées, visualisées, restructurées et recontextualisées, ce qui peut avoir un impact important sur les pratiques personnelles, sociales et institutionnelles. (Traduction personnelle)

²⁶ Analyse de discours médiatisé

D'une façon plus fine, (Torres, 2001) définit l'ADMO (Analyse du Discours Médiatisé par Ordinateur) comme l'étude des manifestations pragmatico-linguistiques propres à la communication médiatisée. L'ADMO contribue à l'amélioration de la communication médiatisée par des apports divers. On peut noter l'analyse des identités qui se déploient sur Internet ou encore de la dynamique des groupes qui constituent les communautés virtuelles par exemple. Dans le domaine de l'ingénierie linguistique, l'ADMO permet d'explicitier les modalités selon lesquelles des personnes interagissent avec d'autres personnes et avec des ordinateurs. Elle participe alors au perfectionnement d'outils de reconnaissance vocale, ou de systèmes de chats (Lewkowicz et Marcoccia, 2004).

Dans le cadre de notre recherche, les productions liées à un outil d'annotation sont de l'ordre de discours médiatisés par ordinateur. Même si nous n'allons pas jusqu'à l'analyse de corpus de discours médiatisés par ordinateur, il est nécessaire de structurer la description du corpus selon les niveaux ci-dessus afin de prévoir une réutilisation possible des données. Nous présenterons donc nos données sous cette forme (chapitre VI).

3.2.2. CMO et Textualité

La CMO traite des genres de conversations de divers types et de tout thème ; du chat au storytelling en passant par le discours politique. Ces genres ont au moins un point commun : Les discours médiatisés ont un style nouveau. E. Guimier de Neef les nomme les NFCE ou Nouvelles Formes de Communication Écrite (Véronis et Guimier de Neef, 2006). Ce terme englobe les différentes formes d'écrit issues des Technologies de l'Information et de la Communication (TIC) dans lesquelles on relève souvent l'usage d'un langage *simplifié*. Cependant (Torres, 2001, section 4.5) note qu'il ne s'agit nullement d'une « transgression de la norme par méconnaissance », mais plutôt de l'usage d'un code simplifié (avec abréviations, liaisons, absence d'accents, etc.) qui permet « d'accroître la vitesse de production du message ». (Leroux, 1999) confirme cette *oralité* de la communication écrite spontanée. Les locuteurs des forums se *parlent par écrit* comme s'ils dialoguaient. Les frontières entre l'oral et l'écrit souvent reconnues comme floues, deviennent ici problématiques. Le texte écrit n'obéit alors plus aux règles classiques de clôture textuelle.

Ainsi, les bornes formelles d'un texte deviennent floues et la clôture du texte est discutable. Le médium numérique ne répond plus aux dichotomies classiques oral/écrit, clôture initiale/finale. Il peut être intéressant alors d'observer les concepts développés en praxématique. La praxématique préfère utiliser le terme de *mise en clôture* désignant « l'ensemble des

opérations mises en oeuvre par les sujets communicants pour structurer l'énoncé comme espace de production de sens » (Détrie *et al.*, 2001). Alors que la pragmatique se focalise sur l'actualisation en discours, c'est à dire l'opération concrète qu'effectue le sujet en acte de parole, la praxématique se concentre sur les processus par lesquels l'actualisation discursive d'un terme sélectionne une acception particulière parmi toutes les potentialités signifiantes capitalisées en langue (Lafont, 1989). S'intéressant aussi bien à l'oral qu'à l'écrit, la praxématique s'est attachée à saisir l'énoncé non pas comme un produit mais comme un processus dynamique qui correspond bien aux problématiques ambivalentes du discours médiatisé et offre un fondement théorique intéressant pour cette discipline.

La médiatisation de la communication comme du texte remet en cause les frontières du discours. L'annotation, toujours en marge du texte mais irrémédiablement lié à celui-ci, devient représentative des activités documentaires médiatisées par son statut même d'artefact de négociation de frontières. Elle est du texte, lié à du texte qui selon le choix de lecture est ou n'est pas constitutif du texte lu. L'annotation et le texte n'ont plus de frontière en CMO, tout message est intrinsèquement lié et n'a plus d'autonomie. Si le fil de discussion d'un forum disparaît, le message n'a plus de sens. Le message est lui aussi *secondaire*. De ce flou de la frontière du texte naissent aussi les problématiques modernes de visualisation des forums et de la cohérence thématique d'une conversation persistante. (Erickson, 1999) salue ainsi l'abandon dans les problématiques de visualisation de la conversation des représentations fondées sur la dichotomie contenu/structure au profit d'une représentation de la conversation comme un phénomène social.

3.2.3. Des annotations en CMO ?

Un fragment textuel est étudié en CMO par (Labbe et Marcoccia, 2005), le billet. Dans cet article, (Labbe et Marcoccia, 2005) présentent une genèse du mél issue du billet. Le billet appartient au genre épistolaire bref (Haroche-Bouzinac, 2000) dont la forme relève du dialogue écrit. Caractérisé principalement par sa brièveté, le billet est un écrit informel, informationnel, séquentiel et relationnel. C'est un document court mettant assez familièrement quelqu'un au courant d'un fait, ou faisant suite à une information reçue précédemment. Il peut aussi avoir une fonction strictement phatique. Le billet, comme le mél, met en relation différents actants de la communication : l'auteur du message, son/ses récepteur(s), et parfois un objet lié au message, par le biais du corps du message. Cette relation peut être implicite, portée par le sens du corps du message ; ou plus formelle, par l'ancrage physique, réel d'un message collé sur un colis, ou

un message accompagnant une pièce jointe. Suivant ces caractéristiques *génétiques* du mél dans le billet, nous considérons que l'annotation, en tant que fragment de document qui met en relation des documents, est elle aussi un élément aux fortes caractéristiques relationnelles, proche du billet et de certains usages du mél. Elle permet non seulement de structurer un document, mais aussi d'organiser plusieurs documents. Le billet est aussi le terme utilisé sur certains blogs et forums francophones pour parler du message posté sur le site (en traduction à *post*).

De même, T. Erickson ou M. Marcoccia étudient les échanges de plusieurs locuteurs dans des espaces de discussion en ligne, les forums. Le forum est structuré autour d'un thème, le fil de la conversation, déposé par un des participants, et construit par l'ensemble des contributions des lecteurs du fil par le biais de *posts*. Un post est un message, en général court, déposé en ligne. Il est possible d'y répondre, ce qui crée une série de messages reliés entre eux et liés à un thème central. Le fil de discussion est le fil central (thème) et ses messages. Un post cite souvent un autre post pour contextualiser la réponse et renforcer les liens au fil de discussion. Le forum est un outil qui permet la communication et le TCAO comme le montrent les travaux de (Barcellini *et al.*, 2005) sur les forums pour la conception d'outil Open-Source. Dans ces outils de CMO, les messages sont reliés à un fil comme une annotation est reliée à un document. (Marcoccia, 2001, p. 5) définit d'ailleurs le forum comme un « document dynamique produit collectivement et interactivement et dont la cohérence du contenu et de la forme est le résultat d'une gestion collective et coopérative ».

L'annotation élaborante est un discours ancré à un fragment (un document, une partie de document, un fil de discussion). Elle est à la frontière de celui-ci et le constitue sans être primordiale. En tant que discours secondaire, il est possible de comparer l'annotation traditionnelle au message électronique (mél ou post) accompagnant une pièce jointe, ou relié par une citation à un message précédent. L'annotation élaborante est rendue possible grâce aux outils de la CMO et il existe ainsi une herméneutique médiatisée par ces outils.

3.3. La boîte à outils de l'annotation élaborante

Les outils d'annotation du WSo sont divers (type chat Messenger, Yahoo, IRc, Wikis, ...). La communication étant médiatisée, nous avons affaire à des conversations persistantes pérennisées via un document numérique. Nous les présentons en faisant l'économie de différencier les logiciels sociaux d'outils de Rédaction Collaborative. Ces différents outils sont

soutenus par des techniques de synchronisation des modifications mises en place pour permettre de rédiger des documents en coopération. Nous les présentons ensuite.

3.3.1. WYSIWYG, blog, Wiki et CMO

Les outils d'annotation sociaux sont en général très simples d'usage. Plusieurs outils se focalisent sur l'interfaçage Web pour l'édition et la diffusion publique d'informations personnelles ou communautaires afin de faciliter la création personnelle sur le Web. Une catégorie très connue des outils du Web Social est celle des WYSIWYG (What You See is What You Get) permettant par une interface simple et des fonctionnalités intuitives de conception, de développer des pages Web comme on rédige un document avec un logiciel bureautique. Cette facilité d'usage a permis à plusieurs outils de proliférer sur le Web, principalement des forums, des blogs et des Wikis (ou parfois des composés comme Bloki²⁷ ou XWiki²⁸ (Fig. 5). Les Wikis (Delacroix, 2005) servent à la fois pour la conception coopérative de site Web et l'écriture coopérative grâce aux blogs et forums souvent liés à ces sites. Ces outils sont très étudiés et nous vous renvoyons aux liens et auteurs cités pour plus de détails.

²⁷ <http://www.bloki.com/>

²⁸ <http://www.xwiki.org/xwiki/bin/view/Main/WebHome>

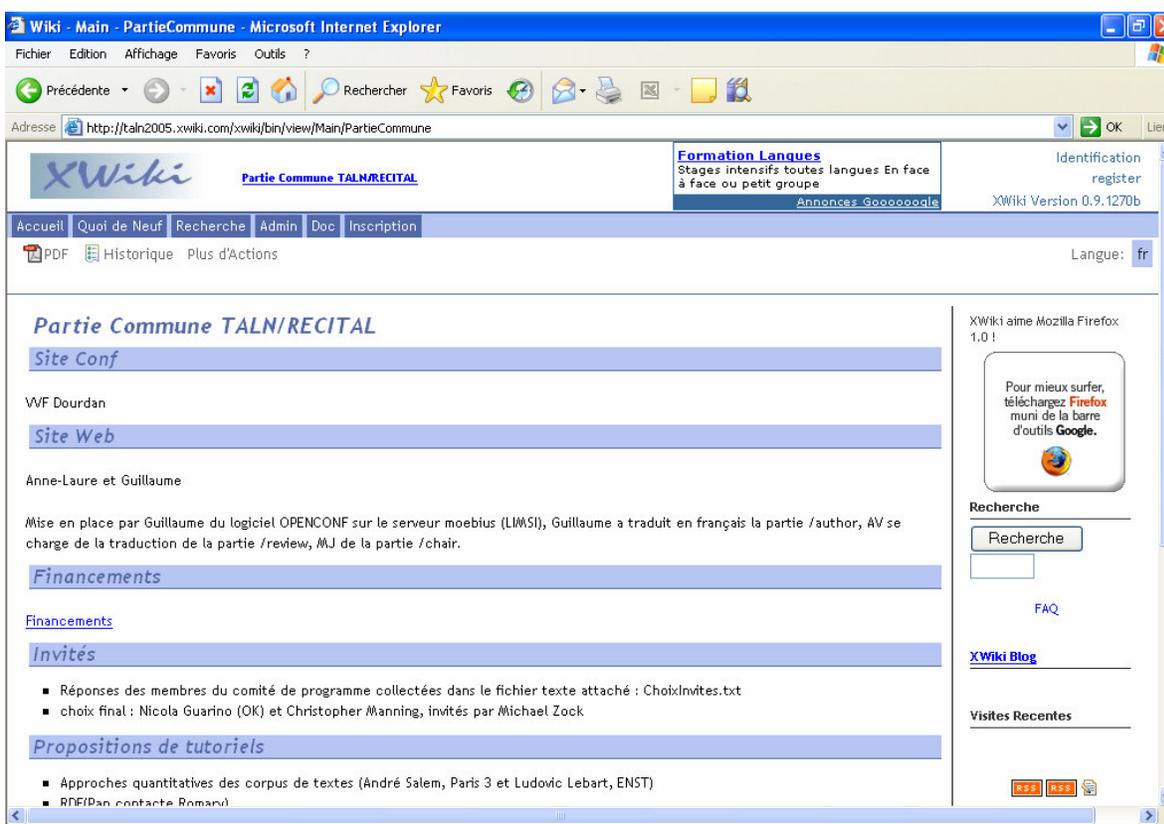


Fig. 5. PAGE D'ACCUEIL DE XWIKI

Ces outils améliorent l'édition en ligne, mais ne gèrent pas réellement la coopération entre les utilisateurs. Pour cela, il est important de se tourner vers des outils permettant la révision de documents électroniques comme BSCW (Bentley *et al.*, 1995), REDUCE (Real-time Cooperative Editing, (Sun *et al.*, 1997)), ou encore Coword (Sun *et al.*, 2004) établi selon les principes et techniques de REDUCE.

3.3.2. Des outils de rédaction coopérative

La Rédaction Coopérative Assistée par Ordinateur met au centre de ses recherches l'écriture en collaboration, naissant d'échanges écrits entre divers participants à cette tâche. Un co-écrivain est un utilisateur qui souhaite écrire en coopération avec d'autres utilisateurs d'un groupe (projet) ou d'une communauté. Dans ce domaine plusieurs méthodes et outils ont été proposés en vue d'aider l'utilisateur à travailler, rédiger, de manière coopérative (Tab.3). Nous nous intéressons plus particulièrement aux outils ayant un client Web, qui permettent de travailler en groupe de manière distribuée et/ou asynchrone autour de documents partagés. Les

outils les plus saillants vont être présentés ci-dessous. Ils sont en général centrés sur un espace partagé pour différents objectifs, principalement la discussion, l'apprentissage et le travail coopératif. L'objectif de la plupart est de permettre les échanges en augmentant la conscience mutuelle (*mutual awareness*, Dourish et Belotti, 1992) dans le groupe. Ils se focalisent pour la plupart sur des techniques permettant la création (*Web authoring*) multi-utilisateur de documents Web.

3.3.2.1. La gestion du temps et du document

Un des principes de base dans la communication des outils de CSCWriting est la mise en place du protocole WebDAV (Web-based Distributed Authoring and Versioning). Il permet une extension du protocole HTTP qui permet aux utilisateurs une édition coopérative et la gestion de fichiers sur des serveurs Web distants. Plusieurs outils de co-création utilisent ces principes tel que PINAS (voir ci-après).

3.3.2.2. Les clients de RAO

Les clients de RAO, présentés ci-dessous et détaillés en annexe, permettent de créer et/ou d'éditer des documents en ligne. REDUCE, Pinas, BSCW ou DocReview partagent des fonctionnalités d'édition et de révision collective. Quick Topic et Doc Review possèdent aussi un forum, un salon réservé au groupe. Ces outils établissent une coopération entre des utilisateurs par le simple partage de document mais aussi parfois par une communication via des notifications aux membres du groupe (comme proposé par des outils propriétaires du type Lotus Notes ou Quick Place). Enfin, Compendium permet de prendre des notes au fil de l'eau et de les organiser suivant des principes de cartes de concepts s'appuyant sur l'argumentation.

Tab. 4. GRILLE DE CLASSEMENT DES OUTILS DE RÉDACTION COLLABORATIVE

Catégorie	Point de mise en relation	Description	Point fort	Point faible	exemple
Gestion de documents	Espace de stockage de document numérique partagé	Plateforme de stockage pour le partage de documents	Travail en temps réel	Pas de gestion des échanges entre utilisateurs par rapport au document Pas de suivi des modifications des documents	Piñas
		Espace de travail partagé	Gestion des droits des utilisateurs, notification	Pas de entre le commentaire et le document	BSCW
		Éditeur de texte coopératif	Édition concurrente	Pas de gestion des échanges entre utilisateurs	REDUCE
Gestion d'utilisateurs	Document partagé	Éditeur de texte coopératif pour la révision	Gestion des documents Lien entre commentaire et document Résumé des commentaires Notification au groupe	Pas de gestion des échanges eux-mêmes Pas de typage des commentaires pour leur gestion (géographique seulement)	DocReview

Catégorie	Point de mise en relation	Description	Point fort	Point faible	exemple
		Espace d'échange en ligne de commentaires à propos d'un document	Notification par liste de diffusion avec commentaires et document envoyés	Commentaires par paragraphe (pas par tag ou mot)	Quick Doc Review
		Plug-in de l'éditeur de texte Word pour la révision coopérative	Gestion des modifications de différents utilisateurs Interface connue	Pas de gestion des échanges eux-mêmes Pas de typage des commentaires pour leur gestion	CoWord
		Outil de structuration argumentative de commentaires interactifs en temps réel ou lié à une chronologie	Rendu des échanges selon la chronologie et l'argumentaire des événements Navigation et manipulation de documents non pérennes	Pas de relation avec des documents annexes Transcription et structuration d'une seule conversation, sans documents de référence Typage argumentatif seul Point de vue d'un seul	Compendium

Catégorie	Point de mise en relation	Description	Point fort	Point faible	exemple
			Interface de qualité	utilisateur (celui qui structure)	

L'intérêt des outils de RCAO se situe non seulement au niveau des techniques utilisées, architectures clients/serveurs, peer to peer pour la révision synchrone distribuée, *authoring* (création) ou encore synchronisation des versions (protocole WebDAV), mais aussi au niveau des interfaces facilitant le travail cognitif de l'utilisateur humain. Cependant, ces outils ne permettent que rarement de suivre la logique de conception des documents et n'utilisent que du versionnage (versioning) de documents très simple (date, comparaison de documents) et ne proposent pas une organisation fine des modifications apportées. Bien qu'ils offrent parfois des possibilités d'annotations du document (mais pas sur des fragments), ils n'offrent pas de possibilité de lier différents fragments par un même index ni même entre eux.

4. Conclusion

Les tableaux suivants présentent de manière synthétique les activités permises par les fonctionnalités des outils présentés dans ce chapitre. Le premier (Tab. 5) permet de comparer les outils développés sur les bases des activités nécessaires au WS dans un objectif soit de communication et de discours (annotation discursive), soit d'indexation (annotation sémantique). Il met en valeur les genres différents d'objets ancrés, de la balise à un fragment de discours selon ces objectifs. Dans le second tableau (Tab. 6), les outils sont répertoriés en fonction des deux grands objectifs qu'ils soutiennent, soit le partage d'un espace commun, soit le partage strict d'un document commun. Dans un cadre médiatisé, ces deux objectifs reviennent au partage d'un document. En effet, l'édition et la publication coopérative d'un espace à partager équivalent à l'élaboration collective d'un document numérique.

Tab. 5. TABLEAU COMPARATIF DES OUTILS D'ANNOTATION EXISTANTS

Objectif	Activité	Activité instrumentée	Objet ancré	Ancre	Outils
Communication Discours	Commentaire	Commentaire Postage	Texte langue naturelle Signe	Fragment discursif	Outils d'annotation discursive
Récupération d'information	Classement	Indexation automatique Étiquetage automatique Annotation automatique	Index Balise Code	Code/données	Outils d'annotation sémantique

Tab. 6. TABLEAU COMPARATIF DES OUTILS EXISTANTS POUR LE CSCWRITING

Objectif	Activité	Activité médiatisée	Spécifications	Espace d'actions	Outils
Création d'un espace partagé Communication	Édition collaborative	Dépôt de documents Créer des pages Web Échanges autour des documents	Interface Web Coopératif asynchrone Architecture distribuée	Public Authentic ation	Wiki Blog Forums

Objectif	Activité	Activité médiatisée	Spécifications	Espace d'actions	Outils
Création de documents partagés Communication	Révision et coopération	Poster des commentaires sur une page Web partagée Notification au groupe	Interface propriétaire ou Web Coopératif synchrone ou asynchrone Architecture distribuée	Espace de groupe privé Nécessité d'authentification	Lotus Notes Quickplace BSCW Project/Course Form Compendium DocReview REDUCE COWORD

Pour soutenir la coopération, l'annotation doit soutenir nombre d'activités présentées ci-dessus. Peu d'outils soutiennent cette annotation sous ses diverses facettes. Nous trouvons d'une part, des outils d'annotation sémantique ou discursive qui s'efforcent de combiner des objectifs annotatifs. Cependant, tout se passe comme si un forum était associé au document. Or, de nombreux travaux soulignent que les discussions en ligne sont souvent désorganisées et confuses, à cause du développement fréquent de multiples fils de discussion et de conversations parallèles. Nous pouvons citer par exemple (Maccoccia, 2004), qui évoque le phénomène de digression thématique à l'intérieur d'un forum, qui se fait progressivement, en parcourant une chaîne de messages introduisant chacun un développement thématique par rapport au message précédent. Le résultat peut être une véritable *décomposition thématique* (topic decay, Herring, 1999). Ce phénomène est largement à prendre en compte dans la mise en place d'un outil d'annotation pour la coopération. Et d'autre part, nous trouvons des outils de RCAO qui permettent intrinsèquement une coopération entre plusieurs utilisateurs autour de documents, mais ne permettent pas de récupération de documents fine et permettent rarement de lier différents fragments à un document par un même index ni même entre eux ou encore d'intégrer un fragment à un document.

Ainsi, après avoir observé différentes définitions de l'objet annotation dans les champs du Web Sémantique, de l'Herméneutique et de la Communication Médiatisée par Ordinateur, nous constatons que des outils y correspondent pour lesquels l'accent est mis soit :

1. sur la production collective de texte, non liée à un texte *d'origine*, cette finalité étant plus ou moins mise en avant (forum vs RCAO)
2. sur le commentaire de texte existant, sans se préoccuper d'élaborer un *tout* à partir de ces commentaires, sans gérer la communauté qui se crée autour de ces commentaires (outils d'annotation discursive)
3. sur l'indexation rendue possible par les annotations (outils d'annotation sémantique).

Ces lacunes sont dues, selon nous, à l'absence de modèle de description de l'activité d'annotation sous-jacent. De là, les outils développés ne soutiennent qu'une partie de l'annotation et restent faibles pour d'autres pratiques annotatives. La plupart des outils souhaitent soutenir le dépôt et le stockage d'*entité* annotations sans définir si cette annotation est sémantique, discursive, ou élaborante, ni si elle est coopérative ou individuelle et surtout sans prendre en compte que l'annotation en tant qu'objet est la production d'une activité d'annotation. Annoter c'est avoir une pratique d'écriture alors que l'on est lecteur, c'est suivre un processus cognitif complexe orienté vers l'échange avec l'Autre.

Or, dans notre problématique de gestion de DoPAs au sein d'un groupe projet, ces trois besoins coexistent. L'outil à définir ou à réutiliser doit donc prendre en compte ces divers besoins.

Chapitre II

Proposition d'une définition de l'annotation pour la coopération

- La science ! Il n'y a que des savants, mon cher, des savants et des moments de savants. Ce sont des hommes... des tâtonnements, des nuits mauvaises, des bouches amères, une excellente après-midi lucide.
Paul Valéry

Dans les pratiques d'écriture médiévales, l'annotation est un élément de partage d'interprétation par plusieurs lecteurs. Quand l'imprimerie met à disposition des copies personnelles, l'annotation devient plus individuelle que partagée (Wolfe, 2002). Elle ne permet plus que de marquer une page, de mémoriser un emplacement, une citation, de noter des idées émergentes. L'annotation pour la coopération disparaît. (Marshall, 1997) montre pourtant un intérêt pour les documents annotés par ses pairs. L'annotation n'est plus destinée à l'autre, mais l'échange d'exemplaire permet un nouveau type de partage. C'est ce phénomène d'échange qui est apparu aussi dans les lieux publics urbains avec le BookCrossing²⁹, l'oubli volontaire de livres que l'on souhaite partager. Les pratiques d'échange de documents physiques marquent cet intérêt toujours vivace de partage. Avec les documents numériques, cet échange est facilité et peut être très étendu. Le document numérique permet ainsi le renouveau des pratiques d'échange et donc, des pratiques d'annotation.

Dans ce chapitre, nous présentons l'annotation coopérative sous trois angles : tout d'abord, l'annotation comme un index, puis comme un fragment de conversation et enfin comme un fragment de document. Nous formons en dernier lieu un continuum de l'annotation et postulons que tracer ce phénomène progressif revient à tracer la logique de communication d'un projet en conception.

²⁹ <http://www.bookcrossing.com/>

1. Un lien vers le document

L'indexation se fait à base d'index. Un index doit indiquer une direction, montrer un emplacement. Pour indexer un document, il suffit de placer un index, c'est-à-dire un petit objet mobile qui sert à distinguer un document dans un ensemble. Tout comme le marque-page ou le signet servent à distinguer une page dans un livre, l'annotation sert à repérer un passage dans une page. Ainsi, cette idée d'indicateur est inhérente à l'annotation. En effet, comme le souligne (Fogli *et al.*, 2005) l'annotation est ancrée à une cible qui est sa base. Une cible est l'objet auquel l'annotation se rapporte (son ancre physique et le contenu du texte d'ancrage). Cette cible peut être une collection de documents, un document, une ou des parties de document, quelle que soit la nature du document (texte, photo, musique, objet, ...). Dans (Marshall, 1998b), C. Marshall effectue un relevé détaillé des points d'ancrage des annotations dans un texte. La première technique de mise en relation graphique est formelle (Marshall, 1998a) et est constituée d'une flèche, ou ligne pour lier l'annotation à sa cible. La seconde utilise des parenthèses, des crochets, des accolades ou ce type de marques pour associer un commentaire au texte. Vu le flou que laisse une accolade dans une marge, le lien effectué ici devient plus flottant, plus informel et non explicite (Marshall, 1998a). En dernier lieu, le lecteur se fonde sur la proximité de l'annotation avec une partie du texte pour les mettre en relation. Dans ce dernier cas, l'ancrage, c'est-à-dire le point de départ ou d'arrivée de la relation entre l'annotation et le document a des variations importantes. Il va sans dire que la plupart des notations ne sont ni explicites, ni formelles et qu'elles obéissent à des besoins cognitifs peu stables de mise en relation, de mise en exergue ou d'aide-mémoire incident. Cette relation d'un objet à un autre à un emplacement temporairement fixe a inspiré en informatique l'adressage et l'indexation des données.

Pour notre propos, nous qualifions d'indexation ce que J. Virbel note classification. Selon nous le terme classification recouvre une autre réalité que l'indexation et exprime plutôt la mise en place d'une catégorisation avec des termes utilisables ensuite pour indexer ce qui permet de classer. Indexer semble être un terme regroupant une plus large réalité. De plus, l'indexation, comme la classification, peut être hautement coopérative comme le montrent (Koivunen et Swick, 2003). Ces auteurs prônent l'échange de signets, de marque-pages, pour une certaine interopérabilité qui peut alors être humaine. (Fogli *et al.*, 2005) suivent aussi cette idée dans la mesure où ils reconnaissent une possibilité de convention tacite sur l'ancrage de l'annotation, c'est à dire un ancrage qui n'a pas de sémantique explicite mais est comprise dans une

communauté. « That is not semantically explicit but represented by symbols, understood only by the members of a community who agreed on its meaning³⁰ ». Il s'agit d'une ancre conventionnelle définie comme les étiquettes normées décrites dans un langage de balises. Nous sommes alors dans un cas d'indexation pour la coopération puisque définie et validée par les pratiques d'une communauté.

L'indexation grâce aux annotations permet une flexibilité dans la relation et dans le type de classement effectué. Elles peuvent être liées à une classification formelle (comme certaines Ressources Termino-Ontologiques - RTOs -) ou informelle (ethnoclassification). L'étiquetage des données contient alors lui-même une information laissée par l'annotateur (qui commence par le choix de la classification utilisée). Cette annotation, selon qu'elle est plus ou moins formelle permet de contextualiser le document annoté en le mettant en relation avec des informations non documentarisées, ou lui offrir un simple cotexte en le mettant en relation avec d'autres documents (textuels, graphiques, archive...) ou avec les éléments d'une classification (qui est documentarisée dans un système informatique).

C. Marshall est une des pionnières de la recherche sur les annotations dans le cadre de documents numériques. Elle travaille principalement sur le développement de nouvelles fonctionnalités pour les pratiques autour des documents numériques. Dans (Marshall, 1997), elle propose une typologie des annotations. Selon elle, les annotations laissées dans des ouvrages à usage scolaire sont de six types :

4. Des signaux de procédure (Annotations as procedural signals) ;
5. Des signets et des aide-mémoire (Annotations as placemarkings and aids to memory) ;
6. Des explicitations dans le contexte de la lecture (Annotations as in situ locations for problem-working) ;
7. Des traces d'une activité d'interprétation (Annotations as a record of interpretative activity) ;
8. Des traces de l'attention portée au cours de la lecture (*Annotations as a visible trace of the reader's attention*) ;
9. Des traces des réflexions incidentes (Annotations as incidental reflections of the material).

³⁰ Ce n'est pas sémantiquement explicite mais représenté par des symboles, compris uniquement par les membres d'une communauté qui se sont mis d'accord sur leur sens. (Traduction personnelle)

Il est vrai que les annotations sont les traces d'une activité de lecture active (Schilit *et al.*, 1998) et de classification (signet). Elles permettent de faciliter les interactions avec les documents numériques. C'est dans cette simple optique que C. Marshall s'intéresse à la médiatisation des annotations. Dans cette optique, les annotations permettent des interactions entre un acteur et un document.

D'un point de vue coopératif, ces travaux tendent à oublier l'importance de l'adresse du but annotatif. Une annotation est déposée dans un but pour un lecteur. Pour pallier cette négligence (Wolfe, 2000) propose une description des annotations en fonction du rôle de l'annotateur (auteur ou non du document annoté) et du rôle des lecteurs visés (l'annotateur lui-même, l'auteur du document annoté ou d'autres lecteurs).

Dans un cadre médiatisé, l'annotation possède un statut attendu mais nouveau, de nature transitoire. En effet, les marques faites au document sont idéalement séparables du document. Elles sont aussi transférables d'un document à un autre, ou effaçables. Dans la même optique, elles peuvent désormais avoir une vie propre. Le principe d'une annotation que l'on peut supprimer, déplacer ou singulariser remet en cause l'indexation informatique comme une balise ancrée physiquement à un document qu'elle modifie. L'annotation transitoire milite en faveur des classifications informelles et de leur mise à jour aisée.

De plus, alors que les membres d'une communauté coopèrent et échangent autour de documents partagés, ils doivent en permanence classer leurs documents, nonobstant leur statut transitoire. Ainsi, des documents d'archive et des fragments de documents ou *proto-documents*³¹ doivent être classés. Les membres doivent donc s'accorder sur une classification qui de surcroît évoluera au fur et à mesure de leur activité et de l'élaboration de nouveaux documents. La classification choisie sera donc significative dans un contexte spécifique, à un instant défini d'un état du projet et selon le point de vue de ses membres à cet instant.

1.1. Folksonomies, ethno-classifications et classifications vulgaires

Cette problématique de classification populaire (ou folksonomy) n'est pas nouvelle. Dans les réflexions sur les bibliothèques numériques (BN), en 1996, Susan Star (1996) souligne

³¹ Collection de matériels documentaires, plus ou moins cohérents et organisés (Pédauque, 2006)

l'intérêt de rendre la sérendipité³² aux BN. (Star, 1996) parle alors d'ethnoclassification pour qualifier les schèmes de classification mis en place individuellement par chaque utilisateur et fusionnés aux schèmes de classification proposés par les documentalistes. La recherche d'information est un processus de sélection actif. Les utilisateurs « prennent et adaptent » pour trouver l'information dont ils ont besoin. Les BN pour être accessibles aux utilisateurs et leur permettre la sérendipité dans leurs recherches, doivent donc comprendre ces schèmes de classification développés par l'utilisateur pour les intégrer dans des classifications plus formelles utilisables en documentation. Là où S. Star évoque l'ethnoclassification, Peter Merholz (2001) parlera plutôt de thésaurus vernaculaires à propos des thésaurus développés pour l'indexation de site Web communautaires. Il s'agit des stocks de mots mis en place dans une certaine communauté pour indexer ses informations. P. Falzon (1989, 1991) dans ces travaux sur les vocabulaires professionnels souligne que dans des contextes communautaires, de nombreux raccourcis sont utilisés entre collaborateurs. En introduisant des phénomènes d'abréviations (Ribeiro, 2006), d'ellipses et de métonymies, les collaborateurs forment un vocabulaire spécifique à leur groupe (lol, bjz) et à leur domaine (« la note pour l'andouillette et le banana split du client de la table 2 » → « l'addition de l'andouillette de la 2 »). Ces phénomènes sont abordables par une communauté tant que les conditions de prédiction sont satisfaites. Lors de la mise en place de schèmes de classification dans une communauté, les ethnoclassifications, ou folksonomies, apparaissent de même, souvent fondées sur le sociolecte partagé dans cette communauté. P. Merholz décrit le terme de T. Vander Wal, folksonomy, qui émerge du Web Social (WSO) puisque le rappel à taxonomy évoque une classification hiérarchisée (type de relation est-un) alors que la folksonomie n'a pas de relations de ce type et que l'étiquetage de ressources n'impose pas d'avoir affaire à une taxinomie. T. Gruber (2005) parle des folksonomies en tant que « the emergent labeling of lots of things by people in a social context³³ », mais considère l'apport de cet étiquetage bottom-up comme important par sa complémentarité avec l'approche en général top-down des ontologies formelles. Les folksonomies s'apparentent ainsi à l'indexation plein-texte (un terme potentiel par dimensions du terme) avec en plus la sélection active d'un index par un utilisateur. Cependant, T. Gruber signale que pour partager ces étiquettes, interopérer ou coopérer dans un écosystème de ressources, il est nécessaire de construire une infrastructure, une ontologie. Grâce à cette ontologie, il serait possible de faire de l'étiquetage coopératif ou du filtrage coopératif selon les étiquettes, alors que les folksonomies ne fournissent que des informations individuelles. Ainsi, la

³² traduction de l'anglais *serendipity*, le don de faire des découvertes heureuses et inattendues par accident, hasard.

³³ L'étiquetage émergent de beaucoup de choses par des gens dans un contexte social. (Traduction personnelle)

folksonomy différerait de la vision d'ethnoclassification puisque cette dernière est établie selon les pratiques d'une communauté et non selon la pratique de chacun des intervenants du Web.

Dans ce contexte plus réduit, on peut considérer (à la suite de (Chibout *et al.*, 2003) dans leurs travaux sur les systèmes de TCAO) que l'ethnoclassification ne porte pas forcément d'ambiguïté puisque « les interlocuteurs partagent un même contexte d'interprétation et les mêmes connaissances sur le domaine ».

Les deux systèmes d'étiquetage libre les plus connus sont del.icio.us³⁴ (Schachter, 2004), pour gérer ses signets de pages Web, et Flickr³⁵ une application de partage et de gestion de photographies. En figure 6, les étiquettes les plus courantes³⁶ de del.icio.us sont disposées par taille.

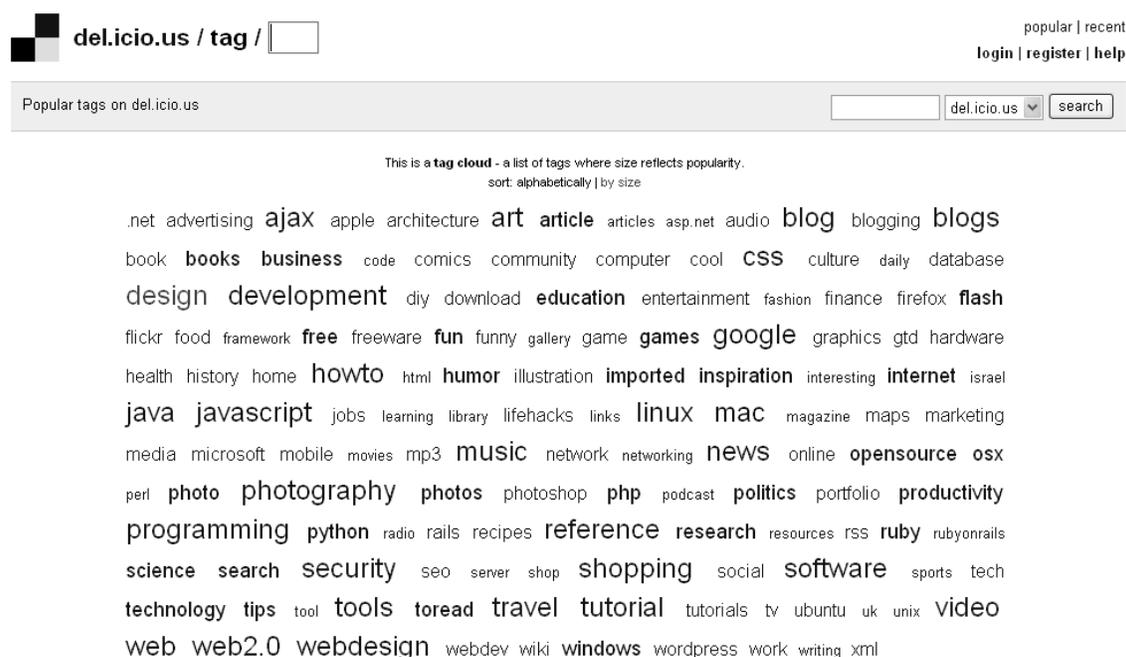


Fig. 6. POPULARITÉ DES ÉTIQUETTES DE DEL.ICIO.US

Vous trouverez sur le gestionnaire de partage de photographies Flickr près de 2 934 360 photographies de moi, c'est-à-dire plus de deux millions de photographies dont l'étiquette est *me*. A. Mathes fait une étude des étiquettes fréquentes dans Flickr dans (Mathes, 2004). On

³⁴ <http://del.icio.us/>

³⁵ <http://www.flickr.com/>

³⁶ Au 15/08/06

retrouve bien entendu les thèmes fréquents et attendus de photographies (cat, friends, dog, sky, sea, park, kids, garden, baby, building, flower, vacation), beaucoup de noms de lieux (Troyes, Goussainville, Coulanges sur Yonne) et d'années (2001, 2002,...). On retrouve aussi nombre de termes spécialisés du domaine photographique (portrait, landscape, contre-plongée). Le trait intéressant de ce relevé est de voir l'utilisation de termes relevant de la néologie tels que ceux des nouvelles technologies, « cameraphone, moblog, fotolog ». Il est visible ici que les classifications sont mises rapidement à jour et sont donc adaptés à des contextes d'innovation comme en conception.

1.2. Attribution du sociolecte

1.2.1. Schibboleth

Schibboleth est un terme hébreu qui permettait d'identifier un individu comme appartenant à une communauté à sa façon de le prononcer. Ce terme est différentiel et rappelle le vocabulaire partagé dans une communauté. C'est le cas des langues de spécialités et des langages techniques.

Wüster (cité dans Slodzian, 2006) s'est le premier penché sur la dimension désignative (dénommer les objets et les qualités) de la terminologie technique. Wüster rêvait d'un langage consensuel construit sur la base d'un accord entre spécialistes du monde entier. « Sa conception du concept est comprise comme résultant d'un consensus arbitraire entre spécialistes (conventionnalisme) et comme le résultat non critique d'une application de la logique du premier ordre (réduction logique aux caractéristiques) » (Slodzian, 2006). Or, l'instabilité des paradigmes scientifiques et le mode de production de connaissances, anarchique et non systémique des domaines techniques pose problème face à la fixité des systèmes terminologiques.

L'assimilation du terme au concept en Ingénierie des Connaissances donne une actualité nouvelle au débat. La notion de terme comme unité de connaissance supérieure au mot et celle de signification (les significations ne sont pas des objets auxquels des mots sont attachés) sont remises en question dans différents domaines. La sémantique linguistique et textuelle tente de récuser non seulement le postulat référentiel posant le terme comme référent à un concept ou à un objet, mais aussi les approches fonctionnalistes des langues de spécialités qui sont insuffisantes puisque les textes nécessitent une caractérisation par une approche multicritères,

mettant en jeu les différents paliers du texte. Dans la structuration d'une RTO, il faut prendre en compte une focalisation de rang supérieur qui n'est pas forcément représentée par les termes des documents, du même ordre que la normalisation comparable à un point de vue de domaine ou de la tâche (Bachimont, 2000).

Les outils d'annotations présentés en chapitre I section 2 établissent le plus souvent leur classement sur des ontologies construites par un expert du domaine, réutilisées ou encore construites semi-automatiquement à l'aide des outils statistiques. Le partage de ce type de RTOs formelles n'est pas facile pour les utilisateurs puisqu'elles ne correspondent ni aux objectifs ni au domaine des utilisateurs. Ces RTOs ne sont pas représentatives de la classification en place dans la communauté ni même de la terminologie utilisée dans la communauté. De plus, elles acceptent difficilement l'ajout ou (la suppression) de nouveaux concepts et donc supportent mal la néologie. En effet, les membres d'un projet ou d'une communauté ne sont pas liés uniquement par un thème commun, un objectif partagé, mais aussi par un langage partagé parfois implicitement, le sociolecte du groupe (Marcellesi et Gardin, 1974). Le sociolecte est non seulement le lexique partagé représentatif du groupe social, mais aussi les connotations sur les termes que les membres partagent. Ce sociolecte se crée au fur et à mesure des échanges dans le groupe, liant à la fois la terminologie technique des différents acteurs du projet et celle qui sera construite pour les besoins du projet. Parallèlement à la négociation de frontières des domaines et des concepts (Lee, 2005), un vocable adapté est mis en place et doit être représenté par la RTO utilisée dans le projet. La RTO permet alors aux utilisateurs de partager un référentiel commun via une classification partagée et de se reconnaître dans une communauté de langage. Ce sociolecte n'est pas toujours explicité dans les groupes projet. Les membres des diverses communautés coopèrent alors sans retour critique par un langage qui n'est pas vraiment partagé. B. Stiegler à la suite d'une étude sur les lecteurs professionnels de grands corpus à la Bibliothèque Nationale de France (Stiegler, 2000, p. 1) observe que ces lecteurs-scripteurs construisent une sémantique située. Il pose alors l'hypothèse que « le système de navigation n'est optimal que lorsqu'il est construit par l'utilisateur lui-même, ou par la communauté des utilisateurs qui devient une société d'auteurs lorsque la construction du système de navigation est assurée par l'usage collectif d'un dispositif d'annotation électronique de documents ». Dans le cadre de notre terrain en conception, nous souhaitons offrir aux membres de la communauté de pratique la possibilité de classer leurs annotations, documents et thèmes de travail par le biais d'un vocabulaire partagé. Finalement, la définition par une communauté de pratique d'un schème de

classification représenté par une RTO permet de saisir les concepts qu'ils utilisent et de les partager explicitement.

Ce langage partagé fonctionne comme une classification sociolectale. Ces RTOs moins formelles, aisément mises à jour au fil de l'eau, permettent d'identifier le processus de création de connaissances par l'observation des nouveaux descripteurs (des nouvelles entrées) ajoutés à la RTO pour la description d'une collection de documents partagés. Les classifications sociolectales peuvent alors permettre de renforcer la communauté non seulement parce qu'elles permettent à leurs membres d'utiliser les termes qu'ils considèrent représentatifs de leurs activités et de leurs domaines, mais aussi parce qu'elles sont évolutives et ouvertes, ce qui permet à une communauté de construire son propre schème de classification et de le mettre à jour au fur et à mesure que son activité évolue. Le partage explicite et négocié d'une classification commune permet aux membres d'un projet d'avoir une activité critique sur leur activité de conception et de suivre la logique de communication du projet. Pour être utilisée de la sorte, notre classification doit tout de même posséder une structure comme préconisée par Gruber. Bien que semi-formelle, elle reste focalisée sur le terme comme les thésaurus.

Dans un cadre de terminologie textuelle, la problématique de la terminologie est déplacée sur les problématiques de relations entre signifiés et de spécificité du fonctionnement des signifiés dans les textes à caractère technique et scientifique. Nous nous appuyerons donc sur les méthodes de la linguistique de corpus pour proposer des listes de candidats termes, sans a priori ontologique. De même, dans une optique plus explicitement non référentialiste, nous ne nous limiterons pas à un usage systématique des relations sémantiques canoniques – incontestablement conceptuelles – dans la constitution de notre RTO et tenterons de prendre en compte aussi bien les relations verticales canoniques que les relations horizontales. Les relations horizontales existent aussi bien au niveau du terme (- co-texte - dans les textes) qu'au niveau de la communauté (- contexte - dans la situation sociale). Le contexte classificatoire des relations horizontales est représenté par les points de vue qu'un utilisateur porte sur son activité et que nous devons donc prendre en compte.

1.2.2. Point de vue

Comme montré dans la section précédente, une terminologie spécifique est utilisée au sein d'une communauté et un schème de classification est développé, fondé sur les connaissances de ses utilisateurs. Puisqu'il est impossible de prédire les connaissances et les schèmes de classification individuels de chaque utilisateur, les membres d'une communauté

doivent avoir la possibilité de construire leur classification et de définir l'étendue de leur(s) point(s) de vue sur un terme, c'est-à-dire la manière dont ils comprennent un concept. Ceci est particulièrement vrai au cours d'activités multidisciplinaires (voir par exemple Bénéel, 2003) comme engagées dans notre terrain de projet de conception mécanique. Un utilisateur peut adopter plusieurs points de vue selon qu'il parle en tant que tel ou tel représentant du domaine. Un utilisateur peut avoir plusieurs rôles au cours d'un projet ; des rôles techniques (mécanicien, chercheur) mais aussi des rôles organisationnels (technicien, concepteur, chef de projet, etc.) ou encore des rôles personnels dus à sa formation, son expérience (un concepteur junior, un expert en mécanique cinématique, de conception, aéronautique, etc.). À chaque fois qu'un membre du projet décide de commenter un document, et d'échanger avec les autres membres, il le fait selon un rôle, en incluant potentiellement des arguments à son commentaire. Le commentateur laisse donc selon son commentaire des traces de planification (prise de rendez-vous, jalons du projet, etc.) ou des traces d'une argumentation (pour et contre, négociation, avantages et inconvénients d'une solution, prise de position, etc.). Ainsi, il peut inclure dans l'échange un point de vue argumentatif ou de planification.

Pour ces diverses raisons, et suivant (Zacklad *et al.*, 2003, et Cahier, 2005), nous considérons que le contexte de classification doit être représenté par plusieurs points de vue, parmi lesquels au moins les points de vue : organisationnel (rôles), argumentatif (négociations), de planification (organisation temporelle, financière, etc.), de domaine. Afin de permettre un classement plus fin de ces annotations, nous proposons donc d'étendre l'indexation de l'annotation pour la coopération non seulement par des points de vue spécifiques aux domaines (thèmes), mais aussi par une dimension cognitive grâce à un point de vue argumentatif (conservant la trace des décisions et des négociations entre les participants) ainsi que par un point de vue organisationnel, se servant du rôle de l'acteur pour souligner l'importance d'une décision. Une annotation n'est bien entendu pas limitée à un point de vue et elle est souvent décrite par une intersection de points de vue.

1.2.2.1. Point de vue organisationnel

Le point de vue organisationnel consiste en l'indexation selon les rôles de l'utilisateur. Celui-ci à un rôle au sein du projet qui est défini officiellement. Au niveau du système, cela se traduit par la définition et la mise à jour d'un profil utilisateur lors de l'ajout d'un utilisateur au système. Ce point de vue officiel est donc récupérable automatiquement. Il permet d'ajouter des descripteurs à notre classification.

1.2.2.2. Point de vue de planification

Le point de vue planificateur consiste en l'ajout de descripteurs ayant trait à la planification des activités d'un projet. Nombre de commentaires laissés au cours d'un projet dans un but de communication relatent des problèmes d'organisation du groupe et de leurs activités comme la mise en place de sessions de travail, l'agencement des phases de travail, le choix de dates et de lieux de réunion, le prélèvement de tel ou tel compte pour régler une facture, etc. Ce sont ces messages qui seront indexés par les descripteurs du point de vue de planification.

1.2.2.3. Point de vue argumentatif

Le point de vue argumentatif consiste en l'indexation des messages ayant un contenu argumentatif. L'argumentation est ici prise au sens le plus large de présenter des arguments. Nous ne limitons pas l'argumentation à la dialectique en tant que raisonnement logique dans l'objectif d'emporter une conviction. Déposer des commentaires pour la définition d'une solution, reporter un problème technique à discuter plus avant ou encore le refus catégorique d'une solution sont des thèmes considérés comme relevant du point de vue argumentatif au même titre que la critique construite d'une solution. Ainsi un grand nombre d'annotations correspondent à ce point de vue puisqu'elles mettent très souvent en place un échange communicationnel sur le plan du développement du produit. Les descripteurs proposés d'emblée doivent donc être placés sur une gamme très large de l'argumentation.

1.2.2.4. Point de vue de domaine

Le point de vue de domaine permet de définir l'étendue de la définition d'un terme ou d'un concept. Le sens d'un terme est inhérent à sa position dans son co-texte et son contexte d'énonciation, entre autres, son énonciateur. Par exemple, le terme flasque pour la plupart d'entre nous rappelle l'adjectif décrivant une texture molle, mais dans le sociolecte de l'équipe projet en conception mécanique, ce terme réfère, sans aucun doute pour chacun des membres, au nom flasque. Encore une fois non pas au premier sens du dictionnaire du mot flasque au féminin³⁷, mais au second sens de un flasque, une pièce mécanique verticale servant de support. Toutefois, les concepts du point de vue du domaine sont le plus souvent présents sous forme de termes dans les documents (ce qui n'est pas le cas de l'argumentation, sauf dans des documents très normés comme les publications académiques). Il est donc possible de les

³⁷ Flacon, poire à poudre, petit flacon plat

récupérer et des les structurer automatiquement selon la démarche et les outils définis en chapitre VI, section 4.

Dans notre système, les ressources organisationnelle et argumentative doivent être construites manuellement. En effet, la première est fondée sur une analyse de l'organisation de la communauté et la seconde sur une analyse cognitive et pragmatique des interactions au sein du groupe projet et sont difficilement automatisables. Les ressources liées au domaine, elles, peuvent être découverte par une analyse conjointe d'outils issus du Traitement Automatique des Langues (TAL) et d'opérations manuelles sur le choix des termes et des concepts. Ces trois éléments de la RTO doivent être stockés sous un format adapté qui permet une représentation d'informations de différentes granularités et une récupération aisée des termes et points de vue par l'utilisateur.

Une description multi-point de vue de l'annotation selon le rôle de l'annotateur, l'argumentation de l'annotateur ou encore le rôle de l'annotation dans la planification du projet met en valeur le statut communicationnel de l'annotation. (Bringay, 2006) reprend cette structuration dans un modèle. Elle structure sa typologie entre la production de l'annotation et sa réception. La réception est séparée en la réception par l'auteur de l'annotation (annotation individuelle) et la réception par un autre lecteur (annotation coopérative). Nous nous plaçons dans cette perspective de soutien de la coopération par l'annotation et souhaitons nous orienter vers une optique communicationnelle et relationnelle en plus de la coloration intentionnelle du but annotatif. Il est alors important de souligner l'intérêt de l'annotation en tant qu'échange avec l'autre.

2. Un fragment de conversation

L'herméneutique et son discours annotatif permettent à une altérité de s'exprimer, donnant naissance à un horizon partagé (Gadamer, 1960). Suivant cette perspective de parole auctoriale (d'auteur) et d'altérité, il est possible de rapprocher le concept d'annotation à ceux de polylogue discontinu (Marras, 2004) ou de conversation persistante (Erickson, 1999). À la croisée de la parole et de l'écrit, l'annotation est un outil intéressant pour permettre la négociation et la création de documents. Dans un cadre de médiatisation, il faut alors observer les moyens disponibles en soutien de cette activité. Dans un forum, plusieurs participants sont impliqués dans un fil de discussion. Nous avons donc affaire non pas à des monologues, ni de simples dialogues, mais des polylogues. Comme montré précédemment, l'annotation possède des

caractéristiques relationnelles fortes qu'elle partage avec les *posts* (messages) d'un forum, reliés par un fil de discussion. L'annotation partage aussi avec les posts d'un forum ses caractéristiques dialogiques ou d'échanges entre une multiplicité de participants.

Dans un contexte de travail distribué, une annotation peut être vue comme un *post-it* collé à un dossier pour informer des collaborateurs facilitant la *conscience des activités* réalisées par les autres participants (*activity awareness* (Carroll *et al.*, 2003)) et permettant de continuer l'activité. Elle est aussi la note en marge d'un document exprimant une opinion et permettant d'engager une négociation qui aide à la construction d'un référentiel commun (Clark, 1996).

La négociation est un mécanisme clé de l'interaction. Il existe un grand nombre de collecticiels qui s'appuient sur les principes de la négociation. Certains collecticiels soutiennent le processus de négociation en lui-même pour de la résolution de problèmes et des outils d'aide à la décision par exemple, quand d'autres utilisent la négociation comme l'un des moyens existants pour favoriser la coopération. La majorité des systèmes coopératifs utilisent la négociation pour coordonner des tâches, pour développer une compréhension partagée d'un sujet, ou pour résoudre des conflits (Dillenbourg et Baker, 1996). Le processus de négociation peut cependant mener à des résultats bien différents selon son utilisation. (Prilla et Ritterskamp, 2006) qualifient de « *system-external behaviour* » et de « *internal state transition* » les systèmes support à la négociation qui peuvent être affectés par les conséquences de la négociation respectivement de l'extérieur ou de l'intérieur. La négociation peut mener à des activités extérieures au système ou modifiant le système lui-même. Dans le cadre d'un collecticiel permettant de mener une négociation via un discours, le discours peut avoir des conséquences sur le document produit suite à une conversation (l'ensemble de posts) ou sur le système lui-même (la définition d'une classification à utiliser dans l'outil).

L'annotation soutient ainsi la communication entre participants à une activité, et permet des échanges à propos d'un document et la construction d'une communauté. Du point de vue de la construction d'une communauté, qui influe sur le système, (Marshall, 1998a) souligne que l'annotation réfère à une audience intentionnelle et s'adresse donc à un individu, un groupe ou au public. C'est le choix de cette audience qui soutient la création d'une communauté en plus de la définition d'une classification et de l'engagement de chacun des membres dans cette construction de classification. L'élaboration d'une interprétation partagée via une négociation au sein d'un collectif participe selon nous à la construction collective du sens (*collective sensemaking*, Weick, 1979). Ce dernier conçoit en effet la construction collective du sens dans les organisations comme un processus de réduction collective de l'ambiguïté perçue d'une

situation. C'est en échangeant, en débattant, que les membres de l'organisation vont clarifier puis partager des compréhensions de situations (retranscrites dans des documents), ce qui construira du sens petit à petit. Les travaux de Weick mettent l'accent sur le processus de création du sens, son émergence et son évolution, et non pas sur la représentation collective du sens. Le sens collectif n'est donc pas seulement un sens partagé par un collectif. Ce processus est le support de l'identité individuelle, car, par le biais de ses interactions, chaque acteur met à l'épreuve et fait évoluer son identité.

L'interprétation collective de documents laisse des traces des actions menées dans l'organisation. Conserver et colliger les annotations, traces de ces interprétations, permet également de tirer parti des documents élaborés, tout en étant capable de les sortir éventuellement du cadre dans lequel ils ont été rédigés. Ainsi, du point de vue du poids de la négociation sur des objets extérieurs au système, il est crucial de noter que l'annotation produite, en tant que fragment textuel, mène à l'élaboration de nouveaux documents.

3. Un fragment de document

Dans une perspective herméneutique, ou d'interprétation de texte, l'annotation est une technique riche de lecture d'un texte. En observant l'évolution des exercices universitaires de l'époque médiévale dans (De Libera, 2000), on constate que l'annotation est un élément de la création de nouveaux textes, utilisé dans un objectif didactique au Moyen-Âge. En herméneutique, l'activité d'annotation est considérée alors comme une activité de médiation d'une réflexion sur des textes (législatif, religieux). Les échanges interprétatifs entre herméneutes (lecteur annotateurs) mènent à la création de nouveaux concepts par l'intermédiaire de nouveaux textes ou fragments de textes. L'annotation est définie comme un fragment textuel étendant le sens du texte par la médiation de la parole de l'autre proposant une nouvelle perspective de lecture. L'évolution diachronique de la glose vers la somme³⁸, se fonde sur le principe d'annotation de texte et montre comment de l'annotation peut naître un document textuel riche grâce à ses références à d'autres interprétations. Les annotations sont des éléments en action impliqués dans la création de documents, des DOcuments Pour l'Action (DoPAs). A. De Libéra (2000) présente l'exercice médiéval de lecture comme un exercice de glose, c'est-à-dire d'ajout de commentaires littéraux à un texte. Ces gloses étaient utilisées dans l'explication de mots dont le sens n'allait pas de soi. Graduellement, ces commentaires

³⁸ Œuvre qui résume toutes les connaissances relatives à une science, à un sujet (in Le Petit Robert *s.v.* « somme »)

paraphrastiques s'étoffent et se détachent du mot et du texte lui-même pour devenir des commentaires organisés. Des séances de commentaires publiques donnent lieu à des joutes oratoires entre étudiants chevronnés ou universitaires qui sont ensuite rapportées par un étudiant en formation. Dans cet exercice, les orateurs interagissent sur un thème (*sententia*) en argumentant leurs points de vue créant ensemble de nouvelles problématiques ou conclusions. De la révision des *sententiae* et des problématiques sont aussi tirées des sommes *summa*. Le genre de la somme est finalement une collection organisée de rapports revus et discutés sur un thème. La somme permet d'homogénéiser la connaissance sur un thème ou un point bien défini. Ces différents types de commentaires sont issus de lectures et reliés à des documents textuels ou à d'autres fragments textuels et, en cela, nous les considérons comme des annotations. De ces descriptions de A. De Libera (2000), nous dégageons plusieurs types d'annotation impliqués dans la création de texte :

- la marque physique permettant de diviser un document pour souligner l'importance de certaines parties ;
- la glose qui est un fragment de texte expliquant une partie du document ;
- la note paraphrasant un point du document ;
- le commentaire amenant de nouvelles idées ;
- le commentaire discursif qui est une argumentation ou un commentaire organisé construit en coopération par les échanges entre annotateurs.

Dans nos travaux, nous nous limiterons aux annotations textuelles (nous excluons par exemple la marque graphique ou de couleur, ou le croquis, traités dans les travaux de S. Guibert (Guibert *et al.*, 2005)), matérialisées sous la forme d'un fragment de texte relié à un document à différents niveaux.

Les différents niveaux d'annotations, de la glose au commentaire discursif, nous permettent d'identifier des phases d'élaboration de ces fragments textuels. Suite à une lecture, l'annotateur met en valeur un fragment du texte en lui liant une annotation strictement explicative, ou peut au contraire organiser une argumentation dans le corps de cette annotation en vue de l'élaboration collective d'une interprétation sur le fragment saillant. Il peut aussi, grâce au dépôt d'une annotation sur un document, donner une information à un destinataire et créer un lien entre son opinion et un document, voire mettre en relation plusieurs documents par son commentaire. La tension entre intensivité et extensivité de l'annotation (Marshall, 1998a) est d'autant plus forte que celle-ci est médiatisée. Cette dichotomie définit la force de la relation

entre l'annotation, le discours secondaire et le document, le discours premier. L'annotation peut aussi bien être liée au document qu'elle annoté qu'elle peut être liée à un autre document en cours de production par exemple.

L'étude diachronique de l'annotation nous permet de tracer un processus de création de document textuel et de repérer l'utilisation de l'annotation dans ce processus. Ces études soulignent aussi l'étendue sociale que prend l'exercice d'interprétation par le biais des annotations en tant que consignation des propos d'une argumentation entre pairs, de la même manière que les échanges dans un forum.

Suite à ce tour d'horizon définitoire de l'annotation en tant que fragment de document, fragment de conversation et chemin vers le document, nous constatons une proximité conceptuelle importante. C'est pour cette raison que nous adoptons une définition unifiée de l'annotation que nous présentons ci-après. Cette définition unifiée de l'annotation nous autorise à procéder à la médiatisation d'une annotation au sens large qui permet de soutenir le travail coopératif médiatisé d'un groupe de travail et de tracer la logique de communication des productions de ce groupe.

4. Une définition unifiée : L'annotation coopérative, une trace de la logique de communication

(Prié et Garlatti, 2004) opèrent un distinguo entre l'annotation et la métadonnée. Ils considèrent que la métadonnée sera plutôt attachée à une ressource identifiée en tant que telle sur le Web et aura plutôt une pertinence a priori. De plus, elle sera plutôt saisie suivant un schéma qui permettra de mettre en place des inférences. A contrario, l'annotation sera plus située au sein de cette ressource et écrite au cours d'un processus d'annotation / lecture. Ils souhaitent ainsi mettre l'accent sur le caractère plus situé au sein de la ressource (du fait de son exploitation par un utilisateur) de l'annotation, par rapport à une métadonnées plus indépendante, voire ressource en tant que telle. Nous ne nous plaçons pas dans la nécessité de différencier l'annotation discursive de l'annotation sémantique, mais plutôt de représenter le continuum entre ces annotations. Faire un lien entre le Web sémantique (WS) dont l'objectif est manifestement l'interopérabilité des systèmes et le Web social (WSo) dont l'objectif est la mise en relation de personnes. Nous nous plaçons dans l'optique d'un Web Socio-Sémantique (W2S) (Zacklad et Barbaud, 2004) comme promu aujourd'hui dans le cadre du Web 2.0. Le Web Socio

Sémantique tel que défini par M. Zacklad vient compléter la définition du Web Cognitivement Sémantique proposée par J. Caussanel (Caussanel *et al.*, 2002). Le Web Cognitivement Sémantique visait à intégrer dans les recherches et les pratiques de développement des applications du WS l'ensemble des activités de conception initiale des représentations, de maintenance au fil de l'eau (au fur et à mesure que les connaissances évoluent) et d'évaluation de la pertinence des résultats des requêtes. Tout en prolongeant cette perspective, le Web Socio-Sémantique (W2S) se positionne vis-à-vis du WSo. Le W2S vise à soutenir des activités de coopération plus structurées dans lesquelles les interactions s'appuient également sur des informations ou des documents partagés par un collectif poursuivant, au moins pour un temps, des objectifs communs. Vis-à-vis de ces objectifs, le W2S doit contribuer à la construction d'une représentation structurée tant du domaine que du collectif.

Nous envisageons l'annotation comme une métadonnée, mais aussi comme un moyen de négocier des frontières pour créer un référentiel commun et favoriser la conscience mutuelle des annotateurs, en enrichissant sémantiquement un document. Le document est enrichi par la relation, le passage créé entre le contenu d'un document et le corps de l'annotation (voire parfois entre deux documents via le corps de l'annotation). Ce contenu extérieur peut être : un contenu calculé informatiquement (visualisation d'un document HTML par le biais d'annotations, de balises conventionnelles), un autre document (un lien hypertexte, un marque-page), ou même le corps textuel d'une annotation (« Je ne suis pas d'accord avec votre définition du WS »). L'annotation peut être considérée comme un continuum de la balise (chemin vers le document - annotation sémantique -) au commentaire (fragment de conversation et de document - annotation discursive, annotation élaborante -) (Voir Fig. 7).

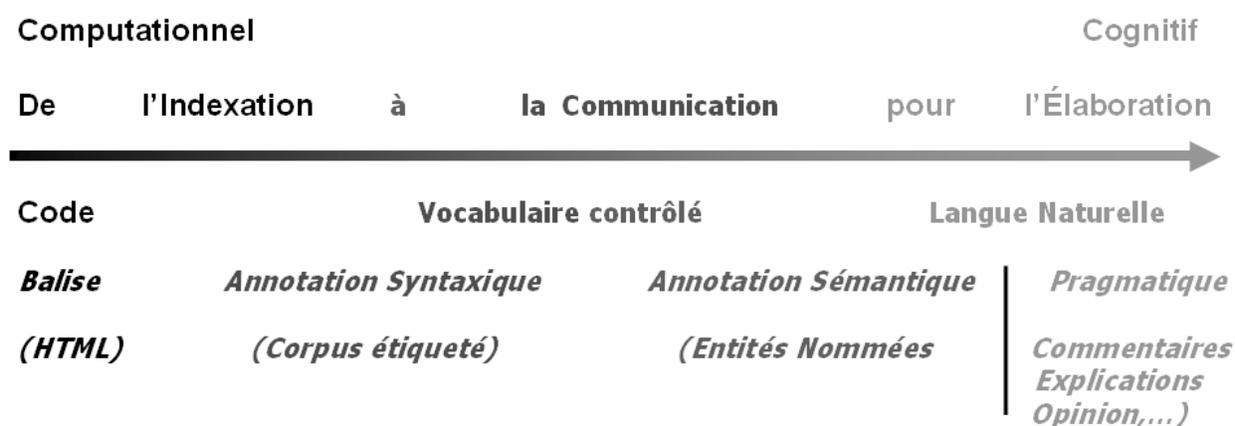


Fig. 7. DÉFINITION DE L'ANNOTATION EN UN CONTINUUM

L'annotation définie comme un fragment de discours relié à un texte est un élément pour la négociation. L'annotation est contextualisée par son environnement textuel d'une part et son environnement de production d'autre part. Son environnement textuel est constitué par son corps (son contenu) et son co-texte (le document auquel elle est ancrée et/ou le document ancré par l'annotation). Son environnement de production est représenté par son auteur (qui l'a écrite, qui l'a rédigée), son destinataire (à quel(s) collaborateurs est-elle originellement destinée), son rôle (quel est son but argumentatif, informatif, etc.). Nous sommes particulièrement intéressées par l'enrichissement cognitif d'un document par ce type d'annotation. Cependant l'annotation est aussi contextualisée par les relations qui la relie à d'autres documents, fragments de documents et annotations. L'activité d'annotation discursive peut alors être soutenue par un outil informatique, dans lequel l'annotation sémantique créée pour l'indexation permettrait de structurer les échanges via des annotations discursives, et donc de tracer la logique de communication permettant de comprendre la conception autour d'un document.

En effet, afin de tracer les choix de conception au cours d'un projet, les concepteurs en mécanique conservent les traces de la logique de conception, i.e. les choix de conception suivis au cours du projet. Cependant, a posteriori, il manque les négociations qui ont mené à ces choix. Nous postulons qu'il est possible de tracer la *logique de communication* d'un projet en structurant l'ensemble des échanges du groupe afin de comprendre les négociations et argumentations intervenues dans la construction d'éléments ou produits du projet. Le document créé grâce au recueil d'annotations et leur mise en forme dans un document plus homogène représente la logique de communication qui permet de comprendre la logique de conception d'une pièce, ou du plan qui représente cette pièce. L'ensemble de ces documents ou fragments de document permettra d'élaborer une somme des nombreux problèmes posés et solutions apportées au cours de notre projet de conception en mécanique aéronautique. Il s'agit, pour que cette somme reste compréhensible a posteriori, d'organiser ces fragments selon leur apport argumentatif ou technique dans le projet. L'organisation des fragments permet d'éviter de laisser aller le document vers une décomposition thématique (Herring, 1999) grâce à l'indexation des fragments de documents selon leur valeur argumentative par exemple. Dans notre contexte de conception coopérative médiatisée par ordinateur, la logique de communication peut être tracée en suivant les négociations menées sur les schèmes de classification du groupe au cours d'activités de classement de documents ou de modifications de la classification par exemple.

La logique de communication est définie par des annotations discursives ayant une valeur argumentative, indexées finement et concourant à la création de document. Ces trois activités d'indexation, de communication, et d'élaboration produisant différents objets annotations, nous mènent à la définition de l'annotation en un continuum du plus computationnel au plus cognitif, de la balise pour l'indexation au commentaire en langue naturelle pour la négociation et l'élaboration discursive. Pour soutenir la coopération au sein d'un groupe, l'annotation doit à la fois soutenir la récupération de document, la conversation et l'élaboration de document.

Nous avons également montré qu'il existe, pour chacune des facettes de l'annotation, des instrumentations possibles : les outils issus du Web Sémantique pour l'annotation index, les outils du Web Social pour l'annotation discursive (ou fragment de conversation), et les outils de rédaction coopérative assistés par ordinateur pour l'annotation élaborante (ou fragment de document). Or, pour concevoir un collectif support à l'annotation, prenant en compte l'annotation comme un continuum, il nous faut relier ces différentes instrumentations. Pour cela il nous est nécessaire de définir un modèle de l'activité d'annotation, qui nous fournira un cadre pour spécifier le collectif. La partie suivante sera donc consacrée à la recherche d'un modèle d'inspiration pour illustrer la coexistence des trois types d'activités liées aux trois facettes de l'annotations listées ci-dessus : la communication, l'indexation et l'élaboration.

3^e partie :

Autour de l'activité annotation

Dans le cadre de la démarche de conception que nous avons présentée en section 2 de la première partie de ce mémoire, nous souhaitons proposer un modèle d'activité d'annotation qui nous permettra de définir un modèle de cette activité médiatisée. Ce dernier servira de base à la prescription de fonctionnalités à mettre en place dans un outil support à l'annotation. L'idée que nous poursuivons ici est que si l'on souhaite proposer un outil qui instrumente une activité qui n'existe pas en tant que telle (l'outil créant de nouveaux possibles), il nous faut puiser notre inspiration non pas d'une observation de pratiques existantes, mais de théories pertinentes pour notre problématique, l'herméneutique.

Dans un premier temps, nous présenterons des modèles proposés dans le cadre de la représentation d'activité de production autour de textes (chapitre III). Dans un second temps, nous présenterons le modèle adopté (chapitre IV). Le modèle est issu d'une description littéraire de l'art mémoriel au Moyen-âge et plus spécifiquement de l'usage de l'annotation en rhétorique médiévale. L'activité modélisée permet de donner une base au développement d'un collecticiel support au travail collaboratif.

Chapitre III

Quel modèle de description de l'activité ?

*[...] un scripteur se « relit », bizarrerie lexicale qui
laisse entendre qu'écrire et lire sont déjà solidaires.
Josette Rey-Debove*

L'herméneutique liant de manière indissociable production d'annotation et compréhension du texte, nous étudions les travaux ayant analysé les activités centrées sur le texte. Nous ne ferons pas un relevé exhaustif des modèles souvent issus de travaux en psychologie cognitive ou psycholinguistique, mais nous nous focaliserons sur les modèles qui ont profondément marqué le champ ou qui amènent des modifications novatrices plus proches de notre problématique. Ainsi, suite à la présentation de modèles cognitifs les plus classiques (Kintsch et Van Dijk, 1978, Kintsch 1988, Hayes et Flower, 1980), nous exposerons aussi des modèles ouvrant sur une conception de la production verbale (Levitt, 1989) plus large, mêlant production verbale orale et écrite (Van Wijk et Sanders, 1999), puis sur des modèles soulignant une problématique interactionniste (Nystrand, 1989). Nous terminerons par la présentation de modèles de la rhétorique (Aristote, - 400) et conclurons sur un modèle possible à adopter.

1. Les modèles cognitifs

Les modèles de lecture ou d'écriture issus des sciences cognitives, principalement de psychologie cognitive, ont pour but une représentation fonctionnelle d'un processus donné, la lecture ou l'écriture. Ils décrivent donc finement les fonctions, éléments et relations d'un processus. Ces modèles sont perfectibles et servent de base à des expériences qui permettent de les valider ou de les réfuter, les faisant évoluer pour plus d'efficacité dans la représentation de processus mentaux.

Les modèles de lecture sont le plus souvent vus comme des modèles de compréhension dont le plus typique est celui de (Kintsch et Van Dijk, 1978). Ce modèle a aujourd'hui un fort impact dans la communauté des sciences cognitives et a déjà donné lieu à bon nombre d'expérimentations ayant le plus souvent renforcé la théorie propositionnelle.

Cependant, les expérimentations menées dans ce cadre n'offrent la plupart du temps pas une activité finalisée. Lorsque celle-ci est finalisée, plus orientée sur la tâche, elle s'apparente alors aux processus de recherche d'information plus qu'au processus strictement de compréhension.

La recherche d'information, sans s'y limiter, peut être considérée comme un type de résolution de problème (Rouet, 2001). Les modèles de résolution de problèmes sont très répandus lorsqu'il s'agit de modéliser le processus d'écriture. Le modèle de ce type le plus connu est celui de (Hayes et Flower, 1980). Ces modèles cognitifs, soit de compréhension, soit d'écriture, sont inscrits dans une perspective de cognition individuelle, largement marquée par les activités écrites de type récit. Ils se focalisent sur un individu qui comprend ou rédige dans un contexte.

1.1. La théorie propositionnelle (Kintsch et Van Dijk, 1978, Kintsch, 1988)

W. Kintsch et T. Van Dijk proposent une analyse de la sémantique des récits fondée sur la proposition. Cette proposition est constituée d'un prédicat (relation d'action) et d'un ou plusieurs arguments. L'analyse propositionnelle d'un récit permet de décrire des micro-structures et des macro-structures (avec des macro-règles).

Une micro-structure est une liste de propositions ou de micro-propositions dans l'ordre d'apparition des prédicats, noyau de la proposition, au cours du récit. Ces propositions sont reliées entre elles en suivant le principe de macro-règles construisant des macrostructures. Une macrostructure permet d'ordonner les propositions, de construire des séquences du récit.

Les macro-règles définies par W. Kintsch et T. Van Dijk sont au nombre de quatre : (1) La généralisation, (2) la suppression, (3) l'intégration et (4) la construction. La généralisation permet de réunir plusieurs micro-propositions en une seule macro-proposition (ex : Popeye mange des brocolis, des haricots et des épinards → Popeye mange des légumes verts). La suppression permet de supprimer certaines micro-propositions considérées comme contingentes (ex : Il y a du fer dans les épinards que Olive a préparés → Il y a du fer dans les épinards). L'intégration permet de former une seule proposition plus globale, par suppression de séquences déjà exprimées ailleurs (ex : Popeye mange ses épinards. Popeye mets la

cuillère d'épinards dans sa bouche → Popeye mange ses épinards). La construction, proche de l'intégration, est une reconstruction de plusieurs propositions en une seule structure de niveau plus général (ex : Popeye mange des épinards trois fois par semaine → Popeye a une bonne hygiène de vie).

Toutes ces règles sont orientées par les buts du lecteur, représentés dans un schéma général. Ce schéma comprend en plus des macro-règles présentées ci-dessus, des processus de mise en cohérence et des opérations de production.

Ce modèle est avant tout un modèle de compréhension de texte spécifique, le récit. Il fonctionne très bien sur les tests menés. Ceux-ci se fondent les plus souvent sur des textes dont la structure est stéréotypée ou fortement codée. Dans notre problématique d'annotation, les textes ne sont généralement pas des récits et la structure n'est pas prédictible comme dans le cas de récits produits selon des règles de rédaction reconnues (par exemple la structure du conte).

La plupart des expérimentations sur la production de récit avec le modèle de W. Kintsch et T. Van Dijk ont été effectuées sur la production de résumé de texte dans laquelle on observe plus les transformations et réorganisation d'un texte présenté auparavant, et les taux de rappel des informations présentées. Même si les principes inférentiels combinés aux opérations de reconstruction considérés comme les inverses des macro-règles semblent intéressantes, il ne reste pas moins que les expériences portent sur la qualité du rappel et non sur celle de l'invention, de la construction d'idées nouvelles et de concepts novateurs. Or il s'agit pour nous d'avantage d'assister la création de prédicats nouveaux que de rappeler des informations présentées et liées à un schéma de production de récit.

1.2. Le modèle de processus d'écriture de Hayes et Flower (1980)

Le modèle de processus d'écriture de J. Hayes et L. Flower est le modèle rédactionnel fondateur largement reconnu par la communauté des sciences cognitives. (Piolat, 2004). Il possède une architecture modulaire. Les trois modules principaux sont les modules d'environnement de la tâche, de mémoire à long terme et celui stricto sensu du processus d'écriture (voir figure 8).

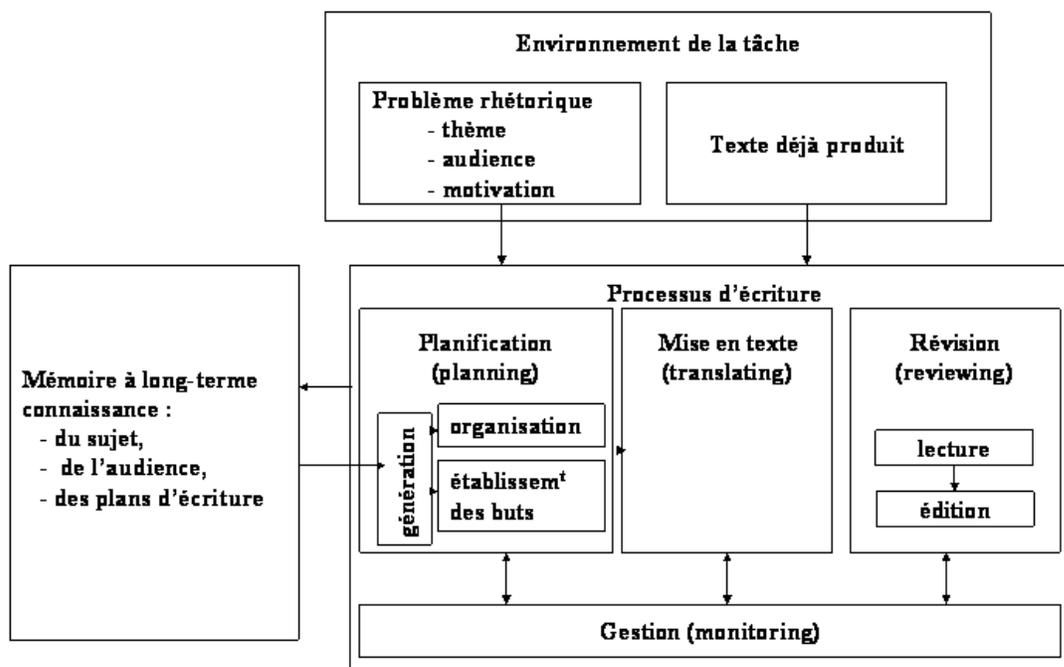


Fig. 8. MODÈLE DE PROCESSUS D'ÉCRITURE DE HAYES ET FLOWER (1980)

Le module du processus d'écriture, décrit dans (Hayes et Flower, 1980, Hayes, 1995), est constitué de trois tâches activées avec un principe récurrence.

On trouve tout d'abord la tâche de planification. La planification consiste en l'organisation des idées du rédacteur selon ses buts et objectifs. La mise en texte est l'encodage du message préverbal formé en phase de planification. Cet encodage peut être de différents niveaux, graphémique, orthographique, lexical, syntaxique...

Le processus de révision est un processus de contrôle. Il est défini dans (Bereiter et Scardamalia, 1987), comme la détection d'une inadéquation par rapport aux contraintes de rédaction ou à la représentation mentale du texte envisagé, sa caractérisation et le choix d'une stratégie de correction. Cependant, J. Hayes et L. Flower (1980) redéfinissent la révision comme un processus comprenant quatre opérations :

1. la définition de la tâche ;
2. l'évaluation du texte et l'identification du problème rencontré ;
3. la sélection d'une stratégie impliquant soit un retour à l'étape précédente, soit une modification du texte ;

4. la modification du texte consistant en une révision ou une réécriture ;

J. Hayes et L. Flower tendent alors à considérer la révision comme un processus de lecture-évaluation, extension du processus de lecture-compréhension, ce qui modifie nettement la place de la production écrite.

De même, dans le modèle proposé par J. Hayes (1995 cité dans Barré de Miniac, 2000) la mise en texte est intégrée dans le processus général de la production du langage oral ou écrit. De ce point de vue, l'écriture n'est plus qu'un cas particulier du traitement de l'information (Barré de Miniac, 2000).

Dans (Hayes, 1995 cité dans Barré de Miniac, 2000), le modèle de production écrite de J. Hayes et L. Flower (1980) est fortement modifié par l'intégration de la compréhension de manière plus conséquente par la composante de révision textuelle. De même, il a accentué l'importance du contexte créant un modèle individuel-environnemental dans lequel l'activité d'écriture est l'actualisation de compétences plus générales. L'interaction entre cognition et environnement est donc soulignée, mais le modèle reste individuel et n'est pas pour autant centré sur la production de l'activité qui n'est qu'un élément de l'environnement. Dans un contexte de conception coopérative médiatisée, l'annotation est une activité fortement empreinte du contexte, et particulièrement du contexte coopératif générant des interactions entre les différents acteurs du projet.

1.3. Modèle de production verbale orale, (Levelt, 1989)

En 1989, W. Levelt (1989) propose un modèle sériel de production verbale orale qui explique comment un locuteur réalise la production d'un énoncé. Ce modèle comprend trois processus principaux : conception, formulation et articulation (voir Figure 9). Les travaux de A. Piolat (2004) reprennent ce modèle. Le conceptualiseur assure la mise en place du message préverbal, le formulateur permet l'encodage des lemmes (unité autonome constituante du lexique, forme conceptuelle du mot en mémoire). L'articulateur permet l'actualisation physique du message, son émission. W. Levelt prévoit de surcroît dans son modèle une gestion du discours avant et après la réalisation phonologique du message.

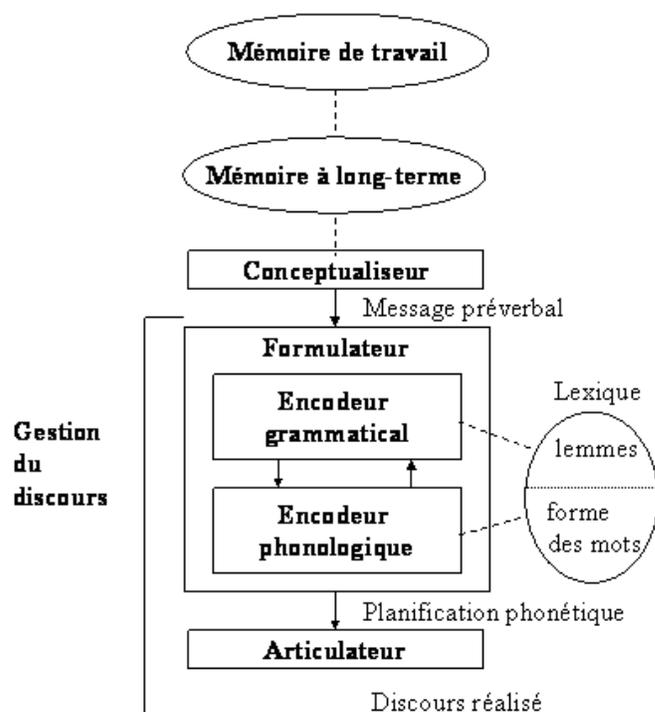


Fig. 9. MODÈLE DE PRODUCTION VERBALE ORALE, (LEVELT, 1989)

Ce modèle issu d'études en psycholinguistique est largement reconnu dans le traitement de l'oral. Il est validé par de nombreuses études sur le *TOT* ou mot sur le bout de la langue. Il a ainsi influencé les modèles de production verbale écrite. Bien que des modifications de ces premières considérations aient été apportées dans un modèle de 1999 (WEAVER), nous ne le présenterons pas ici, puisque ce qui nous intéresse est l'adaptation de ce premier modèle pour l'écrit, principalement par (Van Wijk et Sanders, 1999).

1.4. Van Wijk et Sanders, (1999)

Ce modèle de l'écrit fondé sur le modèle de production verbale orale de W. Levelt propose deux alternatives intéressantes par rapport aux modèles plus traditionnels de processus de production écrite. Tout d'abord, s'inspirant du modèle de W. Levelt (1989), il souligne les similarités de production verbale écrite et orale dont plusieurs aspects fonctionnels sont reconnus communs. Ce modèle pourrait donc, au moins sur le plan de la théorie, s'appliquer à ce que l'on considère comme un nouveau style de langage, le langage médiatisé, mélange entre style oral et style écrit.

Enfin, ce modèle réunit explicitement les phases de compréhension et de production que nous revendiquons lors de la production. Ces deux phases de compréhension et de production sont ici rassemblées sous un seul processus (voir Figure 10).

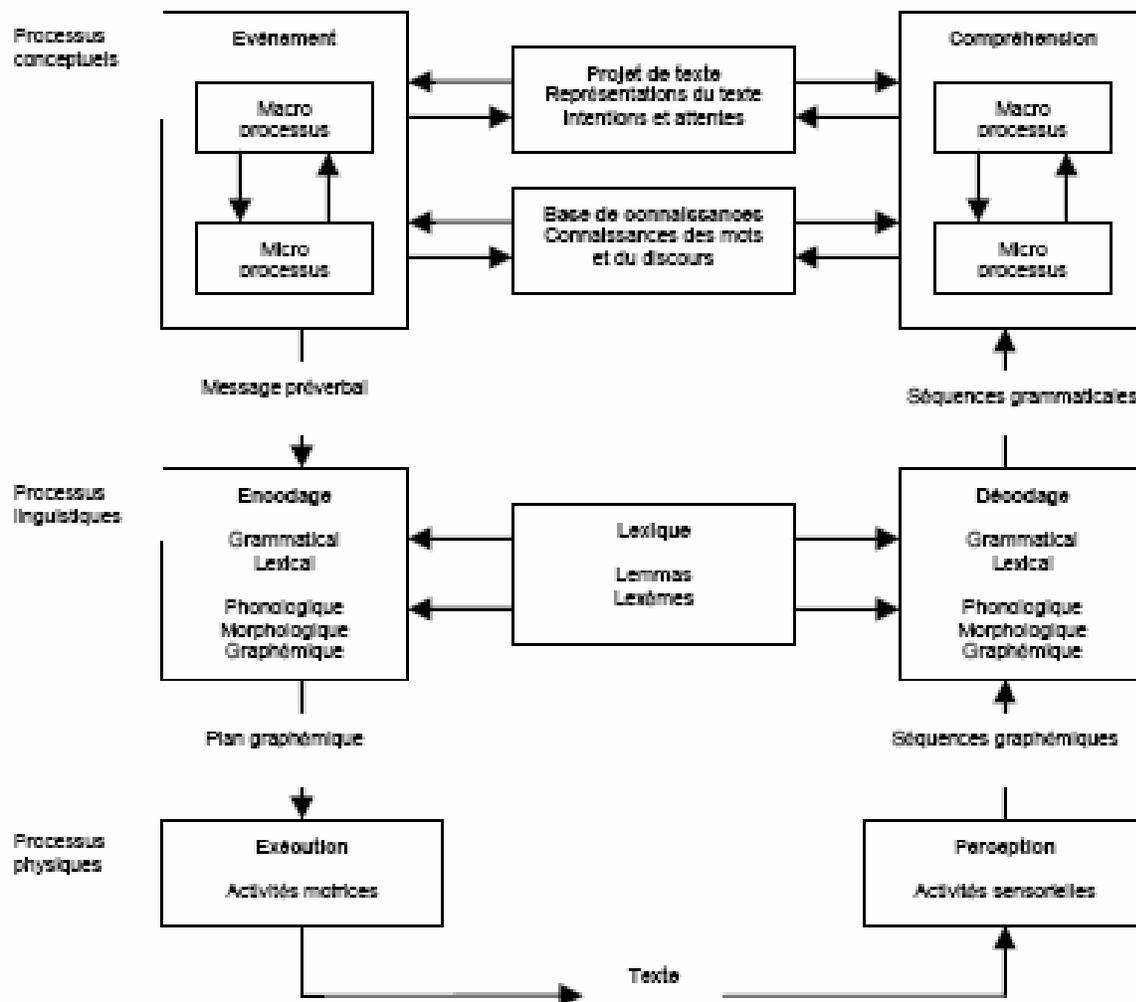


Fig. 10. MODÈLE DE PRODUCTION ÉCRITE SELON VAN WIJK (1999) IN (PIOLAT, 2004)

La faiblesse de ce modèle pour notre propos reste l'absence de processus coopératif dans le processus de production. En effet, il reste centré sur le processus interne de production. On ne met pas l'accent sur un partage des connaissances de l'environnement, sur les interactions entre le rédacteur et ses pairs dans des situations de rédaction coopérative, ni même sur le partage de la production en cours avec de quelconques collaborateurs. Cet aspect serait d'ailleurs difficile à contourner puisque ce modèle est intrinsèquement focalisé sur la méthode de résolution de problème comme la plupart des

modèles cognitifs. Or l'environnement interactionnel, voire socio-culturel, ne peut être négligé dans les problématiques de l'écrit. Cette perspective d'étude de la rédaction de textes ouvre sur de nouveaux modèles que nous appellerons modèles interactionnistes.

2. Les modèles interactionnistes

Plusieurs modèles se centrent ainsi sur les aspects sociaux de l'écriture. Nous ne présenterons ici que celui de M. Nystrand (1989), qui est l'un des premiers modèles à proposer une vision communicationnelle de l'écriture, et le prototype des modèles rhétoriques, le modèle aristotélien.

2.1. Le modèle de l'interaction sociale (Nystrand, 1989)

Dans son modèle, M. Nystrand envisage la production écrite comme une négociation de sens entre un rédacteur et son lecteur afin de créer un cadre de référence commun. Critiquant J. Hayes et L. Flower (1980) pour la pauvreté de la mise en texte au niveau des processus interactionnels, M. Nystrand remet l'accent sur l'interface entre cognition et texte. L'audience qui n'était qu'une contrainte environnementale dans (Hayes et Flower, 1980) devient un élément central de son modèle. La nature de l'écriture partage avec la nature du langage, selon M. Nystrand, la propriété sociale.

Le modèle de l'interaction sociale prend racine dans la communication orale. Il pose donc son modèle dans un contexte d'interaction entre un rédacteur et une audience à laquelle il appartient. Dans ce sens, l'interaction se produit chaque fois qu'un lecteur comprend un texte écrit. On est alors très proche de la situation herméneutique proposée par (Gadamer, 1960).

M. Nystrand précise que le texte n'est pas seulement le résultat d'une composition, il est également un médium de communication. Son modèle (Nystrand, 1989) propose donc une représentation de l'activité de rédaction comme négociation de sens entre un rédacteur et son lecteur. Trois opérations sont principalement effectuées par un rédacteur : l'initialisation du discours, la maintenance/régulation de celui-ci et enfin l'élaboration.

L'initialisation du discours écrit consiste en l'établissement d'un cadre mutuel de référence entre le rédacteur et son lecteur. Il possède donc des caractéristiques conceptuelles comme le sujet de la communication et des caractéristiques métacognitives

(Dolchy et Alexander, 1995) ou métadiscursives (Nystrand, 1986) comme le genre du texte permettant au lecteur de l'interpréter. Pour (Rommetveit, 1974), l'initialisation ainsi constituée résulte en une réalité sociale partagée temporaire (TSSR - temporarily shared social reality).

La régulation du discours écrit est une phase qui prend place après l'initialisation. Elle consiste en l'élaboration du texte lui-même qui agit sur la représentation temporaire partagée. Chaque ajout d'information doit être testé sur le lecteur pour ne pas menacer « la réciprocité rédacteur – lecteur » (Tognotti, 1997, section 2.3.).

Selon M. Nystrand, les troubles de réciprocité sont de trois ordres (Tognotti, 1997, section 2.3) :

1. élaboration inadéquate au niveau du sujet qui résulte en un texte abstrus (« le texte en dit trop sur trop peu de points ») ;
2. élaboration inadéquate au niveau du commentaire qui résulte en un texte ambigu (« le texte en dit trop peu sur trop de points ») ;
3. élaboration inadéquate au niveau du genre qui résulte en une mauvaise interprétation du texte.

Les élaborations de texte du rédacteur sont catégorisées selon M. Nystrand en mots, expressions, phrases et paragraphes et sont de trois types :

1. les élaborations de genre ;
2. les élaborations de sujet ;
3. les élaborations de commentaire.

Ce modèle est intéressant en ce qui concerne les relations d'échange entre un rédacteur et son lecteur. En effet, l'interaction entre pairs est considérée ici comme une nécessité et l'échange est mis au centre du processus.

Cependant, M. Nystrand reste dans une conception de production quasi-parfaite et chronologique où l'initialisation du discours serait une phase en elle-même. Dans le cas de l'annotation, nous n'avons pas affaire à ce type de production. L'annotation n'intervient non pas comme une production structurée du rédacteur à son lecteur, mais plutôt comme la réponse du lecteur au rédacteur. Elle est un élément de la régulation de discours.

Dans un modèle tel que celui de M. Nystrand, l'annotation disparaîtrait au profit d'un texte idéalement structuré. On aurait affaire au document pérenne et non à un document en action ou un Document Pour l'Action (DOPA, Zacklad, 2004).

2.2. Les modèles rhétoriques

Certains modèles classiques de rédaction ont été formulés dans une visée hautement communicationnelle. Il s'agit des modèles de la rhétorique et plus principalement de la rhétorique antique. Le texte à rédiger est en fait un discours oral destiné à un auditoire. Nous avons alors les ingrédients de base pour permettre une potentielle interaction autour d'un texte. Nombre de modèles existent et nombre de travaux sur la structuration des échanges ont utilisé les travaux en rhétorique. Nous nous limiterons à présenter le modèle considéré comme fondateur afin d'indiquer notre positionnement en fonction de celui-ci.

L'art rhétorique le plus fréquemment rencontré est celui défini par les principes d'Aristote (-400). Elle est alors un art de persuader par opposition à l'art de bien dire des sophistes. Les grands principes de ce modèle rhétorique sont représentés sous forme de phases successives allant de (1) l'invention, à (2) la disposition puis (3) l'élocution et (4) la mémoire et enfin (5) l'action. L'invention est la phase d'apport des arguments selon des schémas argumentatifs préconstruits (des lieux). La phase de disposition est la phase d'organisation des arguments et preuves selon un plan. L'élocution procède du choix des figures de style afin de plaire à l'auditoire et être le plus convaincant possible. La phase de mémoire consiste en la mémorisation des arguments, par l'auteur mais aussi par l'auditeur. L'action consiste en le choix de gestes pour accompagner le discours et le rendre plus efficace.

Des théories rhétoriques d'Aristote à celles d'Hugues de St Victor en passant par Cicéron ou Quintilien, la production de discours est enseignée suivant un processus défini. La rhétorique aristotélicienne se focalise sur une production finale de discours oral sans nier pour autant une phase mémorielle nécessaire à toute production. Cette phase de mémorisation est mieux représentée par la rhétorique que nous appellerons mémorielle portée par des penseurs cités dans (Carruthers, 1990), tels que Quintilien (*L'institution oratoire*), Cicéron (*De oratore*, *De inventione*) ou Tullius (*Ad Herennium*) dans l'Antiquité, puis Hugues de St Victor (*Didascalicon*), Fortunatianus (*Artis rhetoricae - libri tres -*) ou Julius Victor (*Ars rhetorica*) au Moyen-Âge. Dans cette approche de la rhétorique, un continuum

entre la partie mémorielle plus *logique* ou *dialectique* et la partie stylistique, rédactionnelle, est observable. La rhétorique est considérée comme une alliance entre structuration et éloquence. La portée argumentative et interactionniste du discours rhétorique mis en place convient bien à l'activité d'annotation, orientée vers la négociation.

Les théories et indications didactiques issues de ces nouvelles pratiques de la rhétorique nous semblent plus adaptées pour décrire l'activité de production d'annotation que les modèles précédemment décrits qui limitent le contexte de production à un seul individu, puisqu'ils sont établis sur des principes de médiatisation par l'écrit et d'interactions dans leur élaboration. Ils ne sont pas uniquement focalisés sur la portée de l'argumentation ce qui intéresse en général les travaux se fondant sur les approches rhétoriques.

Nous présentons donc dans le chapitre suivant le modèle adopté qui prend en compte la dimension interactionnelle et coopérative de l'activité d'annotation.

Chapitre IV

Proposition d'un modèle de l'activité d'annotation

*If you want to read me, you must write, you must do
something other than reading.*
Jacques Derrida

Comme nous l'avons déjà évoqué, notre modèle se doit de décrire un processus coopératif de compréhension-expression ; l'annotation est non seulement une action suite à une compréhension ou une interaction dans un collectif, mais aussi une trace du processus de compréhension suite à une lecture, destinée à être partagée avec le collectif. L'ensemble des traces écrites permet à un collectif de se constituer une réalité temporaire partagée (temporarily shared social reality - TSSR) (Rommetveit, 1974). Dans ce chapitre, nous présentons donc tout d'abord un modèle de description de l'activité issu de la rhétorique médiévale définissant l'annotation en tant qu'activité coopérative de lecture-écriture. Dans un deuxième temps, nous décrivons l'adaptation du modèle en un modèle de l'activité instrumentée. De ce modèle de l'activité d'annotation instrumentée, il est possible de définir quels objets seront primordiaux dans notre système.

1. Un modèle de description de l'activité d'annotation

Pour les raisons exposées au chapitre précédent, notre source d'inspiration pour la modélisation a priori de l'activité d'annotation est la rhétorique médiévale qui lie la lecture et l'écriture. Aucun modèle n'est évidemment développé à ce titre dans la littérature médiévale, cependant, on peut modéliser les règles de bonnes pratiques d'apprentissage sur les textes en vigueur à l'époque. La pratique se base sur le fait que pour apprendre il faut mémoriser, que pour mémoriser il faut structurer et que structurer c'est aussi produire soi-même.

Le processus de production de discours tel qu'il est préconisé comporte deux phases : *Divisio* et *Compositio* (Fig. 11). La *Divisio* se fait au cours de la lecture et représente l'étape de division d'un texte en unités intelligibles, en segments brefs mémorisables. La *Compositio*, elle, est l'assemblage ordonné, l'agencement convenable des *rei* (*res* étant un objet conceptuel aussi bien qu'un objet physique) des segments mémorisés. Ces phases de

mémorisation, la *Divisio*, et de création, la *Compositio*, sont elles-mêmes divisées en étapes soutenues par l'utilisation d'annotations.

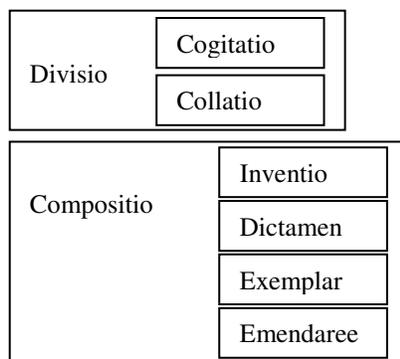


Fig. 11. PHASES PRINCIPALES DE L'ACTIVITÉ DE PRODUCTION DE DISCOURS

La première étape de la *Divisio* est la *Cogitatio* (Fig. 12). C'est une étape mémorielle individuelle qui consiste en l'association d'images par un choix et une remémoration conscients et d'une division chronologique du contenu d'un document en différents lieux mémoriels. Les fragments textuels formant le texte sont alors structurés et deviennent mémorisables facilement.

La *Collatio* est la phase où l'on combine en une structure des fragments textuels reposant en plusieurs lieux de mémoire distincts. C'est dans cette phase que des liens entre les différents lieux de contenu sont créés. Un co-texte est formé entre les nouveaux fragments mémorisés et les fragments précédemment mémorisés, en les liant sémantiquement entre eux. Cette phase n'est pas spécifiquement individuelle même si elle structure une mémoire individuelle, dans la mesure où cette étape peut être liée à des échanges discursifs, des interactions avec d'autres aidant à une structuration enrichie des concepts.

La *Compositio* est divisée en quatre stades d'activité évoquant des stades de création de document. Le stade d'*Inventio* est proche de celui de *Collatio* dans la mesure où il s'agit de créer des liens sémantiques entre divers éléments mémorisés, au niveau de la *res* (objets conceptuels, idée) pas au niveau du mot. Un plan est formé, c'est-à-dire un ensemble hiérarchisé d'idées, une structure argumentative par exemple.

La phase suivante sera celle de mise en mot de ce plan conceptuel, une phase classique de rédaction, le *Dictamen*. Nous voyons à cette étape la création physique du

discours, classiquement sur support encore modulable (brouillon), où seul le style, le choix des termes, donc la forme textuelle du discours peut être modifiée.

La phase d'Exemplar n'est que la mise en support pérenne d'un discours strictement identique à ce que l'on trouve en sortie du processus du Dictamen.

La dernière phase et non la moindre dans cette succession de processus est celle de l'Emendare où la copie finale du discours diffusée est commentée publiquement, par l'ajout des commentaires, *notae* ou arguments d'un auteur au texte original, faisant ainsi du texte une référence, un écrit faisant autorité.

L'annotation d'un document dans un cadre herméneutique consiste donc selon nous à suivre un processus de mise en discours d'idées organisées. Il s'agit en effet, suite à la lecture d'un document, d'engager un processus qui permet d'ajouter une idée, une opinion, structurée sous forme textuelle. Ce processus est synthétisé dans le diagramme d'activité (Fig. 12). Il est comparable aux travaux du GRAAL à Toulouse et de J. Virbel (in (Veron, 1998) et (Evrard et Virbel, 1996)) qui posent huit verbes d'action liés aux pratiques annotatives.

- Hiérarchiser (attribuer un niveau d'importance) ;
- Architecturer (structuration logique du texte, typage de parties du texte) ;
- Contextualiser (borner le sens) ;
- Reformuler ;
- Commenter (critique, idée associée) ;
- Documenter (ajouter un fragment pour en compléter un sens) ;
- Corréler (relier deux fragments) ;
- Programmer (prévoir une action : à traduire, à relire, à analyser).

Ces verbes représentent des buts annotatifs qui soutiennent des buts de lecture. Notre modèle permet d'affiner la typologie précédente en la structurant selon des points de vue dans une optique coopérative, nous obtenons alors :

- Indexer :
 - Classifier (hiérarchiser, architecturer) ;
 - Classifier (construire une classification) ;
 - Étiqueter (typer des fragments selon une classification) ;
 - Co-textualiser (créer une archive, corréler) ;
 - Contextualiser (ajouter des informations non-documentarisées) ;
- Communiquer :
 - Critiquer ;
 - Négocier ;
 - Échanger (des documents, des éléments) ;
 - Planifier ;
- Élaborer :
 - Reformuler ;

Développer ;

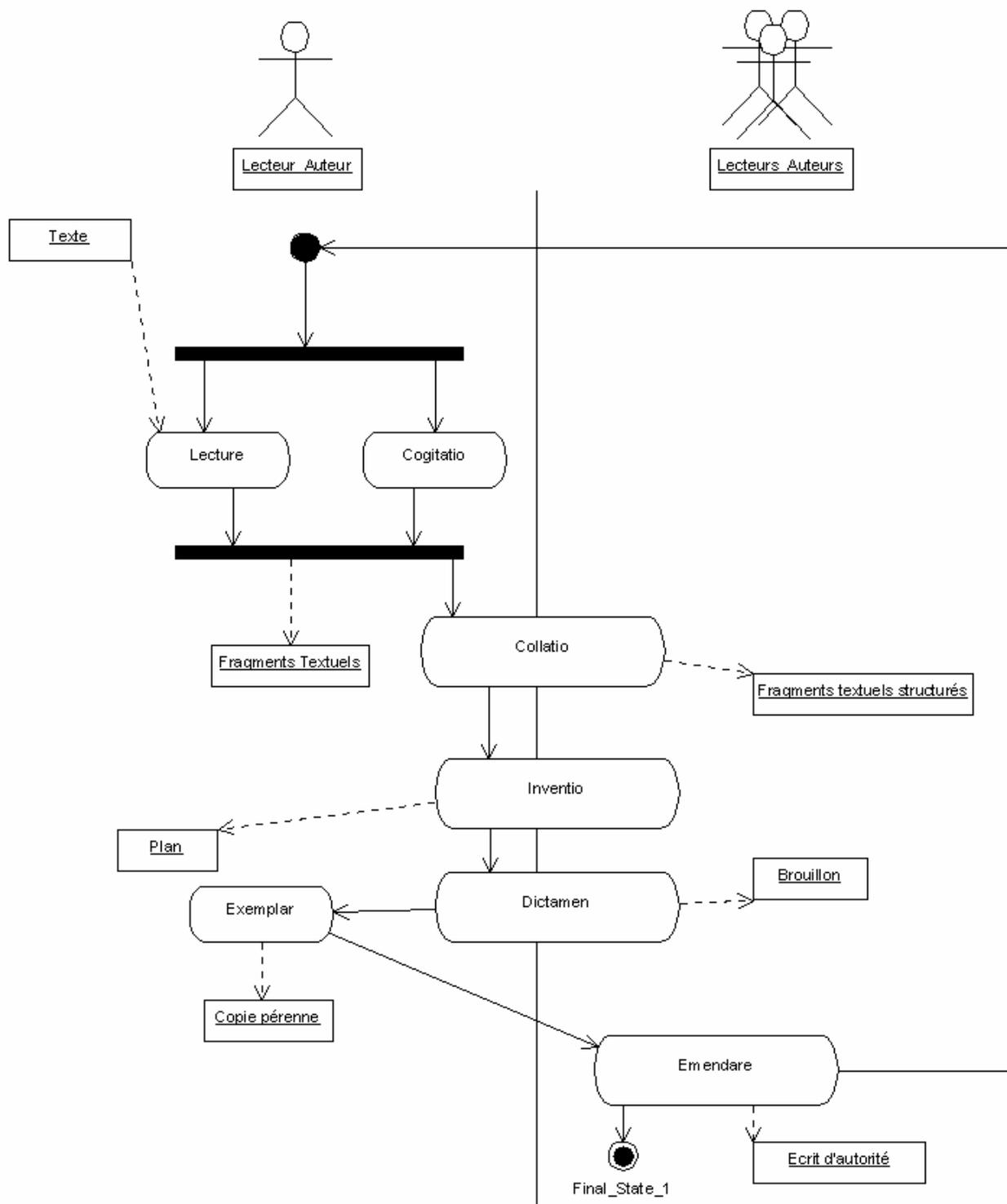


Fig. 12. DIAGRAMME D'ACTIVITÉ DU MODÈLE DE PRODUCTION DE DISCOURS

Ce modèle représente un mode de production de discours fortement soutenu par la mémoire. Dans un contexte de travail coopératif médiatisé par ordinateur, la création discursive doit être soutenue par un outil adéquat permettant de stocker, de créer et partager les informations. Afin de concevoir cet outil, nous souhaitons donc tout d'abord modéliser cette activité de production de discours instrumentée, activité que nous n'analysons ici que dans le cadre d'interactions autour de documents numériques textuels.

2. Adaptation d'un modèle à la médiatisation

Un modèle d'activité instrumentée est un modèle qui décrit une activité lorsqu'elle est opérée via un outil, c'est-à-dire en tenant compte des modifications de cette activité dues à l'instrumentation. Le modèle d'activité est ainsi projeté dans un cadre médiatisé et l'on peut alors définir les fonctionnalités et objets du système. Le modèle proposé ci-dessus relève de pratique hors médiatisation (originellement seule la base de l'activité et l'état final relèvent d'une médiatisation textuelle). Afin de le transformer en modèle instrumenté, il s'agit de mimer l'activité décrite telle que jouée au travers d'un outil.

2.1. Modèle d'activité instrumentée

Pour l'adaptation de notre modèle d'activité en modèle d'activité instrumenté, nous considérons, dans un contexte de travail collectif distribué, la mise en partage d'un document afin qu'il soit commenté. Après une phase de visualisation du texte, d'une lecture, un utilisateur du système mettra en valeur un segment (fragment textuel) de façon à signaler l'ancrage d'un élément discursif en rapport avec ce fragment. Cette mise en valeur pourra se faire par des techniques classiques de surlignage, soulignage, encadrement, colorisation de fragments de tailles variables (du mot, ou d'une partie d'un mot, au paragraphe, ou d'un ensemble disjoint d'éléments). Suite à cette segmentation et ce choix d'élément(s) à annoter, une phase d'indexation pourra intervenir, consistant en une mise en relation des fragments. Il sera alors question de tirer des liens sémantiques entre des éléments pour les structurer entre eux et former un ensemble de fragments textuels organisés selon leur sens, donné par l'utilisateur. L'annotation consistera en effet en un ancrage, une relation géographique, mais aussi en un corps, un discours qui fait sens, qui prend place dans un *co-texte*, l'ensemble des fragments textuels en mémoire indexés par des mots-clés compréhensibles et structurants

pour un utilisateur humain. Lors de la rédaction de cette annotation, une phase d'organisation du discours à écrire sera nécessaire, c'est la structuration des *res*, des concepts en mémoire, qui donnera naissance à un plan constitué d'arguments structurés hiérarchiquement. La phase de rédaction permettra de constituer le corps de l'annotation qui sera lisible par un membre de la discussion herméneutique après sa publication et donc sa diffusion.

Tout comme un texte de référence, l'annotation peut être entérinée grâce à un nouveau lien amené à celle-ci. Une réponse à un commentaire permet alors de participer au fil de discussion initié par la première annotation. Le passage du commentaire individuel à l'annotation discursive marquant la naissance d'une argumentation autour d'un document s'articule grâce à une coopération sous-jacente constituée des interactions créées par l'utilisation d'annotation.

La déclinaison du modèle de production de discours dans un cadre instrumenté nous permet de préconiser les huit étapes suivantes (Fig. 13) pour l'instrumentation de l'activité d'annotation :

Tout d'abord, le document doit être édité pour pouvoir être lu. Au cours de la lecture, l'utilisateur peut sélectionner du texte, des fragments de document, afin de les mettre en mémoire sous la forme de fragments mémorisables. Ces segments sont ensuite indexés automatiquement pour éviter la perte de la structuration du document comme unité. Les métadonnées ajoutées au segment permettent de situer les différents paragraphes, les différents mots sélectionnés dans un paragraphe. Ce type d'indexation concerne toutes les métadonnées associables automatiquement à un élément déposé (localisation, auteur, date, ...). La sélection étant faite, il est possible de déposer un nouveau fragment sous la forme d'une annotation. L'indexation doit également servir à lier les nouveaux fragments déposés à l'ensemble conceptuel existant. Nous obtenons alors un ensemble de segments textuels liés sémantiquement à d'autres segments textuels. La phase de structuration représente un processus de hiérarchisation, d'organisation des idées. Un plan détaillé est défini, contenant toutes les idées nécessaires à la mise en mot du discours. Cette phase peut être coopérative et plusieurs annotations peuvent alors être déposées sur le brouillon en cours de rédaction. C'est aussi la phase où les *rei* (concepts) contenus dans les fragments textuels indexés sont réutilisés et réorganisés en un nouveau document. La phase de rédaction est une mise en texte du plan produisant le discours. Ce discours n'est pas l'objectif final de cette activité dans cette vision de la rhétorique, puisqu'il est ensuite publié pour en faire un objet

amendable par d'autres lecteurs/auteurs de la communauté. Cette phase de commentaire par d'autres membres de la communauté est primordiale pour permettre la validation de l'exemplaire, voire son amélioration, et constituer un écrit d'autorité, un discours faisant référence dans la communauté. Dans ce cadre, il s'agit principalement de publier le document pour permettre sa révision.

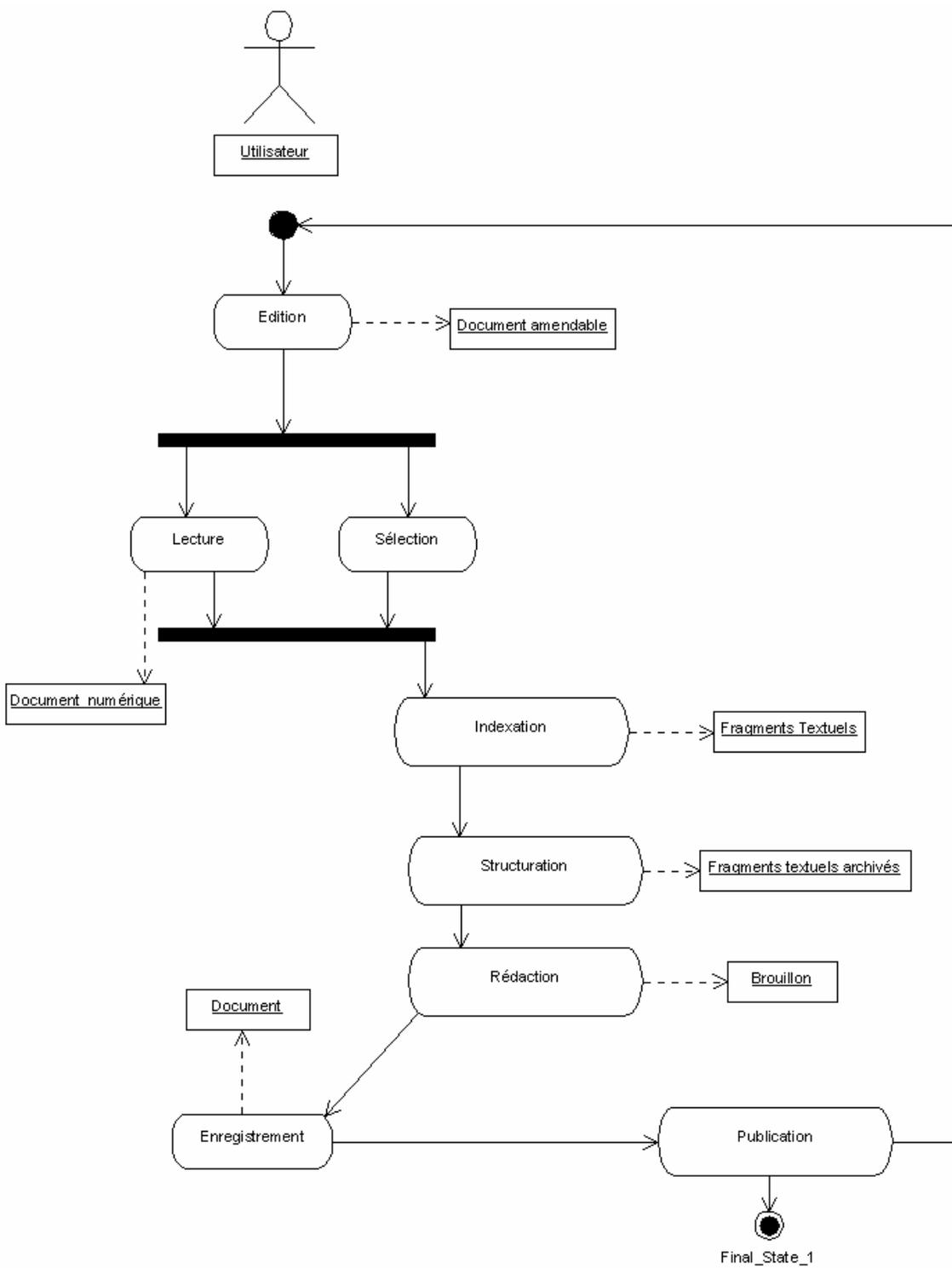


Fig. 13. DIAGRAMME D'ACTIVITÉ DU MODÈLE DE PRODUCTION DE DISCOURS INSTRUMENTÉE

2.2. Objets du système

La représentation de l'activité via le système permet de définir les objets (Fig. 14) qui vont devoir être gérés par le système en lien avec nos trois objectifs d'annotation.

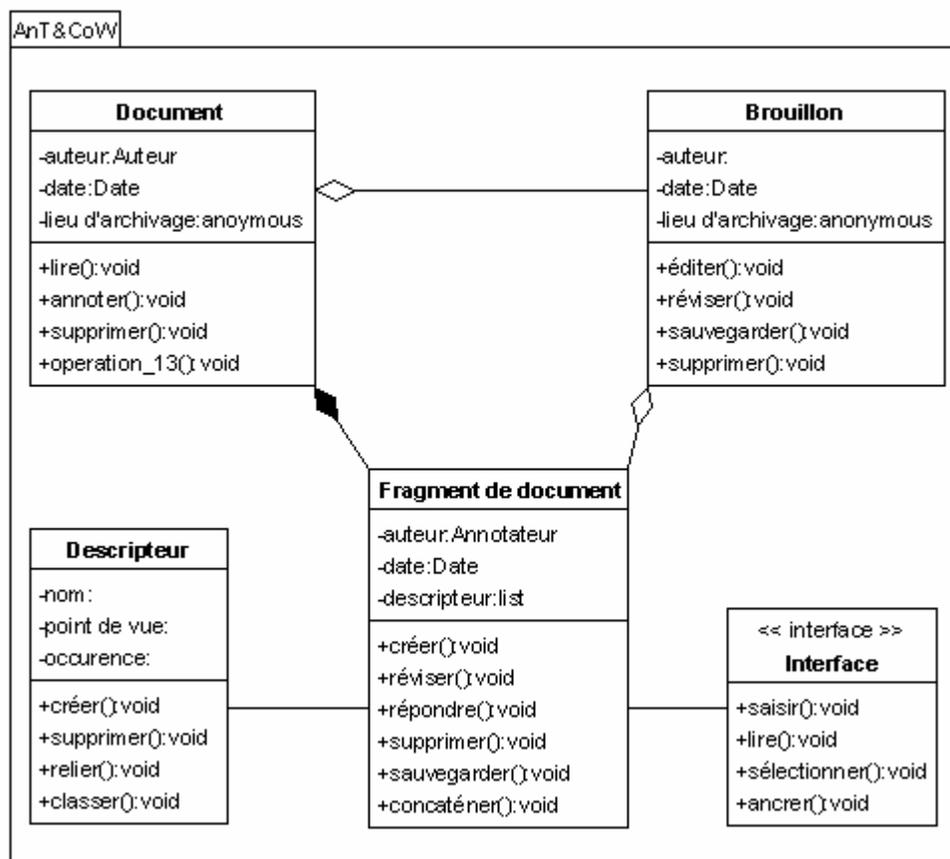


Fig. 14. DIAGRAMME DES CLASSES GÉRÉES PAR L'OUTIL

3. Conclusion

Nous proposons dans ce chapitre un modèle de description de l'activité collective de production de discours inspiré par la rhétorique médiévale. Ce modèle est adapté à l'activité d'annotation pour la coopération puisqu'il décrit une activité comprenant des phases cycliques d'organisation des concepts (classification de rei), d'élaboration de textes (brouillons, documents amendables dans un processus de re-lecture), en interaction constante avec d'autres utilisateurs (communication dans les phases de structuration des

idées, du document, et communication sur le document publié). La traduction de ce modèle de l'activité en modèle de l'activité instrumentée nous permet de spécifier notre collecticiel.

Conformément au modèle que nous préconisons, le collecticiel doit permettre aux utilisateurs de visualiser un document, de le segmenter, de créer des associations de divers types (indexation, assemblage) entre les différents fragments, de rédiger le discours du corps de l'annotation, ou de la publier. La phase de validation nécessite la mise en place d'une fonction spécifique de type *réponse* à une annotation. Le modèle discursif devant permettre un aller-retour constant entre lecture et écriture, la visualisation est également prépondérante.

La partie suivante est consacrée à la description et à la validation des fonctionnalités développées dans notre collecticiel pour l'herméneutique numérique, AnT&CoW.

4^e partie :

AnT&CoW : Description et simulation

Dans cette partie, nous présentons AnT&CoW (Annotation Tool & Cooperative Work), le collecticiel développé pour permettre une coopération effective entre des utilisateurs distants via l'annotation des documents, et sa simulation sur la base de scénarios et d'un corpus d'annotation.

Le chapitre V présente les fonctionnalités définies selon notre modèle d'activité de l'annotation instrumentée et nécessaires à la gestion de Documents Pour l'Action. Les fonctionnalités sont présentées selon les trois objectifs principaux de l'activité d'annotation : la communication, l'élaboration et l'indexation. Tandis que la communication repose principalement sur des fonctionnalités de visualisation et de saisie d'annotations, l'élaboration nécessite des fonctionnalités plus proches de la gestion de document (édition, publication par exemple). Ces objectifs sont tous deux sous-tendus par une indexation fine permettant la sélection d'annotation que ce soit pour les visualiser ou les colliger et tracer la logique de communication du projet. Les fonctionnalités d'indexation reposent, elles, sur l'utilisation d'outils et de techniques de Traitement Automatique des Langues (TAL). Les outils de TAL sont utilisés pour deux propos, la construction de classification permettant l'indexation et l'assistance à l'indexation. La construction de la classification par le module TAL de l'outil AnT&CoW est fondée sur un corpus de référence. Nous postulons qu'une classification fondée sur ce corpus est adaptée à un groupe-projet en conception mécanique aéronautique.

Dans le chapitre VI, nous décrivons le dispositif et les résultats de la simulation de notre outil. Elle s'appuie sur un corpus de travail centré sur l'annotation et sur un ensemble de scénarios mimant les utilisations de l'outil par des concepteurs en mécanique.

Chapitre V

AnT&CoW, un collecticiel pour la gestion des DoPAs dans un groupe projet

La seule chose que je sais, c'est que je ne sais rien.
Socrate

*Tu sais rien. Si ça se trouve, t'as même pas de
cerveau.*
Michel Audiard³⁹

Les classes d'objet définies précédemment correspondent à des éléments incontournables du soutien d'activités coopératives. Le document en cours d'élaboration met en action divers fragments. Le document est utilisé à la fois pour penser, (« writing for thinking » (Nardi et *al.*, 2004, p. 227)) et pour focaliser l'action afin d'induire des évolutions. Les réflexions menées sur les DoPAs, notamment sur leur documentarisation permettent de préciser les fonctionnalités en terme de gestion documentaire de l'annotation discursive, élaborante ou pour l'indexation. Après avoir rappelé ces réflexions sur les DoPAs, nous présenterons donc les spécifications de notre outil d'annotation dans le cadre du soutien d'activités de conception coopératives.

1. Documents pour l'action et documentarisation

(Cahier, 2005) considère que le DoPA est un type d'artefact qui permet d'inscrire de façon dynamique la connaissance en contexte et en situation, en particulier pour tenir compte des situations interactionnelles qui caractérisent l'activité collective ; « les collectifs engagés dans des activités coopératives sont confrontés aux documents dans un contexte d'activités structurellement ouvertes et spatiotemporellement distribuées, impliquant un usage intensif de documents pouvant être extrêmement fragmentés » (Cahier, 2005, p. 135). Cela oblige ces collectifs à mettre en oeuvre une stratégie de documentarisation, en fournissant aux acteurs des repères pour augmenter leur maîtrise des documents dans de telles activités.

³⁹ tiré du film, *La Grande Sauterelle* (1967), film de G. Lautner dialogue de M. Audiard

La stratégie de documentarisation (Zacklad, 2004, p. 8) consiste à « transcrire ou enregistrer les contenus sémiotiques sur des supports pérennes, puis à doter ces supports d'attributs spécifiques permettant :

1. de faciliter leur gestion parmi d'autres supports ;
2. de faciliter leur manipulation physique, condition d'une navigation sémantique à l'intérieur du contenu sémiotique et enfin ;
3. de faciliter l'orientation des récepteurs, mais également de plus en plus des réalisateurs eux-mêmes, à l'intérieur du support en définissant une ou plusieurs cartographies des contenus sémiotiques susceptibles de guider la navigation sémantique. ».

Dans le cadre d'une documentarisation de nos annotations pour favoriser le travail coopératif, nous proposons de combiner des fonctionnalités de visualisation inspirées des applications d'annotation, des fonctionnalités d'élaboration du WSo et des techniques issues des recommandations du W3C et du Web Sémantique. De cette façon, nous comptons soutenir l'interprétation collective de texte, la discussion autour du document, tout en permettant une indexation fine des contenus. Comme montré dans Magpie (Dzbor *et al.*, 2004), l'indexation facilite la visualisation des annotations liées au document, et donc la navigation et la compréhension de la logique de conception d'une part, et d'autre part l'élaboration de l'interprétation collective du document.

2. Spécifications fonctionnelles

L'outil développé doit assister l'interprétation de documents numériques sur la base d'annotations et répondre à des objectifs de communication, d'indexation et d'élaboration de documents pour permettre le travail distant et asynchrone sur des documents partagés. Il doit prendre en charge l'annotation sous sa forme de balise comme sous sa forme de fragment discursif ou de document. Cela signifie que l'outil soutient aussi bien le classement, que la négociation d'une classification commune et la co-élaboration de document. En soutenant ces différents usages de l'annotation, notre outil s'inscrit dans le Web Socio Sémantique (W2S).

Le tableau 7 ci-dessous décrit les principes généraux pour la conception de notre outil. Il reprend les grands objectifs exposés au cours de l'état de l'art (chapitre I). Il s'agit donc

pour nous de lier des utilisateurs par le biais de documents partagés, mais aussi de lier ces documents partagés via les commentaires déposés et leur type. Pour cela, l'outil doit proposer un espace partagé par une communauté. Les fonctionnalités de l'outil doivent répondre à des besoins de classement (mise en place d'une classification, d'un ensemble de classes), de classification (typage des éléments selon la classification construite) et d'élaboration (de la classification, des documents et des commentaires). Pour permettre la coopération, les utilisateurs doivent bénéficier d'une interface Web permettant d'utiliser les différentes fonctionnalités (client) et d'un espace de stockage partagé (serveur).

Tab. 7. SPÉCIFICATIONS GÉNÉRALES D'UN OUTIL D'ANNOTATION POUR LE TRAVAIL COOPÉRATIF

Objectif	Famille	Activité	Activité instrumentée	Objet ancré	Ancre	Spécification
Traçabilité de la communication Gestion des documents et des DoPAs	W2S	Élaboration Classification Classement	Annotation Postage Indexation semi-automatique	Texte en langue naturelle Classe Balise	Documents Fragments de document Fragments discursifs	Interface Web Coopération synchrone et asynchrone Architecture distribuée Espace privé Authentification

Comme présenté avant, nombre d'outils peuvent permettre d'accéder aux activités décrites ci-dessus, mais aucun ne les permet comme un ensemble global d'une même activité d'interprétation coopérative. Il serait cependant dommage de ne pas utiliser des fonctionnalités de base mises en place dans plusieurs outils Open-Source implémentant les standards W3C (protocole Annotea). Des outils présentés en chapitre 1, le plus intéressant puisque le plus fonctionnel et possible à mettre à jour est le client Annozilla.

Les fonctionnalités que nous souhaitons implémenter sont regroupées en fonction des différentes dimensions de l'annotation dans le diagramme de cas d'utilisation en figure 15. Il s'agit de trois paquetages principaux de fonctionnalités permettant la communication, l'indexation et l'élaboration. L'élaboration est permise par la rédaction et la publication des annotations, c'est-à-dire grâce à des fonctionnalités d'édition du fragment ou du document

auquel il est ancré et des fonctionnalités de mise en partage des annotations. La communication est permise par la négociation et la structuration d'idées c'est-à-dire des fonctionnalités d'envoi d'annotation, de réponse à une annotation et de typage des annotations selon une classification construite pour structurer le domaine de la communauté. L'indexation est permise par la segmentation et l'indexation elle-même c'est-à-dire la sélection d'une fragment à annoter et la récupération automatique de ses métadonnées (emplacement géographique, date, auteur) et le typage effectué selon la classification (type argumentatif, du domaine, du rôle,...). Dans un contexte de soutien à l'interprétation de documents textuels, nous proposons de soutenir l'indexation grâce à des outils de Traitement Automatique des Langues (TAL). Le soutien à l'indexation d'annotation implique que les outils de TAL soient capables d'analyser un texte court (l'annotation soumise par l'utilisateur) et de faire correspondre les éléments de ce texte à des concepts stockés dans une carte de connaissances. L'utilisateur décidera ensuite si l'indexation proposée est pertinente.

L'outil développé afin de soutenir la coopération autour de documents partagés dans un groupe s'appelle AnT&CoW (ANnotation Tool & COoperative Work). À partir d'un logiciel existant, Annozilla, des fonctions de communication (la négociation et la structuration), d'élaboration et d'indexation ont été implémentées en un plug-in pour navigateur Web (Mozilla-Firefox). AnT&CoW est un outil défini pour la coopération autour de document entre des utilisateurs distants travaillant de manière asynchrone. L'ajout de ces fonctionnalités est représenté par l'analyse différentielle en vue du développement de l'outil AnT&CoW sur la base de l'outil Annozilla/ZAnnot (tableau 8).

Tab. 8. ANALYSE DIFFÉRENTIELLE ANNOZILLA-ZANNOT / ANT&COW

	Fonctions de Base Annozilla – Zannot	Fonctions d'AnT&Cow
Niveau client	Création et modification d'annotation Téléchargement des annotations et de leurs corps Options de langues et de type	Réponse à annotation Multi-ancrage de l'annotation Module de visualisation orientée Document (édition) et Classification (édition de la ressource pour la classification)

		Module de recherche Indexation semi-automatique (TAL) Export d'un groupe d'annotations au format texte
Niveau serveur	Stockage des annotations selon un schéma RDF Gestion des droits	Modification du schéma de l'annotation Stockage de documents Encapsulation d'un module de TAL Création semi-automatique de Ressources pour la classification

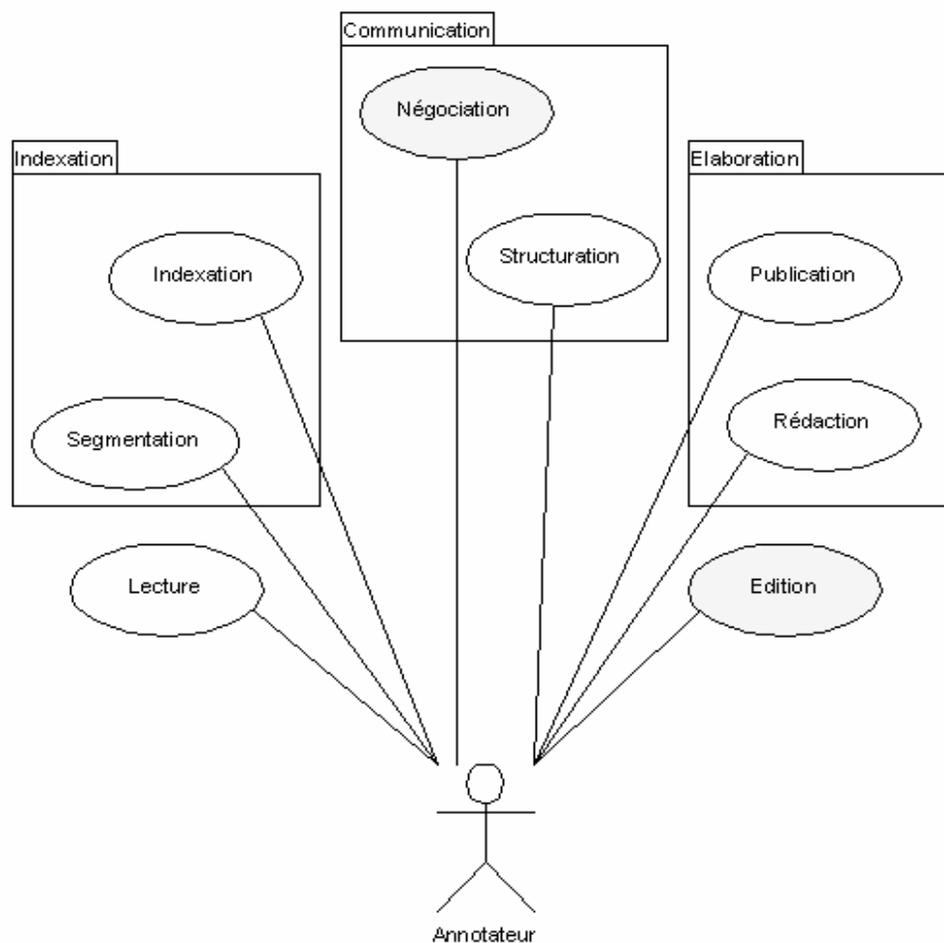


Fig. 15. PAQUETAGES DES CAS D'UTILISATION DE L'OUTIL

Nous allons maintenant présenter l'architecture de notre outil d'annotation et les trois types de fonctionnalités illustrées en figure 15.

3. Une architecture client-serveur

Notre choix de client Web s'étant porté sur Anozilla pour des raisons de développement et de pertinence, nous nous intéressons donc au serveur d'annotation qui lui est dédié, ZAnnot. ZAnnot est un serveur d'annotation disponible sur la plateforme Zope qui sera la base de notre serveur. Nous détaillons ci-dessous l'architecture et les fonctionnalités d'AnT&CoW. Pour mettre en commun les documents, un espace partagé est dédié sur notre serveur et est accessible par tout utilisateur authentifié.

3.1. Le serveur

Le serveur Zannot que nous avons adopté pour la gestion des annotations est développé sous la plateforme Zope (Z Object Publishing Environnement) (Latteier *et al.*, 2003), un environnement Open Source (développé en Python) fournissant des services pour la publication et la gestion d'objets Web. Zope est une plateforme composée de plusieurs objets qui facilitent le développement des applications Web (voir Fig. 16) : un serveur Web (ZServer), une base de données Objet (ZODB) et un ensemble de classes prédéfinies (ZClass) pour accélérer le développement d'application. Zope offre un mécanisme simplifié pour rendre plusieurs applications accessibles par un même site Web. Ce mécanisme permet tout d'abord l'inclusion facile d'une application Zope (Zope Product) et offre la possibilité d'accéder à chaque objet Web (ressource Web) par une URL. Le contenu (objets) de la plateforme Zope est facilement manipulable à travers une interface Web.

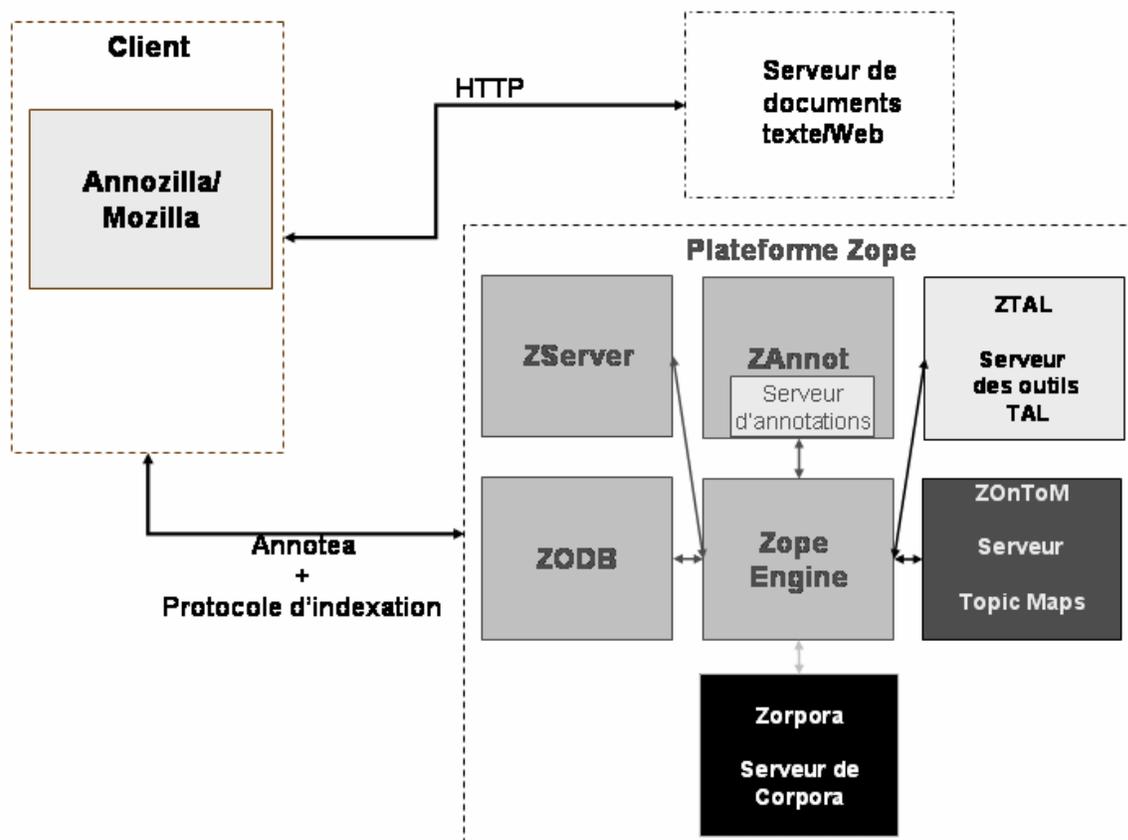


Fig. 16. ARCHITECTURE D'ANT&CoW

ZAnnot est une application de la plateforme Zope, un produit Zope. ZAnnot reprend tous les avantages de Zope et est conforme au standard Annotea. La plateforme Zope nous permet donc de gérer les annotations disponibles sur ZAnnot et nous permet ainsi d'ajouter, de supprimer, de mettre à jour, bref de gérer les annotations. Dans un premier temps, nous avons testé le serveur d'annotation ZAnnot tel que, puis nous avons étendu le protocole Annotea et les fonctions de base de ZAnnot pour l'adapter à notre problématique d'annotation coopérative de documents numériques. Les réponses à une annotation ne sont pas gérées dans la version testée, mais ont été rapidement mises en place sur le serveur. Notre système d'annotations est caractérisé par quatre modules : un module de visualisation, un module de création d'annotations, un module de stockage et un module de recherche.

Les modules de stockage et de recherche sont disponibles d'une façon basique sur l'application serveur ZAnnot. Un cinquième module émerge pour notre système, celui de l'indexation. Ce module doit être présent au niveau de l'application serveur afin de pouvoir

gérer les index et répondre aux différentes requêtes de manipulation des points de vue envoyées par le client. Nous avons donc modifié le protocole Annotea (pour la réponse à annotation, l'indexation et la collection d'annotation) et notre outil communiquera avec le client par le protocole étendu suivant (voir Fig. 17) :

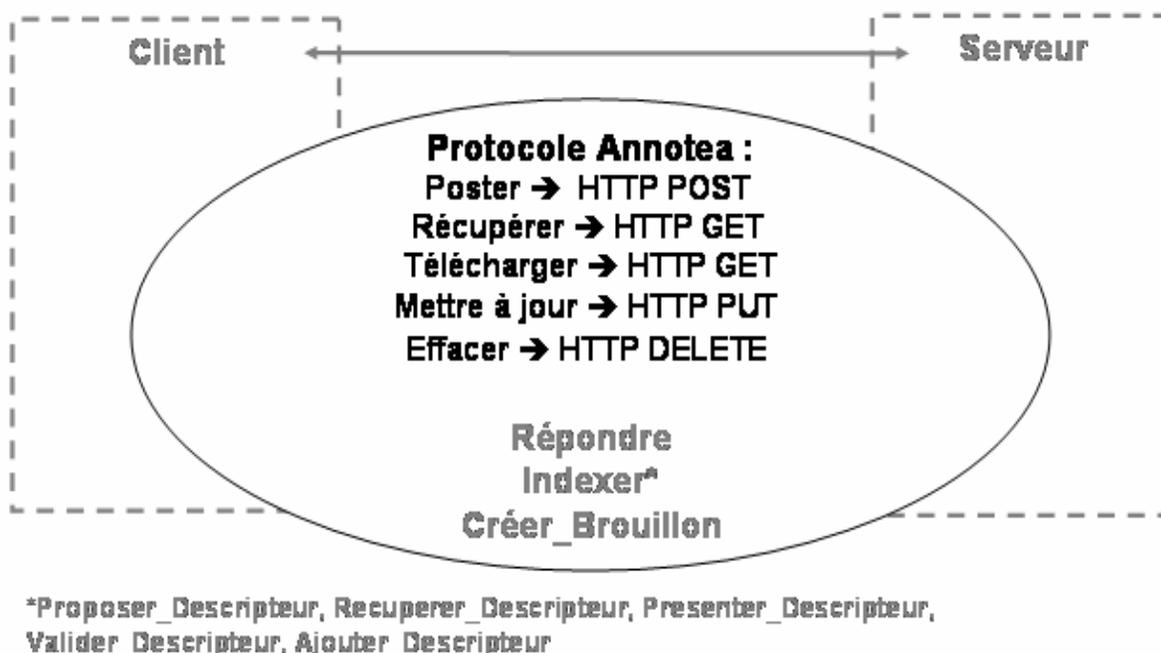


Fig. 17. PROTOCOLE DE COMMUNICATION D'ANT&COW

La réponse est simplement mise en place par la gestion de *inReplyTo* défini dans le standard Annotea. La création de brouillon est une fonctionnalité d'export d'un ensemble d'annotation sous forme d'un fichier homogène accessible en partage. L'indexation est semi-automatique et est permise par un ensemble de messages envoyés du serveur au client tels que : *ProposerDescripteur*, *RécupérerDescripteur*, *PrésenterDescripteur*, *ValiderDescripteur*, *AjouterDescripteur*.

Dans une perspective d'indexation et de soutien à l'indexation, nous avons mis en place un serveur ZTAL qui permet d'encapsuler les outils de TAL sur la plateforme Zope (voir Fig. 18). Ce module nécessitant des documents et produisant une classification semi-formelle, deux modules lui sont adjoints, visibles dans ce schéma, Zorpora, contenant le corpus de documents du projet et ZOnToM, le serveur d'ontologie. Les annotations et les documents sont donc archivés sur des serveurs différents, respectivement ZAnnot et Zorpora. Le document peut ainsi rester intact et être réutilisé pour d'autres propos.

Nous avons implémenté le serveur ZTAL qui met à disposition des outils d'aide à la construction et à la mise à jour de la RTO pour le traitement des annotations et des documents. Il est aussi monopolisé pour la mise en correspondance du texte et de la classification pour l'indexation des annotations. Il traite le texte ou le corpus textuel de l'annotation de son étiquetage à son indexation. Le schéma ci-dessous (Fig. 18) modélise le fonctionnement du module ZTAL et ses interactions avec ZOnToM, Zorpora et ZAnnot.

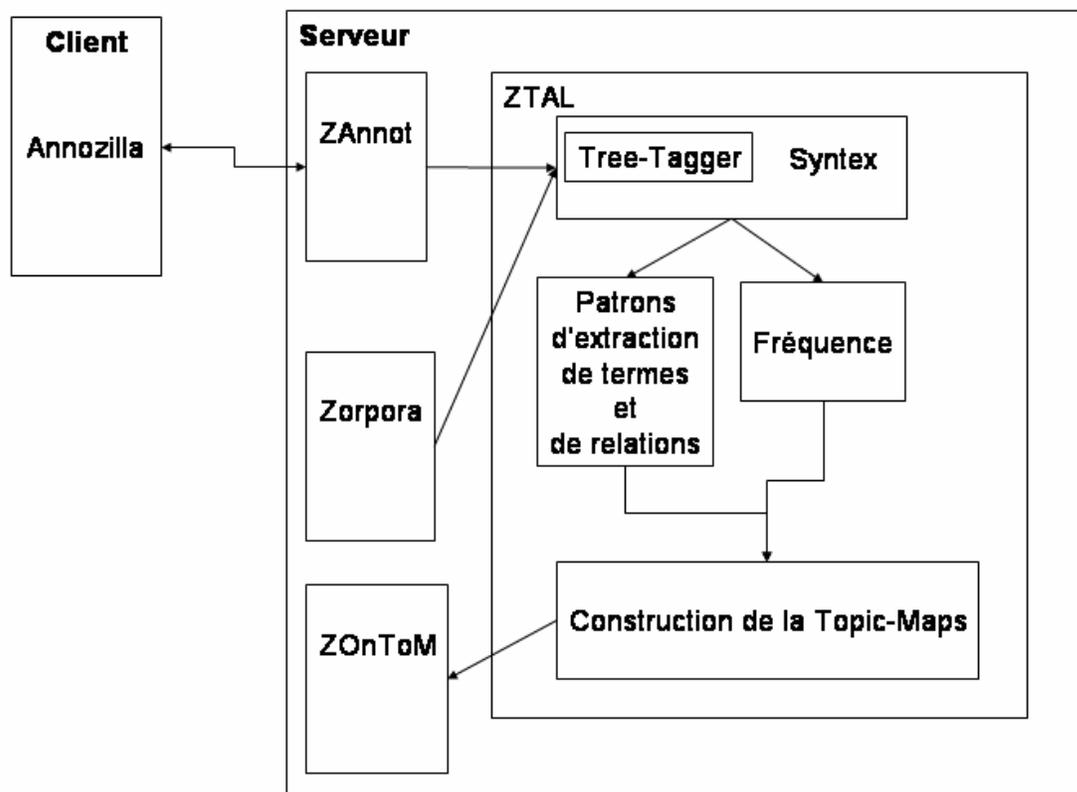


Fig. 18. MODULE ZTAL

Une fois ces différents modules mis en place, les fonctionnalités permettant les interactions Homme-Machine ont été implémentées, non seulement au niveau du client, mais aussi au niveau du serveur pour permettre un partage des documents (bibliothèque, annotations) et de la classification entre tous les utilisateurs.

3.2. Le client

Dans le cadre du travail médiatisé d'utilisateurs non spécialistes de l'informatique, il est primordial d'avoir des fonctionnalités simples et un outil simple d'utilisation. C'est pourquoi les fonctionnalités plus techniques de l'outil sont gérées et lancées du serveur. Le client est ainsi léger et tourne aisément sur toute machine. De même, le client doit être simple d'utilisation et s'intégrer aux pratiques de l'utilisateur. Dans notre contexte d'utilisation, il doit permettre d'annoter des documents partagés via le Web. Le client propose une interface à l'utilisateur qui lui permet de manipuler, de visualiser et d'organiser les annotations. Comme présenté ci-dessus, nous nous sommes tournée vers un des outils développés dans le cadre du groupe de travail W3C pour la coopération, le plug-in de navigateur Web Mozilla (McFarlane, 2003), disponible et fonctionnel, Anzozilla (Fig. 19). Annotea on Mozilla (Wilson *et al.*, 2004) se présente sous la forme d'un plug-in Open Source (Javascript, XUL, XPCOM, HTML) respectant le standard Annotea, à l'interface simple. Il permet de voir et de créer des annotations associées à une page Web. Anzozilla envoie les informations liées aux annotations au format RDF. Il utilise la technologie XPointer pour permettre de localiser les annotations sur la page Web. Ses fonctionnalités de base sont celles du standard Annotea : poster, récupérer, télécharger, mettre à jour, effacer. Ces fonctionnalités ont été étendues (visualisation, indexation) ou développées (collection d'annotations, création d'un brouillon, réponse à l'annotation). Nous allons exposer dans les sections suivantes les différentes fonctionnalités de notre outil, sur le plan de la visualisation pour soutenir la communication, de l'élaboration et de l'indexation.



Fig. 19. LE PLUG-IN D'ANNOTATION

4. Fonctionnalités du client AnT&CoW

4.1. Un espace partagé sur le serveur

L'espace partagé que nous mettons à disposition sur un serveur Web est à la croisée entre la gestion de document et la gestion de contenu. Inspiré des Wikis, le site développé en PHP offre une grande liberté aux utilisateurs. Ce site est aussi proche des sites de CSCWriting quant à la gestion des documents. Les utilisateurs en s'authentifiant (Fig. 20) sur le site ont possibilité de déposer des documents (*.ml, .txt) afin de les partager et de les éditer pour les modifier. Ils créent ainsi une bibliothèque partagée.

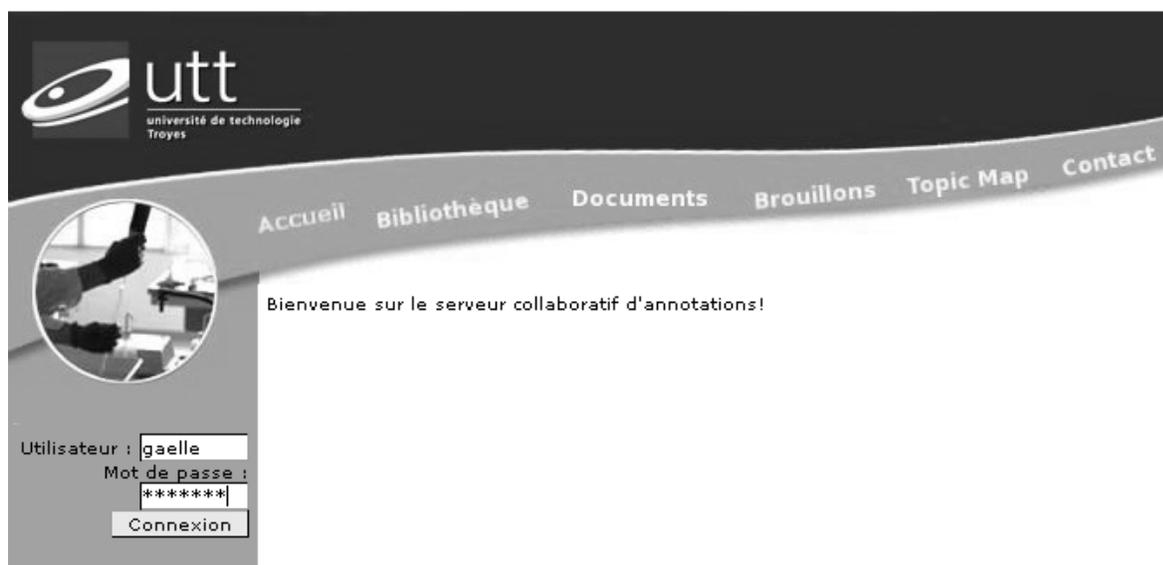


Fig. 20. ACCUEIL SUR L'ESPACE PARTAGÉ

Partager un document

Les utilisateurs peuvent donc ajouter des documents sur le serveur de document Zorpora via l'interface Web suivante (Fig. 21). Ils vont chercher leurs documents en local sur leur machine pour les mettre à disposition de tous sur le serveur (Fig. 22).

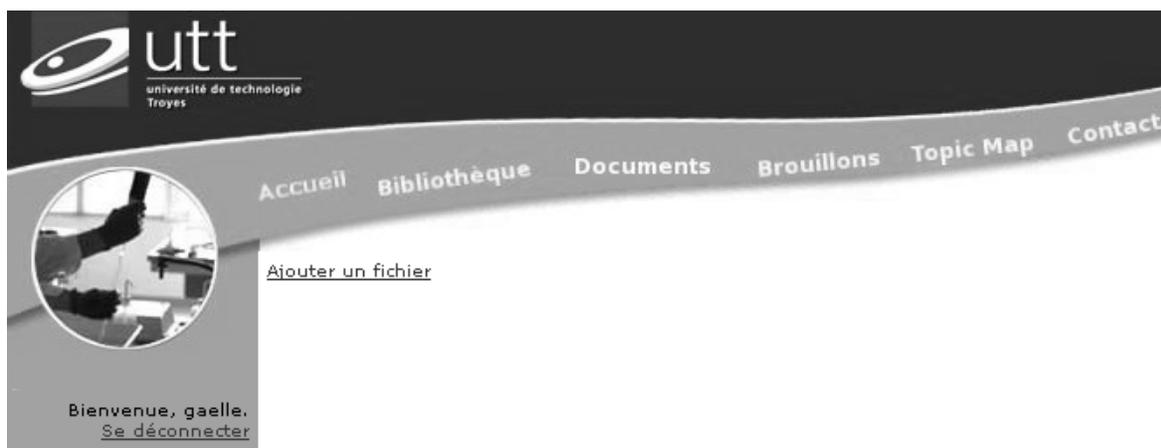


Fig. 21. BIBLIOTHÈQUE DE L'ESPACE PARTAGÉ

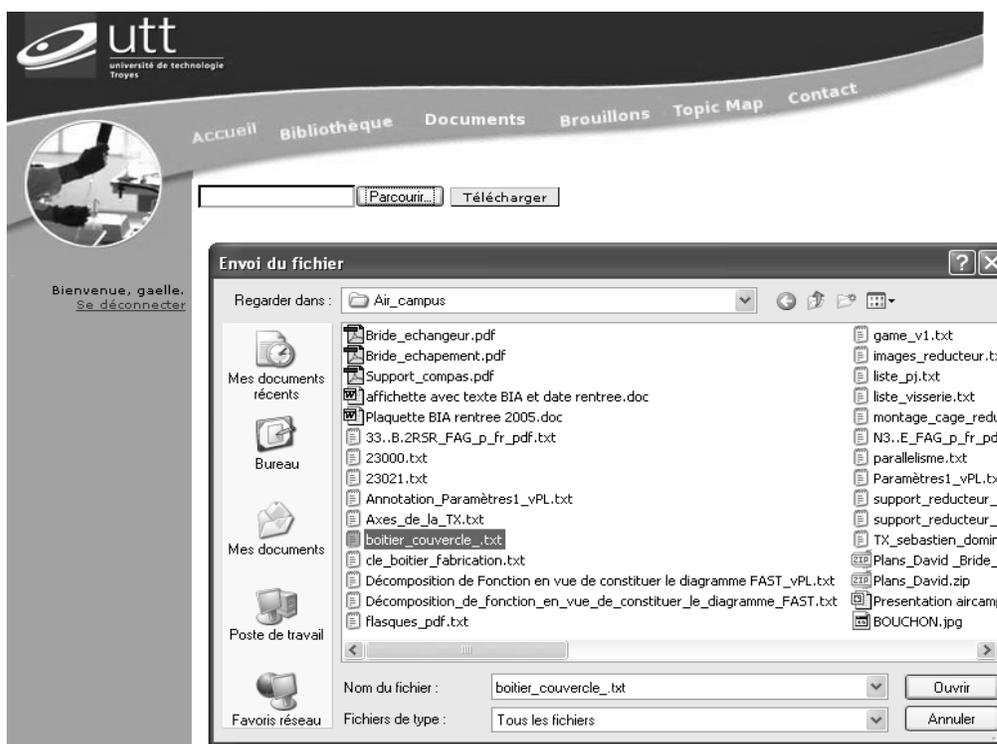


Fig. 22. CHOIX DU FICHIER A TELECHARGER

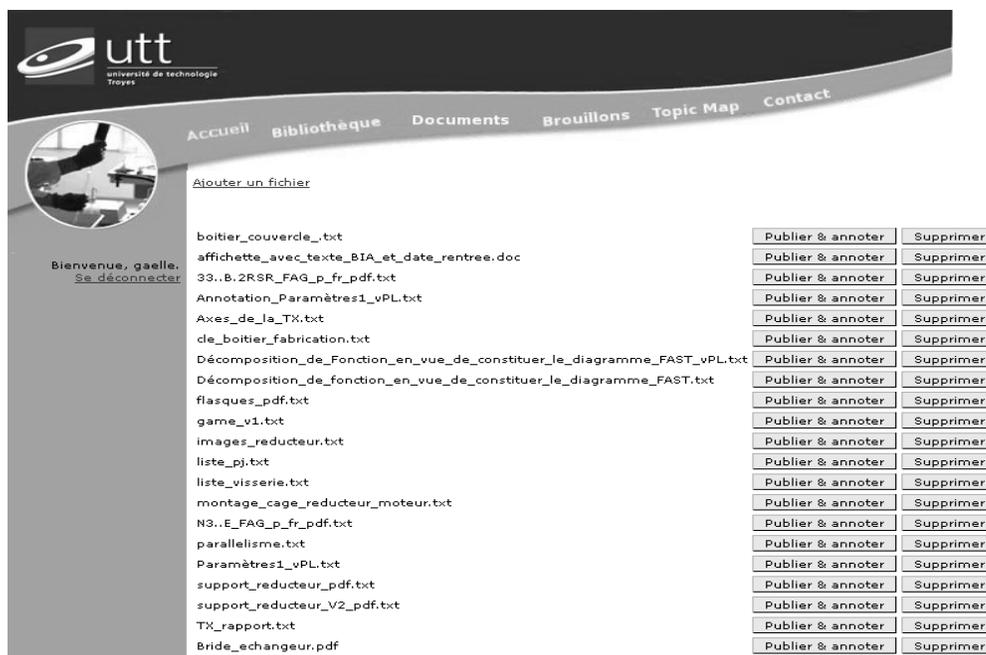


Fig. 23. BIBLIOTHÈQUE DE DOCUMENTS PARTAGÉS

Les documents sont disposés en liste chronologiquement, selon leur mise en partage sur le serveur. Une fois mis en partage, il est possible de les supprimer ou de les afficher pour les annoter. Ils sont disponibles dans la bibliothèque du projet (onglet bibliothèque).

Éditer un document

Pour annoter un document, il faut l'ouvrir en édition via l'interface Web (Fig. 24) par le bouton « éditer et annoter ».

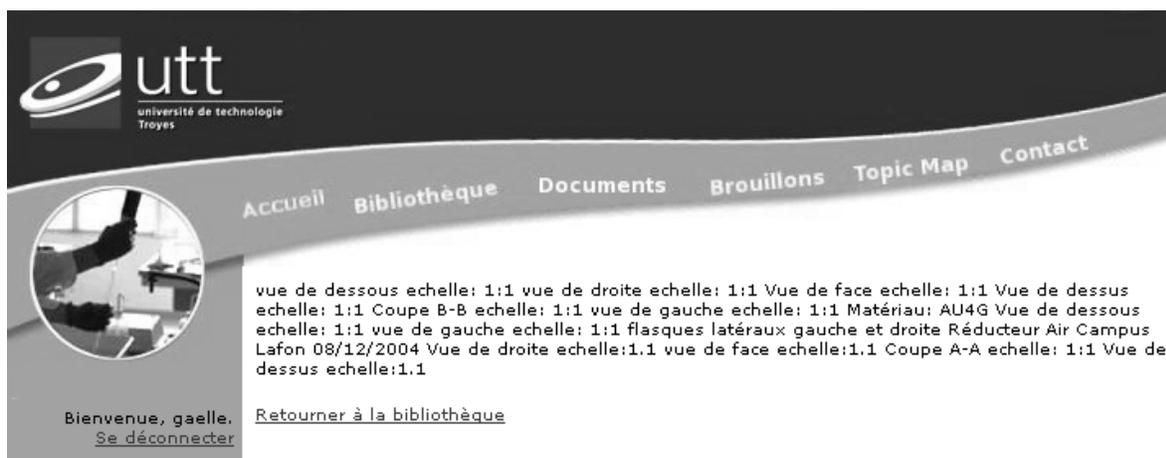


Fig. 24. ÉDITION D'UN DOCUMENT

4.2. Des fonctionnalités pour la communication

Dans le modèle d'activité instrumentée, la fonctionnalité de communication est représentée par des fonctions permettant la négociation et la structuration de schèmes de classification pour l'organisation des concepts partagés par le groupe. Pour structurer ou négocier, il est primordial de pouvoir contextualiser son intervention et donc d'avoir une vision d'ensemble du document lu et de ce qui s'est dit sur ce document. Pour annoter, il est donc nécessaire de pouvoir lire le document et afficher les commentaires déposés auparavant et de pouvoir en sélectionner des parties (surligner, entourer, mettre en couleur, modifier la police,...). Il s'agit ensuite d'ancrer des fragments de texte à ces sélections. L'ancrage représente un commentaire lié à une partie du texte, une réponse à une annotation déposée par un autre utilisateur voire une réponse à des fragments de documents ou de discours tout à la fois.

Communiquer à l'aide d'annotations nécessite des fonctionnalités adaptées de visualisation et de saisie des annotations. Nous soutenons cette communication via les annotations en permettant non seulement la création de fragments de discours par la saisie de texte, mais aussi la visualisation de l'ensemble du fil de discussion par l'affichage des réponses aux annotations par exemple. La première fonctionnalité définie est donc une *fonctionnalité de visualisation* des DoPAs qui permet la sélection de fragments de documents (surligner, entourer, déposer une marque) et l'ancrage d'annotations dans le document (déposer une annotation, répondre à une annotation, multi-ancrage de l'annotation sur différents fragments de documents, ...).

Le plug-in Annozilla permet de visualiser les emplacements des annotations déposées dans un document et de saisir le corps d'une annotation. La visualisation permet l'interaction de l'utilisateur avec les autres utilisateurs via la machine. Les utilisateurs peuvent alors communiquer, négocier et structurer leurs documents et leurs annotations. Pour coordonner leurs actions et mettre en place des interactions, il est nécessaire qu'ils disposent d'une vision d'ensemble du document lu et des annotations déposées précédemment (visualisation du fil de discussion). Il est bien entendu crucial de leur fournir des outils de sélection (surligner, entourer, mettre en couleur, modifier la police,...) pour l'ancrage d'annotation ce que fait Annozilla. Cependant, nous avons étendu cet ancrage puisqu'il est nécessaire qu'il soit de plusieurs types et assez souple pour s'adapter aux usages de l'utilisateur. Nous proposons donc l'ancrage classique à une partie du texte, mais aussi l'ancrage d'une réponse à une annotation déposée par un autre utilisateur, l'ancrage d'une annotation à plusieurs parties de textes... Annozilla propose plusieurs de ces fonctionnalités. Bien que toutes pourraient être optimisées (sélection, visualisation), nous nous sommes focalisée sur l'implémentation de fonctionnalités d'ancrage.

4.2.1. Ancrer une annotation

Afin d'ancrer une annotation à un document, il est nécessaire de sélectionner un document (clic droit sur le document) ou une partie du document (surligner une partie du document). Cette sélection a lieu avant de saisir le corps de l'annotation (Fig. 25).



Fig. 25. ANCRAGE D'UNE ANNOTATION

Cependant, parfois, un argument se construit en se référant à plusieurs parties de documents en opposition. Par exemple, un fait exposé dans une partie A (vitesse élevée possible) semble en contradiction avec un élément d'une partie B (charge élevée possible). Annozilla ne permet pas une explicitation de l'interprétation en fonction de plusieurs éléments (fragments de document, annotation) et donc ne permet pas de partager une interprétation. Nous avons donc proposé une fonctionnalité de multi-ancrage. Celle-ci consiste en la sélection de plusieurs fragments reliés entre eux via une annotation. L'annotation liée à la fois au contexte A et au contexte B pourrait alors être « Une charge élevée ne limiterait-elle pas la vitesse ? ». Le multi-ancrage est utilisable par un aller-retour entre une première sélection et le texte. Les sélections sont visibles dans l'encadré Fig. 26.

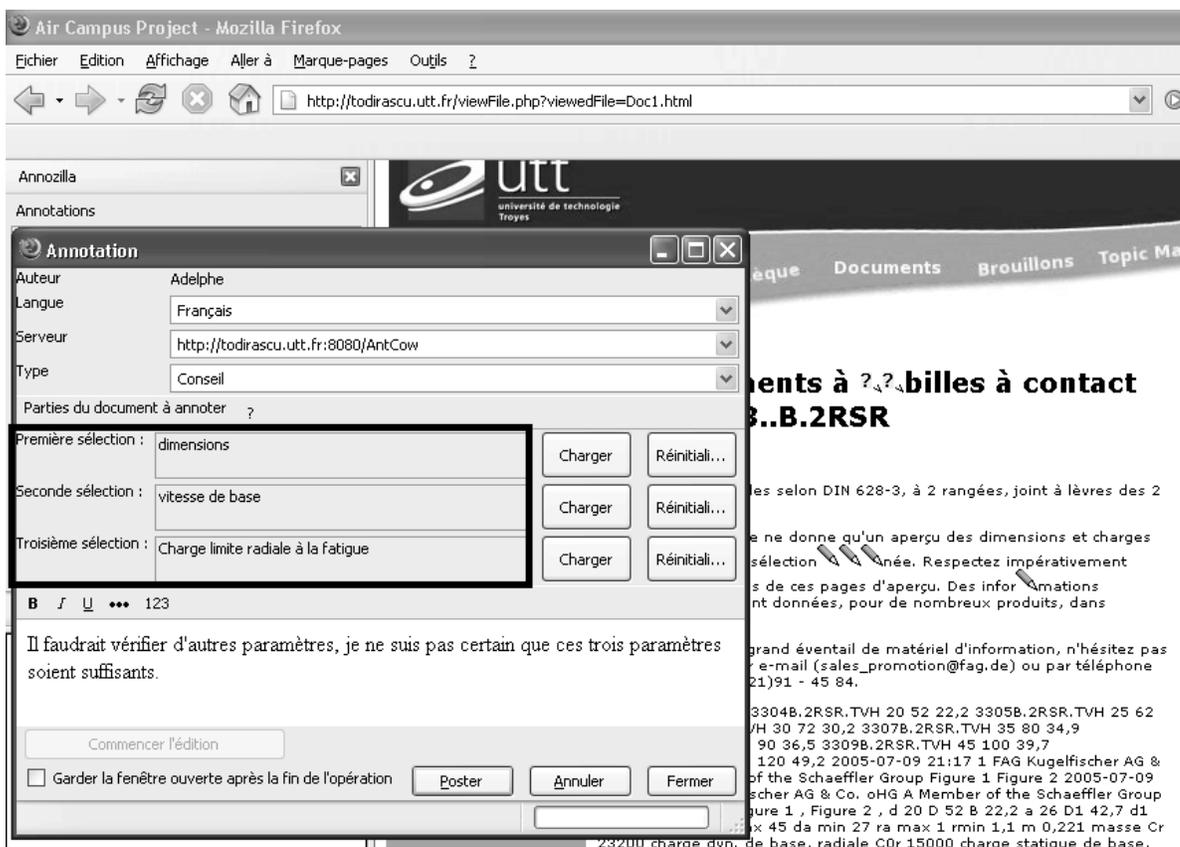


Fig. 26. MULTI-ANCRAGE

Nous avons conservé la fonctionnalité d'archivage proposée par Annotzilla. Les annotations sont stockées sur un serveur différent du serveur de document, ce qui permet de disposer du document original, sans annotation. Il faut sélectionner l'affichage des annotations dans le plug-in (Fig. 27). Cependant, le schéma de l'annotation stockée a été modifié, incluant en plus du multi-ancrage, une possibilité d'ajout de descripteur (voir Fig.40).

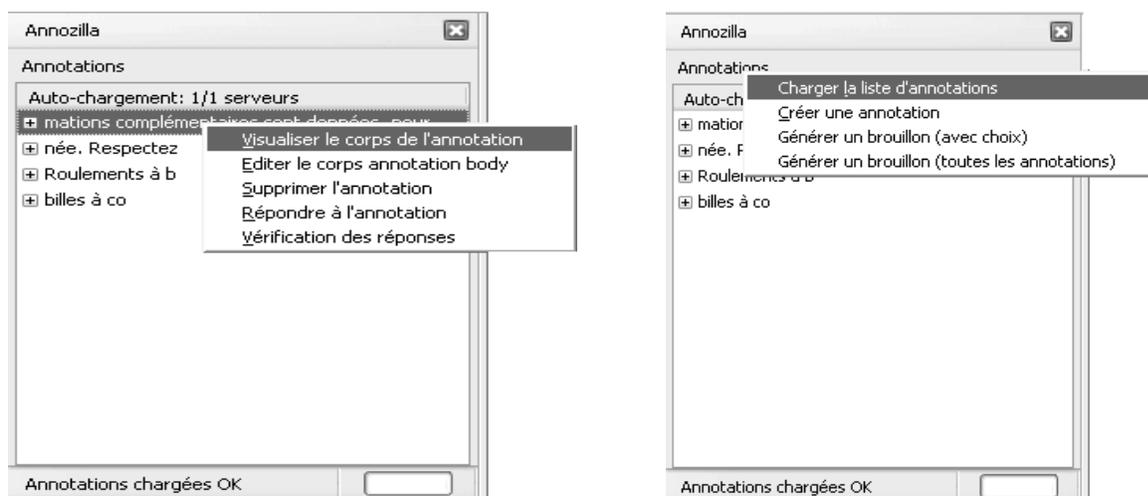


Fig. 27. FONCTIONNALITÉ DU PLUG-IN : AFFICHER, CRÉER, RECHERCHER, GÉNÉRER UN BROUILLON

4.2.2. Saisir une annotation

Suite à une sélection et un clic droit, une fenêtre *pop-up* s'affiche au premier plan sur la fenêtre du navigateur. Elle permet de sélectionner un index, puis de saisir l'annotation. En effet, nous considérons que l'utilisateur peut indexer manuellement la portée argumentative de l'annotation qu'il s'apprête à rédiger. Comme on ne peut prédire les pratiques d'indexation de l'utilisateur, ce dernier peut aussi bien rédiger le corps de l'annotation avant de la typer en cliquant sur le bouton *commencer la saisie* encadré dans la figure 28.

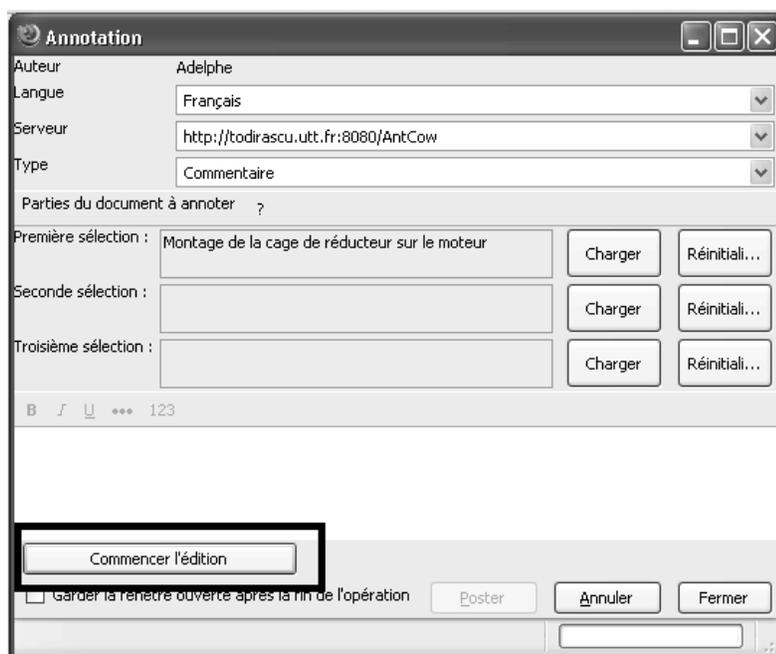


Fig. 28. SAISIE D'UNE ANNOTATION

Il est possible après une saisie de mettre à jour annotation déposée précédemment (voir Fig. 29).

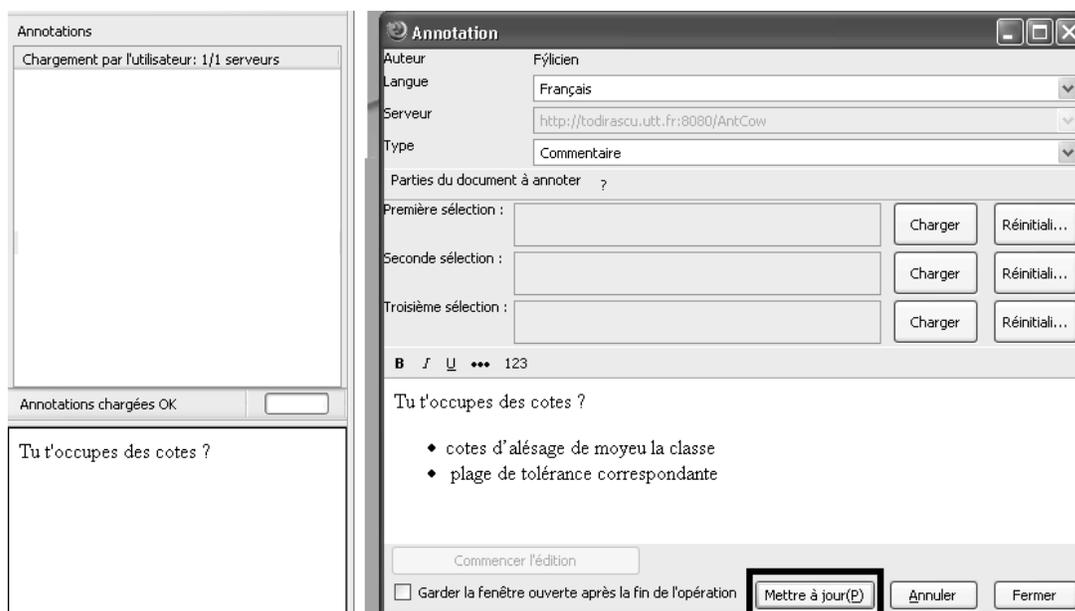


Fig. 29. MISE À JOUR D'UNE ANNOTATION

4.2.3. Visualiser le fil d'annotation

Le fil d'annotation est visible dans la partie droite du plug-in AnT&CoW (encadré Fig. 30).

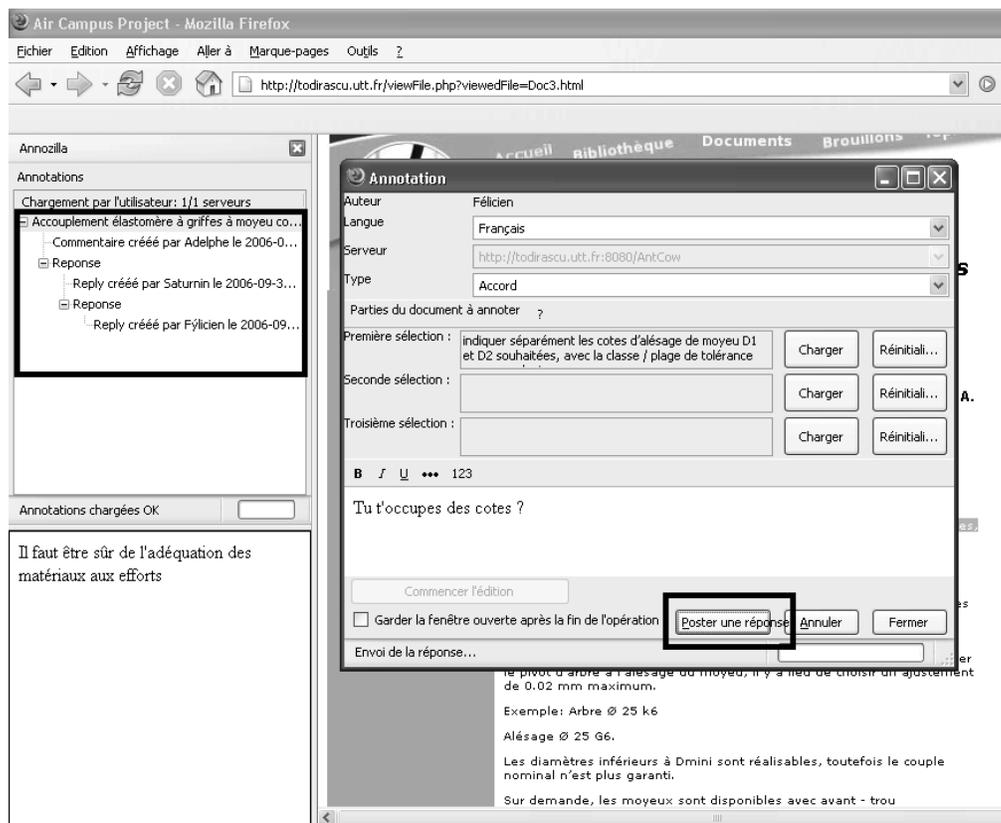


Fig. 30. FIL D'ANNOTATIONS

Il est aussi possible de le voir dans son ensemble en recueillant l'ensemble des annotations dans un document.

4.3. Des fonctionnalités pour l'élaboration

L'analyse de la construction de texte à partir de ses sources écrites est qualifiée par Derrida (1972) de déconstruction. Selon Derrida, les archéo-écrits ou les archéo-traces participent de l'élaboration du texte. Ces considérations rejoignent les observations de la littérature (Chapitre IV, section 1), principalement celles de l'analyse diachronique des activités d'écriture médiévales. En effet, l'annotation déposée est la base de l'élaboration de nouvelles interprétations et d'échanges. Il s'agit donc de pouvoir la réutiliser en vue de

l'élaboration d'un nouveau document récapitulant les interprétations et concepts formés collectivement.

La création de documents revient à faire une collection de concepts recueillis de différents documents, fragments de document ou négociations orales et de les homogénéiser en les présentant en un ensemble reconnu comme un document dans son entier. Les participants du projet doivent avoir la possibilité de sélectionner les fragments produits au cours du projet par une réflexion collective et de les colliger selon leur choix en documents textuels homogènes. Ainsi une fonctionnalité de collection d'annotations permettra la création et l'enregistrement d'un brouillon constitué de fragments choisis (annotations et documents) et son partage pour de nouvelles négociations au sein du groupe.

Cette création de document, comme le décrit la phase d'Emendare ou de validation, n'est pas une simple phase finale d'élaboration. Une fois le document produit coopérativement enregistré, il est possible de l'éditer afin de le commenter, le réviser, etc. et ainsi continuer le processus d'élaboration.

Afin de réutiliser l'annotation en vue de l'élaboration d'un nouveau document, nous avons implémenté une fonctionnalité de collection d'annotations ou fragments de documents. Il est possible à la fois de colliger l'ensemble des annotations, ou de les colliger selon le choix d'un descripteur (auteur, date, thème de la classification).

4.3.1. Recueil d'annotations

Le recueil d'annotation se fait par une liste déroulante dans le plug-in AnT&CoW (Fig. 27 *générer un brouillon*). Le brouillon créé par le recueil d'annotation (Fig. 31) est archivé sur le serveur Zorpara et visible dans l'onglet *Brouillon* du serveur.

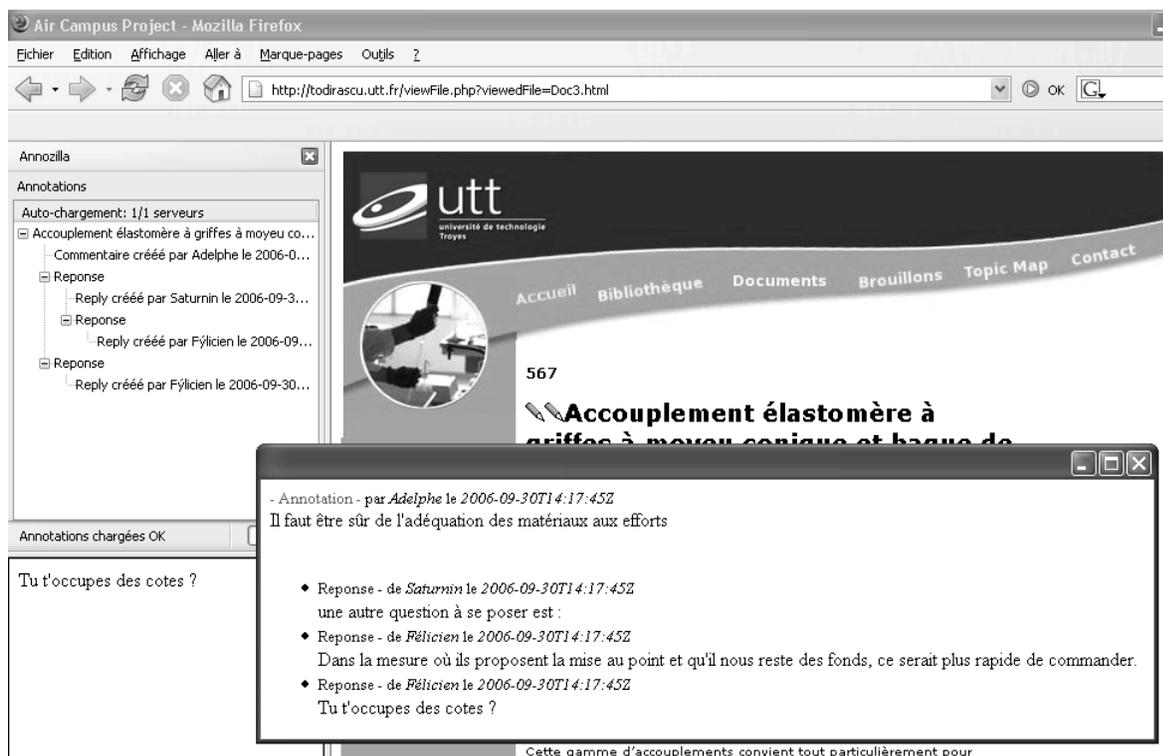


Fig. 31. COLLECTION D'ANNOTATION POUR L'ÉLABORATION D'UN BROUILLON

4.3.2. Sélection suivant l'indexation

Il est possible de collecter les annotations en les choisissant suivant le descripteur qui lui a été lié et de créer ainsi un brouillon thématique (Selon, l'auteur, le domaine, l'argumentation, etc. voir Fig. 32).

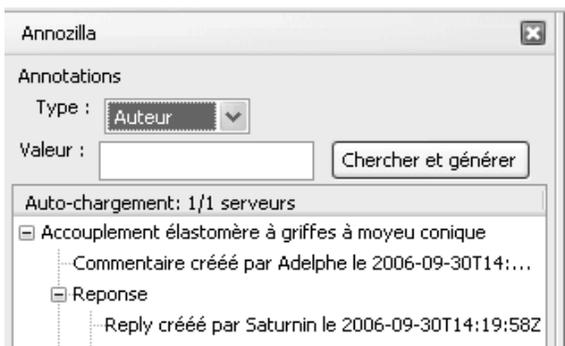


Fig. 32. SÉLECTION SELON UN DESCRIPTEUR

4.3.3. Édition d'un brouillon

Le brouillon créé par la concaténation du corps de plusieurs annotations est éditable afin de le remanier. Il est ainsi possible de partager un brouillon (Fig. 33).

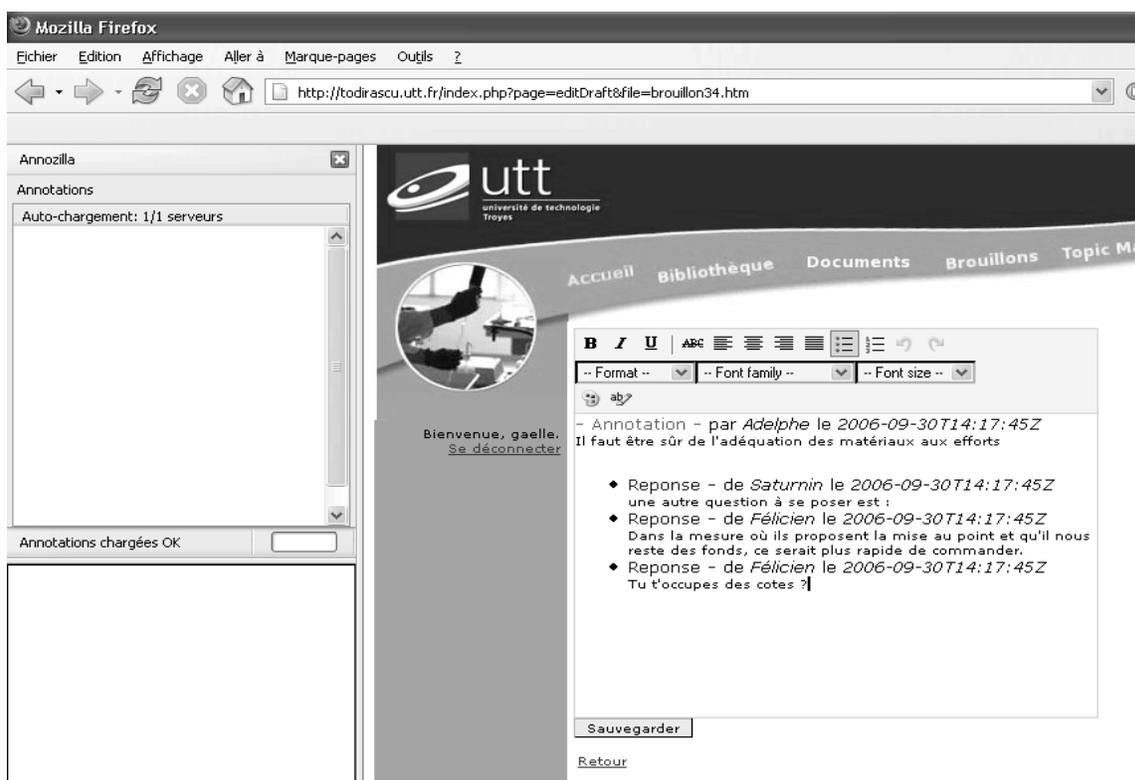


Fig. 33. ÉDITION D'UN BROUILLON

4.3.4. Enregistrement d'un brouillon

Une fois modifié, un brouillon devient un document (voir Fig. 34). Notre outil ne permet pas encore de marquer les différentes étapes du brouillon au document. Cependant, le brouillon est mis en mémoire comme document de travail, non comme document amendable. Les documents amendables constituent la bibliothèque du projet. Le brouillon permet de marquer des phases de négociation. Il serait intéressant de pouvoir indiquer des phases de ratification qui permettent de faire évoluer un brouillon vers un document. En effet, tout comme la négociation, la ratification est un processus qui se fait en coopération pour autoriser la publication d'un document (Pilla et Ritterskamp, 2006).

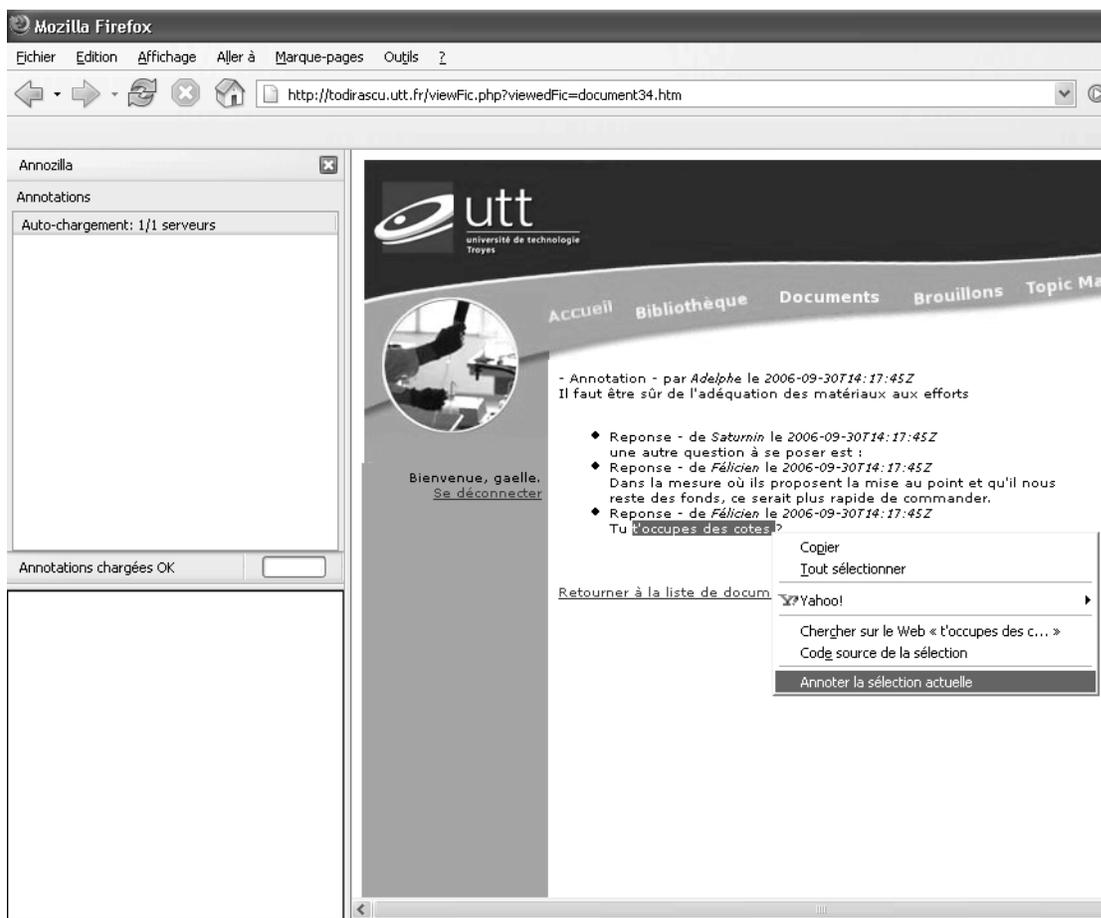


Fig. 34. SAUVEGARDE D'UN BROUILLON, CRÉATION D'UN DOCUMENT PARTAGÉ

4.4. Des fonctionnalités pour l'indexation

La lecture des annotations, et la navigation dans cet ensemble textuel fragmenté ne peuvent être possibles que par une indexation fine des annotations. L'indexation permet la structuration des annotations en une carte de connaissances éditée et lisible par des utilisateurs. Afin d'indexer finement ses fragments, l'utilisateur doit créer son schéma de classification qui sera alors adapté à son vocabulaire et à sa pratique. Cette tâche peut rapidement s'avérer coûteuse ou fastidieuse pour l'utilisateur. Pour remédier à cela, nous utilisons des outils de TAL qui proposent à l'utilisateur des termes liés au domaine et au type argumentatif de l'annotation. L'utilisateur choisit les termes qui lui semblent les plus appropriés. S'il les considère impropres, il peut ajouter ses propres concepts et termes.

L'indexation de l'annotation peut se faire automatiquement par l'outil (date, auteur, codifications des annotations répondues, fil de discussion chronologique automatique) ou manuellement par l'utilisateur, selon trois dimensions : domaine, argumentation, organisation suivant (Zacklad *et al.*, 2003). Dans le second cas (choix d'une valeur représentant le contenu de l'annotation pour chacune des dimensions), l'indexation peut être fastidieuse et nous souhaitons donc soutenir l'utilisateur dans cette tâche en lui proposant des termes (mots-clés) automatiquement grâce à des outils de traitement automatique de la langue. Pour cette opération, nous utilisons un algorithme de mise en correspondance (section 5.2.6.) entre le corps de l'annotation et une classification adaptée. Le système d'annotation propose alors à l'utilisateur des mots-clés pour chacune des dimensions. L'utilisateur décidera ensuite si l'indexation proposée est pertinente et s'il faut alors la conserver comme métadonnée de son annotation. Une fois la validation effectuée par l'utilisateur, l'annotation est stockée avec ses métadonnées sur le serveur d'annotations. La classification utilisée dans la tâche d'indexation doit être mise à jour pour rester pertinente tout au long du projet. Les utilisateurs doivent donc être à même de créer et de mettre à jour la classification ainsi que de l'utiliser pour l'indexation de leurs annotations. Ces fonctionnalités liées à l'indexation permettent à l'utilisateur de comprendre la logique de communication de leur projet.

5. Une classification ajustée

De nombreux types de classification existent. Ils naissent de problématiques d'optimisation de recherche d'information et de normalisation dans une optique WS à l'aide des ontologies, ou de problématiques d'archivages de documents ou de documents numériques comme les thésaurus décrits par une communauté de professionnels de l'indexation, ou encore de problématiques de classification pour l'échange public comme les folksonomies et autres classifications informelles nées des perspectives relationnelles et sociales du WSo. Une classification est composée de concepts et/ou de termes et les organise par des relations plus ou moins formelles. Certaines Ressources Terminologiques (RTOs) ont tenté de prendre en compte aussi bien les questions de normalisation que celles d'ouverture à l'utilisateur, les cartes de thèmes. Ces derniers permettent de représenter les termes propres au sociolecte bien qu'ils restent limités sur le plan des représentations des utilisateurs, supportant en général assez mal la normalisation. Dans cette partie, nous présentons une démarche possible de construction d'une carte de thèmes orientée utilisateur et partage communautaire.

5.1. Une Ressource Termino-Ontologique adaptée : la carte de thèmes

Le WS considère que l'association de métadonnées standardisées avec des objets en réseau est primordiale. (Weibel et Lagoze, 1997, section 1.1) considèrent aussi que « l'association de métadonnées descriptives standardisées avec des objets en réseau offre un potentiel d'amélioration substantiel des possibilités de découverte de ressources : en permettant des recherches selon des champs (ex : auteur, titre), en permettant l'indexation d'objets non textuels et en permettant l'accès à un contenu de substitution, ce qui est différent de l'accès au contenu de la ressource elle-même ». Cette structuration et standardisation des métadonnées va dans le WS jusqu'à la définition d'ontologies (modèle conceptuel structurant des objets, leurs relations et les inférences possibles), quand d'autres se contentent d'un ensemble de métadonnées (système formel qui fournit l'information d'autorité sur la sémantique et la structure de chaque élément) ou encore de thésaurus ou de classifications informelles.

Au vu de ces nombreux types de classification, nous constatons qu'il faut à la fois bénéficier de la souplesse des ethnoclassifications qui reflètent le sociolecte des utilisateurs et permettent de disposer du vocabulaire du domaine comme un thésaurus et d'une structuration du vocabulaire du domaine selon une optique utilisateur. Dans ces travaux, C. Roussey (Roussey *et al.*, 2002, p. 1) utilise le terme de thésaurus sémantique pour qualifier un genre d'ontologie qui « allie à une modélisation du domaine plusieurs terminologies ». Ainsi les termes sont dissociés des notions qu'ils dénotent, ce qui permet de clarifier les relations entre les termes et les notions et d'identifier les relations terminologiques des relations sémantiques. Le thésaurus sémantique de C. Roussey définit deux niveaux de connaissances (Roussey *et al.*, 2002) :

1. Le niveau conceptuel qui modélise le domaine d'étude formé de types de concepts ou de relations ;
2. Le niveau terminologique qui est composé de l'ensemble des termes, le terme étant défini comme la manifestation linguistique d'un concept repéré dans un texte.

Afin de disposer, au niveau terminologique, du vocabulaire représentatif du sociolecte des utilisateurs, il est possible d'utiliser les termes du texte même. Deux choix s'offrent alors à nous : (1) réutiliser un thésaurus sémantique existant pour indexer nos documents et

fragments de documents ou (2) construire une classification adaptée. L'aspect hybride du thésaurus avec ses deux niveaux est intéressant pour notre propos. Cependant, les thésaurus classique ou sémantique existants proposent une liste de types d'association close du fait qu'ils sont créés par des experts du domaine en lien avec des documentalistes dans un objectif spécifique. Or, dans notre problématique de groupe projet en conception mécanique, aucun spécialiste de la classification n'intervient et la mise à jour du thésaurus est cruciale et doit être dynamique. Il serait donc judicieux pour notre propos de créer une classification adaptée qui suit les principes de multi-granularité du thésaurus sémantique dont la mise à jour se fait par les utilisateurs eux-mêmes. Ainsi, nous rapprocherons notre classification du thésaurus sémantique pour l'approche souple au niveau de la granularité, malgré une approche du niveau conceptuel différente. En effet, le niveau conceptuel n'est pas traité de la même façon que C. Roussey. C. Roussey utilise une ontologie formelle de haut niveau et une étape de normalisation. Dans notre cas, le niveau conceptuel est représenté par deux types d'éléments, les points de vue et les concepts (les points de vue permettant une interprétation des concepts dans un contexte spécifié). Pour définir ce niveau conceptuel, nous proposons une approche fondée sur le terme. Nous proposons à l'utilisateur un ensemble de termes structurés par des relations qui sont validés par l'usage. C'est l'utilisateur qui valide les termes en tant que concepts, y adjoignant des occurrences de ce concept issues de sa pratique et donc du corpus qu'il manipule. Il n'y a pas d'étape classique de normalisation puisque le niveau conceptuel est construit par l'utilisateur même.

Cette classification serait réalisable à l'aide d'outils de TAL pour extraire des termes et des relations du texte (Aussenac-Gilles et Condamines, 2004). La base terminologique offerte par les outils de TAL permet de construire, pour nous, la base d'une carte de thèmes sur mesure.

5.2. Démarche pour une construction sur mesure

Nombre de projets se sont tournés vers la construction de ressources terminologiques (RTO) adaptées à leurs projets pour pallier la généralité ou les défaillances de ressources déjà constituées. Cette orientation est motivée par l'existence d'ontologies génériques d'une part, comme SUMO (Nils et Pease, 2001), ou DOLCE (Gangemi *et al.*, 2003) qui ne sont pas adaptées à une application spécifique à un domaine puisqu'elles contiennent peu de définitions de concepts spécifiques à un domaine. Et d'autre part, par l'existence d'ontologies de domaine souvent onéreuses, peu disponibles et dont les

définitions de concepts reflètent le point de vue d'un expert (même si leur portabilité est augmentée par le respect de standards du W3C - OWL, RDF -). Le point de vue de ceux qui ont construit l'ontologie n'est pas toujours partagé par l'ensemble de la communauté. Ainsi, il n'est pas toujours possible de réutiliser une ontologie pour classifier et indexer les documents d'un projet spécifique. En raison du peu de disponibilité d'ontologies de domaine et de l'utilisation difficile d'ontologies génériques, nombre de projets utilisent les techniques de TAL pour extraire semi-automatiquement des termes (des instances de concepts) (Jacquemin et Bourigault, 2003), pour créer des agrégations de termes (concepts) (Cimiano *et al.*, 2004) ou encore pour extraire des relations entre termes (Buitelaar *et al.*, 2004). L'expert doit alors nommer les agrégats (clusters) en tant que concepts et doit éventuellement définir les relations entre concepts. Les différentes techniques de constitution de RTO sont liées entre autres aux différentes sources terminologiques à disposition (Maedche et Staab, 2001). Les sources utilisées pour la construction des ressources terminologiques peuvent être structurées (base de données), semi-structurées (dictionnaire, corpus annoté) ou brutes (texte). De même, la constitution peut être semi-automatique ou guidée par des principes (manuelle) selon le but dans lequel la RTO s'inscrit, respectivement, son intégration dans un système transparent, ou sa construction elle-même comme objectif.

Notre carte de thèmes (CT) est construite dans un objectif d'intégration à un système d'annotation coopératif. La CT est la classification sur laquelle est fondée l'indexation de l'annotation. Elle a pour rôle de permettre de partager et de regrouper les connaissances communes du groupe. L'utilisateur est au centre du projet puisqu'il s'agit de soutenir son activité innovante de conception mécanique. L'outil doit donc lui permettre de créer une classification spécifique (dont le peuplement continue au fil du projet) à l'aide de termes spécifiques au sociolecte et d'indexer ses propres annotations et documents. Notre démarche combine ainsi les deux types de classification (présentés en chapitre I, section 2, les *ontologies* et en chapitre II section 1, les *ethnoclassifications*) afin d'obtenir une CT adaptée à notre projet. La construction de la carte de thèmes se fait au fur et à mesure du projet, par l'intervention directe des membres du projet (qui ne sont pas des ingénieurs de la connaissance ou ontologistes ni des linguistes). Comme préconisé dans la constitution et le peuplement d'ontologie, les termes et les concepts utilisés doivent être clairs et précis pour retrouver facilement les annotations ou les documents et pour être compris par tous les membres du groupe. Cependant, comme effectué dans les ethnoclassifications, l'utilisateur doit avoir la possibilité d'introduire ses propres termes, sans être contraint à utiliser un

ensemble de descripteurs figés, proposés par une ontologie ou par un thésaurus. La consistance de la classification est validée par son *utilisabilité* dans le projet par les utilisateurs qui l'ont créée. La CT présente à l'utilisateur un réseau de termes structurés qui permet le partage des définitions de termes qui sont utilisés. La structuration et le choix de termes et concepts sont très contraignants pour un utilisateur. Nous avons donc adopté une méthode de construction de carte de thèmes semi-automatique en réutilisant des outils légers et robustes, disponibles et encapsulables sur notre plateforme. On se focalisera ici sur la construction de cette première carte de thèmes du domaine et sur son utilisation pour l'indexation. Nous avons adopté une méthodologie classique de construction de réseau terminologique (Bourigault et *al.*, 2004) qui servira comme ressource pour la construction de la carte de thèmes adaptée à notre application. Cette méthode est proche de celle utilisée aussi par A. Reymonet (Reymonet *et al.*, 2006) :

1. constitution de corpus de référence du domaine de la conception en aéronautique ;
2. extraction de termes à partir du corpus de référence, après avoir identifié les types de constituants syntaxiques qui vont être des candidats termes possibles ;
3. extraction des relations paradigmatiques (verticales) et syntagmatiques (horizontales) entre les termes :
 - a. - identification de règles heuristiques d'identification des relations ;
 - b. - identification automatique des relations selon ces règles ;
 - c. - structuration des termes.
4. choix d'un formalisme de représentation de la carte de thèmes adaptée à la navigation.

Nous allons présenter le corpus de référence utilisé pour représenter les termes du domaine, les choix d'outils pour l'extraction de termes et de relations, les modifications effectuées pour les adapter à notre application orientée utilisateur, ainsi que le formalisme adopté pour représenter la carte de thèmes. Le corpus de travail est présenté dans la partie scénarios d'utilisation pour la simulation tout comme la présentation de résultats.

5.2.1. Présentation du corpus de référence

Un corpus se doit d'être documenté. La documentation comprend aussi bien la démarche de constitution et de traitement du corpus que les variables situationnelles de production des données du corpus. Notre corpus étant constitué d'un recueil de discours

médiatisés par ordinateur, nous décrivons notre corpus suivant les variables situationnelles utilisées par (Herring, 2000). Il s'agit donc d'établir une description des acteurs producteurs de ces discours et du projet, des conditions de production, du type de médiatisation de ces informations, de l'objectif et des thèmes de ces communications, des normes respectées, et enfin de la langue et des styles utilisés. Nous rappellerons aussi l'objectif de création du corpus qui influe évidemment sur la constitution et les choix d'un corpus de référence aussi bien que sur la détermination d'un sous-ensemble de ce corpus en tant que corpus d'analyse des annotations.

Objectif de constitution du corpus

Nous avons constitué ce corpus afin qu'il soit représentatif d'un projet en conception mécanique médiatisée par ordinateur et permette l'analyse de l'annotation dans ce domaine.

Domaine, thèmes et objectifs du corpus

Le corpus est constitué de documents issus du domaine mécanique. Il touche à différents sous-domaines de la conception mécanique (mécanique aéronautique), à la fabrication (spécifications de pièces).

Les documents sont à la fois descriptifs d'une conception finalisée ou de plan et maquettes en cours, qu'argumentatifs, concernant les points à développer pour la conception de moteur et de différentes parties de moteur. Cette multiplicité de thèmes est due à un corpus thématique constitué de différents types de documents de travail. L'objectif des collaborateurs est de concevoir un réducteur pour permettre de monter un moteur automobile diesel sur un avion Delvion.

Types de documents

Documents numériques

Dans le cadre de l'observation d'une conception médiatisée, les documents observés sont des documents numériques possibles à partager dans un groupe projet travaillant en asynchrone.

Notre corpus de référence est donc constitué de plans, maquettes, rapports, et documents Web (sites Web) produits au cours du projet et recueillis par les acteurs du projet à des fins de veille dans le cadre du projet.

Échanges médiatisés

Le second type de documents constituant le corpus de référence sont les méls échangés tout au long du projet (nous ne disposons pas des méls de lancement du projet contrairement aux documents de veille pré-projet).

Le corpus de référence recueilli est donc constitué de 22 documents, 247 méls, soient 47 380 mots. De nombreux méls en citent d'autres, ce qui fait non seulement augmenter la taille du corpus en mots, mais aussi incrémente la fréquence de certains termes. Nous considérons que cette augmentation de la fréquence est légitime puisqu'elle est en rapport avec des termes et des passages qui sont importants pour la communauté.

Auteurs des documents

Les documents sont principalement rédigés par des spécialistes en mécanique, mécanique aéronautique, conception ou fabrication.

Conditions de production

Les documents et les messages n'ont pas les mêmes objectifs, pas plus que les mêmes conditions de production. D'une façon générale, les documents produits sont destinés à un grand public (Sites Web) ou un public spécialisé (certaines pages de Site Web, rapports, documents commerciaux de description de pièces). Les messages sont adressés à des collègues, stagiaires, membres de la même association ou de la même organisation (l'université).

Normes, langue et style respectés

L'ensemble du corpus de référence est donc d'un style assez formel, parfois publicitaire dans le cas des documents issus d'entreprise pour la vente de matériel/matériau, en français courant mais relativement spécialisé.

Le corpus présenté ci-dessus est à la base de la construction de la carte de thèmes de notre projet.

5.2.2. Construction de la carte de thèmes

L'objectif de la carte de thèmes que nous allons construire est de faciliter l'activité d'indexation de l'utilisateur, pour retrouver l'information pertinente à un moment donnée, pour

identifier un processus de négociation et l'explication d'une décision prise au fil du projet. La carte de thèmes est une ressource spécifique, reflétant une terminologie spécifique au groupe de travail et dont le peuplement continue au fil de son projet. L'intérêt des participants au projet est focalisé sur l'interprétation des documents de leur domaine dans un objectif de conception et il n'est pas possible de leur demander de proposer des index de leur documents et annotations à un niveau trop profond. Nous soutenons donc les activités des utilisateurs en proposant une CT, créée d'une manière semi-automatique, à partir d'un corpus. Les documents servant à la construction de la carte de thèmes ne sont donc pas étiquetés ou structurés et les outils doivent traiter des documents bruts. Les outils utilisés pour la construction doivent être robustes, à même de traiter des textes dans lesquels des approximations orthographiques et syntaxiques sont relevées (méls, textes issus de formes de discours dites *nouvelles* (Nouvelles Formes de Communication Électronique, comme les messages des blogs ou des forums). Il est aussi primordial que la carte de thèmes prenne en compte le contexte des termes/concepts qui le constituent, puisque le contexte donne la sémantique du concept dans la conception harrissienne (Harris, 1952). Le contexte est hautement pertinent pour construire les relations syntagmatiques et paradigmatisques des termes/concepts et pour permettre une organisation plus cognitive des éléments de la carte de thèmes par les utilisateurs.

Nous nous focalisons maintenant sur la construction de cette première classification ainsi que sur sa mise à jour et sur son utilisation pour l'indexation. Nous explorerons la définition de règles heuristiques pour construire des relations entre les termes, tout en utilisant des extracteurs de termes adaptés à notre application.

5.2.2.1. Extraction de termes

L'extraction de termes représente une première étape importante dans la construction d'une ressource termino-ontologique à partir des corpus. Les termes sont des mots ou des expressions ayant un sens précis dans un contexte donné et représentent les supports linguistiques des concepts. L'identification de termes spécifiques à un domaine où à un langage de spécialité a fait l'objet de nombreux travaux et des outils performants (Bourigault et Jacquemin, 2000) sont disponibles en plusieurs langues. Les outils d'extraction de termes peuvent être regroupés en trois grandes familles répondant à des approches statistiques, linguistiques ou mixtes. Nous avons étudié plusieurs extracteurs de termes appartenant à ces différentes familles dans un objectif d'intégration à notre plate-forme de travail coopératif.

Les extracteurs de textes purement statistiques (Xtract (Smadja 1993), LIKES (Rousselot *et al.*, 2003)) proposent des candidats termes parmi les séquences des mots les plus fréquentes, suivant l'hypothèse que les termes spécifiques au domaine ont une fréquence élevée dans les textes spécialisés. Ces approches ont l'avantage de ne pas utiliser de ressources linguistiques (type grammaires par exemple), et donc peuvent être appliquées facilement sur un nouveau domaine. Cependant, ces outils proposent beaucoup de bruit dans les résultats. De plus, ils nécessitent une phase d'entraînement sur des corpus de grande taille et nous ne pouvons les traiter à la main, le module étant utilisé pour soutenir directement l'utilisateur. Pour notre application, nous ne disposons pas de corpus très volumineux et donc une approche purement statistique ne semble pas très adaptée.

Les outils utilisant une approche linguistique se fondent sur différents paliers d'analyse linguistique. Par exemple, grâce à une analyse au niveau syntaxique, il est possible d'extraire des groupes nominaux qui fournissent en général un grand nombre de candidats termes pertinents. Certains extracteurs reposent donc sur l'exploitation d'une grammaire (Staab *et al.*, 1999), d'autres sur un lexique (utilisation de dictionnaires), d'autres encore sur des analyses distributionnelles (Harris, 1952, Harris, 1988) à base de patrons lexico-syntaxiques (Hearst, 1992) ou de reconnaissance de co-occurrences (Cimiano *et al.*, 2004). Termino (David et Plante, 1990), Lexter (Bourigault, 1996), FASTR (Jacquemin, 1997), sont définis selon des principes linguistiques, par l'identification de patrons ou l'identification de frontières de divers composants syntaxiques. L'analyse syntaxique partielle fondée sur des automates à états finis (XIP (Ait Mokhtar, 2001)) permet de mieux isoler les candidats termes pertinents.

L'extraction de termes selon des approches linguistiques repose en général sur l'exploitation de corpus étiquetés. L'identification des patrons morpho-syntaxiques est réalisée ensuite par l'extracteur de termes. L'étiqueteur d'E. Brill (1992) et Tree Tagger (Schmid, 1994) sont deux étiqueteurs des plus performants et disponibles. Tree-Tagger fonctionne sur des arbres de décision binaires qui permettent l'estimation d'une étiquette grammaticale. L'étiqueteur de E. Brill apprend des règles d'étiquetage à partir d'un corpus annoté manuellement. À chaque étape d'apprentissage, des règles sont modifiées et le résultat de l'étiquetage avec ces nouvelles règles est comparé avec le corpus représentant l'ensemble des annotations justes. Cependant, les étiquettes sont prédéterminées et ne sont pas toujours adaptées aux textes spécialisés. Comme l'étiqueteur Brill nécessite des corpus étiquetés de grande taille pour améliorer les résultats, TreeTagger nous semble plus adapté à notre application.

Les approches statistique et linguistique offrant toute deux des avantages, des approches mixtes pour l'extraction de termes ont été mises en place. Du point de vue outil, nous pouvons souligner ACABIT (Daille, 1999) ou Exit (Roche *et al.*, 2004), qui s'appuient sur des traitements mixtes pour identifier les candidats termes. Ces approches mixtes permettent de filtrer les candidats termes obtenus grâce à des calculs de fréquence et d'affiner la récupération de termes pertinents grâce à des patrons. Elles permettent ainsi non seulement d'extraire des termes, mais aussi de poser une base à leur structuration.

Nous constatons que les extracteurs de termes qui utilisent des approches mixtes pour identifier des candidats termes sont adaptés à notre contexte plus que les outils statistiques. De même, ils sont plus simples à mettre en place que des extracteurs fondés sur des grammaires ou des dictionnaires (lourds à définir et à utiliser dans un système). Nous privilégions donc une approche mixte permise par l'utilisation de l'outil Syntex.

L'analyseur syntaxique Syntex pour l'extraction de termes

L'analyseur syntaxique Syntex (Bourigault *et al.*, 2001) est le descendant de Lexter, un analyseur syntaxique robuste dédié au repérage des syntagmes nominaux. Lexter ne prenant pas en charge la couverture des syntagmes verbaux, Syntex a été développé dans le but de pallier ses faiblesses. Syntex n'est pas dédié à un type de corpus ni à un simple objectif d'analyse syntaxique. Il est orienté construction de RTO et est spécialisé pour des systèmes de traitement de l'information. Syntex est développé pour traiter des corpus de phrases réelles et de taille importante. Il est donc efficace (temps de traitement) et robuste (tolérance aux malformations syntaxiques, aux mots ou structures inconnues, possibilité de rendre des analyses partielles et incomplètes).

Pour analyser, Syntex s'appuie sur des corpus préalablement étiquetés grâce à l'étiqueteur Tree Tagger. Il effectue ensuite une analyse syntaxique en dépendance avec un enchaînement de modules de reconnaissance de relations syntaxiques. Les modules sont constitués d'un ensemble d'heuristiques de parcours de la chaîne étiquetée et partiellement analysée, qui partent d'un régi (resp. recteur) potentiel pour aboutir à son recteur (resp. régi). Ils sont développés à *la main* selon une méthode qui met en oeuvre le recours à la connaissance grammaticale et à des tests nombreux et variés sur des corpus diversifiés. Syntex exploite de façon combinée des procédures d'apprentissage et des ressources lexico-syntaxiques de sous-catégorisation. Les ressources linguistiques nécessaires sont minimales (liste de verbes qui font le passé composé avec être, liste de noms propres).

L'analyse des corpus textuels en dépendances syntaxiques permet de définir le contexte des candidats termes dont nous avons besoin à la fois pour identifier les candidats termes mais aussi des relations entre les termes. Une fois le corpus traité par Syntex, il est possible d'utiliser ses sorties (voir Fig. 35) et de les coupler à un algorithme extrayant des candidats termes et des relations pertinents au sein du corpus.

```

<syntex>
  <SEQ id="_">
    <TXT>
      Roulements à billes à contact oblique dimensions principales selon DIN , à deux
      rangées ,
      joint à lèvres des deux côtés .
    </TXT>
    <tokens>
      <t i="1" l="roulement" f="Roulements" c="Nom?P" p="N"/>
      <t i="2" l="à" f="à" c="Prep" p="O"/>
      <t i="3" l="bille" f="billes" c="Nom?P" p="N"/>
    </tokens>
    <dependances>
      <g r="PREP" s="1" c="2"/>
      <g r="NOMPREP" s="2" c="3"/>
      <d r="PREP" s="2" c="1"/>
      <g r="PREP" s="3" c="4"/>
    </dependances>
  </SEQ>
</syntex>

```

Fig. 35. EXEMPLE DE SORTIE DE SYNTAXE SUR NOTRE CORPUS

Le langage utilisé dans le corpus d'annotations est très proche du langage oral, et contient beaucoup de fautes de frappe ou d'orthographe. C'est pour cette raison que Syntex est l'outil le plus adapté à notre application. En effet, il gère les erreurs de syntaxe (analyse partielle de la phrase).

Les candidats termes proposés par Syntex sont choisis parmi les syntagmes nominaux et les syntagmes verbaux, alors que d'autres outils se limitent à proposer des candidats parmi les syntagmes nominaux. L'étiquetage nécessaire pour le processus d'analyse est réalisé avec TreeTagger, qui fournit des résultats satisfaisants même sur des corpus techniques ou spécialisés.

Une troisième raison pour le choix de Syntex est donnée par les relations de dépendance identifiées au cours de l'analyse. Notre objectif est de créer du matériel terminologique pour constituer un TS, en identifiant les relations possibles entre les termes. Syntex nous permet d'utiliser les éléments syntaxiques dans leur contexte et fournit les relations de dépendances entre les divers composants syntaxiques. Nous allons exploiter ces relations pour établir des relations entre les occurrences.

Une fois le corpus traité par Syntex, il est possible d'utiliser ses sorties et de les coupler à un module d'extraction de candidats termes et de relations pertinents au sein du corpus.

Le plug-out de Syntex pour le choix de termes et la structuration en relation

Pour extraire la liste des candidats termes qui servent à constituer la CT, nous avons développé un module qui recherche des constituants syntaxiques susceptibles d'être intéressants. Nous sélectionnons un ensemble de candidats termes du domaine à l'aide de patrons morpho-syntaxiques, parmi les groupes nominaux complexes (groupes nominaux modifiés par plusieurs groupes prépositionnels) et parmi les groupes verbaux simples (verbe suivi d'un groupe nominal simple ou complexe). Parmi les constituants syntaxiques identifiés dans les textes, plusieurs structures sont possibles, mais nous proposons toujours celles qui ont une couverture maximale (exemple : un groupe nominal suivi de plusieurs modificateurs groupes prépositionnels, des noms modifiés par plusieurs groupes prépositionnels reliés par une conjonction), dans un souci de proposer à l'utilisateur une liste restreinte mais pertinente des index. Les patrons servant à l'extraction de termes sont visibles au Tableau 9. Ils permettent d'extraire des syntagmes du type :

- l'avion avec son montage d'origine
- échange standard
- mener l'étude
- proposition
- moulin d'occase
- la partie technique

Les syntagmes extraits sont conservés selon leur poids calculés sur leur fréquence d'apparition.

Fréquence

Utilisant les techniques de RI, nous plaçons un seuil de recueil des termes. Ce seuil modifiable est placé par l'utilisateur et donné avant la construction initiale de la carte de thèmes. Ce seuil possède un minimum et un maximum entre 0 et 1 puisque la fréquence est calculée selon une formule proche du $tf.idf$ sur l'ensemble des fichiers en mémoire au moment de la construction de la carte de thèmes. Le $tf.idf$ permet de pondérer la fréquence d'apparition d'un terme dans un document en fonction de ses apparitions dans un ensemble de document. Cette formule très utilisée et pertinente pour l'indexation d'un document est très fine. Nous souhaitons donc considérer la fréquence d'un terme en fonction d'un ensemble d'apparitions. Cependant, nous avons besoin d'une pondération de l'occurrence d'un terme dans un ensemble de document. Nous avons donc transformé la formule de base pour obtenir une pondération représentative pour l'ensemble du corpus (Fig. 36).

$$Tf\text{idf}_{\text{remanié}} = \left(\frac{\sum f}{Max(f)} \right) \times \log\left(\frac{N}{n}\right)$$

Fig. 36. FORMULE DE FRÉQUENCE

où :

$\sum f$ est la somme des fréquences du terme x dans le corpus.

$Max(f)$ est la fréquence du terme qui a la fréquence maximale dans le corpus.

N est le nombre de documents dans le corpus.

n est le nombre de documents qui contient le terme x .

Notre module d'identification des relations est fondé non seulement sur la récupération de termes selon leur fréquence, mais aussi sur la définition de patrons morpho-syntaxiques qui s'appliquent sur les résultats fournis par Syntex afin d'extraire des relations entre les candidats-termes.

5.2.3. Construction de relations

L'identification des relations entre des termes permet de relier des classes de termes au niveau sémantique. Pour cette tâche, il existe beaucoup moins d'outils que pour l'extraction de termes. Certains de ces outils construisent des classes sémantiques sur des bases statistiques, et d'autres sur des principes linguistiques. Nous nous pencherons plus particulièrement sur ces seconds. Zellig (Habert et Nazarenko, 1996, Habert et Fabre, 1999)

se fonde sur l'hypothèse harrissienne (Harris, 1988) selon laquelle des régularités syntaxiques dans les contextes des termes permettent de former des classes. Les classes formées par Zellig semblent cependant rester très larges et les regroupements proposés sont plutôt utilisés comme amorce à la constitution de classes sémantiques pertinentes (Bourigault et Jacquemin, 2000). (Capponi et Toussaint, 2000) développent pour leur projet un ensemble de règles de généralisation de structures prédicat-argument définies par exemple sur le plus petit subsumant commun à une classe de structures prédictives. Cependant, leur méthode est fondée sur une forte structuration de leurs données à base de logique de description qui n'est pas adaptée dans notre contexte, puisque la CT est mise à jour par les utilisateurs mêmes. L'outil Caméléon (Séguéla et Aussenac, 1999) aide lui aussi au repérage de relations sémantiques par projection de patrons lexico-syntaxiques sur des textes (voir Fig. 37). Les marqueurs linguistiques de relations, comme les verbes, sont considérés spécifiques au corpus.

- | |
|--|
| <p>(1) Det N1 et Det N2 (qui, adj, p.passé, p.présent)
 (2) Tous les N2 sauf det N1
 (3) Det N1 comme det N2</p> |
|--|

Fig. 37. PATRONS DE RELATIONS CONCEPTUELLES DE (SÉGUÉLA ET AUSSÉNAC, 1999)

Caméléon propose à la fois d'évaluer des connaissances linguistiques générales associées aux relations structurantes d'hyponymie et de méronymie pour les adapter au corpus, et d'assister l'acquisition de marqueurs spécifiques par recherche de récurrences à partir de couples de termes entre lesquels la relation a été identifiée. L'approche de Caméléon est portée par l'utilisateur qui recherche et structure des patrons pour les relations (Fig. 37). Upery (Bourigault, 2002), développé pour les résultats de Syntex est plus proche de nos besoins. Upery permet de rapprocher des couples d'unités retrouvées dans des contextes syntaxiques identiques. Les termes et les contextes syntaxiques peuvent être simples ou complexes. Le module rapproche ensuite les termes, ainsi que les contextes syntaxiques, sur la base de mesures de proximité distributionnelle. Les résultats sont utilisés comme aide à la construction d'ontologies spécialisées.

L'identification semi-automatique des relations entre les candidats termes est une tâche difficile, mais nécessaire au soutien des utilisateurs pour permettre une navigation facile dans la CT, lors de l'indexation de leurs documents et annotations. En plus d'identifier des candidats termes du domaine, nous avons développé un module qui propose une

identification des relations hiérarchiques et horizontales entre les termes du domaine en s'appuyant sur les dépendances syntaxiques établies entre les candidats termes.

Nous proposons de construire notre carte de thèmes par un traitement de notre corpus en conception mécanique par l'exploitation des résultats fournis par Syntex afin d'extraire des candidats-termes structurés syntaxiquement entre eux puis de construire les relations entre nos candidats-termes par des méthodes distributionnelles comme mise en place par Upery et des techniques de classification par patrons (Hearst, 1992, Fig. 38).

- | |
|---|
| <p>(1) NP₀ .such as {NP₁, NP₂,... (and or)} NP_n A tel que X, Y et/ou Z
 (2) such NP as {NP ,}* {(or and)} NP de tels A tels que X, ou Y, et Z.
 (3) NP { , NP } * { , } or other NP A, B ou d'autres X
 (4) NP { , NP } * { , } and other NP A, B et d'autres X
 (5) NP, { , } including {NP ,}* {or and} NP A comprenant, X et ou Y
 (6) NP { , } especially {NP ,}* {or and} NP A et plus spécifiquement X, et ou Y</p> |
|---|

Fig. 38. PATRONS LEXICO-SYNTAXIQUES POUR L'HYPONYMIE DE (HEARST, 1992)

Contrairement à Caméléon, dans notre approche le verbe n'est pas le support de la structuration, mais représente la relation sémantique entre les candidats-termes, identifiée par des règles heuristiques présentées ci-après. Les contextes des candidats termes sont utilisés à la fois pour trouver des candidats termes synonymes pour la construction de la classification et également pour trouver des mots-clés indexant possibles dans les annotations/documents.

Les patrons d'extraction de relations

Afin de structurer les termes issus du corpus par des relations, il faut en premier lieu identifier les relations existantes entre les termes. Pour la reconnaissance de relations hyperonymiques entre les termes, nous utilisons des patrons lexico-syntaxiques de (Hearst, 1992) (Fig. 38), de (Séguéla et Aussenac, 1999) (Fig. 37) ainsi que ceux de (Fuchs, 1982) (Fig. 39). L'ensemble des patrons et des exemples des termes pouvant être extraits sont présentés en Tableau 9. Les patrons sont définis suite à une étude de la littérature et de notre corpus. Bien que provenant d'outils de communication médiatisée, notre corpus ne comporte que peu de spécificités des NFCE. Il n'est remarquable que par l'utilisation d'une terminologie métier liée à la conception mécanique qui influe peu sur le niveau syntaxique. Les patrons présentés sont donc réutilisables pour l'analyse d'autres corpus analogues.

Tab. 9. RELATIONS UTILISÉES POUR L'EXTRACTION DE TERMES

Type de relation	Catégorie de relation	Nom de la relation dans la CT	Patron d'extraction de la relation	Exemple de termes extraits
Relations paradigmatisées	Hyponymie	est-un	<p>D N tel que {D N1, * {et ou}} D N</p> <p>D N tels que {D N1,*} {{ou et} D N2}</p> <p>D N {, D N} * {,} ou d'autres N</p> <p>D N {, D N}* {,} et d'autres N</p> <p>D N, {,} comprenant {D N,}* {ou et} D N</p> <p>D N {,} et plus spécifiquement {D N,}* {ou et} D N</p> <p>Tous les N sauf D N</p>	<p>Une pièce telle que un flasque, un réducteur (,) ou des vis</p> <p>Des pièces telles que des flasques, des réducteurs (,) et des vis</p> <p>Un réducteur, un flasque (,) ou d'autres pièces</p> <p>Un réducteur, un flasque (,) ou d'autres pièces</p> <p>Les pièces comprenant les flasques et les réducteurs</p> <p>Une pièce et plus spécifiquement un réducteur ou un flasque</p> <p>Toutes les pièces sauf le(s) réducteur(s)</p>

	Méronymie Holonymie	partie de	D N appartient {à la au aux} N D N est une partie {de du de la} N D N est composé(e) de N D N contient {un des} N	La vis appartient à la visserie La courroie est une partie du réducteur Le réducteur est composé d'engrenages Un réducteur contient des vis
Relations syntagmatiques	Relations script (ou scénario), analogiques lexicales, liens personnels, nominales,	Terme1Terme2	N V N N A D N comme D N D N et D N qui	arbre entraîne engrenages flasque droit Les réducteurs comme les flasques Un réducteur et une courroie qui
Indéterminable syntaxiquement		Préposition	N préposition N	pièce de fraisage fraise de Felipe pièce à fraiser

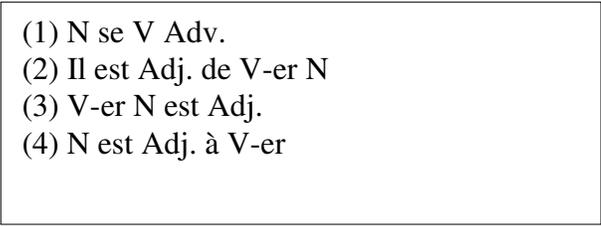
- 
- (1) N se V Adv.
 - (2) Il est Adj. de V-er N
 - (3) V-er N est Adj.
 - (4) N est Adj. à V-er

Fig. 39. PATRONS DE (FUCHS, 1982)

Au delà des structures syntaxiques complexes identifiées, les candidats termes peuvent être groupés en plusieurs catégories (défauts de fabrication, paramètres de fonctionnement des moteurs ou des pièces/composants, emplacement des pièces, éléments de solutions, etc.) qui n'apparaissent pas toujours dans le corpus. Les diverses familles de termes doivent alors être définies manuellement comme des points de vue du domaine par exemple et les utilisateurs classeront eux-mêmes les candidats termes sous le point de vue concerné (composants/pièces, défauts, solutions, paramètres).

5.2.4. Représentation de la carte de thèmes

Comme défini précédemment, la classification construite doit permettre d'indexer les documents et les fragments de documents utilisés ou construits au cours du projet. Il s'agit donc de structurer les termes extraits et structurés en concepts selon différents points de vue selon un formalisme adapté. La représentation doit permettre à des utilisateurs de visualiser leurs documents et les liens entre leurs documents.

5.2.4.1. Langages de représentation d'ontologies

Dans son effort de structuration du Web pour l'interopérabilité des systèmes, le WS a mis en place des recommandations de langages d'assertion qui permettent de mettre en relation des données formelles et donc de représenter des ontologies (Baget *et al.*, 2004), sous un format standardisé. Les connaissances contenues dans les ontologies sont utilisées pour enrichir le contenu des documents et pour les indexer par rapport à leur contenu. Ainsi, les langages de représentation d'ontologies permettent aux agents logiciels ou humains de communiquer et d'accéder au contenu des documents. Les recherches automatiques d'information menées par les agents logiciels sont améliorées par l'utilisation de ces connaissances ajoutées au document, par le biais des inférences logiques sur des connaissances existantes. Les agents humains sont aidés dans leur activité de navigation qui sera guidée par le contenu du document.

Les langages proposés par le WS proposent divers niveaux de formalisation, certains permettent juste de décrire des concepts et leurs propriétés, d'autres permettent des inférences logiques complexes. Parmi les langages proposés pour représenter les ontologies et qui permettent de déduire des connaissances à l'aide des calculs, on peut mentionner RDF, qui est un standard du W3C (Lassila et Swick, 1999) RDFS (Brickley et Guha, 2004), ou OWL (McGuinness et Van Harmelen, 2004). Dans la catégorie de langages de représentation adaptés à la navigation on peut mentionner une norme ISO, Topic Maps (Biezunski *et al.*, 1999) dont la syntaxe XTM (Pepper et Moore, 2001), a été mise en place en 2001.

Tous ces langages permettent de définir des concepts, avec leurs propriétés et leurs relations avec les autres concepts. Les relations peuvent être hiérarchiques (est-un) ou horizontales (relation script, co-occurrence). Pour pouvoir faire des inférences sur ces connaissances (utiliser des propriétés héritées d'un concept plus générique ou déduire des relations qui ne sont pas définies explicitement entre deux concepts), il faut décrire les propriétés des concepts et les relations d'une manière très précise et détaillée. L'intervention d'un expert est nécessaire pour vérifier les incohérences. Or, dans notre contexte, la mise à jour de la CT se fait par des utilisateurs qui ne sont pas ontologistes. De plus, l'objectif de notre outil est de faciliter la communication entre les agents humains, donc nous pouvons nous contenter d'une description des connaissances assez souple et peu détaillée. Parmi les formalismes proposés pour faciliter la navigation, le formalisme des Topic Maps nous semble adapté à notre situation et il sera présenté en détail.

Topic Navigation Maps

Le formalisme Topic Maps (TM) (Norme ISO, Biezunski *et al.* 1999) naît dans les années 90 de travaux sur la gestion des index en documentation et, en 2001, est adopté comme la norme ISO 13250. TM est créé pour permettre d'indexer un ensemble de ressources informatives afin de permettre leur récupération. Une carte de connaissances (TM) met en valeur les relations entre des sujets (topics) dans un certain point de vue (scope). Chaque sujet est aussi relié à ses occurrences (dans les documents), à son nom, à son type... Il y a donc plusieurs types d'assertion en TM, liées aux caractéristiques du sujet. L'assertion minimale en TM est la déclaration du nom d'un sujet (Voir Fig. 40), mais il est aussi possible d'assigner des occurrences trouvées dans un document à un sujet, ou encore des associations avec d'autres sujets (Voir la même figure). Les associations sont typées

librement et chaque association possède un rôle qui permet de différencier par exemple le rôle de *Mémé Talon* par rapport à celui de *Dieudonné Corydon* dit *Alambic* dans leur filiation avec *Achille Talon*. D'une façon plus large, c'est ce rôle qui permettrait de différencier la relation entre deux pages Web de la relation entre une page Web et un lecteur humain.

Déclaration d'un topic



Déclaration d'un Topic avec occurrence



Une association ternaire en TM

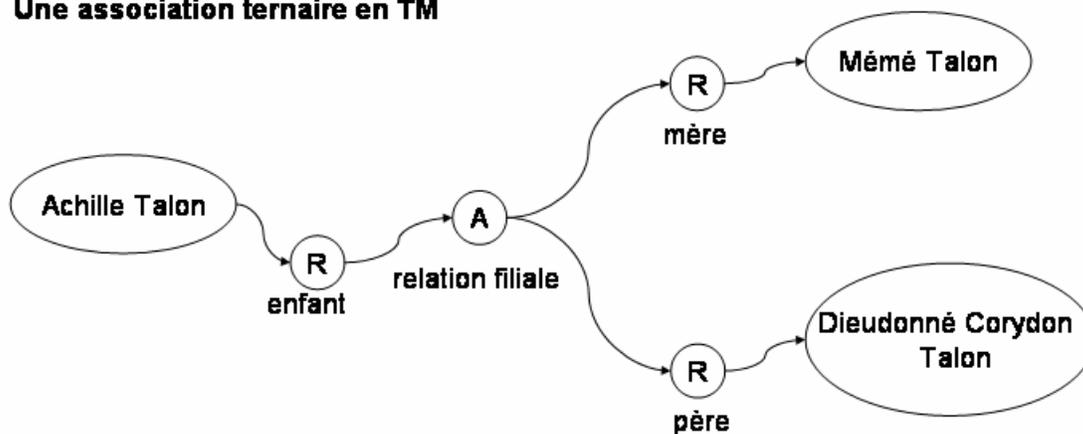


Fig. 40. DÉCLARATIONS EN TM

Nous présentons ici les avantages liés à l'utilisation du formalisme TM pour représenter notre carte de thèmes. Le formalisme TM propose de déclarer un sujet et de l'associer à d'autres sujets ou encore d'illustrer un topic par son utilisation concrète dans un document. De plus, TM simplifie la déclaration des associations puisqu'il permet de faire des associations multiples (où les rôles sont supérieurs à deux) sans passer par des déclarations intermédiaires comme dans les langages dérivés du RDF, ce qui permet plus de souplesse, même si la structuration paraît moins calculable. Les membres du groupe projet seront donc en mesure de mettre à jour la carte sans difficultés.

De plus, la norme TM a été définie dans un objectif de gestion de descripteurs même si cet objectif a été élargi à l'indexation. Un accent a donc été porté sur la signification de l'assignation d'une ressource type URI. En effet, L. Garshol (2003) souligne que cette URI peut aussi bien indiquer la ressource dans laquelle on trouve l'objet, que la ressource indexée par l'objet. Est-ce que « <http://www.dargaud.com/front/albums/series/serie.aspx?id=1608> » indique la ressource que nous obtenons en suivant cette URI, soit une page chez Dargaud, ou l'objet décrit par cette ressource soient les différents albums d'Achille Talon ? Les URIs étant utilisées dans les deux cas, TM différencie les *subject addresses* qui identifient la ressource, des *subject identifiers* qui identifient tout ce qui est décrit par la ressource. De plus, les ressources associées à chaque sujet contiennent des occurrences de ces sujets ce qui permet aux utilisateurs d'avoir une interprétation partagée associée au sujet.

Enfin, TM fournit une possibilité de qualifier les assertions selon un contexte de validité ; ce sont les *scopes* (points de vue). Toute assertion (nom, occurrence, association) possède un point de vue (par défaut un *unconstrained scope*). Par exemple, l'association ternaire présentée Fig. 41, n'est vraie que dans un scope *fiction* alors qu'une association identique avec Greg peut être interprétée comme *auteur* grâce au scope *réalité*. Ainsi, plusieurs points de vue peuvent co-exister en TM.

Selon ces trois points, nous pouvons considérer que les TM sont pertinentes pour notre projet. Non seulement par la souplesse de déclaration qu'elles permettent, mais aussi par leur optimisation pour la navigation, la *carte de connaissances* ou la TM permettant à l'utilisateur une navigation simple et complète dans les concepts couverts par les ressources indexées. De plus, TM étant une norme ISO, nous pouvons garder en vue l'objectif d'interopérabilité recommandée par le WS.

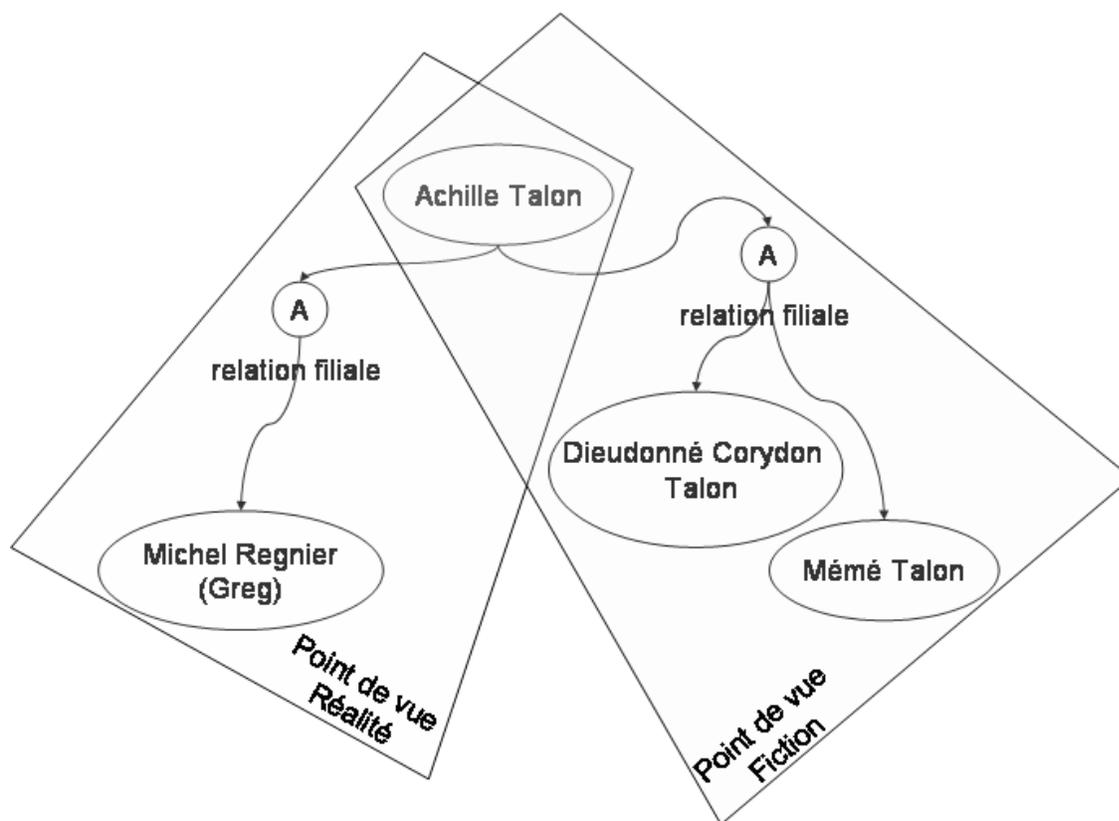


Fig. 41. ASSOCIATIONS SELON DES POINTS DE VUE EN TM

Notre carte de thèmes est utilisée dans le but d'aider l'utilisateur à naviguer dans la base d'annotations et non de construire automatiquement une représentation sémantique du contenu d'un document. Le formalisme TM définit un réseau de sujets couvrant des connaissances de domaine. Pour notre application, les sujets vont être des termes fréquents identifiés dans le corpus de référence. Les sujets (topics) sont définis par de simples URLs, qui gardent les occurrences associées, pour clarifier l'emploi d'un terme. Ces occurrences sont extraites par les outils présentés ci-dessus à partir du corpus de référence et représentent les contextes des sujets trouvés dans les documents.

Les sujets sont reliés par des associations entre les termes qui correspondent aux relations identifiés dans le corpus (relations d'hyponymie, relations horizontales *partie de, utilisé par*). Puisque les TMs sont conçues pour soutenir l'utilisateur dans sa navigation Web, nous avons adopté ce formalisme pour représenter notre CT.

La CT représentée en TM est construite semi-automatiquement du corpus étiqueté par Syntex. Le tableau suivant (Tab. 10) répertorie les patrons d'extraction de termes et leur structuration en relations et en un certain nombre de points de vue de base que l'utilisateur peut étendre.

Tab. 10. PATRONS D'EXTRACTION ET STRUCTURATION EN TM

Point de Vue	Topic	Patron d'extraction des termes
Planification	Temps	NomXXDate ="lundi, mardi, mercredi, jeudi, vendredi, samedi, dimanche" l="janvier, février, mars, avril, mai, juin, juillet, août, septembre, octobre, novembre, décembre" Prep NomXXHeure l="délai, échéance, disponibilité, réunion, créneau, vacances, week-end, demi-journée" Prep= l="dès"
	Lieu	NomXXAdr NomXXMail
	Financière	lemmes="budget, commande, remboursement, facture, subvention, prix, coût, compte, devis, facturation, achat, versement, tarif, RIB, financement, refacturation, paiement" lemmes="commander, facturer, financer " NomXXMon
Organisation	Collaborateur	NomXXTitre
	Organisme	NomPrXXPrenom NomPrXXInc ???NomPr??? ???NomPrXXInc???
	Fonction	Lemmes=chef de projet, ingénieur, étudiant, concepteur

Point de Vue	Topic	Patron d'extraction des termes
Domaine		NON ON NON DNA DNNP DNON DAN DNAONNP ANONA NN NNP NP ONONNP Cc NP NP Cc NP N NA NP N DNON Cc NONODN DNODNON DN DNA NP DN

Point de Vue	Topic	Patron d'extraction des termes
Argumentation	accord démonstration désaccord explication justification confirmation décision mise au point modification proposition récapitulatif suggestion vérification définition commentaire	avis, besoin, choix, critique, difficulté, discussion, doute, ennui, erreur, manque, problème, question, réaction, recherche d'accord, remarque, renseignement, réponse, solution

5.2.5. Création d'une carte

La création de la première classification est automatique, lancée par les utilisateurs, et se fonde sur les documents déposés par les utilisateurs dans leur bibliothèque commune. La construction de la CT est décrite auparavant. L'utilisateur définira le seuil de fréquence des termes à inclure. La carte de thème créée automatiquement peut être corrigée par les utilisateurs dès sa création puis mise à jour au fil de l'eau grâce à un éditeur de Topic Maps.

5.2.5.1. Mise à jour de la carte de thèmes

La CT ne peut répondre à une structuration a priori. Il est nécessaire de la faire évoluer au fur et à mesure du projet. Afin de permettre sa mise à jour, deux techniques sont mises en parallèle. L'une est semi-automatique, l'autre manuelle.

Semi-automatique

La mise à jour de la carte de thèmes a lieu à chaque ajout de document dans la bibliothèque et à chaque ajout d'annotation. En effet, chaque fichier ajouté est traité par le module TAL provoquant la mise en route de l'étiquetage, de Syntex et du programme de mise en forme de la TM. Le dernier programme vérifie la présence des nouveaux thèmes récupérés avant de les intégrer à la TM s'ils en sont absents. Chaque nouveau document est donc pris en compte et incrémente le poids de termes qui est trop faible au départ, mais qui peut devenir suffisant pour permettre leur intégration dans la TM. Des termes apparaissant au fur et à mesure du projet comme des termes devenant cruciaux dans le projet peuvent alors être ajoutés. L'ajout de terme est très simple et ne se fonde que sur la lemmatisation des termes. Aucun calcul de distance sémantique comme de synonymie n'est mis en place mais fait partie des extensions possibles. Il n'y a pas non plus ici d'étape de normalisation, la normalisation se faisant au fil de l'eau selon les choix et les usages l'utilisateur.

Manuelle

Pour la mise à jour manuelle de la CT, nous avons développé un éditeur de TM (Fig. 48) encapsulé dans l'outil. Cet éditeur permet d'ajouter et de supprimer des points de vue, des thèmes, des occurrences et des relations de la TM aux points de vue de base et aux thèmes qui peuplent automatiquement la carte. Ainsi, l'utilisateur détermine ce qui est nécessaire à l'indexation de ses documents et fragments de documents. Il nettoie manuellement les termes proposés automatiquement. La CT, construite sur la base de documents, étant utilisée par un utilisateur humain et non par un système peut être semi-structurée. Cela permet une souplesse d'indexation pour l'utilisateur. La classification sociolectale semi-formelle de la CT facilite la tâche d'indexation de l'utilisateur qui connaît le schème de classification en action. L'outil permet ainsi à l'utilisateur de mettre à jour ce schème de classification au cours de son activité.



Fig. 42. INTERFACE D'ÉDITEUR DE LA CARTE DE THÈMES EN TM

5.2.6. Indexation selon plusieurs index et plusieurs granularités

La structuration en TM de la carte de thèmes permet d'indexer les annotations selon différentes granularités (points de vue, thèmes, sous-thèmes, ...). L'indexation est proposée à l'utilisateur par un algorithme de mise en correspondance du corps de l'annotation avec les éléments de la carte de thèmes (Fig. 43). L'indexation a lieu à différents niveaux. Tout d'abord au niveau du terme, où chaque terme du corps de l'annotation est mis en correspondance avec le niveau thème de la TM. Si la mise en correspondance n'aboutit pas, une seconde indexation au niveau de l'occurrence est lancée en parallèle à une mise en correspondance au point de vue. Lorsqu'un descripteur correspond à un terme de l'annotation, il est proposé à l'utilisateur qui le valide ou non. Si aucune mise en correspondance n'aboutit, il est proposé directement à l'utilisateur d'ajouter son descripteur à la CT. Le descripteur validé est ajouté comme métadonnée de l'annotation. Les descripteurs sont de plusieurs types, appartenant à divers points de vue. Les descripteurs du point de vue argumentatifs sont proposés selon une liste disponible dans un menu déroulant (Fig. 44). Ils serviront à récupérer l'annotation aussi bien qu'à tracer la logique de communication.

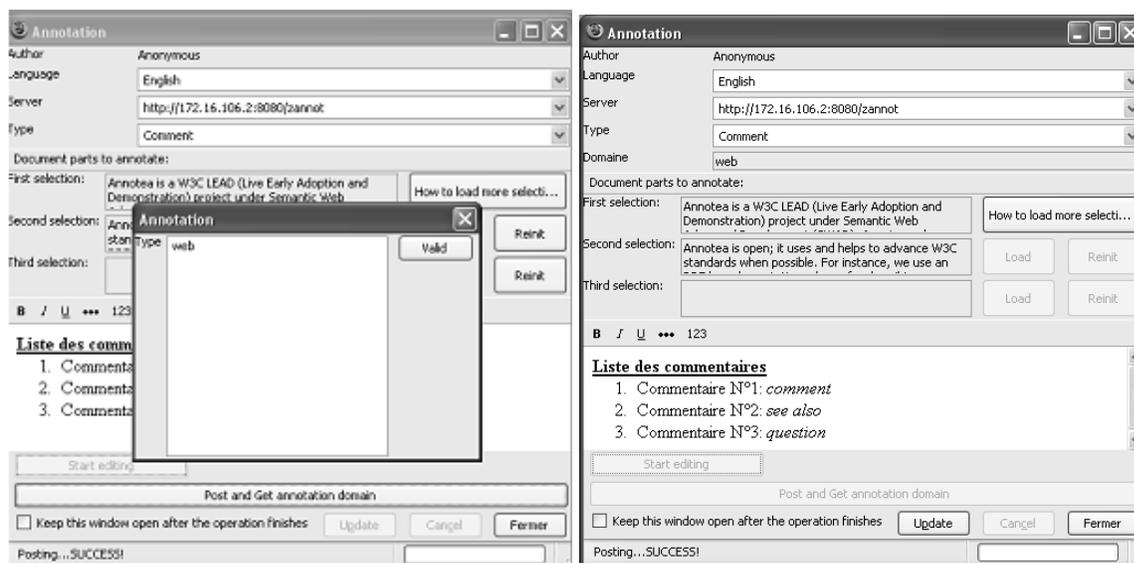


Fig. 43. PROPOSITION ET ENREGISTREMENT D'INDEX

Type

commentaire
avis
besoin
choix
critique
difficulté
discussion
doute
ennui
erreur
manque
problème
question
réaction
recherche d'accord
remarque
renseignement
réponse
solution

Argumentation

accord
démonstration
désaccord
explication
justification
confirmation
décision
mise au point
modification
proposition
récapitulatif
suggestion
vérification
définition

Fig. 44. LISTE DÉROULANTE DES DESCRIPTEURS ARGUMENTATIFS

6. Conclusion

L'outil AnT&CoW est développé pour soutenir les activités documentaires collectives d'utilisateurs au cours d'un projet. Les activités à soutenir dans le projet de conception mécanique étudié sont principalement liées au partage de documents et d'interprétations sur les documents et à la co-édition de documents. L'ensemble du groupe engagé dans le projet commun doit avoir la possibilité de déposer des documents, de les annoter et d'organiser les annotations afin de tracer la logique de communication du projet. La logique de communication peut être représentée sous la forme d'un document édité grâce au recueil d'annotations de différents auteurs. Afin d'organiser les annotations en un fil représentant la logique de communication, il faut auparavant les indexer finement selon les différents points de vue des annotateurs. L'indexation étant une activité fastidieuse, l'outil soutient les utilisateurs dans la construction d'une classification partagée aussi bien que dans l'association de descripteurs à leurs commentaires via un module d'indexation fonctionnant sur la base d'outils de TAL.

Une maquette du collecticiel étant mise en place, il s'agit maintenant de l'évaluer. Le chapitre suivant est consacré à une simulation d'utilisation de l'outil sur la base de scénarios et de leur contextualisation grâce au traitement d'un corpus d'annotations issu du corpus de référence recueilli.

Chapitre VI

Scénarios d'utilisation d'AnT&CoW

Ne pense pas, regarde plutôt!
Ludwig Wittgenstein

L'objet de ce chapitre est de décrire et de présenter les résultats d'une simulation du collecticiel AnT&CoW. Après avoir rappelé le dispositif de simulation fondé sur des scénarios d'utilisation suivant la démarche exposée en introduction, nous présentons le corpus de travail nécessaire à la contextualisation des scénarios. Nous exposons ensuite les scénarios et leur jeu au sein de l'outil AnT&CoW. Nous terminons par la présentation des résultats de la simulation. L'utilité et l'utilisabilité du collecticiel dans le contexte d'un projet en conception mécanique sont examinées sur la base d'un entretien avec les experts du projet.

1. Dispositif de simulation

(Ramage, 1999) définit les systèmes coopératifs comme une combinaison de technologie, de personnes et d'organisations qui facilitent la communication et la coordination, nécessaires aux membres d'un groupe pour qu'ils puissent travailler ensemble en partageant le même but et l'atteindre avec un gain pour chacun des membres. AnT&CoW a en effet pour objectif de soutenir une communauté de projet en favorisant les échanges au cours des tâches de conception coopérative médiatisée. La conception s'appuie sur un échange de documents et des échanges autour de ces documents. Une fois la phase de production effectuée (implémentation de fonctionnalités documentaires dans une maquette pour le soutien de l'activité), une phase de validation des fonctionnalités par simulation est nécessaire.

Cette simulation est menée sur un prototype de l'outil et a pour but d'évaluer l'outil sur des critères de validité, d'utilité et d'utilisabilité. Le test de validité de l'outil se fait par l'évaluation du fonctionnement de chaque fonctionnalité. Ce dernier permet de constater que pour une activité, les fonctionnalités nécessaires sont présentes.

Les critères d'utilité sont établis principalement pour valider le système créé par rapport aux objectifs initiaux de l'annotation. Ils valident les choix de conception exposés précédemment selon un scénario qui reprend les actions des utilisateurs au cours de leur activité. L'utilité considère l'adéquation du système avec les fonctionnalités spécifiées.

Les critères d'utilisabilité sont définis dans un objectif de vérification de l'usage du système développé. Ils permettent de vérifier qu'un utilisateur fait avec cette fonctionnalité ce qu'il veut faire d'une façon adéquate. L'utilisabilité considère l'adéquation du système avec les activités et les actions des utilisateurs.

Dans le cas de notre outil, l'indexation est liée aux sorties du module ZTAL. La validation de la fonctionnalité revient à vérifier si les modules ZTAL génèrent bien une carte de thèmes en TM. L'utilité revient à vérifier que la CT générée et les fonctionnalités développées sur sa base sont utiles à l'activité de l'utilisateur. L'utilisabilité revient à vérifier que l'utilisateur peut aisément utiliser cette fonctionnalité et l'utilise concrètement dans son activité avec un bénéfice pour celle-ci. Dans le cadre de l'évaluation de l'utilité, il serait utile de faire appel aux experts dont les activités ont été observées sur le terrain afin de confirmer l'adéquation de la classification constituée semi-automatiquement. En effet, l'évaluateur est expert de l'outil, non du domaine de la conception mécanique ; il peut valider l'outil, mais non l'utilité et encore moins l'utilisabilité. C'est pourquoi nous présentons ici une simulation sur la base de scénarios contextualisés par le corpus pour valider les fonctionnalités développées.

Ainsi, l'évaluation de l'outil se fera en plusieurs temps. Tout d'abord au niveau de sa validité par (1) le jeu des scénarios par l'acteur-évaluateur et qui a pour objectif de valider les fonctionnalités développées. Puis au niveau de l'utilité de l'outil d'annotation par (2) la construction effective d'une classification et (3) la confirmation d'experts sur l'adéquation de la carte de thèmes constituée et leur activité en conception mécanique. Pour l'évaluation de l'utilité et de l'utilisabilité, il sera nécessaire de mettre en place une expérimentation objective mettant en jeu les acteurs au cours de leur activité réelle. Dans le cadre actuel de notre recherche, nous ne pouvons effectuer cette expérimentation in situ, faute de participants et de temps. Cependant, la validation de la fonctionnalité de création de la carte de thèmes est possible. Pour cette fonctionnalité, notre scénographie prend en compte la langue spontanée utilisée dans la communauté de l'utilisateur au cours de son activité de conception via l'ensemble représentatif et exhaustif de documents utilisés au cours du projet. Ainsi, nous pourrions interroger les experts ayant participé au projet sur l'utilité de la CT.

L'outil conçu est validé par des scénarios construits suite à l'observation des échanges médiatisés effectivement produits et recueillis dans le corpus de référence. Nous allons maintenant présenter le corpus de travail, limité aux annotations, utilisé pour évaluer le module TAL d'AnT&CoW. Suit un exposé des scénarios d'usage utilisés tout d'abord pour la vérification des fonctionnalités puis dans la scénographie de vérification de l'utilisabilité de l'outil. Nous clorons sur la présentation des résultats de l'étape de validation des fonctionnalités de l'outil.

1.1. Présentation du corpus de travail

Après avoir recueilli un corpus large permettant de refléter le vocabulaire du domaine, nous délimitons un corpus de travail. En effet, de l'ensemble du corpus de référence, nous construisons un corpus adapté à notre objectif d'étude. Dans un premier temps, nous documentons ce corpus de travail, centré sur l'annotation. Dans un deuxième temps, nous donnons quelques exemples représentatifs de ce corpus et nous finissons par une description de la démarche de nettoyage et d'étiquetage de ce corpus d'annotations.

Objectif de constitution du corpus d'annotations

Le corpus de travail créé pour notre analyse est un corpus d'annotations. Il doit nous permettre non seulement d'observer les différents points de vue repérés dans la littérature (organisation, planification, argumentation et domaine) mais aussi de prendre en compte les situations d'annotation dans un projet pour mimer un scénario d'usage de notre outil.

Domaine, thèmes et objectifs du corpus

Ce corpus est constitué des annotations déposées par les membres du projet à propos de leurs documents. Il est principalement question de la conception et de la fabrication d'un réducteur pour adapter le moteur diesel automobile disponible sur le Delvion. Le second thème important est celui de la mise en place de phases de travail pour le montage de l'avion.

Des annotations

Le groupe observé ne disposant pas d'outils d'annotations, nous ne récupérons pas directement des annotations. Cependant, les membres du projet accompagnent leurs documents numériques de fichiers, ou les diffusent par mél en laissant leur commentaire en

corps de mél. Nous considérons donc ces messages comme des annotations dans la mesure où le corps du mél :

- est lié à une pièce jointe ;
- commente cette pièce jointe ;
- est lié à un autre message (réponse, citation, sujet identique).

En raison du manque de support à la coopération, de nombreuses pièces jointes ont été commentées en face à face et annotées manuellement. Nous ne les prenons pas en compte et nous limitons à celles qui ont été transmises et annotées dans un cadre médiatisé. Nous disposons ainsi d'un corpus d'annotation de 147 fichiers, 7700 mots diffusés entre janvier 2004 et mars 2006.

Acteurs du projet et auteurs des documents

Ce projet de conception mécanique est monté dans le cadre d'une association aéronautique, entre le bureau de cette association et une équipe de conception mécanique (des chercheurs du domaine ainsi que des techniciens). Au cours des trois années observées, vingt-cinq protagonistes échangent des messages directs et des documents. L'équipe d'ingénierie mécanique est composée d'une dizaine de concepteurs tout au long du projet, qui effectuent ce travail de manière asynchrone (pendant le temps libre de chacun) et distribuée (ses membres sont localisés sur trois sites). Les nationalités et âges des protagonistes sont variés mais en général les auteurs sont francophones natifs et/ou de niveau d'étude supérieur. Les protagonistes sont d'âge adulte, majoritairement de sexe masculin. Ils ont un usage courant du courrier électronique et de l'outil informatique en général. La majorité des messages recueillis pour notre corpus sont issus de l'équipe en conception.

Conditions de production

Les messages sont privés ou adressés à des groupes qui peuvent être constitués très différemment. Certains messages sont adressés à un groupe ou sous-groupe de concepteurs, de techniciens ou au plus large, à l'association. A l'intérieur de ces groupes, plusieurs rôles organisationnels sont représentés, académiques, associatifs, sociaux. Comme l'association est hébergée par une université et le projet développé par des chercheurs de l'université, le cadre organisationnel principal est académique. Les

concepteurs sont hiérarchisés en professeurs, maître de conférences, doctorants, étudiants-stagiaires, étudiants. Cependant, d'une façon plus large, on retrouve des ingénieurs, des techniciens, un chef de projet officiel qui organise les séances de travail et interagit avec le président de l'association d'une part et les membres de l'association d'autre part. Au niveau de l'association, les dirigeants de l'association ont aussi un pouvoir fort par rapport aux étudiants de l'université puisqu'ils ont le plus souvent une double casquette associative-académique. Pour les membres de l'association qui n'appartiennent pas à l'université, ce sont les rapports hiérarchiques sociaux classiques qui s'imposent, principalement par l'âge des protagonistes. L'ensemble des échanges, bien que médiatisés, ayant lieu sous des identités réelles, les forces en opposition agissent assez clairement dans les messages des protagonistes (formule de politesse, exigences d'une personne à l'autre, etc.).

Normes respectées, langue et style

Le ton respecté dans ces messages est très professionnel et reste en général amical. Nous sommes donc dans un style qui répond aux normes de politesse et d'adresse semi-formelles du français standard. Le style reste cependant plutôt oral. Dans l'ensemble, nous avons affaire à un usage restreint d'abréviations ou de signes du type *smileys* mais un usage massif de coquilles et fautes d'orthographe.

L'usage de messagerie électronique ajoute toutefois aux messages une masse importante de données automatiques qui rendent le mél spécifique. Nous considérons comme ajout de données *automatiques* les entête (date, réponse à, suivi de, Cc, Cci, « vous avez écrit », ...) et pied (signature insérées automatiquement, message d'hébergeur de messagerie, d'antivirus, etc.) de mél. Celles-ci contextualisent le message mais ne sont pas volontairement placées dans le corps du mél par le scripteur. Nous ne souhaitons donc pas les prendre en compte dans notre corpus. Nous allons donc nettoyer notre corpus d'annotation et notre corpus de référence pour ne pas être submergée par des données automatiques.

1.2. Recueil du corpus

Le corpus de référence a été recueilli auprès des concepteurs eux-mêmes. Ils nous ont procuré leurs documents de travail et une partie importante de leurs échanges par méls. En effet, nous n'avons pas pu récupérer l'ensemble exhaustif des messages échangés depuis

trois ans puisque beaucoup sont supprimés au fur et à mesure chez certains collaborateurs. De plus, nous n'avons pu contacter tous les protagonistes du projet certains étant distants physiquement ou ne souhaitant pas partager leurs messages privés. D'une façon générale, les concepteurs chez qui le suivi du projet est important conservent l'ensemble de leurs échanges. Dans leur cas, les échanges qui n'ont pu être récupérés sont les échanges en face à face qui ont eu lieu d'une façon informelle le plus souvent, au détour d'un café, d'un couloir, ou d'une séance de travail au hangar de l'aérodrome ou d'une façon plus formelle autour d'un document papier. Les documents qui n'ont pas pu être réutilisés sont des documents issus de logiciel propriétaires (logiciels spécialisés de conception). Dans un souci de respect, le corpus est anonymisé.

1.3. Traitement du corpus

Les données de départ gagnent souvent à être nettoyées et normalisées avant d'être soumises à un outil d'annotation. Bien que Tree-Tagger et Syntex soient robustes, nous avons effectué un nettoyage manuel de nos données de départ. L'étiquetage reste tout automatique.

Dans le cas d'un corpus de mél, les données automatiques sont répétitives et aisées à supprimer. Notre corpus de référence est à la base constitué de 287 734 caractères. Après nettoyage des éléments suivants, nous allégeons notre corpus de 48,5%. Le nettoyage du corpus a lieu par une suppression :

- des chevrons (>)
- des entêtes
- des informations type : re / fwd
- des signatures automatiques
- des messages anti-virus en entête et pied de mél
- des retours chariot en fin de ligne (↵)

Nous conservons par contre les sauts de ligne en fin de paragraphe (¶) et les *sujets* des méls (avec parfois des crochets de signalement ([]) de liste de diffusion ou de thème suivi). Dans le cas de citations, il existe des entêtes intermédiaires pour lesquelles nous appliquons les mêmes règles de conservation et suppression.

Ainsi nous nettoyons notre corpus de près de 150 000 caractères. Celui-ci est alors étiquetable d'une façon plus rapide pour nos outils. Dans le cas d'une évaluation de l'outil en contexte naturel, ce nettoyage des données n'est bien entendu pas possible.

Chaque échange est enregistré sous forme de fichier individuel. Ces fichiers sont ensuite classés chronologiquement et selon les producteurs et destinataires du message. Il est alors facile de comprendre la logique de communication du projet pour un utilisateur humain et de retrouver les rôles des protagonistes en repérant les tensions marquées dans le corps des messages. Cette analyse qualitative du corpus permet de valider la représentativité de la structuration en thèmes et en point de vue des annotations via l'outil.

Nous constatons que le corpus est bien représentatif des trois types d'annotations définies précédemment. Il est possible d'identifier des annotations fragments de conversation, fragments de document et d'indexation.

1.3.1. Des fragments de conversation

De l'éclairage théorique sur l'annotation pour la conversation exposé précédemment, nous pouvons décrire des échanges tels que (1), (2), (3) et (4) (Fig. 45). Saturnin est chef de projet et fait le lien entre le Conseil d'Administration (CA) de l'association qui commande le moteur et les groupes de mécaniciens et bénévoles qui conçoivent, produisent et installent les pièces. Désiré est mécanicien - pilote et a en charge la pose du moteur et les tests de l'avion. Félicien est un ingénieur qui dirige une petite équipe de technicien dont fait partie Adelphe. Adelphe travaille plus particulièrement sur des commandes émanant de concepteurs tels que Léon.

Dans cette série d'échanges, Saturnin essaye d'obtenir les pièces nécessaires au montage du moteur. Félicien lui demande de suivre la procédure normale où le technicien (Adelphe) fabrique certaines pièces bien que lui - même ait les outils nécessaires (fraise acier) disponibles. On voit ici au fur et à mesure des réponses que Saturnin négocie le fraisage par Félicien mais que celui-ci ne répond que par la négative mettant en avant la distribution préconisée des rôles. Les thèmes et reprises de thèmes sont mis en gras.

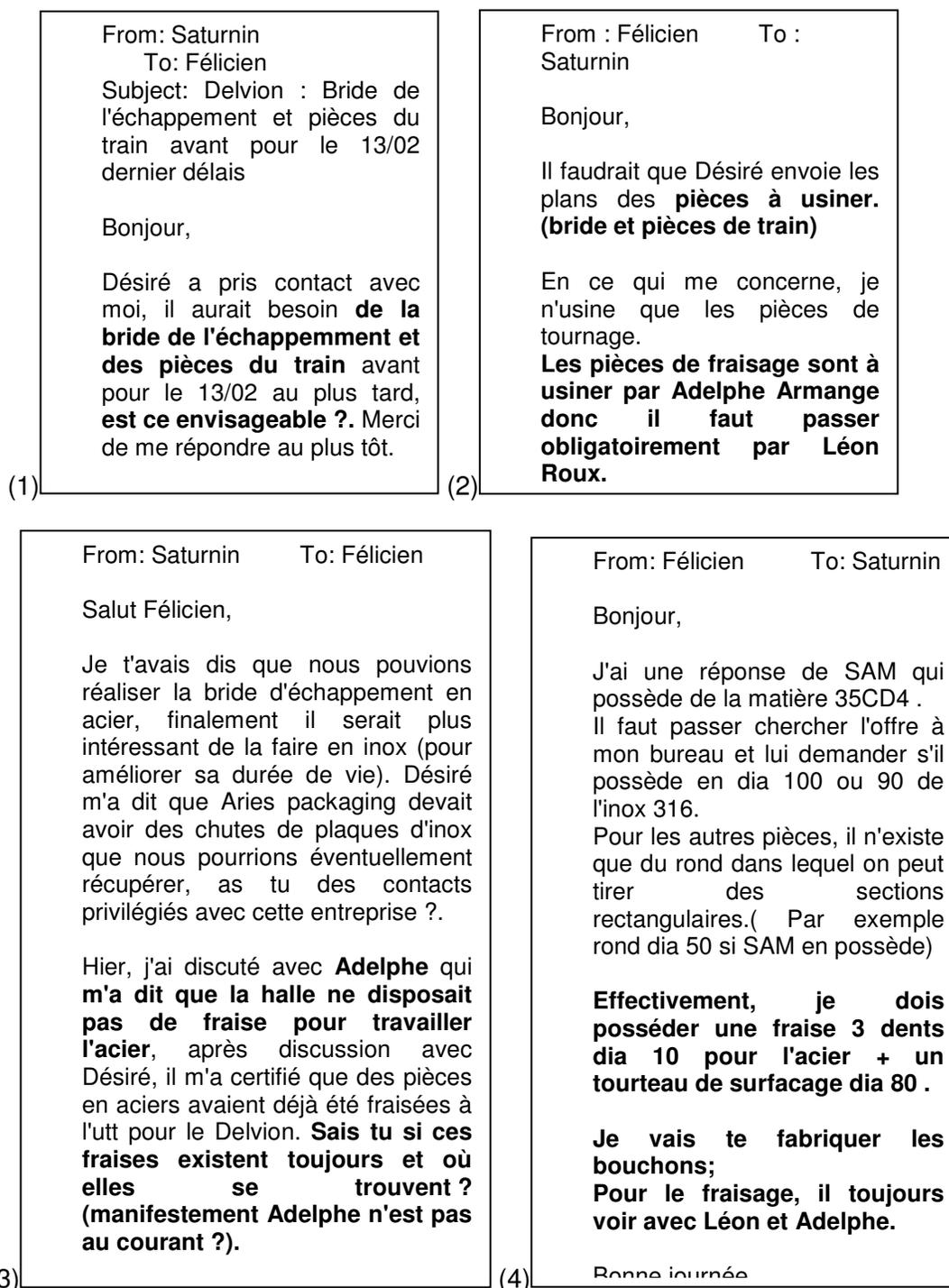


Fig. 45. FRAGMENTS DE CONVERSATION

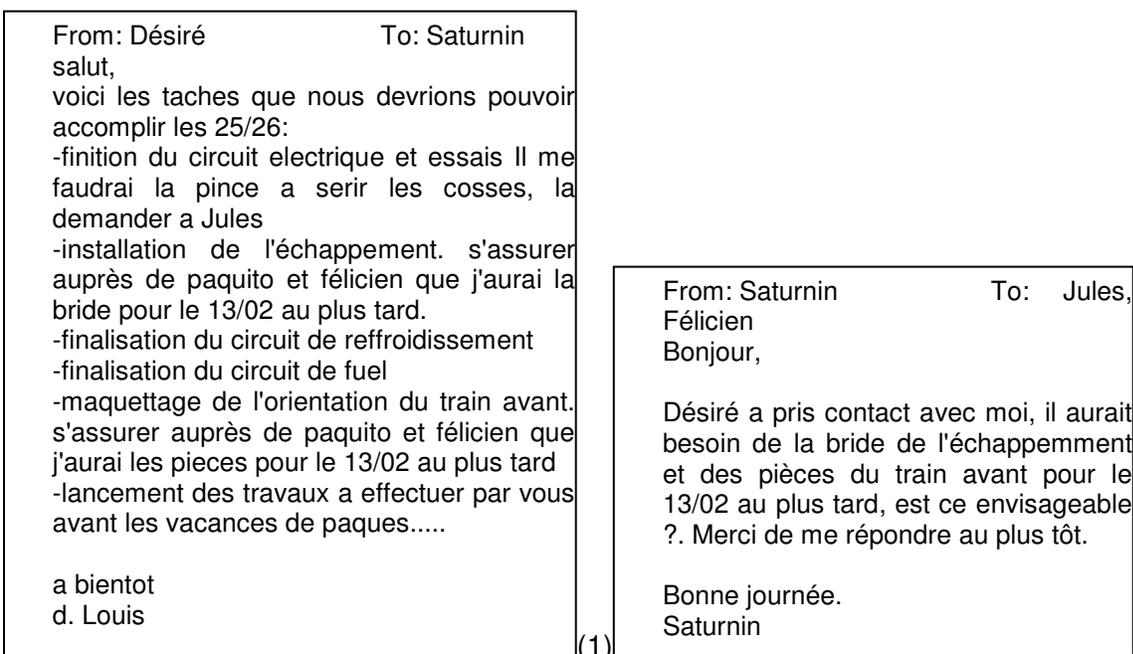
Cette négociation ponctuelle est représentative de nombre d'autres plus larges en général en impliquant plusieurs acteurs. Tout au long du déroulement du projet, on constate

que l'argumentation est importante. Pour soutenir l'annotation sur le plan de l'argumentation, l'outil prévoit une annotation selon le type d'argumentation portée par le corps de l'annotation. Cette typologie est extraite semi-automatiquement du corpus.

1.3.2. Des fragments de document

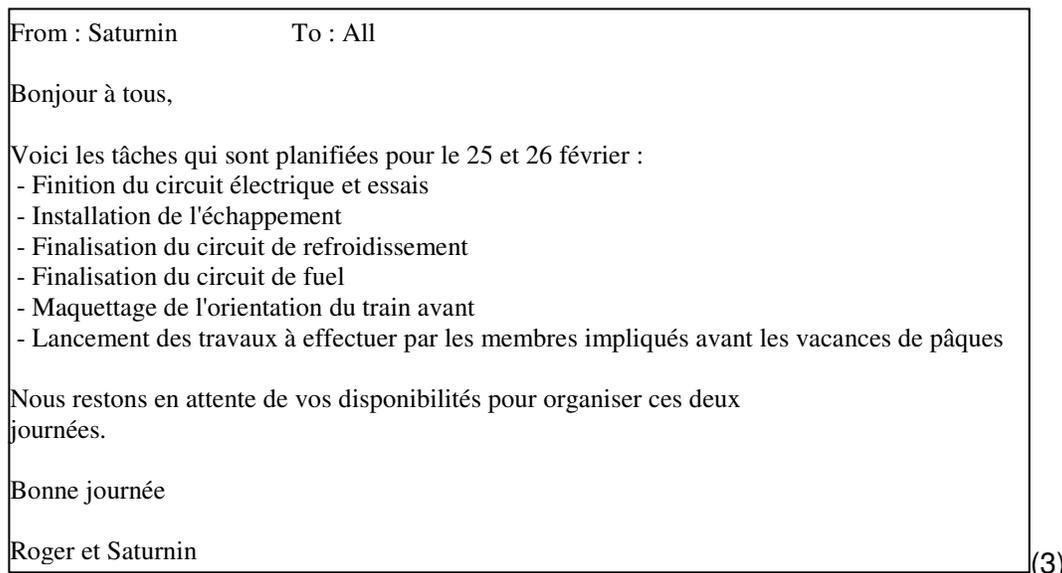
Tout au long d'un projet, les participants échangent des méls dans une visée d'élaboration ; élaboration d'un référentiel commun, de solution ou de documents. En étudiant les échanges de mél tels que présentés ci-après, on constate un grand nombre d'échanges qui ont pour but d'élaborer des documents, même encore transitoires. Dans cet échange de (1), (2) et (3) (Fig. 46), le chef de projet organise une séance de travail avec les bénévoles, et pour cela il doit réussir à mettre en phase des chantiers et des disponibilités de bénévoles. Pour cela, il doit vérifier la faisabilité d'un chantier puis mettre au courant (toujours par mél) tous les membres de l'association. Ces trois méls appartiennent à un échange beaucoup plus long (environ 40 méls tout auteur et destinataire confondus) qui définit la faisabilité des différents chantiers.

Suite à une demande de Saturnin pour relancer le projet en veille depuis le début de l'hiver, Désiré fait une liste des principaux chantiers à mener (1). Chaque chantier dépend de la mise à disposition de pièces et donc de financements. Saturnin le chef de projet doit veiller à la bonne articulation de ces phases. Il lance donc une négociation avec le CA et c'est de la réponse du CA que dépendra la suite du projet. (3) est un en fait un résumé récapitulatif de ce qui va être mis en place.



(1)

(2)



(3)

Fig. 46. DES FRAGMENTS D'ÉLABORATION DE DOCUMENT

1.3.3. Des fragments pour l'indexation

La collection de documents numériques est un élément partagé qui permet d'ancrer des tâches distribuées dans un référentiel commun (Clark, 1996) autorisant ainsi une

médiatisation de ce travail. Les échanges permettent l'élaboration d'un référentiel commun, qui se traduit par une élaboration collective de schèmes de classification lorsqu'il s'agit d'organiser des documents. Dans un projet, un acteur va ainsi participer à la création de documents et à leur classement selon un plan de classement défini.

Lors d'échanges au cours d'un projet, les utilisateurs vont naturellement vers une indexation de leur document. Cette indexation est particulièrement visible dans la structuration des méls. En effet, lors de l'observation de notre corpus, nous constatons l'utilisation des titres de messages (*subjects* dans la figure 46.3) pour indexer le message interne voir les pièces jointes au message. Ces index sont utilisés pour retrouver le thème d'un échange, principalement par rapport au domaine auquel a trait le message (« matières brutes, chargeur, brides,... »), mais aussi pour indiquer la phase de planification abordée (« répartition des tâches ») ou encore pour noter quel type d'argumentation porte le message envoyé (« question »).

Il est ici bien clair que cette indexation est modifiée selon l'étape du projet puisque par exemple sur le thème général « Brides et Support Compas » :

- a) >Subject: Brides et support compas ;
- b) >Subject: Re: Plans Brides et support compas ;
- c) >Subject: RE : Delvion : fabrication des brides et du support compas.

la première occurrence est générale, la seconde est focalisée sur la partie *plan* quant à la dernière, elle indique que nous ne sommes plus dans une phase de conception mais dans la phase suivante de *fabrication*. Ce type de classification ad hoc montre bien l'importance d'avoir une classification évolutive, non figée et adaptée au domaine.

De même au niveau planification, les exemples :

- d) >Subject: Re: Delvion : CR à imputer pour la réalisation ;
- e) >Subject: Re: Imputations sur CR132C.

montrent l'utilisation de titre du message pour classifier au niveau du domaine (« Imputation, CR »), mais aussi pour communiquer des informations (« CR132C ») et marquer l'argumentation (« Re » indiquant une réponse). Comme dans toute classification, il est possible de retrouver des *cagibis*, des ensembles neutres pour classifier d'une façon globale des éléments généraux comme « projet Delvion » ou « Divers Delvion » (Fig. 47) servant à différencier ces messages de messages hors projet.

<p>Domaine :</p> <ul style="list-style-type: none"> >Subject: DAIR-100 >Subject: Delvion : Dimensions des pièces brutes >Subject: Delvion : fraise acier >Subject: Re: Delvion : Pb de fraise >Subject: Tolérances d'ajustement >Subject: Delvion : Support Compas >Subject: Fwd: Matière brute >Subject: Etude de la commande des gazs >Subject: Modèles CAO des pièces >Subject: Poste à soudure TIG >Subject: Re: Chargeur de batterie >Subject: Re: Bride d'échangeur >Subject: Bride d'échangeur >Subject: Devlion : Bride échangeur >Subject: Delvion : Bride de l'échappement et pièces >Subject: Matériau des brides de l'échappement >Subject: Fwd: Delvion : Bride d'échappement >Subject: Devlion : Bride et pièce de train >Subject: Re: Pièces: Bride et train >Subject: Re: Pièces: dride et train >Subject: Brides et support compas >Subject: Re: Plans Brides et suport compas >Subject: RE : Delvion : fabrication des brides et du support compas 	<p>Planification :</p> <ul style="list-style-type: none"> >Subject: Re: Adresses non correctes >Subject: Re: Compléments d'adresse >Subject: Re: Pièces prévues pour le 13 Février >Subject: Re: Répartitions des tâches >Subject: Re: réponse aux questions >Subject: Re: Tâches à accomplir le 25/26 février >Subject: week end de février >Subject: Re: Bilan du Week-end du 25/26 Fevrier >Subject: Re: Delvion : CR à imputer pour la réalisation >Subject: Re: Imputations sur CR132C
	<p>Argumentation :</p> <ul style="list-style-type: none"> >Subject: Question à propos de la bride >Subject: Question à propos du support compas
	<p>Divers :</p> <ul style="list-style-type: none"> >Subject: Re: divers DELVION

Fig. 47. SUJETS DES ÉCHANGES ORGANISÉS PAR POINTS DE VUE SUR LE PROJET

Dans cet ensemble de méls, nous reconnaissons alors bien l'élément joint (la pièce jointe, le texte cité), son annotation (le nouveau message laissé par un auteur), et l'indexation de cet ensemble (le titre, le(s) thème(s) de l'annotation).

Une fois documenté de la sorte, notre corpus avant et après nettoyage, avant et après étiquetage est réutilisable par d'autres et pour d'autres expérimentations pour de la linguistique de corpus manuelle ou semi-automatique. Dans notre cas, nous l'utilisons donc pour simuler un usage de notre outil. Nous simulons le dépôt d'annotations (corpus d'annotations) sur les documents du projet (corpus de référence). Ces scénarios nous

permettent d'évaluer l'usage de notre outil d'annotation dans un contexte de conception médiatisée.

2. Trois scénarios d'annotation dans un projet de conception distribuée

Nous avons développé nos scénarios dans l'optique exposée par (Nardi, 1995). Sa proposition est de générer des scénarios en mêlant des éléments ethnographiques à des artefacts technologiques. Cela revient à combiner des éléments d'observation de l'activité humaine quotidienne en lien avec des outils. Le scénario modélise d'une façon nouvelle l'activité instrumentée. Nous nourrissons nos scénarios de l'expérience *ethnographique* du corpus. Les scénarios développés ci-après se déroulent tous trois dans les mêmes conditions de production que le corpus d'annotation. Nous représentons les échanges tels qu'ils ont eu lieu, c'est-à-dire médiatisés (par le biais d'une messagerie électronique). Leur objectif est identique. Il s'agit pour le groupe - projet de travailler en coopération pour concevoir un moteur d'avion diesel. Cette conception est favorisée par l'utilisation de fonctionnalités d'interprétation, d'élaboration et d'indexation de documents. Alors que les scénarios sont en général utilisés pour déduire des primitives de conception d'outils (Carroll, 2000) dans une optique de modélisation des interactions d'objets et de classes du domaine, nous suivons plutôt les scénarios dans l'optique d'évaluation des fonctionnalités de l'outil contextualisées de J. Bardram (1998).

La scène de base de nos scénarios est donc l'utilisation d'un collecticiel qui permet à des utilisateurs d'optimiser leurs échanges autour de documents via l'annotation de documents. Les participants fondent leur travail sur des plans, des maquettes 2D ou 3D, des documents qu'ils s'échangent en format numérique par pièce jointe de mél ou par courrier. Ces documents sont la base du travail, ils expliquent ou représentent le travail à effectuer durant la phase de fabrication. Ils sont modifiés, hors ligne et individuellement, au fur et à mesure des échanges et il est important que les intéressés soient tenus informés des mises à jour et modifications des documents. Un élément clé de leur activité est l'annotation : Les membres de l'équipe de conception annotent les documents pour les modifier, pour informer les autres membres de l'importance d'une modification ou expliquer une telle modification, accompagnent leurs documents numériques d'une note dans un corps de mél, expliquent par un mél le contenu d'un document joint,... L'observation de leur activité de conception nous a

permis de définir trois scénarios dont les acteurs sont Léon, un chercheur en conception mécanique, Adelphe, un technicien et Pépito le chef de projet. Ils échangent des documents et communiquent à l'aide essentiellement de l'outil AnT&CoW.

2.1. La communication

2.1.1. Scénario d'annotation support à la communication

En tant qu'ingénieurs / chercheurs en conception mécanique, Léon et Pépito vont démarrer le projet par une étude des besoins et de l'existant. Pour ce faire, ils échangent d'une part avec les pilotes et mécaniciens de l'association, et, d'autre part, ils lisent de la documentation sur des projets similaires. Lors de ces lectures, ils échangent entre eux, expliquant à l'autre les points positifs ou négatifs qu'ils trouvent dans leur documentation. En critiquant les rapports de projet et autres documents de conception, ils définissent l'adéquation au projet de certains éléments et en éliminent d'autres. À propos d'un site Web en rapport avec l'adaptation d'un moteur diesel sur un avion, Léon soulignera que le type d'avion n'est pas le même et qu'il faut s'attendre à des problèmes de poids par rapport à leur Delvion. Pépito marquera son accord en répondant qu'il faudrait penser à alléger au maximum le moteur. Après de longues phases de lecture et de négociation sur de nombreux points, Léon le concepteur commence à réaliser les premiers plans de pièces. Il les partage, principalement avec Pépito et ajoute à propos de ces plans de nombreux commentaires, affirmant qu'il existe un problème dans le choix de la matière, expliquant le choix d'une pièce par rapport à une autre par sa disponibilité immédiate, demandant une vérification de calcul sur une autre pièce. Au fur et à mesure de ces échanges sur plans, un plan finalisé est conçu. Adelphe s'y intéresse alors puisque c'est lui qui va mettre la pièce en fabrication. Comme Adelphe n'a pas forcément suivi toutes les négociations, il a besoin de visualiser les plans avec les annotations et leurs réponses. Ces plans annotés permettent à Adelphe de comprendre les raisons de la conception d'une pièce qu'il ne juge pas optimale en qualité. Connaissant la base des décisions, il s'autorise à signaler aux concepteurs que le choix de la matière n'est pas judicieux au niveau du poids et propose un autre matériau, plus léger dont il dispose justement en quantité suffisante suite à un autre projet.

Les annotations permettent ici à la fois d'élaborer collectivement du sens autour d'un document, mais aussi d'avoir un suivi de la négociation et une planification du projet avec des retours possibles entre les différentes phases de conception. L'outil indique explicitement

qui a déposé tel commentaire et pourquoi de façon à mettre en lien les documents et leurs commentaires, mais aussi relier les concepteurs au travers de leurs échanges.

2.1.2. Scène 1 : Échange autour de documents

Nous qualifions de scénographie la mise en perspective des scénarios au travers de l'outil. La représentation de l'activité des annotateurs via notre outil nous permet de vérifier les fonctionnalités développées en fonction d'un contexte d'activité. Nous allons maintenant exposer les copies d'écran des étapes principales le scénario précédent.

La capture en figure. 48 montre les échanges entre Léon et Pépito à propos des plans mis en place. Les échanges sont ancrés en différents lieux du document, comme le montre la chaîne d'annotation dans le plug-in. Les fils d'annotation débutent de « Bride_boitier » et de « Réducteur_Air_Campus ». De même, il est possible d'afficher le fil entier d'annotation avec l'ensemble des réponses apportées à une annotation ou au document ainsi que leur auteur.



Fig. 48. ÉCHANGE ENTRE PÉPITO ET LÉON SUR BRIDE ET RÉDUCTEUR

2.2. L'élaboration

2.2.1. Scénario d'annotation support à la création de documents

Lorsque Léon et Pépito échangent pour créer une maquette précise, pouvant être fabriquée, non seulement ils intègrent des représentations physiques du plan selon les commentaires (modification d'une cote, ajout d'une vis,...) mais ils modifient aussi des documents explicatifs qui accompagnent le plan. Ainsi, Léon explique lors du partage d'une nouvelle version du réducteur du moteur qu'il a ajouté des trous aux flasques droit et gauche comme discuté au cours d'une première évaluation du modèle. Ces documents, ou fragments textuels de document, sont bien sûr les annotations, mais aussi des parties de rapports qui seront transmises avec le plan afin d'éclairer certains points qui peuvent rester confus. Ces parties de rapport sont directement liées aux annotations déposées au cours de la négociation. Elles expliquent des choix et les causes de ces choix par rapport à différents contextes. Le document créé grâce au recueil d'annotations et leur mise en forme dans un document plus homogène représente la logique de communication qui permet de comprendre la logique de conception d'une pièce, ou du plan qui représente cette pièce. L'ensemble de ces documents ou fragments de document sera une somme de nombreux problèmes et solutions posés au cours d'un projet de conception en mécanique aéronautique. Il s'agit, pour que cette somme reste compréhensible a posteriori, d'organiser ces fragments selon leur apport argumentatif ou technique dans le projet. L'organisation des fragments permet d'éviter de laisser aller le document vers une décomposition thématique (Herring, 1999).

2.2.2. Scène 2 : Élaboration de document via des annotations

2.2.2.1. Un fil traçant la logique de communication

La concaténation chronologique voire thématique des annotations déposées par les différents collaborateurs permet de tracer la négociation au cours de la conception. Cela crée un document comme visible en Fig. 49.

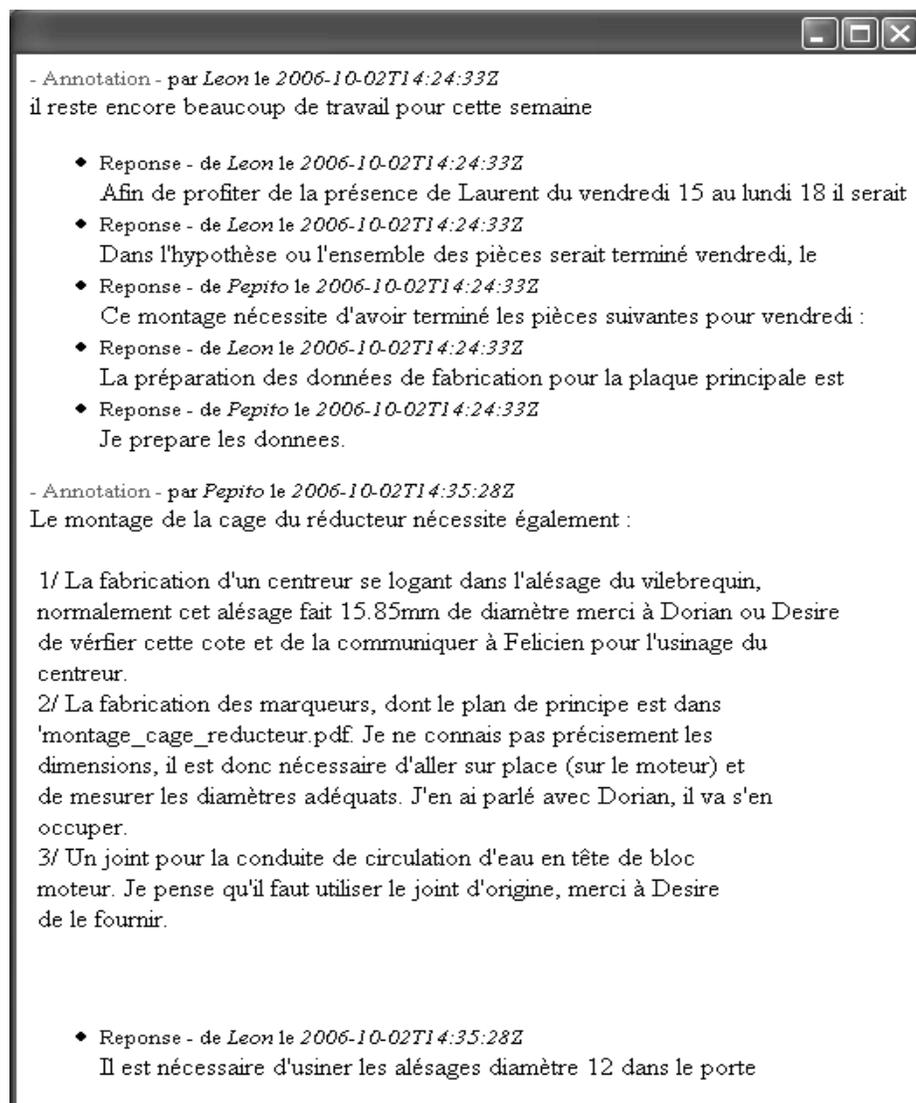


Fig. 49. LOGIQUE DE COMMUNICATION

2.2.2.2. Un document élaboré coopérativement

Les fragments de conversation concaténés produisent un document qui peut être remis en forme afin de permettre de créer un nouveau document, base de nouveaux échanges. Nous avons affaire alors à une concaténation de fragments de documents de divers auteurs en un rapport réutilisable et renégociable. Le document de la figure 50 montre une somme des négociations entre Pépito et Léon. Cette somme est stockée sous l'onglet documents du projet.

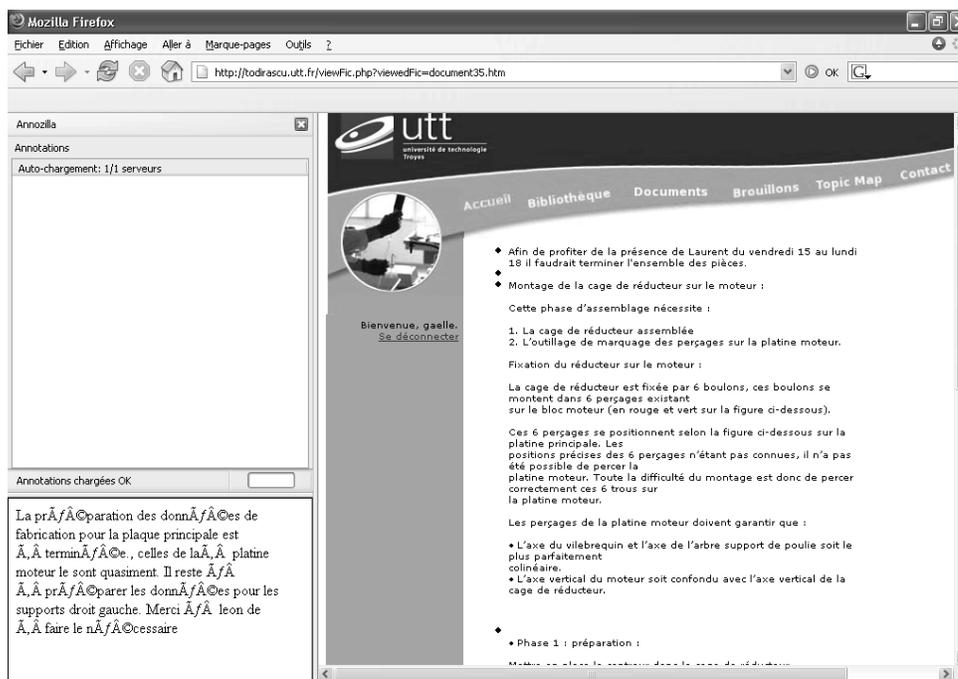


Fig. 50. DOCUMENT CRÉÉ SUR LA BASE D'UN RECUEIL D'ANNOTATION

Le document permet de faire passer les informations à Adelphe qui peut le commenter à sa guise (voir Fig.51) :

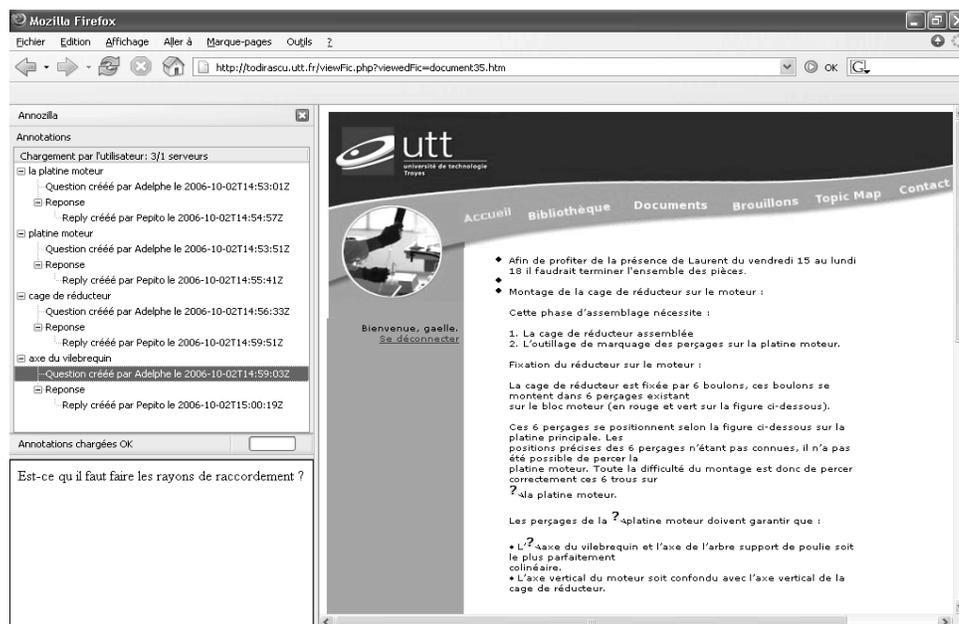


Fig. 51. ANNOTATION D'UN DOCUMENT DU PROJET

À nouveau, il est possible de créer un document qui planifiera les activités à effectuer par Adelphé ainsi que les explications de la nécessité de ces activités (Fig.52).

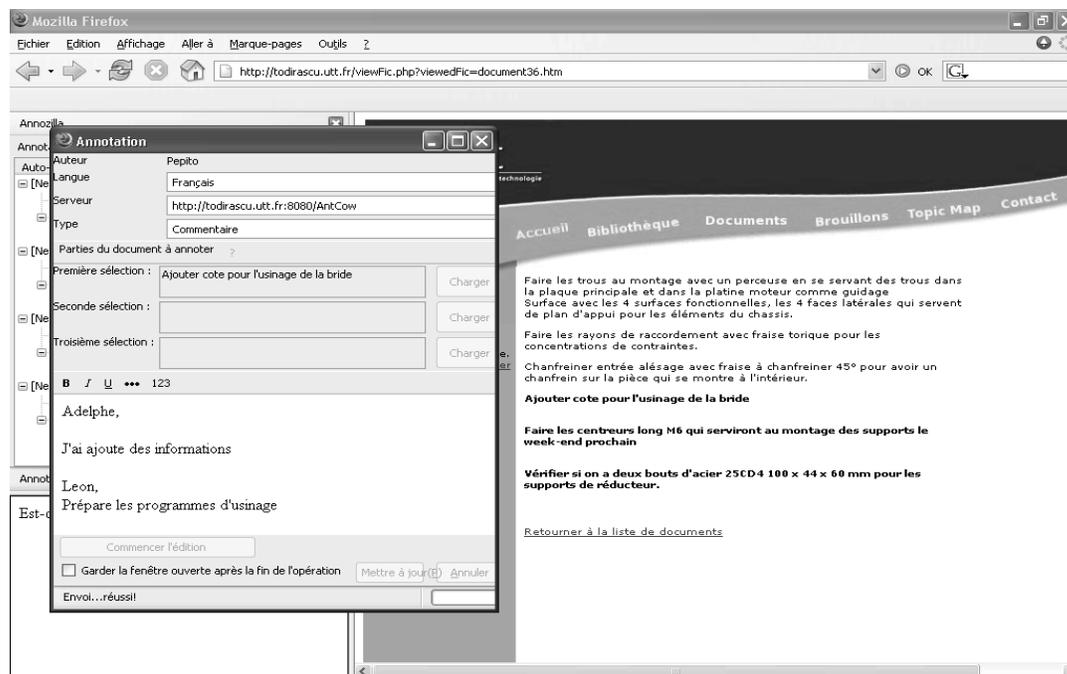


Fig. 52. RECUEIL DE SOLUTIONS POUR LA PLANIFICATION

2.3. L'indexation

2.3.1. Scénario d'annotation support à l'indexation

Pour organiser leur travail, les participants au projet créent naturellement une classification qu'ils partagent le plus souvent implicitement. La première classification opérée est celle effectuée lors de la saisie d'une annotation et de son *objet*, donnant un titre au texte saisi. En donnant un titre à leur message ou à leur document, Léon, Pépito et Adelphé partagent une classification commune. Le plus souvent ce titre est en lien soit avec une pièce (visserie, réducteur), soit avec une phase du travail (plan, maquette, fabrication) soit avec une discussion (problème, ennui, solution) ou en cumulant ces éléments (problème réducteur). De même, ils possèdent une classification individuelle de leurs documents, mais qui tend à partager les principes à la fois des classifications existant en conception (étapes de conception et fabrication) et de la classification inhérente au projet. Celle-ci n'est visible dans leur communication que par les noms des documents en pièces jointes et les noms des

archives envoyées. Elle prend en compte les mêmes points que précédemment en y ajoutant les noms des formats ou des outils utilisés pour générer les documents (modèle STEP, Catia, plan Goélan). Le titre accompagnant un message ou un document est donc révélateur d'une classification que l'on valide en la partageant. Ce titre est un mot-clé qui permet de retrouver un message, de comprendre le point de vue du message et donne des pistes pour la compréhension de la pièce jointe. De même, le message accompagnant une pièce jointe possède des mots spécifiques qui expliquent le contenu du document joint, ses modifications, son envoi. Le corps du mél annote le document, et par ce lien souvent implicite entre le contenu du document et celui de l'annotation, l'un indexe l'autre. Les mots spécifiques de l'annotation permettent d'ores et déjà au lecteur de classifier le document joint au moins en partie, de même que les thèmes du document peuvent permettre d'orienter l'interprétation de l'annotation. L'indexation est un lien complexe qui structure les documents et l'annotation. Les utilisateurs indexent donc leurs documents sagement et légèrement grâce au couplage des outils de traitement du langage au corps de l'annotation.

2.3.2. Scène 3 : Construction de la classification et indexation

2.3.2.1. Un exemple de carte de thèmes pour le projet de conception en mécanique aéronautique

Le carte de thèmes du projet possède plusieurs niveaux. Nous présentons ci-dessous les points de vue (encadré gauche Fig. 50) et plus précisément le point de vue argumentation et les concepts enregistrés avec cette caractéristique. Ensuite, nous présentons un exemple de termes enregistrés comme concepts (Encadré gauche Fig. 51) et le terme/concept réglage encadré de ses occurrences. Enfin nous présentons des associations découvertes automatiquement suivant des patrons (Fig. 52) et plus précisément l'association.

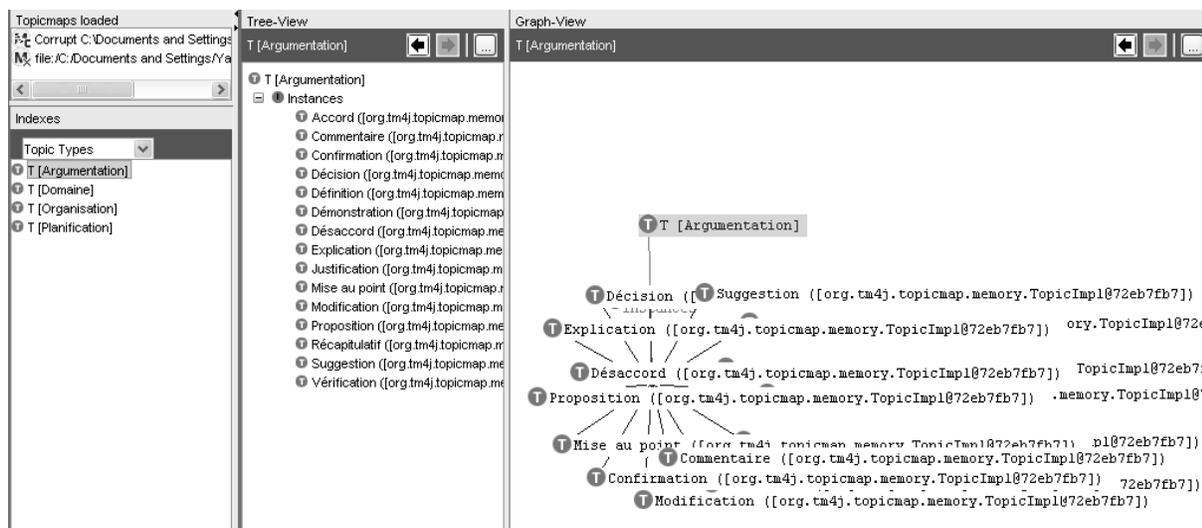


Fig. 53. VISUALISATION DU POINT DE VUE ARGUMENTATION ET DE SES CONCEPTS

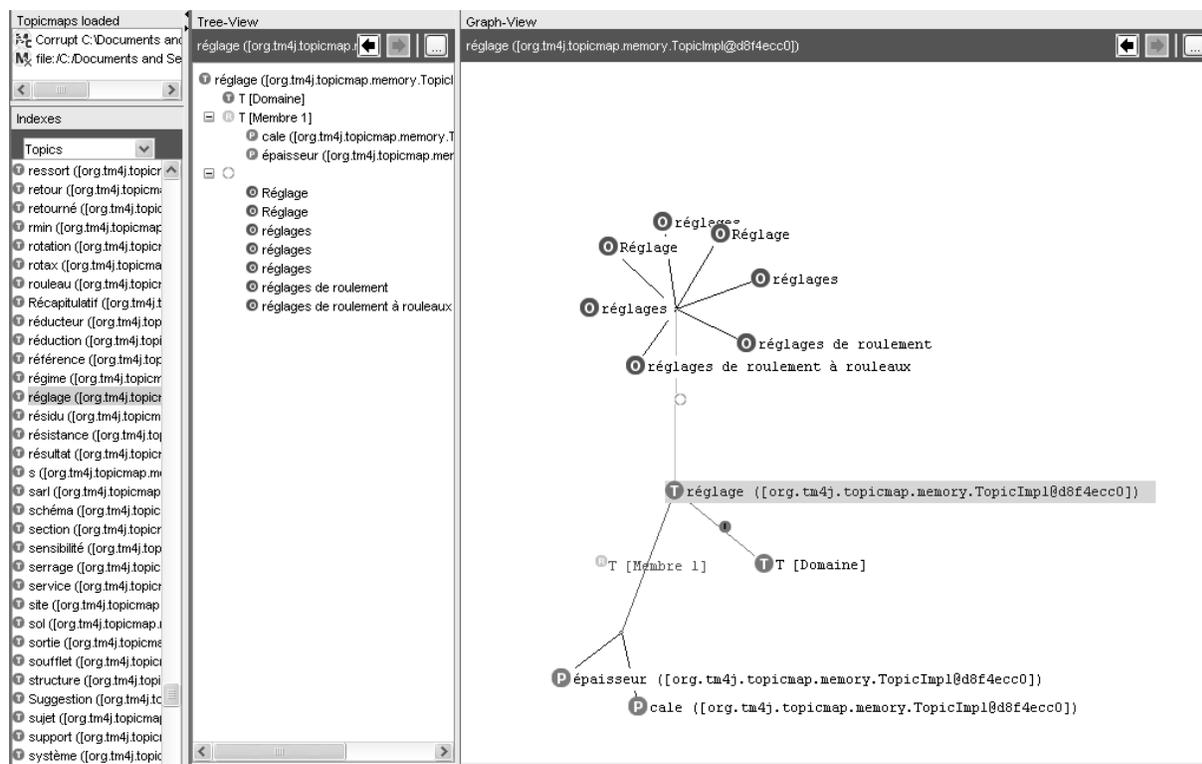


Fig. 54. VISUALISATION DU THÈME RÉGLAGE ET DE SES OCCURRENCES.

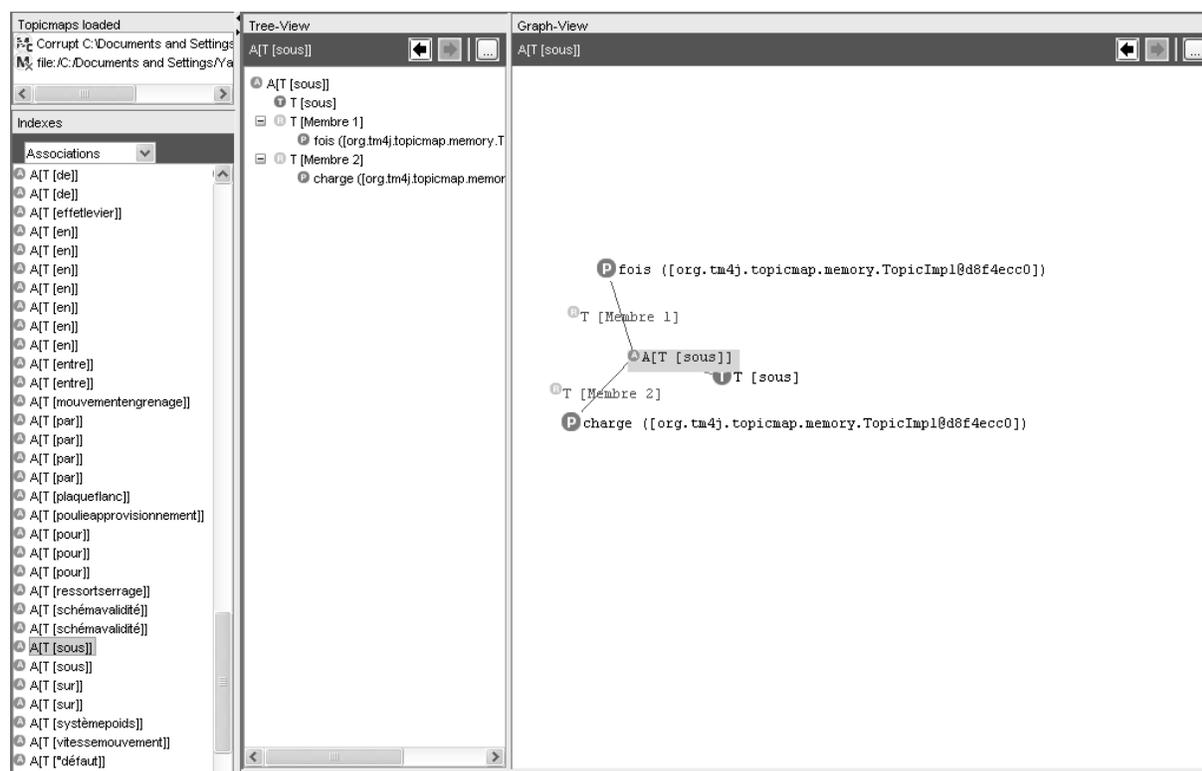


Fig. 55. VISUALISATION D'ASSOCIATIONS DÉCOUVERTES DANS LE CORPUS

3. Évaluation

3.1. Description du protocole

Suite au jeu du corpus dans l'outil, nous avons mis en place un protocole d'évaluation sur trois éléments : l'utilisation des fonctionnalités, la validation des scénarios et l'utilité des fonctionnalités. Au cours d'un entretien d'environ deux heures, les experts, face à l'ordinateur, manipulent l'interface et répondent au questionnaire⁴⁰. Les experts sont les chefs de projet impliqués dans le projet Air Campus et mis en jeu dans le scénario via le corpus.

⁴⁰ Le questionnaire est disponible en annexe.

3.1.1. Utilisation des fonctionnalités d'annotation

3.1.1.1. Manipulation de l'outil

La phase de manipulation consiste en l'utilisation des fonctionnalités de base. Après la découverte de l'outil avec l'expérimentateur (poster une annotation, répondre à une annotation, ancrer une annotation, demander de fournir un descripteur à l'annotation), il s'agit pour les experts de tester individuellement les fonctionnalités offertes pour l'annotation. Dans l'ordre, il s'agit de :

- Multi-ancrer d'une annotation
- Poster une annotation
- Répondre à une annotation
- Générer un brouillon à base d'annotations d'un document
- Éditer et modifier un brouillon
- Annoter un brouillon

Les questions ayant trait à l'interface et l'utilisabilité des fonctionnalités sont les suivantes :

1. Considérez-vous l'interface comme intuitive ?
2. Les fonctionnalités sont-elles simples à retrouver ?
3. Les documents sont-ils aisés à récupérer ?
4. Les annotations sont-elles aisées à visualiser ?

3.1.1.2. Résultats

Les experts considèrent que l'interface est simple d'accès et adéquat pour le partage de document à distance puisque c'est une interface Web. Les fonctionnalités sont disponibles classiquement par des clics droit et un menu déroulant ou un double-clic.

La visualisation du fil d'annotations suit les technologies actuelles de visualisation chronologique de type forum et cette structuration non seulement chronologique mais en plus argumentative dans le brouillon semble intéressante pour les experts qui cherchent à tracer la logique de communication sous-jacente à la logique de conception. Cependant, même si

les annotations sont faciles à voir dans le document elles restent plus difficile à voir dans leur ensemble.

Ainsi, les fonctionnalités correspondent à des besoins actuels. Cependant, l'outil modifie l'organisation du projet. En effet, il existe toujours un fossé important entre l'utilité des fonctionnalités par rapport à l'acceptation d'un (encore) nouvel outil par les concepteurs.

3.1.2. Validation des scénarios

3.1.2.1. Confrontation scénarios - expert

La phase de validation des scénarios selon les experts consiste tout d'abord en la lecture des scénarios puis, dans une deuxième phase, de leur visualisation lorsque les scénarios sont joués (simulation présentée précédemment). Les experts répondent après coup au questionnaire.

Les questions ayant trait aux trois scénarios et permettant leur validation sont les suivantes :

5. Les scénarios sont-ils représentatifs des échanges en conception mécanique ?
6. Les scénarios sont-ils représentatifs d'une activité d'initialisation de projet ?
7. Les scénarios sont-ils représentatifs d'une activité de planification des tâches ?
8. Les scénarios sont-ils représentatifs des rôles des contributeurs ?
9. Pour quels type de tâches les scénarios sont-ils représentatifs (phases du projet) ?
10. Quelles activités seraient à mettre ne scénario pour représenter l'ensemble de processus de conception ?

3.1.2.2. Résultats

Les experts considèrent que les situations décrites dans les scénarios sont vraisemblables au cours de phase de travail asynchrones et distantes. Dans le cas de phases synchrones, le travail en présentiel est plus simple à mettre en place et c'est pourquoi

la plupart des phases de travail⁴¹ du projet Air Campus n'auraient pas suivi ces scénarios. Ce constat était attendu, vu la réutilisation des échanges pour du TCAO.

Tout au long des projets en conception mécanique, des échanges non formalisés entre les phases de conception/fabrication ont lieu. Dans ces échanges, le collecticiel AnT&CoW semble pertinent. En effet, les échanges tacites, le plus souvent effectués en présentiel et en direct avec un interlocuteur désigné par une relation personnelle et non par la hiérarchie, peuvent alors être assistés et enregistrés. Lors de certains projets où la fabrication implique des sous-traitants, il est nécessaire qu'un suivi fin des échanges soit possible même si les échanges sont anonymisés. En effet, la structure hiérarchique en conception est généralement assez flottante : le chef de projet est principalement un coordinateur. Son rôle est reconnu en tant qu'il transfère les informations et qu'il les reçoit. Ainsi, tout se joue sur la possession de l'information plus que la hiérarchie. Le rôle est reconnu dans la prise de décision ou le transfert de l'information, sur la base des connaissances. L'anonymat n'est donc pas rédhibitoire dans la prise de décision au contraire. En effet, pour permettre l'intégration de connaissances nouvelles il est nécessaire que les frontières hiérarchiques soient peu nombreuses et franchissables. Il existe tout de même un processus de validation des documents spécifique et assez figé dans les projets de taille importante. Une fois un document validé/reconnu par un chef de projet, l'information est diffusable et tout le projet y a accès. Malgré un processus de type « Workflow ⁴²», l'information est ensuite diffusée sans autres barrières hiérarchiques et peut être commentée par tous les acteurs du projet.

Les spécifications d'usage définies pour le collecticiel portent principalement sur l'envoi d'échanges et de commentaires à propos de pièces jointes. Les experts considèrent donc les scénarios comme représentatif de ce type d'activité et les fonctionnalités comme utiles pour les comptes-rendus et revues de projet ainsi que pour le retour d'expérience (Design Rationale, prise de décision). Les fonctionnalités devraient s'avérer efficaces pour tracer, structurer la conception puisque les fonctionnalités de communication permettent l'assistance à une activité, celles d'indexation permettent l'analyse et la catégorisation d'activités et celles d'édition permettent la combinaison du support et de la catégorisation, soit un retour sur le projet utile dans le cadre d'un projet de « reconception » par exemple.

⁴¹ Il est aussi intéressant de noter qu'en conception mécanique, « travail synchrone » signifie des activités ayant lieu dans une même phase du processus. Dans ce paragraphe, nous entendons travail synchrone comme « dans une unité de temps » pas « dans une unité de phase ».

⁴² Flux de travail où les tâches sont organisées (circuit de validation, mode de validation, circuit des informations...)

3.1.2.3. Les scénarios instrumentés

Une fois les scénarios validés par les experts, les scénarios instrumentés sont joués dans l'outil par l'expérimentateur.

Les questions ayant trait à la validation des trois scénarios instrumentés sont les suivantes :

11. L'outil fait-il obstacle à la mise en place des échanges (gêne de la communication, difficulté à communiquer par écrit, ...) ?

12. L'outil fait-il obstacle à la compréhension des échanges (modification du sens, contresens, décontextualisation du problème, ...) ?

13. L'outil fait-il obstacle à la compréhension des documents partagés (validité du document, type de documents, ...) ?

14. L'outil fait-il obstacle à la création de documents ?

15. L'outil permet-il de jouer ces scénarios d'une façon naturelle ?

16. Existe-t-il une possibilité d'affiner le scénario ?

3.1.2.4. Résultats

Les experts considèrent que l'outil permet les échanges de commentaire et de documents d'une façon simple et claire. Il assiste efficacement les tâches désignées de communication, de catégorisation ou d'édition. Il est aussi intéressant puisqu'il permet de tracer les échanges. Le partage d'un document annoté est utile à la place des « faire suivre » par méls et permet une homogénéisation des documents du projet qui simplifie l'organisation.

Cependant un problème crucial en conception est de ne pas surcharger le concepteur. Multiplier l'utilisation d'outils différents et modifier les usages pose des problèmes. L'appropriation des outils pour l'utilisateur est souvent problématique.

De même, un problème se pose quant au partage de documents non numériques (les pièces effectivement produites !) ainsi que de documents en format propriétaire. En effet, nombre d'outils propriétaires permettent aux mécaniciens de travailler efficacement sur des plans qui sont des documents normalisés en mécanique et offre un moyen de communication formaté et relativement non ambigu.

Pour finir, les experts restent sceptiques sur l'existence d'un catalyseur possible. Il n'est pas évident de pousser les gens à communiquer ensemble ou via l'outil puisque cela repose sur des relations humaines, hors-projet.

3.1.3. L'utilité de l'outil ou utilité des fonctionnalités dans un usage d'ingénierie mécanique concurrente

La phase d'évaluation de l'utilité des fonctionnalités dans un usage précis consiste en l'observation et la critique des résultats obtenus sur la base du corpus en conception mécanique dans a projet Air Campus. Les experts observe tout d'abord la classification produite et la commente, puis répondent à un ensemble de questions ouvertes sur l'usage de l'outil dans un projet réel. Ces dernières questions permettent d'esquisser une évaluation grandeur réelle.

3.1.3.1. Validation de la classification

Les experts étant au fait du projet, ils sont à même de critiquer la classification obtenue automatiquement et de définir les améliorations à y apporter du point de vue catégorisation. La carte de thème leur est présentée sous l'angle de chacun des points de vue et les experts peuvent naviguer dans les termes issus des documents.

Il est ensuite demandé aux experts de dire si la classification pour chacun des points de vue est suffisamment fine, trop réduite, ou trop large. Les questions ayant trait aux différents points de vue de la carte de thèmes sont les suivantes :

17. Les points de vue pris en compte dans la classification sont-ils adaptés à l'activité ?

18 à 21. La classification pour le point de vue planification / argumentation / organisation / domaine, est-elle :

Suffisante Trop large Trop réduite

→ Une correction manuelle vous semble-t-elle possible ?

oui non

3.1.3.2. Résultats

Les résultats obtenus sont intéressants pour les point de vue domaine, planification et argumentation. Nous présentons donc ces trois points de vue dans cet ordre.

Point de vue *domaine*

Le point de vue de domaine est considéré par les experts comme beaucoup trop large et contenant beaucoup de bruit. Ce résultat était attendu puisque l'outil d'extraction est prévu pour être semi-automatique avec un nettoyage et une validation par l'utilisateur. D'un autre côté, les termes et concepts relatifs et utiles au domaine sont aussi présents.

Selon les experts, pour le domaine mécanique, un certain nombre de mots-clés est déjà défini et normalisé (norme de descripteur ISO), et serait suffisant pour une telle classification. Ces descripteurs sont principalement utilisés pour la documentation technique et sa récupération, c'est à dire a posteriori (par exemple par une personne qui souhaiterait se renseigner sur un sujet). Au cours de la conception, il semblerait que ce type de recherche de document ne soit pas de coutume.

Point de vue *planification*

La classification sous l'angle de la planification porte un intérêt clair dans un projet en conception, celui d'accéder à des documents produits au cours d'une certaine phase du projet. Cela permet à des concepteurs en mécanique d'accéder aux documents et aux échanges selon un processus, par un des jalons du projet. Les données techniques passées d'un acteur du projet aux autres et modifiées au cours de tâches spécifiques sont plus facilement accessibles au sein d'un compte-rendu, ou d'un document en général, modifié suite à des tâches. L'intérêt au point de vue *organisation* est du même ordre, lié à un des rôles d'un acteur du projet. L'organisation est alors importante à mettre en lien avec les phases de travail pour gérer la coopération autour des données.

La classification sous le point de vue planification est donc intéressante puisque les termes classiques représentant le processus d'un projet en mécanique (présents explicitement dans le corpus) ont été extraits mais aussi des termes qui sont liés à différents jalons et offrant un autre type de planification (financière par exemple).

Point de vue *argumentatif*

Les experts interrogés considèrent nécessaire la classification sous le point de vue argumentation pour la structuration des échanges et pour tracer les décisions. Ils considèrent la classification suffisante. Cependant, il faut noter que ce point de vue a été effectué à la main, selon l'analyse du corpus.

En ce qui concerne la classification en général, les experts considèrent qu'il est possible de la mettre à jour manuellement sans perdre de vue problème de surcharge de l'ingénieur. Il est fondamental de ne pas refaire deux fois la même chose pour le concepteur. L'assistance à l'indexation semble donc pertinente, même si l'utilisation de l'outil s'avèrerait sûrement lourde.

3.1.3.3. Validation de l'utilité de l'outil

D'une façon plus générale, nous avons interviewé les experts par un ensemble de questions ouvertes pour permettre de consolider les assertions précédentes et d'élargir les perspectives de développement.

Les questions ouvertes permettent aux experts d'effectuer nombre de commentaires et permettent à l'expérimentateur de rebondir sur ces commentaires. Les questions ayant trait à la validation de l'utilité de l'outil sont les suivantes :

22. Les fonctionnalités sont-elles suffisantes ?
23. Est-ce que l'outil correspond à vos besoins d'annotation ?
24. Est-ce que l'outil correspond à vos besoins de communication ?
25. Est-ce que l'outil correspond à vos besoins de classement de documents ?
26. Est-ce que l'outil correspond à vos besoins de partage de documents ?
27. Pouvez-vous évaluer des limites de l'outil a priori ?
28. Quelles fonctionnalités seraient selon vous possibles à ajouter ?
29. Êtes-vous satisfait par les possibilités de visualisation possible du fil d'annotation ?

Nous présentons en résultat une synthèse focalisée sur les enrichissements proposés par les experts au cours de la discussion.

3.1.3.4. Résultats

L'outil correspond aux besoins des concepteurs en mécanique. Cependant, AnT&CoW est plus générique que leur activité stricte en mécanique et ne prend pas en compte les spécificités formelles de leurs activités. En effet, de nombreux processus sont formalisés dans ce domaine et il serait intéressant d'ajouter par exemple dans la gestion des documents, les différents états de cet objet : créé, modifié (modifications), visé / validé

(version), archivé. Il est possible d'ajouter cela en index du document selon son lieu d'archivage sur le serveur d'AnT&CoW (brouillon, document, bibliothèque).

L'outil correspond à des besoins, mais ne correspond pas forcément à la distribution des rôles dans le projet. Même peu hiérarchisé, l'équipe est structurée et ce n'est pas le même participant au projet qui trace la logique de conception du projet et qui communique des informations par exemple. Il serait donc possible pour éviter certaines limites inhérentes aux fonctionnalités de diviser les trois fonctionnalités (communication, indexation, élaboration) en deux groupes selon les acteurs : La communication et l'indexation technique reviendrait avant tout au concepteur (au lieu d'un extérieur actuellement pour l'indexation – par des documentalistes professionnels-) et l'élaboration reviendrait au chef de projet pour la production de compte-rendus et de revues de projet d'une part, et la reconception d'autre part.

Au niveau de l'élaboration de document, les experts seraient très intéressés par l'assistance à la mise en forme possible du document. En effet, les documents types comptes-rendus ont une structure pré-établie qui pourrait être guidée lors de la prise de note par exemple. La perspective d'assistance à la rédaction est une thématique intéressante nécessitant une recherche spécifique.

Au niveau de la redondance des outils et donc pour éviter de noyer l'utilisateur dans l'utilisation d'un grand nombre d'outils, les experts aimeraient avoir possibilité de coupler l'outil de messagerie électronique à celui d'annotation (en plus du couplage de l'outil de rédaction à celui d'annotation). De nombreux outils permettent déjà de coupler l'annotation à une liste de diffusion par mél. Il serait intéressant d'évaluer les possibilités d'utilisation du plug-in AnT&CoW dans une messagerie électronique Open-Source (Firebird) par exemple.

4. Conclusion

Les trois scénarios présentés en premier lieu reprennent différentes formes de l'annotation. En jouant ces scénarios dans l'outil d'annotation AnT&CoW nous pouvons valider l'adéquation du modèle aux activités décrites et ainsi les fonctionnalités développées. Leur utilisabilité pour l'activité de conception est critiquée par nos experts en conception en mécanique.

Nous constatons en rejouant le corpus de méls dans l'outil que ce dernier facilite non seulement l'ancrage et la contextualisation des commentaires, mais aussi la mise en place de sommes partageables par l'ensemble des collaborateurs. Ces sommes sont créées sur la base des échanges sur des documents partagés, mais permettent une individuation du document créé. Chaque utilisateur peut créer la somme qu'il souhaite, selon le point de vue qu'il souhaite souligner dans son document. D'une façon théorique, l'outil semble bien répondre aux exigences posées et nous avons validé les scénarios et le modèle à travers le jeu de scénarios dans l'outil.

D'une façon pratique, l'utilité des fonctionnalités d'AnT&CoW pour le TCAO distant et asynchrones ou synchrones est validée par les experts interrogés. Les experts invalide cependant l'utilité de l'outil dans le contexte du projet associatif Air Campus. De même, dans l'état actuel des choses, l'usage réel et donc l'évaluation grandeur nature serait difficile. Ceci est principalement dû aux limites de l'outil en termes d'étendue des fonctionnalités. AnT&CoW ne peut pas remplacer les outils propriétaires pour l'édition de documents ni les outils type messagerie électronique ou gestion de l'agenda qui sont déjà entrés dans les habitudes des concepteurs. Pour y remédier, il faudrait permettre au plug-in AnT&CoW de fonctionner dans une messagerie OS ou dans des outils propriétaires (MSWord, Catia, Goëlan) afin de permettre l'accès à des documents partagés et de continuer de tracer la logique de communication au fil de l'eau.

Dans le cadre d'une conception itérative, nous utilisons ces remarques pour la détection d'inconsistances et de faiblesses. Les résultats de ces simulations nous permettent de déterminer les améliorations à apporter à l'outil et de cerner un certain nombre de pistes prospectives.

Conclusion et perspectives

*Cuius rei demonstrationem mirabilem sane detexi.
Hanc marginis exiguitas non caperet*
Pierre Simon de Fermat

1. Conclusion générale

Le propos développé dans ce mémoire est le soutien d'une herméneutique numérique. A ce titre, l'activité d'annotation est apparue comme fondamentale pour rendre possible la coopération autour de documents. Notre objectif était donc de concevoir un outil de coopération permettant d'annoter des documents numériques.

Au terme de notre parcours, (1) nous avons défini l'annotation pour le travail coopératif à la fois comme un objet qui relève de l'étiquette et du commentaire, et comme une activité qui relève de la communication, de l'indexation et de l'élaboration de discours. (2) La modélisation de l'activité instrumentée de production de discours nous a permis de définir les fonctionnalités d'un collecticiel. Enfin, (3) nous avons conçu et validé une maquette de ce collecticiel support à l'herméneutique numérique.

Définition de l'objet et de l'activité d'annotation à soutenir : Il s'agissait tout d'abord de définir l'objet manipulé au cours d'activités annotatives. En effet, l'activité d'annotation est productive d'objets annotation que nous avons définis dans le chapitre 1. Ces objets bien que multiples sont constitutifs d'une activité de lecture-écriture homogène. C'est pourquoi nous appréhendons l'objet annotation au sein d'un continuum. Ce dernier est inspiré (1) des travaux du Web Sémantique (qui utilisent une annotation pour l'étiquetage sémantique de documents), des travaux en herméneutique (qui envisagent une annotation pour l'échange discursif), des travaux en Communication Médiatisée par Ordinateur (qui utilisent des fragments textuels pour l'élaboration de conversations persistantes pouvant être recueillies sous forme de document numérique) et des travaux sur le document (qui définissent le rôle particulier des documents pour l'action, et le processus de documentarisation). Les annotations peuvent être à la fois des index pour la récupération de documents, des fragments de conversation pour planifier la coordination, et des fragments de document pour permettre l'élaboration de DoPAs. L'annotation est fondamentalement un objet pour la

communication. Elle permet donc à une communauté de co-construire un schème de classification afin de partager non seulement les documents mais aussi le principe de leur classification, ou une interprétation pour négocier les objectifs du projet et l'élaboration de documents. Elle est créée lors d'une activité de lecture-écriture que l'on peut modéliser afin de l'instrumenter. Le modèle de description de lecture-écriture adopté souligne la dimension interactionnelle et coopérative de l'activité d'annotation. En effet, il combine des phases d'interprétation individuelle et collective de documents et des phases d'élaboration de document en coopération. La coopération au cours des différentes phases de production de discours est fondée sur des échanges entre lecteurs/annotateurs.

Définition de fonctionnalités pour un collecticiel fondé sur le partage de documents : Il était ensuite question de définir un support à l'utilisateur sur la base de l'adaptation du modèle de description de l'activité de production de discours en un modèle de description de cette même activité médiatisée par un outil informatique. Ce modèle nous a permis de tracer les grandes lignes des fonctionnalités à mettre en place dans un collecticiel support à l'interprétation de documents numériques. Les utilisateurs doivent pouvoir partager des documents, échanger sur ces documents ou encore élaborer de nouveaux documents sur la base d'interprétations partagées. L'herméneutique numérique peut donc être soutenue par de l'annotation dans un but d'indexation de documents, de communication sur ces documents ou encore d'élaboration de nouveaux documents.

Développement du collecticiel AnT&CoW : L'implémentation des fonctionnalités de communication n'a pas donné lieu à des innovations particulières mais leur ajout était crucial. En revanche, les fonctionnalités d'élaboration et d'indexation sont liées à l'existence et donc la construction d'une classification par les utilisateurs pour mieux la comprendre et la partager. Afin qu'elle soit commune, la classification utilisée se doit d'être représentative du domaine ou du type d'organisation dans lequel les utilisateurs travaillent. L'utilisation d'un corpus permet de contextualiser l'activité et de refléter la langue spontanée utilisée dans la communauté de l'utilisateur au cours de son activité de conception.

Construction de la carte de thèmes : Le corpus recueilli auprès des utilisateurs est exploité par un module de Traitement Automatique des Langues. Il est constitué de l'outil Syntex et d'outils d'extraction de relations et de structuration en Topic Maps couplés à ses sorties. La récupération des termes pour en faire des descripteurs et leur structuration syntagmatique ne sont pas suffisants pour représenter les points de vue utilisateur dans la

TM. C'est pourquoi l'utilisateur doit être impliqué dans la construction de sa classification.

Le carte de thèmes résultant de cet effort de structuration est une classification centrée sur des termes organisés selon des ensembles conceptuels de plus haut niveau sémantique et social.

Scénarios d'utilisation de l'outil : L'outil développé est validé sur la base de scénarios. Les résultats nous indiquent que les fonctionnalités de l'outil sont adaptées à l'activité de partage et de rédaction collaborative de documents.

Les apports principaux de ce mémoire se situent sur le plan de l'originalité de notre approche. En effet, nous convoquons des théories de Sciences Humaines et Sociales afin de rapprocher deux univers souvent séparés (WS et SHS) pour définir un processus et un objet complexes ; l'annotation. De même, l'utilisation de corpus, permet de conserver l'utilisateur final au centre de nos préoccupations.

En outre, nous proposons un outil de coopération reposant sur l'emploi d'outils de TAL pour le soutien de la coopération à différents niveaux : soutien de la construction de classification et aide à l'indexation.

Cependant, selon le contexte d'utilisation de cet outil, des adaptations sont à apporter. Les limites et atouts de l'outil, soulignés par la validation des fonctionnalités, nous permettent d'esquisser respectivement des perspectives de développement et d'utilisation présentées ci-après.

2. Limites et perspectives

Les principales limites d'AnT&CoW se trouvent au niveau (1) des développements des fonctionnalités d'indexation, (2) de la structuration des documents élaborés coopérativement et (3) de la visualisation des documents en lien avec la classification. Chacune de ces limites ouvre sur des perspectives de développement pour l'optimisation de l'outil.

L'indexation est établie sur la base de la carte de thèmes comprenant des termes issus du document et des points de vue structurant les termes et les ressources qui y sont liées. L'indexation est soutenue par une proposition de descripteurs par un appariement de la carte de thèmes du corps de l'annotation. Cet appariement reste faible puisqu'il ne prend en compte que la flexion et ne met pas en correspondance des termes de la même famille. Il serait donc intéressant de coupler les sorties de Syntex avec un stemmatiseur simple et léger

(type Portman) afin de tester si la prise en compte de la dérivation améliore l'appariement sans ajouter de bruit.

La visualisation du fil d'annotation est aussi fondée sur la carte de thèmes. Cette visualisation permet la contextualisation des fragments de discours au cours de l'activité. Il serait intéressant d'enrichir la visualisation des annotations dans un document en permettant à l'utilisateur un suivi ponctuel, dans le plug-in, des annotations déposées sur un document selon un point de vue sélectionné. De plus, les documents de la bibliothèque ne sont pas organisés selon leurs descripteurs et il serait utile à l'utilisateur de ne pas avoir une liste chronologique de documents mais plutôt des répertoires structurés selon la carte de thèmes définie. Cette visualisation des ressources est permise par la structuration en TM qui se focalise sur l'indexation de la ressource.

L'annotation comme fragment de document est utilisée à des fins d'élaboration textuelle. Notre outil propose une sélection primaire des annotations à colliger. Cette sélection ne permet qu'une structuration de base du document en cours d'élaboration. Il n'existe pas de modèle sous-jacent de l'argumentation dans notre outil. Un tel modèle autoriserait une élaboration fine selon un plan d'argumentation. L'outil serait alors un artefact approprié de suivi de négociation et d'organisation de l'élaboration de documents. Dans ce cadre, il est possible d'adapter les travaux de résumé automatique (TAL) comme d'analyse lexicale, par exemple de l'argumentation (Charolles).

Les limites présentées ci-dessus sont issues de la simulation de fonctionnalités de l'outil dans le cadre de notre terrain en conception mécanique. En se détachant des limites liées à cette validation, nous pouvons ouvrir sur des perspectives d'utilisation de notre outil au sein de nouveaux terrains. Nous cernons alors des domaines qui utilisent des documents numériques et qui manquent de fonctionnalités d'indexation, de communication, et d'élaboration. Nous nous focalisons principalement sur le domaine des bibliothèques numériques et sur l'analyse de discours sur corpus numérique :

- Les bibliothèques traditionnelles sont des lieux de partage et d'échange d'information entre les différents adhérents. Avec la médiatisation des documents, les bibliothèques numériques prolifèrent et ont pour but de faciliter l'accès à l'information numérisée. L'amélioration de l'accès à l'information se fait majoritairement sur des bases quantitatives (avoir le plus d'informations possibles) (Lagoze *et al.*, 2005). Pourtant, il est possible d'améliorer les échanges de connaissances et les interactions sociales

(Bénel *et al.*, 2001) entre les membres de communautés Web, par un outil permettant l'échange lié à un document. Le plug-in développé au cours de nos travaux pourrait donc permettre d'annoter des documents et d'indexer des commentaires laissés à propos d'une des ressources de la bibliothèque numérique selon un schème de classification communautaire. Les lecteurs pourraient ainsi échanger des opinions, des conseils ou des idées novatrices (Ackerman et Halverson, 2004) et aider ainsi au choix des lectures.

- L'Analyse de Discours Médiatisée par Ordinateur nécessite le recueil d'un corpus de messages « numérisés ». Notre plug-in étant un outil de communication qui permet d'échanger via des annotations, il peut être utilisé afin de recueillir ses fragments de conversation. En effet, les annotations sont enregistrées et indexées sur le serveur d'annotation. Notre outil peut donc être utilisé pour l'observation de communautés par exemple dans une optique d'analyse des pratiques langagières d'une communauté et de l'évolution de leur sociolecte comme de leur schème de classification.

Des perspectives de recherche plus générales sont également envisageables. AnT&CoW est un outil destiné avant tout au travail coopératif médiatisé, et plus précisément à la conception coopérative. Il se situe donc au centre du processus de double itération de J. Bardram (1998), entre co-construction (focus sur l'objet) et coordination du travail (focus sur les conditions du travail). L'argumentation entre les concepteurs que nous avons observés, rendue possible par les annotations, permet à la fois la co-construction de concepts et d'artefacts et la coordination du travail par la planification du processus de conception. Il serait intéressant d'explicitier les mécanismes argumentatifs sous-jacents aux interactions soutenues par notre outil. En effet, l'étude de l'argumentation pour la coopération représente une tâche considérable mais majeure pour le soutien de la conception coopérative.

Bibliographie

- ⁴³Ackerman, M. S., et Halverson, C., (1999), Organizational Memory: Processes, Boundary Objects, and Trajectories, in *Proceedings of HICSS 99*, <http://www.eecs.umich.edu/%7Eackerm/pub/99b26/hicss99.pdf>
- Ackerman, M., et Halverson, C., (2004), *Sharing Expertise: The Next Step for Knowledge Management in Social Capital and Information Technology*, Huysman M. et Wulf V. (Ed.).
- Ait-Mokhtar, S., Chanod, J.-P., et Roux, C., (2001), A Multi-Input Dependency Parser, in *Proceedings of the Seventh IWPT (International Workshop on Parsing Technologies)*, Beijing, China, 17-19 October, p. 201-204.
- Amaya, (2003), <http://www.w3.org/Amaya/>
- Annotea, (2001), <http://www.w3.org/2001/Annotea/>
- Aristote, (- 400), *La rhétorique*, Livre 3, livre de poche : Paris, 1991.
- Atifi, H., Lejeune, C., Ninova, G., et Zacklad, M., (2006), Méthodologie transdisciplinaire de gestion du corpus pour les disciplines de l'interaction : recherche de principes directeurs, in Actes du XVII^e Colloque international d'Albi Langue et Signification (CALS - Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation), Albi, 10-14 juillet 2006. *Textol!* juin 2006, vol. XI, n^o2, p. 175-181.
- Aussenac-Gilles, N., et Condamines, A., (2004), Documents électroniques et constitution de ressources terminologiques ou ontologiques, in J. Charlet et J.-M. Salaun (Eds.) *Information-Interaction-Intelligence*, Volume 4, n^o1, p. 75-93.
- Bachimont, B., (2000), Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances. In *Ingénierie des connaissances, Evolutions récentes et nouveaux défis*, Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault (Eds.), Eyrolles, p. 305-323.
- Baget, J.-F., Canaud, E., Euzenat, J., et Saïd-Hacid, M., (2004), Les langages du Web sémantique, in Web Sémantique Revue Hors-Série 13, *Information - Interaction - Intelligence*, http://www.revue-i3.org/hors_serie/annee2004/index.html
- Baker, C., Fillmore, C., Lowe, J., (1998), The Berkeley FrameNet Project. In *Proceedings of COLING/ACL'98*, p. 86-90.
- Baker, M., (2000), The roles of models in Artificial Intelligence and Education Research: a prospective view, in *International Journal of Artificial Intelligence in Education Research*, Vol 11(2), p. 122-143.
- Balzac (de), H., (1831), *Le chef d'œuvre inconnu*, Folio, Ed.Gallimard, Paris, 1994.
- Bannon, L, et Kuutti, K., (1996), Shifting Perspectives on Organizational Memory: From Storage to Active Remembering, in Proceedings of the 29th IEEE HICSS, vol. III, *Information Systems - Collaboration Systems and Technology*. IEEE Computer Society Press, Washington 1996, p. 156-167.
- Barcellini, F., Détienne, F., Burkhardt, J.-M., et Sack, W., (2005), Thematic coherence and quotation practices in OSS design-oriented online discussions in *Proceedings of GROUP 2005*, p. 177-186.

⁴³ Les adresses Web des documents sont valides au 4/10/2006

- Bardram, J.-E., (1998), *Collaboration, Coordination, and Computer Support - An Activity Theoretical Approach to the Design of Computer Supported Cooperative Work*. Daimi PB 533, University of Aarhus, Aarhus.
- Bardram, J.-E., (2000), Scenario-Based Design of Cooperative Systems – Re-designing an Hospital Information System in Denmark, in *Group Decision and Negotiation*, 9, p. 237-250.
- Barras, C., Geoffrois, E., Wu, Z., et Liberman, M., (2000), Transcriber: development and use of a tool for assisting speech corpora production, in *Speech Communication special issue on Speech Annotation and Corpus Tools*, Vol 33, No 1-2, January 2000, p. 5-22.
- Barré de Miniac, C., (2000), *Le rapport à l'écriture : Aspects théoriques et didactiques*, coll. Savoirs mieux Ed. Septentrion Presses Universitaires, Ch. Barré de Miniac (Ed.).
- Bénel, A., (2003), *Consultation assistée par ordinateur de la documentation en Sciences Humaines : Considérations épistémologiques, solutions opératoires et applications à l'archéologie*, Thèse de doctorat en informatique, INSA de Lyon, décembre 2003. <http://www.porphiry.org/artcadhi/theses/abenel>.
- Bénel, A., (2006), Porphyry au pays des paestans : usages d'un outil d'analyse qualitative de documents par des étudiantes de maîtrise en iconographie grecque, in Actes du colloque international Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Albi, juillet 2006. in *Texto !*, juin 2006, vol. XI, n°2, p. 182-189, http://www.revue-texto.net/Archives/Albi_2006/Actes.html
- Bénel, A., Egyed-Zsigmond, E., Prié, Y., Calabretto S., Mille, A., Iacovella A., et Pinon, J.-M., (2001), Truth in the Digital Library: From Ontological to Hermeneutical Systems, in *Proceedings of the fifth ECDL*, Lecture Notes in Computer Science, Springer-Verlag Ed. p. 366-377.
- Bentley, R., Horstmann, T., Sikkel, K., et Trevor, J., (1995), Supporting Collaborative Information Sharing with the World Wide Web: The BSCW Shared Workspace System, in *The World Wide Web Journal: Proceedings of the 4th International WWW Conference*, Issue 1, December 1995, p. 63-74. O'Reilly.
- Bereiter, C. et Scardamalia, M., (1987), *The Psychology of Written Composition*, Hillsdale, NJ: Erlbaum.
- Berners-Lee, T., Hendler, J., et Lassila, O., (2001), *The Semantic Web*, Scientific American, 2001.
- Biezunski, M., Bryan, M., et Newcomb, S. R., (1999), *Topic Maps*, specification ISO/IEC 13250, 3 December 1999.
- Bird, S. et Liberman, M., (2001), Linguistic Annotation <http://www ldc.upenn.edu/annotation/>
- Boujut, J-F., et Blanco, E., (2003), Intermediary Objects as a Means to Foster Co-operation in Engineering Design in *Computer Supported Cooperative Work* 12(2), p. 205-219.
- Bourguin, G., (2000), *Un support informatique à l'activité coopérative fondé sur la Théorie de l'Activité : le projet DARE*, thèse de l'université de Lille I.
- Bourigault, D. et Jacquemin, C., (2000), Construction de ressources terminologiques, in *Ingénierie des langues*, Jean-Marie Pierrel (Ed.), Hermès, p. 215-233.
- Bourigault, D., (1996) LEXTER, a Natural Language Processing tool for terminology extraction, in *Proceedings of the 7th EURALEX International Congress*, Goteborg.
- Bourigault, D., (2002), UPERY : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus, in *Actes de la 9^e conférence annuelle sur le Traitement Automatique des Langues (TALN 2002)*, Nancy, 2002, p. 75-84.
- Bourigault, D., Aussenac-Gilles, N., et Charlet, J., (2004), *Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas*. In Pierrel, J.M.,

- Slodzian, M. (eds.), *Techniques informatiques et structuration de terminologies*, volume 18/1, Revue d'Intelligence Artificielle, Hermes Sciences.
- Bourigault, D., et Aussenac-Gilles, N., (2003), Construction d'ontologies à partir de textes, in *Actes de la 10^e conférence annuelle sur le Traitement Automatique des Langues (TALN 2003)*, Batz-sur-Mer, 2003, T2, p. 27-50.
- Bourigault, D., Fabre, C., Frérot, C., Jacques, M.-P. et Ozdowska, S. (2005), Syntex, analyseur syntaxique de corpus, in *Actes des 12^{èmes} journées sur le Traitement Automatique des Langues Naturelles*, Dourdan, France.
- Bowker, G. et Star, S. (2000) *Sorting things out: classification and its consequences*, Cambridge, Mass., MIT Press. http://www-personal.si.umich.edu/~rfrost/courses/MatCult/content/Sorting_Things.pdf
- Bresnan, M., Edelman, L., Newell, S., Scarbrough, H. et Swan, J., (2004), The Impact of Social Capital on Project-Based Learning. In: Huysman M. and Wulf V. (Eds.) *Social Capital and Information Technology*.
- Brickley, D., et Guha, R.V., (2004) RDF Vocabulary Description Language 1.0: RDF Schema-<http://www.w3.org/TR/rdf-schema/>.
- Brill, E., (1992), A simple rule-based part of speech tagger, in *Proceedings of the Third Conference on Applied Natural Language Processing, ACL*, Trento, Italie.
- Bringay, S. Barry, C. et Charlet, J., (2006), Annotations: A Functionality to support Cooperation, Coordination and Awareness in the Electronic Medical Record., in Hassanaly P., Herrmann T, Kunau G. and Zacklad M. (Eds.) *Proceedings of COOP'06*, Carry-le-Rouet, May 9th-12th 2006, p. 39-54.
- Bringay, S., (2006), *Les annotations pour supporter la collaboration dans le dossier patient électronique*, thèse de doctorat en informatique de l'université d'Amiens, 4 septembre 2006.
- Buckingham Shum, S., Slack, R., Daw, M., Juby, B., Rowley, A., Bachler, M., Mancini, C., Michaelides, D., Procter, R., De Roure, D., Chown, T., et Hewitt, T., (2006), Memetic: An Infrastructure for Meeting Memory, in *Proceedings of the 7th International Conference on the Design of Cooperative Systems*, Carry-le-Rouet, France, 9-12 May. http://www.memetic-vre.net/publications/COOP2006_Memetic.pdf
- Buckland, M., (1998a), What is a "digital document"? in *Document Numérique*, Paris, 2, no. 2, p. 221-230 <http://www.sims.berkeley.edu/~buckland/digdoc.html>
- Buckland, M., (1998b), What is a "document"? in Hahn, T. B. et Buckland M. (Eds.) *Historical Studies in Information Science*. Medford, NJ: Information Today, 1998, p. 215-220 <http://www.sims.berkeley.edu/~buckland/whatdoc.html>
- Buitelaar, P., (1998), CORELEX: Systematic Polysemy and Underspecification, Ph.D. Thesis, Brandeis University.
- Buitelaar, P., Olejnik, D., Hutanu, M., Schutz, A., Declerck, T., et Sintek, M., (2004), Towards Ontology Engineering Based on Linguistic Analysis, in *Proceedings of LREC'2004*, Lisbon, may 2004, ISBN 2-9517408-1-6, p. 7-11.
- Cabanac, G., Chevalier, M., Chrisment C., et Julien C., (2005), Annotation in cooperative work: from paper-based to the web one, in *International Workshop Annotation for Collaboration*, Paris (France), 24-25 Novembre 2005, p. 31-40.
- Cahier, J-P, (2005), *Ontologies sémiotiques pour le Web socio sémantique*, Thèse de doctorat en informatique de l'Université de technologie de Troyes, 14 décembre 2005.

- Capponi, N. et Toussaint, Y., (2000), Interprétation de classes de termes par généralisation de structures, in *Ingénierie des connaissances, Evolutions récentes et nouveaux défis*, Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault (Eds.), Eyrolles, p. 335-353.
- Carroll, J., (1999), Five reasons for scenario-based design, in *Proceedings of the 32nd Hawaii International Conference on System Sciences*.
- Carroll, J., (2000), *Making Use: Scenario-Based Design of Human-Computer Interactions*, Cambridge: MIT Press.
- Carroll, J., et Rosson, M. B., (1992), Getting around the task-artifact cycle: How to make claims and design by scenario, in *ACM Transactions on Information Systems*, 10(2), p. 181-212.
- Carroll, J., Neale, D., Isenhour, P., Rosson, M., et McCrickard, D., (2003), Notification and awareness: synchronizing task-oriented collaborative activity, in *International Journal of Human-Computer Studies*, v.58 n.5, p. 605-632.
- Carruthers, M., (1990), *The Book of Memory: A Study of Memory in Medieval Culture*, New York: Cambridge University Press.
- Caussanel, J., Cahier, J.P., Zacklad, M., et Charlet, J., (2002), Les Topic Maps sont-ils un bon candidat pour l'ingénierie du Web sémantique. In Bruno Bachimont (Ed.) *Actes des 6èmes Journées Ingénierie des Connaissances*, Rouen, France, 28–30 mai 2002, p. 233–52.
- Charaudeau, P., et Maingueneau, D., (2002), article Discours in *Dictionnaire d'analyse du discours*, Seuil.
- Charlet, J., (2002), *L'ingénierie des connaissances développements, résultats et perspectives pour la gestion des connaissances médicales*, Habilitation à diriger les recherches, Univ. P. et M. Curie, 10 décembre 2002, Paris. <http://estime.spim.jussieu.fr/~jc/Files/CharletHDR2002.pdf>
- Charlet, J., Reynaud, C. et Teulier, R., (2001), Ingénierie des connaissances pour les systèmes d'information. In Cauvet, C., et Rosenthal-Sabroux, C., (Eds.), *Ingénierie des systèmes d'information*, chp. 10. Paris : Hermès.
- Chateauraynaud, F., (2003), *Prospero - une technologie littéraire pour les Sciences Humaines*, CNRS Editions, Paris, 406 pages.
- Chibout K., Kotowicz J.-P., et Briffault X., (2003), Systèmes Collecticiels et Communication Naturelle : Apports Conjointes de l'Informatique Linguistique et de l'Ergonomie, in *Revue d'Interaction Homme-Machine*, Vol 4 - N°2/2003, Jean VIVIER (Ed.), Europa, Paris, 2003, 126 pages (ISSN 1289-2963) ; p. 87-119.
- Choi, Y., et Rasmussen, E., (2006), What do digital librarians do, in *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, p. 187-188.
- Cimiano, P., Hotho, A., et Staab, S., (2004), Clustering Concept Hierarchies from Text, in *Proceedings of LREC'2004*, Lisbon, may 2004, ISBN 2-9517408-1-6, p. 1721-1724.
- Clark, H., (1996), *Using Language*, Cambridge University Press, Cambridge.
- Clark, H., et Brennan, S. E., (1991), Grounding in communication. In L. B. Resnick, J. Levine et S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*, Washington, DC: American Psychological Association, p. 127-149.
- Condamines, A., (2005), *Sémantique et corpus*, Hermès, Londres.
- Cook, S., et Brown, J., (1999), Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organization Knowing, in *Organization Science*, Vol.10, n°4, 1999, p. 381-400.

- Couzinet, V., (2006), Les connaissances au regard des sciences de l'information et de la communication : sens et sujets dans l'inter-discipline, in Harzallah M., Charlet J., et Aussenac-Gilles N. (Eds.), *Actes de la conférence Ingénierie des Connaissances, IC 2006* Nantes, 28-30 Juin 2006, Conférence invitée.
- Daille, B., (1999), Identification des adjectifs relationnels en corpus, in *Actes de la Conférence de Traitement Automatique du Langage Naturel (TALN'99)*, Cargèse, 1999.
- Darses, F., (2001), Converger vers une solution en situation coopérative de conception : analyse cognitive du processus d'argumentation, In F. Darses (Ed.) *Modeling Cooperative Activities in Design Proceedings of the 10th Atelier du Travail Humain*, 27-28 juin 2001, Paris, France: INRIA.
- David, S., et Plante, P., (1990), De la nécessité d'une approche morpho-syntaxique dans l'analyse de textes, *Intelligence Artificielle et Sciences Cognitives au Québec*, 3(3), p. 140-154.
- De Libera, A., (2000), *La philosophie médiévale*, Paris, PUF (« Que sais-je ? » 1044), 4e éd.
- Decouchant, D.; Favela, J.; Martinez-Enriquez, A.M, (2001), PINAS: a middleware for Web distributed cooperative authoring, in *Proc. SAINT'2001, The 2001 Symposium on Applications and the Internet, IEEE Computer Society and IPJ Information Processing Society of Japan*, San Diego, CA (U.S.A), p. 187-194
- Delacroix, J., (2005), *Les Wikis*, M2 éditions.
- Denoue, L., et Vignollet, L., (2000), An annotation tool for Web browsers and its applications to information retrieval, in *Proceedings of RIAO 2000*.
- Derrida, J., (1972), *Positions*, Éditions de Minuit, Paris.
- Derrida, J., (1991), This is not an oral footnote, in Stephen Barney (Ed.) *Annotation and Its Texts*, New York, Oxford.
- Détrie, C., Siblot, P., et Verine, B., (2001), Termes et concepts pour l'analyse du discours, *Pour une approche praxématique*, Paris : Champion.
- Dillenbourg, P., et Baker, M., (1996), Negotiation Spaces in Human-Computer Collaborative Learning, In *Proceedings of the International Conference on Cooperative Systems (COOP'96)*, France, p. 187-206.
- Dingli, A., (2003), Next Generation Annotation Interfaces for Adaptive Information Extraction. In *6th Annual Computer Linguists UK Colloquium (CLUK 03)*, January, 2003, Edinburgh, UK.
- Domingue, J.B., Lanzoni, M., Motta, E., Vargas-Vera, M., et Ciravegna, F., (2002), Mnm: Ontology driven semi-automatic or automatic support for semantic markup. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October.
- Dourish, P., et Belotti, V., (1992), Awareness and coordination in shared workspaces, in *Proceedings Computer Supported Cooperative Work 92*, New York: Association for Computing Machinery, p. 107-115.
- Dubois, J., (1969), Lexicologie et analyse d'énoncé, *Cahiers de lexicologie* 2, p. 115-126.
- Dzbor, M., Motta, E., Domingue, J. B., (2004), Opening Up Magpie via Semantic Services, In *3rd ISWC*, November 2004, Japan.
- Ellis, C. A., Gibbs, S. J., et Rein, G. L., (1991), Groupware: Some Issues and Experiences, in *Communications of the ACM*, 1991, p. 39-58.
- Erickson, T., (1999), Persistent Conversation: Discourse as Document, in the *Proceedings of the Thirty-Second Hawaii International Conference on Systems Science*, J. F. Nunamaker, Jr. R. H. Sprague, (Eds.), January, 1999. Republished as "Persistent Conversation: An Introduction" in the

Journal of Computer Mediated Communication, Vol. 4, June 1999,
<http://jcmc.indiana.edu/vol4/issue4/ericksonintro.html>

- Evrard, F., et Virbel, J., (1996), *Projet de réalisation d'un prototype de station de lecture active et utilisation en milieu professionnel*, rapport final d'activité du contrat état-région n°9300571, Toulouse.
- Falzon, P., (1989), *Ergonomie cognitive du dialogue*, Presses Universitaires de Grenoble.
- Falzon, P., (1991), Les activités verbales dans le travail, in R. Amalberti, M. De Montmollin et J. Theureau, (Ed.), *Modèles en analyse du travail*, Mardaga, Liège.
- Fayol, M., (1997), *Des idées au texte : psychologie cognitive de la production verbale, orale et écrite*, Paris : PUF.
- Fogli, D., Fresta, G., Mussio, P., Marcante, A., et Padula, M., (2005), Annotation in cooperative work: from paper-based to the web one, in *International Workshop Annotation for Collaboration*, Paris (France), Nov. 2005, p. 1-10.
- Folch, H., et Habert, B., (2005), Langages de métadonnées pour Web(s) sémantique(s), in Ihadjadène M. (Ed.) *Méthodes avancées pour les systèmes de recherche d'informations*, Traité des sciences et techniques de l'information, Hermès, chp. 3 p. 55-71.
- Frath, P., Oueslati, R., et Rousselot, F., (2000), Identification de relations sémantiques par repérage et analyse de cooccurrences de signes linguistiques, in J. Charlet, M. Zacklad, G. Kassel, D. Bourigault (Eds.), *Ingénierie des connaissances. Évolutions récentes et nouveaux défis*. Eyrolles, Paris, p. 291-304.
- Fuchs, C., (1982), Éléments pour une approche énonciative de la paraphrase dans les brouillons de manuscrits in *La genèse du texte : les modèles linguistiques collection textes et manuscrits* CNRS (Ed.), p. 73 - 102.
- Gadamer, H-G., (1960), *Vérité et méthode*, Les grandes lignes d'une herméneutique philosophique ; trad. Fruchon, Grondin, Merlo, Seuil, 1996 ; Vol.1 des *Gesammelt Werke*, Mohr, Tübingen, 1986.
- Gangemi, A., Guarino, N., Masolo, C., et Oltramari A., (2003), Sweetening WordNet with DOLCE, i *AI Magazine* 24(3): Fall 2003, p. 13-24.
- Garlatti, S., et Iksal, S., (2000), Méthodologie de conception de documents électroniques adaptatifs sur le Web. in Gaio, M., Trupincide, E., (Eds.) *Document Électronique Dynamique, Actes du troisième colloque international sur le document électronique : CIDE'2000*, .
- Garshol, L.M., (2003), *Living with topic maps and RDF*, <http://www.ontopia.net/topicmaps/materials/tmrdf.html>
- Genette, G., (1982), *Palimpsestes*, Paris, Seuil (Ed.).
- Genette, G., (1987), *Seuils*, Paris, Seuil (Ed.).
- Gil, Y., et Ratnakar, V., (2002), Trusting Information Sources One Citizen at a Time, in *Proceedings of the First International Semantic Web Conference (ISWC)*, Sardinia, Italy, June 9 -12.
- Go, K., et Carroll, J. M., (2004), The Blind Men and the Elephant: Views of Scenario-Based System Design, *Interactions*.
- Goble, C., Bechhofer, S., Carr, L., De Roure, D., et Hall, W., (2001), Conceptual Open Hypermedia = The Semantic Web?, in *SemWeb2001, The Second International Workshop on the Semantic Web*, Hong Kong, May 2001.

- Gómez-Pérez, A., Manzano-Macho, D., Alfonseca, E., Núñez, R., Blacoe, I., Staab, S., Corcho, O., Ding, Y., Paralic, J., et Troncy, R., (2003), *Ontoweb, Deliverable 1.5: A survey of ontology learning methods and techniques*, IST Project IST-2000-29243 OntoWeb, juin 2003.
- Goody, J., (1987), *The Interface Between the Written and the Oral*, Cambridge: Cambridge University Press.
- Greif, I., (1988), *Computer-Supported Cooperative Work: A Book of Readings*, Morgan Kaufman (Ed.), San Mateo, California.
- Gruber, T., (1993), A translation approach to portable ontologies, in *Knowledge Acquisition*, 5(2), p. 199-220.
- Gruber, T., (2005), Folksonomy of Ontology: A Mash-up of Apples and Oranges. In *First on-Line conference on Metadata and Semantics Research (MTR'05)*, <http://mtr.sigsemis.org/>; <http://tomgruber.org/writing/mtr05-ontology-of-folksonomy.htm>
- Gruber, T., et Lytras, M.D., (2004), Interview for the *AIS Special Interest Group on Semantic Web and Information Systems*, Volume 1, Issue 3, 2004.
- Guibert, S., Darses, F., et Boujut, J.F., (2005), Rôle des annotations dans l'élaboration collaborative d'une solution en conception mécanique, in Actes de *EPIQUE'2005*.
- Guilhaumou, J., (2002), Le corpus en analyse de discours : perspective historique, in Corpus Numéro 1, *Corpus et recherches linguistiques*, nov. 2002, <http://revel.unice.fr/corpus/document.html?id=8>
- Habert, B., (2000), Des corpus représentatifs : de quoi, pour quoi, comment ?, In Bilger M. (éd.), *Linguistique sur corpus. Études et réflexions*, Presses Universitaires de Perpignan, Perpignan, p. 11-58.
- Habert, B., et Fabre, C., (1999), Elementary dependency trees for identifying corpus-specific semantic classes, in *Computer and the Humanities*, 33(3), p. 207-219.
- Habert, B., et Nazarenko, A., (1996), La syntaxe comme marche-pied de l'acquisition de connaissances : bilan critique d'une expérience, in *Actes des Journées d'Acquisition de Connaissances (JAC)*.
- Habert, B., Nazarenko, A., et Salem, A., (1997), *Les linguistiques de corpus*, Paris : Armand Colin et Masson (Ed.).
- Handsuh, S., et Staab, S., (2003), Annotating of the Shallow and the Deep Web, in *Annotation for the semantic web*, S. Handsuh, S. Staab, IOS Press p. 25-45.
- Handsuh, S., Staab S., et Ciravegna, F., (2002), S-cream - semi-automatic creation of metadata. In *13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, October 2002.
- Haroche-Bouzinac, G., (2000), Une esthétique de la brièveté, in *Revue de l'AIRE – Recherches sur l'épistolaire*, n° 25-26, p. 49-51.
- Harris, Z., (1952), Discourse analysis, *Language* 28, Traduit dans *Langages* 13, 1969, p. 8-45.
- Harris, Z., (1988), *Language and Information*, Columbia University Press, New York.
- Hayes, J. R., (1995), Un nouveau modèle du processus d'écriture, in J. Boyer, J-P Dionne et P. Raymond (Eds.), *La production des textes, Vers un modèle d'enseignement de l'écriture*, Montréal : Les éditions logiques, p. 49-72.
- Hayes, J. R., et Flower, L. S., (1980), Identifying the organization of writing processes. In L. W. Gregg et E. R. Steinberg (Eds.), *Cognitive processes in writing*, Hillsdale, NJ: Lawrence Erlbaum.

- Haynes, S., Purao, S., et Skattebo, A., (2004), Situating Evaluation in Scenarii of Use, in *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW 2004*, Chicago, November 06-10, p. 92-101.
- Hearst, M. A., (1992), Automatic acquisition of Hyponyms from large text corpora, in *Proceedings of the Fourteenth International Conference on Computational Linguistic, (CoLing'92)*, Nantes, France, July 1992, p. 539-545.
- Heflin, J., et Hendler, J., (2000), Searching the web with shoe, in *Artificial Intelligence for Web Search, papers from the AAAI Workshop*, p. 35-40.
<http://www.cs.umd.edu/projects/plus/SHOE/pubs/#aiweb00>
- Herring, S., (1999), Interactional Coherence in CMC, in *Journal of Computer-Mediated Communication* 4(4). www.ascusc.org/jcmc/vol4/issue4/
- Herring, S., (2000), Linguistic Approaches to Computer-Mediated Communication, *Jornada sobre Comunicació Mediatitzada per Ordinador CMO-Cat*, Universitat de Barcelona, <http://www.ub.es/lincat/cmo-cat>
- Herrmann, T., Jahnke, I., et Loser, K.U., (2004), Designing Community Systems based on Role-Taking and Role-Making, in Darses F., Dieng R., Simone C., Zacklad M., *Proceedings of COOP 2004*, French Riviera, France.
- Ihadjadène, M., (2005), Introduction in Ihadjadène M. (Ed.), *Les systèmes de recherche d'informations*, Traité des sciences et techniques de l'information, Hermès, p. 14-17.
- Jacquemin, C., (1997), *Variation terminologique : Reconnaissance et acquisition automatiques de termes et de leurs variantes en corpus*, Mémoire d'habilitation à diriger des recherches en informatique fondamentale, Université de Nantes.
- Jacquemin, C., et Bourigault, D., (2003), Term Extraction and Automatic Indexing, in Mitkov R. (Ed.), *The Oxford Handbook of Computational Linguistics*, Oxford University Press, p. 599-615.
- Jacquemin, C., et Tzoukermann, E., (1999), NLP for Term Variant Extraction: A Synergy of Morphology, Lexicon and Syntax. In T. Strzalkowski, editor, *Natural Language Information Retrieval*, Kluwer, Boston, MA, 1999, p. 25-74.
- Jauss, H.R., (1988), *Pour une herméneutique littéraire*, Paris, Gallimard (Ed.).
- Jeanneret, Y., (2000), *Y a-t-il vraiment des technologies de l'information ?* Villeneuve d'Ascq : Presses universitaires du Septentrion.
- Joachims, T., (2002), *Learning to Classify Text using Support Vector Machines*, Kluwer Academic Publishers 224 pages.
- Johansen, R., (1988), Current user approaches to groupware, in *Groupware: Computer support for business teams*, R. Johansen (Ed.), The Free Press, New York, p. 12-44.
- Jordan, K., Hauser, J., et Foster, S., (2003), The augmented Social Network: Building identity and trust into the next-generation Internet, in *Firstmonday* vol.8, N.8, august 4th, http://www.firstmonday.org/issues/issue8_8/jordan/
- Kahan, J., Koivunen, M.-R., Prud'Hommeaux, E., et Swick, R.R., (2001), Annotea: an open RDF Infrastructure for Shared Web Annotations, in *Proceedings of WWW10*, May 1-5 2001, Hong-Kong, p. 623-632.
- Kahn, H., (1962), *Thinking about the unthinkable*, New York, Horizon Press (Ed).
- Ka-Ping, Y., (1998), *CritLink: Better hyperlinks for the WWW*, <http://zesty.ca/crit/ht98.html>

- Kintsch, W., (1988), The role of knowledge in discourse comprehension: A Construction-Integration model, in *Psychological Review*, 95, p. 163-182.
- Kintsch, W., et Van Dijk, T. A., (1978), Toward a model of text comprehension and production, in *Psychological Review*, 85, p. 363-394.
- Koivunen, M. J., et Swick, R., (2003), Collaboration through annotations in the Semantic Web, in *Annotation for the Semantic Web*, S. Handschuh, S. Staab (Eds.), IOS Press, p. 46-60.
- Koivunen, M. J., Kahan, J., Swick, R., Prud'hommeaux, E., Shiraishi, N., et al., (2001), *Annotea*, <http://www.w3.org/2001/Annotea/>
- Labbe, H., et Marcoccia, M., (2005), Communication numérique et continuité des genres : l'exemple du courrier électronique, *Texte !*
- Lafont, R., (1989), Le dire et le faire, recueil d'articles édité par le groupe de recherches en praxématique de l'Université Paul Valéry, Montpellier, *Praxiling*, U.P.V.
- Lagoze, C., Arms, W., Gan, S., Hillmann, D., Ingram, C., Krafft, D., Marisa, R., Phipps, J., Saylor, J., et Terrizzi, C., (2002), Core services in the architecture of the National Digital Library for science education NSDL, In *Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries*, Lagoze, C. (Ed.), Portland.
- Lagoze, C., Krafft, D.B., Payette, S., et Jesuroga, S., (2005), What Is a Digital Library Anymore, Anyway? Beyond Search and Access in the NSDL in *D-Lib Magazine*, Nov. 2005, Vol. 11 N°11, ISSN 1082-9873 <http://www.dlib.org/dlib/november05/lagoze/11lagoze.html>
- Lassila, O., et McGuinness, D., (2001), The Role of Frame-Based Representation on the Semantic Web, in Linköping Electronic Articles in *Computer and Information Science*, Issue: Vol. 6 (2001) : 005, <http://www.ida.liu.se/ext/epa/ej/etai/2001/018/01018-etaibody.pdf>
- Lassila, O., et Swick, R., (1999), *Resource Description Framework (RDF) Model and syntax specification, Recommendation, W3C* <http://www.w3.org/TR/REC-rdf-syntax>
- Latteier, A., Pelletier, M., McDonough, C., et Sabaini, P., (2003), *The Zope Book*, Edition 2.6. http://zope.org/Documentation/Books/ZopeBook/2_6Edition/ZopeBook-2_6.pdf
- Laublet, P., Charlet J., et Reynaud C., (2004), Introduction au web sémantique, in *Web Sémantique Revue Hors-Série I3, Information - Interaction - Intelligence*, http://www.revue-i3.org/hors_serie/annee2004/index.html
- Le Moigne, J.-L., (2004), La Modélisation est désormais notre Mot Clé, in *Lettre et InterLettre Chemin Faisant MCX-APC*, nov.2004. <http://www.mcxapc.org/docs/reperes/edil26.pdf>
- Lee, C., (2005), Between Chaos and Routine: Boundary Negotiating Artifacts in Collaboration, in *Proceedings of the Ninth European Conference on Computer Supported Cooperative Work*, p. 387-406.
- Leech, G., (1997), Introduction corpus annotation, in Garside R., Leech G., McEnery A., (Eds.), *Corpus annotation: Linguistic information from computer text corpora*, London: Longman.
- Lejeune C., (2007), Petite histoire des ressources logicielles au service de la sociologie qualitative, in Brossaud C., Reber B. (Eds.) *Les sciences humaines sociales à l'heure des technologies de l'information et de la communication*, Hermès Sciences, Publications, à paraître.
- Leroux, E., (1999), *Analyse linguistique d'échanges de messages électroniques dans un système de communication d'entreprise*, LIMSI-CNRS, Orsay.
- Levelt, W. J. M., (1989), *Speaking: From intention to articulation*, Cambridge, Boston: MIT Press.

- Lewkowicz, M., et Marcoccia, M., (2004), The Participative Framework as a design model for newsgroups: PartRoOM in *proceedings of COOP 2004*, p. 243-256.
- Lewkowicz, M., Soulier, E., et Gauducheau, N., (2006), Collecticiels pour la construction collective du sens – définition, principes de conception, exemple, in Soulier E. (Ed.), *Le storytelling : concepts, outils et applications*, Traité IC2, Hermès, Paris, p. 293-314.
- Ljunggren, C., (2003), *JANNO*, <http://www.claus-ljunggren.dk/janno/web/>
- Maedche, A, et Staab, S. (2001), Ontology Learning for the Semantic Web, in *IEEE Intelligent Systems, Special Issue on the Semantic Web*.
- Manning, C.D., et Schütze, H., (1999), *Foundations of statistical natural language processing*, Cambridge, MA: MIT Press.
- Marcellesi, J.-B., et Gardin B. (1974), Introduction à la sociolinguistique : la linguistique sociale, Paris, Larousse, coll. Langue et langage.
- Marcoccia, M., (2001), L'animation d'un espace numérique de discussion : l'exemple des forums usenet, in *Document Numérique*, Vol. 5, n°3-4, p. 11-26.
- Marcoccia, M., (2004), On-line polylogues: conversation structure and participation framework in Internet newsgroups, in *Journals of Pragmatics*, 36 (2004), p. 115-145.
- Marshall, C., (1997), Annotation: from paper books to the digital library. In *Proc. Digital Libraries '97*, New York: ACM Press, p. 131-140.
- Marshall, C., (1998a), The Future of Annotation in a Digital (Paper) World in *proceedings 35th Annual GSLIS Clinic: Successes and Failures of Digital Libraries*, March 24.
- Marshall, C., (1998b), Toward an ecology of hypertext annotation. In *Proc. Hypertext '98*, New York: ACM Press, p. 40-49.
- Martin, G., Détienne, F., et Lavigne, E., (2001), Analysing viewpoints in design through the argumentation process, in *proceedings of INTERACT'2001*, Tokyo, Japan, July p. 9-13.
- Martin, P., (2000), WinPitch 2000: a tool for experimental phonology and intonation research, in *Proceedings of the Prosody 2000 Workshop*, Kraków, Pologne, 2-5 October 2000.
- Mathes, A., (2004), <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Mayaffre, D., (2006), Rôle et place des corpus en linguistique : réflexions introductives, in *Texto !* http://www.revue-texto.net/Corpus/Publications/Mayaffre_Corpus.html
- Mayaffre, D., L'herméneutique numérique, in *L'Astrolabe, Recherche littéraire et Informatique*, nov. 2002 <http://www.uottawa.ca/academic/arts/astrolabe/articles/art0031.htm/Hermeneutique.htm>
- McCray, A.T., et Gallagher, M.E., (2001), Principles for digital library development, *Commun*, ACM 44, p. 48-54.
- McFarlane, N., (2003), *Rapid Application Development with Mozilla*, in Bruce Perens' Open Source Series (Ed.), December 2003.
- McGuinness, D.L., et Van Harmelen, F., (2004), *OWL Web Ontology Language Overview*, W3C Recommendation 10 February 2004, <http://www.w3.org/TR/2004/REC-owl-features-20040210/>
- Merholz, P., (2001), *Vernacular Thesauri*, Posted on 11/10/2001 on <http://www.peterme.com/archives/00000090.html>
- Miles-Board, T., (2001), *An Annotative Approach To Better Hyperauthoring And Associative Linking*, <http://citeseer.ist.psu.edu/miles-board01annotative.html>

- Narayanan, S., et Harabagiu, S. (2004) *Question answering based on semantic structures*. In *COLING 2004*.
- Nardi, B., (1992), The Use of Scenarii in Design, in *SIGCHI Bulletin* volume 24, Number 4.
- Nardi, B., (1995), Some reflections on scenario, in Carroll, J. M. (Ed.), *Scenario Based Design: Envisioning Work and Technology in System Development*, New York: Wiley & Sons, p. 387-399.
- Nardi, B., Schiano, D., et Gumbrecht, M., Blogging As Social Activity, or, Would You Let 900 Million People Read Your Diary? in *Proceedings of CSCW 2004* (Chicago, November 2004), ACM Press, p. 222-231. <http://home.comcast.net/~diane.schiano/CSCW04.Blog.pdf>
- Niles, I., et Pease, A., (2001), *Towards a Standard Upper Ontology*. In Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Chris Welty and Barry Smith, eds, Ogunquit, Maine, October 17-19, <http://projects.teknowledge.com/HPKB/Publications/FOIS.pdf>
- Nystrand, M., (1986), *The structure of written communication: studies in reciprocity between writers and readers*, Toronto: Academic Press.
- Nystrand, M., (1989), A social-interactive model of writing, in *Written Communication*, 6, p. 65-85.
- OWL, (2004), <http://www.w3.org/2004/OWL/>
- Papert, S., (1993), *Mindstorms: Children, Computers and Powerful Ideas*.
- Pédauque, R.T., (2003), *Document : forme, signe et médium, les reformulations du numérique*, version 3- 8 juillet 2003, http://archivesic.ccsd.cnrs.fr/sic_00000511
- Pédauque, R.T., (2006), *Document et modernités*, 16 mars 2006 – (Document and modernities), working paper, http://archivesic.ccsd.cnrs.fr/sic_00001741
- Pepper, S., et Moore, G., (2001), XML Topic Maps (XTM) 1.0, *TopicMaps.Org Specification*, <http://www.topicmaps.org/xtm/>
- Pery-Woodley, M-P., (2005), Discours, corpus, traitement automatiques, in Condamines A. (Ed.), *Sémantique et corpus*, Hermès, Londres 2005.
- Pfeffer, J., et Sutton, R., (1999), *Knowing "what" to do is not enough: Turning Knowledge into Action*, California Management Review, vol.42, n°1, Fall, 1999, p. 83-108.
- Piolat, A., (2004), Approche cognitive de l'activité rédactionnelle et de son acquisition, Le rôle de la mémoire de travail, *LINX (Linguistique Institut Nanterre Paris X)*, 51, p. 55-74.
- Piolat, A., Farioli, F., et Roussey, J.-Y., (1989), La production de texte assistée par ordinateur, in G. Monteil, et M. Fayol (Eds.), *La psychologie scientifique et ses applications*, Grenoble : Presses Universitaires de Grenoble, p. 177-184.
- Plantin, C., (2005), Pour une archive des langues parlées en interaction, Statuts juridiques, formats et standards, représentativité, in J.L. Lebrave, *La société de l'information et ses enjeux, actes du colloque de bilan la société de l'information 2001-2005*, ENS-LSH, Lyon, 2005.
- Price, M., Schilit, B., et Golovchinsky, G., (1998), XLibris: The active reading machine, in *proceedings of CHI'98 Human factors in computing systems*, Los Angeles, California, USA, vol.2 of Demonstrations: Dynamic Documents, p. 22-23.
- Prié, Y., Garlatti, S., (2004), Métadonnées et annotations dans le Web sémantique, in *Web Sémantique Revue Hors-Série I3, Information - Interaction - Intelligence*, http://www.revue-i3.org/hors_serie/annee2004/index.html

- Prilla, M., et Ritterskamp, C., (2006), Collaboration Support by Co-ownership of Documents, in P. Hassanaly, T. Herrmann, G. Kunau and M. Zacklad (Eds.), in *Proceedings of COOP'06*, Cooperative Systems Design, IOS press, p. 255-269.
- Ramage, M., (1999), *The Learning Way: Evaluating Cooperative Systems*, Unpublished PhD thesis, University of Lancaster.
<http://www.comp.lancs.ac.uk/computing/research/cseg/projects/evaluation/>
- Rastier, F., (1989), *Sens et textualité*, Paris, Hachette Université.
- Rastier, F., (2005), Enjeux épistémologiques de la linguistique de corpus, in G. Williams (Ed.), *La linguistique de corpus*, Rennes, PUR, p. 31-45. http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html
- Reed, D., Le Maitre, M., Barnhill, B., Davis, O., et Labalme, F., (2004), The Social Web: Creating An Open Social Network with XDI in *Planet Network Journal*
<http://journal.planetwork.net/article.php?lab=reed0704>
- Reinert, M., (1986), Présentation du logiciel Alceste à l'aide d'un exemple, in *Psychologie et Éducation* (X/2), p. 58-72.
- Reymonet, A., Aussenac-Gilles, N., et Thomas, J., (2006), Tâche, domaine et application : influences sur le processus de modélisation de connaissances, in Harzallah M., Charlet J., et Aussenac-Gilles N. (Eds.), *Actes de la conférence Ingénierie des Connaissances, IC 2006* Nantes, 28-30 Juin 2006, p.
- Ricoeur, P., (1970), Qu'est-ce qu'un texte? Expliquer et Comprendre, in BUBNER, CRAMER & WIEHL (hrsg.), *Hermeneutik und Dialektik*, Tübingen: Mohr, 1970, p. 181-200
- Roche, M., Heitz, T., Matte-Tailliez, O., et Kodratoff, Y., (2004), EXIT: Un système itératif pour l'extraction de la terminologie du domaine à partir de corpus spécialisés, *Proceedings of JADT'04 (International Conference on Statistical Analysis of Textual Data)*, volume 2, p. 946-956.
- Rogers, Y., (2004), New Theoretical Approaches for HCI, in *ARIST: Annual Review of Information Science and Technology*, no 38, 2004.
- Rommetveit, R., (1974), *On message structure: A framework for the study of language and communication*, London: John Wiley.
- Röscheisen, M., Morgensen, C., et Winograd, T., (1994), *Shared Web Annotations as a Platform for Third-Party Value-Added Information Providers: Architecture, Protocols, and Usage Examples*,
<http://citeseer.ist.psu.edu/roscheisen94shared.html>
- Rouet, J-F., (2001), *Les Activités Documentaires Complexes - Aspects Cognitifs et Développementaux* -, HDR, p. 76-104.
- Rousselot, F., et Montessuit, N., (2003), La station de travail LIKES, INTEX Workshop Bulgarie.
- Roussey C., Calabretto, S., et Pinon J.-M., (2002), Le thésaurus sémantique : contribution à l'ingénierie des connaissances documentaires, in Bachimont (Ed.), *Actes des 6èmes Journées Ingénierie des Connaissances*, Rouen, France, p. 209-20.
- Roussey, C., Calabretto, S., et Pinon, J.-M., (2001), SyDoM: A Multilingual Information Retrieval System for Digital Libraries, in *proc. International Conference ICC/IFIP On Electronic Publishing (ELPUB'2001)*, Canterbury (UK), 5-7 july 2001, p. 150-164.
- Salem, A., (1986), Segments répétés et analyse statistique des données textuelles, *Histoire & Mesure* 1-2, p. 5-28.
- Salembier, P., (2004), *Atelier PRESCOT - Annotation 2004* - <http://www.irit.fr/Annotation-PRESCOT/>, 2004

- Salton, G., et Buckley, C., (1988), Weighting Approaches in Automatic Text Retrieval, in *Information Processing and Management*, n° 24 (5), p. 513–523.
- Salton, G., et McGill, M., (1983), *Introduction to Modern Information Retrieval*, McGraw-Hill.
- Schachter, J., (2004), *Del.icio.us About Page*, <http://del.icio.us/doc/about>
- Schilit, B., et Golovchinsky, G., et Price, M., (1998), Beyond Paper: Supporting Active Reading with Free Form Digital Ink Annotations, in *Proc. CHI98*, New York: ACM Press, p. 249-256.
- Schmid, H., (1994), Probabilistic part-of-speech tagging using decision trees, in *Proceedings of the International Conference on New methods in Language Processing*, Manchester, UK.
- Schmidt, K., et Bannon, L., (1992), Taking CSCW Seriously, in *Computer Supported Cooperative Work Journal*, Kluwer Academic Publishers Volume 1, Numbers 1-2 March 1992 <http://www.idi.ntnu.no/emner/dif8914/kompndium-2004/papers-2004/c5-schmidt92.pdf>
- Schön D., et Bennet, J., (1996), Reflective Conversation with Materials in *Bringing Design to Software*, Winograd T. (Ed.), Stanford University and Interval Research Corporation Addison-Wesley. 1996 chp. 9 <http://hci.stanford.edu/bds/9-schon.html>
- Séguéla, P., et Aussenac-Gilles, N., (1999), Extraction de relations sémantiques entre termes et enrichissement de modèles du domaine, in *Actes de IC'99 (Ingénierie des Connaissances)*, Paris, p. 79-88.
- Shirky, C., (2003), Social Software and the Politics of Groups March 9, 2003 on the *Networks, Economics, and Culture mailing list*.
- Shirky, C., (2005), *Ontology is Overrated: Categories, Links, and Tags* http://www.shirky.com/writings/ontology_overrated.html
- Simone, C., et Sarini, M., (2001), Adaptability of Classification Schemes in Cooperation: What does it mean?, in *Proceedings of ECSCW 2001*, p. 19-38.
- Sinclair, J., (1991), Corpus Creation, in *Corpus, Concordance, Collocation*, Chp. 1, Série : Describing English Language. J. Sinclair et R. Carter (Eds.) Oxford : Oxford University Press.
- Sinclair, J., (1996), *Preliminary recommendations on Text Typology, EAGLES (Expert Advisory Group on Language Engineering Standards)*, <http://www.ilc.cnr.it/EAGLES/browse.html#wg1>, p. 4
- Slodzian, M., (2006), La terminologie, historique et orientations, in Harzallah, M., Charlet, J., et Aussenac-Gilles, N. (Eds.), *Actes de la semaine de la Connaissance (SDC'06)*, Conférence Invitée, Nantes, 28-30 Juin 2006.
- Soulier, E., (2006), *Le Storytelling - Concepts, outils et applications*, Traité IC2 : Information - Commande - Communication Hermès – Lavoisier.
- Sriti, M.-F., (2004), *Modèle de représentation documentaire pour des documents actifs numériques*, Mémoire de DEA, Université de Technologie de Troyes.
- Staab, S., Braun, C., Bruder, I., Düsterhöft, A., Heuer, A., Klettke, M., Neumann, G., Prager, B., Pretzel, J., Schnurr, H.-P., Studer, R., Uszkoreit, H., et Wrenger, B. (1999), *A System for Facilitating and Enhancing Web Search*. In: *Engineering Applications of Bio-Inspired Artificial Neural Networks. IWANN '99 - Proceedings of International Working Conference on Artificial and Natural Neural Networks*. Volume 2. Alicante, ES, June 2-4, 1999, LNCS 1607, Berlin, Heidelberg: Springer, p.706-714.
- Star, S. L., (1989), The structure of ill-structured solutions: Boundary objects and heterogeneous distributed problem solving. In L. Gasser et M. N. Huhns (Eds.), *Distributed artificial intelligence*, London, England: Pitman., p. 37-54.

- Star, S. L., (1996), Slouching Toward Infrastructure, in *Social Aspects of Digital Libraries*, February 16-17, <http://is.gseis.ucla.edu/research/dl/star.html>
- Stiegler, B., (2000), Annotation, navigation, édition électronique: vers une géographie des Informations, *Ec/arts*, 2.
- Strzalkowski, T. (1999), *Natural Language Information Retrieval*, Kluwer Academic Publishers.
- Subrahmanian, E., Monarch, I., Konda, S., Granger, H., Milliken, R., et Westerberg, A., (2003), The N-Dim Group, Boundary Objects and Prototypes at the Interfaces of Engineering Design, in *CSCW Journal* Volume 12, p. 185-203.
- Sumner, T., Buckingham Shum, S., Wright, M., Bonnardel, N., Piolat, A. et Chevalier, A. (2000), Redesigning the peer review process: A developmental theory-in-action, in R. Dieng, A. Giboin, G. De Michelis et L. Karsenty (Eds.), *Designing cooperative systems: The use of theories and models*, Amsterdam : I.O.S. Press, p. 19-34.
- Sun, C., Jia, X., Zhang, Y., et Yang, Y., (1997), REDUCE: a prototypical cooperative editing system, in *Proceedings of the 7th International Conference on Human-Computer Interaction*, San Francisco, USA, Aug. 24-30, p. 89-92.
- Sun, D., Xia, S., Sun, C., et Chen, D., (2004), Operational transformation for collaborative word processing, in *Proceedings of ACM 2004 Conference on Computer Supported Cooperative Work*, Nov 6-10, Chicago, IL USA.
- Tchounikine, P., (2002a), Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain, *Revue I3 Information-Interaction-Intelligence*, Vol. 2, n°1, Cepadues Editions.
- Tchounikine, P., (2002b), Conception des environnements informatiques d'apprentissage : mieux articuler informatique et sciences humaines et sociales, in Baron G.L., Bruillard E. (Eds.), *Les technologies en éducation : Perspectives de recherche et questions vives*, Paris : INRP - MSH - IUFM de Basse Normandie, p. 203-210.
- Tognotti, S., (1997), *Étude d'un dispositif de coopération rédacteur - lecteur pour l'apprentissage de la rédaction technique*, mémoire de master STAF, TECFA.
- Torres Vilatarsana, M., (2001), L'analyse du discours médiatisé par ordinateur : l'apport de la linguistique à la société de l'information in *Actes du colloque La Communication Médiatisée par Ordinateur : un carrefour de problématiques, 69e Congrès de l'Association canadienne-française pour l'avancement des sciences (Acfas)*, Université de Sherbrooke, 15 et 16 mai 2001, <http://grm.uqam.ca/activites/cmo2001/torres.html>
- Uschold, M., et Grüninger, M., (1996), Ontologies: Principles, Methods and 16 applications, in *Knowledge engineering review*, vol.11, N.2., p. 93-136.
- Valéry, P., (1896), *Monsieur Teste*, (première parution *La soirée avec M. Teste*, Le Centaure), Gallimard, Imaginaire (Coll.), 1998, p. 70.
- Van Wijk, C., et Sanders, T. J. M., (1999), Identifying writing strategies through text analysis, in *Written communication: a quarterly journal of research, theory and application*, 16, p. 51-75.
- Veron, M., (1998), *Modélisation de la composante annotative dans les documents électroniques*, Mémoire de DEA, sept. 1998
http://www.enseeiht.fr/fr/recherche/info/Intellig/PUBLICATIONS/VERON/dea_mv.txt
- Véronis, J., (2000), Annotation automatique de corpus : panorama et état de la technique, in J.-M. Pierrel (Ed.), *Ingénierie des langues*, Paris: Editions Hermès, p. 111-129. <http://www.up.univ-mrs.fr/veronis/pdf/2000hermes4.pdf>

- Véronis, J., et Guimier de Neef, E., (2006), Le traitement des nouvelles formes de communication écrite, in Sabah, G. (Ed.), *Compréhension automatique des langues et interaction*, Paris: Hermès Science, <http://www.up.univ-mrs.fr/veronis/pdf/2006-livre-sabah.pdf>
- Voorhees, E., (1999), Natural Language Processing and Information Retrieval. In Pazienza, M., Information Extraction, LNAI 1714, Springer-Verlag, p. 32-49.
- Vossen, P., (1998), *EuroWordNet A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic publishers.
- Weibel, S., et Lagoze, C., (1997), An element set to support resource discovery. In *International Journal on Digital Libraries* 1, p. 176-186.
- Weick, K. E., (1979), *The Social Psychology of organizing*, New York, Random House.
- Weng, C., et Gennari, J. H., (2004), Information sharing and access: Asynchronous collaborative writing through annotations, in *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, p. 578-581.
- Wenger, E., (1998), *Communities of Practice*, Cambridge: Cambridge University Press.
- Wenger, E., McDermott, R., et Snyder, W.M., (2002), *Cultivating communities of practice*, Harvard Business School Press, Boston, Massachusetts. Harvard Business School Press.
- Williams, G., (2005), *La linguistique de corpus*, Rennes, PUR.
- Wilson, M., Daniels, D., et Phillips, J., (2004), Annozilla, <http://annozilla.mozdev.org/>
- Wittgenstein, L., (1953), *Recherches philosophiques*, trad. F. Dastur et al., E. Rigal, Paris Gallimard, 2004, p. 64.
- Wolfe, J., (2000), Effects of Annotations on Student Readers and Writers, in *Proceedings of the Fifth ACM Conference on Digital Libraries*, New York (USA): ACM Press, p. 19-26
- Wolfe, J., (2002), Annotation technologies: a software and research review, in *Computers and Composition*: 19, p. 471-497.
- Zacklad, M., (2004), Processus de documentation dans les Documents pour l'Action (DopA) : statut des annotations et technologies de la coopération associées, in *actes du colloque Le numérique : Impact sur le cycle de vie du document pour une analyse interdisciplinaire*, 13-15 Octobre 2004, Montréal (Québec).
- Zacklad, M., (2006), *Documentarisation processes in Documents for Action (DofA): the status of annotations and associated cooperation technologies*, Journal of Computer Supported Cooperative Work, 0925-9724, June 2006, vol. 15, n° 2-3, p. 205-228.
- Zacklad, M., et Barbaud, X., (2004), Vers une application du Web Socio Sémantique pour la réalisation d'un système d'information destiné aux réseaux de santé, In *Second séminaire francophone du Web Sémantique Médical* - 9 mars 2004, Rouen.
- Zacklad, M., Lewkowicz, M., Boujut, J-F., Darses, F., et Détienne, F., (2003), Formes et gestion des annotations numériques collectives en ingénierie collaborative, in *Actes des journées Ingénierie des Connaissances 2003*, Laval.
- ZAnnot, (2003), <http://www.zope.org/Members/Crouton/ZAnnot/>
- Zweigenbaum, P., et Consortium MENELAS, (1994), MENELAS: an access system for medical records using natural language, in *Computer methods and programs in Biomedicine*, n°45, p. 117-120

Annexes

Index	II
Liste des acronymes et abréviations	II
Glossaire	III
La pondération en RI	IV
Coefficient de similitude de JACARD	IV
Coefficient de DICE	IV
Khi carré	IV
TF.IDF	V
Outils d'annotation	VI
Outils d'annotation sémantique.....	VI
Outils d'annotation discursive	VII
Outils d'annotation élaborante	XI
Corpus	XIII
Extraits du corpus de référence	XIII
Extraits du corpus de travail.....	XXIV
Échantillon de corpus étiqueté avec Syntex.....	XXVII
Extrait de Topic Maps du projet	XXXII
AnT&CoW	XXXIII
Questionnaire d'évaluation d'AnT&CoW	XXXIV
Liste des publications au cours de la thèse	XLI
Chapitres dans ouvrages communs :	XLI
Article dans ouvrages communs :	XLI
Article dans conférence nationale avec publication dans une revue en ligne :	XLI
Articles dans conférences internationales à comité de lecture :	XLI
Article dans conférence internationale avec publication des résumés en ligne :	XLI
Articles dans workshops internationaux à comité de lecture :	XLII
Articles dans conférences nationales à comité de lecture :	XLII
Articles et communication dans conférences doctorales avec comité de lecture : ...	XLII

Index

A

ADMO, Analyse du Discours Médiatisé par Ordinateur, 59
 Annotation, 88, 93, 107
 annotation coopérative. *See* Annotation
 annotation discursive. *See* Annotation
 annotation élaborante. *See* Annotation
 annotation sémantique. *See* Annotation
 Annotea, 43
 AnT&CoW, 121
 Anthropocentré, 28
 Aristote. *See* Rhétorique
 Artefact de négociation de frontières, 14

B

boundary negotiation artifact. See Artefact de Négociation de Frontières
 boundary object. *See* Objet Frontière

C

Classification, 145
 client d'annotation. *See* Outil d'annotation
 CMO, Communication Médiatisée par Ordinateur, 57
 Collecticiel, 6, 11, 12, 15, 25, 85, 116, 121
 Communication, 134
 continuum d'annotation. *See* Annotation
 Corpus, 19, 177

D

Document, 74
 DDocument Pour l'Action, 15, 121
 documentarisation. *See* DDocument Pour l'Action

DoPA. *See* DDocument Pour l'Action

E

Élaboration, 140
 Évaluation, 26

F

folksonomy. *See* Classification
 fragment de conversation. *See* Annotation
 fragment de document. *See* Annotation

H

Hayes et Flower. *See* Modèle de processus d'écriture
 Herméneutique, 49

I

Indexation, 144
 interprétation des textes. *See* Herméneutique
 Interprétation du corpus. *See* Corpus

K

Kintsch & Van Dijk. *See* Théorie propositionnelle

L

Levelt. *See* Modèle de production verbale orale
 Logique de communication, 14, 88, 90, 91

M

modèle. *See* Modélisation
 Modèle d'activité instrumenté. *See* Modélisation

Modèle de description de l'activité d'annotation. *See*

Modélisation

Modèle de l'interaction sociale, 102

Modèle de processus d'écriture, 97

modèle de production de discours. *See* Modélisation

Modèle de production verbale orale, 99

Modélisation, 95

N

Nystrand. *See* Modèle de l'interaction sociale

O

Objet-Frontière, 12

Outil d'annotation, 42, 52, 61, 64

P

Point de vue, 81

R

RCAO, Rédaction Coopérative Assistée par Ordinateur,
63, 64

Recherche d'Information, 36

Rhétorique, 104

RTO. *See* Ressource Termino-Ontologique, Classification

S

Scénario, 23, 187

Schème de classification, 14

schibboleth. *See* Sociolecte

SHS, Sciences Humaines et Sociales, 33, 58

Sociolecte, 79

T

TCAO, Travail Coopératif Assisté par Ordinateur, 11, 27

Technocentré, 28

Théorie des acteurs-réseaux, ANT, 12

Théorie propositionnelle, 96

thésaurus sémantique, 151, *See* Classification

Topic Maps, 163

Topic Navigation Maps. *See* Topic Maps

V

Van Wijk et Sanders, 100

W

Web Cognitivement Sémantique. *See* Web Socio-
Sémantique

Web Sémantique, 35, 88

Web Social, 56, 88

Web Socio-Sémantique, 88

Liste des acronymes et abréviations

ADMO : Analyse de Discours Médiatisé par Ordinateur
ADT : Analyse des textes assistée par ordinateur
ANT : Actor Network Theory
BN : Bibliothèque Numérique
CMO : Communication médiatisée par Ordinateur
CMC : Computer Mediated Communication
CSCWriting : Computer Supported Collaborative Writing
CT : Carte de thèmes
RCAO : Rédaction Collaborative Assistée par Ordinateur
Dopa : DOcument Pour L'Action
FOAF : Friend Of A Friend
GC : Gestion des Connaissances
GED : Gestion Electronique de Documents
HTML : Hypertext Markup Language
IC : Ingénierie des Connaissances
NFCE : Nouvelles Formes de Communication Electronique
OPPA : Objet Pour Penser Avec
OWL : Ontology Web Language
POS : Part Of Speech
RDF/RDFS : Resource Description Framework / Schema
RI : Recherche d'Information
RTO : Ressource Termino - Ontologique
SGML : Standard Generalized Markup Language
SHS : Sciences Humaines et Sociales
SRI : Système de Recherche d'Information
STIC : Sciences et Technologies de l'Information et de la Communication
TAL / TALN : Traitement Automatique des Langues / Naturelles
TCAO : Travail Coopératif Assisté par Ordinateur
CSCW : Computer Supported Cooperative Work
TEI : Text Encoding Initiative
TM : Topic Maps
TS : Thésaurus Sémantique
TSSR : Temporarily Shared Social Reality
W2S : Web Socio-Sémantique
W3C : World Wide Web Consortium
WebDav : Web-based Distributed Authoring and Versioning
WS : Web Sémantique
WSo : Web Social
WYSIWYG : What You See Is What You Get
XML : eXtensible Markup Language
XTM : XML Topic Maps

Glossaire

Anagogique : Se dit d'un sens spirituel de l'Écriture fondé sur un type ou un objet figuratif du ciel et de la vie éternelle.

Anaphore : Élément dont l'interprétation nécessite la recherche d'un référent dans le discours (cotexte)

Architextualité : Ensemble des catégories générales, ou transcendantes — types de discours, modes d'énonciation, genres littéraires, etc. — dont relève chaque texte singulier, (in G. Genette, Introduction à l'architexte, 1979).

Blog : Site web sur lequel une ou plusieurs personnes s'expriment de façon libre, sur la base d'une certaine périodicité. Son expression est décomposée en unités chronologiques, chaque unité est susceptible d'être commentée par les lecteurs et est le plus souvent enrichie de liens externes

Clôture textuelle : Bornes formelles d'un texte (marques typographiques, mise en page spécifique, etc.) à partir desquelles on isole un texte d'autres textes ou du co-texte.

Coréférence : Quand deux syntagmes présents dans un même co-texte désignent le même référent.

Cotexte : Unités textuelles qui précèdent ou suivent l'unité textuelle étudiée.

Forum : (électronique de discussion) Lieu de rencontre et d'échange, accessible en ligne (par mél, par des sites Web spécifiques).

Hypertextualité : Toute relation unissant un texte B (hypertexte) à un texte antérieur A (hypotexte) sur lequel il se greffe d'une manière qui n'est pas celle du commentaire, (in Genette, 1982).

Intertextualité : Relation de coprésence entre deux ou plusieurs textes, c'est-à-dire éidétiquement et le plus souvent, par la présence effective d'un texte dans un autre. Forme explicite et littérale : la citation, (in Genette, 1982).

Messagerie Instantanée : Système logiciel qui permet de dialoguer instantanément par ordinateur avec un interlocuteur distant connecté au même réseau informatique, notamment Internet. Contrairement au courrier électronique, la communication se déroule en direct.

Métatextualité : Relation, dite "de commentaire", qui unit un texte à un autre texte dont il parle, sans nécessairement le citer, (in Genette, 1982).

Palier : Degré de complexité. Les principaux paliers sont le morphème, le syntagme, la période, et le texte. On trouve aussi, mot, phrase, texte.

Paratextualité : Relation que le texte entretient, dans l'ensemble formé par une œuvre littéraire, avec son paratexte : titre, sous-titre, intertitres, préfaces, postfaces, avertissements, avant-propos, etc., notes marginales, infrapaginales, etc., (in Genette, 1989).

Philologie : Étude d'une langue par l'analyse critique des textes ; Étude formelle des textes dans les différents manuscrits qui nous ont été transmis (épigraphie, manuscritologie, paléographie), (in Le Petit Robert, s.v. « philologie »).

Remotivation : Attribution d'un nouveau sens à un terme (Re-sémantisation).

Schibboleth : signifie le courant du fleuve. Aujourd'hui, épreuve décisive qui fait juger de la capacité d'une personne.

Sociolecte : Utilisation d'une langue partagée par des sujets parlants appartenant à une classe socio-culturelle

Uri : Uniform Resource Identifier, identifiant une ressource physique ou abstraite pour exprimer les hyperliens du Web.

Url : Uniform Resource Locator, adresse Web utilisée pour identifier les pages et les sites Web.

Textualité : Ensemble des propriétés de cohésion et de cohérence qui rendent un texte irréductible à une suite d'énoncé. Il est généralement reconnu trois niveaux de textualité : l'intratexte (ce qui est *en deça* du texte, lexies, expressions), le texte et l'intertexte (ce qui est autour du texte et au niveau textuel).

Wiki : Système de gestion de contenu de site Web qui rend les pages Web librement et également modifiables par tous les visiteurs autorisés.

La pondération en RI

Différents indices possibles à utiliser :

Coefficient de similitude de JACARD

$$\text{Jaccard}(i,j) = |\text{Intersection}(i,j)| / |\text{Union}(i,j)|$$

Où i et j sont des documents, on se concentre sur la présence simultanée.

Coefficient de DICE

$$\text{Dice}(i,j) = |\text{Intersection}(i,j)| / (|i| + |j|)$$

Khi carré

Formule statistique basée sur le calcul de la somme des carrés des écarts.

$$f(x^2) = \frac{1}{2^k \Gamma(k/2)} x^{2[(k/2)-1]} e^{-x^2/2}$$

TF.IDF

Term Frequency (TF) considère que l'importance d'un terme t est proportionnelle à sa fréquence dans un document.

$$tf = \frac{n_i}{\sum_k n_k}$$

Tf. améliore le rappel et Idf. la précision. Inverse Document Frequency (IDF) considère que les termes qui apparaissent dans peu de documents sont pertinents

$$idf = \log \frac{|D|}{|(d_i \supset t_i)|}$$

D = nombre de documents et $d_i \supset t_i$ = nombre de documents qui contiennent le terme t .

$$tfidf = tf \cdot idf$$

Outils d'annotation

Outils d'annotation sémantique

OntoMat-annotizer

OntoMat-annotizer (Handsuh *et al.*, 2002) est un outil qui permet de soutenir l'annotation sémantique de pages Web via un browser d'ontologie HTML. Le format WS utilisé pour représenter l'ontologie, ses concepts et ses relations est celui d'OWL (Ontology Web Language, (OWL, 2004). Contrairement à tous les outils qui suivent, OntoMat-annotizer ne s'appuie pas sur un extracteur d'information et ne fait pas d'apprentissage. L'utilisateur identifie des instances des concepts des ontologies dans le texte et il annote manuellement les parties de texte identifiées. Son avantage est la liberté d'annoter avec les concepts requis, mais son utilisation reste fastidieuse.

S-CREAM

Ce client est une extension dans le développement d'OntoMat qui intègre OntoMat et un extracteur d'information Amilcare (voir Melita) pour associer automatiquement les concepts aux fragments du texte (Handsuh *et al.*, 2002).

SHOE Knowledge Annotator

SHOE Knowledge Annotator (Heflin et Hendler, 2000) est une application en Java qui permet d'annoter des pages Web avec l'ontologie SHOE. Les étiquettes ajoutées par l'utilisateur sont intégrées au corps du code HTML. SHOE Knowledge Annotator est principalement une aide à l'utilisateur pour annoter sa page d'une façon standardisée.

SyDoM

C. Roussey (Roussey *et al.*, 2001) a développé un outil d'annotation pour le Web sémantique fondé sur un thésaurus sémantique qui permet les recherches multilingues. L'indexation de pages Web est manuelle et se fait à deux niveaux ; par le choix d'éléments de terminologies d'une part et par le choix d'index issus des connaissances du domaine d'autre part. L'ensemble de ces informations est contenu dans le thésaurus sémantique construit et défini a priori par les membres d'une communauté.

Melita

Melita (Dingli, 2003) est un outil d'annotation fondé sur une ontologie ayant pour objectif la gestion des connaissances. Il tend à l'automatisation totale du processus d'indexation par apprentissage par un entraînement sur des textes déjà annotés, grâce à un extracteur de connaissances, Amilcare. L'utilisateur reste encore

superviseur du processus. Les balises insérées pour l'indexation peuvent provenir d'une ontologie ou de fragments textuels sélectionnés, ce qui permet une souplesse que n'aurait pas un outil fondé sur une ontologie seule.

MnM

MnM (Domingue *et al.*, 2002) permet la visualisation de documents déposés pour indexation et de la base de connaissances (BC) servant à l'ajout de métadonnées pour l'indexation. L'étiquetage sémantique des documents se fait selon une BC à peupler ou à charger. Le peuplement de la BC se fait semi-automatiquement au début, puis selon l'apprentissage effectué il est possible d'aller jusqu'à un peuplement tout automatique (à valider par l'utilisateur, un document annoté servant à l'apprentissage pour le peuplement de la BC).

COHSE

COHSE, The Conceptual Open Hypermedia Project (Goble *et al.*, 2001), est un système plus complexe reposant sur un enchaînement de fonctionnalités issues de différents outils. COHSE fait le lien entre les éléments annotés par l'outil d'annotation Mozilla Annotator. Il lie différentes pages annotées avec le même concept et permet le passage de l'une à l'autre. Un lien hypertexte est créé grâce à l'index. Il permet aussi la visualisation de l'ontologie utilisée et des pages annotées reliées parfois entre elles par leur index.

Magpie

(Dzbor *et al.*, 2004) utilise des annotations sémantiques pour soutenir l'interprétation de pages Web par les humains. Il permet d'appliquer successivement des couches (*layers*) de couleurs surlignant un point de vue que l'utilisateur souhaite mettre en valeur. L'annotation fondée sur une ontologie sert ici à l'appréhension cognitive du document par le lecteur. Elle permet d'ajouter une sémantique au sens « humain » du terme et de le faire apparaître dans le document selon une colorisation sémantique. Il permet une navigation Web facilitée grâce à la visualisation d'annotations, d'index, de plusieurs utilisateurs.

Outils d'annotation discursive

Amaya

(Amaya, 2003) est un client développé dans le cadre du projet Annotea du W3C. C'est un client d'annotation qui fonctionne avec le serveur d'annotation ZAnnot sous (Latteier *et al.*, 2003), adoptant le standard Annotea. Amaya est un client spécifique Open Source, développé en langage C. Il permet la navigation et l'édition de documents (Web editor). Il permet d'annoter une page Web, une partie (sélection) d'un document ou une

annotation (reply sur une annotation). Il dispose aussi la liste des *reply* (réponse à) d'une façon chronologique (voir Fig. 56), c'est-à-dire qu'il met en forme le fil de discussion créé par les annotations et leurs réponses.

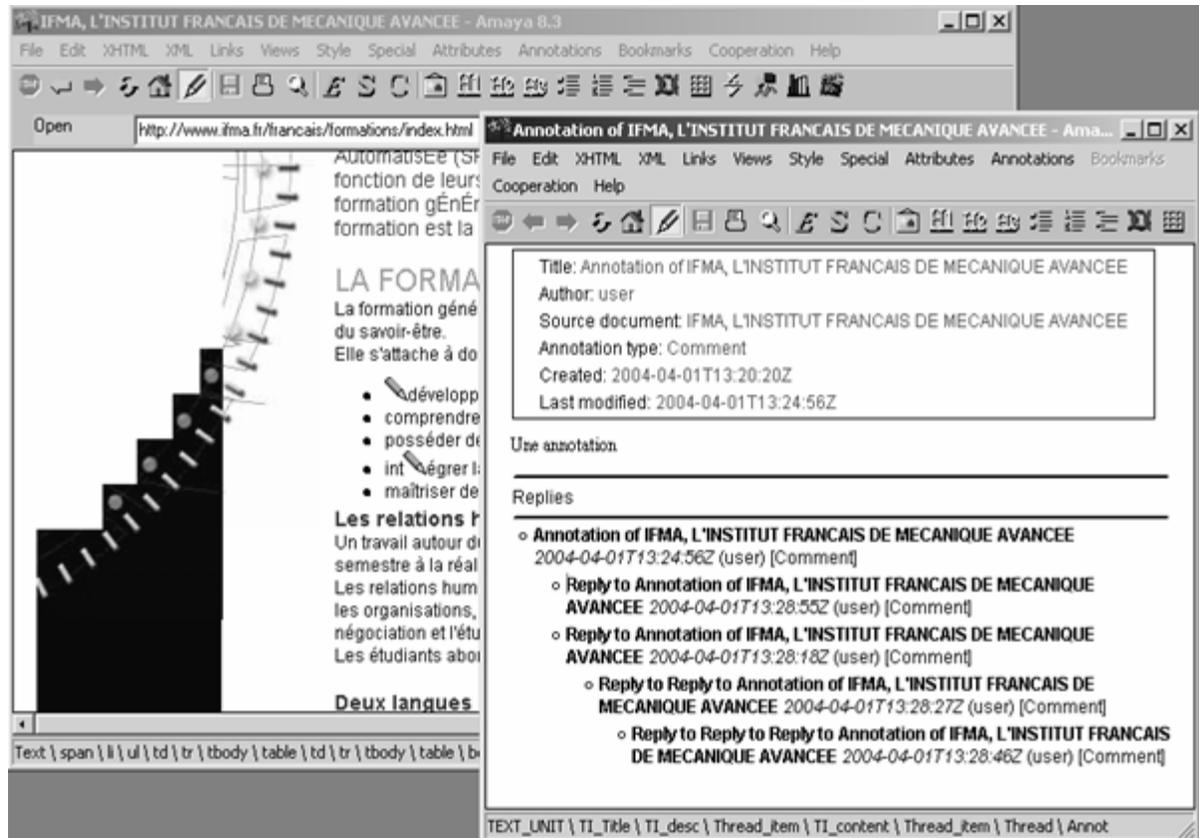


Fig. 56. CAPTURE D'ÉCRAN D'AMAYA

Annozilla

Annotea on Mozilla (Wilson *et al.*, 2004), est un client d'annotation développé pour le navigateur Mozilla - Firefox. Il se présente sous la forme d'un plug-in Open Source (Javascript, XUL, XPCOM, HTML) respectant le standard Annotea, à l'interface simple. Il permet de voir et de créer des annotations associées à une page Web. Annozilla envoie les informations liées aux annotations au format RDF. Il utilise la technologie XPointer pour permettre de localiser les annotations sur la page Web. Les premières versions testées d'Annozilla ne permettaient pas la réponse à annotation.

Janno

Janno (Java et Annotation) est un simple navigateur avec une interface très intuitive développée par Claus Ljunggren (Ljunggren, 2003) dans un cadre universitaire. Il respecte Annotea et partiellement Annotea Bookmark

Schema. Il permet d'afficher une liste d'annotations de n'importe quel serveur. Les annotations peuvent être faites sur un document ou sur une sélection. Il ne permet pas le *reply*, la réponse sur annotation.

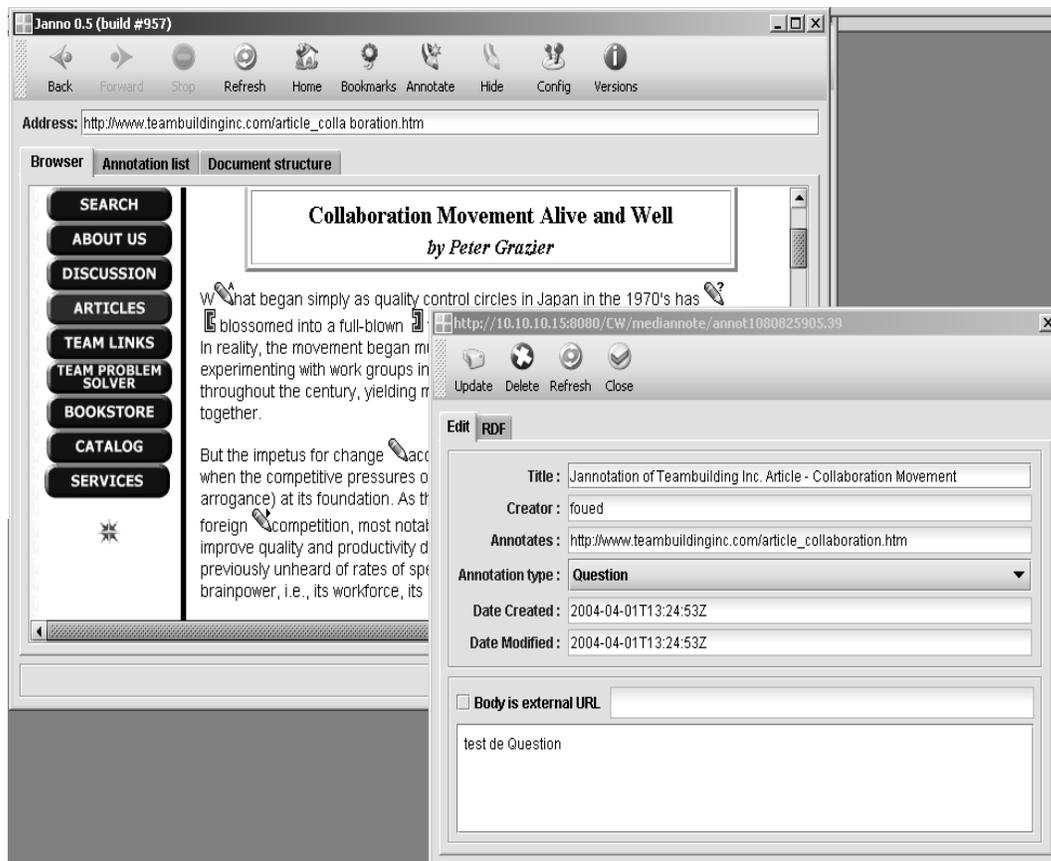


Fig. 57. INTERFACE DE JANNO

TafAnnote

Tafannote (Cabanac *et al.*, 2005) est un outil de recherche développé dans le cadre d'un master au laboratoire IRIT à Toulouse. Descendant d'Annozilla, il permet d'annoter et de répondre à une annotation. Il a pour objectif l'intégration de la tâche d'annotation dans l'environnement de travail de l'utilisateur et une interactivité accrue en ce qui concerne l'accès à l'information. Prenant en compte l'expertise et le jugement de l'annotateur ainsi que différents types d'annotation dont nous donnerons une définition, TafAnnote prend aussi en compte l'élément de note comme *brève appréciation*. De même, une réflexion est menée au niveau de l'adaptation d'indices visuels pour une meilleure compréhension de l'information, en particulier de la structure argumentative du dialogue dans les fils de discussion.

ComMentor

Le système ComMentor (Röscheisen *et al.*, 1994) est un outil d'annotation orienté communication humaine mis en place à Stanford à partir de 1994. Développé en Perl, il permet d'annoter un passage d'une page Web suite à une sélection, de choisir le niveau de visibilité de l'annotation (privé, public ou groupe particulier) ou de répondre à des annotations, ce qui crée des fils de discussions comme ceux des forums. ComMentor permet aussi le partage d'annotations (stockées sur un serveur) ou la navigation en suivant des chemins balisés par des annotations (visite guidée par des commentaires par exemple).

Yawas

Yawas (Yet Another Web Annotation System) (Denoue et Vignollet, 2000) est un outil d'annotation permettant de surligner des parties de pages Web intéressant l'utilisateur. Il fournit aussi des fonctionnalités de recherche d'annotation sur des métadonnées basiques (date de création, auteur ...).

CritLink

CritLink (Ka-Ping, 1998) et CritSuite permettent l'annotation individuelle ou publique de pages Web par l'utilisation de liens hypertextes. Par son usage public, il permet de créer des conversations et échanges médiatisés autour de pages Web. Il est indépendant du format de la page comme du système d'exploitation et se fonde sur des standards du Web largement diffusés (HTML). Il ne nécessite aucun plug-in ou client ou serveur spécifique. Il soutient un multi-ancrage puisqu'il ne fait qu'ajuster les liens hypertextes de la page pour demander d'autres pages liées via le CritLink mediator.

XLibris

XLibris (Price *et al.*, 1998, Miles-Board, 2001) est un système d'annotation qui se fonde sur l'imitation des techniques d'annotation papier pour médiatiser l'activité de lecture critique qu'est l'annotation. Lire c'est voir les pages dans un ordre linéaire mais c'est aussi griffonner, surligner, etc. Un usage collectif de cet outil ergonomique est possible mais non prévu.

Trellis

Trellis (Gil et Ratnakar, 2002) est un outil d'analyse d'information. Il propose à l'utilisateur d'analyser des pages Web puis d'extraire des éléments du texte et de les organiser. Trellis infère ensuite des faits que l'utilisateur peut révoquer. Il permet à l'utilisateur de créer des résumés fiables par le biais d'annotations discursives personnelles et d'extraction manuelle d'information. Trellis permet de déposer des annotations pour soutenir une interprétation individuelle d'un document. Le format de ses informations répond aux recommandations du W3C ce

qui permet une certaine interopérabilité. Cependant, bien que permettant le lien entre des annotations, il ne permet pas l'échange entre utilisateurs.

D3E

Dans la même optique, D3E (Digital Document Discourse Environment) (Sumner *et al.*, 2000) est un outil Web qui permet d'échanger autour de document. Des échanges de messages *autour* d'un document peuvent avoir lieu, mais la conception de cet outil n'étant pas liée à une étude de l'activité d'analyse documentaire, il n'existe pas de réflexion sur le processus coopératif d'interprétation partagée via l'annotation. De plus, la réflexion sur la visualisation et la réutilisation de ces messages est peu approfondie ; en effet, peu de fonctionnalités d'indexation et d'organisation des commentaires sont disponibles (les messages sont présentés de manière arborescente, et indexés selon des attributs standards - date, auteur, titre -).

Outils d'annotation élaborante

PINAS

Piñas est une plateforme (Platform for Interaction Naming And Storage (Decouchant *et al.*, 2001)) permettant une optimisation du travail distribué entre autres à des fins de travail en temps réel ou en synchrone distribué.

BSCW

BSCW (Basic Support for Cooperative Work) (Bentley *et al.*, 1995) est un client Web qui permet de disposer d'un espace de travail partagé par un groupe (shared workspace). BSCW permet principalement de télécharger des fichiers, de notifier le groupe d'un événement et gérer les droits d'un groupe. Il est possible de laisser des commentaires à propos d'un document, mais ceux-ci ne lui sont pas liés et ne sont visibles que sur l'interface Web, pas dans le document.

REDUCE

REDUCE (Real-time Cooperative Editing, (Sun *et al.*, 1997)) est un éditeur de texte coopératif issu de travaux du TCAO. Il est accessible publiquement et est indépendant de toute plateforme. Il permet l'édition concurrente de texte en temps réel et propose un interfaçage personnalisable Web pour le stockage des documents.

DocReview

DocReview est un outil Web pour la révision de documents. Il propose une plateforme universelle par le biais du Web et permet ainsi de partager des documents HTML à critiquer, de créer des glossaires, des bibliographies annotées,... L'éditeur peut sélectionner des options comme modifier les noms des fichiers ou les effacer. Les fichiers peuvent être publics ou protégés par mot de passe. DocReview met en ligne un résumé de tous les commentaires postés sur un document et les commentaires peuvent être ou non affichés sur le document. Lors du postage d'un commentaire, une notification par mél peut être envoyée à une liste de contributeurs.

Quick Topic / Quick Doc Review

Quick Doc Review met à disposition un espace privé pour collecter des commentaires à propos d'un document HTML ou MSWord déposé en ligne. Les commentaires se font par paragraphe et sont visibles par plusieurs utilisateurs. Un forum Web (relayé par une liste mél) est lié à l'espace privé, c'est QuickTopic. Les commentaires sont centralisés avec le document, ce qui évite l'envoi classique de méls avec pièce jointe.

Coword

CoWord (Sun et *al.*, 2004) est un outil qui permet de pallier les faiblesses des commentaires de MSWord. Il permet à plusieurs utilisateurs d'éditer un document MSWord en même temps via le Web. CoWord conserve l'interface connue par une grande majorité d'auteur de documents texte et est donc relativement simple d'accès. CoWord s'appuie sur l'observation de systèmes existants comme REDUCE ou GRACE (GRAphics Collaborative Editing). Il permet aussi de fusionner les modifications de différents utilisateurs et propose un enregistrement de différentes versions en cas de conflit dans l'édition.

Compendium

(Buckingham Shum et *al.*, 2006) développent un outil pour le soutien de pratiques d'interactions éphémères émergeant de l'usage de technologies en ligne (vidéo-conférence par exemple) en les rendant navigables et manipulables. Compendium (figure 58) permet de prendre des notes au fil de l'eau et de les organiser suivant des principes de cartes de concepts s'appuyant sur l'argumentation. Ces cartes permettent de restituer visuellement la structure conceptuelle du discours. De même, un espace est réservé au rendu des échanges ce qui permet de marquer la chronologie des événements. Bien que conçu pour la prise de notes sur des discours oraux, il rejoint les outils présentés auparavant sur le plan de l'objet médiatisé (un commentaire lié à un document audio) et sa gestion.

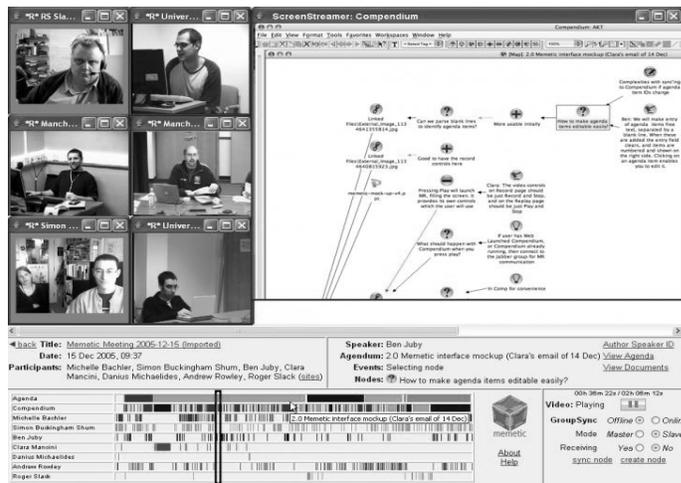


Fig. 58. INTERFACE DE COMPENDIUM

Corpus

Le corpus de référence est disponible sur <http://casanuestra.free.fr/corpus.zip>

Le corpus étiqueté est disponible sur <http://casanuestra.free.fr/corpusSyntax.zip>

Extraits du corpus de référence

Site Web :

DIESELIS

Avion prototype à moteur diesel

Introduction

Depuis l'origine et à de rares exceptions, les moteurs d'avions légers sont des moteurs à essence spécialement conçus, puissants et légers mais construits en petite série et très coûteux.

De leur côté, les moteurs utilisés en automobile, construits en très grandes séries, bénéficient des recherches techniques poussées. Les diesels en particulier sont devenus avec le temps plus puissants, grâce au montage de turbos et d'intercoolers, et relativement moins lourds.

Le Dieselis est un prototype d'avion à moteur diesel qui est basé à l'aéroport de Brest Guipavas. C'est un biplace de fabrication amateur qui a été dessiné spécialement pour recevoir ce type de moteur, et relever ainsi le défi du diesel " trop lourd pour un avion ".

Il a été conçu et réalisé par deux membres de l'Aéro-club du Finistère, Paul Lucas pour la conception et la partie bois, et Serge Pennec pour la motorisation et les pièces métalliques.

La configuration, monoplan à aile basse / moteur à l'avant, est volontairement classique pour ne pas prendre de risques sur la cellule, en plus des inconnues de cette motorisation.

Le moteur choisi pour ce prototype devait être assez puissant (65 cv mini) et pas trop lourd. Après avoir examiné plusieurs moteurs de 1500 à 1900 cc, écarté les plus lourds et les moins puissants, Serge a retenu le turbo 1,5 l de 67 cv monté sur les Opel Corsa. Il a fallu un an pour avionner et tester le moteur au banc, et deux ans et demi pour la conception et la réalisation de la cellule. Le prototype a effectué son premier vol fin septembre 1998, sous l'immatriculation F-PTDI, avec TDI pour Turbo-Diesel Intercooler.

Depuis cette date, le Dieselis a effectué plus de 530 heures de vol sans problèmes (juillet 2000).

Pour mener ce projet, il y avait deux caractéristiques essentielles à prendre en compte : la masse élevée du moteur, la puissance développée peu importante pour un biplace.

La masse du moteur

Le moteur nu (bloc en fonte) pèse 105 kg, au lieu de 60 kg pour un Rotax 912 ou 75 kg pour un dérivé Volkswagen. Avec les accessoires installés sur le moteur on arrive à 140 kg, soit un supplément d'au moins 50 kg sur la masse à vide.

La puissance

La fiche technique du moteur avec turbo, mais sans intercooler, indique 67 cv à 4600 tours, et un couple maximal de 123 Nm entre 2600 et 2800 tours. La courbe de puissance est celle d'un diesel, très plate en haut : on atteint déjà 63 cv à 3600 tours, et il ne reste que seulement 4 cv à prendre dans les 1000 derniers tours. L'intercooler, en refroidissant l'air à l'admission, augmenterait la puissance d'environ 8 à 10 %. Ces puissances sont données en norme DIN, donc comptées avec les accessoires, de façon plus sévère que la norme SAE.

Le rapport poids/puissance

Ce rapport est assez défavorable à première vue : environ 2 kg par cheval. On peut faire une comparaison avec un moteur dérivé d'une base VW, pesant 78 kg et donné pour 76 cv à 2900 t/min, soit 1 kg/cv, donc à priori deux fois mieux que le diesel. Que devient ce rapport en utilisation voyage et en altitude ?

. en vol à 2400 m d'altitude, le VW sans turbo perd 25 % de sa puissance, il reste 57 cv.

. le régime plus élevé, le couple inférieur et donc le diamètre d'hélice plus petit, le capot plus gros et moins bien profilé, la traînée due au refroidissement par air font chuter le rendement de propulsion d'au moins 10 % : il reste 52 cv (et sans doute moins).

. à cause de la chute de rendement d'hélice et du moins bon rendement spécifique du moteur, la masse du carburant à emporter pour une autonomie identique est de 65 kg au lieu de 40, soit un supplément de 25 kg.

Tout considéré, on a finalement 103 kg pour 52 cv, soit 2 kg/cv comme pour le diesel...Au final, en utilisation voyage et en altitude, le rapport poids/puissance du dérivé VW n'est pas meilleur. Il n'en serait pas de même si ce moteur était équipé d'un réducteur et d'un turbo, mais sa masse (et son prix) augmenteraient en même temps.

Le moteur

Monté d'origine sur les Opel Corsa, c'est en fait un moteur Japonais de la marque Isuzu, qui est une autre filiale de General Motors. Cylindrée 1500 cm³ (il existe également en version 1700, 82 cv, de même poids, monté sur les Astra et Vectra). Le bloc est en fonte, ce qui explique le poids élevé du moteur.

Le collecteur d'échappement a été refait en tube inox, car l'original est en fonte, assez lourd.

Le turbocompresseur est d'origine, la surpression du turbo est limitée à 600 grammes.

L'échappement est unique, simple et direct en sortie de turbo. L'intercooler (radiateur pour refroidir l'air à la sortie du turbo) a une section de 260 x 100. C'est une pièce Opel. Il est monté contre la cloison pare-feu. L'admission se fait sur le coté du capot, et la sortie est intégrée au fuselage, au dessus du bord d'attaque de l'aile, dans une zone en dépression. Le passage dans le turbo (échange thermique et compression) élève la température de l'air d'environ 60° en montée. A la sortie de l'intercooler, l'écart de température avec l'air extérieur n'est plus que de 20° en montée, et de seulement 10° en croisière.

Le collecteur d'admission en aluminium est la pièce d'origine, mais très modifiée dans sa géométrie et allégée. Le démarreur est d'origine, alimenté par une batterie de 30 Ah et par une batterie de parc utile en cas de démarrage par temps froid.

Le bâti moteur est de type berceau, en tubes 25CD4S de 16 et de 20, reprenant le moteur par 4 silent blocs situées dans le plan du vilebrequin, donc plus bas que l'axe d'hélice.

Transmission

On a monté un réducteur pour réduire le régime et remonter l'axe d'hélice. Le réducteur est à courroie crantée, donc sans glissement. Un flector (amortisseur de couple) est installé entre le volant moteur (allégé) et le réducteur. La courroie est une GATES, type Poly Chain GT, largeur 62 mm, pas 8 mm. La documentation du fabricant indique, compte tenu du type de moteur et d'entraînement, une puissance admissible en continu de 71 cv à 4000 t/min pour une utilisation de type industriel. Les pignons, 48 et 80 dents, sont en alu anodisé dur, entraxe 240 mm. L'ensemble tourne dans une cage formée par deux platines en alu ép 10 mm reliées par deux tôles en U, le tout assemblé par boulons alu. Le réducteur est fixée à 50 mm du moteur par 6 entretoises reprises sur des fixations de la boîte et du bloc. L'arbre d'hélice tourne sur deux roulements dans un palier prolongateur en alu tourné. Le plateau d'hélice est déporté ainsi de 15 cm en avant du réducteur pour affiner l'avant du capot. Le prolongateur peut être déplacé dans ses fixations sur la platine avant pour régler la tension de la courroie. La masse du réducteur est de 14 kg.

Le rapport de réduction est 48/80, soit 1/1.66 ou 60 % . En croisière, entre 2800 et 3200 t/min, l'hélice tourne à moins de 1900 t/min, au lieu de 2800 t/min pour un moteur en prise directe.

Refroidissement

Le radiateur d'eau est le modèle qui est monté sur la voiture, section 650 x 268 soit 17 dcm², un accessoire difficile à faire rentrer sous le capot avant. On l'a disposé en arrière des sièges sous la tablette arrière, ce qui a permis de reculer le centrage, mais obligé par contre à installer un véritable " chauffage central " dans l'avion. La

masse totale radiateur, tuyaux, vase d'expansion et liquide de refroidissement (7.5 litres) atteint 13 kg.

L'avantage du radiateur à l'arrière est la place disponible à cet endroit pour installer un circuit d'air divergent/convergent optimal pour la traînée. L'admission se fait par une prise d'air Naca sous le fuselage, dans une zone de pression. La section d'entrée est de 60 x 250, soit 1.5 dcm². La sortie d'air comporte (comme sur le Mustang...) un volet réglable pour ajuster le débit et donc la température. Son ouverture permet de gagner 10°.

La température d'eau est d'environ 80° en montée, 70° en palier, et 60° en descente.

Comme sur beaucoup de moteurs en automobile, l'huile est refroidie dans un échangeur avec le circuit d'eau situé à la base du filtre à huile.

Le compartiment moteur est refroidi par une prise d'air sous le capot dirigée vers les zones les plus chaudes, qui sont le collecteur d'échappement et le turbo. La sortie se fait en haut du capotage (et non pas en bas comme cela se fait le plus souvent car il est intéressant de profiter de la dépression existant en montée au dessus du capot pour aider à évacuer les calories).

L'hélice

Elle est de grand diamètre (1.76 m) pour la puissance du moteur, grâce au réducteur et au couple élevé du diesel. Le calage des sections a été établi pour une vitesse de 170 km/h.

Fabrication mixte bois et composites, forme mère en sapin lamellé et stratification en carbone unidirectionnel, 4 couches dégressives posées à +/-30 degrés (flexion/torsion) et à 0 degrés (flexion), masse finie 3,2 kg.

Le diamètre balayé important et la faible puissance requise (moins de 30 cv en croisière) donnent une faible charge favorable au rendement de propulsion. L'étroitesse du bloc moteur a permis d'affiner le capotage en aval de l'hélice. Les performances atteintes ont permis de calculer que le rendement d'hélice est d'au moins 0.85. L'adaptation hélice paraît bonne, mais le mérite en revient principalement à la courbe de couple du diesel (le couple augmente quand le régime diminue), ce qui rend cette adaptation particulièrement facile.

La poussée au décollage est assez élevée pour la puissance, la vitesse de montée est de 5 m/s au lieu de 2,5 à 3 m/s, ou même moins, pour des avions à hélice en prise directe de masse et de puissance installée comparables. Grâce à la faible vitesse de rotation, on a très peu de bruit au sol et en vol. Le cône d'hélice est en carbone /kevlar, diamètre 290 mm, masse 280 grammes.

Aérodynamique

- Pour obtenir de bonnes performances on a réduit les traînées autant que possible.
- . traînée induite : elle est réduite par l'allongement assez élevé de l'aile (8.5).
- . interaction aile /fuselage : on sait la réduire en jouant, quand c'est possible, sur le décalage des maîtres couples, l'idéal étant d'avoir une section de fuselage réduite au niveau de l'épaisseur maximale de l'aile, et une section de fuselage maximale au niveau du bord de fuite de l'aile. Ceci suppose de reculer la verrière par rapport au longeron de l'aile.
- . traînées parasites :

carénages très ajustés autour des roues et des jambes de train, commandes des gouvernes intégrées.

. traînée de refroidissement : la vitesse de l'air qui traverse le radiateur est environ 10 fois plus faible que la vitesse de l'avion, ce qui est souhaitable pour avoir une traînée de refroidissement interne très faible (grande surface d'échange et petite vitesse d'écoulement).

. écoulements internes bien canalisés : intercooler, ventilation du moteur, ventilation de la cabine. Ce sont des gains en traînée souvent négligés, car ils ne se " voient " pas.

. l'échappement sort sous l'avion dans un demi-tunnel caréné qui passe en oblique à travers le plancher de la cabine. La sortie se fait presque horizontalement, pour récupérer la poussée des gaz d'échappement, estimée à 4 % de la poussée de l'hélice en croisière.

- Portance de l'aile

Bien que la surface portante soit réduite de 14 % par rapport à celle de l'ATL, la vitesse de décrochage est plus basse, 80 au lieu de 90. Cela ne peut s'expliquer que par une meilleure interaction ailes/fuselage, le rétreint marqué du fuselage de l'ATL et la flèche avant faisant décrocher prématurément l'emplanture de l'aile en dessous de 100 km/h.

Capotage

Le capotage moteur est en trois parties. Le capot inférieur est attaché au bâti moteur. Les capots supérieurs droit et gauche sont reliés par une charnière continue, et fixés à la cloison pare-feu et au capot inférieur par des vis rapides dzuss. Fabrication en carbone/époxy (résine tenant la température) sur une forme perdue en plâtre réalisée directement sur l'avion, moteur et cône d'hélice en place. La masse des capots est de 3.3 kg.

Fuselage

- Cabine, disposition des pilotes.

Pour équilibrer la masse du moteur, les passagers sont légèrement reculés en arrière du CG.

Les longerons d'ailes passent sous les genoux et non sous les cuisses, ce qui a permis de descendre les sièges et de réduire la hauteur du fuselage. Les ferrures de jonction des longerons sont sous la console.

Largeur extérieure du fuselage 1.14 m, intérieure 1.135 au niveau des coudes.

Les sièges sont encastrés entre les flancs de la console centrale (largeur 14 cm) et les flancs latéraux structurels formant accoudoirs. La largeur d'assise est de 41 cm. Les sièges pourraient être réglables, mais ont été fixés une fois pour toute à la taille standard des deux constructeurs/ pilotes : 1.76 m.

Le chauffage de la cabine se fait par une prise d'air chaud juste en arrière du radia.

Ventilation cabine : prise d'air sur le capot, deux arrivées d'air en haut du tableau de bord, évacuation haute en arrière de la verrière.

- Fabrication

Fuselage en bois, tous collages à l'époxy (résine Excell 755 équivalente à la safe-T, chargée au flox pour les collages et aux micro-ballons pour les congés et les enduits).

La cloison moteur est en sandwich, soit 2 cp de 2 mm collés sur une mousse klégécell de 12 mm avec des inserts en cp acajou au niveau des ferrures. Tôle pare-feu en acier 0.25 mm.

Les flancs sont en contre-plaqué okoumé, épaisseur 2.5 mm à l'avant, doublés par les flancs parallèles entre lesquels sont posés les sièges. Les sièges sont en sandwich, 2 cp okoumé 1.2 mm et une mousse de 10 mm, renforcés par un tissu kevlar au niveau de l'assise.

Le plancher avant est de type caisson avec des renforts internes reliant en triangle les appuis inférieurs du bâti moteur à la ferrure centrale de jonction des ailes.

Les flancs sont en cp okoumé ép. 1,4 mm à en arrière de la verrière. Les lisses principales sont en spruce, les lisses secondaires en cèdre ou 'red cedar', densité 0.35. Les cloisons arrières sont en mousse styrodur de 20 mm tous les 35 cm. Toutes les surfaces sont développables, sauf l'angle inférieur de fuselage en arrière de l'aile qui est à double courbure, réalisé en composite mousse/strate verre-époxy.

Le fuselage pèse 40 kg. Finition extérieure : le fuselage n'est pas marouflé "à l'ancienne" (toile dacron/enduit cellulosique), mais stratifié. Tissu de verre 100 gr/m², résine époxy, enduit époxy allégé au micro-ballons, apprêts et laque deux éléments donnent l'aspect lisse et brillant d'une machine en composites.

Réservoir de 60 litres en composite verre /kevlar époxy (1.8 kg), stratifié sur une forme mâle, fixé en avant du tableau de bord.

Bloc à instruments central, étagères latérales, rangements dans la console centrale.

La verrière est collée sur un cadre en tube. Le cintre avant portant les charnières est en tube acier, le reste en alu. Elle est articulée à l'avant, l'effort de levage étant assisté par des vérins pneumatiques. Le verrouillage se fait à l'intérieur par des leviers latéraux, actionnés de l'extérieur par des poignées profilées.

Les ailes

Elles proviennent d'un ATL, étant mises au rebut pour un défaut d'étanchéité des réservoirs intégrés au bord d'attaque.

Technologie bois et toile, profil Naca 43 015, vrillage 3.2°, corde à l'emplanture 1.42 m, corde à l'extrémité 0.85 m, effilement 0.60. Avec les saumons d'extrémité, l'envergure atteint 9.34 m et l'allongement 8.5. Le longeron est à 32 % de la corde. La liaison des longerons se fait à l'axe, alors qu'ils sont fixés au côté du fuselage sur l'ATL. Ceci a permis de réduire l'envergure d'un mètre dix et de gratter un peu sur la surface portante et sur la masse. Les nervures sont en cp bouleau, placées tous les 20 cm en avant du longeron, et tous les 40 cm en arrière. Le bord d'attaque est coffré en cp okoumé. Les ailerons et les volets sans fente sont d'origine, en tôle alu de 0.5. L'aile étant raccourcie à l'emplanture, les volets sont plus courts de 60 cm. Contrairement à l'ATL, l'aile n'a pas de flèche avant, ce qui fait que les nervures sont légèrement en biais (7.5°). La masse totale des ailes équipées et finies est de 63 kg.

Le revêtement en toile est seulement collé aux chapeaux de nervures (largeur 40 mm), ce qui explique sans doute que la VNE de l'ATL (et donc celle du Dieselis) soit limitée à 195 km/h.

Empennage horizontal

Stabilisateur monobloc de technologie bois et toile, envergure 2.88 m, surface 1.80 m².

Le bras de levier par rapport au foyer de l'aile est de 3.66 m, soit 3.36 fois la corde moyenne de l'aile (1.10 m). Le compensateur monté au bord de fuite sert également d'anti-tab, car il pivote automatiquement dans le même sens que le stabilisateur pour donner un effort au manche. Commande de trim : moteur électrique et vis sans fin agissant par un levier sur la fixation de la bielle de commande du tab, côté fuselage. Contrepoids d'équilibrage interne au fuselage.

Commande par câbles et bielle, différentiel de débattement 12° à cabrer / 6° à piquer.

Empennage vertical

Le plan fixe est tenu par deux longerons repris sur des cloisons de fuselage, les nervures sont en mousse et le coffrage en cp okoumé 1.2 mm. Le volet de direction est entoilé, son guignol de commande est interne au fuselage et accessible par la pointe arrière du fuselage.

Le train

Le train principal est une lame en carbone pré-imprégné, polymérisée à 160° sous vide. Epaisseur moyenne 20 mm, masse brute 5.8 kg. Les jambes sont très fines et profilées, et les jonctions des jambes au fuselage sont carénées. La lame est fixée aux flancs de fuselage par des brides en U, usinées en dural, l'attache avant étant reprise au niveau des ferrures de longerons. La reprise d'effort dans le plan axial se fait sur la ferrure centrale de longeron.

Roues de 320 x 120, sur jantes en alu de 4 pouces. Ce sont des roues avant de Rallye.

Freins à disque diamètre 140 à commande hydraulique, commande par un levier latéral.

Carénages de roues en stratifié verre kevlar/époxy, réalisés sur mousse perdue.

Train avant : jambe télescopique en tube acier 30 x 2, suspension par blocs élastiques intégrés dans le tube, débattement environ 60 mm. La commande de direction se fait par une biellette reliée au palonnier à travers la cloison pare-feu. Roue avant 270 x 80.

Commandes de vol

Double commande complète, manche, gaz et palonnier. Les câbles de profondeur et de direction passent dans la console centrale, puis au dessus du radiateur situé sous la tablette arrière. Guignol de profondeur à l'arrière, et bielle de commande sous le stabilisateur. Les ailerons sont commandés par câbles et par bielles. Les volets sont commandés par des bielles internes au fuselage, l'accrochage au volet étant caché sous le congé karman de raccordement aile-fuselage (bielles et leviers à l'extérieur sur l'ATL). Ces bielles sont actionnées par un tube de

torsion en alu passant sous les sièges, avec un levier de commande le long de la console. Les pédales ne sont pas réglables ; le palonnier très léger, en tubes de 10, pèse 650 gr.

La commande du volet de direction se fait par des câbles en kevlar de 3.5 mm (masse 30 gr).

L'instrumentation est de type VFR, avec pour le moteur la pression turbo, pression dans le circuit d'eau et les températures d'échappement, d'eau et d'huile. Masse des instruments : 3.1 kg. Radio et GPS portables.

Centrage

Le réservoir se situe assez près du CG, son niveau de remplissage influe peu le centrage. Le centrage est en avant (19 %) en solo et normal en double (25 % de la corde moyenne). On peut emmener jusqu'à 25 kg de bagages sur la tablette arrière, le centrage passe alors à 29 %.

Les roues arrières sont peu reculées par rapport au CG. La roue avant est peu chargée, portant moins de 10 % du poids.

Devis des masses / chasse au poids

Il fallait gagner quelque part les kilos en trop du moteur, environ 50 kg de plus qu'une base VW, ou 65 kg de plus qu'un Rotax 912. La chasse aux grammes a été lancée dès le début de la construction de façon systématique :

- . meulage des bossages moteur inutiles, allègement systématique des ferrures
- . fuselage très léger en échantillonnage, cp okoumé et spruce, stratifications très fines
- . lame de train, capots moteur, renforts de fuselage et cône d'hélice en carbone
- . réservoir et carénages de roues en kevlar/époxy
- . axes d'aile et visseries diverses en titane

Au final, l'avion pèse 355 kg (345 sans la batterie), soit 25 kg de moins qu'un ATL.

Le devis des masses et le centrage prévus au départ du projet ont été respectés.

Performances mesurées, avec deux pilotes à bord et 30 litres de carburant :

- . décollage sur piste en dur en 150 mètres, 12 secondes.
- . vitesse verticale : 5 m/sec à 3600 t/min, à 125 km/h
- . croisière économique : 150 km/h à 2850 t/min, 5.3 l/h
- . croisière rapide à 8000 pieds : 200 km/h (vitesse indiquée 175 km/h)
- . finesse max. en palier : 16 à 125 km/h
- . taux de chute à 120 km/h, moteur au ralenti ou calé : 2.6 m/s, soit une finesse de 12.8
- . vitesse d'approche : 110, vitesse d'atterrissage : 90
- . vitesses de décrochage : 80 en lisse, 78 avec les volets à 15°, 75 avec les volets à 25° ou 35°

- . autonomie : 7 à 10 heures
- . distance franchissable : 1200 à 1600 km
- Conduite moteur.

Le régime de décollage est de 3600 t/min. La puissance utilisée en montée est environ 68 cv. En palier, le régime de 3850 n'a pas été dépassé pour respecter la Vne. En ballade le moteur est utilisé entre 2800 et 3200 t/min, ce qui donne entre 26 et 35 cv, soit environ la moitié de la puissance disponible. A la différence d'un moteur à essence, un diesel n'est pas vraiment fait pour tourner à haut régime, mais plutôt pour être exploité sur le couple.

Le régime de ralenti est de 1200 t/min, soit 700 t/min à l'hélice. Grâce à la compression élevée du diesel, on profite moteur réduit d'un freinage aérodynamique d'hélice important réduisant la finesse en approche à 12.8 au lieu de 16.

Une sonde de température d'échappement indique 700° en montée et 300 à 450° en palier. La pression de turbo indiquée est de 500/600 gr en montée, et de 100 à 300 grammes en croisière.

Tableau de caractéristiques moteur en fonction de la vitesse :

[...]

Rapports :

TX : Montage d'une hélice d'avion sur un moteur diesel

La première chose à faire est de définir le besoin en interrogeant les personnes concernées.

Il faut ensuite faire : Une analyse fonctionnelle :

- Déterminer les fonctions,
- Déterminer les contraintes environnementales (les différences de conditions atmosphériques à hautes altitudes peuvent énormément influencer sur le comportement du moteur, par exemple),
- Déterminer les contraintes dimensionnelles. L'avion faisant partie de l'environnement du système, il est nécessaire de trouver une solution qui puisse s'adapter aux dimensions de celui-ci. Il faut donc trouver un compromis entre un fonctionnement optimum et pouvoir insérer le tout dans l'avion,
- Déterminer les limites économiques. En effet, dans ce genre de travail, le gain de poids est une des priorités, or un tel projet vise à rendre les vols plus économiques, il faut donc trouver une solution technologique fiable, légère et également économique, ces termes étant toujours en contradiction.
- Réunir des connaissances sur différentes solutions technologiques (sans oublier de voir s'il est possible d'adapter la solution technologique déjà employée pour les moteurs AVGAS).
- Envisager plusieurs solutions technologiques.
- Sélectionner une solution technologique.
- Dimensionner la solution (rendre la plage de fonctionnement du moteur compatible avec celle de l'hélice).
- Affiner le dimensionnement avec une ou plusieurs techniques d'optimisation.

Ce projet semble tout de même très complexe et dense à premières vues. En effet, la collecte d'informations pour l'analyse du besoin, l'utilisation d'outils encore peu connus... demandent à priori beaucoup de temps. Cependant, le sujet étant intéressant et surtout très concret, celui-ci me motive tout particulièrement surtout si je parviens à l'aboutissement total du projet : le montage d'une hélice d'avion sur un moteur diesel en tenant compte des nombreuses contraintes énoncées précédemment.

[...]

Documentation commerciale :

567

Accouplement élastomère à griffes à moyeu conique et bague de serrage

Matière, finition:

Système d'accouplement: polyuréthane, dureté Shore 98 - A.

Moyeu: aluminium.

Bague conique: acier de traitement bruni.

Exemple de commande: nlm 23021-010, D1 = 14H6 D2 = 14H6

(indiquer séparément les cotes d'alésage de moyeu D1 et D2 souhaitées, avec la classe / plage de tolérance correspondante).

Nota:

Cette gamme d'accouplements convient tout particulièrement pour l'utilisation dans des motorisations de broches principales ou de broches de perçage, qui requièrent des régimes élevés. Avant l'enfichage, les deux moyeux à bague de serrage doivent être fixés sur les pivots d'arbres respectifs au couple de serrage spécifié. Un léger huilage de l'étoile minimise l'effort de montage. Consignes de montage: Pour relier le pivot d'arbre à l'alésage du moyeu, il y a lieu de choisir un ajustement de 0.02 mm maximum.

Exemple: Arbre Ø 25 k6

Alésage Ø 25 G6.

Les diamètres inférieurs à Dmini sont réalisables, toutefois le couple nominal n'est plus garanti.

Sur demande, les moyeux sont disponibles avec avant - trou

Variante disponibles sur demande:

23021 norelem nlm

Référence Taille Couple Moment Résistance Déport maximal Force de Couple de Régime g nominal d'inertie statique ressort serrage max.

Nm (10-3 kgm²) á la torsion radiale des vis min-1 Nm/rayon axial ± latéral N/mm Nm 23021-010 10 10
 0,015 0,04 0,5 0,1 600 1,8 20000 110 23021-017 17 17 0,05 0,24 0,5 0,1 2100 4 18000 280 23021-043 43 43
 0,19 0,4 0,5 0,1 2500 8 14500 400 23021-060 60 60 0,28 0,6 0,5 0,1 2600 8 13000 600 23021-150 150 150 0,65
 1,05 1 0,1 3300 8 11000 900 23021-320 320 320 2 2 1 0,12 4500 35 9000 1900 23021-500 500 500 5,6 5,8 1
 0,15 5900 67 7500 4500 23021-700 700 700 13 7 1 0,15 7000 115 6000 7000

Taille D1/D2 D3

+0,2 D4 D5

+0,5 D6 L L1 L2 L3 B B1 C

(DIN min. max. 912-12.9) 10 6 14 17 32 8,5 10,5 50 18,5 13 15,5 10 2 4 x M3 17 9 19 22 40 9,5 18 66 25 16 21 12
 3 6 x M4 43 12 24 29 50 12,5 27 78 30 18 25 14 3 4 x M5 60 12 26 30 55 12,5 27 78 30 18 25 14 3 4 x M5 150 17
 36 40 65 14,5 30 90 35 20 30 15 4 8 x M5 320 20 40 46 80 16,5 38 114 45 24 40 18 4 4 x M8 500 22 48 58 100
 20,5 47 138 55 28 49 22 5 4 x M10 700 25 60 72 120 22,5 58 155 61 33 54 25 6 4 x M12

Etoile élastomère

Taille D1/D2 avec avant-trou

10 6 17 9 43 10 60 12 150 12 320 18 500 20 700 24

Extraits du corpus de travail

sujet : sujet : Air Campus _ clé de tension en CN+ plans

Salut Adelphe,

Pourrais-tu réaliser sur le centre cette clé de tension à découper dans une tôle d'acier doux ep 15mm.

Ci-joint le plan et le modèle step. Les perçages seront réalisés par Félicien.

Ton emploi du temps te permet-il de faire cette clé pour le milieu de la semaine prochaine ?

Merci.

Pépito.

De: "Adelphe Armange" <adelphe.arrive@utt.fr>

Objet: Modèle STEP du boitier.

Date: mardi 13 septembre 2005 15:24

Salut Adelphe,

Ci-joint le modèle STEP du boitier ainsi que le plan 2D afin de préparer la FAO pour l'usinage de l'excentrique.

Merci de me renvoyer les photos de la plaque amortisseur que tu as usiné hier.

Pépito.

Sujet:

Air Campus ..

De:

Pépito LAFON <pépito.lun@utt.fr>

Date:

Wed, 04 May 2005 12:33:55 +0200

Pour:

adelphe.arrive@utt.fr, Léon Roux <léon.roux@utt.fr>

Salut Adelphe,

Comme convenu ce matin ci-joint les pièces restant à réaliser ...

A+.

Pépito.

Plateau_armortisseur.stp

Platine_moteur.stp

Support_demarreur.stp

Support_reducteur_droit_v2.stp

Support_reducteur_gauche_v2.stp

Tole_inf.stp

Tole_sup.stp

Sujet:

AIR CAMPUS : platine moteur

De:

Pépito LAFON <pépito.lun@utt.fr>

Date:

Wed, 04 May 2005 12:39:34 +0200

Pour:
adelphe.arrive@utt.fr, Léon Roux <léon.roux@utt.fr>

Attention,

La poche est plus profonde : elle fait 1 mm de profondeur comme sur le modèle STP que tu as reçu.

Vois avec Félicien pour les bruts des supports gauche et droit (25CD4) et pour ceux des tôles.

Ci-joint les plans des supports et des tôles.

Bon courage.
Pépito.

Support_reducteur_v2.pdf
Content-Type:
application/pdf
Content-Encoding:
base64
toles.pdf
Content-Type:
application/pdf
Content-Encoding:
base64

Sujet:
AIR CAMPUS
De:
Pépito LAFON <pépito.lun@utt.fr>
Date:
Wed, 11 May 2005 16:13:54 +0200
Pour:
adelphe.arrive@utt.fr, léon.roux.@utt.fr

Salut Adelphe,

Ci-joint les modèles STEP avec le second perçage pour faciliter la mise en position. J'ai ajouté un plan avec le diamètre mini du brut cylindrique contenant les supports.

A+.
Pépito.
Support_demarrreur.zip
Content-Type:
application/x-zip-compressed
Content-Encoding:
base64
Support_reducteur_v2.pdf
Content-Type:
application/pdf
Content-Encoding:
base64

Adelphe Armange a écrit :

>At 16:54 01/06/2005 +0200, vous avez écrit:

>
>>Salut Adelphe,
>>
>>As-tu reçu mon mail, pourras-tu usiner la clé de boitier et si oui dans
>>quel délai ?
>>
>>Merci de me répondre.
>>Pépito.
>>
>Salut Pépito,
>Je vais usiner la clé de boitier lundi.
>Bonne journée
>A +
>
>Adelphe

Merci en de problèmes sur la FAO contacte moi.

pépito.

>At 16:54 01/06/2005 +0200, vous avez écrit:

>
>>Salut Adelphe,
>>
>>As-tu reçu mon mail, pourras-tu usiner la clé de boitier et si oui dans
>>quel délai ?
>>
>>Merci de me répondre.
>>Pépito.
>>

----- Original Message -----

From: "Pépito LUN" <pépito.lun@utt.fr>
To: <jules.dupuis@utt.fr>; <félicien.willem@utt.fr>;
<léon.roux@utt.fr>; "Désiré Louis" <d.louisact@voila.fr>; "Dorian
Santo" <dorian.santo@utt.fr>
Sent: Monday, April 11, 2005 12:29 PM
Subject: Réduction Delvion

> Bonjour,
>
> Le réducteur est en cours de fabrication. Le document joint
> "game_v1.pdf" fait le point sur l'état de fabrication. Vous constaterez
> qu'il reste encore beaucoup de travail pour cette semaine. Afin de
> profiter de la présence de Louis du vendredi 15 au lundi 18 il serait
> intéressant de pouvoir effectuer le montage du réducteur sur le moteur,
> opération assez délicate.
> Dans l'hypothèse ou l'ensemble des pièces serait terminé vendredi, le
> montage complet pourrait avoir lieu. Si ce n'est pas le cas il faudrait
> impérativement que l'on puisse faire le montage de la "cage du
> réducteur" sur le moteur (cf détail dans le document joint
> Montage_cage_reducteur.pdf).
>
> Ce montage nécessite d'avoir terminé les pièces suivantes pour vendredi :
> 1/ Plaque principale.
> 2/ Platine moteur.
> 3/ Les centreurs M6 et la visserie nécessaire.
> 4/ Support réducteur droit et gauche.
>

> La préparation des données de fabrication pour la plaque principale est terminée., celles de la platine moteur le sont quasiment. Il reste à préparer les données pour les supports droit gauche. Merci à léon de faire le nécessaire.

>

> Dans le fichier liste_visserie.txt se trouve l'ensemble des vis nécessaires au montage complet. Désiré doit pouvoir s'occuper de trouver les vis nécessaires (finition bi-chromatée).

>

> Le montage de la cage du réducteur nécessite également :

>

> 1/ La fabrication d'un centreur se logant dans l'alésage du vilebrequin, normalement cet alésage fait 15.85mm de diamètre merci à Désiré ou Dorian de vérifier cette cote et de la communiquer à Félicien pour l'usinage du centreur.

> 2/ La fabrication des marqueurs, dont le plan de principe est dans 'montage_cage_reducteur.pdf. Je ne connais pas précisément les dimensions, il est donc nécessaire d'aller sur place (sur le moteur) et de mesurer les diamètres adéquats. J'en ai parlé avec Dorian, il va s'en occuper.

> 3/ Un joint pour la conduite de circulation d'eau en tête de bloc moteur. Je pense qu'il faut utiliser le joint d'origine, merci à Désiré de le fournir.

>

> Je ne pourrais pas malheureusement être là ce week end avec vous à cause d'un déplacement prévu de longue date. Je reste joignable au 06 18 95 40 60 jusqu'à la fin de la semaine (dimanche 17 avril) et dans tous les cas ici à l'UTT jusqu'a mardi soir. Merci de me contacter si vous souhaitez des explications supplémentaires.

> Bon courage.

> Pépito.

>

> P:S: Il est nécessaire d'usiner les alésages diamètre 12 dans le porte hélice pour les centreurs. Félicien s'en occupe. Si cela pose un problème contactez moi avant mardi soir.

Échantillon de corpus étiqueté avec Syntex

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE syntex SYSTEM "SyntexX.DTD">
<syntex>
  <SEQ id="_">
    <TXT>Reunion Aircampus</TXT>
    <tokens>
      <t i="1" l="réunion" f="Reunion" c="Nom?P" p="N"/>
      <t i="2" l="aircampus" f="Aircampus" c="NomInc" p="N"/>
    </tokens>
    <dependances>
      <g r="EPI" s="1" c="2"/>
      <d r="EPI" s="2" c="1"/>
    </dependances>
  </SEQ>
  <SEQ id="_1">
    <TXT>Bonjour Jacques ,</TXT>
    <tokens>
      <t i="1" l="bonjour" f="Bonjour" c="Nom?S" p="N"/>
      <t i="2" l="Jacques" f="Jacques" c="NomPrXXPrenom" p="NP"/>
      <t i="3" l="," f="," c="Typo" p="T"/>
    </tokens>
    <dependances>
```

```

<g r="EPI" s="1" c="2"/>
<d r="EPI" s="2" c="1"/>
</dependances>
</SEQ>
<SEQ id="_2">
<TXT>Je suis passé à ton bureau début de semaine pour te voir mais apparemment il y avait une réunion .</TXT>
<tokens>
<t i="1" l="je" f="Je" c="Pro" p="P"/>
<t i="2" l="être" f="suis" c="VCONJS" p="V"/>
<t i="3" l="passer" f="passé" c="PpaMS" p="V"/>
<t i="4" l="à" f="à" c="Prep" p="O"/>
<t i="5" l="ton" f="ton" c="DetMS" p="D"/>
<t i="6" l="bureau" f="bureau" c="NomMS" p="N"/>
<t i="7" l="début" f="début" c="Nom?S" p="N"/>
<t i="8" l="de" f="de" c="Prep" p="O"/>
<t i="9" l="semaine" f="semaine" c="NomXXDate" p="N"/>
<t i="10" l="pour" f="pour" c="Prep" p="O"/>
<t i="11" l="te" f="te" c="Pro" p="P"/>
<t i="12" l="voir" f="voir" c="VINFP" p="V"/>
<t i="13" l="mais" f="mais" c="CCoord" p="Cc"/>
<t i="14" l="apparemment" f="apparemment" c="Adv" p="R"/>
<t i="15" l="il" f="il" c="Pro" p="P"/>
<t i="16" l="y" f="y" c="Pro" p="P"/>
<t i="17" l="avoir" f="avait" c="VCONJS" p="V"/>
<t i="18" l="un" f="une" c="Det??" p="D"/>
<t i="19" l="réunion" f="réunion" c="NomFS" p="N"/>
<t i="20" l="." f="." c="Typo" p="T"/>
</tokens>
<dependances>
<d r="SUJ" s="1" c="2"/>
<g r="AUX" s="2" c="3"/>
<g r="SUJ" s="2" c="1"/>
<g r="PREP" s="3" c="4"/>
<g r="PREP" s="3" c="10"/>
<d r="AUX" s="3" c="2"/>
<g r="NOMPREP" s="4" c="6"/>
<d r="PREP" s="4" c="3"/>
<d r="DET" s="5" c="6"/>
<g r="DET" s="6" c="5"/>
<g r="EPI" s="6" c="7"/>
<d r="NOMPREP" s="6" c="4"/>
<g r="PREP" s="7" c="8"/>
<d r="EPI" s="7" c="6"/>
<g r="NOMPREP" s="8" c="9"/>
<d r="PREP" s="8" c="7"/>
<d r="NOMPREP" s="9" c="8"/>
<g r="NOMPREP" s="10" c="12"/>
<d r="PREP" s="10" c="3"/>
<d r="NOMPREP" s="12" c="10"/>
<d r="SUJ" s="15" c="17"/>
<g r="SUJ" s="17" c="15"/>
<g r="OBJ" s="17" c="19"/>
<d r="DET" s="18" c="19"/>
<g r="DET" s="19" c="18"/>
<d r="OBJ" s="19" c="17"/>
</dependances>
</SEQ>
<SEQ id="_3">
<TXT>Alors voilà mon point de vue concernant l' organisation :</TXT>
<tokens>
<t i="1" l="alors" f="Alors" c="Adv" p="R"/>
<t i="2" l="voilà" f="voilà" c="Adv" p="R"/>
<t i="3" l="mon" f="mon" c="DetMS" p="D"/>
<t i="4" l="point" f="point" c="NomMS" p="N"/>
<t i="5" l="de" f="de" c="Prep" p="O"/>
<t i="6" l="vue" f="vue" c="Nom?S" p="N"/>

```

```

<t i="7" l="concerner" f="concernant" c="Ppr" p="V"/>
<t i="8" l="le" f="l" c="Det??" p="D"/>
<t i="9" l="organisation" f="organisation" c="Nom?S" p="N"/>
<t i="10" l=":" f=":" c="Typo" p="T"/>
</tokens>
<dependances>
<d r="DET" s="3" c="4"/>
<g r="DET" s="4" c="3"/>
<g r="PREP" s="4" c="5"/>
<g r="ADJ" s="4" c="7"/>
<g r="NOMPREP" s="5" c="6"/>
<d r="PREP" s="5" c="4"/>
<d r="NOMPREP" s="6" c="5"/>
<g r="OBJ" s="7" c="9"/>
<d r="ADJ" s="7" c="4"/>
<d r="DET" s="8" c="9"/>
<g r="DET" s="9" c="8"/>
<d r="OBJ" s="9" c="7"/>
</dependances>
</SEQ>
<SEQ id=" 4">
<TXT>Moi je me suis engagé pour travailler sur ce Delevion , il n' y donc pas de problème de disponibilité .</TXT>
<tokens>
<t i="1" l="moi" f="Moi" c="Pro" p="P"/>
<t i="2" l="je" f="je" c="Pro" p="P"/>
<t i="3" l="me" f="me" c="Pro" p="P"/>
<t i="4" l="être" f="suis" c="VCONJS" p="V"/>
<t i="5" l="engager" f="engagé" c="PpaMS" p="V"/>
<t i="6" l="pour" f="pour" c="Prep" p="O"/>
<t i="7" l="travailler" f="travailler" c="VINF" p="V"/>
<t i="8" l="sur" f="sur" c="Prep" p="O"/>
<t i="9" l="ce" f="ce" c="DetMS" p="D"/>
<t i="10" l="Delevion" f="Delevion" c="NomPrXXInc" p="NP"/>
<t i="11" l="," f="," c="Typo" p="T"/>
<t i="12" l="il" f="il" c="Pro" p="P"/>
<t i="13" l="ne" f="n" c="Adv" p="R"/>
<t i="14" l="y" f="y" c="Pro" p="P"/>
<t i="15" l="donc" f="donc" c="Adv" p="R"/>
<t i="16" l="pas" f="pas" c="Adv" p="R"/>
<t i="17" l="de" f="de" c="Prep" p="O"/>
<t i="18" l="problème" f="problème" c="Nom?S" p="N"/>
<t i="19" l="de" f="de" c="Prep" p="O"/>
<t i="20" l="disponibilité" f="disponibilité" c="Nom?S" p="N"/>
<t i="21" l="." f="." c="Typo" p="T"/>
</tokens>
<dependances>
<d r="SUJ" s="2" c="4"/>
<d r="REF" s="3" c="4"/>
<g r="SUJ" s="4" c="2"/>
<g r="REF" s="4" c="3"/>
<g r="AUX" s="4" c="5"/>
<g r="PREP" s="5" c="6"/>
<d r="AUX" s="5" c="4"/>
<g r="NOMPREP" s="6" c="7"/>
<d r="PREP" s="6" c="5"/>
<g r="PREP" s="7" c="8"/>
<d r="NOMPREP" s="7" c="6"/>
<g r="NOMPREP" s="8" c="10"/>
<d r="PREP" s="8" c="7"/>
<d r="DET" s="9" c="10"/>
<g r="DET" s="10" c="9"/>
<d r="NOMPREP" s="10" c="8"/>
<g r="NOMPREP" s="17" c="18"/>
<g r="PREP" s="18" c="19"/>
<d r="NOMPREP" s="18" c="17"/>
<g r="NOMPREP" s="19" c="20"/>

```

```

<d r="PREP" s="19" c="18"/>
<d r="NOMPREP" s="20" c="19"/>
</dependances>
</SEQ>
<SEQ id="_7">
<TXT>Le démontage du moteur à l' IUT a confirmé la vision que j' avais de David : l' homme ne me paraît pas très
diplomate , certes mais simplement il se plaint de ne pas avoir d' aide de la part d' étudiants .</TXT>
<tokens>
<t i="1" l="le" f="Le" c="Det??" p="D"/>
<t i="2" l="démontage" f="démontage" c="NomMS" p="N"/>
<t i="3" l="de" f="du" c="Prep" p="O"/>
<t i="4" l="moteur" f="moteur" c="NomMS" p="N"/>
<t i="5" l="à" f="à" c="Prep" p="O"/>
<t i="6" l="le" f="l'" c="Det??" p="D"/>
<t i="7" l="l'ut" f="IUT" c="NomPrXXInc" p="NP"/>
<t i="8" l="avoir" f="a" c="VCONJS" p="V"/>
<t i="9" l="confirmer" f="confirmé" c="PpaMS" p="V"/>
<t i="10" l="le" f="la" c="Det??" p="D"/>
<t i="11" l="vision" f="vision" c="NomFS" p="N"/>
<t i="12" l="que" f="que" c="ProRel" p="P"/>
<t i="13" l="je" f="j'" c="Pro" p="P"/>
<t i="14" l="avoir" f="avais" c="VCONJS" p="V"/>
<t i="15" l="de" f="de" c="Prep" p="O"/>
<t i="16" l="David" f="David" c="NomPrXXPrenom" p="NP"/>
<t i="17" l="," f="," c="Typo" p="T"/>
<t i="18" l="le" f="l'" c="Det??" p="D"/>
<t i="19" l="homme" f="homme" c="Nom?S" p="N"/>
<t i="20" l="ne" f="ne" c="Adv" p="R"/>
<t i="21" l="me" f="me" c="Pro" p="P"/>
<t i="22" l="parer" f="paraît" c="VCONJS" p="V"/>
<t i="23" l="pas" f="pas" c="Adv" p="R"/>
<t i="24" l="très" f="très" c="Adv" p="R"/>
<t i="25" l="diplomate" f="diplomate" c="Nom?S" p="N"/>
<t i="26" l="," f="," c="Typo" p="T"/>
<t i="27" l="certes" f="certes" c="Adv" p="R"/>
<t i="28" l="mais" f="mais" c="CCoord" p="Cc"/>
<t i="29" l="simplement" f="simplement" c="Adv" p="R"/>
<t i="30" l="il" f="il" c="Pro" p="P"/>
<t i="31" l="se" f="se" c="Pro" p="P"/>
<t i="32" l="plaindre" f="plaint" c="VCONJS" p="V"/>
<t i="33" l="de" f="de" c="Prep" p="O"/>
<t i="34" l="ne" f="ne" c="Adv" p="R"/>
<t i="35" l="pas" f="pas" c="Adv" p="R"/>
<t i="36" l="avoir" f="avoir" c="VINFP" p="V"/>
<t i="37" l="de" f="d'" c="Prep" p="O"/>
<t i="38" l="aide" f="aide" c="Nom?S" p="N"/>
<t i="39" l="de la part de" f="de la part d'" c="Prep" p="O"/>
<t i="40" l="étudiant" f="étudiants" c="Nom?P" p="N"/>
<t i="41" l="," f="," c="Typo" p="T"/>
</tokens>
<dependances>
<d r="DET" s="1" c="2"/>
<g r="DET" s="2" c="1"/>
<g r="PREP" s="2" c="3"/>
<g r="PREP" s="2" c="5"/>
<d r="SUJ" s="2" c="8"/>
<g r="NOMPREP" s="3" c="4"/>
<d r="PREP" s="3" c="2"/>
<d r="NOMPREP" s="4" c="3"/>
<g r="NOMPREP" s="5" c="7"/>
<d r="PREP" s="5" c="2"/>
<d r="DET" s="6" c="7"/>
<g r="DET" s="7" c="6"/>
<d r="NOMPREP" s="7" c="5"/>
<g r="AUX" s="8" c="9"/>
<g r="SUJ" s="8" c="2"/>

```

```

<g r="OBJ" s="8" c="11"/>
<d r="AUX" s="9" c="8"/>
<d r="DET" s="10" c="11"/>
<g r="DET" s="11" c="10"/>
<d r="OBJ" s="11" c="8"/>
<d r="REL" s="11" c="12"/>
<g r="REL" s="12" c="11"/>
<d r="OBJ" s="12" c="14"/>
<d r="SUJ" s="13" c="14"/>
<g r="SUJ" s="14" c="13"/>
<g r="OBJ" s="14" c="12"/>
<g r="PREP" s="14" c="15"/>
<g r="NOMPREP" s="15" c="16"/>
<d r="PREP" s="15" c="14"/>
<d r="NOMPREP" s="16" c="15"/>
<d r="DET" s="18" c="19"/>
<g r="DET" s="19" c="18"/>
<d r="SUJ" s="19" c="22"/>
<d r="ADV" s="20" c="22"/>
<g r="ADV" s="22" c="20"/>
<g r="ADV" s="22" c="23"/>
<g r="ADV" s="22" c="24"/>
<g r="SUJ" s="22" c="19"/>
<g r="OBJ" s="22" c="25"/>
<d r="ADV" s="23" c="22"/>
<d r="ADV" s="24" c="22"/>
<d r="OBJ" s="25" c="22"/>
<d r="SUJ" s="30" c="32"/>
<d r="REF" s="31" c="32"/>
<g r="REF" s="32" c="31"/>
<g r="SUJ" s="32" c="30"/>
<g r="PREP" s="32" c="33"/>
<g r="NOMPREP" s="33" c="36"/>
<d r="PREP" s="33" c="32"/>
<d r="ADV" s="34" c="35"/>
<g r="ADV" s="35" c="34"/>
<d r="ADV" s="35" c="36"/>
<g r="ADV" s="36" c="35"/>
<g r="PREP" s="36" c="37"/>
<g r="PREP" s="36" c="39"/>
<d r="NOMPREP" s="36" c="33"/>
<g r="NOMPREP" s="37" c="38"/>
<d r="PREP" s="37" c="36"/>
<d r="NOMPREP" s="38" c="37"/>
<g r="NOMPREP" s="39" c="40"/>
<d r="PREP" s="39" c="36"/>
<d r="NOMPREP" s="40" c="39"/>
</dependances>
</SEQ>

```

Extrait de Topic Maps du projet

```

<?xml version="1.0" encoding="UTF-8"?>
<topicMap id="x1mp6rmrod-0"
  xml:base="c:\gaelle\corpus_reference-syntax.xml"
  xmlns="http://www.topicmaps.org/xtm/1.0/"
  xmlns:xlink="http://www.w3.org/1999/xlink">
  <topic id="Justification">
    <instanceOf>
      <topicRef xlink:href="#Argumentation"/>
    </instanceOf>
    <baseName id="x1mp6rms74-c">
      <scope>
        <topicRef xlink:href="#Argumentation"/>
      </scope>
      <baseNameString>Justification</baseNameString>
    </baseName>
  </topic>
  <topic id="Financière">
    <instanceOf>
      <topicRef xlink:href="#Planification"/>
    </instanceOf>
    <baseName id="x1mp6rms74-2">
      <scope>
        <topicRef xlink:href="#Planification"/>
      </scope>
      <baseNameString>Financière</baseNameString>
    </baseName>
    <occurrence id="x1mp6rms74-32a">
      <resourceData>commande </resourceData>
    </occurrence>
    <occurrence id="x1mp6rms74-14ed">
      <resourceData>coût </resourceData>
    </occurrence>
    <occurrence id="x1mp6rms74-14f7">
      <resourceData>le coût de ses vols de tourisme </resourceData>
    </occurrence>
    <occurrence id="x1mp6rms74-14f8">
      <resourceData>réduire le coût </resourceData>
    </occurrence>
    <occurrence id="x1mp6rms74-17d0">
      <resourceData>prix </resourceData>
    </occurrence>
  </topic>
  <topic id="phénomène">
    <instanceOf>
      <topicRef xlink:href="#Domaine"/>
    </instanceOf>
    <baseName id="x1mp6rms74-1577">
      <scope>
        <topicRef xlink:href="#Domaine"/>
      </scope>
      <baseNameString>phénomène</baseNameString>
    </baseName>
    <occurrence id="x1mp6rms74-1578">
      <resourceData>phénomènes d' ondes de chocs </resourceData>
    </occurrence>
    <occurrence id="x1mp6rms74-1579">

```

```

    <resourceData>phénomènes de délaminage </resourceData>
  </occurrence>
  <occurrence id="x1mp6rms74-157e">
    <resourceData>phénomènes d' ondes </resourceData>
  </occurrence>
  <occurrence id="x1mp6rms74-1581">
    <resourceData>Des phénomènes de délaminage </resourceData>
  </occurrence>
  <occurrence id="x1mp6rms74-1584">
    <resourceData>des phénomènes d' ondes </resourceData>
  </occurrence>
  <occurrence id="x1mp6rms74-1585">
    <resourceData>phénomènes </resourceData>
  </occurrence>
  <occurrence id="x1mp6rms74-1591">
    <resourceData>phénomènes </resourceData>
  </occurrence>
  <occurrence id="x1mp6rms74-1596">
    <resourceData>produit des phénomènes </resourceData>
  </occurrence>
</topic>
<association id="x1mp6rms74-19c0">
  <instanceOf>
    <topicRef xlink:href="#de"/>
  </instanceOf>
  <member id="x1mp6rms74-19c1">
    <roleSpec>
      <topicRef xlink:href="#Membre 1"/>
    </roleSpec>
    <topicRef xlink:href="#phénomène"/>
  </member>
  <member id="x1mp6rms74-19c2">
    <roleSpec>
      <topicRef xlink:href="#Membre 2"/>
    </roleSpec>
    <topicRef xlink:href="#délaminage"/>
  </member>
</association>
</topicMap>

```

Ant&CoW

Tous les fichiers sources et les documentations de l'outil sont disponibles sur <http://casanuestra.fee.fr/AntCow.zip>

Questionnaire d'évaluation d'AnT&CoW

1. Considérez-vous l'interface comme intuitive ?

Tout à fait

Moyennement

Pas du tout

Commentaires :

.....
.....

2. Les fonctionnalités sont elles simples à retrouver ?

Tout à fait

Moyennement

Pas du tout

Commentaires :

.....
.....

3. Les documents sont-ils aisés à récupérer ?

Tout à fait

Moyennement

Pas du tout

Commentaires :

.....
.....

4. Les annotations sont-elles aisées à visualiser ?

Tout à fait

Moyennement

Pas du tout

Commentaires :

.....
.....

5. Les scénarios sont-ils représentatifs des échanges en conception mécanique ?

Tout à fait

Moyennement

Pas du tout

Commentaires :

.....

.....

6. Les scénarios sont-ils représentatifs d'une activité d'initialisation de projet ?

- Tout à fait Moyennement Pas du tout

Commentaires :

.....
.....

7. Les scénarios sont-ils représentatifs d'une activité de planification des tâches ?

- Tout à fait Moyennement Pas du tout

Commentaires :

.....
.....

8. Les scénarios sont-ils représentatifs des rôles des contributeurs ?

- Tout à fait Moyennement Pas du tout

Commentaires :

.....
.....

9. Pour quels type de tâches les scénarios sont-ils représentatifs (phases du projet) ?

.....
.....
.....
.....

10. Quelles activités seraient à mettre ne scénario pour représenter l'ensemble de processus de conception ?

.....
.....
.....

.....

11. L'outil fait-il obstacle à la mise en place des échanges (gêne de la communication, difficulté à communiquer par écrit, ...) ?

- Oui Non Dans une certaine mesure⁴⁴

Commentaires :

.....
.....

12. L'outil fait-il obstacle à la compréhension des échanges (modification du sens, contresens, décontextualisation du problème, ...) ?

- Oui Non Dans une certaine mesure

Commentaires :

.....
.....

13. L'outil fait-il obstacle à la compréhension des documents partagés (validité du document, type de documents, ...) ?

- Oui Non Dans une certaine mesure

Commentaires :

.....
.....

14. L'outil fait-il obstacle à la création de documents ?

- Oui Non Dans une certaine mesure

Commentaires :

.....
.....

15. L'outil permet-il de jouer ces scénarios d'une façon naturelle ?

⁴⁴ Validez par des commentaires dans le champ ci-dessous

Oui Non Dans une certaine mesure

Commentaires :

.....
.....

16. Existe-t-il une possibilité d'affiner le scénario ?

Oui Non Dans une certaine mesure

Commentaires :

.....
.....

17. Les points de vue pris en compte dans la classification sont-ils adaptés à l'activité ?

Tout à fait Moyennement Pas du tout

Commentaires :

.....
.....

18. La classification pour le point de vue *planification*, est-elle :

Suffisante Trop large Trop réduite

→ Une correction manuelle vous semble-t-elle possible ?

oui non

Commentaires :

.....
.....

19. La classification pour le point de vue *argumentation*, est-elle :

Suffisante Trop large Trop réduite

→ Une correction manuelle vous semble-t-elle possible ?

oui non

Commentaires :

.....
.....

20. La classification pour le point de vue *organisation*, est-elle :

- Suffisante Trop large Trop réduite

→ Une correction manuelle vous semble-t-elle possible ?

- oui non

Commentaires :

.....
.....

21. La classification pour le point de vue *domaine*, est-elle :

- Suffisante Trop large Trop réduite

→ Une correction manuelle vous semble-t-elle possible ?

- oui non

Commentaires :

.....
.....

22. Les fonctionnalités sont-elles suffisantes ?

.....
.....

23. Est-ce que l'outil correspond à vos besoins d'annotation ?

.....
.....

24. Est-ce que l'outil correspond à vos besoins de communication ?

.....

.....

25. Est-ce que l'outil correspond à vos besoins de classement de documents ?

.....

.....

26. Est-ce que l'outil correspond à vos besoins de partage de documents ?

.....

.....

27. Pouvez-vous évaluer des limites de l'outil *a priori* ?

.....

.....

28. Quelles fonctionnalités seraient selon vous possibles à ajouter ?

.....

.....

29. Etes - vous satisfait par les possibilités de visualisation possible du fil d'annotation ?

.....

.....

Liste des publications au cours de la thèse

Liste des publications au cours de la thèse

Toutes les publications sont disponibles sur :

http://casanuestra.free.fr/presentation_gaelle_fr.html#publication

Chapitres dans ouvrages communs :

Zacklad M., Bénel A., Barcellini F., Barry-Gréboval C., Boujut J.-F., Bringay S., Burckhardt J.-M., Charlet J., Curé O., Darses F., De Saint Léger M., Détienne F., Guibert S., Lewkowicz M., Lortal G., Pierrat M.-J., Sack W., Salzano G., Todirascu A., Treins M., Turner W., Processus d'annotation dans les documents pour l'action : textualité et médiation de la coopération, in J.Charlet (Ed.), *Documents et Contenu : création, indexation, navigation*, Hermès, Lavoisier, Paris (à paraître).

Lortal G., Lewkowicz M., Todirascu-Courtier A., 2006, Des activités d'annotation : De la glose au document in P. Salembier et M. Zacklad (eds) : *Annotations dans les documents pour l'action*, Hermes Publishing, Londres-Paris.

Article dans ouvrages communs :

Lewkowicz, M., Lortal, G., Todirascu, A., Zacklad, M., Sriti, M.F., (2004), A web-based annotation system for improving cooperation in a care network, in Matera, M., Comai, S., *Engineering Advanced Web Applications, International Conference on Web Engineering, ICWE 2004* Rinton Press ISBN 1-58949-046-0 (p. 227-239).

Article dans conférence nationale avec publication dans une revue en ligne :

Lortal G., Todirascu-Courtier A., Lewkowicz M., (2006), Pour une herméneutique numérique : Médiatiser l'activité d'annotation in *Actes du XVII^e Colloque international d'Albi Langue et Signification (CALS)*; Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation, Albi, 10-14 juillet 2006. *Texto!* juin 2006, vol. XI, n°2. p. 59-69. http://www.revue-texto.net/Archives/Albi_2006/Actes.html

Articles dans conférences internationales à comité de lecture :

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2006), AnT&CoW: Share, Classify and Elaborate Documents by means of Annotation, in Mathew T.C. and Pichappan P. (Eds.), *Proceedings of the IEEE - 1st International Conference on Digital Information Management, ICDIM 06*, Bangalore, India, December 06th-08th, 2006 (to be published).

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2006), An annotation service for e-library: enhancing collaborative reading, in James M. Tien and J. Chen (Eds.), *Proceedings of IEEE - International Conference on Service Systems and Service Management*, Troyes, France, October 25th-27th, 2006 (5 pages).

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2006), Annotations: A Way to Capture Experience, in *Proceedings of Knowledge-Based Intelligent Information and Engineering Systems: 10th Intl Conf. KES2006*, Bournemouth, October 9-11, LNAI, Springer-Verlag (8 pages).

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2006), Enabling communication rationale via annotations: a document-based cooperation in, Hassanaly P., Herrmann T., Kunau G. and Zacklad M. (Eds.) *Proceedings of COOP'06* (short paper), Carry-le-Rouet, May 9th-12th 2006, p. 75-82.

Article dans conférence internationale avec publication des résumés en ligne :

Lortal G., Todirascu-Courtier A., Lewkowicz M., (2005), Sorting annotations to trace interactions, *Computational Linguistics in Netherlands (CLIN)*, December the 16th, <http://www.science.uva.nl/events/CLIN2005/Program/Abstracts/abstract-lortal.html>

Lortal G., Todirascu-Courtier A., Lewkowicz M., (2006), Annotation as negotiation and collaborative writing means, *10th International Conference of the EARLI Special Interest Group on Writing*, september 2006, <http://webhost.ua.ac.be/sigwriting2006/viewabstract.php?id=87&cf=1>

Articles dans workshops internationaux à comité de lecture :

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2005), Annotation: Textual Media for Cooperation, in *Proceedings of Annotation for Cooperation Workshop* November 24-25th (p. 41-50).

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2005), AnT&CoW, a tool supporting collective interpretation of documents through annotation and indexation, in Dieng R., et Matta N., in *Actes du Workshop Knowledge Management and Organizational Memories, KMOM, International Joint Conference on Artificial Intelligence, IJCAI05* (p. 43-54).

Articles dans conférences nationales à comité de lecture :

Lortal G., Todirascu-Courtier A., Lewkowicz M., (2006), Soutenir la coopération par l'indexation semi-automatique d'annotations in Harzallah M., Charlet J., et Aussenac-Gilles N. (Eds.), in *Actes de la conférence Ingénierie des Connaissances, IC 2006* Nantes, 28-30 Juin 2006, (p. 61-70).

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2005), Modélisation de l'activité d'annotation discursive pour la conception d'un collecticiel support à l'herméneutique, in Jaulent M.-C. *Actes de la conférence Ingénierie des Connaissances, IC 2005*, PUG, ISBN 2-7061-1284-0 (p. 169-180).

Lortal G., Grau B., Zock M., (2004), Système d'aide à l'accès lexical : trouver le mot qu'on a sur le bout de la langue in *Actes de Traitement Automatique du Langage Naturel, TALN04*, Fez, Maroc (p. 259-268).

Articles et communication dans conférences doctorales avec comité de lecture :

Lortal G., Lewkowicz M., Todirascu-Courtier A., (2005), L'annotation comme support à la collaboration autour de documents : l'outil AnT&CoW in *Actes des 7èmes Journées Doctorales Informatique et Réseau (JDIR)* 13-15 décembre, (p. 2-11).

Lortal G., (2005), Médiatiser l'annotation discursive support à l'herméneutique in *Actes du 9^e Atelier des Doctorants en Linguistique* 17-18 octobre, (p. 174-181).

Lortal G., (2005), Médiatiser l'annotation discursive support à l'herméneutique, Communication Orale, Journée Jeunes Chercheurs en Linguistique Appliquée, organisée par l'Association Française de Linguistique Appliquée, 29 octobre 2005, Paris.