



HAL
open science

**Français et diglossie. Décrire la situation linguistique
française contemporaine comme une diglossie:
arguments morphosyntaxiques**

Benjamin Massot

► **To cite this version:**

Benjamin Massot. Français et diglossie. Décrire la situation linguistique française contemporaine comme une diglossie: arguments morphosyntaxiques. Linguistique. Université Paris VIII Vincennes-Saint Denis, 2008. Français. NNT: . tel-00726999

HAL Id: tel-00726999

<https://theses.hal.science/tel-00726999v1>

Submitted on 31 Aug 2012

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 Paris 8 — Vincennes → Saint-Denis

École doctorale Cognition, Langage, Interaction

U.F.R. Sciences du Langage

Numéro attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

THÈSE

Nouveau régime

Pour obtenir le grade de

Docteur en Sciences du Langage

Discipline : Linguistique Générale

Présentée et soutenue publiquement

par

Benjamin MASSOT

le 28 novembre 2008

FRANÇAIS ET DIGLOSSIE

DÉCRIRE LA SITUATION LINGUISTIQUE FRANÇAISE

CONTEMPORAINE COMME UNE DIGLOSSIE :

ARGUMENTS MORPHOSYNTAXIQUES

Directeur de recherche(s) :

Mario BARRA JOVER

Jury :

Alain KIHM	CNRS	(pré-rapporteur)
Danielle LEEMAN	Université Paris 10	(présidente du jury)
Patrick SAUZET	Université de Toulouse	(pré-rapporteur)
Achim STEIN	Université de Stuttgart	
Anne ZRIBI-HERTZ	Université Paris 8	

*À tous mes „Hü“ et tous mes „Hott“.
À vous mes guides permanents.*

Et à mon index gauche, qui s'est sacrifié sans hésitation pour permettre la finalisation de la rédaction de cette thèse.



Baptiste Morge

image : © Jean-Baptiste Monge

Remerciements

Tout d'abord, un grand merci à Alain Kihm, Danielle Leeman, Patrick Sauzet, Achim Stein et Anne Zribi-Hertz d'avoir accepté de faire partie de mon jury. Merci à Danielle Leeman d'avoir accepté d'en prendre la présidence, et merci à Alain Kihm et Patrick Sauzet d'avoir accepté de prendre les fonctions de pré-rapporteurs.

Un air de déjà-vu

Ses enseignements, ses idées, ses paroles et ses écrits, ses enthousiasmes, ses colères, ses doutes, ses peurs et ses engagements ont trouvé un tel écho en moi que j'en finis par m'approprier ce qui ne seraient, si l'on suit le mythe de la connaissance, que des savoirs retrouvés. Les références à son travail dans cette thèse, déjà nombreuses, ne donnent pourtant qu'une petite idée du pillage autorisé auquel je me suis adonné pendant toutes ces années. Je ne lui dois pas tout, certes. Sûrement. En cherchant bien. Vraiment bien. En fait, si : je lui ai même piqué son sujet !

Liberté et confiance

Madame Laca (Brenda), Achim (monsieur Stein), Anne (madame Hertz), Joaquim (monsieur Brandão de Carvalho), Florence (madame Villoing), monsieur Sauzet (Patrick), Mario (monsieur Barra (Jover)), vous avez mis en moi une telle confiance dans mes enseignements et mes recherches que je n'ai jamais senti ma liberté mise en jeu. Même, vous l'avez assurée, et protégée. Je vous suis à tous profondément reconnaissant pour ce bien si précieux que vous m'avez offert.

Générosité neuronale

Un grand merci à tous ceux qui ont stimulé mes neurones en donnant à voir l'activité des leurs, à ceux qui n'ont pas toujours ou même jamais été d'accord avec moi autant qu'à ceux qui épousent les mêmes idées que moi : Nesrine Al-Zahre, Mario Barra, Jean-Louis Aroui, Mario Barra, Mathieu Avanzi, Mario Barra, Nora Boneh, Mario Barra, Leïla Boutora, Mario Barra, Gilles Boyé, Mario Barra, Johannes Kabatek, Mario Barra, Peter Koch, Mario Barra, Fabienne Martin, Mario Barra, Pandelis Mavrogiannis, Mario Barra, Gerhard Schaden, Mario Barra, Achim Stein,

Mario Barra, Anne Zribi, Mario Barra, les camarades de CIEL8, Mario Barra, et du DFGK, Mario Barra, les auditoires des colloques et ateliers auxquels j'ai participé, Mario Barra, mes étudiants de *Grammaire française (1 et 2)* de Paris 8, Mario Barra, les étudiants de mes deux Hauptseminare de Stuttgart, Mario Barra, et enfin tous les profanes, les non-linguistes, qui ont souhaité « au moins comprendre le titre en entier ! »

Patience et coopération

Il me faut également citer ceux dont j'ai éprouvé la patience. Je pense ici, entre autres, à mes relecteurs. Pour la forme : Papa, mon petit frère (ça t'apprendra à nous avoir fait relire, en ton temps, tes travaux de *droit*, non mais !), Valou, Nicole (pour une aide pour les traductions de et vers l'anglais), et la fidèle clique des Scheiß-Franzosen de Stuttgart (Julie en „Vorfeld“, Valkenyan, Sylvia, Hary avec un A, et Vincents).

Merci à Nora Boneh d'avoir accepté (et de s'être même proposée) de relire et critiquer des parties de cette thèse.

Laurent Roussarie en premier chef, mais aussi Gerhard Schaden et Achim Stein ont joué un rôle important dans l'aventure au début acrobatique mais finalement si jouissive de ma conversion à \LaTeX . Un grand merci également à Achim Stein et Berndt Möbius de m'avoir donné accès et initié aux outils informatiques de l'IMS (institut de traitement automatique du langage) de Stuttgart.

Et les généreux informateurs : Maurice Dadou pour le poyaudin, Mario Barra pour l'espagnol, Patrick Sauzet pour l'occitan, les anglophones du Stuttgart Rugby Club pour l'anglais, et pour l'allemand, mes étudiants Lorena Killmann, Dennis Keifer et Jan Muckenfuß, et mes collocataires Tanja Becker et Laurence Lachnit, lesquelles ont tenu bon et m'ont retenu d'exagérer des convergences et des différences entre l'allemand et le français qui m'arrangeaient trop.

Les innocents et les coupables

Je remercie les innocents qui m'ont généreusement, involontairement et spontanément fourni des énoncés merveilleusement démotiques, classiques, ou même les deux. Parmi ceux dont je me souviens, il y a Nono le prof de maths (qui m'a bien aidé), Maman, Céline, Anne-Laure la psy-

chomot', Damien le caviste de Vézelay, Tristan „Brüllaffe“, et surtout, surtout, surtout Mustapha, innocent parmi les innocents (qu'il trouve ici toute ma gratitude).

Enfin, une pensée vexée mais bienveillante pour les culottés qui ont osé détecter *chez moi* des démotismes qui avaient échappé à mon auto-observation ! Chapeau à Valkenyan et à Arnaud H. (H comme Hussard).



Heidinger

Massot

Wespe

Jäckh

Die Klasse von 2008

Résumé

On a depuis longtemps fait le constat de nombreuses variantes grammaticales en français (la négation avec et sans *ne*, SV(O) vs. la dislocation à gauche, etc.), les unes étant valorisées et les autres stigmatisées. Dans ce travail, on défend l'idée que l'on a affaire à une situation de diglossie. Cela suppose que les locuteurs intériorisent deux grammaires : l'une, le *français démotique*, est acquise « sur les genoux de la mère » et l'autre, le *français classique tardif*, est acquise à l'école et à travers les institutions qui exigent son emploi.

On place cette problématique dans un cadre qui requiert l'étude des productions spontanées, l'abandon de l'opposition oral-écrit et de son caractère explicatif, et une transcription phonologique des données. À travers l'étude du nombre, de la négation, et des alternatives à SV(O), on montre l'intérêt descriptif et typologique de décrire deux grammaires. Chaque grammaire ainsi décrite est plus consistante. Enfin, une étude de corpus observe un locuteur diglosse, qui n'active bien qu'une grammaire à la fois : il mélange par exemple la dislocation à gauche avec la négation sans *ne*, mais jamais avec la négation avec *ne*.

Abstract

Grammatical variants in French have long been a well-known phenomenon (negation with or without *ne*, SV(O) vs. left-dislocation, etc.), the first being valued, the latter being stigmatised. This work deals with arguments in favour of a case of diglossia. This assumes that the speakers internalise two grammars: the first, *Demotic French*, is acquired at home and the second, *Late Classical French*, is acquired at school and through the institutions that require its use.

This problematic is developed in a framework requiring the study of spontaneous productions, the renunciation of the oral-written opposition and of its accounting properties, and a phonological transcription of the data. Through the study of Number, Negation, and the alternate structures to SV(O), the descriptive and typological advantage of describing two grammars is shown. Thus, both grammars become more consistent. Finally, a corpus study shows a diglossic speaker who indeed activates only one grammar at a time : for example, he mixes left-dislocation with the negation without *ne*, but never with the negation with *ne*.

Sommaire

Remerciements	iv
Résumé	viii
Abstract	ix
Table des figures	xxi
Liste des tableaux	xxiii
Conventions typographiques	xxv
Introduction	1
Le vif du sujet	1
Le contenu de la thèse	6
Quelques absents de la thèse	12
Attention, polémique!	14
I Les protagonistes	15
1 Objet	19
1.1 La langue	19
1.1.1 Langue externe, langue interne	19
1.1.1.1 La langue externe	20
1.1.1.2 La langue interne	21
1.1.2 Autour du débat sur la syntaxe	24
1.1.3 Ici : le point de vue de la grammaire générative	28
1.2 Le français	30
1.2.1 Une métaphore redoutable	30
1.2.2 Langue, progrès, détérioration, évolution, perfection	33

Sommaire

1.3	Conclusion : mon objet d'étude	34
2	Tangentes	35
2.1	Le problème du canal	36
2.1.1	Canaux de communication	36
2.1.1.1	Définition	36
2.1.1.2	Canal et encodage linguistique	37
2.1.1.3	Encodage secondaire	39
2.1.2	Et le 'canal écrit' ?	42
2.2	Écritures, et dépendance et autonomie de l'écrit	43
2.2.1	Histoire et typologie des écritures	43
2.2.2	Dépendance et autonomie des langues écrites	45
2.2.2.1	Dépendance	45
	L'écriture chinoise recyclée	45
	La lecture	46
2.2.2.2	Autonomie de l'écrit	48
	L'holistique	48
	La prise de recul	50
2.2.3	Conclusion	50
2.3	L'orthographe du français : une observation critique	51
2.3.1	L'orthographe : une définition	51
2.3.2	Orthographe et langue : la grande illusion	52
2.3.2.1	« L'orthographe, c'est la langue »	52
2.3.2.2	Un non-patrimoine national	53
2.3.3	Le système graphique du français : un pot-pourri	54
2.3.3.1	La fonction phonographique	55
2.3.3.2	Les fonctions idéographiques	57
	La fonction logographique	58
	Les fonctions morphographiques	60
2.3.4	Conclusion	65
2.4	Compétence linguistique et connaissances métalinguistiques	67
2.4.1	Définitions	67
	Compétence linguistique	67
	Connaissances métalinguistiques	68
	Activités linguistiques et métalinguistiques	68

2.4.2	Exemples	68
2.4.2.1	Cas consensuels	68
	Phonologie	68
	Morphosyntaxe	69
	Lexique	70
	Sociolinguistique	70
2.4.2.2	Cas controversés et polémiques	71
	Jugements	71
	Écrire	71
2.4.2.3	Le passé simple et le subjonctif imparfait en français et en espagnol	73
2.4.2.4	L'exemple de l'accord du participe passé Les PP adjectifs	74
	Les PP dans les passifs et dans les temps composés des verbes conjugués avec <i>être</i>	76
	Les autres PP : dans les temps composés des verbes conjugués avec <i>avoir</i> et dans les constructions pronomi- nales	77
2.4.2.5	« OMAR M'A TUER » et le laxisme ortho- graphique	79
2.5	La transcription adoptée	81
2.5.1	Rendre justice à la morphosyntaxe	81
2.5.2	Analyse en phonèmes du français	82
2.5.3	Découper en mots	83
2.5.4	Quelles informations suprasegmentales ?	85
2.5.5	La liaison	85
3	Sujet	89
3.1	La variation morphosyntaxique en français	89
3.1.1	Définitions	89
3.1.2	La variation diatopique	91
3.1.3	La variation diachronique	92
3.1.4	La variation diastratique	93

Sommaire

3.1.5	La variation diaphasique	94
3.1.6	La variation diamésique	94
3.1.7	La variation libre	96
3.1.8	Variation entre idiolectes et accomodation	97
3.2	Constat initial et problématique retenue	99
3.3	Hypothèses écartées	100
3.3.1	Les bons et les méchants	100
3.3.2	La variation en question est à corrélérer à l'opposition parlé – écrit	102
3.3.2.1	Reformuler parlé-écrit en spontané-planifié	102
3.3.2.2	Ce qui s'explique	103
3.3.2.3	Ce qu'on ne cherche pas à expliquer	104
3.3.2.4	Les limites de l'opposition spontané-planifié	104
	Les interrogatives	105
	La parataxe	106
3.3.2.5	Synthèse	108
3.3.3	La variation est sociolinguistique	108
3.3.4	Vers une troisième voie	109
3.4	Hypothèse adoptée : la diglossie	110
3.4.1	Définition (stricte)	110
3.4.2	Application à la situation française contemporaine	113
3.4.2.1	L'aspect proprement linguistique	113
3.4.2.2	L'aspect proprement sociolinguistique	114
3.4.3	Les choix terminologiques	118
3.4.4	Les zones zribi-hertziennes	121
3.5	Conclusion	122
4	L'argumentation	125
4.1	Soulager le théoricien et le typologue	126
4.2	Soulager le typologue et l'empiriste	127
4.3	Conclusion	128

II Morphosyntaxe du nombre dans le domaine no-

minal	131
5 Morphologies du nombre en français	137
5.1 Les déterminants	137
5.1.1 Les suffixes synthétiques	138
5.1.2 Le suffixe de pluriel	139
5.1.3 Synthèse	140
5.2 Les adjectifs pré-nominaux	140
5.3 Les noms et les adjectifs post-nominaux	144
5.3.1 Le suffixe de pluriel	144
5.3.2 Les formes supplétives lexicales	146
5.4 Variété et variation dans la morphologie du nombre . . .	148
5.4.1 Variété	148
5.4.2 Variation	150
5.4.3 Conclusion	151
6 Les statuts morphosyntaxiques du nombre	153
6.1 Un modèle morphologique : Barra Jover (2008)	154
6.1.1 Le modèle espagnol	155
6.1.2 Le modèle anglais	157
6.1.3 Le modèle poyaudin et le modèle basque	158
6.2 L'interface syntaxe-sémantique du nombre	161
6.2.1 L'actualisation	161
6.2.2 Actualiser sans déterminant : les noms nus	163
6.2.3 Corréler la morphosyntaxe du nombre et l'actuali- sation par les NN	165
6.2.3.1 Dobrovie-Sorin et Laca (2003) sur les NNE	165
6.2.3.2 Dobrovie-Sorin et Laca (2003) sur les NNG	167
6.2.3.3 Corréler la sémantique à la syntaxe . . .	168
6.2.3.4 Corréler la syntaxe à la morphologie . .	170
6.2.3.5 Synthèse des statuts morphosyntaxiques du nombre	176
6.3 Enfin, le français	179
6.3.1 Le statut du nombre du FD	180
6.3.1.1 Représenter le DP du FD	181
6.3.1.2 Quel statut pour le nombre du FD ? . . .	185

Sommaire

6.3.2	Le statut du nombre du FCT	186
6.3.2.1	Si le nombre du FCT avait un statut . . .	186
6.3.2.2	Et si le nombre du FCT n'avait pas de statut	192
6.3.2.3	Le français écrit : une langue naturelle?! . . .	193
6.4	Synthèse	194
6.4.1	Résultats	194
6.4.2	Le nombre et l'hypothèse de la diglossie	195

III Éléments de syntaxe 197

7 La négation du verbe 199

7.1	Introduction : (n) fet pa atãsiõ a nẽ	199
7.2	La négation verbale et la négation de constituant	203
7.2.1	La négation du verbe	203
7.2.2	La négation verbale	204
7.2.3	La négation de constituant	207
7.2.4	Les deux syntaxes de 'V pa'	208
7.3	La négation, verbale ou de constituant?	212
7.3.1	L'autonomie de <i>pa</i> sur le verbe	213
7.3.2	Négation et quantification universelle	216
7.3.2.1	Sans <i>n</i>	216
7.3.2.2	Avec <i>n</i>	219
7.3.3	L'impératif	220
7.4	Synthèse	222
7.4.1	Les propositions	222
7.4.2	<i>pa</i> et les zones zribi-hertziennes	223
7.4.2.1	Le grand diglosse	223
7.4.2.2	Le petit diglosse	224
7.4.2.3	Le chaotique	225
7.4.3	Des aspects que ce chapitre n'a pas traités	226
7.4.3.1	La diachronie	226
7.4.3.2	<i>pa</i> et les autres	226
7.4.3.3	Mélanger la négation et d'autres variables	226

8 Syntaxe et structure informationnelle en variation 229

8.1	la dislokasiõ è-ser a kwa ?	229
8.2	Les travaux de Lambrecht : SI et topique en français . . .	232
8.2.1	Terminologie	232
8.2.2	Lambrecht (1994) : généralités sur la SI et le topique	233
8.2.2.1	Généralités	233
8.2.2.2	Le topique	234
8.2.2.3	L'identifiabilité des référents	236
8.2.3	Le topique en français	236
8.2.3.1	Topique et Anti-Topique	237
8.2.3.2	Les constructions relatives de présentation	239
8.2.3.3	Les propositions Tout-Focus	240
8.2.3.4	Les propositions Arrière-Plan	240
8.2.3.5	Le type (T) pro-V (X) (AT)	242
8.3	Syntaxe, SI et variation	244
8.3.1	La norme face aux allophrases de Lambrecht . . .	244
8.3.2	La variation autour de la SI	245
8.3.2.1	Corréler variation et SI	245
8.3.2.2	La SI en variation diamésique	246
8.3.2.3	La SI en variation sociolinguistique . . .	247
8.4	Le français, SV(O) ou (T) pro-V (X) (AT)?	248
8.4.1	Le petit diglosse	249
8.4.2	Le grand diglosse	250
8.4.3	Le chaotique	250
8.4.4	Et la dislocation ? Et le clitique sujet ?	251

IV Est-on cohérents ? 253

9 Quoi chercher ? 257

9.1	But et méthode	257
9.1.1	But	257
9.1.2	Méthode	260
9.1.2.1	La méthode du conditionnel (Barra Jo- ver, 2007)	260
9.1.2.2	Auto-observation du schéma diglossique	262

Sommaire

9.1.2.3	Observation sur corpus du schéma diglossique	265
9.1.2.4	Qualitatif et quantitatif	266
9.2	Une cohérence grammaticale auto-observée	268
9.2.1	/pa{z}/ à /pA/	268
9.2.2	Et PLUS	271
9.2.3	PAS, PLUS... et NE	271
9.2.3.1	$Z \times Neg$	271
9.2.3.2	La difficulté de boucler la boucle	273
9.2.4	Synthèse	274
10	Étude de corpus	277
10.1	Présentation du corpus HH	278
10.2	Les variables retenues	281
10.2.1	La négation du verbe	282
10.2.2	La première personne du pluriel	283
10.2.3	Le clitique sujet impersonnel	284
10.2.4	Le sujet lexical	284
10.2.5	Une liaison	292
10.3	Mélangeons	293
10.3.1	$Neg \times 1pl$	294
10.3.2	$Neg \times Expl$	295
10.3.3	$Neg \times SujLex$	298
10.3.4	$Etre \times SujLex$	304
10.3.5	$1pl \times SujLex$	305
10.3.6	$1pl-2 \times Etre$	307
10.3.7	Le choix de la proposition comme unité de cohérence	308
10.4	Synthèse	310
10.4.1	La cohérence du FCT et du FD de UE	310
10.4.2	Des agrammaticalités révélées par l'étude	311
10.4.3	Quelques questions ouvertes	313
10.4.3.1	HH, un 'petit' corpus	313
10.4.3.2	Le domaine de portée des variables	314
10.4.4	Conclusion	316
11	Clashes	319

11.1 Un flagrant délit de mélange interdit	319
11.2 S'expliquer le contre-exemple	321
11.3 Un scénario reproductible	322
11.4 L'écrit est décidément particulier	322
11.5 Définitions empiriques du FCT et du FD	325
Conclusion	327
Bilan du programme de la thèse	329
Hypothèses générées	331
Anomalies générées	332
Manques	333
Ouvertures	334
Quel futur pour les données du FD ?	335
Le mot de la fin	335
Le second mot de la fin	339
Bibliographie	341
Annexes	351
A Présentation sommaire du poyaudin	353
Nomenclature	353
La Puisaye	353
Le poyaudin	354

Table des figures

1.1	La boîte noire du langage	23
1.2	La boîte noire du langage et ses multiples interfaces . . .	24
1.3	La <i>stricte</i> boîte noire du langage	25
1.4	La <i>pure interface</i>	27
1.5	La <i>boîte blanche</i>	27
1.6	Un modèle modulaire (restreint) de la faculté de langage	29
1.7	Le français, une plante qui a crû, et qui donc décroitra! .	32
2.1	Unilinéarité, multilinéarité, non-linéarité	39
2.2	Langue support et encodage secondaire	40
2.3	Correspondances graphèmes-sons dans le mot CUIR . . .	55
3.1	Les zones 1, 2 et 3 de Zribi-Hertz	122
9.1	Les variantes de mon FCT et de mon FD, auto-observées .	274
10.1	Le FCT et le FD de UE d'après le corpus HH	310

Liste des tableaux

1.1	Deux points de vue sur la langue	20
2.1	La communication par le langage naturel	36
2.2	La ‘liste de Martinet’	37
2.3	Réviser la ‘liste de Martinet’ pour les LS?	38
2.4	Les phonèmes consonantiques du français	82
2.5	Les phonèmes vocaliques du français	84
2.6	Les symboles pour transcrire les voyelles	84
5.1	Les déterminants du français (formes phonologiques) . . .	141
5.2	Les liaisons post-nominales et le niveau de langue (Ågren, 1973:127)	146
5.3	Les noms à pluriel supplétif	147
5.4	Les ‘noms et adjectifs en -al/-aux’	149
6.1	Quelques déterminants du poyaudin	159
6.2	L’interface morphologie-syntaxe-sémantique et le nombre	178
8.1	La SI en variation et les zones zribi-hertziennes	249
9.1	La variable <i>Z</i> de PAS et le « niveau de langue », Ågren (1973:102)	272
9.2	Les grammaires partielles de mon FCT et de mon FD . . .	275
10.1	<i>Neg</i> × <i>1pl</i>	294
10.2	<i>Neg</i> × <i>(il)ya</i>	296
10.3	<i>Neg</i> × <i>(il)faut</i>	296
10.4	<i>Neg</i> × <i>SujLex</i>	299
10.5	<i>Etre</i> × <i>SujLex</i>	304
10.6	<i>1pl</i> × <i>SujLex</i>	306

Liste des tableaux

10.7 $1pl-2 \times Etre$ 307

Conventions typographiques

- Entre guillemets simples, les expressions ‘familiales’.
- Les exemples sans glose sont transcrits dans le système de transcription phonologique présenté à la section 2.5 pour le français et le poyaudin¹, et dans la graphie officielle en caractères penchés pour les autres langues.

- (1) a. *ẽn egzãpl d frãse*
b. *en ot ã pwajodẽ*
c. *a last one in English*

- Les exemples glosés ajoutent une ligne de gloses dans une police plus petite et une éventuelle traduction entre guillemets simples.

- (2) *zum* *Beispiel*
PRÉP.le_{DAT} exemple
‘par exemple’

- Les gloses sont soit des mots, soit des informations grammaticales, lesquelles sont alors en petites majuscules. Un tiret sépare les morphèmes. Si un mot est glosé par deux mots ou traits grammaticaux, alors ils sont séparés par un point.
- La liste des informations grammaticales comprend les abréviations

1. Il s’agit d’un parler d’oïl, qui sera évoqué à plusieurs reprises. Je le présente rapidement à l’annexe A.

- suivantes : 1-2-3 : 1^{re}, 2^e ou 3^e personne ; ART : article ; AUX : auxiliaire ; DAT : datif ; DÉM : démonstratif ; DÉT : déterminant ; ERG : ergatif ; FÉM : féminin ; GÉN : génitif ; NOM : nominatif ; PASS : passé ; PL : pluriel ; SG : singulier ; SUJ : marqueur sujet ; TOP : topique ;
- Les constituants syntaxiques de type XP sont des sigles anglais : AP : Adjective Phrase ; CoordP : Coordination Phrase ; DP : Determiner Phrase ; DemP : Demonstrative Phrase ; IP : Inflexion Phrase ; NP : Noun Phrase ; NegP : Negation Phrase ; NumP : Number Phrase ; PolP : Polarity Phrase ; QP : Quantifier Phrase ; VP : Verb Phrase.
 - Les éléments saillants PROSODIQUEMENT sont en petites majuscules s'ils sont orthographiés, et suligne s'ils sont transcrits phonologiquement.
 - Un exemple comme *Beispiel*, introduit directement dans le texte, est en caractères penchés, éventuellement transcrit phonologiquement.
 - « Entre guillemets français », on trouve les propos rapportés. S'ils ne sont pas accompagnés d'une référence explicite, alors il s'agit de propos folkloriques ou personnels.
 - L'italique marque une *emphase*.
 - L'italique gras marque une ***notion importante***.
 - Les petites majuscules font référence explicitement à l'ORTHOGRAPHE d'un mot.
 - Le style machine à écrire dénote le CONCEPT associé à un mot ou une expression, ou l'ENTRÉE lexicale d'un mot.
 - Les exemples sont parfois accompagnés de jugements de grammaticalité. L'étoile * indique un énoncé mal formé, agrammatical. Les points d'interrogation, doubles (??) et simples (?), signalent des degrés moins forts d'agrammaticalité. Le dièse # signale un énoncé ininterprétable, ou inadéquat dans le contexte discursif où l'énoncé apparaît. Et le symbole de pourcentage % signale un énoncé bien

formé pourvu d'un sens, mais pas le sens voulu.

Introduction

Le vif du sujet

C'est des bêtes de phrases,
t'as vu, je comprends rien,
elles sont trop longues.

Krimo, dans L'esquive

Pour entrer rapidement dans le vif du sujet, observons en (1) un extrait du film *L'esquive*, d'Abdelatif Kechiche (70^e min du film). Dans cet extrait, deux adolescents de 15 ans, Krimo et Lydia, répètent une scène de théâtre devant leurs camarades de classe et leur professeur de français. La pièce en question est *Le jeu de l'amour et du hasard* de Marivaux (1730). La scène confronte Arlequin (joué par Krimo) et Lisette (jouée par Lydia). Les deux personnages sont tombés amoureux plus tôt dans la pièce et se retrouvent dans cette scène après qu'ils ne se soient pas vus pendant quelques temps. Comme l'indique l'intervention de l'enseignante, Krimo ne fait que réciter péniblement des mots appris par cœur. Au contraire, Lydia joue parfaitement son personnage.

(1) Krimo : *Enfin ma reine, je vous vois et je ne vous quitte plus, car j'ai trop pâti d'avoir manqué de votre présence et j'ai cru que vous esquiviez la mienne.*

Lydia : *Il faut vous avouer, monsieur, qu'il en était quelque chose.*

Introduction

Krimo : *Comment donc ma chère âme, élixir de mon cœur, avez-vous entrepris la fin de ma vie ?*

Lydia : *Non mon cher, la durée m'en est trop précieuse.*

Krimo : *Ah que ces paroles me fortifient.*

l'enseignante (très agacée) : [...] *Tu me dérites tes répliques y a pas de cœur y a rien là. « Ah ! que ces paroles me fortifient ! » on l'a déjà vu ça. [...] Il faut vous av- il en était quelque chose monsieur, oui c'est vrai je vous esquivais. Ah bon vous m'esquiviez ? Mais pourquoi ? Vous voulez plus me voir ?*

Comment comprendre l'intervention de l'enseignante ? Tout d'abord, elle veut voir Krimo jouer la comédie et non réciter un texte. Elle l'enjoint donc de mettre en pratique le jeu qui a été discuté auparavant (*on l'a déjà vu ça*). Puis, elle repose le problème : manifestement, non seulement Krimo ne parvient pas à jouer un personnage, mais en plus, il ne comprend pas ce qui se dit. Donc elle se lance dans la traduction des répliques de la scène. Ainsi, elle traduit *il faut vous avouer, monsieur, qu'il en était quelque chose* par *oui c'est vrai je vous esquivais*, puis *Comment donc ma chère âme, élixir de mon cœur, avez-vous entrepris la fin de ma vie ?* par *Ah bon vous m'esquiviez ? Mais pourquoi ? Vous voulez plus me voir ?*

Plusieurs points ici rencontrent directement le propos de mon travail. Tout d'abord, il est possible qu'un Français de 15 ans ne comprenne pas le français classique. C'est pourtant peu ou prou ce français classique (ou son descendant proche, le français standard actuel) qui est enseigné en France dès l'apprentissage de la lecture.

Ensuite, cette enseignante de français reste longtemps dans l'incompréhension du problème de Krimo (commencée bien avant cette scène précise), avant d'envisager qu'il s'agisse avant tout d'un problème de

compréhension. Le heurt est fort entre l'imaginaire de l'enseignante et la réalité de la langue de l'adolescent. Quand enfin l'enseignante comprend le problème (et je dirais que c'est tout à la gloire du personnage du film), que fait-elle exactement ? On pourrait s'attendre à une simple traduction des mots qui ont changé de sens ou dont l'emploi est obsolète ou rare aujourd'hui (*pâtir, esquiver, élixir...*).

Mais le trait le plus frappant à mes yeux est le basculement immédiat dans une autre grammaire. Dans les répliques, on repère plusieurs traits grammaticaux marquant un français formel. On trouve la construction impersonnelle *il en était quelque chose*, la négation avec *ne* dans *je ne vous quitte plus*, l'inversion dans l'interrogative *avez-vous entrepris la fin de ma vie ?* et l'exclamation introduite par (*ah*) *que* dans *ah que ces paroles me fortifient*.

À l'opposé, les traductions que donne l'enseignante proposent notamment deux variantes grammaticales informelles à ces structures : l'interrogation sans inversion, à deux reprises (*vous m'esquiviez ?* et *vous voulez plus me voir ?*), et la négation sans *ne* (*vous voulez plus me voir ?*). C'est bien ce changement grammatical qui me fait parler de traduction dans la langue de Krime, et pas seulement d'adaptation à son « langage des banlieues » (les traductions proposées par l'enseignante n'ont d'ailleurs rien d'un tel langage).

Un des thèmes du *Jeu de l'amour et du hasard* est justement le langage. Lisette et Arlequin sont des servants qui se font passer pour des bourgeois. Pour cela, ils en viennent entre autres à imiter le langage des bourgeois. Puisqu'ils ne le maîtrisent pas, ils produisent des formes burlesques : *j'ai trop pâti d'avoir manqué de votre présence*. Krime est à des années-lumières de pouvoir saisir cet enjeu du langage de la pièce, et de pouvoir en jouer et représenter un personnage qui se ridiculise en imitant un autre langage que le sien. Comment le pourrait-il seulement, alors que lui-même

Introduction

ne saisit pas cet autre langage (je dirais même cette autre langue) qu'est l'écriture de Marivaux ?

A contrario, alors que l'école impose à Krimo (l'étude d') une langue qui n'est pas la sienne, est-ce que l'on peut imaginer qu'elle se penchera sur sa grammaire à lui ? Est-ce que Krimo aura un jour la chance de découvrir la grammaire cachée par exemple derrière ses expressions telles que (2)² (exemples tous tirés du film) ?

- (2) a. *des bêtes de phrases*
'des phrases très compliquées'
- b. *un bête de film*
'un très bon film'
- c. *une bête de meuf*
'une fille géniale'

Est-ce qu'il saura que ces exemples entrent dans un schéma grammatical apparemment de plus en plus productif où *bête* est simplement un nouveau nom d'une classe de noms pouvant prendre un rôle prédicatif dans la première position d'un syntagme nominal de la forme *Dét N₁ de N₂* sans imposer leur genre grammatical au déterminant, comme le font déjà *espèce* et *putain* (cf. (3) et (4)) ?

- (3) a. *des espèces de phrases*
'des phrases d'un genre particulier'
- b. *un espèce de film*
'un film d'un genre particulier'
- c. *une espèce de meuf*

2. Les paraphrases que je propose ne rendent que partiellement justice aux sens des expressions en (2), (3) et (4). Et je ne connais pas de nom de la classe de noms décrite ci-après (*bête*, *espèce*, *putain*, ...) qui soit masculin.

‘une fille d’un genre particulier’

- (4) a. *des putains de phrases*
‘des phrases incroyables’
b. *un putain de film*
‘un film génial’
c. *une putain de meuf*
‘une fille géniale’

Est-ce qu’il apprendra que la stigmatisation de l’absence de la conjonction de subordination *que* dans (5) (exemples également tirés du film) est arbitraire, puisque cette même absence est parfaitement standard en anglais et en allemand³ (cf. (6) et (7)), et qu’elle est même franchement châtiée en espagnol⁴ (8) ?

- (5) a. *Je savais pas t’allais me demander ça.*
b. *Je croyais on était potes.*
c. *J’imaginai pas t’allais me demander ça.*
d. *Tu crois tout de suite là je peux te dire la réponse ?*
e. *Y en a ils me disent t’es en chien sur une meuf.*
- (6) a. *I didn’t know you were going to ask me that.*
b. *I thought we were friends.*

3. La classe de verbes qui permettent l’emploi sans la conjonction de subordination en allemand est un peu différente, et la négation *y* joue un rôle important.

4. La classe de verbes qui permettent cette absence en espagnol est limitée aux verbes de souhait, et cet emploi se confine presque systématiquement aux lettres intimes. Au contraire, ce qui est stigmatisé (« ça fait plouc » selon mon informateur) est le fait d’ajouter la préposition *de* (phénomène de *déquéisme*).

- (i) *Pienso (de) que todo va a salir bien.*
je.pense de que tout va à réussir
‘Je pense que tout va bien marcher.’

Introduction

[...]

- (7) a. *Ich wußte, du würdest es mich fragen.*
je savais tu AUX ça me demander
'Je savais que tu allais me demander ça.'
- b. *Ich dachte, wir wären Freunde.*
je pensais nous serions amis
'Je croyais que nous étions amis.'

[...]

- (8) a. *Espero estés bien.*
j'espère tu.sois bien
'J'espère que tu vas bien.'
- b. *Deseo vengas.*
je.souhaite tu.viennes
'Je souhaite que tu viennes.'

Le contenu de la thèse

Le cadre et la problématique

La situation décrite ci-dessus demande quelques précisions. Tout d'abord, le français de Marivaux est distinct du français standard actuel (ce que j'appellerai plus tard le *français classique tardif*, et qui fera partie de mon objet d'étude, contrairement au français de Marivaux). Et il n'est pas évident que le personnage de Krimo soit aussi incapable de comprendre ce français standard actuel que le français classique de Marivaux. Malgré cela, il semble clair que Krimo ne maîtrise pas la langue que l'école lui enseigne. Cette situation est assez extrême, mais sûrement pas exceptionnelle.

Mais la situation qui occupera une place centrale dans cette thèse est différente. Il s'agit de la cohabitation d'un français courant (que je nom-

merai ultérieurement le *français démotique*), tel le français de Krimo, et du français standard actuel, au sein des productions d'un même locuteur du français. En effet, d'une part, de nombreux locuteurs du français acquièrent effectivement la langue de l'école et sa norme standard, et d'autre part, même parmi ces locuteurs, très peu (virtuellement aucun) ne s'expriment jamais dans un français courant, c'est-à-dire un français avec des caractéristiques grammaticales officiellement exclues de la norme actuelle du français standard. Un point important de cette thèse sera de formuler un cadre terminologique et théorique précis pour énoncer des hypothèses concernant cette situation et sa description par la linguistique.

Tant du point de vue grammatical que du point de vue sociolinguistique, la situation de variation qui nous occupe est connue des linguistes depuis plusieurs décennies. Plusieurs écoles variationnistes se proposent d'en rendre compte au sein de la communauté scientifique des linguistes. J'en citerai trois.

La première que je citerai est l'école du *français parlé*. Sa principale préoccupation est de mettre en avant les productions orales des locuteurs. L'ignorance de ces productions, notamment les productions de ce que j'ai provisoirement appelé du français courant, est supposée être la cause de la description du français standard seul, et de l'oubli des caractéristiques grammaticales des productions non-normées. De plus, cette école émet l'hypothèse que le fait que le français courant existe principalement dans les productions *orales* peut avoir une force explicative quant à la forme même de sa grammaire.

La deuxième école est l'école *sociolinguistique*. Son objectif principal est de décrire et corrélérer l'emploi des variantes grammaticales du français à des facteurs sociolinguistiques, sous forme de réponses à la question « dans quelle situation est-ce qu'un locuteur emploiera une certaine va-

Introduction

riante ? » ou bien « qu'est-ce qui, dans une situation donnée, permet ou favorise, ou bien empêche ou défavorise l'emploi d'une variante donnée ? »

Parallèlement à leurs programmes respectifs (décrire toutes les productions orales, et décrire les conditions de production de chaque variante), ces deux écoles proposent de modéliser la grammaire des Français de façon à ce qu'elle produise toutes les variantes rencontrées, en faisant de cette modélisation une unique et grande grammaire variationniste.

La troisième école est celle que j'appellerai l'école de *la diglossie*. Son principal point de divergence avec les deux autres écoles est qu'elle suppose que les locuteurs qui produisent à la fois des variantes normées et des variantes non-normées parlent deux langues, et non une seule. C'est ce que le terme de *traduction* que j'ai employé au début de cette introduction voulait suggérer. Dans cette hypothèse de la diglossie, le fait que les deux langues en question soient très semblables, au point que de nombreuses variantes soient considérées à la fois standard et courantes, est responsable de l'illusion, chez les locuteurs, d'avoir à faire à une seule langue. C'est la voix de cette école que j'aimerais porter ici. Et c'est par des arguments strictement grammaticaux que j'aimerais convaincre le lecteur que le français standard et le français courant n'ont pas la même grammaire, même lorsqu'un seul locuteur les parle tous les deux.

Les considérations présentées jusqu'ici seront l'objet de la première partie (*Les protagonistes*). Le chapitre 1 (*Objet*) donnera à voir une façon de cerner l'objet de la linguistique (la langue et sa grammaire). Le chapitre 2 (*Tangentes*) explorera le contenu de l'opposition souvent citée entre la langue écrite et la langue orale, la pertinence duquel contenu sera contestée, avant de proposer une description critique de l'orthographe du français puis d'exposer le système de transcription à caractère phonologique que j'adopterai pour l'ensemble des productions de français qui

apparaîtront dans cette thèse. Le chapitre 3 (*Sujet*) sera consacré à l'exposé du sujet de la thèse, la variation morphosyntaxique intra-locuteur et ses modélisations, ainsi qu'à l'hypothèse retenue pour cette modélisation, la diglossie. Enfin, le chapitre 4 (*L'argumentation*) présentera l'organisation des arguments développés dans la suite.

Les arguments théoriques et typologiques

La partie II et la partie III apporteront des arguments typologiques et théoriques pour convaincre que postuler une diglossie permet une description plus adéquate des données de variation du français dans le langage de la linguistique.

La partie II (*Le nombre dans le domaine nominal*) consacrera un chapitre à la description des faits de morphologie du nombre à l'intérieur des syntagmes nominaux canoniques (chapitre 5, *Morphologies du nombre en français*), et un chapitre à la comparaison du nombre du français avec le nombre d'autres langues à travers les lunettes d'un petit modèle typologique des noms nus (chapitre 6, *Les statuts morphosyntaxiques du nombre*).

La partie III (*Éléments de syntaxe*) sera divisée en deux chapitres. Le chapitre 7 (*La négation du verbe*) s'attachera à convaincre que la négation du français courant (c'est-à-dire la négation sans *ne*) est mieux décrite si l'on souhaite la décrire avec d'autres moyens que ceux utilisés pour décrire la négation du français standard (celle avec *ne*).

Le chapitre 8 (*Syntaxe et structure informationnelle en variation*) fait d'abord le constat que de nombreuses alternatives sont possibles lorsque le sujet d'une proposition est exprimé lexicalement et est le topique grammatical de cette proposition. Par exemple, en (9b), la dislocation à gauche alterne avec la construction SV(O) de (9a).

Introduction

(9) l'enseignante de Krimo :

- a. *Les riches tombent amoureux de qui dans la pièce ?*
- b. *Et les pauvres ils tombent amoureux de qui dans la pièce ?*

Le constat de ces alternatives sera ensuite interprété en lien avec l'organisation de la structure informationnelle des propositions, et en particulier avec l'identifiabilité du référent du sujet lorsqu'il est le topique de la proposition. Puis, puisque, comme en (9), il n'en reste pas moins vrai que des situations identiques peuvent être rendues avec SV(O) comme avec une alternative (ici la dislocation à gauche), on proposera que cette variation soit également une variation entre deux grammaires.

Les arguments empiriques

La dernière partie (partie IV, *Est-on cohérents ?*) est décisive dans l'argumentaire. Le chapitre 9 (*Quoi chercher ?*) remarque que, si les Français parlent vraiment deux langues distinctes, alors ils ne peuvent en activer qu'une seule à la fois. C'est-à-dire qu'ils ne sont pas censés mélanger des variantes seulement présentes dans le français standard et des variantes interdites dans la norme.

L'énoncé cité plus haut de la prof de français de Krimo (*vous voulez plus me voir ?*) illustre cela. En effet, on trouve deux variantes interdites dans la norme, la négation sans *ne* et l'interrogation sans inversion. L'hypothèse de la diglossie prédit que, puisque la locutrice a activé sa grammaire courante, elle n'aurait pas pu mélanger l'une de ces deux variantes avec la variante normée de l'autre. C'est-à-dire qu'elle n'aurait pas dit *vous ne voulez plus me voir ?* ni *voulez-vous plus me voir ?* Et si elle avait au contraire activé sa langue standard, alors elle aurait seulement pu dire *ne voulez-vous plus me voir ?*

Le chapitre 10 (*Étude de corpus*) propose de vérifier ces prédictions

sur un corpus oral spontané où le locuteur étudié produit à la fois du français standard et du français courant. On constatera qu'il n'active effectivement qu'*une grammaire à la fois*, puisqu'à l'intérieur de chacune de ses propositions, aucune variante non-courante n'apparaît avec aucune variante non-standard.

Le chapitre 11 (*Clashes*) constatera qu'il existe pourtant bien des cas manifestes de mélanges de variantes contradictoires, laissant ainsi penser que l'on peut activer deux grammaires à la fois, ce qui rend caduque l'idée même de double-grammaire. Pourtant, je proposerai de comprendre comment la situation particulière dans laquelle ces exemples ont été trouvés (ce sont des écrits informels) permet de suggérer que le mélange supposé interdit est le fruit d'une interférence entre l'activité d'écriture et l'informalité de la situation. En particulier, la situation d'écriture permettrait au locuteur-scripteur de contrôler et normaliser partiellement ses propres productions, ce qu'une situation parfaitement spontanée ne permet pas.

La conclusion

La conclusion tirera évidemment le bilan de chacun des arguments exposés, et évaluera leurs forces de conviction dans le cadre de l'hypothèse de la diglossie. On verra que les considérations de la partie II sur le nombre montrent que la morphologie du nombre comme son statut dans l'ensemble de la grammaire est très similaire en français standard et en français courant. Ceci rend l'hypothèse diglossique trop forte : postuler deux langues alors qu'elles sont presque identiques est inutile.

Par contre, la partie III aura montré que des considérations sur la syntaxe sont beaucoup plus convaincantes et que l'hypothèse d'une double-grammaire permet une description plus simple et plus adéquate des données concernant la négation du verbe et la syntaxe associée à la structure informationnelle.

Introduction

Ensuite, on reviendra sur l'argument majeur de cette thèse, objet de la partie IV. Dans les faits, les deux grammaires des locuteurs sont une réalité. En effet, ils produisent bien du français standard et du français courant, mais n'activent bien qu'une de ces deux langues au sein d'une proposition.

Puis, on présentera sous forme de listes les hypothèses avancées dans cette thèse, et des anomalies que ces hypothèses génèrent. Suivront quelques considérations sur les manques de ce travail, et quelques propositions d'ouverture sur d'autres thèmes de recherche.

Finalement, un article d'Alain Kihm paru dans *Libération* en janvier 2008 viendra faire écho à l'extrait du film *L'esquive* présenté au début de cette introduction.

Quelques absents de la thèse

Deux questions sont absentes de ce travail. La première est une limite que je me suis donnée alors que cette question a bien un rôle à jouer dans le débat. Il s'agit de l'adéquation de l'hypothèse de la diglossie au sein des considérations strictement sociolinguistiques autour de la variation en français. La seconde limite est une limite naturelle, puisque je ne proposerai pas d'éclairage sur l'enjeu *extra*-linguistique (c'est-à-dire l'enjeu de société) du constat des linguistes de la variation morphosyntaxique non-standard du français.

Et mon travail se limite à des enjeux franco-français.

La sociolinguistique

L'hypothèse de la diglossie suppose une situation sociolinguistique très particulière. Je présenterai seulement brièvement cet enjeu au chapitre 3. Je laisserai au lecteur le soin de se faire sa propre opinion, et je me res-

treindrai bien à des questions purement grammaticales. Le point principal est de savoir si la variation étudiée est bipolaire (une variante est \pm formelle — ou \pm normée, ou toute autre conceptualisation de cette bipolarité) ou bien multipolaire (une variante est *recherchée ou ordinaire, soutenue, courante ou familière, bourgeoise ou populaire, centrale ou régionale ...*).

Le débat externe à la linguistique

La variation morphosyntaxique en français est le constat qui réunit tous ceux qui veulent, au sein de la linguistique, se pencher sur ces productions hors de la norme et sur leurs structures grammaticales. Antérieurement même au débat entre les écoles variationnistes du français, on est vivement tenté de se demander quel rôle la linguistique variationniste et son constat pourraient jouer face à la situation d'ignorance, ou de déni, voire de mépris, de la réalité de la grammaire de la langue quotidienne des Français de la part de ceux que Kihm (2008) appelle « ceux qui font les promotions ».

C'est bien le débat interne à la linguistique qui sera donné à voir dans cette thèse, et non le constat que je viens de citer, lequel constat, je crois, les linguistes dans leur ensemble s'accordent à déplorer. Donc, à ce titre, l'entrée en matière comme le final de cette thèse seront des entorses à cette volonté de ne pas mettre les pieds dans le monde extra-linguistique.

Le français de France

Enfin, il faut préciser que ce travail se limite strictement à l'étude du 'français des francophones de France'. C'est-à-dire qu'il ne sera pas question de prétendre que les hypothèses avancées sont applicables telles quelles aux situations de confrontation entre le français et des langues et dialectes distincts de celui-ci (les patois, les langues régionales, les créoles, ...), ni entre le français standard et des variétés de français hors de la

Introduction

France métropolitaine (le français québécois, le français d’Afrique, ...).

Attention, polémique !

Il prêche le faux
pour savoir le vrai.

dicton populaire

Dans l’ensemble de cette thèse, j’adopte un ton polémique. J’aimerais que le lecteur ne se méprenne pas sur mes motivations. D’abord, c’est simplement une écriture que je maîtrise mieux que l’écriture académique classique. Je m’y sens plus à l’aise. Ensuite, cela s’inscrit dans une démarche d’interaction. Je cherche des réponses à mes questions. Et en formulant ces questions et les réponses que je propose en termes polémiques, je souhaite avant tout trouver un moyen pour mettre au jour et identifier mes idées et ce qui peut les opposer à d’autres idées et d’autres réponses. C’est simplement le ton grâce auquel j’ai l’impression de le mieux me vérifier. C’est un garde-fou dans le sens où, je crois, les défauts, les manquements, et les contradictions de mes propos ne pourront pas se cacher derrière ce que je ressentirais comme une diplomatie que je ne maîtrise pas.

Première partie

Les protagonistes

DANS cette première partie, je cherche à poser le plus clairement possible les fondements théoriques et techniques de ma thèse.

Dans le chapitre 1, je définis l'objet de mon étude (la langue française) comme il sera compris ensuite.

Dans le chapitre 2, je prends la tangente et fais quelques détours par des sujets qui me permettent de mieux comprendre cette fameuse opposition entre *langue écrite* et *langue parlée/orale* et de m'y positionner. Une description critique de l'orthographe du français m'amènera à la considérer comme inadéquate pour transcrire les données du français en rendant correctement justice à leur morphosyntaxe. Je présenterai alors le système de transcription que j'utiliserai par la suite, en considérant qu'il permettra de donner la commune mesure entre toutes les productions de français.

Dans le chapitre 3, j'expose le sujet précis de ma recherche — la variation morphosyntaxique intralocuteur en français. Je liste succinctement des hypothèses qui répondent à ma problématique mais qui me posent problème, en essayant d'explicitier les difficultés qu'elles présentent. Enfin, je présente l'hypothèse que je vais défendre au long de cette thèse : la **diglossie**.

Dans le chapitre 4, je présente l'organisation de l'argumentation que je vais suivre dans les parties suivantes.

Chapitre 1

Objet

Ce n'est pas en ouvrant la gorge du rossignol qu'on découvrira le secret de son chant.

Jean Cocteau

DANS ce chapitre, je présente mon objet d'étude, le français comme langue particulière, tel qu'il se postule dans mon cadre théorique de travail. Après avoir introduit la langue comme objet d'étude (section 1.1), j'introduirai la façon que j'adopte d'aborder l'objet *langue française* (section 1.2).

1.1 La langue

1.1.1 Langue externe, langue interne

Deux points de vue s'opposent et se complètent pour définir la langue. D'un côté, le point de vue externe considère la langue comme un objet social. D'un autre, on trouve le point de vue interne, qui voit dans la langue un objet cognitif, individuel. On identifie finalement fréquemment

Chapitre 1 Objet

un auteur en fonction de son point de vue, et certaines problématiques se posent mieux selon un des deux points de vue¹. (cf. tableau 1.1)

<hr/>	
langue externe	langue interne
langue — objet social	langue — objet individuel
langue collective	vs. langue individuelle
langue labovienne	langue chomskyenne
niveau macroscopique	niveau microscopique
<hr/>	

TABLEAU 1.1: Deux points de vue sur la langue

1.1.1.1 La langue externe

Pour cerner la langue externe, j'utiliserai la métaphore de la forêt, empruntée à Barra Jover (1999:52). Cette métaphore consiste à suggérer de parler des langues comme on parle des forêts.

Par exemple, concernant le passage du latin au français (et aux langues romanes en général), l'utilisation de deux identifiants distincts (*latin* et *français*) est problématique. Du coup, on est contraint de se demander quand on est passé de l'un à l'autre, alors que la réponse est « jamais » mais que le résultat est bien là : on est passé du latin à l'ancien français en Gaule dans une période allant du IV^e au IX^e siècle. C'est certainement

1. Voir le débat dans Keller (1994:125 et 133) : “Should linguists conceptualise and investigate language as language faculty, or should they, disregarding the speaker, consider language as a relatively autonomous entity?” “The distinction between I-language and E-language is useful in that it invites linguists to ask what the actual object of their study is.” [traduction (BM) :] « Est-ce que les linguistes doivent conceptualiser et investir la langue comme la faculté de langage, ou bien est-ce qu'ils doivent, sans tenir compte du locuteur, considérer la langue comme une entité relativement autonome? » « La distinction entre la langue interne et la langue externe est utile dans la mesure où elle invite les linguistes à se demander quel est l'objet exact de leur étude. »

parce qu'on conceptualise comme des entités autonomes ce qui ne devrait pas l'être (cf. Michel Banniard (c. p.) : « Il n'y a pas d'un côté le latin qui est en train de mourir dans une pièce, et de l'autre le français qui est en train de naître dans une autre pièce. »). La métaphore de la forêt aborde le problème autrement : si le latin et le français étaient des forêts, ils auraient le même nom. Il ne s'agirait que d'étiquettes désignant des états de la même forêt à des périodes distinctes. « La forêt X d'aujourd'hui est différente de ce qu'elle était il y a deux mille ans, mais il s'agit bien dans les deux cas de la forêt X ». Comme on s'imagine peu changer le nom d'une forêt au cours du temps, on pourrait choisir de garder le même nom pour désigner une langue dans toutes ses périodes différentes. Ainsi, on parlerait latin actuellement en France (ou indo-européen, ou ...). C'est d'ailleurs ce que l'on fait en conservant le terme *français* entre le IX^e siècle et la période actuelle. Les adjectifs qualifiant le français sont là pour évoquer des périodes successives de l'histoire :

L'ancien français désigne la période allant du IX^e au XIII^e siècle ;

Le moyen français désigne la période allant du XIV^e au XVI^e siècle ;

Le français classique désigne la période allant du XVII^e au XIX^e siècle ;

Le français moderne désigne la période actuelle, depuis le XIX^e siècle.

Du point de vue typologique, cette unité terminologique n'est pas justifiée : selon Böhme-Eckert (2004), le moyen français entre dans un « type à part » de celui de l'ancien français et des autres langues romanes.

1.1.1.2 La langue interne

Comme toutes les sciences, la linguistique se donne un objet d'étude (les énoncés produits dans les langues naturelles), effectue des mesures sur cet objet et cherche à en établir un modèle prévisionniste. Les éléments

Chapitre 1 Objet

mesurables sont principalement de deux natures. D'abord, on effectue des mesures physiques (c'est l'objet de la phonétique), ensuite, des 'mesures de sens' (c'est l'objet de la sémantique). Les modèles proposés veulent décrire et prédire comment des ensembles de manifestations physiques correspondent à tel ou tel sens, et inversement, pour les locuteurs d'une langue donnée.

Les linguistes générativistes ajoutent un instrument de mesure. Il s'agit du *jugement de grammaticalité*. Ce que postulent les générativistes, c'est que les locuteurs ont, parmi leur savoir linguistique, la capacité de reconnaître la bonne formation d'un énoncé dans leur langue. Il s'agit d'un postulat très controversé, surtout parce qu'il s'appuie sur la notion d'*intuition* du locuteur sur sa langue. On peut se faire une idée de son intérêt en observant l'existence d'exemples contradictoires pour lesquels les locuteurs ont envie de dire qu'ils sont mauvais tout en reconnaissant que leur sens est clair. C'est le cas de l'exemple (1a) du français, où l'on comprend nettement le sens de (1b), mais avec « quelque chose qui cloche » (l'énoncé est alors noté avec une étoile).

- (1) a. *ty vø k ty reysis
b. ty vø reysir

La langue interne, c'est donc la capacité qu'a un locuteur d'une langue donnée d'échanger des énoncés dans sa langue avec d'autres locuteurs, en étant suffisamment d'accord sur la façon dont ces énoncés doivent mettre en relation des sons et du sens. Bouchard (2003:60) :

« L'étude du langage considère traditionnellement comme évident que celui-ci est une relation entre une forme et un sens, produite par des êtres humains. »

Les linguistes appellent cela la *compétence linguistique* des locu-

teurs, qui produisent et reconnaissent des énoncés bien formés dans leur langue, un code commun de mise en relation entre le *signifiant* — les sons — et le *signifié* — le sens. La tâche scientifique des linguistes consiste à caractériser ce savoir des locuteurs. À moins de se dire connexioniste, ce qui n'est pas mon cas, on considère que l'on n'a pas accès directement au langage dans le cerveau, mais, au mieux, seulement à certaines manifestations neurologiques macroscopiques.

Pour cette raison, la caractérisation de la langue interne, c'est d'abord l'établissement d'un modèle formel qui imite et prédit le savoir des locuteurs, concernant le signifiant, le signifié, éventuellement la grammaticalité, et leurs mises en relation.

Pour ce qui est de la langue interne, on choisira d'utiliser l'image de la boîte noire (cf. figure 1.1). Une boîte noire, c'est l'idée d'un mécanisme interne pour lequel on n'a accès qu'à des manifestations externes, et lequel peut être décrit et modélisé par un algorithme qui propose de mettre en relation des données d'entrée (*input*) et des données de sortie (*output*). Dans le cas du langage, le signifiant et le signifié sont successivement l'input et l'output (modèles de compréhension et de production).



FIGURE 1.1: La boîte noire du langage

1.1.2 Autour du débat sur la syntaxe

« La forme est contrainte par les propriétés de notre système articulatoire-perceptuel (AP), et le sens est restreint à des notions relevant de notre système conceptuel-intensionnel (CI). » (Bouchard, 2003:60)

Une fois admises les contraintes imposées par les deux interfaces du langage, il reste à savoir quelles autres éventuelles contraintes jouent un rôle. En gros, les linguistes se répartissent selon leur point de vue sur la syntaxe et l'autonomie du langage sur les modules non-langagiers (cf. figure 1.2). Il en ressort trois points de vue sur le langage et la langue :

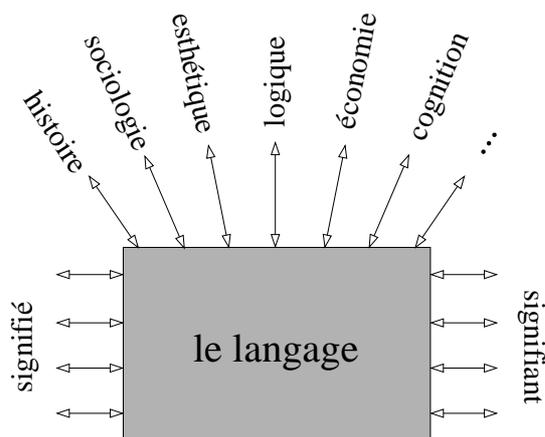


FIGURE 1.2: La boîte noire du langage et ses multiples interfaces

1. la boîte noire contient des propriétés propres et arbitraires :
 - a) pas toutes déductibles des deux interfaces : on a alors une linguistique *avec syntaxe* ; le langage est une interface entre les sons et le sens, et met en jeu des propriétés propres indépendantes de ces deux interfaces (cf. figure 1.3). Les générativistes

du programme *Principes et Paramètres* soutiennent ce point de vue. Le *Programme minimaliste* entend recentrer les propriétés formelles du langage dans les traits formels présents dans le lexique, tout en gardant comme fondement l'idée que le langage a bien des propriétés formelles propres *indépendantes*.

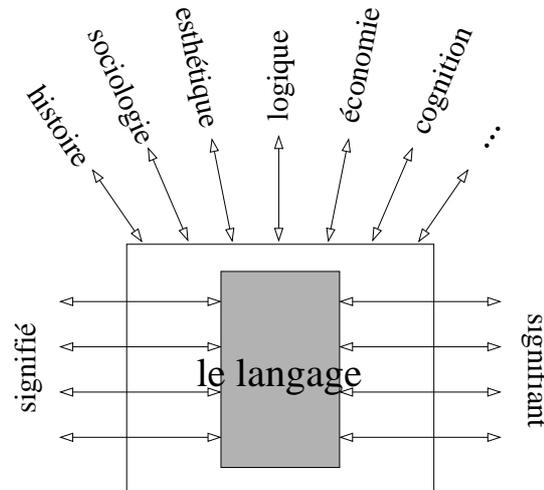


FIGURE 1.3: La *stricte* boîte noire du langage

- b) toutes déductibles des deux interfaces : on a alors une linguistique sans syntaxe ; la langue est une pure interface entre les sons et le sens (cf. figure 1.4). Bouchard (2003:61) explique ce programme :

« Une théorie linguistique basée exclusivement sur des propriétés logiquement antérieures des interfaces serait parfaite : tout découlerait de la constitution physique et cognitive des humains, et ne seraient axiomatiques que des éléments incontournables, logiquement

antérieurs ; il n’y aurait aucun besoin de poser des notions axiomatiques propres au langage. La plupart des théories considèrent que le langage s’écarte de cette perfection de façon significative. »

Sans forcément nier l’existence d’une autonomie formelle du langage, les cognitivistes veulent mettre en avant ses propriétés non autonomes. Par exemple, Haiman (1985:8, je souligne) se dit que « linguistic structures are *often* similar to non-linguistic diagrams of our thoughts, such that “selection of a point along the linguistic dimension determines and signals a point along the non-linguistic dimension”² ». Langacker (1991:1, je souligne) est plus radical : « Grammatical structures *do not* constitute an autonomous formal system or level of representation : they are claimed instead to be inherently symbolic, providing for the structuring and conventional symbolization of conceptual content.³ »

2. la boîte ne contient pas de propriétés propres, ou des propriétés non-arbitraires : on a alors une linguistique sans formalisme propre (cf. figure 1.5). C’est le point de vue d’Alain Rey : « Je pense que la fiction qui consiste à isoler la mécanique du langage de ses conditions de fonctionnement reste et restera toujours une fiction à moins qu’on n’arrive à déceler le fonctionnement matériel du langage au niveau neurologique, ce qu’on est très loin de pouvoir faire⁴ ».

2. « Les structures linguistiques sont souvent similaires aux diagrammes non-linguistiques de nos pensées, de telle façon que “la sélection d’un point le long de la dimension linguistique détermine et signale un point le long de la dimension non-linguistique” » (traduction : BM).

3. « Les structures grammaticales ne constituent pas un système formel ou un niveau de représentation autonome : on prétend au contraire qu’elles sont inhéremment symboliques, fournissant la symbolisation structurante et conventionnelle du contenu conceptuel » (traduction : BM).

1.1 La langue

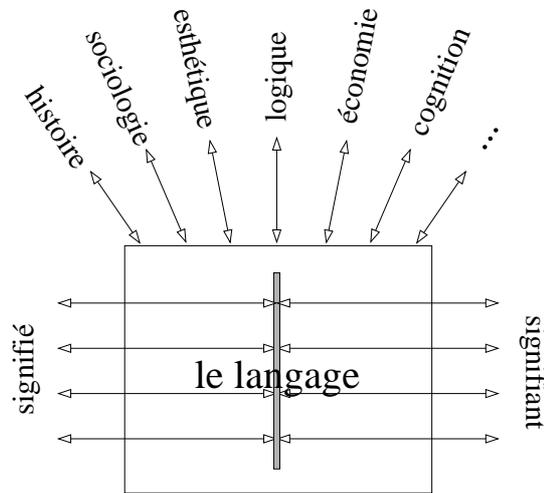


FIGURE 1.4: La *pure interface*

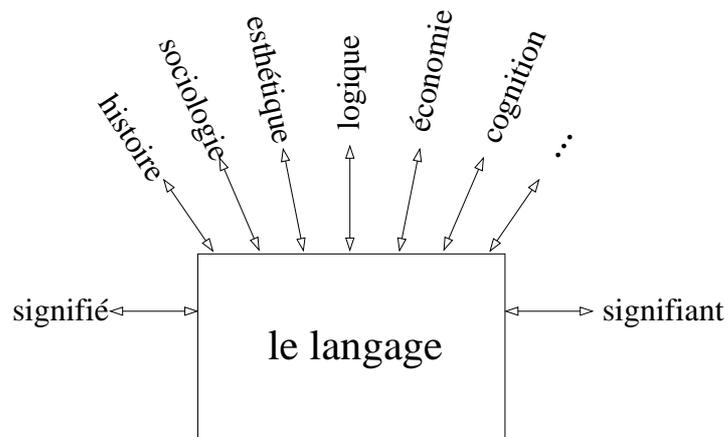


FIGURE 1.5: La *boite blanche*

4. Émission Libre-cours, France Inter, le 18 décembre 2005, en réponse à une question de François Mouret : « Un certain nombre de linguistes admettent qu'on peut décrire certains aspects de la langue, pas tous, sans avoir à se prononcer sur les inter-

1.1.3 Ici : le point de vue de la grammaire générative

La grammaire générative s'attache à décrire la langue interne. En général, on conçoit plusieurs aspects à la linguistique, qui s'articulent comme des modules. Ainsi, la figure 1.6 se lit de la façon suivante :

- la façon dont les sons sont organisés dans les langues est l'objet de la *phonologie* ;
- l'organisation linguistique du sens (des ontologies) est l'objet de la *sémantique formelle* et de la *sémantique lexicale* ;
- l'ensemble des items que l'on imagine mémorisés par les locuteurs est désigné comme étant le *lexique* ;
- et enfin, la *syntaxe* et la *morphologie* (souvent perçues comme un seul module : la *morphosyntaxe*) étudient l'agencement linguistique des autres modules.

actions de ce système avec d'autres systèmes, notamment ses structures historiques, les systèmes sociaux, les systèmes biologiques. Est-ce que c'est une position que vous considérez comme complémentaire de l'approche historique que vous avez développée, ou est-ce que vous pensez qu'on peut vraiment pas décrire une langue sans prendre en compte son histoire... ? » Fin de la réponse : « Et tous les systèmes formels, je pense en particulier à Chomsky, pour qui j'ai une très grande admiration, parce qu'il a montré les limites du formalisme précisément, tout en prétendant envahir tout de formalisme. Et le point de résistance majeure à la formalisation du langage et le point d'existence majeure de la socialisation et de la psychologisation du langage, c'est le lexique ».

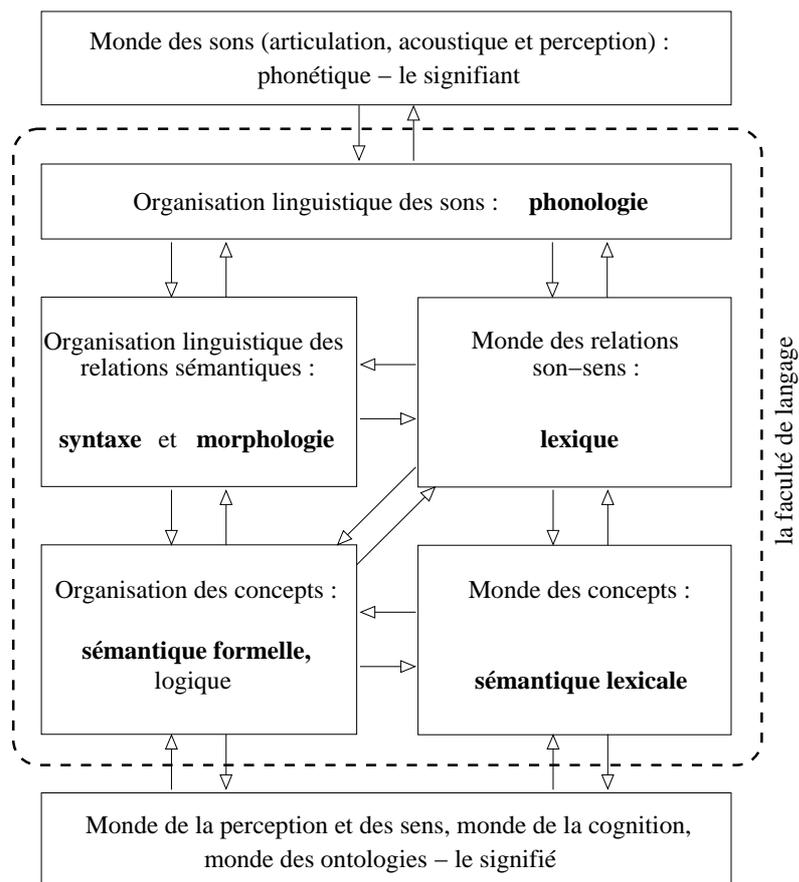


FIGURE 1.6: Un modèle modulaire (restreint) de la faculté de langage

1.2 Le français

1.2.1 Une métaphore redoutable

La métaphore de la forêt, proposée à la section 1.1.1.1, entend s'opposer à une autre métaphore, omniprésente dans les discours sur le français en tant que langue externe, la métaphore de la plante individuelle. Cette dernière sous-tend, par exemple, le langage des historiens de la langue. En effet, ils tiennent fréquemment des propos téléologiques, ramenant les changements linguistiques en France à un processus qui devait finalement aboutir au français classique, celui du XVII^e siècle.

Au lieu d'être de simples étiquettes, les adjectifs qualifiant le français au cours de son histoire ramènent cette histoire à une suite de stades de croissance d'une plante (cf. figure 1.7). La période classique est perçue comme une apogée, comme l'est le stade de la floraison, et l'ancien et le moyen français ne sont que des passages intermédiaires obligés et non-aboutis pour parvenir à cette floraison. On peut ressentir les effets de cette métaphore par exemple en pensant au titre de l'ouvrage d'Allières (1996) : *La formation de la langue française* ; ou aux titres de parties de Chaurand (2006:15) : « *Ébauche* d'une morphosyntaxe », et Chaurand (1999:17) : « Les structures latines *ébranlées* » ; ou encore aux personifications de Chaurand (2006:9) : « définitivement *dégagée* du latin, la nouvelle langue *s'affirme* comme étant de plus en plus différente [...] », et de Picoche et Marchello-Nizia (1996:185) : « Des transformations capitales sont déjà *acquises* ».

On sent bien le problème qu'il y a à poser des états de langue qui ne seraient pas consistants en synchronie, comme si on pouvait croire que les locuteurs de l'ancien et du moyen français ne parlaient comme ils parlaient que parce qu'ils n'étaient pas encore parvenus à faire de

leur langue le français classique, bien qu'ils aient réussi à en faire autre chose que le latin. La métaphore de la forêt évite le problème puisqu'elle permet de décrire des états successifs consistants (chaque état est une forêt à part entière), qui ne dépendent, quant à leur identité de 'forêt', ni de leur passé, ni de leur avenir.

Et la métaphore de la plante ne s'arrête pas là : comment parler du français moderne ? Les paysans le diront : « le vrai stade utile d'une plante, c'est la maturité, pas la floraison ». Alors, on accorde au français classique le droit d'avoir su effectuer quelques derniers ajustements, pour devenir une langue adaptée à la modernité, le français moderne. Une fois encore, la métaphore de la plante se « casse la figure ». Car les paysans poursuivent : « La maturité, c'est l'âge où il faut moissonner, puisqu'après, la plante se dénature définitivement ». Moissonner ? Mais alors il faudrait ensuite repartir de zéro, après tous les efforts consentis ? À moins que la *relève* ne soit déjà là (Raymond Queneau, 1965:67, je souligne) :

« Les partisans du français *correct* et académique ne pourront l'empêcher de se *corrompre* s'ils persistent à le vouloir unique. Au contraire, si on laisse tout le dynamisme, tout le foisonnement de nouveautés au français nouveau, au néo-français – les *impuretés* de l'un devenant la *correction* de l'autre – alors le français proprement dit, indemne des *attaques* du temps, conservera sa *pureté* naturelle. Pour qu'il survive, il faut l'*embaumer*. »

Comme on le voit, on ne peut plus penser le changement sans rupture.

Chapitre 1 *Objet*

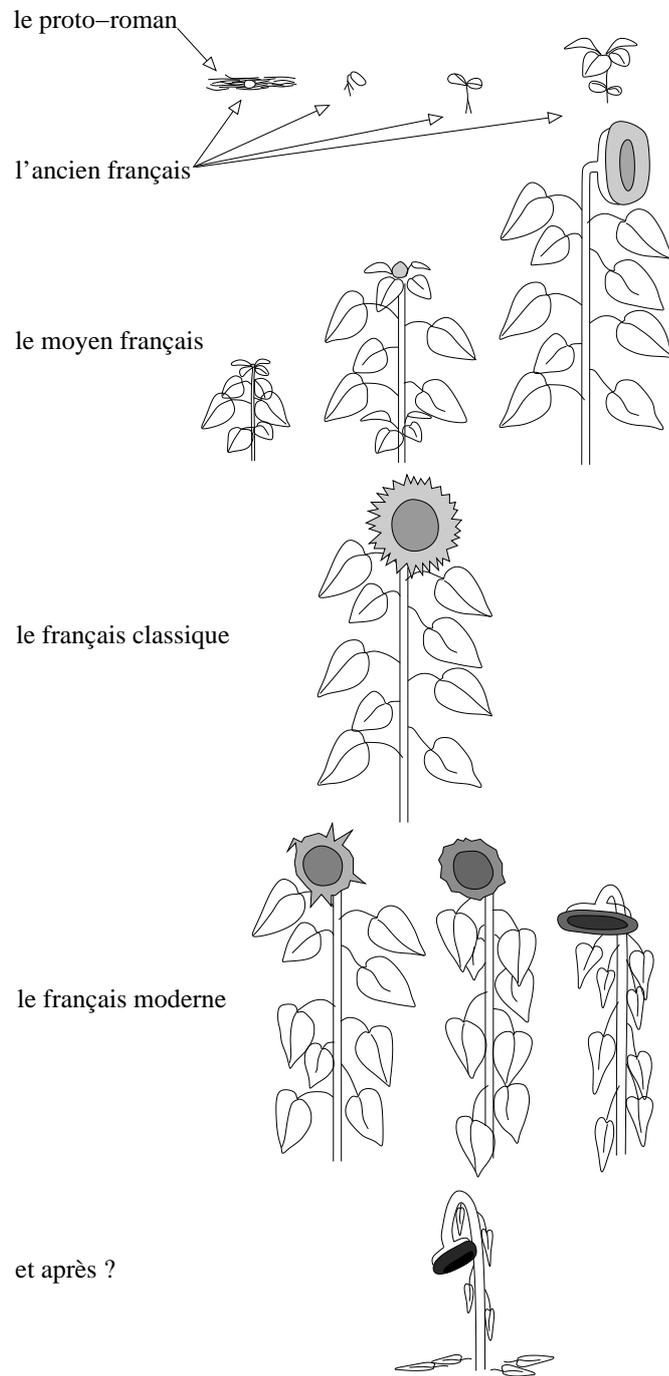


FIGURE 1.7: Le français, une plante qui a crû, et qui donc décroitra !

1.2.2 Langue, progrès, détérioration, évolution, perfection

À cause de la métaphore de la plante (cf. section 1.2.1), penser le français en termes d'évolution est douloureux. On se voit contraint de disserter sur le progrès et la détérioration de la langue et de la grammaire : « Tout changement n'est pas une détérioration, loin de là, sinon il faut mettre la notion de progrès à la poubelle. » (Josette Rey-Debove, citée par Gadet, 2003:106)⁵. Dans ce contexte, les profanes sont contraints de voir les locuteurs comme des usagers de la langue, et la langue comme un livre, tandis que la *main invisible* de Grevisse est garante du respect dû à cet objet fragile, et qu'il faut défendre contre les attaques du temps.

« Il me semble que toute langue est, pour le locuteur qui la parle dans un moment quelconque de son histoire, “parfaite” dans le sens aristotélicien du terme (« Parfait se dit d'une chose en dehors de laquelle il n'est plus possible de rien trouver qui lui appartienne », *Métaphysique*, Livre Δ, XVI). Seule la comparaison technique avec une autre grammaire attestée ou idéalisée produit l'impression d'imperfection. » (Barra Jover, 2008:note 1)⁶. Partant de là, aucune question langagière franco-française ne pourra toucher clairement la faculté universelle de langage

5. Cf. également Farandjis (1983:9-10) : « Compétence linguistique ! Elle est nécessaire. En effet, les noyaux durs de la syntaxe et de la prononciation doivent être constamment enrichis d'enveloppes nouvelles mais ne peuvent être en aucun cas oubliés sous peine de donner comme progression dans la communication et la sociabilité, ce qui ne serait en fait qu'une régression ; “aller chez le coiffeur” ou “aller au coiffeur”, la distinction de sens pour scabreuse qu'elle soit, n'en demeure pas moins réelle ! » « De Rabelais à nos jours, la langue française a connu bien des avatars. Elle peut et doit s'assouplir, s'enrichir mais son progrès, pour le plus grand progrès de tous, ne peut se faire en régressant au niveau du borborigme ou du code “flash et spot”. »

6. Montaigne exprimait également cette idée (*Essais*, III, 9, cité par Huchon, 2002:16) : « [...] Nous disons que [notre langage] est asture parfait. Autant en dit du sien chaque siècle. »

(les propriétés universelles et particulières des boîtes noires). Tout ce qui n'est pas en interaction avec les 'modules non-langagiers' (cf. figure 1.2) est évident, inconscient, et donc n'apparaît que dans des études interlinguistiques. Sans cet apport comparatif, les jugements langagiers que les locuteurs du français émettent se limitent généralement à la perception sociale, esthétique, ou d'intelligence que les énoncés suggèrent de ceux qui les prononcent.

1.3 Conclusion : mon objet d'étude

Mon objet d'étude sera donc l'ensemble des énoncés produits en français (en France) actuellement, en considérant que, comme objet social, le français possède des propriétés suggérées par la métaphore de la forêt, et que, comme objet individuel, il possède les propriétés d'une boîte noire qui sont à explorer dans une perspective interlinguistique. Je me centrerai sur l'étude de cette boîte noire.

Chapitre 2

Tangentes

C E CHAPITRE sera l'occasion de proposer quelques détours pour présenter certains présupposés qui ont une influence et des implications essentielles dans mon travail.

Dans la section 2.1, je pose ce que j'appelle le problème du canal et je propose une réponse. En particulier, j'aboutirai à la constatation banale mais nécessaire que l'écrit n'a pas le statut de l'oral, simplement parce que le graphique n'a pas le statut du vocal.

La section 2.2 me permettra de présenter les systèmes graphiques et leurs propriétés, ainsi que d'insister sur le caractère fallacieux de l'idée d'une *langue écrite* indépendante de la langue orale.

L'orthographe du français sera analysée de façon critique en détails, sans toutefois chercher l'exhaustivité, dans la section 2.3.

La section suivante sera l'occasion d'aborder à travers des exemples la distinction que je propose de suivre entre la compétence linguistique et les connaissances métalinguistiques des locuteurs (section 2.4).

Enfin, j'exposerai à la section 2.5 le système de transcription utilisé ultérieurement dans la thèse, lequel système se veut donner la commune mesure de toutes les productions de français.

2.1 Le problème du canal

Pour cette section, on admettra le modèle de la communication par le langage naturel du tableau 2.1. On rappelle que la linguistique se consacre, selon ce schéma, à l'encodage linguistique, selon un modèle qui porte bien sûr à débat (voir section 1.1).

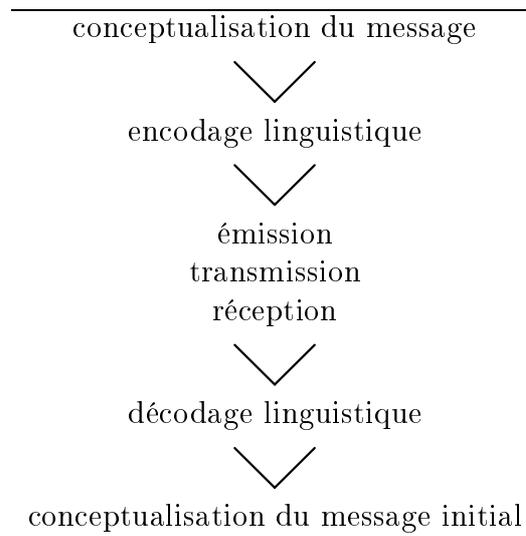


TABLEAU 2.1: La communication par le langage naturel

2.1.1 Canaux de communication

2.1.1.1 Définition

Le *canal primaire de communication* est ce qui sépare l'encodage linguistique de son décodage. Il définit la nature du matériau physique du signifiant de la langue en question. Ainsi, les *langues vocales* (LV, à

signifiant phonétique) utilisent le *canal audio-oral*, et les *langues des signes* (LS, à signifiant gestuel) utilisent le *canal visuo-gestuel*.

2.1.1.2 Canal et encodage linguistique

La nature et les propriétés fondamentales de l'encodage linguistique ne sont pas débattues ici, mais on prendra à titre illustratif la 'liste de Martinet' (tableau 2.2), développée à partir des langues vocales, pour interroger cette problématique.

1.	vocalité
2.	arbitraire du signe
3.	linéarité
4.	double articulation
5.	discrétion (\neq continu)

TABLEAU 2.2: La 'liste de Martinet'

La propriété 1 est clairement liée au canal, puisqu'elle le définit. Est-ce que les propriétés 2 à 5 sont des conséquences de cette vocalité ? Les signes linguistiques des LV sont-ils arbitraires précisément parce que ces langues sont vocales ? Quid également de la linéarité, de la double articulation, de la discrétion ? Pour certains chercheurs sur les LS, c'est toute la faculté de langage qui est implémentée en fonction du canal. C'est ainsi que, concernant le lexique de la langue des signes française (LSF), Cuxac (2000:55) conclut que « l'iconicité s'avère être une donnée structurante ». Et, en effet, on peut s'imaginer que l'arbitraire des LV est imputable à la faible motivation qu'il y a à trouver dans les sons. À part les bruits, quoi imiter facilement et intuitivement par les sons de la voix ? Malgré toutes les trouvailles du phonosymbolisme, il est difficile de concevoir que

l'iconicité sonore de la voix motive et structure les signes linguistiques des LV dans leur ensemble¹.

Il faut peut-être alors repenser les propriétés du langage quand celui-ci est implémenté pour et par le canal visuo-gestuel (tableau 2.3).

1. gestualité
2. motivation du signe ? (grâce à l'iconicité du canal)
3. multilinéarité ?
4. simple articulation ?
5. continuité ?

TABLEAU 2.3: Réviser la 'liste de Martinet' pour les LS ?

Après la possible remise en cause de l'arbitraire du signe dans les LS, sans nécessairement remettre en cause sa conventionalité, on peut se réinterroger sur la linéarité du langage. Le canal visuo-gestuel, quadridimensionnel, laisse ouverte la possibilité d'une multilinéarité, difficile à mettre en place par le canal audio-oral, réputé unidimensionnel. Je dirai cependant qu'on oppose là *multilinéarité* à *unilinéarité*, sans que cela ait de franches conséquences, alors qu'une opposition entre une *non*-linéarité et une linéarité entraînerait, elle, de vraies différences de nature (figure 2.1).

Dans une langue iconique, est-ce que la double articulation est encore nécessaire ? est-ce que la description des unités minimales pourrait aboutir, à chaque fois ou au moins significativement fréquemment, à des éléments gestuels indécomposables encore porteurs de sens ? est-ce que les unités minimales de réalisation sont plus ou moins systématiquement

1. On pourrait voir, dans des parties identifiées du lexique, quelques cas d'exploitation de l'iconicité, comme dans l'exploitation des onomatopées. Par exemple, en anglais, à partir de l'onomatopée [*bu:*] imitant une foule huant, on obtient le verbe (*to*) *boo* ('huer'). En français, on pourrait voir le même phénomène dans 'atchoumer', tiré du vocabulaire enfantin.

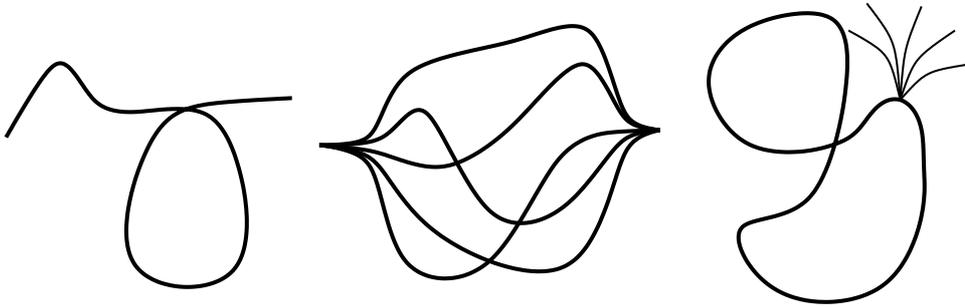


FIGURE 2.1: Unilinéarité, multilinéarité, non-linéarité

porteuses de sens? À nouveau, Cuxac (2000:55, je souligne) est assez convaincu : « Une forte majorité du lexique de la LSF présente des caractéristiques iconiques. On fait l'hypothèse que l'évolution signifiante de ce lexique est limitée par le jeu d'une contrainte de maintien d'iconicité et d'une contrainte de compatibilité iconique. *Ce phénomène ne permet donc pas d'assimiler théoriquement les unités sublexicales de la LSF à des équivalents de phonèmes.* La compositionnalité lexicale relève en fait d'une très dense organisation morphémique qui donne lieu à une *analyse morphophonétique* où l'iconicité s'avère être une donnée structurante. »

Enfin, est-ce que, dans une langue simplement articulée, on pourrait observer des phénomènes de continuité entre le signifiant et le signifié?

Je voudrais simplement conclure sur cette problématique pour dire qu'elle soulève des 'questions existentielles' sur la nature du langage, auxquelles on n'a pas encore trouvé de réponse consensuelle.

2.1.1.3 Encodage secondaire

Il est possible de réencoder le message encodé dans une *langue support* pour le transmettre par un autre canal de communication (un *canal*

secondaire de communication) (voir figure 2.2).

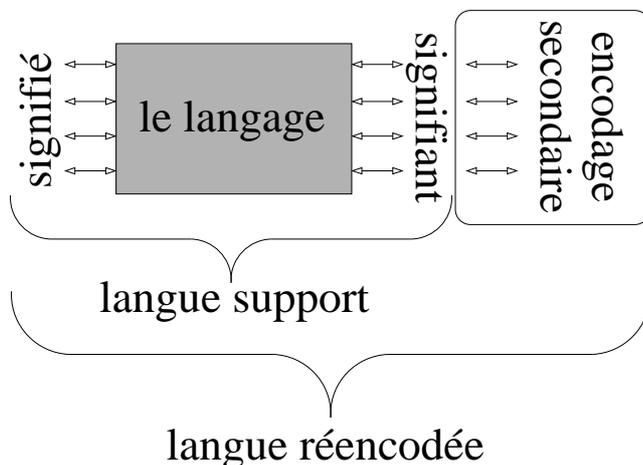


FIGURE 2.2: Langue support et encodage secondaire

Ainsi, les langues sifflées utilisent le sifflement pour encoder la phonologie d’une langue vocale support. Ceci offre un canal, *audio-sifflé*², qui permet d’établir une communication sur de longues distances, par exemple d’une montagne à une autre, grâce à sa puissante intensité acoustique. Sans entrer dans les détails, la technique de réencodage par sifflement consiste à sélectionner quelques propriétés acoustiques saillantes et suffisantes de la langue vocale support³, propres à être sifflées. Par exemple, dans le cas d’une langue support non-tonale, on peut encoder les voyelles en sifflant leur deuxième formant, et les consonnes en imitant la déformation qu’elles provoquent sur les formants des voyelles adjacentes. On trouvera une étude détaillée et nuancée du phénomène chez

2. Il me semble que je propose là un néologisme.

3. cf. Rialland (2005) : “A further respect in which true whistled languages differ from other types of whistled communication is that they encode auditory features of spoken languages by transposing key components of speech sounds.”

Rialland (2005).

Elle conclut ceci concernant les langues sifflées basées sur des langues support non-tonales (je souligne) : “Formant-based whistled languages transpose *the signal envelope of the spoken language, which forms the frame of the whistled message*⁴.” C’est-à-dire que, même si la transposition en question met en jeu des réinvestissements des propriétés acoustiques du message parlé pour les traiter avec une certaine autonomie dans le sifflement, les différentes propriétés de la technique d’encodage sont étudiées *sur la base de l’acoustique du message telle qu’elle aurait été si ce message avait été produit vocalement*.

Ainsi, le Silbo, langue sifflée de l’île de La Gomera aux Canaries et décrite par Rialland (2005), est de l’*espagnol sifflé*. Les propriétés de l’encodage linguistique de cette langue sont celles de sa langue support, l’espagnol. Pour comprendre un message en Silbo, il faut trouver et comprendre le message en espagnol qui en est le support. Ceci semble suffisamment évident pour que l’on n’étudie pas (à ma connaissance) la morphosyntaxe ou le lexique du Silbo (on en connaît suffisamment à travers nos connaissances de l’espagnol), en tout cas pas *en partant de zéro*, mais que l’on s’intéresse au tout premier chef à cette technique d’encodage qu’est le sifflement.

Alors, en tout état de cause, ***les langues sifflées sont des langues vocales***, possédant les propriétés des langues vocales, auxquelles on ajoute un encodage secondaire, pour permettre une transmission du message à travers un canal secondaire. Mais, cette utilisation d’un canal secondaire ne soulève pas les ‘questions existentielles’ que pose l’utilisation d’un autre canal primaire.

4. « Les langues sifflées sur la base des formants transposent l’enveloppe du signal de la langue orale, laquelle forme le cadre du message sifflé. » (traduction : BM)

2.1.2 Et le ‘canal écrit’ ?

À la suite de Saussure, et de sa ‘primauté de l’oral’, on considèrera ici que l’écrit, c’est d’abord l’usage du *canal graphique* pour *réencoder une langue vocale*. En ceci, les langues écrites apparaissent, du point de vue des langues vocales, comme des équivalents des langues sifflées (un simple encodage secondaire), et non comme des équivalents des langues des signes (une implémentation différente de la faculté de langage).

Pour le vérifier, on peut reprendre la ‘liste de Martinet’ (cf. tableau 2.2, p. 37) pour se rendre compte que les langues écrites ne remettent pas en question les propriétés du langage comme le font les langues des signes, mais, au contraire, entérinent en général ces propriétés (voir la section 2.2.1 (p. 43) pour les détails évoqués ci-après sur les fonctionnements et les propriétés des écritures).

En effet, l’iconicité du graphisme (cf. propriété 2.) est marginalement exploitée par les langues écrites. Elles restent des langues à signes linguistiques arbitraires. Il en va de même de la bidimensionalité de la surface graphique (cf. propriété 3.) : elle n’est pas utilisée systématiquement. Et même, la prosodie, qui est pourtant réputée être un phénomène de multilinéarité des langues vocales, est représentée unilinéairement, quand elle l’est, par des signes de ponctuation. On ne parle pas non plus de simple articulation (cf. propriété 4.), puisque les écritures, même celles à base idéographique, donnent des informations sur le contenu phonique du message, c’est-à-dire des informations non-directement porteuses de sens. Enfin, les langues écrites sont bien composées de signes discrets (cf. propriété 5.).

On peut désormais assumer l’idée de l’écrit comme canal secondaire de communication et décrire les écritures comme des techniques d’encodage de messages linguistiques vocaux. On souhaite montrer au passage

2.2 Écritures, et dépendance et autonomie de l'écrit

l'intérêt d'une telle démarche, par opposition à une description de l'écrit comme fait linguistique parfaitement indépendant de toute langue vocale support.

2.2 Écritures, et dépendance et autonomie de l'écrit



2.2.1 Histoire et typologie des écritures

De nombreuses langues vocales ont développé des codes graphiques (écritures) très divers. Boutora (2003) propose une synthèse historique et une typologie des écritures, que l'on va rapidement tirer ici à grands traits.

À l'origine des écritures, les pictogrammes, qui « suggèrent approximativement le contenu d'un énoncé » (Boutora, 2003:22). Mais ceux-ci ne sont pas là pour représenter les éléments linguistiques d'un énoncé, seulement le contenu du message. Ainsi, « la pictographie peut ne pas être considérée comme un système d'écriture, mais comme étant à l'origine de l'écriture » (Boutora, 2003:22).

Chapitre 2 Tangentes

Puis arrivent les *idéogrammes*, premières véritables manifestations d'une écriture, puisqu'ils notent directement les mots d'un énoncé linguistique. Le principe du graphique pour encoder un message linguistique formulé dans une langue support est alors acquis.

Dans le trop grand foisonnement des idéogrammes, ceux-ci finissent par prendre des valeurs phonétiques pour « faciliter la lecture de certains idéogrammes trop ambigus, [avant de devenir] autonomes » (Boutora, 2003:23). Une fois ces signes devenus autonomes, ils sont désormais des *phonogrammes*, pouvant noter « selon les langues, [des] syllabes graphiques, [des] racines consonantiques, ou [...] des] unités sonores minimales de la langue, les consonnes et les voyelles » (Boutora, 2003:23).

Dans leur état synchronique, les écritures peuvent être comprises comme des systèmes de symboles graphiques remplissant essentiellement deux fonctions de représentation d'unités linguistiques. D'une part, un signe graphique peut encoder une unité linguistique signifiante, remplissant alors une *fonction idéographique*. Et d'autre part, il peut encoder une unité linguistique non-signifiante, remplissant une *fonction phonographique*.

Les écritures idéographiques développent de la phonographie, devenant en ça hybrides, idéophonographiques⁵, tandis que les écritures phonographiques peuvent développer des éléments à fonction idéographique⁶. L'étude de l'orthographe du français, à la section 2.3, illustrera ce der-

5. « Chaque système qui fut idéographique à l'origine a dû nécessairement intégrer une certaine dose de phonographie pour continuer à se développer, comme on a pu l'observer dans les écritures cunéiformes, l'écriture hiéroglyphique et l'écriture chinoise. On qualifiera ces systèmes d'idéophonographiques [...]. » (Boutora, 2003:24)

6. Je ne conclurai pas, malgré la suggestion de Boutora (2003:24), à la nécessité de la fonction idéographique dans les écritures phonographiques : « un système purement phonographique n'est pas apte non plus à rendre pleinement compte de la langue (homophonie) et nécessite à son tour un soutien du principe « opposé », autrement dit idéographique. »

2.2 Écritures, et dépendance et autonomie de l'écrit

nier phénomène et permettra de la situer dans la typologie des systèmes d'écriture.

Finalement, la très théorique dichotomie entre écritures idéographiques et phonographiques réside surtout dans l'opposition suivante : il y a les écritures où la fonction idéographique des signes est directe, et la fonction phonographique dérivée, et des écritures où c'est l'inverse qui vaut.

2.2.2 Dépendance et autonomie des langues écrites

La question cruciale qu'on repose ici est la suivante : *est-ce que le canal graphique influence les propriétés linguistiques (en particulier, morpho-syntaxiques) des langues vocales support ? si oui, est-ce par nécessité ou contrainte, ou plutôt par tendance ou facilitation ?*

2.2.2.1 Dépendance

Comme on a présenté les écritures, c'est-à-dire des techniques d'encodage secondaire de langues vocales support, on a supposé qu'elles étaient dépendantes des langues support. Ceci s'est traduit par une analyse des écritures et des signes graphiques en termes de fonction idéographique et fonction phonographique. Si les fonctions graphiques sont bien seulement des techniques d'encodage des unités linguistiques d'un énoncé formulé dans une langue support, alors les écritures ne sont pas censées innover en matière de propriétés de l'encodage linguistique.

L'écriture chinoise recyclée

Un argument montrant que les écritures s'appuient sur des langues vocales, c'est le cas de l'écriture chinoise. En effet, elle est réputée idéographique, et peut donner l'impression de pouvoir encoder directement

Chapitre 2 Tangentes

un contenu conceptuel, presque en dehors de toute formulation linguistique. C'est ce qui la rendrait utilisable pour transmettre un message formulé dans une autre langue. Et l'écriture chinoise a été en effet utilisée par les Japonais, les Coréens et les Vietnamiens. Boutora (2003:65ss.) montre que, dans ces trois cas, l'écriture a subi une adaptation forte à la nouvelle langue support. Ainsi, malgré l'impression d'universalité de son idéographie, l'écriture chinoise était faite *pour écrire le chinois*. Les propriétés du japonais, du coréen et du vietnamien ont dû s'imposer à ce système d'écriture.

Mair (2003), dans sa préface à un ouvrage de Unger, se réjouit de la démonstration proposée par l'auteur :

“In this informative and entertaining book, once and for all, J. Marshall Unger thoroughly demolishes the notion that Chinese characters directly convey meaning without any reference to specific languages and cultural contexts. [...] the author's] conclusion: in reality, there is no such thing as an ideogram.”⁷

La lecture

Un autre argument vient des études du traitement cognitif de la lecture. Elles montrent à quel point elle est intimement liée à la langue vocale support de l'écrit.

D'après Wren, lire, c'est également écouter :

7. « Dans ce livre informatif et distrayant, J. Marshall Unger démolit absolument et une fois pour toutes l'idée que les caractères chinois portent du sens sans référence à des langues particulières ou des contextes culturels. [...] conclusion de l'auteur : en réalité, l'idéogramme n'existe pas. » (traduction : BM)

2.2 Écritures, et dépendance et autonomie de l'écrit

“Surprisingly, your temporal lobe (particularly on the left side of your brain if you’re right handed) is also active right now, processing all of the “sounds” associated with reading — even though you’re reading silently to yourself, the areas of the brain that process speech sounds are active just like they would be if you were listening to somebody speak⁸.”

Wren (2006) présente le modèle de Philip Gough. Ce modèle, élaboré avec l'idée de venir concrètement en aide aux enseignants de la lecture, en particulier dans le cas de l'anglais, aboutit fermement à la conclusion suivante : « *l'aptitude à la compréhension [écrite] [dépend de] la capacité de décodage et [de] celle à comprendre la langue orale.* »

Comment est-ce que l'on peut, dans ce cas, supposer qu'un écrit ne soit pas un encodage secondaire, alors que la lecture est clairement un décodage secondaire, et s'appuie clairement sur la langue vocale support ?

Enfin, je citerai le fait que l'on n'étudie pas une langue, même lorsqu'on a accès seulement à des représentations graphiques de celle-ci, comme dans le cas des langues mortes, sans lui attribuer une représentation sonore⁹. Que cette représentation sonore soit la bonne ou pas, elle est nécessaire. Et c'est bien elle qui permet de maîtriser une langue étrangère (même vivante) seulement à l'écrit, lorsque l'on n'est confronté qu'à celui-ci.

Finalement, pour envisager des cas d'écriture autonome, on ne trouve plus que des cas de messages écrits, sans formulation linguistique sup-

8. « Étonnamment, votre lobe temporal (particulièrement sur le côté gauche si vous êtes droitier) est également actif en ce moment, en train de traiter tous les “sons” associés à la lecture — même si vous êtes en lecture silencieuse, les aires de votre cerveau qui traitent les sons de la parole sont en activité de la même façon qu'ils le seraient si vous étiez en train d'écouter quelqu'un parler. » (traduction : BM)

9. Le seul cas qui m'a été rapporté comme exception est le cas du *linéaire B* (Mario Barra Jover, c.p.).

port, et interprétables comme des *langages* formels, universellement traduisibles dans une langue naturelle. En voici un :

$$(2) \quad 3^2 = 9$$

(Au fait, est-ce que vous êtes parvenu à lire l'exemple précédent sans lui donner une formulation vocale ?...)

2.2.2.2 Autonomie de l'écrit

À travers la transmission d'un énoncé linguistique par un autre canal que le canal primaire, un encodage secondaire a une ou plusieurs finalités. Par exemple, en utilisant le sifflement (cf. section 2.1.1.3), on cherche à obtenir une communication à longue distance, rendue possible par la puissance acoustique de ce sifflement. Ce sont toutes les propriétés induites par le canal secondaire que l'on va regrouper, dans le cas des langues écrites, sous le terme '*autonomie de l'écrit*'.

« Tous les auteurs s'accordent sur le fait que les fonctions principales de l'écriture sont *le stockage de l'information* et *la prise de recul* qu'elle permet d'opérer sur la langue. » (Boutora, 2003:20, je souligne). Ces deux propriétés sont, je pense, les principales motivations de l'utilisation de l'écrit, et les principales sources des faits d'autonomie de l'écrit par rapport aux langues vocales support. Chacune de ces deux fonctions a des effets importants.

L'holistique

En stockant l'information, on la rend accessible globalement à l'œil du lecteur. Pour ce qui est de l'écrit, Françoise Gadet¹⁰ parle de l'association

10. Intervention lors de l'École de Linguistique des Corpus Oraux, Nantes, Juin 2006.

2.2 Écritures, et dépendance et autonomie de l'écrit

de la linéarité et d'une holistique. Une des conséquences de cette propriété qui démarque nettement l'écrit de l'oral est que le locuteur–allocutaire peut traiter des énoncés plus longs et plus complexes et enchevêtrés à l'écrit qu'à l'oral. Les écrits des philosophes allemands ou de Marcel Proust ont cette réputation. Pour obtenir cet effet holistique à l'oral, il faut s'appuyer sur la mémoire immédiate du locuteur–allocutaire. Il y a pour cela plusieurs possibilités. Soit il faut du temps pour connaître par cœur l'énoncé, ce qui peut être réalisé dans les histoires courtes transmises oralement, comme les blagues. Soit il faut tolérer une certaine inexactitude dans la reproduction de la forme, et se contenter de reproduire au mieux le sens. Jean-Louis Aroui (c.p.) rapporte le cas d'un locuteur d'une langue à tradition orale qui venait d'entendre un récit réalisé en 2000 vers, et qui s'est proposé de le reproduire « à l'identique » ; ce qu'il a fait immédiatement, mais en 6000 vers !

Comme on le voit, l'écrit *favorise* une certaine complexité des énoncés. Mais ceci est indirectement obtenu par la facilité qu'il offre de pallier au faible stockage possible dans la mémoire à court terme du locuteur–allocutaire. Un bon entraînement de cette mémoire à court terme, ou une redéfinition de la notion d'*énoncé identique* affaiblit cette particularité. De plus, dès lors que la possibilité de relecture offerte au locuteur comme à l'allocutaire disparaît, les phénomènes holistiques sont très contrariés. Essayons de lire Marcel Proust ou d'écrire à sa façon au moyen d'un de ces panneaux d'affichage qui font défiler des messages sur une seule ligne, en laissant visible une vingtaine de caractères seulement, et qui ne s'arrêtent pas ni ne reviennent en arrière. . . ça paraît compromis, non ?¹¹

11. Exemple proposé par Mario Barra Jover (c.p.).

La prise de recul

Grâce au stockage de l'information, l'écrit permet une prise de recul sur la langue. Il aide à réfléchir sur celle-ci. Cette réflexion peut être inconsciente, comme lorsqu'un scripteur se relit et modifie son énoncé, sans qu'il puisse pour autant expliciter ses motivations. Elle peut devenir consciente et s'apparente alors à un propos métalinguistique. À nouveau, il me semble qu'il ne s'agit pas d'une dichotomie entre l'oral et l'écrit, mais que l'écrit est un simple *facilitateur* de la prise de recul, même s'il est très efficace. On n'est pas surpris que l'écrit, comme canal secondaire, soit archi-privilegié face à l'oral pour mener l'étude d'une langue. Même dans le cas de données orales, on leur préfère une transcription graphique, sans laquelle l'étude de ces données est périlleuse. L'écrit est, *mais n'est que*, beaucoup plus pratique que l'oral pour prendre de la distance par rapport à l'énoncé.

2.2.3 Conclusion

La littérature décrit les écritures comme des encodages graphiques de langues vocales. Les écritures remplissent principalement des fonctions phonographiques, dans lesquelles elles représentent les éléments signifiants de la langue, et des fonctions idéographiques, dans lesquelles elles représentent les éléments signifiés. Toute tentative, telles celles rapportées dans Arrivé (1993), de décrire l'écrit totalement indépendamment du vocal et de l'oral, ne peut qu'échouer devant les contradictions proposées par l'écriture chinoise et l'impossibilité de concevoir la formulation d'un message linguistique sans lui donner une forme sonore.

Comme elle a été décrite, l'autonomie de l'écrit est donc un ensemble de propriétés particulièrement prononcées dans les contextes écrits, sans

2.3 L'orthographe du français : une observation critique

être exclusives de l'écrit, et n'est donc pas de l'ordre de la compétence linguistique.

Pour ce qui sera de la présente étude, je considérerai que le français oral comme le français écrit sont des langues vocales. Je devrai donc leur donner une représentation graphique équivalente (section 2.5).

2.3 L'orthographe du français : une observation critique

2.3.1 L'orthographe : une définition (adaptée de Blanche-Benveniste et Chervel (1978))

L'orthographe, c'est la place que laisse un code graphique (ici celui du français) à des formes non-canoniques d'un mot, *jugées fautives*, sans que le mot soit véritablement perdu. Blanche-Benveniste et Chervel (1978:85–86) donnent l'exemple de MYTHES. Le code graphique du français laisse la possibilité d'écrire ce mot MITHES, MITES, MYTTHE, . . . voire MIT, qui sont des formes fautives — il n'y a toujours qu'une seule forme admise par la norme parmi celles possibles dans le code. Ces formes peuvent être comprises comme étant le mot MYTHES, mais 'sans qu'on n'y ait mis l'orthographe', tandis que les formes MOTHES, RYTHES ou MYDHES ne respectent pas le code graphique et ne peuvent pas être prises pour MYTHES.

A contrario, dans la graphie espagnole, MODAS ('modes') ne peut s'écrire qu'ainsi. Le code graphique de cette langue ne laisse pas la place à d'autres formes graphiques pour ce mot. MODA, MODDAS, MODASSE, MAUDAS, etc., ne sont pas MODAS, et ne peuvent pas être pris pour ce mot. C'est pour cela que l'on dit que l'espagnol n'a pas d'orthographe.

D'ailleurs, en espagnol, le code est régulièrement révisé pour empêcher l'apparition d'orthographe, et ce depuis les débuts des études grammaticales sur cette langue.¹² Antonio de Nebrija, dans sa *Gramática Castellana* de 1492 pose ainsi les principes qui guident encore aujourd'hui les règles de la graphie espagnole : « Il faut écrire comme on prononce et, inversement, prononcer ce que l'on écrit » (cité par Tollis, 1998:78).

2.3.2 Orthographe et langue : la grande illusion

2.3.2.1 « L'orthographe, c'est la langue »

Les clichés sur l'orthographe française vont bon train. Un des plus forts est celui selon lequel, en touchant à l'orthographe, on toucherait directement à l'histoire de France, et il faut défendre notre orthographe comme on défendrait encore aujourd'hui notre territoire et nos frontières. Très rapidement assimilée à la langue, l'orthographe a droit aux mêmes égards que le français classique. Sa complexité est une nécessité de son état de maturité. Et toucher à sa complexité, ce serait là aussi risquer la fénaison.

À ce titre, le dossier consacré à l'orthographe par Le Monde de l'Éducation de mai 2004 est très révélateur. On y trouve quantité de propos qui mélangent l'orthographe à des problèmes étrangers à celle-ci. En particulier, on la confond avec la langue elle-même :

- lorsqu'une correctrice d'édition dit que « les écrivains font beaucoup de fautes [*d'orthographe*] », elle trouve que « c'est normal, notre langue est très difficile » (p. 26) ;

12. À l'heure actuelle se pose par exemple la question de 'régler le compte' de B et V : ils codent le même phonème, /b/, et ne transcrivent même pas chacun un des deux allophones (B ↔ [b] et V ↔ [β]). De même, il reste des cas d'indécidabilité entre G et J, H et Ø, LL et Y.

2.3 *L'orthographe du français : une observation critique*

- lorsque la graphie des SMS « engendre un nouveau type d'orthographe », la problématique devient non pas « une menace pour l'orthographe traditionnelle ? » mais carrément « une menace pour la langue ? » (p. 34)
- on rapporte que la réforme de l'orthographe allemande avait pour but de « rendre plus logique et cohérente la langue » (p. 38) ; et si on se demandait d'abord si une réforme de *l'orthographe* réussissait à rendre *l'orthographe* plus logique et cohérente ?
- Renée Honvault, secrétaire générale de l'Association pour l'information et la recherche sur les orthographe et les systèmes d'écriture, « clame son amour pour l'orthographe, mais un amour objectivé » ; et pour illustrer cette objectivité, elle « recommande la régularisation [. . . d]es pluriels en -aux, par exemple » (p. 39) ; mais est-ce que le fait de dire (ou écrire) LES CHEVALS est vraiment un problème d'orthographe ?

Finalement c'est Pilote le Hot, slameur, qui parvient, dans ces colonnes ma foi fort édifiantes, à distinguer orthographe et langue : selon lui, « il faut décomplexer. On peut commencer par le bas, **en faisant des fautes, ça n'empêche pas d'écrire de belles choses**. L'important, c'est de donner du sens » (p. 40).

2.3.2.2 **Un non-patrimoine national**

Les ouvrages de Blanche-Benveniste et Chervel (1978) et Catach (1998) montrent bien que la graphie du français a une histoire d'abord interne. Ils montrent aussi et surtout que ce n'est que l'histoire récente qui lui a consacré son caractère officiel (légal), lequel donne de cette orthographe une impression d'immuabilité depuis les débuts lointains du français, et par là d'éternité. Au contraire, disent ces auteurs, le système graphique

du français a toujours été varié, attaché à des groupes sociaux, voire à des personnes, et le dictionnaire de l'Académie n'a fait loi qu'à partir du XIX^e siècle, consacrant alors la *faute d'orthographe*, qui n'existaient pas auparavant — ni l'orthographe, ni la faute.

2.3.3 Le système graphique du français : un pot-pourri

Je reprends ici largement les formidables descriptions de Blanche-Benveniste et Chervel (1978) et Catach (1998). Les éléments que je retiens et mets en valeur sont des choix personnels, correspondant aux points que je souhaite ensuite exploiter. Je renvoie le lecteur aux ouvrages cités pour des descriptions exhaustives.

Historiquement, le système graphique du français est alphabétique. C'est-à-dire un système phonographique où les unités graphiques, *les graphèmes*, codent des unités sonores de la taille des phonèmes. Mais l'alphabet latin s'est révélé inadapté à la phonologie du français. En particulier, on est passé d'un système à cinq timbres vocaliques en latin ([i u e o a]) à un système qui a pu avoir seize timbres ([i y u e ø o ε œ ɔ a ɑ ə ẽ œ ã õ]). Et la conservation de cet alphabet, porteur de la langue de la bible donc revêtant un caractère sacré — il a fallu un millénaire pour y ajouter les trois petites lettres J, K, et V! —, a créé une situation conflictuelle. Ainsi, alors que le latin avait un système graphique où les unités avaient des fonctions phonographiques régies par des règles qui allaient au plus simple, le système graphique du français a produit dès le début des 'monstres' phonographiques, qui se sont souvent développés en idéogrammes.

2.3.3.1 La fonction phonographique

Les graphèmes ont une fonction phonographique lorsqu'ils représentent un matériau phonique. On parle alors de *phonogrammes*. Par exemple, le graphème T du mot TALONNEUR code le son [t] (ou le phonème /t/, selon le point de vue), le digraphe EU de PLAQUEEUR code le son [œ] (ou le phonème /œ/), et le graphème X de DÉSAXER code les sons [ks] (ou les phonèmes /ks/). Le cas le plus simple repose sur deux points : (i) les graphèmes sont seulement des monogrames, sans signe diacritique, et un graphème code toujours le même son ou phonème, en même temps qu'un son ou phonème est toujours codé par le même graphème, et (ii) dans un mot, les graphèmes et les sons sont en bijection ordonnée, c'est-à-dire que le premier graphème code exactement le premier son, le deuxième graphème, le deuxième son, et ainsi de suite. Le mot CUIR fonctionne ainsi (voir figure 2.3).

C	U	I	R
[k	ɥ	i	ʁ]

FIGURE 2.3: Correspondances graphèmes-sons dans le mot CUIR

Il faut ici s'arrêter dans la description du système graphique du français pour constater qu'il existe des systèmes graphiques qui se limitent à la fonction phonographique élémentaire ; c'est-à-dire qui ne se donnent aucune autre prérogative que celles décrites au paragraphe précédent (c'est le cas des systèmes graphiques du turc, du roumain, ...).

À partir de ce point, les 'dérives' du système graphique standard du français prennent place. À commencer par les dérives de la fonction phonographique.

Déjà évoqués sont les cas où les graphèmes et les sons ne sont pas en nombre égal. Ensuite viennent les cas où les valeurs phonographiques des graphèmes ne sont pas stables. Ils sont tellement nombreux en français que l'on ne recense qu'un cas parfaitement stable : le son [u] est toujours codé par le graphème U, comme dans APPUI. Tous les autres sons peuvent être codés par des graphèmes différents et tous les graphèmes peuvent coder des sons différents¹³. Voici quelques exemples pour illustrer cela. Le son [ɛ] peut être représenté par E dans INTERUVALLE, par È dans PROTÈUGE-DENTS, par Ê dans FÊUTE, par AI dans VESTIAURES, par EI dans BEUIGNE, par É dans PROTÉUGERA, et même marginalement par A dans CAUKE et EA dans BREUAK. Le graphème E peut représenter [ə] dans DÉBORUDEMENT, [e] dans USSAI, [ɛ] dans PERUCÉE, ∅ dans VASEULINUE, et même [a] dans MOUELLE.

Ces faits sont problématiques, aussi parce que s'y mêlent des questions de valeurs de position, puis des questions de valeurs variables, et de valeurs indécidables :

- le graphème S a une valeur de base [s], dans SAUTÉE et PASTS;
- puis il peut prendre la valeur de position [z] entre deux voyelles (graphiques), dans CROISSÉE;
- ou la valeur ∅ en fin de mot, dans BRASS;
- etc ... — hélas¹⁴.

13. Les consonnes aussi peuvent toutes être codées de plusieurs façons si l'on prend en compte les consonnes doubles

14. Pour illustrer une dernière fois la douce folie de la phonographie du français, on citera deux cas :

- Si le o de MOUELLE \sim /mwul/ code le son /w/, alors il le code aussi dans le trigraphe OIN \sim /wẽ/, et du coup aussi dans le digraphe OI \sim /wa/. Donc dans ce dernier, I code /a/! Pour éviter cette dernière ineptie, on peut prendre le problème à l'envers. OI est un digraphe qui code irréductiblement /wa/, donc, dans OIN, OI code /wẽ/ et N code la nasalisation (ou bien OIN code irréductiblement /wẽ/). Mais alors on perd le rapprochement pourtant bienvenu entre le o de OIN et le o de MOUELLE, et entre le IN de OIN et le IN de PIN.

2.3 L'orthographe du français : une observation critique

PASTIS_̄ et BRAS_̄ montrent que la valeur de S est variable. Ce qui se vérifie complètement dans BUS_̄. Dans ce cas, sa valeur est indécidable si le mot est isolé. Il faut le reconnaître autrement que par sa seule graphie, ce que l'on fait dans les phrases *ils ont voyagé en bus et au cours de la troisième mi-temps, les jus d'orange ont été bus moins vite que les bières.*

Conclusion On peut commencer à décrire comment les graphèmes du français remplissent une fonction phonographique. On doit rapidement accepter de compliquer les règles pour rendre compte de phénomènes très banals (dans le sens où ils sont fréquents). Et finalement, chose qui s'illustrera principalement avec l'étude des graphèmes muets, on doit se résoudre à faire appel au sens à transmettre, et pas seulement aux sons.

2.3.3.2 Les fonctions idéographiques

L'exemple de BUS nous lance dans ces considérations idéographiques, tellement nécessaires à la description et à la compréhension du système graphique du français.

À côté de la fonction phonographique apparaissent donc des graphèmes dont la présence ne s'explique que par une référence à autre chose que le signifiant des mots où ils se trouvent. On distingue en général les références au sens du mot isolé (la fonction logographique), et les réf-

Il n'y a pas moyen de conserver une consistance sur tout le raisonnement. Cette contradiction peut même conduire à cette anecdote qui m'a été rapportée d'élèves de primaire à qui l'on interdisait de citer le mot VOITURE comme exemple de mot contenant le son [a]!

- Le E final de PLATE_̄ ~ /plat/ a une fonction phonographique très particulière : il ne code aucun son, mais il permet l'activation de la fonction phonographique du T, puisque sans E, on prononce PLAT ~ /pla/. Mais que faire alors du E de MATE_̄ ~ /mat/ puisque MAT se prononce /mat/ également ?

rences au mot dans sa structure morphologique et dans la syntaxe de la phrase (la fonction morphographique). Là encore le système graphique du français se complique non seulement parce qu'il laisse une place à l'idéographie, mais encore parce qu'il le fait très vite n'importe comment. En particulier, une grande difficulté vient du fait que les symboles graphiques utilisés pour remplir les fonctions idéographiques sont déjà ceux utilisés pour la phonographie : ce sont les lettres de l'alphabet dans les deux cas. Comme si cela ne suffisait pas, on fait bien souvent porter ces fonctions idéographiques à des graphèmes qui sont déjà des phonogrammes !

La fonction logographique

Les *logogrammes* sont des graphèmes qui ont pour fonction d'identifier une unité lexicale directement. C'est-à-dire en court-circuitant le passage par le signifiant. Le graphème A dans ANGLE a une fonction phonographique : il participe au codage du son [ã] à l'intérieur du digraphe AN. Mais le choix de A par l'orthographe — et non de E : ENGLE serait aussi un codage phonographique possible pour [ãg] — supprime la variabilité¹⁵. Ainsi, ce choix propose l'identification de l'unité lexicale à la lecture, et exige cette identification à l'écriture¹⁶. Ce choix graphique est essentiel puisqu'un lecteur intransigeant lisant ENGLE peut se refuser à identifier le concept d'ANGLE en se disant : « Si celui qui a écrit ENGLE n'a pas fait le bon choix graphique, je le soupçonne de ne pas savoir quel

15. Comme cela a déjà été signalé, c'est cette *exigence* qui constitue le fait *orthographique*. Sans cette exigence, la variabilité graphique offerte par les faiblesses du code phonographique sont de simples *habitudes graphiques*, telles que celles qui ont cohabité jusqu'au XIX^e siècle.

16. On pourrait ainsi paraphraser le lecteur-scripteur exigeant : « Je reconnais ce mot parce que c'est bien celui qui a la prononciation [ãg], et *a fait le choix de la graphie* A. Et je code la prononciation du mot [ãg], et le concept d'ANGLE plus directement par le fait que je choisis la graphie A »

2.3 L'orthographe du français : une observation critique

concept il a exprimé ; s'il avait lui-même bien identifié le concept, il n'aurait pas pu se tromper dans la graphie. Dans le doute, je ne me prononce pas moi-même en tant que lecteur. »

Les logogrammes sont également très fréquemment des graphèmes muets. Ils n'opèrent donc même pas de choix entre deux codages phonographiques possibles, mais sont des signes uniquement et directement associés au concept du mot qui les insère dans sa graphie. Le S de CARQUOIS est ainsi un pur logogramme. Avec ces logogrammes muets, on peut pousser le raisonnement jusqu'à dire que toute absence de logogramme muet est aussi un choix, et une façon d'identifier un concept.

Pour justifier la présence des logogrammes, on évoque fréquemment la 'nécessaire' différenciation graphique d'homonymes lexicaux¹⁷. ANCRE et ENCRE sont distingués par ce moyen, tout comme FOIS, FOIE et FOI. Le choix d'un logogramme peut être justifié de différentes façons. Mais toutes les justifications sont finalement historiques, dans le sens où les choix sont déjà tous faits et ne souffrent plus guère de discussion.

L'étymologie, latine, grecque, ou autre, est invoquée. Par exemple, l'ATILF propose ces étymologies pour ANCRE et ENCRE :

ancre : 'emprunt au latin *ancora*'

encre : 'du bas latin *encau(s)turn*. Le dictionnaire de l'Académie proposait ANCRE et ENCRE en 1694, mais plus à partir de 1718.'

On voit bien que l'étymon en EN de ENCRE a été adapté en AN pour entrer dans le système graphique du français — pour permettre à EN de rester spécialisé dans le codage de [ɛ̃] — avant que la fonction logographique

17. Cette nécessité semble toute relative, vu le nombre non négligeable d'homonymes également homographes, comme SON, MODE, VASE, pour ne donner que des exemples de mots de même catégorie grammaticale . . . D'autre part, le français oral est étranger à cette nécessité, lui qui se passe fort bien de distinguer les homonymes.

ne reprenne définitivement le pas.

Conclusion Les logogrammes ont la fonction d'identifier des unités lexicales de la langue. Ils personnifient graphiquement les mots. Mais, aucune règle, en particulier synchronique, ne permet de les prédire. Ils sont parfaitement arbitraires.

Les fonctions morphographiques

On trouve un bon exemple d'un *morphogramme* simple et robuste (c'est-à-dire pour lequel il n'est pas utile de se soucier d'éventuelles exceptions) dans le système graphique allemand. Il s'agit de la majuscule que l'on trouve au début des noms — propres et communs. Ainsi, on sait que 'enden' (finir) n'est pas un nom, et que 'Enden' (des fins) en est un. Cette information est directe : elle n'est pas fournie par les sons codés, et cette majuscule (ou non) n'influence pas le son codé. Et surtout cette information est grammaticale : il s'agit ou non d'un mot appartenant à la catégorie grammaticale des noms, donc susceptible d'apparaître dans certaines positions syntaxiques, ... On notera bien une qualité de ce morphogramme que ne partagent pas les morphogrammes du français : la fonction phonographique — portée par l'identité du graphème — et la fonction morphographique — portée par la casse du graphème (majuscule ou minuscule) — sont autonomes.

Dans le système graphique du français, on trouve d'une part des morphogrammes indiquant l'appartenance à une famille lexicale de mots — les *morphogrammes dérivationnels*, et d'autre part les morphogrammes indiquant des catégories grammaticales comme le temps, la personne, le genre ... — les *morphogrammes grammaticaux*.

2.3 L'orthographe du français : une observation critique

La fonction morphographique dérivationnelle Lorsqu'on se demande ce qui explique un choix de graphème, si la réponse réside dans l'identification de l'appartenance d'un mot à une famille lexicale de mots, alors ce graphème a une fonction morphographique dérivationnelle.

- (3) a. LENTŒ ~ [lã] car LENTŒEUR, LENTŒEMENT, RALENTŒIR ... ;
b. PAIN ~ [pã] car PANŒIFIER ;
c. NUIT ~ [nuj] car NOCTŒURNE ;
d. DOIGŒT ~ [dwa] car DIGŒTAL.

Certains choix de morphogrammes dérivationnels sont douteux. Le phonogramme associé dans les mots de la même famille est légèrement différent, comme en (4a), ou bien le lien le plus spontané n'est pas celui choisi, comme en (4c), ou bien encore le lien lexical est étymologiquement éronné (4d).

- (4) a. DOUŒ et DOUCŒEUR ;
b. FRANŒC et FRANŒCHISE ;
c. SIROŒP et SIRUŒPEUX, et non SIROŒTER ;
d. POIDŒS et PONDŒÉRER, malgré une origine commune avec PESER.

La fonction morphographique dérivationnelle n'existe que si le choix du graphème se justifie *phonographiquement* dans la famille de mots ; sinon, il reste un logogramme à l'échelle d'une famille de mots. C'est le cas du E de ENCRE : s'il se justifie par le E de ENCRER et de ENCREUR, il n'est alors toujours pas plus justifié, puisque ENCRER et ENCREUR n'ont aucune raison supplémentaire, en particulier phonographique, d'avoir un E, et non par exemple un A.

Conclusion La morphographie dérivationnelle implique des choix graphiques en fonction des familles lexicales. Son principe a l'air assez clair, mais on ne peut pas l'appliquer strictement, sans transiger, et sans glisser vers l'arbitraire : pourquoi ne pas écrire CHOUCHOUT, puisque son féminin est CHOUCHOUTE ?

La fonction morphographique grammaticale Nous y voilà ! Car c'est bien lorsqu'on aborde l'orthographe grammaticale que les esprits se braquent et se refusent le plus à relativiser 'l'évidence et la nécessité' d'une orthographe, et de l'orthographe dans son état actuel. Il faut avouer que la moindre faute dans ce domaine ruine presque automatiquement n'importe quel CV ou lettre de motivation, quelque excellentes que soient leurs qualités par ailleurs¹⁸.

Que sont donc ces morphogrammes grammaticaux si redoutés ? De même que les logogrammes peuvent être définis comme des différenciateurs d'homonymes *lexicaux*, les morphogrammes grammaticaux peuvent être compris comme des différenciateurs d'homonymes *grammaticaux*. Ils répondent au phénomène suivant : il existe des mots qui cachent sous une forme phonologique unique des homonymes dont la seule différence réside dans le fait qu'ils portent des traits grammaticaux différents. Les exemples sont très fréquents dans les formes verbales où le phonème /e/ peut représenter des ensembles de traits de temps, mode et personne très

18. Si on retient essentiellement la morphographie grammaticale, ce n'est pas parce que des fautes de phonographie, de logographie ou de morphographie dérivationnelle soient particulièrement plus pardonnées, mais c'est parce qu'aucun dictionnaire ou correcteur orthographique ne viendra au secours d'une défaillance dans la morphographie grammaticale automatiquement et sans une intervention active et avertie du scripteur.

Ces fautes sont d'ailleurs bien les plus sévèrement sanctionnées dans les dictées des élèves, et les scripteurs qui se font une fierté de bien orthographier ont encore plus honte de ces fautes-là que des autres.

2.3 L'orthographe du français : une observation critique

variés — voir (5abcd), surtout si on ne différencie pas [e] et [ɛ] à la finale de mot — ajouter (5efg).

(5) /fãt-e/

- a. CHANT-ER : infinitif présent
- b. CHANT-É : participe passé
- c. CHANT-AI : 1^{re} personne du singulier de l'indicatif passé simple
- d. CHANT-EZ : 2^e personne du pluriel (indicatif présent)
- e. CHANT-AIS : 1^{re} ou 2^e personne du singulier de l'indicatif imparfait
- f. CHANT-AIT : 3^e personne du singulier de l'indicatif imparfait
- g. CHANT-AIENT : 3^e personne du pluriel de l'indicatif imparfait

Comment ces morphogrammes grammaticaux sont-ils justifiés ?

Il y a deux points à justifier :

1. D'où vient la certitude que les exemples (5) sont bien des homonymes, et non simplement une forme très ambiguë ? À cette première question viennent deux réponses :

- a) La forme phonologique des exemples (5) n'est en fait pas identique : beaucoup de ces formes contiennent une consonne latente, facultativement réalisable en contexte de liaison (/fãte(r)/ pour (5a), /fãte(z)/ pour (5de), /fãte(t)/ pour (5fg), et /fãte/ pour (5bc)). On obtiendrait alors quatre formes non homonymes, que le simple encodage phonographique implique de distinguer.

Mais le doute est permis, puisque les liaisons facultatives invoquées interviennent surtout lors de la lecture à voix haute, ou lorsque le locuteur possède une très bonne compétence écrite, comme si c'étaient les morphogrammes grammaticaux

eux-mêmes qui induisaient ces liaisons. Le serpent se mord la queue ?

- b) Ces formes apparaissent dans des paradigmes distincts. L'opposition entre (5a), (5b), (5c), et (5defg) est révélée par le verbe *vendre*, qui substitue à (5a) /vãdr/, à (5b) /vãdy/, à (5c) /vãdi/, et à (5defg) /vãde/.

Il reste donc à montrer l'homonymie (5defg). L'opposition entre présent et imparfait est établie par rapport au paradigme complet du verbe à ces temps : il existe un /fãte/ qui alterne avec /fãt/ et /fãtõ/, tandis que les autres alternent avec /fãtjõ/ et /fãtje/.

Comment, enfin, justifier l'homonymie et non l'ambiguïté en personne de (5efg) ? Le singulier et le pluriel de la troisième personne s'opposent, puisque (5f) alterne avec /fãtra/ là où (5g) alterne avec /fãtrõ/. En revanche, l'opposition au singulier entre les 1^{re} et 2^e personnes et la 3^e personne est à peu près injustifiable. Tous les paradigmes qui ne confondent pas les trois personnes du singulier distinguent la 1^{re} des 2^e et 3^e personnes et non la 1^{re} et la 2^e de la 3^e : l'indicatif futur fait /fãtre/ /fãtra/ /fãtra/, et à l'indicatif présent, *aller* fait /ve/ /va/ /va/, *avoir* fait /e/ /a/ /a/, et *être* fait /sui/ /e/ /e/.

2. Une fois cette certitude obtenue (ici seulement partiellement), comment se fait le choix des graphèmes pour chaque cas ?
- a) Ici, on ne peut pas se passer d'évoquer une part importante d'arbitraire. En effet, puisqu'on a fortement limité la portée des liaisons facultatives, le choix des consonnes muettes finales, R, Z, S et T, est historique : elles sont ainsi parce qu'elles étaient ainsi dans le passé. Et même si elles ont pu être justifiées

2.3 L'orthographe du français : une observation critique

phonographiquement, ce n'est plus le cas actuellement.

Il en est de même pour le choix de EN(T) pour la marque de la 3^e personne du pluriel. Il n'a aucune motivation synchronique, à moins de le relier à la nasalité de la voyelle de /ɔ̃/, /sɔ̃/, /fɔ̃/ et /vɔ̃/.

- b) Ensuite, il faut trouver une motivation pour le choix du codage phonographique de /e/. On ne trouve là aussi que des raisons de conservation historique.

Au final, les morphogrammes grammaticaux entrent dans la distinction possible de l'ensemble des traits grammaticaux exprimés morphologiquement en français : on trouve des marques orthographiques de temps et mode, de personne et nombre, dans le domaine verbal, et de nombre et de genre dans le domaine nominal. Le cas particulier du nombre dans le domaine nominal sera traité précisément dans la partie II.

Conclusion De même que les logogrammes personnifient des mots lexicaux, de même les morphogrammes grammaticaux personnifient des traits grammaticaux. Lesquels et comment ? On a montré des exemples qui laissent voir que ces graphèmes ne sont pas toujours évidents. D'une part, leur forme n'a pas de fondement synchronique clair, d'autre part, le choix des traits grammaticaux à représenter et distinguer peut être clairement infondé dans la langue.

2.3.4 Conclusion

La réputation de complexité de l'orthographe du français n'est certes pas usurpée. Cependant, on a vu que les études des quarante dernières années en ont permis une description précise et structurée. En s'appuyant

Chapitre 2 Tangentes

sur la synchronie *et* la diachronie du français, on se rend compte que l'orthographe n'invente rien qui ne soit déjà peu ou prou dans la langue.

Je retiendrai quelques propriétés qui ressortent de ces études et qui montrent que l'orthographe est bien un *choix* :

- a posteriori, presque toute l'orthographe du français est descriptible et généralisable ;
- a priori, c'est-à-dire synchroniquement, de très nombreux points restent arbitraires et non-justifiés :
 - la « douce folie » de la fonction phonographique ;
 - le choix des items lexicaux à personnifier par la fonction logographique ;
 - le choix des logogrammes eux-mêmes ;
 - le choix des items lexicaux qui « méritent » un morphogramme dérivationnel ;
 - le choix des morphogrammes dérivationnels eux-mêmes ;
 - le choix des traits grammaticaux à identifier ;
 - le choix des morphogrammes grammaticaux pour les identifier ;
- et enfin, il faut noter la présence d'exceptions !

Une démonstration par l'absurde pourrait ressembler à cela :

CONTER et COMPTER sont des homonymes. Heureusement que l'orthographe les distingue, sinon les enfants apprendraient à « CONTER sur leurs doigts » ! Il est donc tout aussi urgent de distinguer VOLER₁ et VOLER₂, afin de désambiguïser « VOLER un faucon ». Par exemple, « le voleur a VOLÉ le faucon » et « le fauconnier a *VOLLÉ le faucon ».

2.4 Compétence linguistique et connaissances métalinguistiques

« Nous n'avons pas appris, ni comme enfants ni comme adultes, à dire ce que nous faisons lorsque nous parlons : les mouvements que nous faisons pour produire des sons, les structures de sons et de grammaire que nous utilisons. Mais nous avons appris à parler de ce que nous faisons en écrivant ; c'est pourquoi lorsqu'on demande à quelqu'un de parler de sa langue, il parle en fait de l'écriture de sa langue. »

*Leonard Bloomfield, 1927, cité par
Blanche-Benveniste (2002b:8)*

Avant de poursuivre, je souhaite faire un dernier détour autour des notions de *compétence linguistique* et de *connaissances métalinguistiques*. La première et les dernières sous-tendent chez les locuteurs respectivement leur *activité linguistique* (ou performance linguistique) et leurs *activités métalinguistiques*.

Il sera important de distinguer ce qui ressortit à l'une de ce qui ressortit aux autres dans l'ensemble des faits évoqués dans mon travail.

2.4.1 Définitions

Compétence linguistique

La compétence linguistique d'un locuteur est l'ensemble des mécanismes inconscients intériorisés dans sa « boîte noire » (cf. section 1.1).

Chapitre 2 Tangentes

La compétence acquise par les locuteurs d'une langue seconde qui développent une parole plus ou moins spontanée est une forme de compétence linguistique, objet de l'étude de « l'acquisition langue seconde ». Cependant, on s'en tiendra ici à l'étude et à la caractérisation de la compétence linguistique des locuteurs natifs d'une langue.

Connaissances métalinguistiques

Les connaissances métalinguistiques d'un locuteur sont l'ensemble des savoirs explicités, théorisés, conscients qu'il a de sa langue et de sa compétence linguistique. Ces connaissances ne sont plus réservées aux seuls locuteurs natifs. Tenir le propos « Pour moi qui suis français, l'espagnol est plus facile que l'allemand » exprime typiquement une connaissance métalinguistique. Ces connaissances sont évidemment à prendre avec des pincettes.

Activités linguistiques et métalinguistiques

Les activités linguistiques et métalinguistiques sont toutes les activités qui supposent la mise en œuvre concrète respectivement de la compétence linguistique et des connaissances métalinguistiques par un locuteur.

2.4.2 Exemples

2.4.2.1 Cas consensuels

Phonologie

Dans le domaine de la phonologie, la compétence linguistique fait confondre les allophones et distinguer les phonèmes, ou encore adapter les emprunts à la phonotactique de sa langue maternelle.

2.4 Compétence linguistique et connaissances métalinguistiques

- (6) a. français : [ʁ] = [χ] = /r/
b. arabe : /ʁ/ ≠ /χ/ ≠ /r/
- (7) (anglais) ['wi:kɛnd] > (français) [wi'kɛn]

Au contraire, la conscience phonologique (savoir quels sons composent un mot) est un savoir métalinguistique, nécessaire par exemple pour proposer une transcription phonétique ou phonologique, ou encore un encodage phonographique.

Morphosyntaxe

Dans le domaine de la morphologie, la compétence linguistique permet de fabriquer des néologismes et de dériver les mots de la famille d'une nouvelle racine. Elle implique également la flexion, par exemple des verbes. En syntaxe, on trouve les phénomènes d'accord ou les contraintes sur l'ordre des mots.¹⁹

- (8) à partir de /tøf/ (= 'fête' en verlan), on dérive le verbe /tøfe/ (= 'faire la fête') et les noms /tøfør/ et /tøføz/ (≃ 'fêtard').
- (9) a. il-a perdy/*perdr se kle
b. tu-vø vwar/*vwa/*vy l rezylta
- (10) a. lez asiet sõi/*e sal
b. el-e mort/*mor la smen dernier
- (11) a. il-a kase mõi stilo
b. *il-a kase stilo mõi
c. *i kase a mõi stilo
d. *a kase mõi stilo il

19. Le système de transcription sans crochets ni barres obliques utilisé à partir de l'exemple (9) est présenté à la section suivante (2.5).

Chapitre 2 Tangentes

Les connaissances métalinguistiques liées à la morphosyntaxe s'expriment dans plusieurs domaines. On trouve là la classification des mots en parties du discours, où l'identification des relations syntaxiques. Énoncer des règles de grammaire, concevoir un exercice de grammaire avec ses exemples ou fabriquer un néologisme sur commande sont aussi des activités typiquement métalinguistiques.

Lexique

Dans le lexique, la compétence linguistique consiste à connaître les relations sons-sens. Apporter sciemment un couteau et pas une fourchette quand on nous demande d'apporter un *kuto* (et pas une *furfet*) ressortit à cette compétence (associée à une connaissance du monde partagée avec notre interlocuteur). Un cas plus formel est par exemple celui des traits lexicaux arbitrairement distribués dans le lexique. Cela nous fait produire (12).

(12) il-a transforme yn banal anegdot ãn istwar emuvã*(t)

Expliciter le fait que *istwar* est un nom féminin (et même simplement employer le terme « féminin ») est une activité métalinguistique. On peut également penser à l'écriture ou la consultation d'un dictionnaire.

Sociolinguistique

En matière de sociolinguistique, il convient peut-être de parler de compétence sociolinguistique et non simplement linguistique, voire de parler de connaissances métasociolinguistiques !

Une compétence sociolinguistique et de savoir changer de registre de langue selon la situation. Et une connaissance métasociolinguistique est

2.4 Compétence linguistique et connaissances métalinguistiques

de connaître et attribuer les étiquettes comme *familier, courant, soutenu, populaire, argotique, technique, régional*, etc. tant pour le vocabulaire que pour les tournures de phrase.

2.4.2.2 Cas controversés et polémiques

Jugements

La linguistique générative, qui s'intéresse en priorité à la compétence linguistique, s'est proposé un accès tout particulier à cette compétence. Il s'agit de demander à un locuteur natif des jugements de grammaticalité sur des énoncés de sa langue maternelle, c'est-à-dire de dire si des énoncés sont 'bien ou mal formés' dans sa langue.

Tout irait bien si le doute n'existait pas qu'un jugement de grammaticalité n'active des activités métalinguistiques de la part des locuteurs. On peut craindre que ces jugements ne se transforment en jugements langagiers, et que des locuteurs du français rejettent la grammaticalité de (13a), qu'ils produisent pourtant, et accordent à (13b) une grammaticalité bien généreuse puisqu'ils ne la produisent certainement pas spontanément ni de façon productive.

- (13) a. ʒ-ã-kone plẽ de ʒuør d rygbi
b. n-l-rãkõtrãrõ-nu pa ylteriørMã ?

Écrire

Qu'est-ce qui, dans la production d'un énoncé par écrit, ressortit à la compétence linguistique, ou aux connaissances métalinguistiques ?

On a vu précédemment que le simple fait d'explicitier les sons d'un mot doit être appris. Identifier les mots pour les séparer graphiquement relève

également d'une activité métalinguistique. De même que l'orthographiage d'un énoncé.

Quand un locuteur produit par écrit des structures morphosyntaxiques qu'il ne produirait pas spontanément par oral, comment l'interpréter ? Il me semble qu'il y a deux cas à distinguer. Soit il s'agit d'une structure morphosyntaxique que le mode de production écrit facilite (et même encourage, cf. section 2.2.2.2). Soit les deux structures morphosyntaxiques sont strictement équivalentes, et on est dans le cas suivant. Par écrit, et par réflexe conditionné — l'association systématique de l'enseignement de l'écriture et de certaines structures morphosyntaxiques non spontanées —, les locuteurs se lancent dans une activité métalinguistique qui leur fait produire des énoncés d'une façon comparable à la production d'énoncés dans une langue étrangère. Et de même que les bons locuteurs alloglottes peuvent acquérir une production spontanée, de même les meilleurs scripteurs peuvent atteindre une compétence écrite telle qu'elle cache le fait que les structures morphosyntaxiques en question ne sont pas issues de leur compétence linguistique de locuteurs natifs, mais de locuteurs entraînés aux productions formelles, surtout s'ils parviennent à transférer cette compétence à l'oral.

Il me reste à préciser que cette compétence particulière des situations formelles écrites (et orales) n'empêche pas l'émergence d'une intersubjectivité forte, comme dans le cas du français normé entre deux locuteurs performants, ce qui ne serait pas le cas si la situation impliquait l'emploi d'une langue morte comme le latin actuellement.

2.4.2.3 **Le passé simple et le subjonctif imparfait en français et en espagnol**

Mario Barra Jover (c.p.) relève un contraste fort entre les élèves espagnols et les élèves français en ce qui concerne l'apprentissage de connaissances métalinguistiques.

Quand un élève d'espagnol langue maternelle en arrive à la leçon sur le passé simple ou sur l'imparfait du subjonctif, il doit simplement apprendre à repérer et étiqueter correctement ces deux paradigmes verbaux, qu'il produit absolument naturellement et sans effort. Il doit simplement mettre un nom métalinguistique sur une compétence linguistique déjà acquise.

Contrastivement, l'élève de français langue maternelle, pour la leçon sur le passé simple ou l'imparfait du subjonctif, ne peut s'appuyer sur sa compétence linguistique pour apprendre à identifier ces deux paradigmes, puisqu'il ne les maîtrise pas par le truchement de la simple acquisition de sa langue maternelle. Sa démarche devra se baser entièrement sur des activités métalinguistiques qui s'apparentent à l'enseignement d'un paradigme verbal d'une langue étrangère. Il faut décrire les emplois et la morphosyntaxe de ces temps, et parvenir à les mettre en œuvre, ce qu'un petit Espagnol n'a pas à faire. L'élève français doit repérer qu'il faut employer un passé simple, puis réussir à le faire, alors que l'élève espagnol doit simplement repérer qu'il a employé une forme à étiqueter comme un passé simple.

Si certains enfants français acquièrent bien une compétence concernant le passé simple comme temps du récit non daté (sur le modèle des contes de fées, et tout en produisant fréquemment des formes verbales erronées), il n'en est pas de même pour le passé simple comme temps du passé, et encore moins pour l'imparfait du subjonctif, ces deux derniers étant

nettement en dehors de la compétence linguistique des enfants.

2.4.2.4 L'exemple de l'accord du participe passé²⁰

Après le passé simple et l'imparfait du subjonctif, j'ai choisi d'examiner l'exemple de l'accord du participe passé (ci-après PP), car il illustre sous plusieurs angles comment s'y prendre pour démêler ce qui ressortit au linguistique et au métalinguistique. Comme le nombre présente des difficultés d'analyse, que l'on discute dans la partie II, on se contentera de dissenter ici sur l'accord en genre.

Les PP adjectifs

Commençons par le cas des PP adjectifs. L'accord des PP adjectifs présente le même patron que l'accord des adjectifs : les PP adjectifs reçoivent le trait de genre du 'nom auquel ils se rapportent', qu'ils soient attributs ou épithètes. On distingue ensuite entre les PP épïcènes²¹, qui ne reflètent pas morphologiquement leur trait de genre, et les autres, qui le font.

- (14) a. (i) il-a rsy ã mesaz ãvwaje par sa sør
(ii) il-a rsy ã mesaz ekri par sa sør
b. (i) il-a rsy yn letr ãvwaje par sa sør
(ii) il-a rsy yn letr ekri*(t) par sa sør

20. Je suis reconnaissant envers mes étudiants de Paris 8 du 2^e semestre 2006-2007 de « Grammaire française 2 » qui ont participé activement à l'élicitation des idées contenues dans les paragraphes qui suivent, en particulier en exhibant l'exemple de *mor(t)*, seul PP non-épïcène parmi les verbes qui se conjuguent avec *être* (merci Béatrice!).

21. Il semble que le terme 'épïcène' soit parfois considéré comme impropre dans certaines acceptions, dont celle utilisée ici. Je l'emploierai pourtant.

2.4 Compétence linguistique et connaissances métalinguistiques

C'est la compétence linguistique des francophones qui leur fait produire l'accord du PP adjectif. Et c'est en revanche l'activité métalinguistique des grammairiens qui nomme et décrit les données en (14) comme *l'accord du PP adjectif*. Enfin, il est demandé aux personnes qui reçoivent un cours de grammaire française d'acquérir la connaissance métalinguistique que les données en (14) sont à nommer et décrire comme *l'accord du PP adjectif*.

Cette connaissance métalinguistique est ensuite sollicitable. C'est ce que fait la morphographie du genre : elle exige que la graphie des PP adjectifs reflète le trait de genre qu'ils ont reçu du nom dont ils dépendent, comme illustré en (15).

- (15) a. (i) IL A REÇU UN MESSAGE ENVOYÉ_ PAR SA SŒUR
(ii) IL A REÇU UN MESSAGE ÉCRIT_ PAR SA SŒUR
b. (i) IL A REÇU UNE LETTRE ENVOYÉE_ PAR SA SŒUR
(ii) IL A REÇU UNE LETTRE ÉCRITE_ PAR SA SŒUR

Obtenir la graphie ENVOYÉE_ de (15b-i) demande une technique métalinguistique. Il s'agit ici de repérer le caractère épïcène du PP *ãvwaje*²², puis de le remplacer par un PP non-épïcène, comme *ekri(t)*, pour détecter la présence du trait de genre masculin en (15a) et féminin en (15b), et enfin de traduire ce trait de genre morphographiquement (et par convention) par Ø pour le masculin et par E pour le féminin.²³

Ce qu'il est important de constater, c'est que plusieurs choix métalinguistiques, donc discutables, sont effectués avant d'obtenir la graphie

22. Ceci, bien sûr, après avoir repéré la nature de PP de *ãvwaje* parmi tous ses homophones, voir p. 62.

23. Les graphies ÉCRIT_ et ÉCRITE_ peuvent être décrites dans ces mêmes termes, ou bien en cherchant une justification simplement phonographique, même si elle doit alors être légèrement complexifiée.

ENVOYÉE de (15b-i). Par exemple, on peut se demander s'il est justifié de représenter graphiquement le trait de genre d'un PP qui ne le marque pas morphologiquement, ou même s'il est judicieux de décrire les PP épiciènes comme portant un trait de genre invisible.²⁴

Les PP dans les passifs et dans les temps composés des verbes conjugués avec être

Lorsque le PP fait partie d'un passif, il s'accorde en genre avec le sujet du verbe au passif (16a). Lorsque le PP fait partie d'un temps composé d'un verbe qui se conjugue avec être, il s'accorde en genre avec le sujet du verbe (16c). Dans les deux cas, si le PP est épiciène, l'accord est invisible (16bd).

- (16) a. (i) l fwa final sra fe dmě
(ii) la desiziõ final sra pri*(z) dmě
b. (i) sõ kur e fwazi par trè pø d etydiã
(ii) l o gazøz e vãdy ã butej d dø litr
c. (i) l pti fa e mor la smen dernier
(ii) sa grãmer e mor*(t) presipitamã
d. (i) i-sra syrmã parti avã mwa
(ii) el-ete deza vny isi

Comme dans le cas des PP adjectifs, les locuteurs font appel à leur compétence linguistique, à laquelle ils ajoutent une technique métalinguistique, pour orthographier correctement les exemples (16). Et comme dans le cas des PP adjectifs, la règle de grammaire officielle (comme les éventuelles règles officieuses), et l'orthographe, tout en s'appuyant sur

24. Prétendre que ce trait de genre n'est pas invisible puisqu'il se voit graphiquement ne répond évidemment pas à la question...

2.4 Compétence linguistique et connaissances métalinguistiques

une compétence linguistique claire, ne manquent pas d'introduire une part d'arbitraire.

Les autres PP : dans les temps composés des verbes conjugués avec avoir et dans les constructions pronominales

Le problème rencontré dans ce dernier cas est la non-convergence entre les données produites par la compétence linguistique des francophones et l'analyse métalinguistique officielle de ces données. Les données prises en compte par l'analyse officielle sont — partiellement — énumérées en (17). Les données en (18), pourtant attestées, sont simplement ignorées.

- (17) a. il-a kury
b. (i) il-a vy la foto
 (ii) il-l-a vy
c. (i) il-a pri la foto
 (ii) il-l-a priz
d. (i) ʒ-e fe devlope le foto
 (ii) ʒ-lez-e fe devlope
e. (i) el-s-sõ ekri de kart postal
 (ii) el-s-le-sõ ekrit
f. (i) el-s-sõ permi d vnir
 (ii) el-s-sõ kōpriz
g. (i) le foto s-sõ priz fasilmã
 (ii) el-s-sõ epriz dy nuvo
- (18) a. il-a kury
b. (i) il-a vy la foto
 (ii) il-l-a vy
c. (i) il-a priz la foto

Chapitre 2 Tangentes

- (ii) il-l-a pri
- d. (i) ʒ-e fe devlope le foto
- (ii) ʒ-lez-e fet devlope
- e. (i) el-s-sõ ekri de kart postal
- (ii) el-s-le-sõ ekri
- f. (i) el-s-sõ permiz d vnir
- (ii) el-s-sõ kōpri
- g. (i) le foto s-sõ pri fasilmã
- (ii) el-s-sõ epri dy nuvo

Comme l'analyse officielle 'précède' les données, il s'agit en fait d'une prescription. Cette prescription croit si fort en la 'validité' scientifique, historique, logique, esthétique. . . de sa règle que les données qui y contreviennent font du locuteur qui les produit un quasi-paria, lui qui spolie toute cette belle horlogerie, celle des « COD qui précèdent (ou pas) », des « vrais-faux COD des verbes de mesure », des « verbes à *construction* pronominale » et des « verbes *intrinsèquement* pronominaux ou pronominaux à *valeur passive* »...

Au final, pour produire les données en (17), les locuteurs du français peuvent, il me semble, former trois groupes :

- soit leur compétence linguistique leur fait produire les données en (17) en excluant les données en (18) (ils sont pour le moins rares), et toutes leurs activités métalinguistiques autour de cet accord sont facilitées ;
- soit leur compétence linguistique leur fait manquer des accords de la norme et produire des accords hors la norme, en pouvant varier d'une fois sur l'autre ; alors produire les données de la norme, et les orthographier selon cette norme, exige pour eux de contredire leur compétence linguistique ; on est alors plus proche d'une production

2.4 Compétence linguistique et connaissances métalinguistiques

- alloglottale, naturellement facilitée dans une production différée (typiquement à l'écrit, cf. section 2.2.2.2) ;
- soit leur compétence ne leur fait produire aucun des accords en (17) et (18), ce cas se rapprochant alors du précédent.

2.4.2.5 « OMAR M'A TUER » et le laxisme orthographique

En recevant un courriel disant « CONTENT QUE TU SOIS BIEN RENTRER », on peut adopter plusieurs attitudes devant la faute d'orthographe grammaticale rencontrée. La plus radicale est de n'accorder aucun sens à la phrase. Et donc de ne pas entendre la satisfaction exprimée par notre interlocuteur. Une attitude intermédiaire est de tolérer le message (et donc d'en accepter le contenu) malgré la faute. On pense alors que notre interlocuteur se réjouit effectivement qu'on soit bien rentré, même s'il a mis un infinitif au lieu d'un participe passé.

Mon attitude personnelle a été bien plus généreuse envers mon ami. Je me suis dit qu'il avait pêché seulement au moment d'orthographier la forme verbale à partir de l'explicitation de la catégorie grammaticale du mot en question. Il peut y avoir plusieurs origines à cette faute d'orthographe. Soit mon ami a bien identifié explicitement le participe passé, mais il s'est trompé sur l'orthographe à adopter pour cette forme. Soit il a oublié d'identifier la forme en question et n'importe quel codage phonographique de /e/ a fait l'affaire. Soit enfin il a simplement identifié à tort la forme comme un infinitif.

Quelle que soit la source de la faute, j'affirme pourtant qu'il a bien employé un participe passé ! Et ce pour la raison suivante : j'accorde peut-être beaucoup de crédit à mon ami — quoique je trouve que je lui accorde là une compétence linguistique qui semble bien le minimum lorsque l'on a affaire à un locuteur natif du français — mais je suis persuadé que,

Chapitre 2 Tangentes

s'il s'était réjoui de la bonne tenue de mon *départ*, il n'aurait pas écrit « CONTENT QUE TU SOIS BIEN PARTIR ».

Il serait donc ridicule de conclure à partir de la faute d'orthographe observée que la compétence linguistique du locuteur qui l'a produite est défaillante ! C'est bien seulement son savoir et son activité métalinguistiques qui font un léger acte de dissidence. Comme le montrent très clairement Blanche-Benveniste et Chervel (1978), à la fin de leur ouvrage, à partir de l'étude de productions d'élèves ayant des difficultés en orthographe, le laxisme orthographique, puisqu'il concerne une activité métalinguistique, n'a aucune corrélation nécessaire avec un quelconque laxisme linguistique.²⁵

Le droit a également donné ces dernières années en France un exemple a contrario de la nécessité de ce laxisme orthographique. Comme toute discipline axiomatique et déductive, l'activité qui consiste à dire le droit ne tolère aucune faute de langage. On ne s'attendrait alors pas au fait suivant. Dans un procès qui a rendu célèbre la faute d'orthographe présentée, on s'est demandé si l'auteur du fameux « OMAR M'A TUER » était bien la victime du crime, et s'il avait menti, etc. Mais, évidemment, aucun protagoniste du procès, et pas même la défense, n'a émis le moindre doute sur le fait que le message véhiculé par l'inscription était bien « OMAR M'A TUÉE ». On pourrait insister, en supposant que, s'il avait été écrit « OMAR M'A TUÉ », avec le PP orthographié au masculin, personne n'aurait suggéré que l'auteur de cet écrit ne pouvait pas être la victime, « puisque celle-ci était une femme » . . . Même le droit ne confond donc pas les fautes de *langage* avec les fautes de *métalangage*.

25. Et d'ajouter que le laxisme métalinguistique — donc également orthographique — est bien le moins que l'on puisse attendre des linguistes, eux qui se font une activité professionnelle que de relativiser la qualité de toute connaissance métalinguistique, qu'elle soit leur ou d'autres, passée ou présente, officielle ou dissidente.

2.5 La transcription adoptée

2.5.1 **Rendre justice à la morphosyntaxe**

Ce chapitre jusqu'ici a proposé de se méfier de l'idée d'une langue écrite opposable à une langue orale à travers la contestation du statut parfois accordé à cette différence de canal. Puis, on a montré que l'orthographe repose sur des choix métalinguistiques, donc qu'elle n'est pas recevable comme représentation écrite donnant accès à la compétence linguistique cachée derrière les productions de français. Elle ne peut pas être considérée comme une commune mesure entre toutes ces productions.

J'adopterai donc un autre système de transcription, dont je suggère qu'il peut donner cette commune mesure. Il ne sera pour autant pas dénué de caractéristiques théoriques, puisqu'il fera des choix que l'on peut remettre en cause. Par exemple, la façon de découper les énoncés en mots est bien un choix de représentation qui n'est pas neutre. Ou simplement le fait de représenter des informations phonologiques plutôt que phonétiques.

Cependant, je ne pousserai pas plus loin cette question, et j'espère seulement que la transcription que j'adopte permet une étude sereine des faits de morphosyntaxe. En tout cas, je suis persuadé que cette transcription rend mieux justice à la morphosyntaxe des productions linguistiques de français que l'orthographe standard ou une forme dérivée de cette orthographe, puisqu'il ne paraît pas envisageable dans ces cas de neutraliser la controversée orthographe grammaticale.

En espérant que le lecteur pardonnera de sacrifier à la « lisibilité » perdue dans cette « autre représentation [qui lui demandera] un grand effort » (Blanche-Benveniste, 2004:132-133), et appréciera au contraire l'observation d'énoncés bien plus 'déshabillés'.

2.5.2 Analyse en phonèmes du français

La première étape est le choix de transcrire principalement l'information phonologique des énoncés, à partir d'une analyse en phonèmes que je pense consensuelle. Le résultat sera non pas une transcription exactement phonologique, mais une transcription intermédiaire. Ceci sera rendu visible par une notation sans les crochets de la phonétique ni les barres obliques de la phonologie. Ces crochets et ces barres obliques seront réintroduits pour signaler explicitement du matériau phonétique ou phonologique. Par ailleurs, les crochets seront également utilisés pour marquer des constituants syntaxiques. On ne les confondra pas avec les crochets de la phonétique puisqu'ils n'empêcheront pas de transcrire des mots séparés.

Les phonèmes consonantiques du français font l'objet d'un consensus assez large quant à leur dénombrement. On retiendra directement les symboles de l'API. On obtient le tableau 2.4.

			labiales	coronales	palatales/ vélares
occlusives	orales	sourdes	p	t	k
		voisées	b	d	g
	nasales		m	n	ɲ
fricatives	sourdes		f	s	ʃ
	voisées		v	z	ʒ
liquides	latérale			l	
	vibrante			r	
semi-consonne					j

TABLEAU 2.4: Les phonèmes consonantiques du français

2.5 La transcription adoptée

Les voyelles sont plus problématiques. En effet, on peut encore souhaiter opposer les voyelles mi-hautes et les voyelles mi-basses. Ou bien opposer deux voyelles basses. Ou encore transcrire quatre voyelles nasalisées plutôt que trois. Pour garder une transcription la plus simple possible, j'ai choisi d'utiliser deux diacritiques, à employer à la discrétion du transcriteur. Ainsi, on notera e dans tous les contextes si l'on considère qu'il n'existe plus d'opposition $/e \sim \varepsilon/$, ou bien, on notera e dans les contextes neutralisants (ce qui me semble plus simple qu'un archiphonème $/E/$), et \acute{e} ou \grave{e} (pour $[e]$ et $[\varepsilon]$) dans les contextes considérés comme oppositifs. Les voyelles nasales sont réputées absentes de la structure phonologique abstraite, mais noter les nasalisations est bien pratique. On pourra considérer que le $\tilde{}$ n'est qu'une abbréviation de la notation de la présence d'une consonne nasale dans les contextes de nasalisation ($\tilde{e} \leftrightarrow eN$). ə est conservé non pas pour sa valeur phonologique dans les alternances e/\emptyset (on notera la valeur neutralisée dans ces cas : afet/afte) mais pour quelques cas marginaux où son apparition n'est pas prévisible ($\text{dəor}/\text{pələz} \dots$). Enfin, on emploiera le w dans le but de représenter deux diphtongues, wa et par similitude $w\tilde{e}$. Ceci est justifié puisqu'il faut par exemple distinguer $[\text{t}\chi\text{wa}]$ (*trois*) et $[\text{t}\chi\text{u.a}]$ (*troua*), qui seront donc transcrits respectivement trwa et trua . En revanche, ɥi face à yi est superflu puisqu'aucun $[\text{ɥi}]$ ne s'oppose à un $[\text{y.i}]$ bisyllabique.

À partir du tableau 2.5, on obtient les symboles du tableau 2.6.

2.5.3 Découper en mots

C'est un enjeu théorique et descriptif important que de faire un choix de découpage en mots des énoncés. J'ai finalement choisi, faute d'une réflexion approfondie sur le sujet, de suivre d'assez près le découpage traditionnel. Ce qui s'en écarte légèrement, c'est d'abord le choix de noter

Chapitre 2 Tangentes

		[-ARR]				[+ARR]	
		[-ROND]	[+ROND]			[-ROND]	[+ROND]
[+HAUT]	[-BAS]	[+ATR]	i	y	ə		ɥ
[-HAUT]		[-ATR]	e	ø			o
		[+BAS]	(ɛ)	(œ)		(ɔ)	
diphthongues		wa / wɛ̃					

TABLEAU 2.5: Les phonèmes vocaliques du français

		[-ARR]				[+ARR]	
		[-ROND]	[+ROND]			[-ROND]	[+ROND]
[+HAUT]	[-BAS]	i	y	ə			ɥ
[-HAUT]			e (é è) ě		ø (ø̇ ø̈) (ø̃)		o (ó ò) õ
		[+BAS]	a (á à) ã				
diphthongues		wa / wɛ̃					

TABLEAU 2.6: Les symboles pour transcrire les voyelles

2.5 La transcription adoptée

tous les clitiques du verbe avec des tirets les reliant entre eux et au verbe : *tu ne m'appelles pas* est transcrit par *ty-n-m-apel pa*. Ensuite, j'ai parfois fait le choix de noter des locutions avec des tirets absents dans l'orthographe, comme pour *pø-a-pø*. Enfin, quelques groupes de mots ont été notés par un seul mot : *est-ce que/est-ce qui* par *esk/eski*, *peut-être* par *pøtetr*, et *parce que* par *parsk*. Il ne s'agit là que de suggestions intuitives.

2.5.4 Quelles informations suprasegmentales ?

Pour transcrire des informations suprasegmentales, il faudrait une théorie claire de la prosodie du français, que je ne possède pas. C'est pourquoi je ne propose presque aucun signe suprasegmental. On se contentera d'un point d'interrogation signalant une prosodie interrogative, d'un point d'exclamation pour une prosodie exclamative, et d'un soulignement pour un mot à accent contrastif. Pour le reste, et en particulier pour repérer les diverses propositions d'un énoncé, on supposera que l'intuition du lecteur, s'il est lui-même francophone, suffit. Enfin, dans le cas d'énoncés spontanés, comme ceux tirés du corpus HH (cf. section 10.1), le symbole + signalera une rupture syntaxique.

2.5.5 La liaison

- A: sa-fãzra pøz-a-pø
B: pøt-a-pø!
C: õ di pa pøz-a-pø ni pøt-a-pø
õ di pø-a-pø
lə aʃ e aspire!

J'utilise une formulation précise du phénomène de *liaison* :

Chapitre 2 Tangentes

1. Une **consonne de liaison** est une consonne située à la fin de la forme phonologique d'un mot et qui possède la propriété d'être réalisée phonétiquement ou non en fonction du contexte (cf. (19)).
 2. Un **contexte de liaison** est un lien syntaxique particulier existant entre des mots adjacents. Je ne me prononcerai pas sur les propriétés de ce lien syntaxique. On dira alors que le (z) de *viej(z)* est dans un contexte de liaison en (20a) et hors contexte de liaison en (20b).
 3. Une consonne de liaison hors contexte de liaison n'est pas réalisée phonétiquement.
 4. Une consonne de liaison dans un contexte de liaison est obligatoirement ou facultativement réalisée phonétiquement devant les mots à initiale vocalique, dit jonctifs, et est tronquée devant les mots à initiale consonantique, dit disjonctifs (et dont font partie certains mots à initiale vocalique).
 5. Le caractère obligatoire ou facultatif de la réalisation de la consonne de liaison peut être attribué à la consonne elle-même : certains mots possèdent une consonne de liaison obligatoire, et d'autres une consonne de liaison facultative, comme /trè(z)/ et /for{t}/ en (21).
 6. Ce caractère obligatoire ou facultatif peut également être attribué à la 'qualité' du lien syntaxique du contexte de liaison : le contexte de liaison *pronom sujet + verbe* rend la liaison obligatoire, comme en (22a), et le contexte de liaison *verbe + attribut du sujet* rend la liaison facultative, comme en (22b).
- (19) a. trè fer
b. trèz onerø
c. /trè(z)/
- (20) a. (i) le viej dam

2.5 La transcription adoptée

- (ii) le viejz ami
- b. (i) le viej dormè
(ii) le viej arivè
- (21) a. (i) trè rapid
(ii) trèz ěteresã
(iii) /trè(z)/
- b. (i) for rapid
(ii) for(t) ěteresã
(iii) /for{t}/
- (22) a. (i) il-vien
(ii) ilz-ariv
(iii) /il(z)/
- b. (i) il-e prezã
(ii) il-e(t) apsã
(iii) /e{t}/

Puisque les transcriptions ne sont pas phonologiques, mais intermédiaires, le système note seulement la réalisation effective d'une consonne de liaison. Si l'on souhaite transcrire les deux possibilités (réalisation ou non), on le signale en plaçant la consonne entre parenthèses, comme en (22b-ii).

Chapitre 3

Sujet

APRÈS avoir présenté un cadre théorique et terminologique permettant d'envisager l'objet langue française et les moyens d'y accéder, on peut se tourner vers le sujet de cette thèse. Il s'agit donc dans ce chapitre de présenter la problématique de la variation morphosyntaxique en général et en français en particulier (sections 3.1 et 3.2), puis de donner un bref aperçu de trois attitudes fréquentes à l'égard de cette variation (section 3.3), et enfin d'exposer l'hypothèse de la diglossie qui est celle retenue dans ce travail pour répondre à la problématique posée (section 3.4).

3.1 La variation morphosyntaxique en français

3.1.1 Définitions

Définition 1 (la variation morphosyntaxique) *La variation morphosyntaxique est le fait qu'il existe des structures morphosyntaxiques*

Chapitre 3 Sujet

*synonymes au sein d'une même langue externe.*¹

Le concept de synonymie doit être pris dans un sens relativement souple. C'est-à-dire que l'on n'attend pas de deux constructions grammaticales qu'elles expriment le même sens en toutes circonstances, mais seulement au moins dans certains contextes. En (23), les deux variantes sont synonymes, en (24), elles ne le sont pas.

(23) en constatant qu'il fait très froid :

- a. sa-ʒel
- b. i-ʒel

(24) pour décrire des gouttelettes d'eau qui se transforment en givre :

- a. sa-komãs a ʒle
- b. %i-komãs a ʒle

Par ailleurs, le concept de synonymie est étendu au domaine de la pragmatique : deux constructions sont synonymes seulement si elles sont également pragmatiquement interchangeable. Par exemple, (25a) et (25b), dans certains contextes, n'accomplissent pas le même acte de langage (Coveney, 1996). (25a) est une vraie requête d'information (le locuteur veut savoir si son interlocuteur a compris, et attend qu'il le lui dise), alors que (25b) est une question rhétorique : le locuteur ne cherche pas une réponse à sa question, mais transmet à l'interlocuteur son intention de poursuivre son discours, en supposant que jusqu'ici son interlocuteur comprend. (25b) introduit un biais vers une réponse positive que (25a) n'introduit pas.

(25) a. esk vuz-ave kōpri ?

1. Voir la section 1.1 pour une définition de la langue externe.

3.1 La variation morphosyntaxique en français

- b. vuz-ave kōpri ?

La question qui se pose lorsqu'on observe une variation morphosyntaxique est de trouver un cadre conceptuel pour la décrire et la modéliser. On utilisera le mécanisme suivant pour identifier et classer les différents types de variation. Tout d'abord, rappelons que les variations doivent être localement synonymes sémantiquement et pragmatiquement. C'est-à-dire que la stricte observation linguistique des données ne permet pas de les interpréter. Ensuite, on cherchera la motivation d'un locuteur qui souhaite corriger un énoncé en proposant de le remplacer par une variante synonyme de cet énoncé. En somme, on se demandera comment une variation est corrélée extralinguistiquement.

Les sous-sections suivantes proposent d'identifier quelques grands types de variation morphosyntaxique, avant de préciser lesquels seront abordés dans mon travail.

3.1.2 La variation diatopique

- (26) a. ʒ-la-kone sofi
b. ʒ-la-kone a sofi

Lorsqu'on reproche à un locuteur de produire (26b) et non (26a), on lui reproche de colorer son langage régionalement. Ici, il s'agit d'identifier une variante régionale repérable comme provenant du sud de la France (par exemple de Toulouse). On pourrait bien sûr au contraire suggérer à un locuteur de produire (26b) plutôt que (26a) pour qu'il s'identifie comme toulousain et non bourguignon, par exemple. Dans cet exemple, on notera que l'une des deux variantes seulement est clairement marquée diatopiquement : (26a) est possible en Bourgogne et à Toulouse, et probablement dans tout l'espace francophone, tandis que (26b) n'est pas

possible en Bourgogne.

L'impression d'étrangeté devant une variante diatopique peut être forte, même pour une variante géographiquement proche. La variante (27b), produite par exemple dans le Morvan, en Bourgogne, n'est simplement pas comprise par les habitants de la Puisaye, une autre région de Bourgogne proche du Morvan, jusqu'à ce qu'on leur explique que (27b) présente simplement une forme particulière de clitique objet direct dont le référent est non-animé (quel que soit son genre et son nombre).²

- (27) a. (la pòm) ʒ-l-e mǎʒe ijer
b. (la pòm) ʒ-i-e mǎʒe ijer

Une variante diatopique est donc une variante identifiable géographiquement. Une variation qui met en jeu une variante diatopique est une *variation diatopique*.

3.1.3 La variation diachronique

- (28) a. ʒ-la-vø kōprãdr
b. ʒ-vø la-kōprãdr
- (29) a. la sien ide
b. sōn ide

Une variante identifiable temporellement est une variante diachronique et entre en jeu dans la *variation diachronique*. (28a), par exemple, est marquée comme une variante du français classique, perçue aujourd'hui comme vieillie, voire agrammaticale. De même pour (29a). Ces variantes illustrent pour certains locuteurs le fait qu'une variante diachronique

2. Les Poyaudins s'accoutument volontiers du sens, mais tout en constatant naïvement que les Morvandiaux « mettent des [i] partout ! »

3.1 La variation morphosyntaxique en français

peut rester compréhensible sans être productible (sauf à le faire exprès), et pour d'autres le fait qu'une variante obsolète peut être parfaitement en dehors du système synchronique.

On peut aussi constater qu'une variante est perçue comme innovante, et peut être préférée (ou pas) pour cette raison. (30b) est par exemple une construction innovante par rapport à (30a), comme (31b) et (32b) par rapport à (31a) et (32a) respectivement.

- (30) a. il-a l atityd dy sørfør
b. il-a la sørfør-atityd
- (31) a. la tuf fisiakyzien (la touche fusciasusienne)
b. la fwasi-tøtj (la 'Foissy-touch')
- (32) a. i-m-ěsylv
b. i-m-tret

3.1.4 La variation diastratique

- (33) a. si el-ave py è-sre vny
b. si el-ore py è-sre vny
- (34) a. il-a y pør biě k i-swa abitye
b. il-a y pør malgre k i-swa abitye

Les variantes (b) en (33) et (34) sont produites dans des groupes sociologiques particuliers, que l'on pourrait intuitivement étiqueter comme populaires. Certains groupes sociologiques au contraire s'identifient parce qu'ils ne produisent que les variantes (a). En (35) et (36), c'est au contraire la variante (a) qui est marquée sociologiquement, comme non-produite par les groupes populaires.

- (35) a. si el-y py el-fy vny
b. si el-ave py è-sre vny
- (36) a. k set mezõ e bel
b. k esk set mezõ e bel

Les variantes qui marquent l'appartenance du locuteur à un groupe sociologique mettent en œuvre le phénomène de *variation diastratique*. Ici, les variantes sont assez clairement hiérarchisées par les locuteurs, qui émettent rapidement des jugements de valeur comparatifs.

3.1.5 La variation diaphasique

La *variation diaphasique* est le phénomène dans lequel un même locuteur peut produire des variantes marquées contradictoirement. Par exemple, il peut produire à la fois des variantes toulousaines et des variantes bourguignonnes, des variantes contemporaines et des variantes obsolètes, des variantes populaires et des variantes bourgeoises. Ou encore, il peut s'exprimer selon différents genres textuels.

C'est cette variation individuelle qui sera au cœur des préoccupations de mon travail.

3.1.6 La variation diamésique

On a affaire à une *variation diamésique* lorsqu'une variante est marquée comme étant produite lorsqu'un canal de communication (un média) particulier est utilisé. Je pense ici à seulement deux types d'exemples. Le premier est la syntaxe particulière des 'petits mots laissés sur la table en partant', qui produisent la variante (37b).

- (37) a. 3-syi parti fer le/de kurs

3.1 La variation morphosyntaxique en français

b. SUIS PARTI FAIRE COURSES

Le second type d'exemple concerne les affiches qui désignent ce sur quoi elles sont affichées, comme en (38b).

(38) a. lez orer d uvertyr sō le syivã

b. HORAIRES D'OUVERTURE

Ce qui me semble remarquable, c'est que dans les deux cas la variante (b) n'est pas susceptible d'être produite oralement (sinon à lire le texte, tout en le laissant dans son contexte de 'petit mot laissé sur la table en partant' ou d'affiche). Produire (39a) ou (39b) en dehors de ces contextes paraît exclu.

(39) a. *syi parti fer kurs

b. ??orer d uvertyr

On pourrait être tenté de modéliser de nombreux cas de variation dans le cadre de la variation diamésique. Mais, comme on l'a décrite au chapitre 2, l'opposition entre l'oral et l'écrit est non-exclusive : une structure morphosyntaxique peut être typique de l'un, elle n'en est pas pour autant exclue de l'autre. De ce point de vue, les cas décrits en (37) à (39) sont exceptionnels, et mériteraient un traitement à part, en dehors de ce qui est souvent traité comme variation écrit-parlé.

On reconnaîtra alors une variante diamésique au fait qu'un autre média ne rende pas justice à cette production langagière. Une transcription phonologique ne convient pas, et même une reproduction exacte de la graphie du message ne peut pas se passer d'une explicitation du contexte. Seul le support original rend parfaitement justice directement à cette variante.

3.1.7 La variation libre

Je finirai en évoquant le cas des variations qui ne trouvent pas de corrélation. On appellera cela la *variation libre*. Aucune des variantes à disposition ne peut être corrélée d'aucune des façons évoquées ci-dessus. À proprement parler, on pourrait trouver que cette variation ne sert à rien. C'est cette idée qui appelle le principe énoncé par Raimo Anttila :

(40) « The mind shuns purposeless variety. »³

Pourtant, je n'opposerai ici aucune résistance à l'idée que la variation peut parfaitement être gratuite.

Pourquoi est-ce que les locuteurs comme les linguistes résistent souvent à cette idée ? Je suggère l'interprétation suivante. La variation nous semble toujours corrélable, mais c'est parce que nous parlons de celle dont nous avons conscience. Et nous en avons conscience justement parce qu'elle est corrélable. La vraie, la variation libre, nous est inconsciente. Dès lors, comment la repérer (puisqu'elle n'est pas corrélable) ? C'est notre esprit conscient, notre activité métalinguistique, qui ne supporte pas l'idée de la variation libre. L'inconscient n'y voit peut-être rien à esquiver. Et dans les communautés linguistiques où s'exerce une pression normative comme le monde francophone, notre prise de conscience métalinguistique se dépêche de corréler (parfois arbitrairement) les variantes nouvellement repérées. Comme le dit Gadet (1997a:10) : « on connaît [...] la manie grammairienne d'attribuer un sens à toute variation de forme ».

Lorsque deux locuteurs se disputent pour savoir quelle variante employer parmi plusieurs possibles, si aucun ne peut justifier sa préférence,

3. « L'esprit évite la variation inutile » (traduction : BM), principe rapporté par exemple dans Lambrecht (1994:257).

3.1 La variation morphosyntaxique en français

alors la variation sera dite libre.

- (41) a. kõtiny a ferfe
b. kõtiny d ferfe

- (42) a. õ-s-aswa u ?
b. õ-s-asie u ?

À ce titre, l'exemple (23), répété en (43), est révélateur.

- (43) en constatant qu'il fait très froid :
a. sa-ʒel
b. i-ʒel

J'ai d'abord été convaincu que les deux variantes avaient une nuance de sens. L'une impliquerait que la température soit inférieure à zéro, l'autre non. En reprenant les exemples pour les présenter à des étudiants, je me suis rappelé la nuance de sens, mais, avant vérification, j'ai spontanément inversé les deux variantes ... comme si la nuance supposée avait été seulement provoquée par la volonté de la voir.

3.1.8 Variation entre idiolectes et accomodation

Lorsque deux locuteurs présentent une variation grammaticale de type interlocuteur (diatopique, diachronique, ou diastratique, et même une variation libre), la problématique est de décrire leurs grammaires respectives *sachant qu'ils se comprennent*. Le cadre conceptuel auquel on pense ici est celui de l'accomodation. C'est-à-dire que l'on décrit chaque idiolecte séparément, sans référence particulière à l'autre, et en se concentrant sur la compétence de production de chacun.

Est-ce que le fait que chacun puisse comprendre les variantes produites

Chapitre 3 Sujet

par l'autre implique qu'il faille les prendre en compte dans la description de leur grammaire ? On rendra compte plutôt de cette intercompréhension en proposant que chacun peut *accomoder* le sens des variantes de l'autre qu'il ne peut pas produire lui-même. L'*accomodation* comme elle est comprise ici peut être définie comme le fait qu'un locuteur traduise en un sens une variante qu'il ne produit pas dans son propre idiolecte et accepte ainsi d'en comprendre le sens.

Ce qui, éventuellement, est modélisable et demande de prendre en compte les compétences de compréhension, c'est l'envie d'interpréter, et de décrire et modéliser les contrastes de difficultés d'apprentissage d'une variante à l'autre. S'il est plus facile pour un Bourguignon d'apprendre à accomoder la variante toulousaine (44b), que la variante morvandelle (45b) pour un Poyaudin⁴, est-ce qu'on va décrire ce contraste en cherchant à proposer des descriptions grammaticales de chaque idiolecte plus proches pour (44a) et (44b) que pour (45a) et (45b) ?

- (44) a. ʒ-la-kone sofi
b. ʒ-la-kone a sofi
- (45) a. (la pòm) ʒ-l-e mǎʒe ijer
b. (la pòm) ʒ-i-e mǎʒe ijer

Ou bien, est-ce que l'on peut se contenter de suggérer que l'exposition à chacune des deux variantes (44b) et (45b) contraste quantitativement suffisamment pour que l'apprentissage de l'accomodation soit biaisé ?⁵

4. C'est là mon intuition de Bourguignon poyaudin.

5. Là-aussi, c'est mon intuition, malgré la distance plus grande entre la Bourgogne et Toulouse qu'entre la Puisaye et le Morvan. On pourrait simplement imaginer que le nombre de locuteurs produisant (44b) (possiblement l'ensemble des locuteurs du sud de la France) est tellement plus grand que celui des locuteurs produisant (45b), qu'il assure à la première variante une diffusion bien supérieure à celle de la deuxième.

3.2 Constat initial et problématique retenue

Malgré l'intérêt de discuter cette problématique, on s'arrêtera bien à l'idée que la grammaire interne d'un locuteur sera à modéliser sur la seule base des énoncés qu'il peut produire.

3.2 Constat initial et problématique retenue

Mes investigations se concentreront sur un aspect particulier de la variation : le fait qu'il y ait de la variation au sein même des productions d'un locuteur, et les façons d'envisager cette variation dans le cas du français.

L'ensemble des développements de cette thèse se fonde sur le constat suivant (Blanche-Benveniste *et al.*, 1990:207) : « Il nous paraît nécessaire de distinguer au moins deux sortes de savoir [grammatical] ; l'un qui est indiscutable, et semble acquis par les francophones avant leur entrée à l'école primaire ; un autre, aléatoire, qu'ils acquièrent plus ou moins bien, tout au long de leur vie, et parfois jamais. L'un correspond à ce que nous appelons la "grammaire première" et l'autre à la "grammaire seconde". »

On propose alors ici d'identifier deux ensembles parmi les énoncés produits en français. Pour l'instant, leur existence est justifiée seulement par l'intuition de ces grammaires première et seconde. Les exemples en (46) seront considérés comme définitoires. Soit les énoncés sont à rapprocher de (46a), et appartiennent exclusivement à un ensemble étiqueté *français classique tardif* (ci-après FCT), correspondant à la grammaire seconde évoquée ci-dessus. Soit ils sont à rapprocher de (46c), et appartiennent exclusivement à un ensemble étiqueté *français démotique* (ci-après FD), correspondant alors à la grammaire première. Soit enfin ils sont à rapprocher de (46b), et appartiennent à ces deux ensembles à la fois.

(46) a. a kel ør arivt-il ? ∈ (FCT \ FD)

- b. a kel \emptyset r esk il-ariv ? \in (FCT \cap FD)
- c. il-ariv a kel \emptyset r ? \in (FD \setminus FCT)

Pour l'instant, FCT et FD sont de simples étiquettes, auxquelles on accorde une valeur heuristique, celle des deux types de savoirs et de productions linguistiques en français. On reprendra ces deux termes en détail à la section 3.4.3, où on les justifiera.

La problématique retenue à partir de ce constat est alors la suivante :

Problématique 1 *Décrire, interpréter, représenter la variation morphosyntaxique individuelle des locuteurs du français, ou : pourquoi est-ce que l'espace de variation interne en français est ce qu'il est ?*

3.3 Hypothèses écartées

Les hypothèses formulées pour répondre à la problématique que je vais écarter dans mon travail ont un point commun. Elles supposent toutes que les locuteurs ont une seule grammaire. Blanche-Benveniste *et al.* (1990:12) se positionnent clairement : « après avoir étudié le sujet pendant des années, nous ne pensons pas qu'il faut poser une grammaire spéciale pour le français parlé ».

3.3.1 Les bons et les méchants

Dans la communauté linguistique française (et peut-être francophone en général), la croyance est forte en l'existence d'un « français correct » et un « français incorrect » (un bon usage et des mauvais usages, ...).

Cette façon de considérer la variation est purement axiomatique. On postule des données correctes, et on justifie a posteriori de l'existence des autres données. Les instances qui ont autorité sur les constructions

3.3 Hypothèses écartées

morphosyntaxiques se prétendent motivées par la logique, l'esthétique, l'économie, l'utilité, la francophonie, mais l'arbitraire règne en maître.

- (47) a. (i) vu-n-ave pa kōpri
(ii) vuz-ave pa kōpri
b. (i) a kel ør arivt-il
(ii) il-ariv a kel ør
c. (ii) est un ratage de (i)
(ii) est une simplification de (i)
(i) est une amélioration de (ii)
(i) est un embellissement de (ii)
(i) est une complexification de (ii)

Les deux exemples canoniques de la négation et de l'interrogation directe (ici très partiellement évoqués) nous donnent l'illustration de la faiblesse scientifique de l'hypothèse des bons et des méchants. En effet, bien des interprétations listées sous (47c) seraient aussi valables si on inversait les variantes (i) et les variantes (ii). D'autre part, ces interprétations manquent cruellement de contenu conceptuel : qu'est-ce qu'une amélioration ? une simplification ? ... et comment repérer objectivement un embellissement ? Il n'y a donc que peu de contenu informatif dans ces simples stigmatisations arbitraires.

Mario Barra Jover (c.p.) fait également remarquer que cet arbitraire est flagrant dans le fait que certains auteurs puissent être cités comme exemples de bon usage, et quelques pages plus tôt ou plus loin comme contre-exemples. Quelle peut être la cohérence grammaticale d'un tel corpus ? Pendant combien de siècles est-ce qu'on peut être cité comme bon usage contemporain ?

L'attitude présentée dans cette section n'est depuis longtemps plus

portée par la majorité des linguistes. C'est pourtant hélas celle qui aujourd'hui encore pénètre le plus le grand public.

3.3.2 La variation en question est à corrélérer à l'opposition parlé – écrit

Dans cette section, on souhaite discuter et critiquer l'interprétation de la variation étudiée dans les termes parlé-écrit de l'école du *français parlé* (Blanche-Benveniste (1983), Blanche-Benveniste et Jeanjean (1987), Blanche-Benveniste *et al.* (1990), Blanche-Benveniste (1997, 2003, 2004) , et l'ensemble des travaux du Groupe Aixois de Recherche en Syntaxe). En particulier, on donnera des indices de la limitation de son caractère explicatif à certains cas bien cernés de variation.

3.3.2.1 Reformuler parlé-écrit en spontané-planifié

Comme on a présenté le français écrit au chapitre 2, l'opposition entre un énoncé parlé et un énoncé écrit, qui pourrait être comprise ainsi :

- (48) a. [yn'v^fafɛb^fwawa'p^fad^{wj}y'le]
b. UNE VACHE NE BOIT PAS DE LAIT.

est reformulée comme suit, c'est-à-dire en fait par une transcription qui donne la commune mesure entre les deux énoncés :

- (49) a. yn vaf è-bwa pa dy le
b. yn vaf n-bwa pa d le

Les moyens que l'on s'est donnés pour parler de la variation de (49) se formulent en termes d'*énoncés spontanés* et d'*énoncés planifiés*. Chaque variation que l'on a envie de lire par ce prisme devra être compa-

tible avec l'idée que la planification d'un énoncé porte la responsabilité de cette variation.

Dans ce qui suit, on gardera parfois les termes 'écrit' et 'parlé' originaux des auteurs évoqués, tout en gardant à l'esprit qu'on les interprète dans les termes 'planifié' et 'spontané'.

3.3.2.2 Ce qui s'explique

La planification d'un énoncé a des conséquences possibles sur cet énoncé. Puisqu'on peut envisager une planification dans n'importe quelle langue, on observera ici des tendances possiblement universelles des énoncés planifiés.

Söll (1985:63-66) rapporte des études qui montrent la plus haute fréquence des nominalisations à l'écrit qu'à l'oral, et ce dans plusieurs langues.

- (50) a. i-sō parti avā k ma sør swa arive
b. i-sō parti avā l arive d ma sør

La description habituelle des nominalisations dérive ces noms des verbes qu'ils 'déverbalisent'. Donc, produire une nominalisation, c'est produire un verbe *puis* le nominaliser. Cette modélisation s'accorde bien avec les observations sur la fréquence des nominalisations supérieure dans les énoncés planifiés, justement là où l'on peut plus facilement se permettre des productions demandant des calculs (computations) plus compliqués et plus longs.

Mais, la variation en français connaît bien plus que ces quelques cas, et notamment de nombreux cas qui semblent ne pas relever d'une planification (à moins de le décider arbitrairement, voir la section 3.3.2.4) ni d'une simple tendance encourageant certaines variantes sur d'autres,

mais qui relèvent d'une fracture idiosyncratique.

3.3.2.3 Ce qu'on ne cherche pas à expliquer

Les pauses, hésitations, reprises, corrections, ruptures et autres abandons sont des avatars connus des énoncés spontanés. Cependant, comme argumenté par exemple dans Blanche-Benveniste et Jeanjean (1987), le but de la description grammaticale n'est pas de modéliser une grammaire pouvant générer (51a). Générer (51b) suffit.

- (51) a. yn y- yn vaf s-e pa dy s-e pa è-bwa pa dy le
b. yn vaf è-bwa pa dy le

On pourrait pourtant vouloir établir une typologie de ces faits, voire essayer d'en tirer des conséquences pour la description grammaticale, mais je m'arrêterai ici à l'idée que la grammaticalité de (51a) est exactement la grammaticalité de (51b). Et qu'en aucun cas, (51a) n'a la grammaticalité de (52).

- (52) yn vaf n-bwa pa d le

Blanche-Benveniste et Jeanjean (1987) diraient que (51a) est « l'avant-texte » de (51b), et pas de (52).

3.3.2.4 Les limites de l'opposition spontané–planifié

Les tendances naturelles des énoncés parlés, comme leur nom l'indique, sont universelles. On s'attend à trouver 'naturellement' des nominalisations moins fréquentes à l'oral qu'à l'écrit, ou bien des énoncés moins longs et complexes et enchassés.

Et puisqu'il existe des variantes parlées et des variantes écrites du

français, alors il faut bien que cette variation se reproduise à travers les langues. Sans quoi, la motivation diamésique de cette variation perd sa crédibilité, et les étiquettes ‘parlé’ et ‘écrit’ perdent leur force explicative.

Pour montrer que l’opposition diamésique n’a pas de force explicative (suffisante) dans la variation en français, on s’essaiera à appliquer naïvement et froidement deux cas de variations du français à une autre langue, ici l’allemand⁶. On verra alors que cette application souffre de contradictions.

Les interrogatives

Il est connu, en termes folkloriques, que les interrogatives avec mouvement Qu- et inversion verbe-sujet du français écrit (53a) sont simplifiables, en deux étapes ordonnées, en supprimant l’inversion (53b), puis en supprimant le mouvement Qu- (53c).

- (53) a. kel ør et-il ?
b. kel ør il-e ?
c. il-e kel ør ?

En comparaison, il est frappant de voir que ces ‘simplifications’ n’ont pas lieu en allemand :

- (54) a. *Wieviel Uhr ist es ?*
 combien heure est ce
 ‘Quelle heure est-il ?’

6. Mon choix de l’allemand n’est pas anodin. Il est lié à une expérience personnelle forte : l’impression (naïve et révisable), quand j’étais lycéen en séjour en Allemagne, que « les Allemands sont fous ! Ils parlent comme dans les livres de grammaire ! » alors que les Français, c’est bien connu, « se contentent à l’oral d’une grammaire plus simple qu’à l’écrit ».

Chapitre 3 Sujet

- b. **Wieviel Uhr es ist ?*
combien heure ce est
- c. **Es ist wieviel Uhr ?*
ce est combien heure

Bien sûr, on doit préciser que l'inversion apparente en (54a) n'est que le résultat de surface de la contrainte V2 de l'allemand⁷, et donc que l'agrammaticalité de (54b) ne doit rien à la non-simplifiabilité de l'inversion, mais à la stricte contrainte V2. Cependant, il reste extrêmement peu justifiable dans les termes parlé-écrit comme spontané-planifié que le mouvement Qu- soit inévitable, comme montré en (54c)⁸.

La parataxe

À l'oral, c'est bien connu, la syntaxe peut presque disparaître, en particulier la syntaxe qui relie les propositions. On trouve facilement des cas de parataxe, c'est-à-dire des propositions reliées sémantiquement, mais pas formellement (ni coordination, ni subordination).

- (55) a. s il-ave verifie l ěfo i-nuz-ore eparŋe yn emosiõ
b. il-ore verifie l ěfo i-nuz-ore eparŋe yn emosiõ
- (56) a. kã il-ariv i-kri deza
b. il-ariv e i-kri deza
c. il-ariv i-kri

En (55b), on peut se passer de la subordination explicite, même si on doit remarquer que dans ce cas la forme verbale au conditionnel passé

7. En allemand, l'élément conjugué du verbe se trouve toujours dans la deuxième position d'une proposition indépendante ou principale.

8. Il n'est pas fait mention ici des questions-échos, qui en allemand comme en français n'exigent pas le mouvement Qu-.

3.3 Hypothèses écartées

(*ore verifie*) plutôt qu'au plus-que-parfait (*ave verifie*) pourrait être l'indicateur formel du lien sémantique entre les deux propositions. Et en (56), la subordination de (56a) alterne avec la coordination de (56b), et la parataxe de (56c), qui peut même se passer de l'indice lexical *deza*.

(55b) et (56c) seraient alors moins élaborés (si la parataxe est bien un défaut d'élaboration) parce qu'ils seraient produits spontanément, sans pouvoir être planifiés.

Là aussi le contraste avec l'allemand est assez net, puisqu'il n'est pas question, dans cette langue, de lier sémantiquement deux propositions sans expression formelle, de subordination ou de coordination. (57c) ne peut être compris que comme une suite de faits, c'est-à-dire qu'il n'exprime pas la contradiction entre les deux propositions que l'on trouve dans (57a) et (57b), et dans l'exemple parataxique (56c) du français.⁹

- (57) a. *Als er herkommt, schreit er schon.*
comme il arrive crie il déjà
- b. *Er kommt her, und schreit schon.*
il arrive et crie déjà
- c. *%Er kommt her, er schreit.*
il arrive il crie

L'idée que « la langue parlée [...] aurait pour particularité de ne pas indiquer formellement les relations sinon par des marqueurs prosodiques » (Blanche-Benveniste, 2002b:95) doit être reformulée par « la prosodie *du français* donne les moyens de signaler des relations sémantiques sans indice formel ». Et on doit se demander non pas pourquoi l'oral permet la parataxe, mais pourquoi le français et sa prosodie la permettent très

9. On doit nuancer ce propos, puisqu'une informatrice (parmi plusieurs informateurs) n'a pas exclu qu'une prosodie très marquée puisse exprimer ce lien en (57c), en particulier en présence de *schon*.

facilement, et non l'allemand.

3.3.2.5 Synthèse

Selon l'école du français parlé, la variation qui est à corrélérer par le concept de « langue du dimanche » et « langue de tous les jours » (Blanche-Benveniste *et al.*, 1990:211) est expliquée dans le constat que la langue du dimanche est principalement écrite (ou planifiée) et la langue de tous les jours principalement parlée (ou spontanée). Les tendances qui se dégageraient de cette opposition seraient dues précisément à cette différence, et les caractéristiques de chaque variante seraient en un sens prévisibles par cette différence de mode de production.

Malgré le pouvoir parfois explicatif de ce point de vue, on a donné quelques indices du caractère pourtant généralement idiosyncratique de la répartition des variantes entre le français écrit et le français parlé. En particulier, ce qui est typique de l'écrit ou du parlé en français ne l'est souvent pas dans d'autres langues. Tout ne s'explique pas comme s'expliquent les nominalisations. Et ce sont ces cas que l'on mettra en avant.

3.3.3 La variation est sociolinguistique

Le projet de l'hypothèse des variantes marquées sociolinguistiquement, qu'on appelle l'école *sociolinguistique*, est double. D'une part, il souhaite proposer une description fine des contraintes sociolinguistiques qui pèsent sur les choix des variantes. On trouve là des propositions multipolaires, qui multiplient les registres de langue à travers une batterie d'étiquettes, comme *soutenu, courant, familier, ordinaire, populaire, ...*, que l'on retrouve par exemple dans l'œuvre de Gadet (1992, 1997b, 2003).

D'autres propositions se concentrent sur une bipolarisation des contraintes sociolinguistiques, comme dans la tradition allemande autour

3.3 Hypothèses écartées

de Söll (1985) puis Koch et Oesterreicher (1990), Koch (1997), Koch et Oesterreicher (2001) qui, à partir des étiquettes initiales *langue parlée–langue écrite* développent les étiquettes *oralité–scripturalité* puis *langue de la proximité–langue de la distance*¹⁰. La bipolarisation est présente également dans l’interprétation sociolinguistique à la française des étiquettes *parlé–écrit* (cf. Blanche-Benveniste *et al.*, 1990).

L’autre aspect du projet sociolinguistique concerne la description grammaticale. On trouve là-aussi deux tendances. La première souhaite donner à ces étiquettes un pouvoir explicatif à l’intérieur de la grammaire. On a vu à la section précédente (3.3.2) pourquoi je souhaite écarter le pouvoir explicatif des étiquettes *parlé–écrit*. La seconde tendance prend son contrepied, en assumant au contraire l’arbitraire de la répartition des variantes dans les registres de langue.

Dans ce cas, soit on décrit chaque variante individuellement, sans la corrélérer aux autres variantes, en suivant par exemple la démarche générale de Gadet (1992, 1997b), soit on souhaite proposer un modèle générant toutes les variantes en mettant en avant leur structure sous-jacente commune. C’est là par exemple le projet de Berrendonner (1983a).

3.3.4 Vers une troisième voie

Pour dépasser les réticences à ces approches que l’on a présentées, il faut une hypothèse qui combine différemment le constat de la variation morphosyntaxique et du patron sociolinguistique de cette variation. D’une part, on cherche à respecter l’intuition d’une variation bipolaire. D’autre part, on veut croire que cette bipolarisation n’explique pas la répartition de la variation, sous quelque formulation de cette bipolarité

10. Les termes allemand originaux sont : *gesprochene Sprache–geschriebene Sprache*, *Mündlichkeit–Schriftlichkeit*, et *Sprache der Nähe–Sprache der Distanz*.

que ce soit, c'est-à-dire que la variation doit garder un caractère idiosyncratique. Enfin, c'est bien la structure de la langue interne des locuteurs qui sera à modéliser, c'est-à-dire la structure des idiolectes, plutôt qu'une grammaire de la langue externe, coiffant tous les idiolectes possibles.

On propose alors de prendre au mot l'idée de la « grammaire première » et de la « grammaire seconde » : « un Français lettré vit [...] une double vie grammaticale » (Blanche-Benveniste, 2004:132). Cette approche s'exprime pleinement dans l'hypothèse que je me propose de soutenir : l'hypothèse de la diglossie, présentée ci-après en détail.

3.4 Hypothèse adoptée : la diglossie

3.4.1 Définition (stricte)

Ferguson (1959) popularise le concept de *diglossie* pour caractériser un type particulier de situation linguistique :

“DIGLOSSIA is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.”(Ferguson, 1959:22)¹¹

11. “La diglossie est une situation linguistique relativement stable dans laquelle, en plus des dialectes primaires de la langue (qui peuvent comprendre un standard ou des standards régionaux), il y a une variété superposée très divergente, hautement codifiée

3.4 Hypothèse adoptée : la diglossie

Depuis, on n'a eu de cesse de dissenter sur la pertinence de ce concept (Fernández (1993) consacre 472 pages à une bibliographie sur la diglossie!). D'une part, on s'est demandé si les quatre situations données en exemple par Ferguson entrent bien dans les critères que lui-même a listés (Dejean (1983) conteste la validité de l'exemple du français et du créole en Haïti). D'autre part, on a proposé d'élargir la notion de diglossie et donc de qualifier des situations diverses de diglossiques, principalement par rapprochement avec le fait qu'elles mettent en jeu deux langues ou variétés de langue dans une communauté linguistique (cf. Fishman, 1971).

Par ailleurs, on a pu remettre en cause certains critères caractéristiques de la diglossie, ou du moins la façon dont Ferguson les aborde, comme la controversée *complexité grammaticale*, ou le rapport exact qu'entretiennent la variété haute — H(igh) — et la variété basse — L(ow). Enfin, on s'est interrogé sur la véritable plausibilité d'une troisième place entre la simple variation au sein d'une langue et le franc bilinguisme, l'existence de laquelle place est bien le centre du propos de Ferguson.

On rejettera ici tout assouplissement de la définition de la diglossie (suivant Hawkins, 1983:228-231, contre les propositions de Fishman, 1971), et on conservera les trois critères définitoires suivants (repris de Koch, 1997:225, d'après la notion de *diglossie prototypique* de Lüdi, 1990:311) :

(58) Une situation linguistique est à décrire comme une diglossie si et seulement si :

- (i) **A:** Es besteht ein starker Abstand zwischen den beiden Sprachformen. Da man dies eher bei verschiedenen Sprachen erwartet

(souvent grammaticalement plus complexe), véhicule d'une quantité de littérature écrite grande et respectée, provenant soit d'une période antérieure soit d'une autre communauté linguistique, qui est apprise grandement à travers l'éducation formelle et qui est utilisée dans la plupart des écrits et des paroles formelles, mais par aucun secteur de la communauté dans la conversation ordinaire." (traduction: BM)

(wenn auch nicht notwendigerweise antreffen muß), erscheint der Überschneidungsbereich mit Parameter VV aus der Sicht von A als peripher.

- (ii) **VV**: Die beiden Sprachformen sind nicht nur miteinander verwandt, sondern werden sogar als Varietäten einer Sprache empfunden. Da man bei Varietäten einer Sprache eher das Gemeinsame im Vordergrund sieht, erscheint der Überschneidungsbereich mit Parameter A aus der Sicht von VV als peripher.
- (iii) **L//H**: Zwischen den beiden Varietäten herrscht eine strikte Funktionstrennung nach L und H, was wiederum eher typisch für verschiedene Sprachen ist.¹²

À notre avis, ce que l'hypothèse de Ferguson capture se situe essentiellement autour de ces trois propriétés et de leurs contradictions, que je formule ainsi dans le cadre théorique général de cette thèse :

Définition 2 (la diglossie) *Il existe des conditions sociolinguistiques particulières, fondées sur le choix des formes linguistiques à employer*

12. Traduction (BM et SH) :

- (i) **A** : Il existe un fort écart entre les deux formes de langue. Comme on s'attend à cela plutôt de la part de langues distinctes (ce qui n'est pas nécessairement le cas), l'intersection avec le paramètre VV paraît marginale du point de vue du paramètre A.
- (ii) **VV** : Les deux formes de langues ne sont pas seulement apparentées, mais elles sont même perçues comme des variétés d'une même langue. Comme dans le cas de deux variétés d'une langue on perçoit en premier lieu ce qu'elles ont en commun, l'intersection avec le paramètre A paraît marginale du point de vue du paramètre VV.
- (iii) **L//H** : Entre les deux variétés règne une fracture fonctionnelle stricte qui suit l'opposition entre L et H, ce qui nous ramène vers ce qui est typique de langues distinctes.

3.4 Hypothèse adoptée : la diglossie

ou non pour remplir respectivement les fonctions hautes et les fonctions basses de la langue, qui permettent et provoquent l'existence durable de situations contradictoires où ce qui est perçu comme une langue externe unique est implémenté par deux langues internes distinctes par les locuteurs.

L'hypothèse de la diglossie implique au final de se satisfaire du paradoxe suivant : le *continuum macroscopique* — la langue externe unique — n'empêche pas le *discontinu microscopique* — la production de ce continu par deux grammaires identifiables et discontinues.

3.4.2 Application à la situation française contemporaine

Parmi les trois propriétés énumérées ci-dessus en (58-i-ii-iii), la première est très linguistique, la seconde et la troisième sont au contraire sociolinguistiques.

3.4.2.1 L'aspect proprement linguistique

Pour parler de diglossie en France, il faut qu'entre les formes linguistiques des énoncés remplissant les fonctions H et celles remplissant les fonctions L, il y ait un écart grammatical suffisamment *grand* (cf. critère (i)). C'est cette problématique qui sera débattue dans les parties II, III et IV. Et on défendra l'idée que l'écart en question entre le FCT et le FD est bien de la taille requise, et que les énoncés de français dans leur ensemble sont générés par deux grammaires distinctes, donnant à voir le discontinu microscopique évoqué ci-dessus.

3.4.2.2 L'aspect proprement sociolinguistique

Pour parler de diglossie en France, il faut une fracture fonctionnelle entre les formes linguistiques du FCT et celles du FD, ressenties comme appartenant à la même langue, ou à deux variétés de la même langue (cf. critères (ii) et (iii)).

On n'investira pas cette problématique ici. Il faudrait pour cela examiner et discuter en détail l'ensemble de l'article fondateur de Ferguson. Simplement, on laissera au lecteur le soin de se laisser convaincre ou pas, de l'adéquation du concept fergusonien de diglossie pour décrire la face sociologique de la variation en français, à travers la lecture de quelques longs extraits traduits et adaptés de l'article de Ferguson (après remplacement de H par FCT et de L par FD). À mon avis, c'est très parlant, et répond bien à l'intuition d'une bipolarité de la répartition des variantes du français¹³.

[pp. 328, 329] **1. Fonction.** Un des traits les plus importants de la diglossie est la spécialisation des fonctions pour le FCT et le FD. Dans un ensemble de situations, seulement le FCT est approprié, et dans un autre, seulement le FD, les deux ensembles se recouvrant très peu. [...] On peut difficilement surestimer l'importance d'utiliser la bonne variété dans la bonne situation. Un étranger qui apprend à parler couramment le FD et ensuite l'utilise dans un discours formel est l'objet de ridicule. Un membre de la communauté linguistique qui utilise le FCT dans une situation purement conversationnelle ou dans une activité informelle, comme faire les courses, est également l'objet de ridicule. [...] (*traduction arrangée*

13. On me pardonnera de tordre ainsi le propos de Ferguson dans la forme, que je on pense respecter dans le fond.

3.4 Hypothèse adoptée : la diglossie

*de : « 1. **Function.** One of the most important features of diglossia is the specialization of fonction for H and L. In one set of situations only H is appropriate and in another only L, with the two sets overlapping only very slightly. » etc.)*

[pp. 329, 330] **2. Prestige.** [...] les locuteurs considèrent le FCT comme supérieur au FD sur un certain nombre de points. Parfois le ressenti est tellement fort que le FCT est le seul considéré comme réel et le FD est dit « ne pas exister ». [...]

Même si le sentiment d'une réalité et d'une supériorité du FCT n'est pas si fort, il y a habituellement la croyance que le FCT est d'une certaine façon plus beau, plus logique, mieux à même d'exprimer les idées importantes, et ainsi de suite. Et cette croyance existe également chez des locuteurs dont la maîtrise du FCT est limitée.

[p. 330] **3. Héritage littéraire.** [...] il y a une quantité considérable de littérature écrite de FCT qui est tenue en haute estime par la communauté linguistique, et les productions littéraires contemporaines en FCT par les membres de la communauté linguistique sont ressenties comme faisant partie de cette littérature existant par ailleurs.

[p. 331] **4. Acquisition.** [...] les adultes utilisent invariablement le FD pour parler aux enfants, et les enfants utilisent le FD pour parler entre eux. Ainsi, le FD est invariablement appris par les enfants d'une façon qui pourrait être considérée comme l'apprentissage « normal » d'une langue maternelle. Les enfants peuvent entendre le FCT de temps en temps, mais son apprentissage effectif est principalement accompli au moyen d'une éducation formelle [...].

Cette différence dans la méthode d'acquisition est très importante. Le locuteur se sent à l'aise en FD à un degré qu'il n'atteint quasiment jamais en FCT. La structure grammaticale du FD est apprise sans discussion explicite de concepts grammaticaux ; la grammaire du FCT est apprise en termes de « règles » et de normes à imiter.

[pp. 331, 332] **5. Standardisation.** [...] il y a une forte tradition d'études grammaticales du FCT. Il y a des grammaires, des dictionnaires, des traités de prononciation, de style, etc. Il y a une norme établie pour la prononciation, la grammaire, et le vocabulaire, qui autorise la variation seulement dans certaines limites. L'orthographe est bien établie et connaît peu de variation. Par contraste, les études descriptives et normatives du FD sont soit non-existantes soit relativement récentes et en faible quantité. [...] Il n'y a pas d'orthographe en place et il y a une large variation dans la prononciation, la grammaire, et le vocabulaire.

[pp. 336, 337] Comme elle a été caractérisée ici, la diglossie diffère du plus répandu « standard-et-ses-dialectes » en ce qu'aucun groupe de la communauté diglossique n'utilise véritablement le FCT comme média de la conversation ordinaire, et que toute tentative de le faire est ressentie soit comme pédante et artificielle, soit en un sens comme déloyale envers la communauté.

[p. 338] La diglossie semble acceptée et ne pas être considérée comme un « problème » par la communauté dans laquelle elle fait force, jusqu'à ce que certaines tendances apparaissent dans la communauté. Elles incluent des tendances vers (a) une littéracie plus répandue [...], (b) une communication plus

3.4 Hypothèse adoptée : la diglossie

vaste entre les différents groupes régionaux et sociaux de la communauté [...], (c) le désir d'une langue standard « nationale » complète comme attribut d'autonomie ou de souveraineté.

Quand ces tendances apparaissent, des leaders de la communauté commencent à appeler à l'unification de la langue, et pour cela, des tendances effectives vers l'unification commencent à avoir lieu. Ces individus tendent à soutenir soit l'adoption du FCT soit d'une forme de FD comme standard, moins souvent l'adoption d'un FCT ou FD modifié, une variété « mixte » d'une quelconque sorte.

Les tenants du FCT argumentent que le FCT doit être adopté parce qu'il connecte la communauté avec son passé glorieux ou avec la communauté mondiale, et parce que c'est un facteur unificateur naturel, par opposition à la nature divisante des dialectes de FD. À côté de ces deux arguments solidement fondés, il y a habituellement des prétextes fondés sur les croyances de la communauté dans la supériorité du FCT : il est plus beau, plus expressif, plus logique, [...] ou quelques croyances spécifiques qu'elle ait.

Les tenants du FD argumentent qu'une variété du FD doit être adoptée parce qu'il est plus près de la pensée et du ressenti réels des gens ; ça facilite le problème de l'éducation puisque les gens en ont déjà acquis une connaissance basique dans leur enfance ; et c'est un outil de communication plus efficace à tous les niveaux. En plus de ces arguments solidement fondés, il y a souvent une grande mise en avant de points de moindre importance, telle la vitalité de la métaphore dans la langue courante, le fait que d'autres « nations modernes »

écrivent grandement comme elles parlent, etc.

[. . .] Sans quoi, le FCT s'éteint et devient une langue de culture ou liturgique étudiée seulement par les érudits ou les spécialistes, et non activement employée dans la communauté. [...]

3.4.3 Les choix terminologiques

Il existe de nombreux termes pour désigner les différentes variétés de français marquées sociolinguistiquement, et certains ont déjà été proposés pour désigner les ensembles de variantes qui remplissent respectivement les fonctions hautes et les fonctions basses de la langue, ou des variantes de cette façon de décrire la variation sociolinguistique individuelle en termes bipolaires.

Pourtant, plusieurs raisons m'ont amené à forger de nouveaux termes. Certains termes qui pourraient être facilement employés en termes de bipolarité sont aussi souvent employés dans le cadre de descriptions multipolaires, ou apolaires. Et les termes habituels ne sont pas neutres dans le langage courant, et sont rapidement assimilés aux notions de bon et de mauvais français, ce que l'on souhaite éviter absolument.

populaire Le terme est également associé à des descriptions multipolaires, et exprime dans le public non-spécialiste des jugements sociaux qui empêchent une description dépassionnée de la situation grammaticale.

familier, courant, ordinaire, soutenu, littéraire . . . Ces termes ne sont pas adaptés à des descriptions bipolaires et font l'objet de considérations sociales passionnées.

écrit vs. parlé/oral Ces termes invoquent des motivations internes de la variation qui ne sont pas retenues ici comme suffisamment pré-

3.4 Hypothèse adoptée : la diglossie

dictives. De plus, ils proposent une bipolarisation qui n'est pas celle retenue ici.

standard vs. non-standard Ces termes sont tout à fait compatibles avec l'hypothèse de la diglossie. Seulement, il est possible que le concept de standard puisse évoquer l'idée naïve d'une opposition entre une grammaire qui serait bien organisée, et une autre qui n'aurait pas de bonne organisation et serait anarchique. Ou bien, on pourrait imaginer que les énoncés non-standard seraient pas banals, peu courants, exceptionnels. Aucune de ces deux idées n'est impliquée par l'hypothèse, au contraire.

standard vs. dialectal Ces termes, utilisés par exemple par Zribi-Hertz (2006), conviennent également pour décrire des situations de diglossie, par exemple par comparaison avec l'arabe standard et l'arabe dialectal. Mais on pourrait imaginer des implications spontanées non-souhaitées dans le sens de différenciations grammaticales fortes entre les dialectes non-standard, alors qu'on souhaite plutôt mettre en avant leur relative cohérence entre eux, au moins dans le cas de la France, et leur forte divergence face à la langue standard.

français du dimanche vs. français de tous les jours Blanche-Benveniste *et al.* (1990:211ss.) présentent le terme de 'français du dimanche', par métaphore avec les habits du dimanche, pour décrire la capacité des locuteurs du français de se placer dans un registre soutenu dans certaines situations formelles, par opposition aux situations informelles où ils gardent leur langue de tous les jours (comme on porte ses habits de tous les jours en dehors des situations particulières qui exigent un code vestimentaire particulier). Il y a le risque que cette idée de « langue d'apparat » (Blanche-Benveniste *et al.*, 1990:211) puisse laisser penser que les

Chapitre 3 Sujet

variantes étudiées sont intrinsèquement reconnaissables comme formelles ou non, c'est-à-dire qu'une forme grammaticale aurait par sa forme même plus d'apparat qu'une autre.

Le choix est donc fait de forger deux termes nouveaux. On souhaite qu'ils respectent quelques caractéristiques essentielles :

bipolarité sociale Les termes devront évoquer la bipolarisation sociale telle que suggérée par Ferguson (1959) (les fonctions hautes vs basses) ou par Koch et Oesterreicher (1990) (la communication distante vs. proche).

neutralité grammaticale Les termes ne devront pas impliquer de biais grammatical concernant les faits compris dans les deux ensembles polarisés, en dehors du hasard historique.

neutralité médiale Les termes excluront toute idée de tendance ou exclusivité des pôles envers l'écrit ou l'oral.

neutralité de jugement Les termes doivent éviter tout réflexe de jugement extragrammatical quelconque à l'encontre des locuteurs.

Les termes choisis sont *français classique tardif* et *français démotique*, que l'on justifie ainsi :

classique tardif Langue issue du français classique, indentifiée historiquement et cernée grammaticalement par la normativisation.

démotique Ce terme fait référence à la situation grecque, où la *demotiki* (le grecque moderne) a été reconnue comme langue à part entière face à la *katharevousa* (le grecque 'pur') (elle a même atteint le statut de langue officielle). Le terme évoque également le peuple, et ainsi cet aspect de langue acquise « sur les genoux de la mère »

3.4 Hypothèse adoptée : la diglossie

(Ferguson, 1959), transmise informellement entre les locuteurs, par opposition à la langue administrative, académique, littéraire, officielle, transmise et entretenue formellement par les institutions et l'éducation.

diachronie Bien que cette thèse ne souhaite pas s'investir dans des considérations diachroniques, on peut se réjouir que les termes choisis se prêtent volontiers à des hypothèses concernant le rapport génétique entre les deux grammaires à décrire, puisqu'un tel rapport est explicitement évoqué par Ferguson (1959). Pour accentuer cet aspect, on pourrait d'ailleurs insister en qualifiant le français démotique de contemporain.

3.4.4 Les zones zribi-hertziennes

Massot (2003:79) et Zribi-Hertz (2006:4) proposent un schéma pour visualiser l'idée de la diglossie comme une situation où deux ensembles d'énoncés sont distincts tout en ayant une intersection non-négligeable. On reproduit le schéma de Zribi-Hertz (2006) sous la figure 3.1, à la petite différence que l'on substitue ses termes 'dialectal' et 'standard' par 'démotique' et 'classique tardif', et on utilisera ses termes de zone 1, zone 2 et zone 3 (Z1, Z2 et Z3) pour classer les énoncés en fonction de leur appartenance à l'un ou l'autre des deux ensembles, tels qu'ils ont été définis plus haut par les exemples reproduits en (59).

- (59) a. a kel ør arivt-il ? \in (FCT \setminus FD = Z1)
b. a kel ør esk il-ariv ? \in (FCT \cap FD = Z2)
c. il-ariv a kel ør ? \in (FD \setminus FCT = Z3)

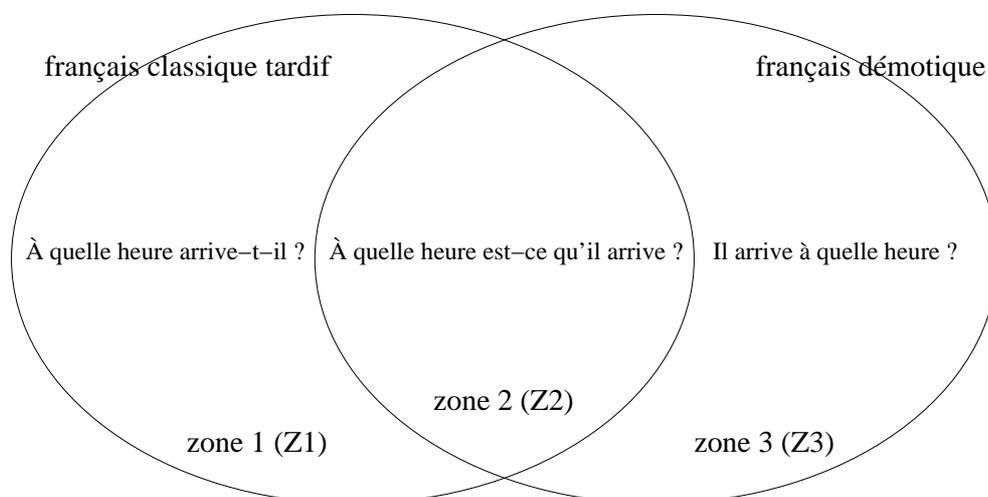


FIGURE 3.1: Les zones 1, 2 et 3 de Zribi-Hertz

3.5 Conclusion

Je me propose d'étudier la variation morphosyntaxique diaphasique des locuteurs francophones (de France). Je chercherai à soutenir l'hypothèse d'une diglossie, telle qu'elle a été décrite par Ferguson. Parmi les trois critères retenus pour pouvoir parler de diglossie, je laisse au lecteur le soin d'apprécier leurs aspects sociolinguistiques (deux langues ou variétés d'une langue, ici le FCT et le FD, perçues comme une seule langue, et la fracture fonctionnelle entre elles, ici le FCT pour remplir les fonctions H, et le FD pour remplir les fonctions L), et je me concentrerai sur le premier critère (l'écart grammatical significatif entre le FCT et le FD). Je le reformule en termes de grammaires internes distinctes chez les locuteurs. Mon objectif sera donc d'explorer l'idée suivante, et d'argumenter en sa faveur : pour représenter la compétence qu'ont les locuteurs du français concernant la variation morphosyntaxique, on a intérêt à la décrire

3.5 Conclusion

comme l'implémentation de deux grammaires distinctes, l'une produisant du FCT, l'autre produisant du FD.

Chapitre 4

L'argumentation

LA VALEUR de l'hypothèse de la diglossie doit se mesurer à plusieurs paramètres. Le premier de ces paramètres sera la parcimonie. L'hypothèse de la diglossie est particulièrement mal notée de ce point de vue, puisqu'elle complexifie les axiomes du modèle (« les locuteurs ont deux grammaires » vs. « les locuteurs ont une (seule) grammaire »). Les autres paramètres devront donc montrer des avantages suffisants pour compenser ce manque de parcimonie, voire pour le renverser (en montrant que sans cette hypothèse, la modélisation de la grammaire gagne en complexité, et donc perd en parcimonie).

Les avantages qu'on propose de présenter dans cette thèse sont de deux natures. Les premiers sont internes à la description théorique. On les résumera en disant que l'hypothèse « soulage le théoricien et le typologue ». Les seconds concernent l'adéquation à l'empirie. L'hypothèse sera dite « soulager le typologue et l'empiriste ».

Épistémologiquement, on constate que l'hypothèse prédit plus de données négatives. Elle est donc plus informative et mieux à même d'être falsifiée que l'hypothèse d'une grammaire unique (ce qui ne suppose pas en soit qu'elle soit meilleure dans les faits).

4.1 Soulager le théoricien et le typologue

Dans la partie II, consacrée au nombre dans le domaine nominal, on décrira une grammaire du nombre en français, à partir de deux ensembles de données légèrement distincts. Le premier ensemble contient les données en (60) tandis que le second les exclut, comme indiqué en (61).

- (60) a. le $fi\check{e}(z)$ $\check{a}raze$
b. $fi\check{e}(z)$ e fa $s-s\check{o}$ baty tut la nyi

- (61) a. le $fi\check{e}(*z)$ $\check{a}raze$
b. $*fi\check{e}(z)$ e fa $s-s\check{o}$ baty tut la nyi

À partir des descriptions de la morphologie du nombre de ces deux ensembles (chapitre 5), et à partir d'un petit modèle comparatif des statuts du nombre autour du phénomène des noms nus (chapitre 6), on proposera d'accorder un certain statut au nombre du français (une faible autonomie, seulement morphologique, et pas d'autonomie syntaxique). On suggèrera que, malgré les nuances visibles entre (60) et (61), il est sûrement peu intéressant de supposer deux grammaires distinctes du nombre en FCT et en FD.

Dans la partie III, le chapitre 7 est consacré à la négation. Il proposera de réviser la description de paires d'exemples comme (62) au moyen d'une structure sous-jacente commune.

- (62) a. il-n-a pa $k\check{o}pri$
b. il-a pa $k\check{o}pri$

On suggèrera au contraire qu'il est souhaitable d'y voir deux structures syntaxiques fondamentalement distinctes, une négation verbale face à une

négarion de constituant.

Le chapitre 8 présente les travaux de Lambrecht sur la syntaxe de la structure informationnelle et ses applications au français. On confrontera ses propositions au sein des différentes hypothèses consacrées à la variation morphosyntaxique, pour conclure que la diglossie est vraisemblablement la plus compatible avec ses résultats, et avec le constat que le FCT et le FD ne présentent pas la même variation syntaxique en lien avec la structure informationnelle.

4.2 Soulager le typologue et l'empiriste

Dans les parties II et III, la séparation des données en deux ensembles est peu justifiée empiriquement. On se contente d'y invoquer des jugements sociolinguistiques supposés consensuels, et des jugements de grammaticalité acrobatiques (« Est-ce que cette phrase est bien formée *dans une situation informelle* ? »).

La partie IV se consacre à l'étude de la cohérence grammaticale des énoncés, c'est-à-dire à vérifier que les locuteurs n'activent bien qu'une seule des deux grammaires que l'hypothèse leur prête au sein du même énoncé (section 9.1 du chapitre 9).

On fait appel aux jugements de locuteurs dans un premier temps (section 9.2 du chapitre 9). Mais cette fois, au lieu de demander simplement si deux variantes appartiennent au même « registre », on testera concrètement la compatibilité de deux variantes, vues comme variantes du même registre ou non, *au sein du même énoncé*.

Puis, dans un deuxième temps (chapitre 10), et puisqu'on aura constaté que les jugements de locuteurs ne sont pas parfaitement fiables, on partira à la recherche sur corpus de ces énoncés combinant des variantes prédites

compatibles, en espérant ne pas tomber sur des énoncés combinant des variantes prédites incompatibles.

Enfin, le chapitre 11 proposera d'interpréter des énoncés attestés déviant au regard de la prédiction du chapitre 9, notamment à travers l'idée d'interférences.

On espère ainsi soulager l'empiriste qui souhaite que les données soient suffisamment belles pour qu'elles aient pu induire l'hypothèse soutenue à elles seules. Et soulager à nouveau le typologue, qui commençait à trouver le mélange des ingrédients de la recette de la variation en français de plus en plus original et à douter des qualités culinaires et de la comestibilité linguistique du résultat.

4.3 Conclusion

Si l'hypothèse de la diglossie simplifie nettement les descriptions, si elle évite de décrire un type original seulement pour le français, alors, même si elle implique des axiomes plus complexes, elle permet de proposer un modèle plus simple, et plus parcimonieux.

Le nombre ne constituera pas un cas privilégié d'une telle simplification. En revanche, les considérations sur la négation et la structure informationnelle feront de l'hypothèse de la diglossie un modèle avantageux descriptivement et typologiquement.

Et, à chaque occasion manquée de trouver un contre-exemple à l'idée que les locuteurs n'activent qu'une grammaire à la fois, les considérations empiriques rapprocheront de la conviction que les locuteurs ont bien une double-grammaire.

On pourra conclure finalement que la diglossie est provisoirement adéquate et même supérieure aux autres hypothèses pour rendre compte de

4.3 Conclusion

la variation morphosyntaxique individuelle du français.

Deuxième partie

Morphosyntaxe du nombre dans le domaine nominal

Au pluriel on dit des « cristaux »
car il y a plusieurs cristaux.

anonyme (perles du bac 2007)

CETTE partie concerne la *grammaire* du nombre. Il s'agit donc d'observer les comportements morphologiques et syntaxiques liés au nombre, et non de disserter sur la réalité sémantique exprimée sous l'étiquette du nombre.

Pour bien distinguer grammaire et sémantique du nombre, on observera les faits suivants. Ils cherchent simplement à montrer que repérer un trait grammatical de nombre n'est ni nécessaire ni suffisant pour prédire l'interprétation d'un groupe nominal.

Le pluriel a la réputation de dénoter les quantités supérieures ou égales à deux. Le dialogue en (1) montre pourtant qu'une morphologie de pluriel est compatible avec un singulier.

- (1) A : t-a aŋte de tomat ?
B : ui yn.

La lecture générique a vraisemblablement une interprétation unique, fortement liée à ce qu'on appelle la lecture d'espèce (la dénotation de tous les membres d'un ensemble défini par des propriétés). Cette lecture d'espèce est obtenue avec le singulier pour les noms massiques (*l vĕ*) et avec le pluriel pour les noms comptables (*lez āfā*). Mais on peut obtenir l'interprétation générique d'un nom comptable avec une morphologie de pluriel (2ab) ou de singulier (2cd).

- (2) a. (i) le likorn sa-a k yn korn
(ii) le likorn ez-ō k yn korn
b. (i) de likorn sa-a k yn korn
(ii) de likorn ez-ō k yn korn

- c. (i) yn likorn sa-a k yn korn
- (ii) yn likorn el-a k yn korn
- d. la likorn el-a k yn korn

Certaines morphologies de pluriel sont simplement incompatibles avec un dénombrement. En (3), B aimerait savoir combien de « chantiers » de travaux A a faits (par exemple dans la salle-de-bains, dans la cuisine, dans le sous-sol...) ou combien de « types » de travaux (par exemple des travaux de peinture, de menuiserie...). Mais il n’y a pas de façon simple de formuler la question, en tout cas sans se servir de ce que l’on pourrait appeler un classificateur — comme *chantier* ou *type*. De ce point de vue, *de travo*, dont la morphologie est explicitement plurielle, se comporte strictement comme un singulier massique.

- (3) A : ʒ e fe de travo je mua
 B : #kõbiẽ?

Comme l’objectif ici n’est pas d’induire l’hypothèse de la diglossie à partir des observations, mais au contraire de la tester face à ces dernières, on assumera de bout en bout l’idée d’une grammaire du nombre du FD, avec ses propres données, indépendante de celle du FCT et de ses données.

Enfin, on attire une dernière fois l’attention sur la limitation principale et volontaire de cette partie : il ne s’agira bien que du nombre dans le domaine nominal, à l’intérieur de syntagmes nominaux canoniques, sans aucune considération sur les phénomènes d’accord à l’extérieur de ce domaine, ni même de considération sur les pronoms, supposés proportionnels à ce domaine.

Cette partie s’organise en deux chapitres. Le chapitre 5 expose l’ensemble des faits morphologiques et leur pertinence. Puis, le chapitre 6 prend le parti d’aller chercher des réponses sur cette variété et cette va-

riation dans la syntaxe et l'interface syntaxe-sémantique. En particulier, on essaiera de placer le FCT et le FD dans une brève typologie des noms nus qui met en relation la morphologie, la syntaxe et la sémantique. On conclura que, dans l'état actuel de cette étude, il est préférable de décrire une seule grammaire du nombre en français, et de s'expliquer autrement les quelques points de divergence entre le FCT et le FD.

Chapitre 5

Morphologies du nombre en français

DANS ce chapitre, on détaille l'ensemble des faits morphologiques de nombre observables dans le domaine nominal en français. Ils seront présentés en trois sections, dans l'ordre linéaire de leur apparition dans le syntagme nominal : les déterminants, puis les adjectifs pré-nominaux, enfin les noms et adjectifs post-nominaux. On séparera, chaque fois qu'il y aura lieu de le faire, les données de FCT et les données de FD. Une quatrième section synthétisera les points pertinents : la variété et la variation de ces morphologies.

5.1 Les déterminants

Les déterminants montrent deux faits morphologiques qu'on distinguera pour faciliter la présentation. Le tableau 5.1 (p. 141) donne une vue d'ensemble des formes phonologiques des déterminants du français.

5.1.1 Les suffixes synthétiques

Tout d'abord, les déterminants du français portent pour les principaux d'entre eux une marque suffixale de nombre, qui se synthétise avec la marque de genre en la neutralisant au pluriel (4a), ou bien en la supprimant complètement (4b). Certains déterminants ne sont pas concernés par cette morphologie (4c).

- (4) a. (i) mō balō
(ii) ma bal
(iii) me balō/bal
b. (i) notr balō/bal
(ii) no balō/bal
c. lør balō/bal

Le singulier et le pluriel sont donc explicitement marqués. C'est-à-dire en particulier que le marquage de singulier n'est pas obtenu par l'absence de marque de pluriel, mais bien par une morphologie de singulier explicite (-ō ou -a vs. -e en (4a)).

Chacun des ensembles de marques est suffisamment peu productif pour qu'on se pose la question d'une possible supplétion. Mais, d'une part, la productivité ne peut pas être vue comme une nécessité dans une classe fermée. Et d'autre part, les consonnes initiales de ces déterminants sont d'une constance suffisante pour qu'on souhaite distinguer une 'racine' d'un suffixe : $ma = m+a$. Ce que Csécsy (1971:95) résume ainsi :



5.1.2 Le suffixe de pluriel

Ensuite, les déterminants montrent un suffixe de pluriel. Il apparaît sous la forme d'un [z] conditionné par l'entourage phonologique, ce qu'il est convenu d'appeler une consonne de liaison¹. Il s'agit d'une liaison obligatoire. L'adjectif possessif de 3^e personne du pluriel, comme observé en (5), possède ainsi la forme phonologique sous-jacente /lør/ au singulier et /lør(z)/ au pluriel, où la consonne de liaison *obligatoire* est notée entre parenthèses.

- (5) a. lør propozisiõ
 b. lør ide
 c. lørz ide

Tous les déterminants présentent ce /(z)/, mais il faut pourtant distinguer, si l'on veut être précis, entre deux classes de déterminants. Il y a d'un côté ceux, tels *leur(s)*, pour qui la consonne de liaison est un suffixe autonome dans le sens où il s'oppose à sa simple absence, laquelle révèle une marque de singulier — (5c) vs. (5b). De l'autre côté, on a les déterminants pour lesquelles la consonne de liaison semble synthétisée avec le suffixe de nombre-genre observé en (4). Comme le montre (6), /no/, obtenu en retirant la consonne de liaison de /no(z)/, n'est pas un déterminant du français.

- (6) no*(z) ide

On affine alors ainsi les représentations : /notr/ possède le trait SG et /no(z)/ le trait PL tandis que /lør+∅/ possède le trait -PL et /lør+(z)/ le trait +PL. Des adjectifs indéfinis montrent des distributions lacunaires,

1. Cf. section 2.5.5 pour le traitement de la liaison adopté ici.

certains n'ayant de formes qu'au singulier ou qu'au pluriel. On y trouve aussi des cas de distinction de genre au pluriel (cf. tableau 5.1).

5.1.3 Synthèse

Le tableau 5.1 regroupe l'ensemble des déterminants du français classés selon leurs traits de nombre (et de genre) et représentés par leurs formes phonologiques.

Ces faits sont identiques en FD et en FCT. On pourra simplement remarquer que certains items sont pour le moins rares en FD : *nul*, *maints*, voire l'emploi de *tout* comme adjectif indéfini. De plus, on notera l'avènement (plutôt en FD) du déterminant $z\emptyset$, emprunté à l'anglais. Tandis que sa nature et son emploi restent encore à décrire, on notera déjà qu'il neutralise le genre et le nombre (même si son emploi au pluriel ne semble pas acquis) :

(7) s-e vrēmã $z\emptyset$ āduj s mek / set nana / ?se mek/ ?se nana

5.2 Les adjectifs pré-nominaux

Les adjectifs pré-nominaux présentent également une morphologie de nombre sous la forme d'un [z] de liaison. On en voit les effets en (8).

- (8) a. notr nuvel propozisiõ
b. no nuvel propozisiõ
c. notr nuvel ide
d. no nuvelz ide

(8) résume les faits de FCT sur ce point : cette liaison y est obligatoire. Mais les faits de FD en (9) demandent examen.

5.2 Les adjectifs pré-nominaux

	MASC SG	FÉM SG	MASC PL	FÉM PL
article défini	l	la (l)	le(z)	
article indéfini	ě(n)	yn	de(z)	
article partitif	dy (d l)	d la (d l)		
adjectif démonstratif	s (set)	set	se(z)	
adjectif possessif	mõ (mõn)	ma (mõn)	me(z)	
	tõ (tõn)	ta (tõn)	te(z)	
	sõ (sõn)	sa (sõn)	se(z)	
	notr		no(z)	
	votr		vo(z)	
	lør		lør+(z)	
adjectifs indéfinis	okě(n)	okyn	×	
	ƒak		×	
	nyl		×	
	tel		tel+(z)	
	tu(t)	tut	tu+(z)	tut+(z)
	×		diver+(z)	divers+(z)
	×		sertě+(z)	serten+(z)
	×		diferã+(z)	diferãt+(z)
	×		mě+(z)	mět+(z)
	×		plysiør+(z)	
×		kelk+(z)		
cardinaux			dø(z) trwa(z) katr. . .	

(détachés et entre parenthèses, les allomorphes apparaissant devant voyelle)

TABLEAU 5.1: Les déterminants du français (formes phonologiques)

(9) FD :

- a. (i) l otr abi
(ii) lez otr abi
(iii) lez otrz abi
- b. (i) la nuvel ide
(ii) le nuvel ide
(iii) le nuvelz ide
- c. (i) set ãsien ami
(ii) sez ãsien ami
(iii) sez ãsienz ami
- d. (i) s ptit ãzẽ
(ii) ?se ptit ãzẽ
(iii) se ptiz ãzẽ
- e. (i) yn zoli abej
(ii) ??trwa zoli abej
(iii) trwa zoliz abej

La réalisation de la liaison est influencée par le contexte phonologique gauche. Intuitivement, disons que plus le contexte gauche contient de consonnes, plus il est possible de ne pas réaliser la liaison. Cependant, l'influence du contexte gauche n'est pas connue comme facteur influençant la liaison. Au contraire, Ågren (1973:127-130) montre même, sur corpus, que dans un autre cas possible de rôle du contexte gauche, les cas NomPl+Adj, ce paramètre n'a aucune influence.

Je propose alors trois interprétations de ces faits de variation, les deux dernières ayant un aspect problématique :

1. On peut envisager une variation lexicale : le phénomène observé, s'il est avéré diachronique, atteint déjà certains items du lexique

5.2 Les adjectifs pré-nominaux

mais pas d'autres. Je n'ai pas exploré cette hypothèse plus avant, et je m'arrêterai à son évocation.

2. Tous les adjectifs pré-nominaux portent ce suffixe en toutes circonstances (c'est-à-dire qu'il est présent dans leur représentation phonologique). La variation est alors phonologique. Ce qui peut être couteux descriptivement, c'est de devoir décrire un fait nouveau autour de la liaison, sans qu'il puisse servir à des cas similaires ailleurs dans la grammaire.

On dira dans ce cas que cette liaison est obligatoire (dans le sens donné à la section 2.5.5), comme le montre (9e), mais que *sa réalisation peut être annulée par des consonnes dans le contexte gauche*.

3. On parle d'une variation morphosyntaxique : certains adjectifs (tels *zoli*) reçoivent systématiquement cette consonne, d'autres la reçoivent variablement, en fonction du poids du groupe consonantique à la finale de la racine de l'adjectif. Mais cette description viole le *Principe d'Inaccessibilité de la Phonologie par la Syntaxe*².

La violation du PIPS me semblant plus problématique que l'originalité des propriétés de cette liaison, je n'intégrerai pas dans mes analyses l'idée que, au moins en FD, une variante ne marque pas le nombre sur les adjectifs pré-nominaux.

On retient donc les formes phonologiques suivantes (où les chevrons notent ce nouveau type de consonne de liaison) :

- (10) FCT :
- singulier : /otr-∅/ /nuvel-∅/ /ãsien-∅/ /pti(t)-∅/ /zoli-∅/
pluriel : /otr-(z)/ /nuvel-(z)/ /ãsien-(z)/ /pti(t)-(z)/ /zoli-(z)/

2. Cf. par exemple Miller *et al.* (1992) pour une discussion sur le PIPS à partir de faits du français.

(11) FD :

singulier : /otr-∅/ /nuvel-∅/ /ãsien-∅/ /pti(t)-∅/ /zoli-∅/

pluriel : /otr-<z>/ /nuvel-<z>/ /ãsien-<z>/ /pti(t)-<z>/ /zoli-<z>/

5.3 Les noms et les adjectifs post-nominaux

Les noms et les adjectifs post-nominaux présentent deux faits morphologiques liés au nombre. L'un est un fait de liaison, l'autre ne l'est pas.

5.3.1 Le suffixe de pluriel

En FCT, les noms et les adjectifs post-nominaux sont fléchis en nombre au pluriel par la suffixation d'un /{z}/ de liaison *facultatif*.

(12) FCT

a. la persòn prezãt

b. le persòn prezãt

c. la persòn aze e depãdãt

d. le persòn(z) aze(z) e depãdãt

e. singulier : /persòn-∅/ /prezãt-∅/ /aze-∅/ /depãdãt-∅/

pluriel : /persòn-{z}/ /prezãt-{z}/ /aze-{z}/ /depãdãt-{z}/

Au contraire, et c'est ici un point central de ma description, ce suffixe n'apparaît pas en FD. Ce jugement est issu de mon auto-conviction, d'une petite expérience, et de données de corpus de seconde main.

Donc, à part le fait que je suis convaincu de ne pas fléchir les noms en nombre dans ma parole informelle, j'ai observé le fait suivant. À plusieurs

5.3 Les noms et les adjectifs post-nominaux

occasions, des groupes d'étudiants de première année de linguistique ont été interrogés. En imaginant de produire (13) dans une situation formelle, comme les informations radiophoniques, ils ont accepté de marquer le nom *fiẽ* avec un *z*, sans que ce soit nécessaire pour ressentir la formalité de la situation.

(13) FCT : *plysiør fiẽ(z) āraze õt ete rperesamã*

En revanche, lorsque (14) est produit dans une situation informelle, comme lorsqu'une personne, au lever, cherche ses vêtements, ils n'ont pas souhaité envisager de marquer le nom *foset* d'un *z*.

(14) FD : *merd ʒ truv ply me foset(*z) orãʒ*

Par ailleurs, l'étude de Ågren (1973) semble corrélérer ce sentiment. Il répartit son corpus de conversations radiophoniques en trois registres : un registre *soutenu*, un registre *courant*, et un registre intermédiaire entre les deux. Pour ce classement, il se réfère à Colette Stourdzé qui « propose de distinguer trois sous-niveaux du bon usage [. . .] : *langue familière*, *langue courante* et *langue soignée* » (Ågren, 1973:18, note 16). Ågren admet donc que son corpus ne contient pas de français familier, en tout cas pas de façon consistante à l'intérieur d'une conversation. Et les résultats sont intéressants (tableau 5.2).

Étant donnée la faible fréquence des liaisons dans le niveau courant, notamment par comparaison avec le niveau soigné, on peut facilement s'imaginer que, si Ågren avait eu un corpus de niveau familier, il aurait pu observer une absence totale de liaisons, fait compatible avec l'idée que le FD (compris comme consistant avec le français familier ou le français des situations informelles suggérées à mes étudiants) ne fléchit pas le nom en nombre.

	L[iaisons]	N[on-]L[iaisons]
S[oigné]	108	146
[Intermédiaire]	50	127
C[ourant]	12	72

TABLEAU 5.2: Les liaisons post-nominales et le niveau de langue (Ågren, 1973:127)

D'où les représentations phonologiques ci-dessous :

- (15) FD :
 singulier : ma /*ʃo*set/ /*orã*ʒ/
 pluriel : me /*ʃo*set/ /*orã*ʒ/

5.3.2 Les formes supplétives lexicales

Un petit garçon se promène avec sa maman dans les rues de Paris (où, à notre époque, on n'a pas souvent l'occasion de voir les quadrupèdes en question) :

- Regarde, maman, des cheval !
- Non, mon petit, des chevaux.

Le petit, après un instant de réflexion :

- Et pourtant, ils ressemblent drôlement aux autres.

anecdote rapportée dans Csécsy (1971:100)

Il existe un certain nombre de noms masculins marqués pour le nombre. La plupart des cas sont de la stricte supplétion figée (tableau 5.3), ce qui pourrait donc être négligé, vu comme non pertinent pour la description des processus grammaticaux productifs.

Mais un cas pose la question de la possibilité d'y voir un mécanisme grammatical de morphologie flexionnelle. Il s'agit d'un ensemble de noms

5.3 Les noms et les adjectifs post-nominaux

singulier	pluriel
øf	ø
bøf	bø
siel	siø
aiøl	aiø
øj	ziø
òs	o
msiø	mesiø
madam	medam
madmwazel	medmwazel
bonòm	bonòm ou bõzòm
travaj	travaj ou travo
koraj	koro
vitraj	vitro
emaj	emo
baj	bo
aj	aj ou o
supiraj	supiro

TABLEAU 5.3: Les noms à pluriel supplétif

et d'adjectifs post-nominaux masculins présentant le couple de terminaisons *-al/-o* associé aux valeurs SG/PL. On est devant deux interprétations possibles. Soit il s'agit d'un ensemble de marques purement lexicales et figées, soit c'est une morphologie flexionnelle active dans la grammaire. Un début de réponse peut se trouver dans l'expérience suivante. Aux étudiants de première année de linguistique déjà mentionnés, on a demandé si un néologisme se terminant en *-al* (par exemple *ẽ tokal*) aurait un pluriel en *-o* (donc *de toko*). Ils ont répondu par l'affirmative si on présente ce néologisme comme du français « standard ». Mais, quand on leur a demandé de s'imaginer utiliser ce néologisme dans une situation de français « non standard », alors ils se sont vus très volontiers ne pas produire un pluriel en *-o*, mais au contraire laisser le nom non-marqué quant au nombre (soit *de tokal*).

Ainsi, alors que l'on pourrait chercher à modéliser un mécanisme produisant les couples *-al/-o* en FCT, on s'en passera en FD, et je suggère de supposer que *fval* et *fvo* sont en un sens stockés indépendamment dans le lexique, avec des caractéristiques de nombre et de comptabilité idiosyncratiques, comme peuvent en avoir des mots comme *fyneraj*, ou des comportements face au nombre tels les comportements face au genre des paires telles *vera/tryi*.

5.4 Variété et variation dans la morphologie du nombre

5.4.1 Variété

Un premier point qui mérite attention, c'est le fait que les différents mécanismes morphologiques de marquage du nombre en français sont

5.4 Variété et variation dans la morphologie du nombre

	singulier	pluriel
noms	zurnal fval animal metal ...	zurno fvo animo meto ...
adjectifs	mātal normal rezional myral ...	māto normo reziono myro ...
noms	bal etal fakal vādal petal fytal ...	
adjectifs	fraktal natal bākal sal oval kozal ...	
noms	eto blero ero sizo ...	
adjectifs	rigolo nigo faio pyso ...	

TABLEAU 5.4: Les ‘noms et adjectifs en -al/-aux’

variés. On trouve un mécanisme synthétique sur les déterminants, dont le mélange évoque par exemple les suffixes de cas-genre-nombre du latin ou du grec ancien. Ici, on modélise le nombre par un trait à deux valeurs équivalentes : SG et PL.

Et on trouve un autre mécanisme, cette fois un suffixe de pluriel, qui s'oppose à sa propre absence. On a alors un cas de trait binaire \pm PL. On pense par exemple au suffixe de pluriel de l'anglais³. Cependant, ce suffixe, décrit dans les termes du phénomène de liaison, n'en est pas homogène pour autant. On trouve des cas où il est une consonne de liaison obligatoire, et d'autres où c'est une liaison facultative. Le cas des adjectifs pré-nominaux du FD est encore différent, puisqu'il innove dans le domaine de la liaison en laissant jouer un rôle au contexte gauche.

Enfin, un mécanisme qui n'a pas été soumis en détail à la modélisation concerne des terminaisons lexico-grammaticales. Autant le *o* de *fvo* semble là seulement pour la marque grammaticale du pluriel, autant le *al* de *fval*, que le *o* remplace, semble motivé également lexicalement, comme le suggère la dérivation adjectivale *fvalresk*.

Bien sûr des faits idiosyncratiques sont à signaler, mais comme l'annonce leur dénomination, on s'arrêtera à leur simple citation.

5.4.2 Variation

La morphologie du nombre montre une légère variation, autour de la productivité des marques de nombre sur les noms et les adjectifs post-nominaux. On aurait pu décrire cette variation comme étant phonologique, autour de la notion de liaison facultative. Mais je préfère évidemment y voir une distinction entre le FD et le FCT. Seul le second a une morphologie de nombre sur le nom, sous la forme très légère d'une

3. Quelques détails à ce sujet au chapitre 6.

5.4 Variété et variation dans la morphologie du nombre

consonne de liaison facultative.

5.4.3 Conclusion

Ce chapitre m'amène à proposer le résumé en (16) et (17) des morphologies du nombre dans le domaine nominal du français.

(16) FCT :	singulier	Dét _{SG} Dét-∅	Adj-∅	Nom-∅	Adj-∅
	pluriel	Dét _{PL} (z) Dét-(z)	Adj-(z)	Nom-{z}	Adj-{z}

(17) FD :	singulier	Dét _{SG} Dét-∅	Adj-∅	Nom	Adj
	pluriel	Dét _{PL} (z) Dét-(z)	Adj-<z>		

Est-ce que ces observations, ainsi formulées, ont une portée dépassant la simple morphologie ? Le chapitre suivant s'attachera à le montrer, en cherchant à corrélérer typologiquement et dans la syntaxe et la sémantique ces faits de morphologie à travers un petit modèle des noms nus. On s'y demandera également si les différences entre le FCT et le FD ont une corrélation au sein de ce modèle.

Chapitre 6

Les statuts morphosyntaxiques du nombre

DANS ce chapitre, j'aborde la question des statuts morphosyntaxiques du nombre. L'idée est d'élargir les considérations sur la morphologie du chapitre 5 aux autres domaines de la grammaire liés au nombre, en particulier la syntaxe et l'interface syntaxe-sémantique, et en donnant à la question une dimension comparative centrale. Je souhaite ainsi montrer, après d'autres, que l'enjeu est plus grand que la simple question « Est-ce que les noms du français ont toujours un S au pluriel ou pas ? »

La première section présentera le modèle morphologique du nombre des langues romanes de Barra Jover (2008) que j'appliquerai à quatre langues : l'anglais, l'espagnol, le poyaudin et le basque.

La deuxième section se consacrera à la description d'un modèle corrélant la morphologie du nombre à la syntaxe et à la sémantique à travers la question des noms nus. Il permettra de proposer une petite typologie des statuts morphosyntaxiques du nombre dans les quatre langues considérées.

La troisième section partira à la recherche du statut morphosyntaxique

du nombre du FD et du FCT à la lumière de ce modèle, qui servira de point de contrôle. Cette section remettra au jour la question de la présence de noms sans déterminant coordonnés en FCT, sans toutefois avancer de réponse claire. Et elle sera l'occasion d'une remarque sur le statut possiblement métalinguistique du S de pluriel de l'orthographe grammaticale.

Enfin, la quatrième section synthétisera ce chapitre ainsi que cette deuxième partie, notamment en remarquant qu'elle n'aboutit pas à exiger le besoin de décrire deux grammaires du nombre, mais qu'elle a peut-être permis d'accorder un statut serein et motivé au nombre du français.

6.1 Un modèle morphologique : Barra Jover (2008)

Dans Barra Jover (2008), on trouve un ensemble de propositions issues du cadre théorique de la morphologie distribuée pour représenter le nombre du domaine nominal en français et dans un choix de langues romanes, en particulier en ce qui concerne les variations morphologiques observables. J'utiliserai son modèle pour décrire des morphologies plus variées encore que celles qu'il décrit lui-même. Mon idée est de faire complètement le tour, en dehors du français, des manifestations morphologiques de nombre que l'on trouve en français et qui ont été répertoriées au chapitre 5. Ensuite, je présente l'application de ce modèle au français, ce que fait déjà Barra Jover (2008), mais en insistant sur un point différentiel : tandis que son français oral informel et mon français démotique sont à mes yeux équivalents, son français écrit n'est pas mon français classique tardif. La dernière section sera l'occasion de revenir sur ce point.

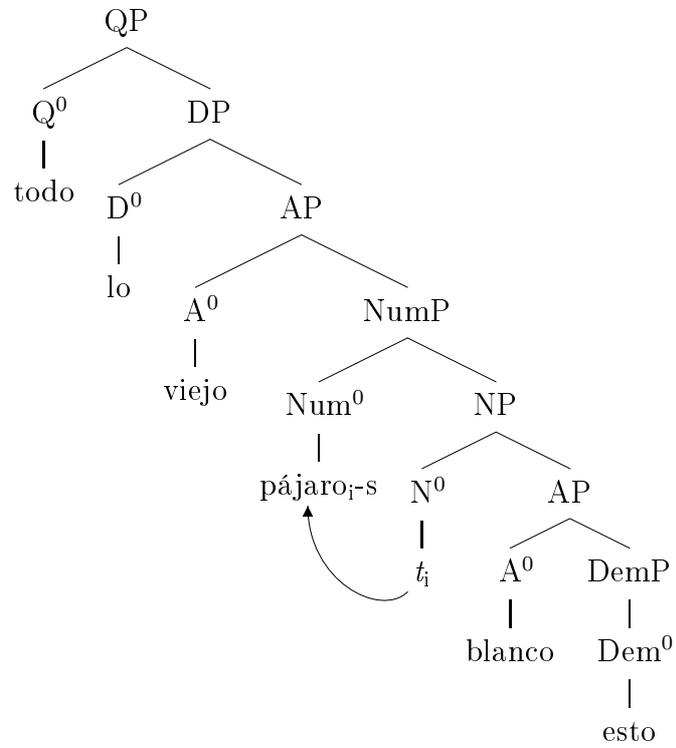
6.1.1 Le modèle espagnol

La morphologie du nombre de l'espagnol, que Barra Jover (2008:15) considère comme « la structure romane commune », sera ici résumée ainsi : chaque élément du syntagme nominal contient un suffixe de nombre, \emptyset au singulier, *-s* au pluriel :

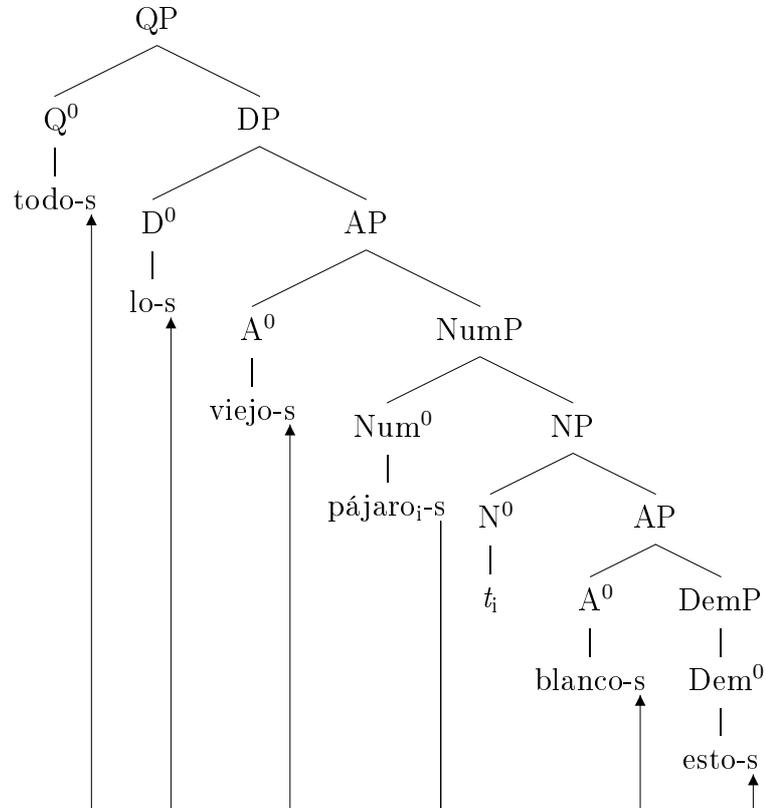
- (18) a. *toda- \emptyset la- \emptyset vieja- \emptyset ropa- \emptyset blanca- \emptyset tuya- \emptyset*
toute--PL la--PL vieille--PL robe--PL blanche--PL tienne--PL
'toute ta vieille robe blanche'
- b. *todo-s lo-s viejo-s pájaro-s blanco-s esto-s*
tout-+PL le-+PL vieux-+PL oiseau-+PL blanc-+PL DÉM-+PL
'tous ces vieux oiseaux blancs'

On dira alors dans les termes de Barra Jover (2008) que ce suffixe de nombre est engendré dans une tête morphosyntaxique autonome où le nom vient se fléchir (étape (19)), et se répend dans le DP par le mécanisme de l'accord (étape (20)).

(19)



(20)

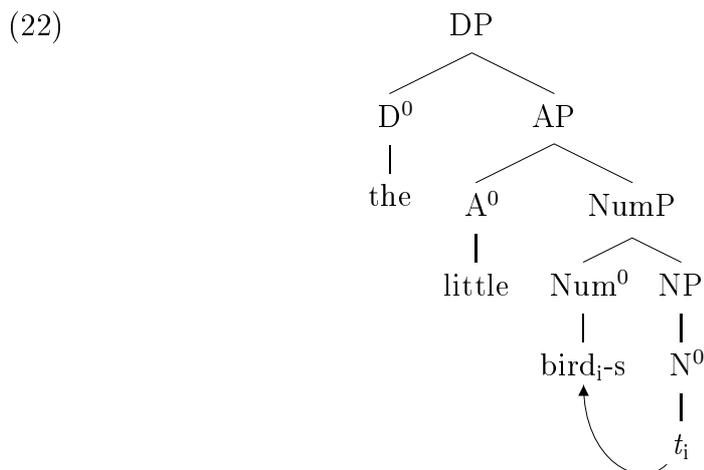


6.1.2 Le modèle anglais

La morphologie du nombre anglais est limitée à un suffixe de pluriel sur le nom. Pour la simplicité de l'exposé, on négligera ici l'exception des déterminants démonstratifs (*this/these* et *that/those*) et on retiendra l'exclusivité des noms sur le suffixe de pluriel (cf. Bouchard, 2003:56). Ce cas est modélisé en utilisant une projection fonctionnelle de nombre dans laquelle vient se fléchir la tête lexicale N^0 , selon un schéma similaire à la montée du verbe dans la ou les têtes fonctionnelles générées sous IP. Ceci correspond à la première étape de la flexion de nombre de l'espagnol, et il n'y a pour l'instant pas de justification à ce que l'anglais ne connaisse

pas la seconde étape, celle de l'accord. Je reviendrai sur ce point à la section 6.2.3.

(21) *the little bird* -Ø/s
-PL/+PL



6.1.3 Le modèle poyaudin et le modèle basque

Dans le DP poyaudin, seul le déterminant porte une morphologie de nombre, qui est similaire à celle observée en français (voir tableau 6.1). Aucun élément ne rappelle, de près ou de loin, le suffixe de pluriel ouest-roman *-s*, pas même sous une quelconque forme de consonne de liaison.

(23) a. *mon/me āsiē ami almā*
'mon/mes ancien(s) amis(s) allemand(s)'

C'est-à-dire, donc, que le nombre est un trait dépendant du déterminant, c'est-à-dire un des traits du faisceau de traits généré par la tête D^0 .

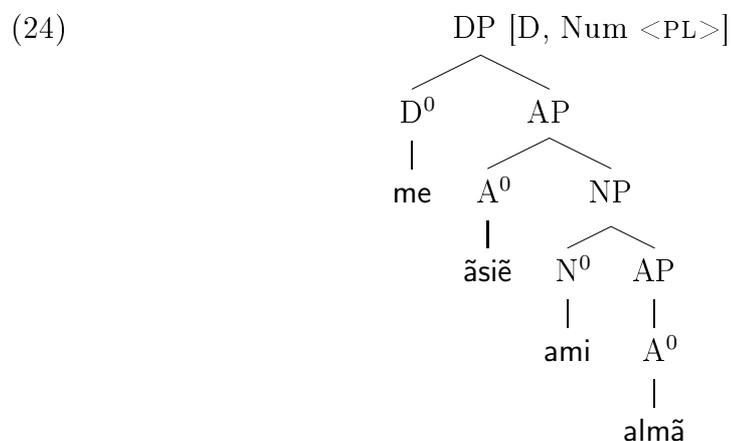
6.1 Un modèle morphologique : Barra Jover (2008)

	MASC SG	FEM SG	PL
article défini	l	la (l)	le
article indéfini	ẽ (en)	yn	de
article partitif	dy (d l)	d la (d l)	
adjectif démonstratif	s (st)	st	sø
adjectif possessif	mõ (mon)	ma (mon)	me
	tõ (ton)	ta (ton)	te
	sõ (son)	sa (son)	se
	not		no
	vot		vo
		iø	

entre parenthèses, les allomorphes apparaissant devant voyelle

TABLEAU 6.1: Quelques déterminants du poyaudin

Ceci donne la représentation en (24), sans tête fonctionnelle pour le nombre, et avec un nom réalisé in situ.



En basque, d'après l'étude d'Artiagoitia (2002), similairement à ce qu'on trouve en poyaudin, le nombre n'apparaît morphologiquement que sur l'article, lequel est un suffixe généré à droite de l'élément le plus

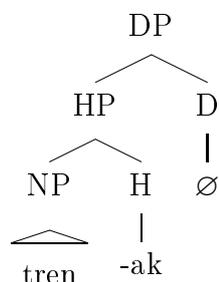
à droite du groupe nominal, quel que soit cet élément, comme esquissé en (25). Comme le basque est une langue à tête finale, cet article est bien l'élément placé le plus haut dans le DP. Je reproduis ci-dessous les exemples pertinents d'Artiagoitia (2002), avec ses gloses anglaises, ainsi que ses représentations arborescentes. On reviendra sur la distinction qu'il fait entre les lectures existentielles et les descriptions spécifiques dans la section 6.2.3.4.

(25) basque (d'après Artiagoitia, 2002:76-77) :

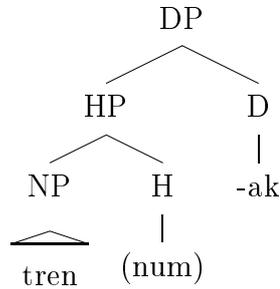
- a. *tren-a* – *tren-ak*
 train-ART – train-ART.PL
 '(the) train' – '(the) trains'
- b. *tren urdin-a* – *tren urdin-ak*
 train blue-ART – train blue-ART.PL
 '(the) blue train' – '(the) blue trains'

(26) basque, Artiagoitia (2002:94)

- a. lectures existentielles :



- b. descriptions spécifiques :



6.2 L'interface syntaxe-sémantique du nombre

6.2.1 L'actualisation

Comme point de départ de mon étude de l'interface syntaxe-sémantique autour du nombre, on observera l'opposition basique, dans les langues avec articles, entre les interprétations des noms sans et avec article. Ainsi, en (27a), le nom sans article *plakar* n'implique pas l'existence d'un quelconque placard, alors que (27b) et (27c) sont redevables d'une telle existence, quel que soit la qualité de l'identification du placard en question.

- (27) a. la port d plakar
 b. la port dy plakar
 c. la port d ě plakar

À titre de comparaison, on peut remarquer en (28) que cette opposition se réalise un peu différemment en allemand, puisque le nom sans article *Schrank* est incorporé au nom tête pour former un nom composé, alors qu'il est réalisé comme complément au génitif (allemand standard) ou

Chapitre 6 Les statuts morphosyntaxiques du nombre

complément prépositionnel (allemand familier), lorsqu'il apparaît avec un article. Les interprétations correspondent bien à celles de (27).

- (28) a. *die Schranktür*
la placard porte
- b. *die Tür des Schrank* – *die Tür vom Schrank*
la porte le.GÉN placard.GÉN – la porte du.DAT placard
- c. *die Tür eines Schrank* – *die Tür von einem Schrank*
la porte un.GÉN placard.GÉN – la porte de un.DAT placard

Guillaume (1919/1975:22ss.) décrit cette opposition en termes de valeur en langue et valeur en discours du substantif. Pour lui, l'emploi d'un substantif introduit sa valeur en langue (les propriétés définitives des objets du monde potentiellement identifiables par ce substantif) dans l'énoncé, et l'emploi d'un article donne à ce substantif une valeur en discours, c'est-à-dire que l'énoncé fait alors référence à des objets du monde du discours identifiés par les propriétés du substantif et selon un mode de calcul de l'identification porté par l'article lui-même.

Bouchard (2003:56) décrit cette opposition en termes d'atomisation : selon lui, l'opposition discutée correspond au fait de « déterminer qu'un actant est en jeu dans l'événement ».

Dobrovie-Sorin et Laca (2003:235-237) parlent de dénnotations de propriétés et de dénnotations d'individus.

Dans cette terminologie foisonnante, je retiendrai ici le terme d'*actualisation* pour décrire le passage d'un *sens prédicatif* à un *sens actualisé* d'un substantif (tel le passage de (27a) à (27bc)). Parmi tous les sens actualisés possibles, je me restreindrai à la distinction entre les lectures existentielles, les lectures génériques, et les autres lectures (dont par exemple les descriptions spécifiques dont parle Artiagoitia 2002).

6.2.2 Actualiser sans déterminant : les noms nus

Une littérature abondante traite d'un phénomène présent dans certaines langues avec articles, à savoir la possibilité d'actualiser un nom sans aucun déterminant. Je retiens quatre références pour présenter le débat : Artiagoitia (2002), Dobrovie-Sorin et Laca (2003), Bouchard (2003), Roodenburg (2004).

On y apprend que les noms de masse et les noms au pluriel peuvent, sans déterminant, recevoir une interprétation actualisée en position d'argument : ce qu'on appelle alors des *noms nus* (NN). En (29), le nom pluriel *horses* est actualisé dans une lecture générique, et en (30a), *niños* reçoit une lecture existentielle. On fait également souvent remarquer que la lecture existentielle est parfois contrainte syntaxiquement : (30b) montre que la position de sujet préverbal n'est pas accessible aux noms nus existentiels de l'espagnol.

(29) anglais : (Roodenburg, 2004:7)
Horses are extinct.

(30) espagnol : (Dobrovie-Sorin et Laca, 2003:236)
 a. *En la calle jugaban niño-s*
 dans la rue jouaient enfant-+PL
 'Des enfants jouaient dans la rue.'
 b. **Niños jugaban en la calle.*

Puis, on observe que la présence de noms nus génériques (ci-après NNG) dans une langue implique la présence de noms nus existentiels (ci-après NNE), mais pas le contraire (Dobrovie-Sorin et Laca, 2003:246), ce qu'illustre le fait que l'anglais a des NNE comme en (31), mais l'espagnol pas de NNG, cf. (32).

Chapitre 6 Les statuts morphosyntaxiques du nombre

(31) *John is eating apples.* (Dobrovie-Sorin et Laca, 2003:245)

(32) *Juan destesta *(los) gatos.* (Dobrovie-Sorin et Laca, 2003:246)
Juan déteste les chats

Et pour compléter cette implication, des langues sans NNE n'ont pas non plus de NNG, cas souvent illustré par le français, mais que, pour garder encore un peu le suspense pour ce qui est du français, j'illustrerai avec le poyaudin et le basque. En poyaudin, aucune lecture existentielle ou générique n'est possible sans article :

(33) poyaudin :

- a. **(le) tomat a-mørisõ*
les tomates elles-mûrissent
lecture générique (et lecture 'définie')
- b. **(de) tomat a-mørisõ*
des tomates elles-mûrissent
lecture générique seulement
- c. *i-a *(de) tomat k a-mørisõ*
il y a des tomates que elles-mûrissent
lecture existentielle seulement

De son côté, Artiagoitia (2002) montre que le basque n'a pas de NNE, et on peut supposer qu'il n'a pas non plus de NNG, même si Artiagoitia n'illustre pas les lectures génériques, mais seulement ce qu'il appelle des descriptions spécifiques : en (34), on voit que les lectures existentielles et les descriptions spécifiques ne peuvent pas se passer de l'article.

(34) basque (d'après Artiagoitia, 2002:77) :

- a. *Tren-*(ak) heldu dira.*
train-ART arrive AUX
'(The) trains arrived.'

- b. *Tren-*(ak) ikusi ditut.*
train-ART see AUX
'I saw (the) trains.'
- c. *Ardo-*(a) edan dut.*
wine-ART drink AUX
'I drank (the) wine.'

6.2.3 Corréler la morphosyntaxe du nombre et l'actualisation par les NN

Chacun de leur côté, l'anglais, l'espagnol, et le poyaudin et le basque ont des morphologies du nombre distinctes et des propriétés de leurs NN également. Ceci amène à proposer l'hypothèse suivante :

- (35) **hypothèse de la corrélation morphologie-syntaxe-sémantique du nombre** : Les interprétations existentielles et génériques (des NNE et des NNG) sont prévisibles sur la base de la syntaxe du nombre, et cette syntaxe est elle-même prévisible à partir de la morphologie du nombre.

Pour comprendre les implications de cette corrélation, je détaille le propos de Dobrovie-Sorin et Laca (2003).

6.2.3.1 Dobrovie-Sorin et Laca (2003) sur les NNE

Ces auteurs décrivent les NNE comme dénotant des propriétés (comme tous les noms sans déterminant). Or, les NN, donc a fortiori les NNE, sont restreints aux noms de masse et aux pluriels. C'est-à-dire, comme le font remarquer ces auteurs, que les NNE dénotent plus précisément les propriétés des membres d'un domaine de quantification cumulative. Ce qui définit un domaine de quantification cumulative, c'est la propriété de

Chapitre 6 Les statuts morphosyntaxiques du nombre

ses membres de pouvoir s'associer indéfiniment sans sortir du domaine, ce qui est illustré pour l'anglais par (36), où (36ab) contrastent avec (36c).

- (36) a. $rice(x) \& rice(y) \rightarrow rice(x + y)$
b. $tables(x) \& tables(y) \rightarrow tables(x + y)$
c. $table(x) \& table(y) \rightarrow \neg table(x + y)$

La lecture existentielle des NNE, toujours selon Dobrovie-Sorin et Laca (2003), est alors obtenue lorsque qu'un NNE est l'argument de ce qu'elles appellent un prédicat existentiel. En effet, elles remarquent que les NNE se sont pas possibles avec tous les types de prédicats. Pour l'espagnol, elles proposent les contrastes entre (37) et (38) (Dobrovie-Sorin et Laca, 2003:256 et 258).

- (37) a. *Estaban presentes médicos.*
étaient présents médecins
'Des médecins étaient présents.'
b. *Este manuscrito contiene errores.*
ce manuscrit contient erreurs
'Ce manuscrit contient des erreurs.'
- (38) a. **María adora café.*
Marie adore café
b. **Juan detesta manzanas.*
Juan déteste pommes

Elles fournissent de plus un critère discriminant ces deux ensembles de prédicats : les premiers peuvent localiser spatialement leur argument, les seconds ne le peuvent pas, ce que révèle le test de (39) (Dobrovie-Sorin et Laca, 2003:257).

- (39) a. *¿Dónde estaba presente Juan?*
où était présent Juan

6.2 L'interface syntaxe-sémantique du nombre

‘Où est-ce que Juan était présent ?’

- b. ??¿ *Dónde estaba triste / inquieto / enfermo Juan ?*
où était triste / inquiet / malade Juan

6.2.3.2 Dobrovie-Sorin et Laca (2003) sur les NNG

Dans des langues comme l'anglais, les NN peuvent recevoir une lecture générique. Comme les NN dénotent à la base, selon Dobrovie-Sorin et Laca (2003), des propriétés, ces auteurs proposent de décrire les NNG comme un changement de type sémantique, qui fait passer les NN de la dénotation de propriétés à la dénotation d'entités maximales, c'est-à-dire la dénotation d'espèces.

Ce changement de type, possible en anglais, ne l'est pas en espagnol, comme le montre le contraste entre (40) et (41) (tirés de Dobrovie-Sorin et Laca, 2003:262).

- (40) a. *Cats are on the verge of extinction.*
b. *Cats are intelligent.*
c. *John hates cats.*
- (41) a. **Gatos se están extinguiendo.*
chats se sont éteignant
b. **Gatos son inteligentes.*
chats sont intelligents
c. **Juan odia gatos.*
Juan déteste chats

Une dernière remarque que les auteurs font est que ce changement de type est similaire à celui qui a lieu lorsqu'un nom apparaît avec un article défini, ce que montrent les exemples corrigés de l'espagnol en (42), où la lecture générique est permise par l'article *los*.

- (42) a. *Los gatos se están extinguiendo.*
‘Les chats sont en voie d’extinction.’
b. *Los gatos son inteligentes.*
‘Les chats sont intelligents.’
c. *Juan odia los gatos.*
‘Juan déteste les chats’

6.2.3.3 Corréler la sémantique à la syntaxe

Roodenburg (2004:19) annonce le programme de l’interface syntaxe–sémantique : « c’est la structure syntaxique qui détermine en grande partie la sémantique [...] la syntaxe ne détermine pas directement les propriétés sémantiques mais crée les conditions nécessaires pour qu’une certaine propriété sémantique soit disponible ».

Selon ce programme, que je vais suivre, il faut pouvoir proposer une structure syntaxique différente pour chaque propriété sémantique observée, lesquelles structures syntaxiques seront les conditions nécessaires à l’instanciation de ces propriétés sémantiques (c’est-à-dire ces différentes lectures). Les résultats que je propose sont principalement inspirés des résultats que l’on peut voir chez Roodenburg (2004), et ne s’en écartent pas fondamentalement, si ce n’est que je n’ajouterai pas quelques finesses d’analyse sémantique qu’il propose. On se réfèrera par exemple à son « échelle sémantique » (95) ainsi qu’à son schéma de la page 112 pour prendre connaissance de ces nuances supplémentaires.

La première lecture à corréler dans la syntaxe est la lecture prédicative. La littérature est assez unanime à ce sujet : cette lecture est obtenue grâce aux noms sans déterminant (tel *plakar* dans mon exemple (27a)). Et elle est instanciable dans les positions syntaxiques prédicatives, comme la position prédicative N_2 du SN de la forme N_1 de N_2 de (27a), ou la

position prédicative de *profesør* en (43).

(43) *farl e profesør*

Quelle est la représentation syntaxique associée à ces noms prédicatifs ? Là aussi, je suivrai l'idée répandue qu'il s'agit de syntagmes nominaux sans projection fonctionnelle (des NP).

La seconde lecture à corrélérer est la lecture existentielle des NNE. On a vu que cette lecture était réservée aux noms pouvant dénoter des propriétés d'un domaine de quantification cumulative, laquelle dénotation est réservée aux noms de masse et aux pluriels. Il semble que cette dénotation soit permise, dans le cas des pluriels, par la réalisation du nombre. Cette propriété du nombre de jouer un rôle dans la caractérisation de l'opposition massique-comptable est connue, et puisque les singuliers massiques peuvent être considérés comme réalisant le nombre singulier, on peut proposer que la structure syntaxique qui permet la dénotations des NNE est une projection fonctionnelle du nombre, c'est-à-dire un NumP. Et cette lecture est instanciable par les prédicats existentiels, indentifiés en (37) et (38).

La troisième lecture est la lecture d'espèce des NNG. Comme cela a été rapidement suggéré par les contrastes en (40) à (42), la lecture d'espèce peut être permise syntaxiquement par une structure équivalente à celle des noms accompagnés d'un article défini, c'est-à-dire aux syntagmes nominaux projetant un DP. Et la lecture d'espèce des NNG est instanciée lorsqu'ils sont associés à des prédicats permettant ou exigeant une quantification générique.

Enfin, j'évoquerai rapidement les lectures dont on n'a pas parlé jusqu'ici. Qu'il s'agisse de lectures définies, indéfinies (spécifiques ou non), etc., le détail de leurs représentations syntaxiques ne sera pas investi, et

je m'arrêterai à l'idée qu'il s'agit de structures, pour le dire naïvement, 'au moins aussi grosses qu'un DP'. Ceci sort du cadre du débat sur les NN, et pour ce qui m'intéresse, du débat sur le rôle du nombre dans ce même débat.

6.2.3.4 Corréler la syntaxe à la morphologie

Similairement au programme de l'interface syntaxe–sémantique, je propose de formuler le programme de l'interface morphologie–syntaxe en ces termes : la structure morphologique détermine en grande partie la structure syntaxique, en lui fournissant les traits formels nécessaires pour qu'une structure syntaxique donnée soit instanciable.

Mon hypothèse concernant le rôle du nombre dans le débat sur les NN est à nouveau calquée sur ce qui a été proposé dans ce débat (voir à cet effet Roodenburg (2004:56ss.) et Bouchard (2003)). Je la formulerai ainsi : les structures syntaxiques permettant l'instanciation des différentes lectures étudiées varient à travers les langues, et cette variation est directement projetable à partir de la morphologie exacte du nombre dans ces langues. Cette hypothèse doit pouvoir corréler les propriétés syntaxiques distinctes des langues en question à leurs morphologies du nombre, ce qui aurait un certain pouvoir explicatif.

Les lectures prédicatives, pour commencer, sont obtenues dans les quatre langues en question de la même façon, à l'aide d'un NP sans nombre. Ce n'est pas étonnant, puisque ces quatre langues peuvent morphologiquement générer un nom sans nombre. On trouve là les prédications secondes de l'anglais et du basque, en (44a) et (44d), et les attributifs de l'espagnol et du poyaudin, en (44b) et (44c).

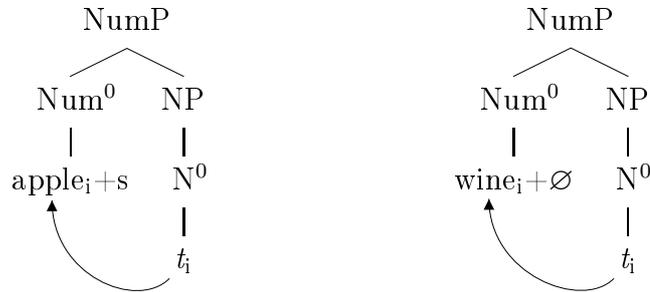
- (44) a. *Charles was elected chair of the department.*
b. *Juan es profesor.* (espagnol)

6.2 L'interface syntaxe-sémantique du nombre

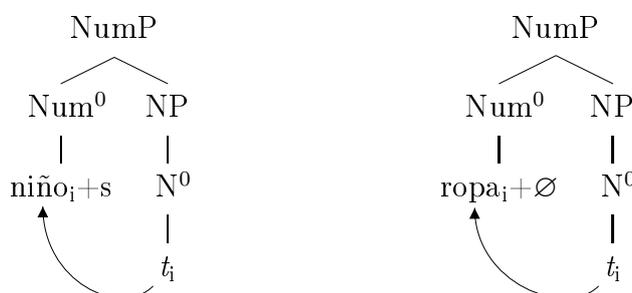
- ‘Juan est professeur.’
- c. *farl il-e profesø* (poyaudin)
 ‘Charles est professeur.’
- d. basque (Artiagoitia, 2002) :
*Landideek Ana sailburu-(*a) aukeratu dute*
 workmate.ART/ERG Ana chair.(ART) elect AUX
 ‘Workmates elected Ana chair of the department.’

Ensuite, l'espagnol et l'anglais ont des NNE, mais pas le poyaudin et le basque. Or, on a supposé que les NNE exigent une projection NumP, c'est-à-dire une réalisation minimale du nombre. Le contraste entre ces deux groupes est corrélable au fait suivant : la réalisation minimale du nombre est un nom fléchi en nombre en anglais et en espagnol, puisque leurs morphologies de nombre le permettent, d'où leurs NNE. Les arbres de (45) représentent ce cas.

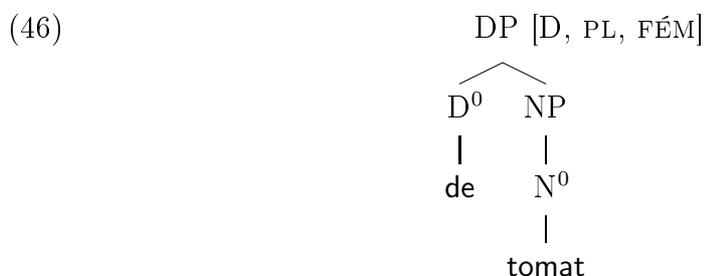
- (45) a. les NNE de l'anglais :



- b. les NNE de l'espagnol :



Au contraire, les morphologies du nombre du poyaudin et du basque ne leur permettent pas de générer un simple nom fléchi en nombre. Dans ces deux langues, il faut générer un déterminant dès que le nombre est exigible par la syntaxe, laquelle exigence lui est transmise par la sémantique. Il me semble alors opportun de proposer la représentation en (46) des lectures existentielles du poyaudin, ce qui représente déjà un syntagme nominal maximal (un DP). Et, sur le mode du « qui peut le plus peut le moins », ce DP crée les conditions suffisantes à la dénotation de propriétés d'un domaine de quantification cumulatif, même si cette dénotation ne va pas jusqu'à exiger un DP.



Pour ce qui est du basque, on a vu qu'Artiagoitia (2002) proposait des structures syntaxiques distinctes pour les lectures existentielles et les descriptions spécifiques (que j'ai rapidement proposé d'assimiler aux structures permettant la lecture d'espèce). Si le principe du « qui peut le plus peut le moins » est intéressant, alors il invite à ne pas générer

6.2 L'interface syntaxe-sémantique du nombre

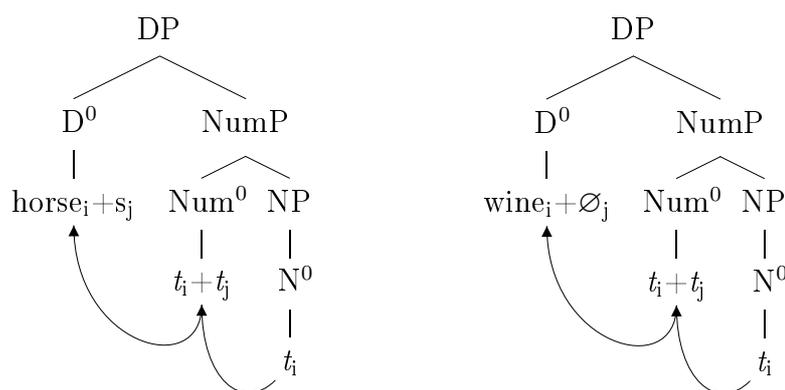
de projection intermédiaire entre NP et DP (le HP d'Artiagoitia des arbres sous (26)). C'est pourquoi je préférerai proposer pour le basque l'équivalent du *poyaudin* :



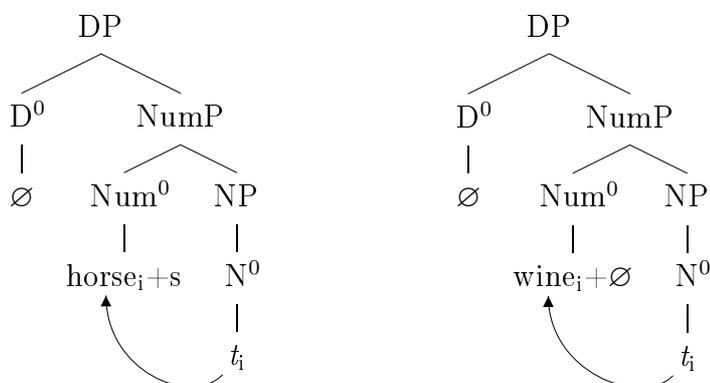
À l'étape suivante, celle des NNG, on ne trouve plus que l'anglais. Les NNG exigent une projection DP. Comment est-ce qu'elle est représentable sans déterminant ? Deux propositions circulent dans la littérature : soit on a un déterminant vide, soit le nom fléchi en nombre peut monter dans la tête D^0 et satisfaire l'exigence de déterminant. Comme ces deux solutions me semblent aussi formelles l'une que l'autre, je ne rentrerai pas dans le débat, suivant ainsi Dobrovie-Sorin et Laca (2003:265), et je représenterai les deux solutions :

(48) les NNG de l'anglais :

a. par montée du nom en D^0 :



b. avec un déterminant vide :

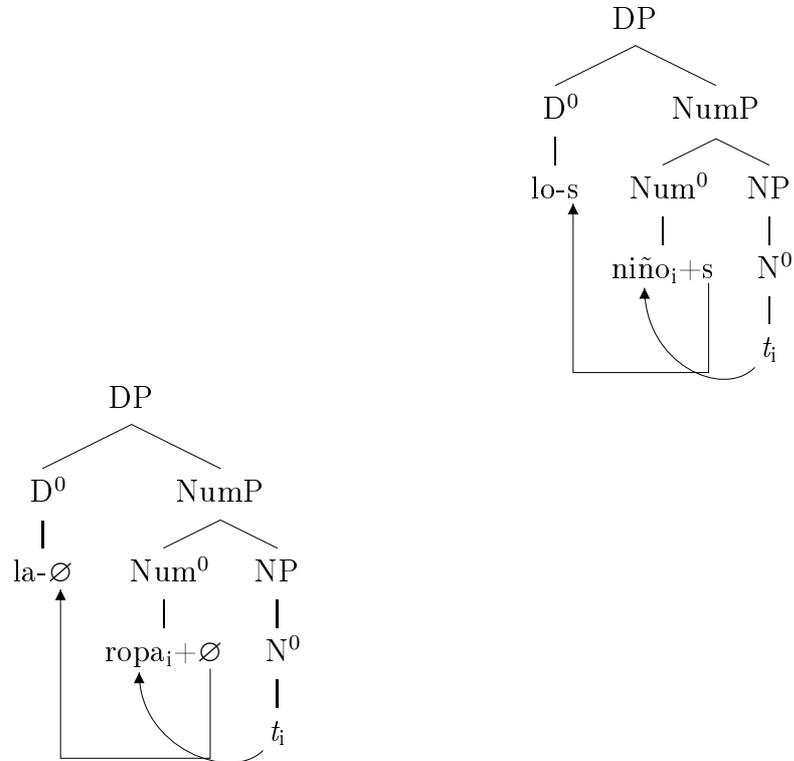


La morphologie du nombre de l'anglais permet cette projection DP sans générer de déterminant, alors que celle de l'espagnol ne le permet pas. Qu'est-ce qui, dans leurs morphologies respectives, justifie cette différence ? La différence morphologique est que le trait de nombre de l'espagnol se répartit à tous les membres du syntagme nominal, alors qu'il se cantonne au nom en anglais. Jusqu'ici, je n'ai pas rencontré dans la littérature un cadre conceptuel qui permette d'envisager cette incompatibilité entre le phénomène d'accord et la possibilité de montée du nom dans D^0 (ou la disponibilité d'un déterminant vide). Il n'en reste pas moins vrai que cette différence syntaxique est corrélée morphologiquement.

Comment est-ce que les trois autres langues fournissent la structure syntaxique exigible pour la dénotation d'espèce nécessaire à la lecture générique ? Pour le poyaudin et le basque, le DP a déjà été généré pour les lectures existentielles, donc on peut conserver les mêmes représentations qu'en (46) et (47). Pour l'espagnol, il faut ajouter à (45) une projection DP contenant l'article défini, qui reçoit son trait de nombre par accord (flèche rectiligne) :

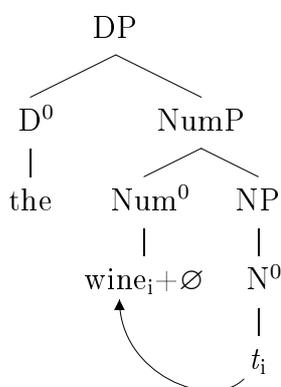
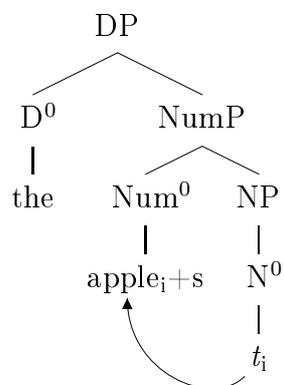
6.2 L'interface syntaxe-sémantique du nombre

(49)



Enfin, il reste à présenter les structures syntaxiques qui permettent les autres dénominations. On supposera simplement qu'elles exigent toutes au moins un DP, dont la génération a déjà été présentée pour le poyaudin, le basque et l'espagnol. Sans surprise, les descriptions définies (par exemple) de l'anglais sont générées ainsi dans le modèle :

(50)



6.2.3.5 Synthèse des statuts morphosyntaxiques du nombre

Pour conclure cette section, je voudrais synthétiser les résultats en remarquant qu'ils encouragent l'hypothèse présentée à l'encadré (35), et qu'ils permettront de s'interroger, dans la section suivante, sur la place du français dans cette petite typologie des NN.

Tout d'abord, on est bien parvenu à proposer une représentation syntaxique pour chacune des interprétations sémantiques étudiées : un NP pour les sens prédicatifs, un NumP pour les lectures existentielles, et un DP pour les lectures génériques.

Ensuite, on a corrélé la possibilité ou l'impossibilité, pour chacune des langues étudiées, de produire ces structures en se passant de détermi-

6.2 L'interface syntaxe-sémantique du nombre

nant à la morphologie de nombre qui est en œuvre dans chacune de ces langues : pas de NN possible en poyaudin et en basque, langues où le nombre n'apparaît que sur le déterminant, des NNE mais pas de NNG en espagnol, où le nombre est présent sur chaque élément du syntagme nominal par un phénomène d'accord, et enfin des NNE et des NNG en anglais, où le nombre n'apparaît que sur le nom.

On peut maintenant répondre à la question du titre du chapitre : « quels sont les statuts morphosyntaxiques du nombre ? » Autrement posée : « quelle est la force syntaxique et sémantique du nombre ? » Dans ce qu'on a pu observer, on peut proposer trois statuts morphosyntaxiques possibles. Le nombre du poyaudin et du basque, pour commencer, n'a pas de statut morphosyntaxique par lui-même. Morphologiquement, il est indissociable du déterminant, et n'a aucune autonomie. Syntaxiquement, il ne peut donc satisfaire seul un NumP. Et sémantiquement, il n'est alors pas capable d'être associé seul à une dénotation quelconque, en particulier dans le cas des lectures existentielles et génériques.

Le second statut morphosyntaxique est celui du nombre de l'espagnol. Dans cette langue, le nombre a une petite autonomie morphologique : il est identifiable (un suffixe \pm PL) et se répend par accord sur tout le syntagme nominal. Du coup, syntaxiquement, il satisfait à l'établissement d'un NumP, mais pas encore d'un DP (il atteint la position D^0 par accord, donc seulement si cette position est occupée par un déterminant à même d'accueillir cet accord). Et sémantiquement, il peut donc offrir les conditions nécessaires à la lecture existentielle, mais pas à la lecture générique. Dans cette langue, le nombre a un petit statut morphosyntaxique.

Enfin, le statut morphosyntaxique du nombre anglais. Morphologiquement, il est identifiable comme suffixe \pm PL, comme en espagnol, à ceci près qu'il n'a pas à se répendre sur tout le syntagme nominal pour que celui-ci soit marqué en nombre dans son ensemble. On peut proposer que

c'est cette propriété morphologique qui lui permet de satisfaire syntaxiquement non seulement un NumP, mais aussi un DP. Et ainsi, il permet à lui seul à la fois la lecture existentielle et la lecture générique. C'est donc un statut morphosyntaxique plus grand que celui du nombre espagnol.

En anglais comme dans les trois autres langues étudiées, le statut morphosyntaxique du nombre s'arrête là. Il ne peut pas satisfaire seul les conditions nécessaires à des lectures associées à des DP plus complexes, c'est-à-dire des lectures associées à des déterminants plus particuliers (les lectures définies, déictiques, ...).

Le tableau 6.2 propose une visualisation de ces résultats.

dénotation	structure syntaxique minimale exigible	morphologie de nombre nécessaire à une réalisation sans déterminant	langues le permettant
propriétés	NP	aucune	toutes
propriétés des membres d'un domaine de quantification cumulative	NumP	fléchir le nom en nombre	anglais espagnol
dénotation d'espèce	DP	exclusivité du nom pour porter une marque de nombre	anglais
autres dénominations	DP	impossible	aucune

TABLEAU 6.2: L'interface morphologie-syntaxe-sémantique et le nombre

6.3 Enfin, le français

« Un morphème inaudible dans la parole n'en reste pas moins un morphème existant en pensée : il suffit que sa réalisation soit possible, si rare soit-elle effectivement. [...] D'autre part, la forme écrite du langage a pris une telle importance dans une langue de civilisation comme le français, et dans la pensée d'un peuple à qui la graphie de la langue a été enseignée avant toute autre chose, que les morphèmes graphiques du pluriel, *-s*, *-x*, *-nt* s'intègrent plus ou moins, selon les individus, à l'image mentale des mots, et sont donc psychiquement présents même s'ils ne correspondent à rien de réalisé dans la parole. »

Gérard Moignet, cité par Eschmann (1976:7)

« Aussi croyons-nous que, puisque les noms français n'ont pas tenu en général à garder leur *-s* pour marquer le pluriel, les liaisons ont de leur côté le rôle d'empêcher l'hiatus dont le français a horreur – plutôt que de noter ledit pluriel. »

Eugène Tanase, cité par Eschmann (1976:11)

Dans cette section, je propose de reprendre la représentation du DP du français oral informel de Barra Jover (2008), en supposant qu'il puisse s'agir de ce que j'identifie comme le FD. Et puisque la section précédente a permis d'établir un petit modèle des statuts morphosyntaxiques du nombre dans des langues dont les réalisations morphologiques de nombre rappellent de plus ou moins près celles décrites pour le français au chapitre 5, je me référerai à ce modèle pour contrôler la conformité du statut morphosyntaxique accordé au nombre en français.

Après avoir établi une représentation du DP du FD, et avoir donné un statut morphosyntaxique au nombre de ce FD, grâce au modèle proposé, je poserai la question du statut morphosyntaxique du nombre du FCT au regard de ses différences morphologiques d'avec le nombre du FD. J'essaierai en particulier de comprendre quel rôle les noms sans déterminant coordonnés du FCT, que je présenterai, peuvent jouer pour établir ce statut morphosyntaxique au regard de la présence possible de marques de nombre sur les noms et les adjectifs post-nominaux du FCT.

Enfin, comme je montrerai que la tentation de proposer un statut morphosyntaxique distinct pour le nombre du FCT conduit à des considérations 'tirées par les cheveux' et pose des questions auxquelles je n'apporterai pas de réponse claire, j'envisagerai comment on peut conceptualiser la flexion possible du nom en nombre à travers la notion de connaissance métalinguistique. Ceci m'amènera enfin à discuter la 'naturalité' de ce qu'on aime à nommer 'le français écrit' et à conceptualiser comme une langue autonome avec une grammaire à décrire à partir de la réalité graphique.

6.3.1 Le statut du nombre du FD

D'après les faits morphologiques tels qu'observés au chapitre 5, on va pouvoir examiner la proposition de représentation du DP du français oral informel de Barra Jover (2008). On verra que cette représentation rend adéquatement compte de la morphologie du nombre du FD, et également qu'elle est en adéquation avec le statut morphosyntaxique du nombre que l'on souhaitera proposer au regard du modèle établi et du comportement du FD quant aux NN.

6.3.1.1 Représenter le DP du FD

Je suis ici presque à la lettre les propositions de Barra Jover (2008:16ss.), concernant à la fois la structure du DP et ses justifications sur la base de la morphologie.

En FD, il faut modéliser le nombre synthétiquement présent sur les déterminants, et le suffixe de nombre qui touche toute la partie pré-nominale du DP. Pour cela, il faut placer un trait de nombre parmi le faisceau de traits générés par la tête D^0 . Dans les exemples (51) et (52), ces autres traits sont le type de détermination (ici DÉM, pour démonstratif), et le genre (ici FÉM). Le trait de genre est dépendant morphologiquement du nombre, et les traits de genre et nombre sont dépendants de la détermination (cf. section 5.1.1), donc le faisceau de traits contiendra dans cet ordre D, Num, et Gn.

Ensuite, le suffixe de nombre présent sur les adjectifs pré-nominaux et le déterminant est généré comme suit : on suppose une tête Num^0 , projetée dans un NumP juste au-dessus du nom, et qui va monter dans chaque adjectif jusqu'à atteindre le déterminant, où elle satisfera le trait de nombre déjà présent. Pour chaque tête rencontrée pendant son déplacement de Num^0 à D^0 , la tête Num^0 dépose son épel, qui est \emptyset pour le singulier (-PL) et /**(z)**/ pour le pluriel (+PL).

Les arbres ci-dessous proposent une visualisation en trois étapes de la dérivation des formes de surface des DP avec un adjectif pré-nominal et un adjectif post-nominal. La première étape génère les projections syntaxiques, leurs traits et leurs têtes. Puis, la tête Num^0 joint l'adjectif pré-nominal et y dépose son épel. Enfin, elle rejoint la tête D^0 , y dépose son épel, et satisfait le trait de nombre de la projection DP, qui peut alors générer la bonne forme pour sa tête D^0 .

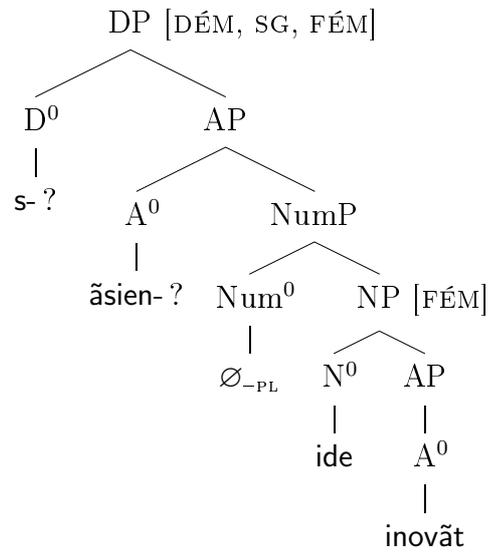
Je laisse de côté la dérivation du genre, qui n'est pas mon propos ici.

Chapitre 6 Les statuts morphosyntaxiques du nombre

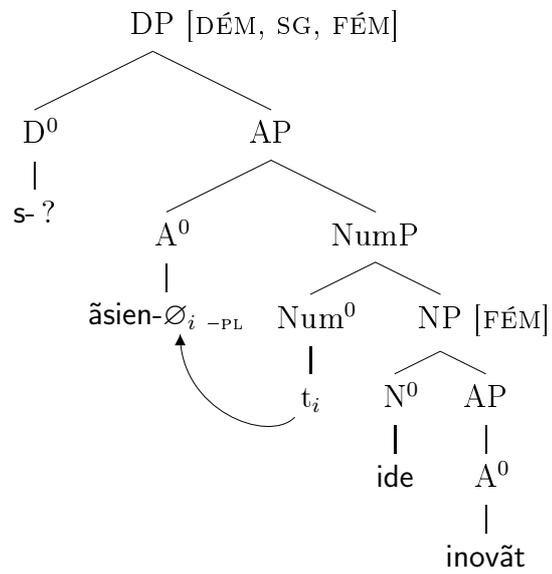
(51) est la dérivation d'un DP singulier, (52) celle d'un DP pluriel.

(51) FD : set ãsien ide inovãt

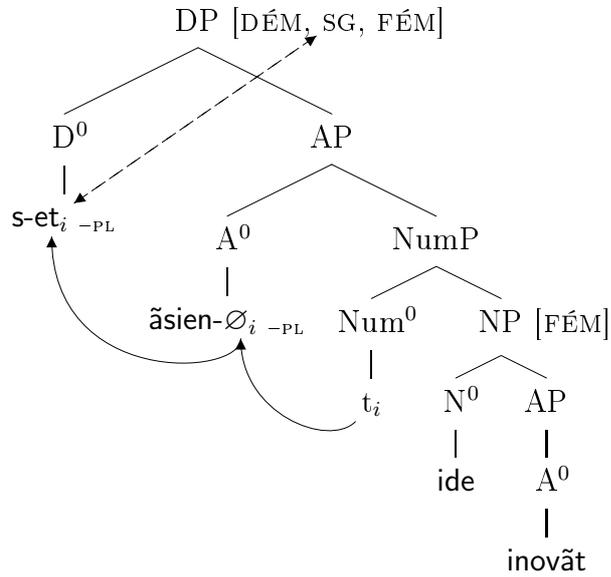
a.



b.

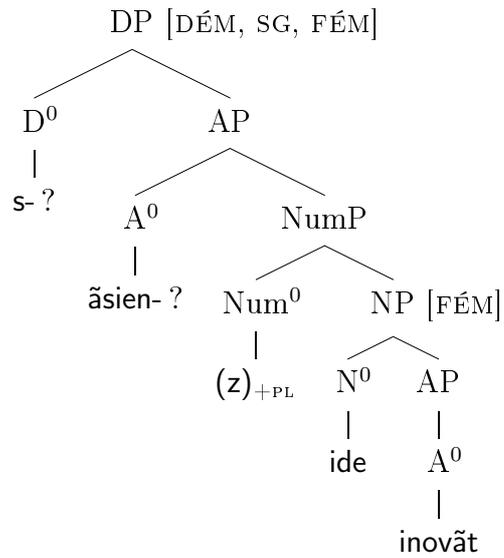


c.

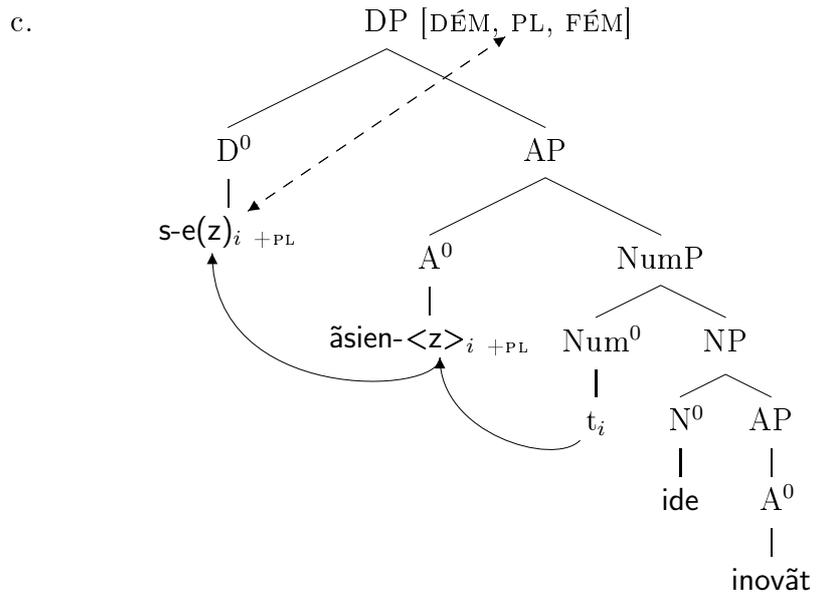
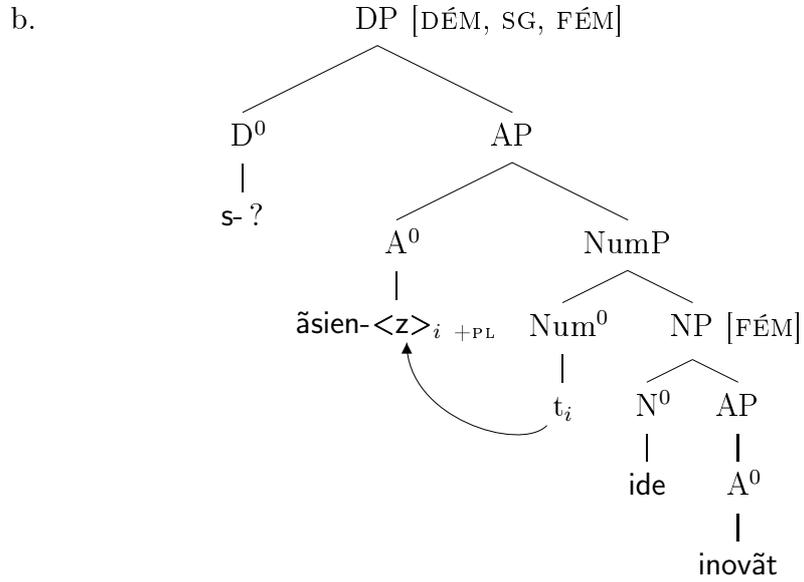


(52) FD : sez äsien(z) ide inovãt

a.



Chapitre 6 Les statuts morphosyntaxiques du nombre



6.3.1.2 Quel statut pour le nombre du FD ?

Si ces représentations se justifient sur la base des faits de morphologie du nombre, on peut se demander ce que le modèle exposé à la section précédente prédit quant au statut morphosyntaxique du nombre du FD. Pour cela, on sera amené à poser la question des NN du FD, pour vérifier la corrélation posée comme hypothèse à l'encadré (35).

Le premier point sera la question de l'autonomie du nombre sur le déterminant. Si comme en anglais le nombre du FD a une quelconque autonomie sur la présence d'un déterminant, alors le FD doit avoir des NNG. On pourrait éventuellement le penser, si on prend au sérieux l'observation que le suffixe de nombre peut apparaître sur l'adjectif pré-nominal. Un syntagme nominal sans déterminant ni adjectif pré-nominal ne peut pas réaliser morphologiquement le nombre, donc un nom seul ne peut pas servir de NNG, ce que montre (53a). Est-ce qu'un nom accompagné d'un adjectif pré-nominal le peut ? (53b) montre que ce n'est pas le cas non plus.

- (53) a. *il-em abriko
 b. *il-em ptiz abriko

Ceci montre que le nombre, bien qu'il soit généré dans une tête fonctionnelle autonome, n'a pas la possibilité de légitimer les NNG. Son statut est donc tel qu'il ne puisse pas fournir un DP. Sa présence dans D^0 est redevable d'un déterminant.

Est-ce que, pour autant, le nombre du FD peut légitimer un NumP, ce qui revient à dire que les NNE seraient possibles en FD ? (54) montre que ce n'est pas le cas, même avec un adjectif pré-nominal.

- (54) a. *abriko sō tōbe d l arbr ijer

b. *ptiz abriko sō tōbe d l arbr ijer

Ces faits ne sont pas surprenants, puisque, pour le cas de la lecture générique, le FD marque le déterminant en nombre, comme le fait l'espagnol, qui n'a pas non plus de NNG, et contrairement à l'anglais. Pour le cas de la lecture existentielle, il faut d'abord observer que la tête Num⁰ ne peut pas se réaliser sans une projection au-dessus d'elle. Ensuite, la projection de l'adjectif pré-nominal (AP), puisqu'elle est lexicale et non fonctionnelle, ne peut pas se substituer à NumP pour légitimer la lecture existentielle, laquelle exige, selon cette idée, et, il me semble, selon le programme de l'interface syntaxe-sémantique, une vraie projection fonctionnelle la légitimant, ce que fournit au contraire un NumP qui accueille la montée du nom.

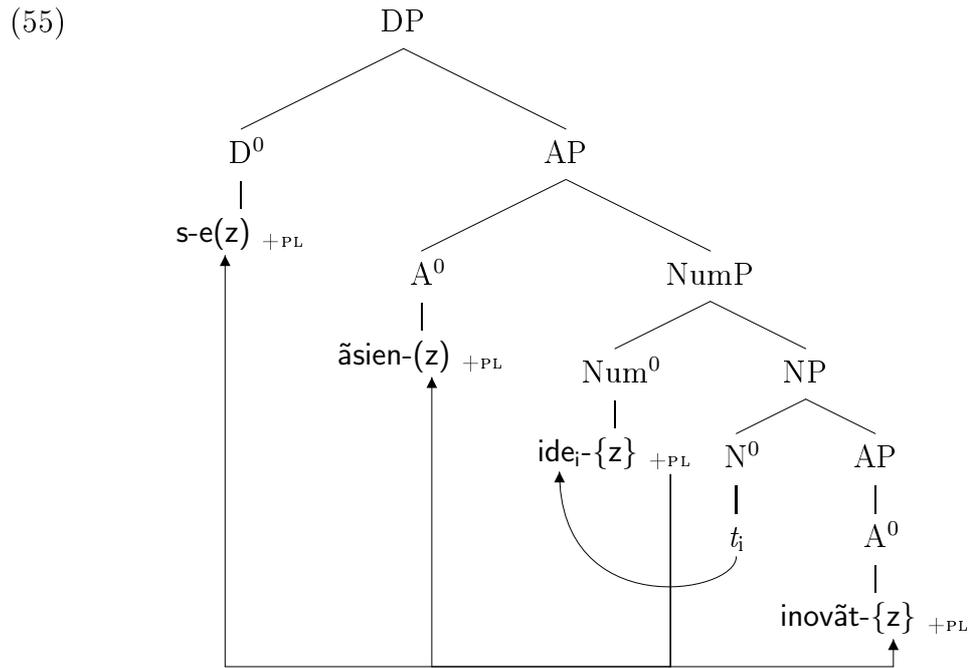
Je propose alors de conclure que le nombre du FD n'a finalement aucun statut morphosyntaxique, comme ceux du poyaudin et du basque. Le fait qu'il soit généré dans une projection propre ne lui donne pas de statut, ce que je formule ainsi : Num⁰ a pour seule raison d'être grammaticale d'aller satisfaire le trait de nombre de la tête D⁰. Sa faible force grammaticale peut aussi être corrélée au fait que son épel soit phonologiquement léger, voire ignorable sur certains adjectifs pré-nominaux.

6.3.2 Le statut du nombre du FCT

6.3.2.1 Si le nombre du FCT avait un statut

Puisque le FCT a la possibilité de marquer le nombre sur tous les éléments du syntagme nominal, on peut poser la question suivante : est-ce que sa morphologie donne à son nombre le statut du nombre espagnol ? C'est-à-dire, est-ce que le nom peut n'être pas réalisé in situ mais monter dans la tête Num⁰ ? Et est-ce que le suffixe de nombre se répend dans l'en-

semble du syntagme nominal par accord ? Si c'est le cas, alors le nombre du FCT a le statut du nombre de l'espagnol, et on donne à son DP la représentation en (55).



Cette position n'est pas tenable telle quelle puisque le FCT ne connaît ni les NNG (comme l'espagnol, dont il partage la propriété de marquer le nombre sur le déterminant), ni non plus les NNE :

- (56) a. *il-em abriko
 b. *abriko sã tõe d l arbr ijer

Pourtant, il n'en reste pas moins vrai que le FCT n'a pas exactement la même morphologie de nombre que le FD. Pour lui donner un statut malgré l'absence de NNG et de NNE, il faudrait trouver des situations où le nom sans déterminant permette une lecture actualisée. Un fait du FCT

s'en rapproche, et on va tenter de voir quel rôle il peut jouer pour accorder un statut grammatical au nombre du FCT. Ce fait en question, c'est la possibilité de produire en FCT ce qu'on appelle des noms sans déterminant coordonnés (ci-après N-et-N). (57) à (59) en présentent quelques exemples, tous tirés de Roodenburg (2004).

(57) N-et-N à lecture spécifique :

- a. [...] Jiẽ e ʃa ave tu dø l er sal
- b. zur d marʃe ã vil – kliã(z) e kuriø s-promen otur dez etalaz

(58) N-et-N à lecture générique :

- a. le ʒapone õ tuzur apresie velo(z) e skutør d tut mark
- b. [...] ministr(z) e o-fõktioner n-vwajaʒ pa(z) ã sgõd klas

(59) N-et-N à lecture existentielle :

- a. kom ʃak ane l kěz fevrie ʒã a plãte begonia(z) e ʒasõt o fõ d sõ ʒardẽ
- b. il-i-a velo(z) e skotør dã l garaz

Ma première remarque sera que ces N-et-N sont très marqués sociolinguistiquement, et qu'il est tout à fait plausible de les exclure de toute production informelle, c'est-à-dire du FD.

Devant ces exemples, je vois deux attitudes à adopter. Soit on les accepte tous et on suit Roodenburg en les considérant comme des NN parce qu'ils admettent la lecture existentielle (et d'autres propriétés sémantiques de cette lecture que Roodenburg (2004:31ss.) détaille). Soit, comme moi-même, on n'accepte pas la lecture existentielle de (59), et le débat se pose en d'autres termes. Je commencerai par explorer la première attitude.

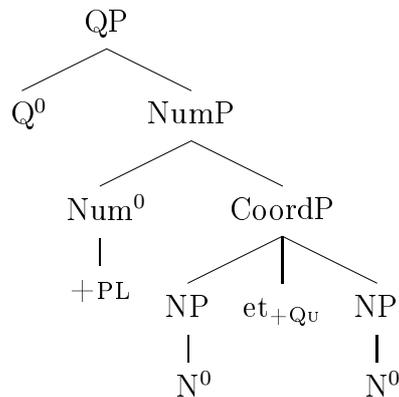
Dans le cas où il faut décrire une grammaire du FCT avec des N-et-N à lecture existentielle, Roodenburg (2004) reprend certaines propositions de la littérature sur les N-et-N et leurs rapports aux NN pour justifier leur lecture existentielle possible. Puisqu'en général les NNE sont considérés comme légitimés par une projection autonome de nombre, il faut comprendre comment les N-et-N du FCT peuvent également légitimer la lecture existentielle.

L'idée rapportée et exploitée dans Roodenburg (2004) est que la coordination, sous la forme de la conjonction *e*, est un épel possible du trait de nombre +PL. Sa représentation s'écarte de celles présentées ici, ou du moins les étend dans la mesure où Roodenburg introduit une projection QP en plus de DP, qui affine la structure du syntagme nominal au-dessus de NumP, laquelle partie du syntagme nominal ne m'a pas préoccupé, et que j'ai donc laissée sous la simple forme d'un DP sans en détailler ni le contenu syntaxique ni le contenu sémantique.

(60) Roodenburg (2004:110) :

$[_{QP} Q^0 [_{NumP} Num^0 [_{CoordP} [_{NP} N^0] et_{+Qu} [_{NP} N^0]]]]$

(61)



Chapitre 6 Les statuts morphosyntaxiques du nombre

(60) est la structure proposée par Roodenburg, à laquelle j'ajoute sa représentation arborescente en (61). Un point ne permet pas d'intégrer facilement cette représentation dans le modèle que j'ai voulu suivre. En effet, la légitimation de la tête Num⁰ par la coordination s'écarte des légitimations précédemment proposées en lien direct avec la morphologie du nombre présente sur les noms et les déterminants.

Enfin, si cette description est intéressante, il reste cependant à justifier que la coordination du FD (ou du *poaudin*) ne puisse pas servir d'épel au trait de nombre. C'est-à-dire que cette proposition garde un caractère idiosyncratique, ce qui la laisse affaiblie en l'état au sein de l'hypothèse de la corrélation morphologie–syntaxe–sémantique.

Si au contraire on trouvait une description des N-et-N existentiels qui fasse intervenir la marque de nombre sur le nom, il faudrait réussir à justifier qu'elle n'implique pas des NNE non-coordonnés en FCT. Je ne connais pas une telle description.

Comme je l'ai signalé, la lecture existentielle me semble loin d'être acquise pour les N-et-N du français, ce qui m'amène à adopter parallèlement une autre attitude. Pour illustrer et bien identifier ce doute face aux N-et-N existentiels, je propose l'exemple forgé suivant. Dans un contexte où, au salon de l'agriculture, un paysan producteur de viande reçoit la visite d'une personnalité qui ne connaît pas sa région d'origine et complimente ses produits, le paysan formule l'invitation suivante :

(62) ma reziõ propoz d nõbrø prodyi d kalite e 3-vuz-ẽvit biẽ kordialmã a
pase a l okaziõ dã notr koperativ u nu-degystrõ par egzãpl #(dy) pẽ
e #(dy) fromaz akõpane d ẽ kry lokal

J'ai noté par des # l'inadéquation que je trouve à ce N-et-N *pẽ e fromaz* dans ce contexte où l'inattendu du contenu exact de l'invitation empêche

une pré-identification des référents de *pě e froma₃*, et donc une lecture définie, et où l’ancrage temporel *a l okaziõ*, associé au temps futur du verbe, le sens du prédicat *degyste* et la locution adverbiale *par egzãpl*, ne permettent pas non plus de lecture générique.

La conséquence de ce rejet des lectures existentielles des N-et-N est que le débat s’en trouve déplacé. En effet, les N-et-N deviennent de simples alternatives à la présence d’un déterminant pour des lectures qui exigent la projection DP, et non plus un moyen d’introduire une projection NumP sans déterminant. Les N-et-N posent alors seulement la question de savoir pourquoi ils donnent un statut de DP au syntagme nominal en FCT et pas en FD. Et comme cette question n’est pas reliée au nombre, je ne la traiterai pas ici.

Quoi tirer finalement de ce débat sur les N-et-N du FCT? Soit ils montrent qu’il faut donner un petit statut morphosyntaxique au nombre du FCT, c’est-à-dire que, grâce à la coordination, et selon un processus qui reste à mes yeux encore à définir de façon convaincante, ils créent les conditions d’une lecture existentielle, comme s’ils fournissaient un NumP. Les autres lectures possibles des N-et-N seraient alors à inscrire dans la problématique évoquée de la propriété des N-et-N de fournir un DP sans déterminant.

L’autre conclusion possible de ce débat est que le nombre du FCT n’a pas plus de statut morphosyntaxique que celui du FD, et la représentation du DP du FD de (51) et (52) conviendra au FCT, qui réalise alors le nom in situ également. Et cette conclusion est en contradiction avec le constat du chapitre 5 que le FCT marque possiblement le nom en nombre et pas le FD. Il manque aussi, comme dans la conclusion précédente, un cadre conceptuel expliquant la présence de N-et-N en FCT et pas en FD.

Puisque cette partie de la morphologie du nombre du FCT concernant la présence du suffixe de nombre sur les noms et adjectifs post-nominaux

n'a éventuellement pas trouvé de corrélation satisfaisante dans le modèle présenté, je propose ci-après une autre explication, qui tranchera nettement puisqu'elle se reposera sur des considérations extra-grammaticales.

6.3.2.2 Et si le nombre du FCT n'avait pas de statut

Le cadre conceptuel de cette section concerne le statut d'une catégorie précise de liaisons facultatives : les liaisons facultatives liées à des suffixes grammaticaux, c'est-à-dire les suffixes représentés par un morphogramme grammatical dans l'orthographe. On compte notamment parmi ces liaisons les marques de personne sur le verbe fléchi et le S de pluriel.

Au paragraphe sur les morphogrammes grammaticaux de la section 2.3.3.2, j'ai rapidement mis en doute le fait que les marques telles {z} et {t} fassent clairement partie de la compétence linguistique des locuteurs qui réalisent possiblement ces liaisons. Pour cela, j'ai remarqué que ces liaisons étaient d'abord favorisées par la lecture à haute voix, ou qu'elles étaient produites par des locuteurs très entraînés, probablement plus ou moins lecteurs et scripteurs experts, et dans des situations très formelles.

Je reformule cette même idée ici dans d'autres termes. Je suggère de faire de cette catégorie de liaisons facultatives et de leur réalisation le résultat d'une activité métalinguistique provoquée par la présence de l'écrit ou par la formalité d'une situation donnée. Cette activité métalinguistique doit reposer sur des connaissances métalinguistiques fortes, puisque l'orthographe grammaticale, selon sa réputation, est l'activité métalinguistique associée à l'activité d'écriture la plus difficile à apprendre et à maîtriser.

Si cette idée est conceptuellement satisfaisante, le FCT perd cette morphologie de nombre qui a posé problème à la fin de la section précé-

dente pour proposer un statut morphosyntaxique à son nombre qui soit consistant avec le modèle utilisé de corrélation morphologie–syntaxe–sémantique. La liaison de pluriel post-nominale ne demanderait alors plus à être intégrée à une modélisation de la compétence linguistique (c’est-à-dire de la grammaire interne) d’un locuteur du FCT.

6.3.2.3 Le français écrit : une langue naturelle ? !

Aux chapitres 2 et 3, j’ai déjà émis quelques griefs contre l’idée d’une langue écrite à considérer comme le fruit d’une compétence linguistique banale, donnant un accès légitime et non perturbé à la grammaire interne d’un locuteur. Et avant de remettre une couche à la section 11.4, je présente un autre grief.

Le français écrit correctement orthographié, compris comme une production langagière native présente des caractéristiques typologiquement contradictoires dans le cadre du modèle de la corrélation morphologie–syntaxe–sémantique de la section 6.2.3 :

- Sa morphologie de nombre se distance très peu de celle de l’espagnol. C’est-à-dire que, à part quelques remarques à faire sur la neutralisation du genre dans le pluriel des déterminants, on peut calquer les descriptions du français écrit sur celles de l’espagnol.
- Sa grammaire des NN est au contraire, à l’exception des N-et-N (seulement dans la mesure où ces N-et-N doivent bien être intégrés dans la grammaire des NN), équivalente à celle du FD.

En l’attente de trouver d’autres langues présentant ce patron grammatical, on restera sceptique quant à sa naturalité. D’aucuns souhaiteraient même en profiter pour justifier les difficultés et la résistance d’apprenants, pourtant natifs, du français écrit (ou de l’écriture du français) face à son orthographe, comme face à sa norme morphosyntaxique. À

ce titre, l'étude de copies d'étudiants de Barra Jover (2008:28ss.) invite clairement à douter d'une compétence linguistique native et spontanée cachée derrière bien des *s* de pluriel.

6.4 Synthèse

Ce chapitre a été l'occasion de réunir des idées et des propositions glanées dans la littérature. J'y fais bien peu de propositions originales. En effet, la plupart des considérations présentées ont déjà été formulées ailleurs : les représentations de différentes morphologies du nombre, la typologie des NN, leur place dans l'interface syntaxe-sémantique, l'intuition du rôle du nombre dans cette typologie, et les différents statuts morphosyntaxiques possibles du nombre à travers différentes langues avec articles.

6.4.1 Résultats

L'intérêt de cette étude a surtout été de proposer qu'il est conceptuellement attrayant que le statut à donner au nombre dans une langue particulière soit à inscrire dans une description consistante réunissant la morphologie du nombre, la syntaxe de certains syntagmes nominaux, et la sémantique des lectures associées à ces syntagmes.

Ceci a permis de dépasser la simple question « est-ce qu'il y a encore un *s* au pluriel en français ? » en la transformant en une autre question : « est-ce que la réalité toujours productive du suffixe de nombre du français lui donne un statut morphosyntaxique, et si oui, lequel ? »

La réponse apportée dans le cas du FD est clairement que le nombre n'y a pas de statut morphosyntaxique. Dans le cas du FCT, la réponse est plus complexe. Dans le premier cas de figure observé, le nombre a un

rôle morphosyntaxique à jouer. Mais ce rôle semble être plus à relier et conceptualiser autour de ce que la syntaxe des N-et-N, plus précisément à travers la coordination, peut fournir comme projection syntaxique liée au nombre et légitimant les N-et-N existentiels, plutôt qu'à appuyer sur la flexion possible des noms du FCT en nombre. Dans le second cas de figure, le débat a été déplacé hors de toute considération sur le nombre, lequel a alors exactement le statut du nombre du FD.

Ce refus d'accorder un statut à la liaison à droite du nom a enfin été conceptualisé autour de l'idée que les liaisons facultatives porteuses de marques grammaticales peuvent être vues comme des phénomènes métalinguistiques, donc sans statut aucun dans la grammaire interne des locuteurs du FCT.

6.4.2 Le nombre et l'hypothèse de la diglossie

En consacrant deux chapitres (une partie entière) au nombre, je suis finalement parvenu à la conclusion que le nombre du FD et le nombre du FCT ont une grammaire comparable, si ce n'est identique. Ceci peut paraître surprenant et décevant dans un travail qui veut fournir des arguments à l'hypothèse de la diglossie.

Cependant, je voudrais (évidemment) tirer un bilan positif de cette partie : la démarche postulant une diglossie a été intéressante. Elle a permis de se libérer de l'idée d'un S de pluriel homogène en motivant l'identification précise de « qui est qui » parmi la variété des marques de nombre en français. Elle a ainsi permis de trouver une position intermédiaire entre le déni du suffixe de nombre et son omniprésence, sans tomber dans la mollesse descriptive sur le mode « ce n'est pas grave si le nombre n'apparaît pas sur le nom, puisqu'une liaison est parfois possible (on a toujours une marque implicite qui peut parfois être explicitée phonolo-

Chapitre 6 Les statuts morphosyntaxiques du nombre

giquement) et que le déterminant compense en général ce manque (dans les faits, il y a très peu de cas vraiment ambigus et problématiques) ».

Ensuite, l'élargissement de la question morphologique aux questions syntaxiques et sémantiques et la dimension comparative ont permis, grâce au modèle présenté, de proposer un statut grammatical motivé et typologiquement consistant du nombre en français.

Enfin, les considérations sur les N-et-N du FCT ont mis à jour une autre problématique investissable dans le cadre de l'hypothèse diglossique, mais qui dépasse le cadre de cette partie : « quelle description pour les N-et-N du FCT, et, au-delà, quelle modélisation de ce point de variation entre FCT et FD ? »

Troisième partie

Éléments de syntaxe

Chapitre 7

La négation du verbe

LA NÉGATION est un sujet classique et souvent traité de la grammaire du français, que ce soit en termes variationnistes ou non. La description de la variation en question semble être assez simple (section 7.1). Pourtant, après s'être penché sur les notions de *négation verbale* et de *négation de constituant* (section 7.2), on avancera quelques exemples qui amèneront à remettre en question l'idée d'une syntaxe unifiée entre la variante avec *n* et la variante sans *n* de la négation du verbe (section 7.3). La synthèse (section 7.4) reprendra les propositions faites, notamment le fait de vouloir décrire la négation avec *n* comme une négation verbale et la négation sans *n* comme une négation de constituant, puis projetera les conséquences de cette description au sein de l'hypothèse de la diglossie.

7.1 Introduction : (n) fet pa atãsiõ a nø

Il est fort connu que les énoncés négatifs du français présentent une variation qui consiste, formulée naïvement, en l'optionalité du clitique de négation *n*, comme en (1).

- (1) a. ʒ-syi universiter me sa-vø pa dir grã ʒoz (UE : 27)

b. 3-syi universiter me sa-n-vø pa dir grã foz

Du point de vue sociolinguistique, on peut dire simplement que la présence de ce clitique apporte à l'énoncé une valeur ajoutée.

Cette proposition d'optionalité se trouve dans nombre d'ouvrages, sous différentes formes, et sans plus de justification que le simple constat d'exemples tels que (1).

Milner (1982b:186, note 1) fait de cette variation des *usages* : « La plupart des sujets ont à leur disposition les deux usages : avec *ne* dans la langue écrite, sans *ne* dans la langue parlée. Mis à part la présence ou l'absence de *ne*, les deux usages ne manifestent aucune propriété différentielle quand au système de la négation. » Il insiste en rapprochant ce cas de celui « du *e* muet en phonologie, qu'on s'accorde généralement à traiter par une règle d'effacement. Cela donne à penser qu'il existe aussi, dans le français parlé, une règle d'effacement de *ne*. »

Dans Recourcé (1996:75), cette optionalité corrélée au registre de langue est exprimée ainsi : « La présence d'un forclusif¹ [...] entraîne obligatoirement (en français standard) la réalisation locale d'une forme verbale en *ne* » tandis que (précision en note) « [o]n décrira la langue orale en supprimant cette contrainte de marquage du verbe dans la description des forclusifs ».

Zanuttini (1997:5 et 155, note 5) abandonne l'idée de se concentrer sur le français (« In this book I generally set aside the case of French, which is rather complex and would require me to draw not only dialectal distinctions among the varieties spoken in different geographical areas but also genre distinctions between formal and non-formal registers² »), et laisse

1. Le terme *forclusif* désigne les items négatifs de la classe de *pa*.

2. « Dans ce livre, je laisse en général de côté le cas du français, qui est plutôt complexe et me demanderait de distinguer non seulement des distinctions dialectales parmi les variétés parlées dans différents lieux géographiques, mais aussi des distinc-

7.1 Introduction : (*n*) *fet pa atãsiõ a nã*

simplement entendre, par des références, le même phénomène d'omission : « The pre-verbal negative marker *ne* is part of standard written French [...] It is often omitted in colloquial French spoken in France and Switzerland [...] and is present as a stylistic variable in the French spoken in Montréal [...] »³. Zanuttini n'est pas spécialiste du français, et on peut bien comprendre sa confusion devant les propos contradictoires tenus à ce sujet. Difficile d'exploiter des données de seconde main dans ces conditions.

Abeillé et Godard (1997) sont peu précises à ce sujet : l'exemple *ne lit pas Proust* introduit dans le texte (p. 2) est repris immédiatement dans un arbre syntaxique où il apparaît sous la forme (*ne*)-*lit pas Proust*, sans aucune précision sur le sens de ce parenthésage de *ne*, et sans que ce parenthésage soit jamais repris par ailleurs dans leur article. Est-ce qu'il faut seulement comprendre que les auteurs ne traitent pas les exemples tels (1a) ? ou alors très implicitement, et similairement aux autres auteurs ?

L'ensemble des traitements à base d'optionnalité ainsi rapidement esquissés ci-dessus reposent sur une hypothèse plus ou moins implicite selon les auteurs⁴, mais pas nécessairement innocente. Je résumerai cette hypothèse ainsi :

- (2) « La structure syntaxique de la négation du verbe en français est en un sens unique, la réalisation morphologique (ou phonologique) de cette structure est variable, en ceci que l'élément morphologique (ou

tions de genre entre les registres formel et informel. » (traduction : BM)

3. « Le marqueur négatif préverbal *ne* appartient au français écrit standard [...] Il est souvent omis en français parlé familier en France et en Suisse [...] et il est présent comme variable stylistique dans le français parlé à Montréal [...] » (traduction : BM)

4. Milner (1982b), cité plus haut, est le plus explicite, et ma formulation ne s'écarte pas fondamentalement de la sienne.

Chapitre 7 La négation du verbe

phonologique) *n* peut y être réalisé ou non sans changer la structure. »

Ce n'est peut-être pas un luxe que de rappeler qu'une autre hypothèse se propose *naturellement* :

- (3) « La négation du verbe du français est réalisable au moyen de deux structures syntaxiques distinctes ; l'une que l'on reconnaîtra au fait qu'elle mette en jeu l'élément *n*, l'autre au fait que cet élément n'y intervient pas. »

Je ne souhaite pas discuter ici les diverses propositions concernant la syntaxe de (1b), mais j'aimerais simplement inviter le lecteur à considérer les conséquences de la décision de décrire (1a) dans le cadre de l'hypothèse (2), et ce quelle que soit la description de (1b).

Pour la suite de ce chapitre, je retiens le terme '*pa* avec *n*' pour désigner les exemples du type (1b), et le terme '*pa* sans *n*' pour ceux du type (1a). Les autres items négatifs (*ply*, *riě*, *persòn*...) seront ignorés.

La section 7.2 présente et caractérise les notions de négation verbale et de négation de constituant, et on aura soin de distinguer à la lecture le terme générique 'négation du verbe' et le terme plus spécifique 'négation verbale'. Des exemples où la négation ne porte pas sur le verbe seront utilisés au cours de l'argumentation, mais la problématique de ce chapitre se concentrera bien autour des cas de négation du verbe.

La section 7.3 investit directement la problématique, et mènera, après examen de trois séries d'exemples, à la conclusion qu'il est préférable de décrire une négation de constituant pour la variante sans *n*, donc structurellement distincte de la variante avec *n*.

Enfin, la section 7.4 synthétise les résultats et les propositions, met en avant les avantages descriptifs de ces propositions, issues de l'hypothèse

(3), sur les propositions issues de l'hypothèse (2), puis fait le lien entre ces résultats et l'hypothèse de la diglossie, et rappelle quelques faits sur la négation qui n'auront pas été traités dans les limites de ce chapitre.

7.2 La négation verbale et la négation de constituant

À la base du raisonnement de ce chapitre, il faut comprendre les notions de *négation verbale* et *négation de constituant*, telles qu'elles sont présentées par exemple dans Barra Jover (2004a:113-116). Dans cette section je présente donc ces notions après avoir précisé une autre notion, celle de négation du verbe. Et je termine par la problématique à laquelle je répondrai à la section suivante : la négation du verbe par *pa* sans *n* est-elle une négation verbale ou bien une négation de constituant ?

7.2.1 La négation du verbe

Ce chapitre traite centralement de la négation du verbe, c'est-à-dire des énoncés où un item négatif est syntaxiquement placé sur le verbe. Et cette section cherche à cerner cette notion de *négation du verbe*. Le rapport syntaxique exact entre l'item négatif et le verbe est variable d'une langue à l'autre, et, d'après ce qu'on en dira pour le français, éventuellement au sein d'une même langue. Avant d'entrer dans les détails syntaxiques, je souhaite signaler que la négation du verbe possède une propriété qui permet, au moins en français, de la distinguer et de la définir par rapport aux négations non-verbales.

Lors d'une négation du verbe, même si la négation est attachée syntaxiquement au verbe, elle peut porter sémantiquement sur divers éléments

de l'énoncé. En (4), en fonction du focus pragmatique de la phrase (là où réside l'assertion pragmatique, cf. chapitre 8), on peut comprendre les diverses périphrases de (5)⁵.

(4) ʒ-n-syi paz ale o restorã italiẽ ijer avek ma fam

- (5) a. Dans un contexte discursif éventuellement vide, il est faux que je sois allé au restaurant italien hier avec ma femme.
b. Parmi ce qu'il est pertinent de dire sur moi, il est faux que je sois allé au restaurant italien hier avec ma femme.
c. Parmi ce qu'il est pertinent de dire à propos d'être allé au restaurant italien hier avec sa femme, il est faux que je l'aie fait.
d. Parmi ce qu'il est pertinent de dire à propos de ce que j'ai fait hier soir avec ma femme, il est faux que je sois allé au restaurant italien.
e. Parmi ce qu'il est pertinent de dire à propos de ce que je n'ai pas fait hier soir avec ma femme, il est vrai que je ne suis pas allé au restaurant italien.
f. ...

7.2.2 La négation verbale

La *négation verbale* (appelée parfois *négation d'énoncé*) est une structure syntaxique qui associe un item négatif au verbe dans un rapport particulier et exclusif. C'est-à-dire que la négation est un attribut du verbe sous forme d'une projection maximale qui ne peut être légitimée que par un verbe : pas de verbe, pas de négation possible.

5. Ceci sous-entend que (4) est la transcription unifiée d'énoncés distincts présentant des accents focaux différents. Le français n'est d'ailleurs pas la langue idéale pour exemplifier ce fait, puisque l'accent focal n'est pas le moyen privilégié pour refléter différentes structures informationnelles.

7.2 La négation verbale et la négation de constituant

La corrélation principale de la propriété d'un item négatif d'être engendré uniquement par un verbe est que, lorsqu'on veut nier un autre élément de l'énoncé, il faut malgré tout engendrer cet item autour d'un verbe, ou bien engendrer un autre item négatif, qui ne présente pas cette propriété de négation exclusivement verbale. Les exemples sous (6) montrent comment l'espagnol, dont la négation est préverbale, s'y prend pour nier un adjectif.

- (6) a. *Quiere ser feliz.*
il.veut être heureux
'Il veut être heureux.'
- b. **Quiere ser [no feliz].*
il.veut être NÉG heureux
- c. *Quiere [no ser] feliz.*
il.veut NÉG être heureux
'Il veut ne pas être heureux.'
- d. *[No quiere] ser feliz.*
NÉG il.veut être heureux
'Il ne veut pas être heureux.'
- e. *Quiere ser infeliz.*
il.veut être malheureux
'Il veut être malheureux.'

L'occitan⁶ a une négation post-verbale, ce qui complique un peu les analyses de (7).

- (7) a. *Vòl èsser aürós.*
il.veut être heureux
'Il veut être heureux.'
- b. ?*Vòl èsser pas aürós.*
il.veut être NÉG heureux

6. Les données de l'occitan présentées dans ce chapitre m'ont été aimablement fournies par Patrick Sauzet. Il s'agit ici de l'occitan central et oriental.

Chapitre 7 La négation du verbe

- c. *Vòl pas èsser aürós.*
il.veut NÉG être heureux
'Il veut ne pas être heureux.'/'Il ne veut pas être heureux.'
- d. *Vòl èsser malaürós.*
il.veut être malheureux
'Il veut être malheureux.'

En (7b), il n'est pas naturel de comprendre *pas aürós* comme un tout. Mais le tout *èsser pas aürós* est possible, surtout si on l'améliore avec *brica* ou *jamai pus*, comme en (8).

- (8) *Vòl èsser pas (brica/jamai pus) aürós.*
il.veut être NÉG morceau/jamais plus heureux
'Il veut être pas du tout/plus jamais heureux.'

Et puisque la négation post-verbale est placée avant un infinitif, (7c) est ambigu entre une négation de l'infinitif et une négation du modal, ce qu'indique (9).

- (9) a. *Vòl [pas èsser] aürós.*
il.veut NÉG être heureux
'Il veut ne pas être heureux.'
- b. *[Vòl pas] èsser aürós.*
il.veut NÉG être heureux
'Il ne veut pas être heureux.'

Ce que montrent principalement les données de l'espagnol et de l'occitan, c'est que leurs négations *no* et *pas* exigent de 's'accrocher' à un verbe.

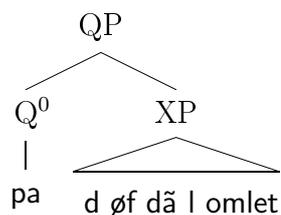
La négation *pa* avec *n* du français a la réputation d'être une négation verbale. En effet, (10) montre des contraintes similaires à celles mises en place en (6) et (7).

- (10) a. il-v \emptyset etr $\emptyset r \emptyset$
 b. *il-v \emptyset etr [n-pa $\emptyset r \emptyset$]
 c. il-v \emptyset [[n-pa etr] $\emptyset r \emptyset$]
 d. il-[[n-v \emptyset pa] etr $\emptyset r \emptyset$]
 e. il-v \emptyset etr mal $\emptyset r \emptyset$

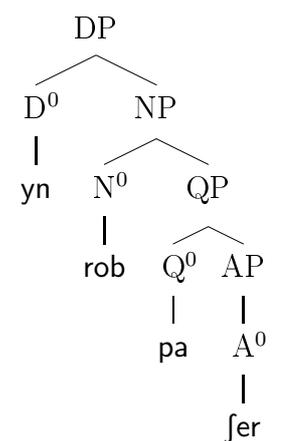
7.2.3 La négation de constituant

La *négation de constituant* est au contraire une négation indépendante de la génération d'un verbe. Il s'agit simplement de construire un QP avec un adverbe de quantification négative à sa tête. Ce QP gouverne alors à peu près n'importe quel type de constituant, le constituant verbal en étant un comme un autre. (11a) présente la négation d'un syntagme nominal, et (11b) la négation d'un adjectif.

- (11) a. pa d $\emptyset f$ dā l omlet



- b. yn rob pa fer



La suite du chapitre sera l'occasion de présenter d'autres exemples de négations de constituant du français, en se concentrant notamment sur leurs rapports au verbe de l'énoncé.

7.2.4 Les deux syntaxes de 'V *pa*'

Lorsque la négation sans *n* est placée syntaxiquement sur le verbe de l'énoncé, deux représentations structurales sont possibles. Tout d'abord, on peut y voir une instance de négation verbale, dans une structure unificatrice avec la négation avec *n*, ce qui correspond à l'hypothèse formulée sous (2). La représentation syntaxique proposée en (16) met en jeu deux projections syntaxiques : PolP et NegP. PolP est une projection syntaxique de la polarité de la proposition, qui génère *pa*. On la conçoit comme une des sous-projections de IP, située bas dans la hiérarchie, puisque *pa* apparaît à droite du verbe fléchi, et à gauche du verbe à l'infinitif. PolP est justifiable par l'existence de l'item *biě*, qui peut représenter le pendant positif de *pa*, et qui a la même distribution autour du verbe, ce qu'on voit en (12) à (14).

- (12) a. i-viědra pa
b. i-viědra biě

- (13) a. il-e pa vny
b. il-e biě vny

- (14) a. pās a pa kupe l gaz
b. pās a biě kupe l gaz

La projection NegP génère *n*. Elle est la projection qui permet au verbe de satisfaire son trait de polarité négative en montant dans la tête Neg⁰ où *n* se cliticise en tête du verbe. Il s'agit d'une contrainte mor-

7.2 La négation verbale et la négation de constituant

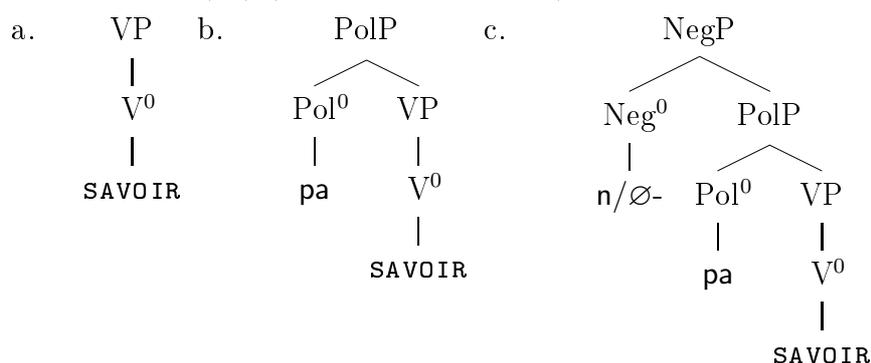
phosyntaxique, puisque la projection PolP porte à elle seule le contenu sémantique de la négation. Le mécanisme présenté autour de NegP reflète le propos déjà cité de Recourcé (1996:75) : « la présence d'un forclusif [...] entraîne obligatoirement (en français standard) la réalisation locale d'une forme verbale en *ne* », et « [o]n décrira la langue orale en supprimant cette contrainte de marquage du verbe dans la description des forclusifs ».

Lorsque *pa* apparaît sans *n*, est-ce que NegP est projetée avec une tête vide, ou bien est-ce qu'elle n'est simplement pas projetée ? Il semble que la réponse soit purement formelle, puisque la projection NegP elle-même a une motivation formelle et non sémantique. Si l'on considère que le verbe hérite d'un trait de polarité négative qu'il doit satisfaire, on peut dire soit qu'il peut le satisfaire par une tête vide, soit que cette satisfaction elle-même n'est pas obligatoire. Quitte à rapprocher *pa* sans *n* et *pa* avec *n*, on représentera la possibilité que la tête Neg⁰ puisse être réalisée par *n* ou par \emptyset .

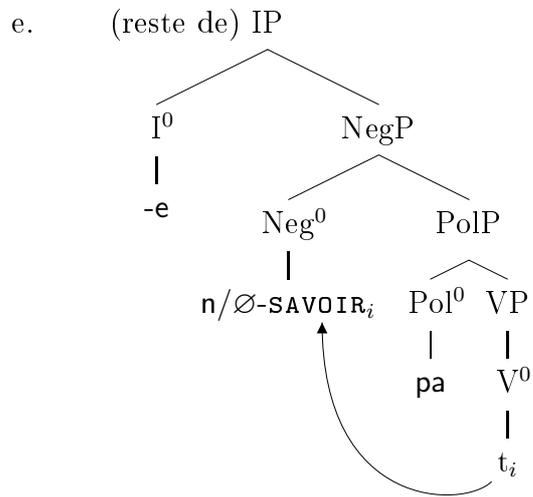
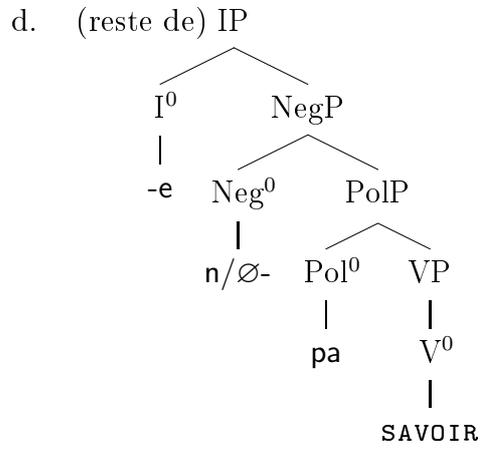
Enfin, *n* et *pa* sont générés directement dans leur ordre de surface, ce qui correspond à l'intuition justifiée dans Hirschbüller et Labelle (1993).

(15) ty-(n)-save *pa*

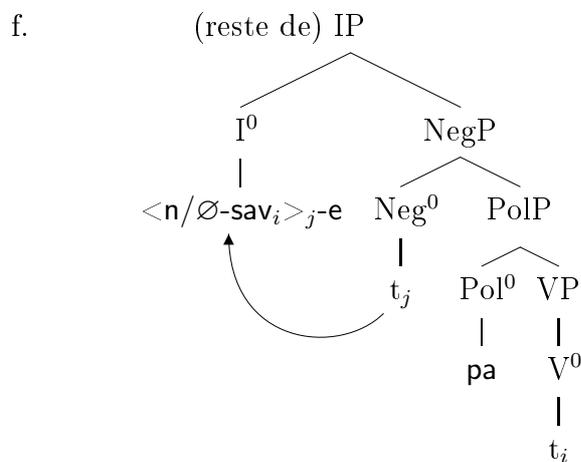
(16) dérivation de (15) (sans le clitique sujet)



Chapitre 7 La négation du verbe



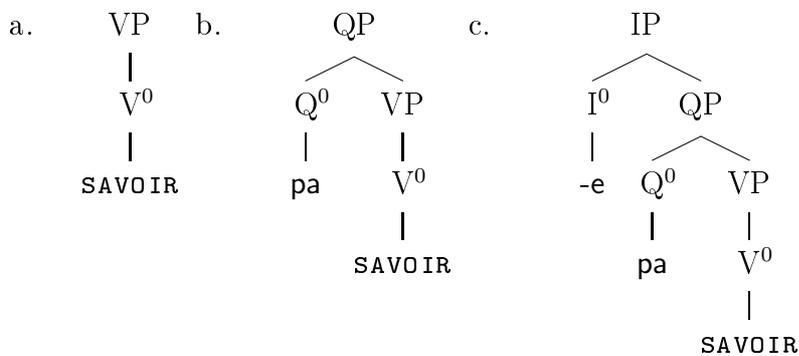
7.2 La négation verbale et la négation de constituant

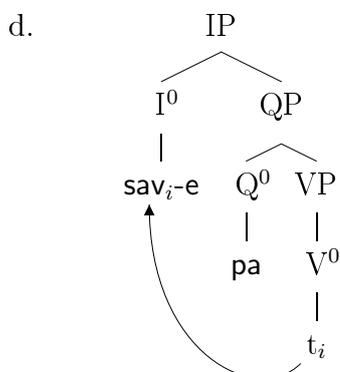


Mais on peut voir au contraire dans *pa* sans *n* une négation de constituant, où *pa* est simplement la tête d'un QP au-dessus de VP, similairement aux cas en (11). Ceci correspond alors à l'hypothèse sous (3), c'est-à-dire où *pa* sans *n* n'a pas la structure de *pa* avec *n*.

(17) *ty-save pa*

(18) dérivation de (17) (sans le clitique sujet)





On peut donc ici répéter la problématique de ce chapitre :

Problématique 2 *Quel est le rapport grammatical de **pa** sans *n* avec le verbe de l'énoncé ? Est-ce le même que celui de **pa** avec *n*, ou bien est-ce un rapport différent ?*

7.3 La négation, verbale ou de constituant ?

On a vu à la section précédente que la séquence 'V **pa**' était dérivable en surface à partir au moins de deux structures profondes. L'avantage de la négation verbale est d'unifier les négations avec et sans *n*. Étant donné que de nombreux locuteurs produisent visiblement relativement indifféremment l'une ou l'autre variante, notamment au sein d'un même discours, cette solution est intuitivement séduisante. On aurait une variation minimale, (morpho)phonologique et non pas structurale, autour de l'item *n*.

Cependant, les cas présentés ci-après plaident en faveur de la négation de constituant, puisqu'ils montrent des cas où la négation avec *n* n'est pas une variante possible de la négation sans *n*, et au contraire invitent à unifier la séquence 'V **pa**' avec d'autres séquences, soit simplement des séquences avec d'autres quantifieurs que **pa**, soit des séquences où

7.3 La négation, verbale ou de constituant ?

pa quantifie négativement sur d'autres constituants que le constituant verbal.

7.3.1 L'autonomie de *pa* sur le verbe

Comme le montrent les exemples suivants, *pa* peut se déplacer dans l'énoncé. On remarque, dans les cas proposés, qu'il se déplace sans *n*, ce qui invite à envisager sérieusement de décrire *pa* sans *n* comme une autre négation. Cette argumentation est fortement inspirée de Barra Jover (2004a:115-116), dont je tire certains exemples.

- (19) a. sa (n-)va pa etr evidã
b. sa (*n-)va etr pa evidã
- (20) a. mi∅ vo k el-(n-)krev pa
b. mi∅ (*n-)vo pa k el-(*n-)krev (Muller 1991 in Barra Jover 2004a)
- (21) a. e-(n-)s-abij pa pur fer
b. e-(*n-)s-abij pur pa fer (Muller 1991 in Barra Jover 2004a)

Mais, on pourrait vouloir décrire *pa* sans *n* également comme une négation verbale, simplement distincte de *pa* avec *n*. Outre la difficulté qu'on aurait à unifier les cas précédents autour d'une négation verbale de *pa* sans *n* (*pa* devraient être modélisés avec des propriétés très flottantes), les cas suivants montrent que ce n'est pas nécessaire, puisque *pa* peut très bien être généré sans verbe, dans ces énoncés averbaux.

- (22) A : ty v∅ yn pòm ?
B : pa mẽtnã
- (23) A : ma bel-mer el-e frãfmã sãpa
B : pa posibl!

Chapitre 7 La négation du verbe

(24) en commentant un tir qui rate sa cible de peu :

pa lwě

(25) pa d nuvel d pier

(Barra Jover 2004a)

Le cas de l'espagnol pousse également à croire à une négation non-verbale de *pa* sans *n*. En effet, sa négation pré-verbale en *no* n'a pas la mobilité de *pa*, ni son autonomie par rapport au verbe.

(26) a. *No ha comido rapido.*

NÉG a mangé vite
'Il n'a pas mangé vite.'

b. **Ha comido no rapido.*

a mangé NÉG vite

(27) a. *No vive lejos.*

NÉG vit loin
'Il n'habite pas loin.'

b. **Vive no lejos.*

vit NÉG loin

(28) ¿ *Donde vive ?*

'Il habite où ?'

a. *No *(vive) lejos.*

NÉG vit loin
'Il n'habite pas loin.'

Le cas de *muy* est à ce titre une exception, puisqu'il peut être modifié directement par *no*, c'est-à-dire sans l'intermédiaire d'un verbe.

(29) a. *Ha comido no muy rapido.*

a mangé NÉG très vite
'Il n'a pas mangé très vite.'

b. *Vive no muy lejos.*

vit NÉG très loin

7.3 La négation, verbale ou de constituant ?

‘Il n’habite pas très loin.’

(30) ¿ *Donde vive?*

‘Il habite où?’

a. *No (vive) muy lejos.*

NÉG vit très loin

‘(Il n’habite) pas très loin.’

En conclusion, je dirai que cette section a été l’occasion de découvrir dans les faits la propriété de *pa* d’être la tête d’un QP, et pas seulement une négation verbale. Ceci suggère de décrire la négation de *pa* sans *n* comme une négation de constituant *même lorsque ce constituant est le constituant verbal*. Mais comme il s’agit d’une simple suggestion, on se trouve encore sans réponse claire devant les structures où *pa* suit immédiatement le verbe fléchi.

(31) *ty-save pa*

a. par rapprochement avec la variante avec *n* :

$$[IP [I^0 < \emptyset -sav_i >_j -e] [NegP [Neg^0 t_j] [PolP [Pol^0 pa] [VP [V^0 t_i]]]]]$$

b. par rapprochement avec *pa* en tête d’autres constituants :

$$[IP [I^0 sav_i -e] [QP [Q^0 pa] [VP [V t_i]]]]]$$

La section suivante s’attachera à faire un pas de plus en direction de la négation de constituant. Pour cela, on observera l’interaction de la négation portant sur le verbe avec le quantifieur flottant *tu* lorsque celui-ci porte sur le sujet de l’énoncé tout en se plaçant, comme *pa*, entre l’auxiliaire et le participe passé des temps composés.

7.3.2 Négation et quantification universelle

7.3.2.1 Sans *n*

Observons une position possible du quantifieur universel en (32a). Malgré sa position, il a portée sur le sujet, puisque la forme logique de cet énoncé est (32b).

- (32) a. *ez-õ* tut parle
b. pour tout x parmi elles, x a parlé

Observons maintenant en (33) que *pa* apparaît dans la même position.

- (33) *ez-õ pa* parle

De trois choses l'une :

- Soit *pa* porte seulement sur le participe passé, et la trace de l'auxiliaire se trouve devant, comme indiqué en (34a). On a alors une négation de constituant sur le participe passé.
- Soit *pa* porte sur l'auxiliaire. Il est alors
 - soit une négation de constituant, comme indiqué en (34b).
 - soit une négation verbale, comme en (34c).

- (34) a. *ez-õ_i t_i [_{QP} pa parle]*
b. *ez-õ_i [_{QP} pa t_i parle]*
c. *ez-õ_i [_{PolP} pa t_i parle]*

Les formes logiques possibles pour (33) dépendent de la portée de *pa*. Si *pa* porte sur le participe passé, on obtient (35a). S'il porte sur le verbe, on obtient (35a), (35b) ou (35c) (cf. les portées possibles en (4)).

- (35) a. elles ont \neg (parlé)

7.3 La négation, verbale ou de constituant ?

- b. elles \neg (ont parlé)
- c. \neg (elles ont parlé)

Je laisse ici de côté la question de savoir si ces formes logiques sont équivalentes, c'est-à-dire si elles ont les mêmes conditions de vérité. Je supposerai simplement qu'elles le sont, puisque (33) me semble bien avoir toujours les mêmes conditions de vérité que (36), qui échange la négation de (33) avec l'antonyme de *parle*.

(36) el-s-sõ ty

On note de plus que la négation de (33) ne peut pas avoir portée exclusivement sur le sujet. (37) n'est pas une forme logique possible de (33).

(37) (\neg elles) ont parlé

Il semble que la syntaxe et la sémantique de *pa* soient à rapprocher de celles des adverbes de quantification tels *pø* ou *tro*, puisque les exemples en (38) ont les formes logiques en (39).

- (38) a. ez-õ pø parle
b. ez-õ tro parle

- (39) a. (i) elles ont [parlé] (en petite quantité)
(ii) elles [ont parlé] (en petite quantité)
(iii) [elles ont parlé] (en petite quantité)
b. (i) elles ont [parlé] (en trop grande quantité)
(ii) elles [ont parlé] (en trop grande quantité)
(iii) [elles ont parlé] (en trop grande quantité)

Chapitre 7 La négation du verbe

Notons bien que les quantifieurs de (38), comme le *pa* de (33), n'ont pas de portée stricte possible sur le sujet (contrairement à *tut* en (32)), ce qui les rapproche également. Les exemples sous (38) ne sont pas équivalents aux paraphrases de (40).

- (40) a. *i-ãn-a pø ki ã parle*
b. *i-ãn-a tro ki ã parle*

On peut maintenant essayer de combiner syntaxiquement la négation et la quantification universelle. Si on place la négation au-dessus de la quantification universelle, on obtient (41a), avec la forme logique (41b).

- (41) a. *ez-õ pa tut parle*
b. \neg (pour tout x parmi elles, x a parlé)

Ici, la négation prend obligatoirement la portée sur tout l'énoncé, c'est-à-dire que le quantifieur universel *tu*, en montant en tête de la forme logique, entraîne *pa* avec lui, qui continue d'avoir portée sur lui.

Si maintenant on place la quantification universelle au-dessus de la négation, on obtient (42a), de forme logique (42b), c'est-à-dire où cette fois c'est la quantification universelle qui a portée sur la négation, quelle que soit la place de la négation dans la forme logique parmi les trois possibles observées en (34).

- (42) a. *ez-õ tut pa parle*
b. pour tout x parmi elles, x a \neg (parlé)
ou : pour tout x parmi elles, x \neg (a parlé)
ou : pour tout x parmi elles, \neg (x a parlé)

Comment interpréter ce contraste ? Admettons tout d'abord que dans (32), la possibilité pour le quantifieur de porter sur le sujet vient du fait

7.3 La négation, verbale ou de constituant ?

que ce dernier est généré bas, avec le verbe, et donc dans la portée du quantifieur, avant de monter dans la flexion, ce que représente (43).

(43) $ez_j-\tilde{o}_i$ tut t_j-t_i parle

Donc, en insérant une négation, elle peut d'une part se trouver au-dessus de *tut* et aura portée directement sur *tut*, qui lui-même aura toujours portée sur le sujet.

(44) $ez_j-\tilde{o}_i$ pa tut t_j-t_i parle

Et d'autre part, cette négation peut être insérée en-dessous de *tut*. Dans ce cas, qu'elle soit au-dessus ou en-dessous des traces, elle gardera sa portée habituelle, sans pouvoir avoir portée sur le sujet, comme on l'a vu en (35).

(45) a. $ez_j-\tilde{o}_i$ tut pa t_j-t_i parle
b. $ez_j-\tilde{o}_i$ tut t_j-t_i pa parle

Ma suggestion sera de voir en (44) une négation verbale, ce qui prédit sa portée sur tout l'énoncé, et en (45) une négation de constituant, ici le constituant verbal, dont la portée ne dépasse pas le domaine de ce constituant. C'est-à-dire que la position de surface de la négation en (33), répété ci-dessous, est ambiguë entre les deux structures syntaxiques en question.

(33) $ez-\tilde{o}$ pa parle

7.3.2.2 Avec n

Si, comme le dit Milner, « les deux usages ne manifestent aucune propriété différentielle quand au système de la négation », la variante (46)

Chapitre 7 La négation du verbe

de (33) doit montrer les mêmes propriétés lorsqu'elle est combinée avec la quantification universelle, suggérant la même ambiguïté syntaxique.

(46) *el-n-õ pa parle*

Or, cette variante n'est compatible avec la quantification universelle que lorsque *pa* reste collé au verbe fléchi, comme en (47a). (47b) montre au contraire que *pa* hors de la portée de *n* entraîne une agrammaticalité.

(47) a. *el-n-õ pa tut parle*
b. **el-n-õ tut pa parle*

Donc, le *pa* de (33) était bien au moins ambigu entre deux structures syntaxiques, alors que celui de (46) ne l'est pas.

(48) $\langle \text{el-n-}\tilde{\text{o}}_i \rangle_j$ [_{NegP} t_j [_{PolP} *pa* [_{VP} t_i *parle*]]]

Ainsi, on a une raison supplémentaire d'émettre des hypothèses proposant des propriétés structurelles distinctes pour le *pa* sans *n* et pour celui avec *n*.

7.3.3 L'impératif

La négation verbale des impératifs suggère clairement que les deux variantes ont des comportements syntaxiques distincts. Les clitiques objets se placent après le verbe dans l'impératif affirmatif (enclitiques), alors qu'ils sont proclitiques à l'indicatif (comparer (49) et (50)).

(49) indicatif :
a. **ty-repet-lə*
b. *ty-l-repet*

7.3 La négation, verbale ou de constituant ?

(50) impératif affirmatif :

- a. *repet-lə*
- b. **l-repet*

En présence de la négation sans *n*, on trouve la version proclitique comme la version enclitique :

(51) impératif négatif (sans *n*) :

- a. *repet-lə pa*
- b. *l-repet pa*

Ce dernier exemple contraste étonnamment en fait avec les données proposées par la négation avec *n*, qui bloque la variante avec enclitique.

(52) impératif négatif (avec *n*) :

- a. **n-repet-lə pa*
- b. *n-l-repet pa*

Donc il semble que la syntaxe de négation mise en place en (51), plus précisément en (51a), d'une certaine façon, ne doive rien à *n*.

La comparaison avec d'autres quantifieurs en (53) est révélatrice : la variante *repet-lə* est compatible avec les quantifieurs et pas avec la négation verbale, et la variante *l-repet* n'existe que sous l'influence de la négation verbale — un quantifieur autre que *pa* (ou pas de quantifieur) ne peut légitimer seul le proclitique à l'impératif. D'où les structures proposées en (54).

(53) d'autres quantifieurs et l'impératif :

- a. *repet-lə pø/tro/boku*
- b. **l-repet pø/tro/boku*

- (54) a. (n-)l-repet pa
 $[_{IP} \langle n/\emptyset\text{-l-repet}_i \rangle_j [_{NegP} t_j [_{PoIP} pa [_{VP} t_i]]]]$
 b. (*n-)repet-lə pa
 $[_{IP} \langle \text{repet-lə} \rangle_i [_{QP} pa [_{VP} t_i]]]$

En conclusion, on dira d'un côté que *l-repet pa* est une négation verbale, de ces négations reconnaissables à leurs rapports syntaxiques particuliers au verbe, ici la propriété de légitimer le proclitique/bloquer l'enclitique de l'impératif. Et de l'autre côté, *repet-lə pa* est une négation de constituant, c'est-à-dire sans influence syntaxique particulière sur le constituant nié, au même titre que d'autres quantifieurs.

7.4 Synthèse

7.4.1 Les propositions

De ce chapitre ressortent trois propositions principales concernant la syntaxe de la négation du verbe en français.

1. *ty-n-save pa* est une négation verbale, avec une variante (morpho)phonologique sans *n*, avec des propriétés typiques des négations verbales.
2. *ty-save pa* est une négation verbale, avec une variante (morpho)phonologique avec *n*, avec des propriétés typiques des négations verbales.
3. *ty-save pa* est une négation de constituant, donc une variante syntaxique, structurellement éloignée, de *ty-(n-)save pa*, avec une distribution de simple quantification.

En termes folkloriques, il ne semble plus possible de pouvoir certifier que produire *pa* sans *n* soit, en toute situation, « un oubli, une simplification, un manquement... », dès que l'on parle d'un locuteur qui produit les exemples qui ont mené à décrire une négation du verbe par *pa* en termes de négation de constituant et non de négation verbale.

7.4.2 *pa* et les zones zribi-hertziennes

Comment ces propositions s'inscrivent-elles dans l'hypothèse de la diglossie, en particulier en termes de zones zribi-hertziennes ? On se rappellera ici que l'hypothèse de la diglossie a été formulée en termes de grammaires individuelles, donc que tous les locuteurs ne présentent pas forcément la même répartition des variantes dans Z1, Z2 et Z3.

Je présenterai trois types de locuteurs, à titre illustratif, sans prétendre ni qu'ils existent tous les trois, ni qu'il n'y en a pas d'autres. Simplement, ils me semblent être les plus plausibles. J'exclurai a priori les locuteurs qui ne produisent que la variante avec *n*, ou que les variantes sans *n*, puisque ces cas ne sont pas informatifs dans le cadre de mon hypothèse.

7.4.2.1 Le grand diglosse

Quel serait le locuteur idéal au sein de l'hypothèse de la diglossie ? C'est celui qui aurait une Z2 vide, et des variantes Z1 et Z3 les plus éloignées grammaticalement. La négation verbale avec *n* serait sa variante Z1. Et sa variante Z3 serait la négation de constituant. Un tel locuteur produirait les exemples qui montrent l'existence de cette négation de constituant, répétés en (55), et exclurait les énoncés tels (56), ce qui permettrait d'exclure toute négation verbale de son FD.

(55) a. sa va etr pa evidã

- b. miø vo pa k e-krev
- c. è-s-abij pur pa fer
- d. pa d nuvel d pier
- e. il-a mǎže pa (trè) vit
- f. ez-õ tut pa parle
- g. repet-lə pa

(56) *I-repet pa

Les descriptions de la variation à base d'optionalité de *n* sont inadéquates pour ce type de locuteur. Supposer au contraire que les variantes sont générées différemment et par des grammaires implémentées indépendamment est beaucoup plus satisfaisant.

7.4.2.2 Le petit diglosse

Si un locuteur produit la négation verbale avec *n* dans des situations informelles, alors il faut envisager que cette variante soit chez lui en Z2. Pour le confirmer, il faudrait encore vérifier systématiquement que *n* se combine bien avec tout type de variante (Z1, Z2 et surtout Z3), ce qui ne me semble pas être une évidence.

Si par ailleurs ce locuteur ne produit aucun des exemples proposés en (55), et nettement marqués comme en bas de l'échelle des valeurs sociolinguistiques, qui induisent l'existence de la négation de constituant, alors ses variantes sans *n* sont des négations verbales.

Pour ce locuteur, sa variation autour de la négation du verbe est grammaticalement minimale. Son FCT contient en Z2 la négation verbale avec *n*, et son FD contient deux variantes : la négation verbale avec *n* en Z2, et la négation verbale sans *n* en Z3.

La simple optionalité de *n* décrit très bien ce cas, ce qui fait que l'idée

de deux grammaires autonomes pour générer les deux variantes donne l'impression de se forcer à enfoncer une malheureusement petite pointe avec une masse.

7.4.2.3 Le chaotique

Le locuteur chaotique est celui qui présente les trois variantes de la négation du verbe.

Comme pour les deux types précédents, la négation verbale avec n entre dans sa zone Z1 ou Z2, en fonction de sa possibilité de la produire en combinaison avec des variantes Z3⁷.

Puisqu'elles sont exclues de la norme, les deux variantes sans n de ce locuteur sont en Z3, et rendent la caractérisation de la variation de son FD plutôt chaotique. En effet, à l'intérieur de son FD, on trouve deux variantes structurelles assez éloignées, que l'on aurait préféré décrire dans deux grammaires distinctes, tandis que l'une des deux variantes, la négation verbale sans n , est structurellement équivalente à sa variante du FCT, bien que les deux variantes soient par hypothèse générées par deux grammaires autonomes.

J'emploie le terme 'chaotique' pour suggérer non pas que la situation de ce type de locuteur est anarchique, sans régularité, mais au contraire qu'elle est, telle un chaos stable, dynamique dans son organisation.

7. Le chapitre 10 sera l'occasion d'observer le cas d'un tel locuteur ne produisant pas cette variante combinée à une variante purement démotique telle que l'emploi de \bar{o} comme première personne du pluriel.

7.4.3 Des aspects que ce chapitre n'a pas traités

7.4.3.1 La diachronie

Comme dans l'ensemble de cette thèse, je n'ai envisagé la variation que comme une dynamique synchronique de la grammaire. L'hypothèse de la diglossie prédit également une dynamique diachronique dans les rapports génétiques entre le FCT et le FD. On trouve une telle démarche dans Barra Jover (2004a), qui décrit la situation actuelle, avec ses trois variantes, comme une sortie du cycle de Jespersen, puisque celui-ci prévoit la négation post-verbale à l'occitane, mais pas l'émergence d'une négation de constituant à partir de cette négation post-verbale.

7.4.3.2 *pa* et les autres

Dans ce chapitre, je me suis restreint à l'étude de *pa*. Mais la même démarche serait nécessaire pour les autres items négatifs combinables avec *n* : *ply*, *riě*, *persòn*, *zame*, *okě*... Rien ne dit en effet qu'ils présentent la même distribution. On pourrait imaginer par exemple, en termes diachroniques comme en termes synchroniques, que l'implémentation de la négation de constituant face à la négation verbale ne s'étend pas à toute la classe des items négatifs.

7.4.3.3 Mélanger la négation et d'autres variables

Enfin, il me reste à annoncer le programme de ce qui pourrait corrélérer les propositions exposées dans ce chapitre. Un énoncé combine facilement la négation avec d'autres variables, par exemple la première personne du pluriel en *nu* ou en *õ*. Quelles variantes de la négation se combinent avec quelles variantes des autres variables ? Est-ce que tous les mélanges sont possibles ? Dans ce cas, l'hypothèse de la diglossie serait très affaiblie.

Est-ce que certains mélanges sont impossibles ? Dans ce cas, est-ce que les mélanges possibles laissent les supposées grammaires du FCT et du FD consistantes ?

Un premier indice est déjà donné par la distribution du mélange entre négation et placement du clitique de l'impératif, puisque, d'après cette distribution, on peut établir les deux sous-grammaires consistantes où d'une part le proclitique est en Z2 et la négation avec *n* en Z1 (grammaire du FCT), et d'autre part le proclitique en Z2 et l'enclitique et les négations sans *n* en Z3 (grammaire du FD).

- (57) a. n-l-repet pa
 b. l-repet pa
 c. *n-repet-lə pa
 d. repet-lə pa

Ce sera le propos du chapitre 10 que de tenter systématiquement d'observer dans les faits si les grammaires induites par ces distributions partielles restent consistantes.

Chapitre 8

Syntaxe et structure informationnelle en variation¹

8.1 la dislokasiõ è-ser a kwa ?

Ce chapitre est né d'une longue interrogation autour de la dislocation à gauche du sujet, en termes normatifs « la redondance du pronom sujet ».

Dans le soucis de légitimer les démotismes comme celui-ci, on a pu proposer des explications functionalistes, dans des termes qui permettent de contredire l'idée de redondance — c'est-à-dire d'inutilité, voire d'absence de logique chez les locuteurs de cette variante — du discours folklorico-normatif. Je citerai deux de ces propositions, avec mes principaux griefs.

1. La dislocation à gauche a une fonction 'focalisante', ce qui justifie

1. Je souhaite remercier sincèrement les étudiants des deux Hauptseminare de linguistique du français que j'ai donnés à Stuttgart (*Linguistique du français démotique*, semestre d'hiver 2007-2008 et *Syntaxe de la structure informationnelle en français*, semestre d'été 2008) qui, s'ils ont appris quoi que ce soit, ne se doutent pas à quel point j'en ai sûrement appris bien plus qu'eux, grâce à leur engagement tout au long de cette année. Et merci à Achim Stein de m'avoir laissé les deux fois le choix du thème du cours. Sans ces cours, je n'aurais pas pu développer et structurer les idées de ce chapitre.

cette variante comme n'en étant pas une (c'est en fait une variante pragmatique, non synonyme de la variante sans dislocation, cf. section 3.1.1). Il me semble que la notion de focalisation est issue à nouveau de jugements normatifs, puisque mes intuitions propres m'indiquent qu'il est rare en fait que je produise cette dislocation pour « insister sur le sujet ». Les situations non-focalisantes restent alors injustifiées et toujours aussi illégitimes. De plus, et on le verra au cours de ce chapitre (section 8.2.3.1), la notion de focalisation n'est pas précise, voire mal venue, puisque ce qui est mis en avant ici n'est pas le focus tel qu'il est en général compris dans la littérature sur la structure informationnelle, mais en fait recouvre la notion de contraste.

2. La dislocation à gauche n'est pas issue d'une cliticisation du sujet. Le pronom sujet n'est pas un vrai pronom, mais est un marqueur de personne préfixal (Lambrecht, 1981), à comparer avec les désinences de personne de l'espagnol ou de l'ancien français, ou la notion de marqueur sujet développée pour la description des langues polysynthétiques. Cette approche manque à mon avis un fait crucial : si l'accord sujet-verbe était réalisé par les clitiques sujets, alors les cas pourtant attestés de non-dislocation en FD (cf. sections 8.2.3.4, 10.2.4 et 10.3.3) seraient des cas sans accord sujet-verbe, ce qui ferait du FD une langue curieuse au sein de la typologie de cet accord. On ne peut pas simplement comparer le *il-* de (58) au *-a* de (59) (exemple espagnol), puisqu'il est possible, en (58), de supprimer cet accord, et pas en (59). De même, le marqueur sujet *zi* du chicheŵa², langue bantoue, en (60) n'est pas omissible (Mchombo, 2003:142).

2. Les chiffres 10 et 2 dans les gloses indiquent des classes nominales.

- (58) mǒ jǐě (il)-abwa boku
- (59) *El perro mío ladr-*(a) mucho*
le chien mien aboie-3_{SG} beaucoup
'Mon chien aboie beaucoup.'
- (60) *Njúchí *(zi)-na-lúmá alenje*
10.abeilles 10.SUJ-PASS-piquer 2.chasseurs
'Les abeilles ont piqué les chasseurs.'

Pour aborder cette question, j'ai finalement choisi, dans la diversité des approches, de m'en remettre aux travaux de Lambrecht, présentés ci-après, qui lient la dislocation à gauche et d'autres constructions à la structure informationnelle (ci-après SI) dans sa dimension pragmatique³. L'hypothèse de travail est que la dislocation à gauche comme les autres constructions que l'on verra sont à décrire autour de la notion de topique de phrase, telle que Lambrecht la traite. Dans ce qui suit, les généralités sur la SI sont tirées de Lambrecht (1994), et les applications de la SI au français de Lambrecht (1981, 1987, 1994, 2000, 2002).

Dans la section 8.2, après avoir retiré de ces travaux des informations sur la SI en général et le topique en particulier, on verra une présentation, partielle et simplifiée pour les besoins de mon propos, des constructions syntaxiques liées à diverses structures topicales en français. Puis, on envisagera d'interpréter la variation, observable et évaluable en termes sociolinguistiques, autour de certaines de ces structures topicales (sec-

3. L'idée d'aller chercher du côté de la SI m'a été soufflée par Anne Zribi-Hertz (c.p.) qui a suggéré de corréler le contraste syntaxique en (i) au contraste entre jugement thétique et jugement catégorique.

- (i) a. 1) l aviõ a ete deturne
2) ãn aviõ a ete deturne
b. 1) l aviõ il-a ete deturne
2) ??ãn aviõ il-a ete deturne

tion 8.3). Enfin, la section 8.4 est consacrée à la tentative de répartition des variantes dans les zones zribi-hertziennes.

8.2 Les travaux de Lambrecht : SI et topique en français

8.2.1 Terminologie

Puisque j'adopte la description et le cadre théorique de Lambrecht, j'adopte également dans les limites de ce chapitre sa terminologie.

proposition L'unité de description de la SI de Lambrecht est la proposition.

structure (informationnelle) Une structure (informationnelle) est le rapport pragmatique qu'entretiennent certains éléments d'une proposition avec l'ensemble de celle-ci. Son nom portera une majuscule (Topique, Anti-Topique, ...).

construction Une structure morphosyntaxique particulière et corrélée avec une structure informationnelle donnée sera une construction.

Ainsi, en (61), la *construction* clivée en *c'est* introduit la *structure d'identification Focus-Fond*, où le constituant *mõ frer* est en *relation de focus* avec la proposition.

(61) s-e mõ frer k a komãse a rigole

8.2.2 Lambrecht (1994) : généralités sur la SI et le topique

Lambrecht (1994) développe un modèle fin et presque complet de la notion de structure informationnelle pragmatique. Notons bien qu'il ne s'intéresse pas aux effets sémantiques de la SI, c'est-à-dire aux effets liés aux conditions de vérité d'un énoncé, mais au contraire aux ancrages pragmatiques de la SI, c'est-à-dire à la pertinence de l'information et de sa structuration dans un énoncé au sein d'un contexte discursif particulier. Dans la suite du chapitre, j'omettrai fréquemment de préciser cette préoccupation pragmatique et non sémantique.

8.2.2.1 Généralités

Lambrecht part du constat qu'il existe dans les langues ce qu'il appelle des *allophrases*. Il s'agit de propositions dont le contenu strictement sémantique est identique, mais dont la pertinence pragmatique au sein d'un discours donné est différente, comme on peut le voir intuitivement en (62).

- (62) a. mĩ frer a komãse a rigole
b. mĩ frer il-a komãse a rigole
c. il-a komãse a rigole mĩ frer
d. il-a komãse a rigole
e. i-a mĩ frer k a komãse a rigole
f. 3-e mĩ frer k a komãse a rigole
g. s-e mĩ frer k a komãse a rigole

Quelles sont ces pertinences pragmatiques ? Lambrecht en définit principalement deux : la relation de *focus* et la relation de *topique*. Je

laisse de côté le focus, et je me concentre sur le topique, qui permettra à lui seul de présenter de nombreuses structures informationnelles et leurs constructions.

8.2.2.2 Le topique

Lambrecht définit le topique en termes d'*aboutness* de la proposition, ce que je traduirai par l'*à-propos* de la proposition. Selon cette idée, la pertinence d'une proposition peut être de tenir un propos concernant un à-propos. La structure informationnelle d'une telle proposition sera dite de type *Topique-Commentaire*. Lorsqu'une langue possède le trait grammatical de topique, elle marque l'à-propos d'une proposition à travers une construction qui permettra d'identifier cet à-propos en le mettant en relation de topique avec la proposition. Il ne faut pas assimiler le topique d'une proposition avec le concept de topique de discours : le topique d'une proposition est grammaticalement identifié à l'intérieur de la proposition, par une construction de la proposition. Le fait que tenir un propos concernant un à-propos soit pertinent à un instant du discours n'implique pas que cet à-propos soit le thème actuel du discours.

Pour ce qui est de ce chapitre, je m'en tiendrai aux cas où le référent du topique de la proposition est également le référent de son sujet syntaxique. Les exemples en (63), de Lambrecht (1994:137), illustrent trois grandes réalités morphosyntaxiques de marquage de topique dans trois langues distinctes.

(63) structures Topique-Commentaire en anglais, italien et japonais :

- a. My neck HURTS.
- b. *Il collo mi fa* MALE.
le cou me fait mal

8.2 Les travaux de Lambrecht : SI et topique en français

- c. *Kubi wa* ITAL.
cou TOP fait-mal

Lambrecht s'empresse d'opposer ces structures Topique-Commentaire aux structures Tout-Focus, dans ces mêmes langues, lesquelles structures ont le même contenu sémantique, mais ne marquent pas le sujet comme ayant une relation de topique avec la proposition, mais au contraire signalent que la proposition n'a pas d'à-propos⁴.

(64) structures Tout-Focus en anglais, italien et japonais :

- a. My NECK hurts.
b. *Mi fa male il* COLLO.
me fait mal le cou
c. KUBI *ga* ITAL.
cou NOM fait-mal

On peut remarquer ici que les structures morphosyntaxiques proposées par ces trois langues pour opposer les structures Topique-Commentaire et Tout-Focus semblent unifiables autour de l'idée que la structure Tout-Focus est grammaticalement marquée par rapport à la structure Topique-Commentaire, elle grammaticalement canonique. En effet, la prosodie de l'exemple anglais en (64a) est 'déplacée' par rapport à celle de (63a), l'ordre canonique SV de l'italien en (63b) est 'inversé' en (64b), et la prosodie particulière de (64c), ainsi que la marque de nominatif *wa* sur *kubi*, contrastent avec celles de (63c).

4. Cette opposition rappelle son équivalent de la tradition sémantique qui oppose le jugement catégorique et le jugement thétique (Kuroda, 1972, Sasse, 1987, Laduslaw, 1994).

8.2.2.3 L'identifiabilité des référents

Une autre notion importante pour le topique, c'est l'*identifiabilité des référents*. En effet, pour qu'un référent puisse entrer en relation de topique avec une proposition, il faut qu'il soit identifié, au moins une fois la proposition énoncée. Lambrecht distingue plusieurs classes de référents sur une échelle d'identifiabilité. On trouve en bas de cette échelle les référents qui ne sont pas identifiés dans le discours au moment d'énoncer la proposition dans laquelle ils sont employés. Juste au-dessus se trouvent les référents qui ont été identifiés dans le discours, mais qui ne sont pas actifs au moment de les énoncer, soit parce qu'ils ont été identifiés trop longtemps auparavant, soit parce que d'autres référents sont plus actifs qu'eux. Ensuite, il y a les référents qui sont actifs dans le discours, soit parce qu'ils viennent d'être identifiés, soit parce qu'ils sont déictiquement facilement récupérables dans le contexte de l'énonciation. Enfin, le haut de l'échelle est occupé par les référents qui sont des topiques établis dans le discours, par exemple parce qu'ils viennent d'être identifiés comme l'à-propos d'une proposition.

8.2.3 Le topique en français

En français, lorsque le sujet est l'à-propos d'une proposition, il peut être encodé de différentes façons dans la construction de la proposition, en fonction de son identifiabilité dans le contexte discursif. On distinguera deux ensembles de cas principalement : les cas où le référent du sujet est identifiable, et les cas où il ne l'est pas. On présentera ensuite deux cas où il n'y a pas de topique dans la proposition : le cas des propositions Tout-Focus, et le cas des propositions Arrière-Plan. Enfin, on décrira ce que Lambrecht appelle le type (T) pro-V (X) (AT), qui semble être

la structure syntaxique unifiant toutes les allophrases présentées. Les notations crochétées et indicées ci-après ([]_t, []_T, []_{AT}, ...) indiquent les constituants à observer. Elles sont reprises de Lambrecht.

8.2.3.1 Topique et Anti-Topique

Le premier cas sera celui où le référent du sujet de la proposition est un topique établi dans le discours. Il n'est alors pas nécessaire d'y référer lexicalement. Le clitique sujet suffit, comme dans la réponse de B en (65).

- (65) A : t-a biěn ävwaje l koli ?
 B : ui normalmä [il]_t-arivra dă dø zur

Puis, moins le référent du sujet est facilement identifiable, plus il doit être identifié comme topique par la construction de la proposition, à commencer par le fait d'être évoqué lexicalement. Un topique dont le référent est actif dans le discours peut être encodé au moyen de la dislocation à droite, ce que Lambrecht (1981, 1994) appelle un Anti-Topique.

- (66) [il]_t-arivra dă dø zur [l koli]_{AT}

Si maintenant le référent du sujet n'est pas actif dans le discours, mais récupérable, par exemple parce qu'il a déjà été déjà évoqué mais n'est plus actif, la dislocation à gauche signale ce Topique :

- (67) [l koli]_T [il]_t-arivra dă dø zur

Signalons ici un phénomène important : chaque construction évoquée a sa fonction pragmatique à l'intérieur de la proposition, tout en ayant sa pertinence pragmatique déterminée par le contexte. Le contexte en lui-même ne provoque pas d'agrammaticalité, mais un effet pragmatique

d'inadéquation : lorsque l'encodage du topique de la proposition ne correspond pas au degré d'identifiabilité du référent du topique, l'interlocuteur doit *accomoder* ce référent, ce qui crée éventuellement des problèmes d'interprétation de la pertinence pragmatique de l'énoncé, ou bien simplement des jeux plus ou moins volontaires par effet de surprise, comme dans cet extrait de Pierre Desproges, où le topique établi du discours n'est pas celui qu'on croit :

- (68) la zɛstasiõ ʃe la fam dyr dø-sã-swasãt-di zur o kur dekel el-s-ãpifr s-ãledi zemi vagmã tut ã kõtribyã a fer grẽpe le kurb d l apsãteism dã l ãtrpriz
o bu d se nøf mwa l ptit òm viët o mõi
l akuʃmã e dulurø
øøzmã la fam tiẽ la mẽ d l òm
#ẽsi [il]_t-sufɾ mwẽ

Dans cet extrait, le discours établit progressivement la femme enceinte comme thème général. On s'attend alors à ce que l'avant-dernier énoncé *øøzmã la fam tiẽ la mẽ d l òm* en fasse un topique établi pour l'énoncé suivant. Mais ce dernier utilise *l òm* comme topique établi, ce qui provoque une surprise à la source d'un effet que l'auteur veut humoristique.

Il ne faut pas confondre la relation de topique, et le caractère contrastif qu'il peut avoir, lequel lui est orthogonal : ce qui est topical n'est pas nécessairement contrastif, et ce qui est contrastif n'est pas nécessairement topical. En (69), le premier topique est contrastif, pas le second. Et en (70), le premier focus est contrastif, pas le second.

- (69) A : t-e frãse twa d tut fasõ
B : nõ
[ma mer]_T el-e frãsez

[mwa]_T ʒ-e pa la nasionalite

(70) A : avek tɔ̃ per ki e frãse s-e normal k ty-parl biẽ

B : nɔ̃

s-e [ma mer]_F ki e frãsez

me ui s-e [mɔ̃ per]_F ki a vuly k ʒ-swa bilẽg

8.2.3.2 Les constructions relatives de présentation

Se reposer sur la capacité d'accommodation est risqué, et il n'est pas toujours question de jouer sur les effets de surprise. Donc, il reste un cas particulier de construction encodant un topique de proposition : le cas où le référent du topique est parfaitement inconnu du contexte discursif, ou dont l'identifiabilité est trop faible pour ne pas poser problème. Lambrecht (2000, 2002) propose de considérer un ensemble de structures propositionnelles comme ayant la fonction pragmatique d'introduire un référent dans le discours pour pouvoir en faire immédiatement l'à-propos d'un propos (le topique d'un commentaire). Il s'agit de ce que Lambrecht appelle *les constructions relatives de présentation* (CRP). Les CRP sont organisées en deux parties. La première est une proposition principale existentielle qui introduit un référent lexicalement, et la seconde partie est une proposition subordonnée relative qui prend le référent introduit dans la principale comme antécédent et prédique un propos sur lui. Lambrecht signale également que ces référents sont ensuite des topiques établis idéaux. Parmi toutes les CRP qu'il présente, je retiendrai pour mon exposé le prototypique *il était une fois*, et la clivée en *avoir*, qui retiendra plus précisément mon attention par la suite.

(71) a. il-etet yn fwa yn prẽses ki vive dãz ã grã fato

[el]_t-ete promiz a ã bo mariaʒ avek l prẽs d la kɔ̃tre vwazin

- b. atã yn sgõd t-a tõi gato ki va krame
sor-[lə]_t dy fur
- c. ʒ-e me parã k õ arete d fyne
dy ku [i]_t-m-met la presiõ pur k ʒ-aret osi
- d. õn-a y l fãpiona d ørop k a y liø a liõ (Philippe Candeloro)

8.2.3.3 Les propositions Tout-Focus

Lorsque l'assertion concerne l'ensemble de la proposition, sans qu'elle ait d'à-propos, la clivée en *il y a* signale cette structure de Tout-Focus⁵. Son but est de retirer toute idée de topique à l'intérieur de la proposition, c'est-à-dire de signaler que la proposition ne contient pas de pré-supposition pragmatique. On dit que ces propositions rapportent un fait ('event-reporting'). On illustre généralement ce cas à l'aide d'une question portant sur l'ensemble d'un état de chose.

- (72) A : qwa d nøf ?
B : i-a mõ fiẽ k a ãkor fe de sien

8.2.3.4 Les propositions Arrière-Plan

Face à toutes ces constructions sans aucun sujet lexical (c'est le topique qui y est exprimé lexicalement, pas le sujet), Lambrecht (1987) constate

5. Lambrecht ne différencie pas systématiquement les clivées en *avoir* avec pronom personnel et les clivées en *il y a*. Pour lui, (ia) et (ib) sont des CRP comme des structures Tout-Focus :

- (i) a. ʒ-e ma vvatyr k e ã pan (Lambrecht (1994))
- b. i-a l telefõn ki sòn (Lambrecht (2002))

Dans la suite, je désignerai ces deux types de clivées sous le terme générique *clivées en avoir*.

8.2 Les travaux de Lambrecht : SI et topique en français

que ces sujets lexicaux (c'est-à-dire des syntagmes nominaux en position strictement pré-verbale, sans clitique sujet) se trouvent parfois, et dans des conditions pragmatiques particulières. Il décrit ainsi la fonction informationnelle de ces cas, qu'il nomme les structures d'Arrière-Plan :

« Lexical subject NPs in spoken French strongly tend to have referents whose topic status is low, and they often occur in backgrounded portions of the discourse. »⁶ (Lambrecht, 1987:235)

Pour montrer cette propriété des sujets lexicaux d'être en arrière-plan du discours, Lambrecht (1987) propose des exemples de début de récit où se succèdent des personnages introduits comme sujets lexicaux dans ce qui constitue le décor du récit avant d'entrer dans le vif du récit et d'introduire les personnages principaux par des Topiques ou des CRP. À côté de ces cas où le statut faiblement topical des référents est à comprendre sur l'ensemble du discours, Lambrecht remarque des situations où ce statut topical bas est inscrit à l'intérieur même d'un énoncé : il est assez fréquent de trouver des sujets lexicaux plutôt que des Topiques ou des Anti-Topiques dans les subordonnées, lesquelles introduisent assez naturellement l'arrière-plan de leur principale. Pour cela, Lambrecht observe un extrait de dialogue où deux référents en avant-plan contrastent avec un référent en arrière-plan :

(73) Lambrecht (1987:236), exemple du corpus François (1974) :

A : *ya des écoles aussi ?*

B : *ah ya des écoles, ah oui ya des écoles (...)*

A : *et **le lycée d'Argenteuil** (T_i)...*

6. « En français parlé, les NP sujets lexicaux ont tendance à avoir des référents au statut topical bas, et ils apparaissent souvent dans des parties en arrière-plan du discours. » [traduction : BM]

- B : *ça y est*
A : *... ça va être par là ?*
B : *oui*
C : *non, écoute, moi je n'sais plus, parce que là **ce lycée que tu me dis sur le boulevard** (T_j)*
B : *oui sur le boulevard*
C : *bah alors, **ce lycée** (T_j) on m'a dit qu'il était pas terminé et qu'i'serait à peine terminé pour quand **les enfants** (S) prendraient au lycée, alors je voudrais...*
B : *alors peut-être*
C : *... savoir où est-ce qu'il est, **ce lycée** (AT_i), si c'est celui qu'tu m'dis qu'ya déjà les enfants dedans*

Dans ce dialogue, il s'agit principalement, pour le locuteur C, de contester l'identité supposée par B entre *le lycée d'Argenteuil* (indiqué i) et *le lycée [du boulevard]* (indiqué j). Ces deux référents sont introduits par une structure de Topique, et le premier est même repris comme Anti-Topique, montrant ainsi sa continuité topicale. En revanche, le référent *les enfants* apparaît dans une subordonnée en *quand*, et comme il n'est clairement pas le topique de l'énoncé (c'est *ce lycée*), il est introduit sous la forme d'un sujet lexical. Le référent *les enfants* n'est pas là pour prendre le rôle de topique de sa proposition, puisque sa proposition elle-même n'est que l'arrière-plan de la principale.

8.2.3.5 Le type (T) pro-V (X) (AT)

Lambrecht (1987) veut voir dans l'ensemble des constructions observées, à l'exception de l'exceptionnel SV(O), des réalisations d'un patron unique d'ordre des mots. Pour cela, il remarque que, dans toutes les propositions de ces allophrases, le verbe est précédé d'un clitique sujet

8.2 Les travaux de Lambrecht : SI et topique en français

(Lambrecht préfère le terme *pronom lié*) : le pro-V. Puis, on voit que les éventuels autres éléments syntaxiquement internes à la proposition sont placés après le verbe : le (X). Enfin, il est possible de détacher syntaxiquement un Topique ou un Anti-Topique : le (T) et le (AT).

- (74) a. [mõ frer]_T [il]_{pro}-[a]_V [komãse a rigole]_X
 b. [il]_{pro}-[a]_V [komãse a rigole]_X [mõ frer]_{AT}
 c. [il]_{pro}-[a]_V [komãse a rigole]_X
 d. [i]_{pro}-[a]_V [mõ frer]_X [k]_{pro} [a]_V [komãse a rigole]_X
 e. [ʒ]_{pro}-[e]_V [mõ frer]_X [k]_{pro} [a]_V [komãse a rigole]_X
 f. [s]_{pro}-[e]_V [mõ frer]_X [k]_{pro} [a]_V [komãse a rigole]_X

Pour appuyer la pertinence de cette observation, Lambrecht (1987) évoque brièvement, sans présenter de données, des études du sacapultec, du samoan et du sesotho montrant l'existence de ce patron dans ces langues. Plus proche de nous, je citerai le poyaudin, qui présente par exemple les allophrases en (76), dont le contenu sémantique est celui de l'exemple français (75).

(75) no tomat myris

(76) poyaudin :

- a. *no tomat a-mørisõ*
 nos tomates elles-mûrissent
 b. *a-mørisõ no tomat*
 elles-mûrissent nos tomates
 c. *i-a no tomat k a-mørisõ*
 y-a nos tomates que elles-mûrissent
 d. *ʒ-õ no tomat k a-mørisõ*
 nous-avons nos tomates que elles-mûrissent

- e. *s-e no tomat k a-mørisõ*
c-est nos tomates que elles-mûrissent

8.3 Syntaxe, SI et variation

Cette présentation succincte et légèrement idéalisée des travaux de Lambrecht dresse un tableau qu'il est possible d'interpréter en termes de variation.

8.3.1 La norme face aux allophrases de Lambrecht

Parmi les constructions du français présentées, certaines reçoivent un jugement sociolinguistique plus ou moins marqué.

La dislocation à gauche est admise par la norme dans la mesure où elle marque un topique contrastif, ce qui n'est pas la fonction pragmatique qui lui a été accordée ici. On peut donc avancer que la dislocation à gauche comme structure de Topique est exclue de la norme, qui lui préfère le classique SV(O).

De même, les clivées en *avoir* sont peu valorisées par la norme, à nouveau face à SV(O). Pourtant, certaines CRP ne sont pas stigmatisées, comme la CRP en *il était une fois*.

On pourra remarquer qu'apparemment, les CRP admises par la norme montrent une prédication seconde, puisque la relative peut y être omise, alors que les CRP stigmatisées sont bien des clivées, qui sans relative perdent leur pertinence pragmatique, c'est-à-dire leur fonction présentative ((77) vs. (78)).

- (77) a. *il-etet yn fwa yn prèses (ki abite dāz ã viø fato)*
b. *vwala l faktør (ki ariv)* (Lambrecht (2000))

- (78) a. ʒ-e ma vvatyr #(k e ã pan) (Lambrecht (1994))
b. i-a l telefòn #(ki sòn) (Lambrecht (2002))

Les allophrases de Lambrecht peuvent ainsi se prêter à une interprétation variationniste, ce à quoi je vais passer maintenant.

8.3.2 La variation autour de la SI

De trois choses l'une :

1. soit les allophrases étudiées par Lambrecht sont une variation simplement à corrélérer avec la SI, et on peut se contenter de se dire étudier le français tout court ;
2. soit ces allophrases, à corrélérer avec la SI, sont également à corrélérer avec la variation diamésique (l'opposition spontané-planifié, cf. section 3.1.6) ;
3. soit, enfin, la syntaxe corrélérable avec la SI varie sociolinguistiquement, et on peut émettre l'hypothèse d'une simple variation sociolinguistique, ou bien l'hypothèse d'une diglossie.

Dans ce qui suit, je vais proposer un éclairage des trois positions, pour finir en suggérant que la diglossie, si elle n'est pas forcément supérieure en tout point à l'hypothèse de la diamésie, est au moins aussi intéressante et rend compte seule de certains faits.

8.3.2.1 Corréler variation et SI

C'est le propos de Lambrecht que de comprendre comment les allophrases en français sont corrélables à la SI. Mais, comme indiqué précédemment, certaines allophrases sont exclues par la norme, et ne font donc pas partie du FCT. C'est-à-dire que SV(O) et la dislocation à gauche, ou

SV(O) et les clivées en *avoir*, s'ils semblent porteurs d'un contraste pragmatique à l'intérieur du FD, ne le sont plus si on les considère comme porteurs d'un simple contraste sociolinguistique, l'un et l'autre étant alors pragmatiquement interchangeables.

8.3.2.2 La SI en variation diamésique

Dans cette hypothèse, une situation de communication spontanée implique un investissement pragmatique du locuteur, qui peut alors modifier la construction de base SV(O) pour refléter la SI de ses énoncés.

Cette hypothèse a un certain pouvoir explicatif, mais se heurte à certaines contradictions, que je voudrais présenter :

1. a) Si on prend l'opposition oral-écrit comme explicative, on ne rend pas compte des éventuelles absences de SI à l'oral formel, et des éventuelles présences de SI à l'écrit informel.
- b) Si on reformule cette opposition dans les termes spontané-planifié, on colle probablement mieux aux données, mais il reste assez acrobatique de justifier ainsi un investissement supérieur dans la SI dans les situations spontanées par rapport aux situations planifiées. Ici, on pourrait inverser la formulation en disant qu'un locuteur peut s'impliquer plus avant dans la *syntaxe*, en faisant par exemple cumuler la fonction syntaxique de sujet et la fonction pragmatique de topique.
- c) Si enfin on reformule l'opposition dans les termes proche-distant de Koch et Oesterreicher (1990), on peut continuer de voir la SI comme un investissement du locuteur pour refléter formellement la proximité de la situation d'énonciation.
- d) Pourtant, il reste des cas difficilement traitables en ces termes, puisqu'on trouve des discours spontanés sans SI, comme dans

le corpus *L'aventure du français fondamental* (in Blanche-Benveniste, 2002a:132-135), où le locuteur principal réalise presque exclusivement des propositions SV(O), alors qu'il semble bien parler spontanément.⁷

2. a) Si on admet malgré tout le pouvoir explicatif de l'opposition diamésique, on se heurte à des difficultés typologiques. En effet, il faut encore s'expliquer pourquoi la spontanéité ne provoque pas cet investissement dans la SI dans d'autres langues SV(O), ou, si elle le fait, pourquoi ces autres langues SV(O) ne projettent pas la SI avec des constructions équivalentes à celles du français. En effet, en anglais, les dislocations à gauche ou à droite sont plutôt mal venues (en tout cas, elles sont tout sauf systématiques) :

- (79) a. *My dad is a great dad.*
 b. *He's a great dad.*
 c. *??/? My dad he's a great dad.*
 d. *??He's a great dad my dad.*

- b) Il devient alors plus simple de décrire, avec Lambrecht, un type distinct pour le FD.

8.3.2.3 La SI en variation sociolinguistique

L'adoption ou non de certaines allophrases (la dislocation à gauche, ou les clivées en *avoir*, au lieu de SV(O)) peut être éclairée par des considé-

7. Il manque à cette démonstration un point de comparaison où le même locuteur peut produire des alternatives à la structure SV(O). On aurait ainsi la preuve qu'il est bien à la fois locuteur spontané du FCT et du FD, et que son choix de produire du FCT n'est pas expliqué par la spontanéité de la situation, ni par son incapacité à produire du FD.

rations sociolinguistiques. Soit que ces considérations se contiennent dans une grammaire, soit qu'elles se répartissent diglossiquement. Je prendrai bien sûr le parti d'y voir la présence d'énoncés produits par deux grammaires distinctes. Il y a ici plusieurs avantages à cela.

1. Il ne paraît plus nécessaire de tout dériver de SV(O). (T) pro-V (X) (AT) peut être généré pour et par lui-même, ce qui promet une modélisation de la grammaire nettement plus simple.
2. Une corrélation explicative acrobatique entre la syntaxe des allophrases et la situation d'énonciation n'est plus nécessaire non plus.
3. Décrire une grammaire du FD seul permet de le situer plus sereinement dans une typologie de l'ordre des mots canonique à travers les langues : le FD devient une langue (T) pro-V (X) (AT), un type apparemment connu, alors que le FCT est bien du type classiquement décrit SV(O).
4. Puisque le FD n'est pas SV(O), on s'explique mieux pourquoi l'explication diamésique ne s'applique pas aussi simplement à d'autres langues SV(O).

8.4 Le français, SV(O) ou (T) pro-V (X) (AT) ?

Je voudrais conclure ce chapitre en comparant ses résultats à travers les lunettes de la diglossie aux résultats du chapitre 7. En effet, interpréter la variation liée à la SI en termes de zones zribi-hertziennes va rappeler ce qu'on avait conclu pour la négation.

Comme on s'était demandé si le FD et le FCT sont des langues à négation verbale ou négation de constituant, on se demandera ici si le FD et

8.4 Le français, SV(O) ou (T) pro-V (X) (AT) ?

le FCT sont des langues SV(O) ou bien des langues (T) pro-V (X) (AT). Les résultats de cette question se laissent assez bien visualiser dans le tableau 8.1.

	FCT		
SI	Z1	Z2	Z3
Topique établi		pro-V X	
Topique	SVO		T pro-V X
Anti-Topique	SVO		pro-V X AT
Présentation		CRP en <i>il était une fois</i>	clivée en <i>avoir</i>
Arrière-Plan		SVO	
Tout-Focus	SVO		clivée en <i>il y a</i>
		FD	

TABLEAU 8.1: La SI en variation et les zones zribi-hertziennes

Comme dans le cas de la négation, la répartition des variantes observées doit être établie pour chaque locuteur. À nouveau, je propose d'imaginer et détailler trois types de locuteurs, qui se distinguent notamment autour de la place de la structure syntaxique SV(O) dans un idiolecte donné.

Dans les trois cas, la modélisation du FCT pourra par exemple retenir la construction SV(O) comme structure générique des propositions, puis suggérer la dérivation de pro-V (X) à partir de SV(O) (par cliticisation), et enfin décrire certaines CRP, comme celle en *il était une fois*.

8.4.1 Le petit diglosse

Le petit diglosse sera le locuteur qui laisse une grande place à la construction SV(O) dans son FD, c'est-à-dire que, par exemple, la dislocation ne sera pas systématique pour encoder les topiques exprimés lexicalement, ou bien les clivées en *avoir* pour les énoncés Tout-Focus.

À côté de son FCT conforme à la norme, son FD ne contiendra que peu des constructions telles que la dislocation, et elles pourront être dérivées également à partir de SV(O). Il faudra également décrire parallèlement plus de constructions clivées que dans son FCT.

Le petit diglosse invite finalement plus à élargir la modélisation du FCT qu'à proposer une modélisation indépendante pour le FD.

8.4.2 Le grand diglosse

Le grand diglosse a ceci de particulier que son FD ne présente aucune construction SV(O), même dans le cas des propositions Arrière-Plan, qui seront du type (T) pro-V (X) (AT). Son FCT se modélisera comme celui du petit diglosse.

En revanche, son FD innovera nettement, et pourra être modélisé exclusivement autour du type (T) pro-V (X) (AT), parallèlement à la description de nombreuses constructions clivées.

Le grand diglosse invite nettement à décrire un FD autonome, un FD où tous les verbes finis sont accompagnés d'un clitique sujet, lequel clitique pourra être l'objet du débat sur sa nature éventuelle de simple marque d'accord sujet-verbe.

8.4.3 Le chaotique

Comme le chaotique de la négation, le chaotique de la SI a ceci de troublant que les variantes SV(O), dans l'idéal confinées au FCT comme chez le grand diglosse, sont toutes possiblement de type Z2, à commencer par les propositions Arrière-Plan, ensuite éventuellement les propositions Tout-Focus, moins probablement les propositions avec Topique ou Anti-Topique. Ceci ne change pas la description du FCT, mais il faut revoir le

8.4 Le français, SV(O) ou (T) pro-V (X) (AT) ?

FD qu'il y a alors à modéliser.

Le FD du chaotique induit une modélisation acrobatique, puisqu'il tient du type SV(O) comme du type (T) pro-V (X) (AT). Il faut décrire et engendrer les deux types dans le même modèle grammatical, sans pour autant qu'il soit intuitivement optimal de dériver dynamiquement et synchroniquement l'un à partir de l'autre.

8.4.4 Et la dislocation ? Et le clitique sujet ?

La description proposée dans ce chapitre de la variation autour de la structure informationnelle permet d'envisager la question du statut de la dislocation et du clitique sujet sous un angle particulier. Lambrecht (1981) veut clairement voir un accord verbal dans le clitique sujet, un *marqueur sujet* dans les termes des descriptions de langues polysynthétiques. Le concept même de dislocation n'y trouve plus sa place, puisqu'il n'est pas question de voir en (T) pro-V (X) (AT) une transformation à partir de SV(O). Au contraire, (T) pro-V (X) (AT) est une construction grammaticale dans le sens donné à ce terme par la *grammaire de construction*, engendrée directement par le modèle grammatical.

Pourtant, seul le cas du grand diglosse laisse envisager cette solution parfaitement sereinement. Le chaotique, en particulier, résiste à une généralisation descriptive autour du seul type (T) pro-V (X) (AT), qui devra être envisagé parallèlement au type SV(O). Les réalisations de surface telles que Topique et Anti-topique restent syntaxiquement ambiguës entre la dislocation et la construction (T) pro-V (X) (AT). Et il est également indécidable de décrire le clitique sujet comme un accord verbal ou comme une cliticisation du sujet syntaxique.

En face de cette situation du FD, celle du FCT est plus facile à envisager : SV(O), construction générique, a parfois l'occasion de se dislo-

quer, entraînant ainsi la seule véritable interprétation répertoriée dans la norme, l'interprétation de topique contrastif. Le FCT est bien une langue SV(O), et le clitique sujet a bien le rôle syntaxique de sujet.

Une dernière observation rapproche le cas de la SI de celui de la négation : nombreux sont les énoncés syntaxiquement ambigus entre une analyse SV(O) et une analyse (T) pro-V (X) (AT). Ce sont là tous les cas où l'on ne trouve pas de sujet ni de topique exprimé lexicalement :

(80) [il]_{t/s}-a dy pẽ syr la plãf

- a. (T) pro-V (X) (AT), obtenu par omission de la référence lexicale au topique :

[mõ kwafør]_T [il]_t-a dy pẽ syr la plãf

- b. SV(O), obtenu par cliticisation du sujet :

[mõ kwafør]_S a dy pẽ syr la plãf

Ceci n'est pas sans rappeler les énoncés structurellement ambigus entre une négation verbale et une négation de constituant.

Enfin, une dernière citation de Lambrecht (1987:255), que je sors de son contexte argumentatif puisque je ne souhaite pas en discuter ici, laisse voir quels genres de révolutions dans la description linguistique du FD sont envisageables dès lors qu'on laisse décrire ce FD sans laisser prédominer la construction SV(O) :

« It is but a small step from here to the conclusion that there exists an *ergative undercurrent* in the syntax of spoken French. »⁸

8. « Il n'y a qu'un petit pas à franchir pour aboutir à la conclusion qu'il y a un *courant sous-jacent ergatif* dans la syntaxe du français. » [traduction : BM]

Quatrième partie

Est-on cohérents ?

CETTE partie se propose de plaider en faveur de l'hypothèse de la diglossie non pas en suggérant qu'elle permet une description plus intéressante théoriquement et typologiquement de la variation morpho-syntaxique comme dans la partie III, mais en la fondant empiriquement, en explorant une méthode à même d'induire la diglossie à partir des données empiriques.

La méthode est présentée au chapitre 9, avant d'être confiée aux jugements de grammaticalité sur quelques petits points de variation de mon propre idiolecte. Je proposerai à cette occasion une rapide critique du bien-fondé de s'en remettre entièrement à de tels jugements.

Puis, cette méthode est l'objet d'une étude empirique sur corpus au chapitre 10, ce qui constituera le point culminant de cette partie, où, après avoir induit sans soucis deux sous-grammaires partielles pour le FCT et le FD du locuteur étudié, je ferai quelques remarques sur les limites et limitations de l'application empirique de cette méthode.

Enfin, cette partie sera conclue par un chapitre qui exposera le cas concret d'une donnée apparemment problématique pour l'hypothèse de la diglossie, et suggèrera une explication à cette donnée contradictoire en termes d'interférences entre les deux grammaires des locuteurs qui l'ont produite (chapitre 11).

Chapitre 9

Quoi chercher ?

DANS ce chapitre, j'avance une méthode empirique (section 9.1), adaptée de Barra Jover (2007), qui doit proposer l'induction de l'hypothèse de la diglossie, et notamment sa représentation zribi-hertzienne, à partir de données empiriques. Cette méthode est explorable ensuite en interrogeant les jugements de grammaticalité des locuteurs (section 9.2) ou bien en scrutant un corpus oral spontané (chapitre 10). Chacune de ses deux applications a ses propres problèmes, que je signalerai quand ils se présenteront. La méthode employée prévoit aussi la possibilité que des données réfutent l'hypothèse testée. Je suggère brièvement une attitude possible à adopter face à de telles données (section 9.1.2.4), laquelle attitude sera exemplifiée par le chapitre 11.

9.1 But et méthode

9.1.1 But

À la section 3.4, j'ai exposé la formulation de l'hypothèse de la diglossie en termes de double grammaire interne. Puis j'ai présenté la visualisation de cette hypothèse à l'aide des zones zribi-hertziennes. Ainsi présentée, la

Chapitre 9 Quoi chercher ?

diglossie fait une prédiction forte : les productions des locuteurs diglosses doivent être cohérentes. C'est-à-dire qu'elles doivent être produites entièrement par une des deux grammaires, soit celle du FCT, soit celle du FD. Un locuteur « n'active » qu'une grammaire à la fois (cf. Zribi-Hertz, 2006). Si un locuteur, au contraire, produit des énoncés qui relèvent à la fois de son FCT et de son FD, alors l'idée que sa grammaire interne est structurée selon le schéma diglossique est simplement caduque.

Le but de cette partie est donc le suivant :

- (1) Les productions d'un locuteur supposé diglosse FCT-FD doivent être grammaticalement cohérentes. Ce locuteur doit activer une seule grammaire à la fois dans un énoncé (soit son FCT soit son FD, mais pas des deux à la fois). Qu'est-ce qu'il en est exactement ?

Il n'est pas dit dans la problématique qu'un locuteur doive être cohérent à l'échelle d'un discours. On sait d'ailleurs de longue date que cette cohérence n'existe pas, et le chapitre 10 le montrera, si c'est encore nécessaire. Simplement, puisque l'on prédit une cohérence *grammaticale*, elle sera observée à l'échelle des unités de la grammaire, et non des unités du discours. Quelle est cette unité de la grammaire à l'intérieur de laquelle l'hypothèse prédit une cohérence ? Il faudrait pour cela une théorie précise, à laquelle je n'ai pas accès. Je considérerai provisoirement qu'il s'agit de l'énoncé, sans d'ailleurs lui donner de définition (je n'en ai pas non plus). Mais, je serai amené à réviser ceci à la lumière de l'étude du chapitre 10. Plus exactement, je prendrai le parti d'exiger la cohérence grammaticale au sein des propositions produites par le locuteur que j'étudierai.

Pourquoi la proposition comme domaine maximal de cohérence gram-

maticale? Pas pour des motivations théoriques, puisque je n'en ai pas. Ni par intuition, je n'en ai pas plus. Mais simplement parce que ça marche, alors que le domaine de l'énoncé n'est pas cohérent. Comme j'ai effectué ce choix a posteriori, je ne peux pas le justifier ici. Dans le chapitre 10, je reviendrai en détail sur le constat que le domaine de la proposition fonctionne (c'est-à-dire que les productions sont grammaticalement cohérentes au sein de toutes les propositions du corpus), alors que le domaine de l'énoncé est trop grand (section 10.3.7). Et je suggérerai quelques considérations théoriques autour de cette problématique à la section 10.4.3.2. Dans les limites du présent chapitre, je me restreindrai à des énoncés mono-propositionnels, donc la question n'y aura pas de conséquence majeure.

Dans la section suivante (9.1.2), j'exposerai une méthode pour observer cette cohérence grammaticale, si elle existe. On verra que l'on peut l'appliquer en s'appuyant sur des jugements de grammaticalité. Mais on verra également que cette application a ses faiblesses. On verra ensuite qu'une application de la méthode est envisageable pour vérifier cette production sur des données de corpus. Ensuite, le chapitre se terminera sur un fragment d'étude auto-observée (donc à base de jugements de grammaticalité) de cette cohérence. Le chapitre suivant pourra alors se consacrer à une étude plus détaillée d'un corpus oral spontané, dont au moins un des intérêts sera de pallier au manque de clareté et d'adéquation du fait de confier la méthode à l'auto-observation comme cela est fait à la section 9.2.

9.1.2 Méthode

9.1.2.1 La méthode du conditionnel (Barra Jover, 2007)

Barra Jover (2007) présente le principe de cohérence grammaticale. Il l'applique dans Barra Jover (2002) à un cas où les jugements de grammaticalité font défaut : le cas des corpus diachroniques. Le principe est le suivant : on se doit de décrire des grammaires cohérentes. C'est-à-dire, dans le cas de la diachronie, ne pas confondre l'apparition d'une variante innovante dans les productions d'un locuteur particulier (dans une grammaire interne) et l'établissement de cette variante comme caractéristique d'une langue externe. Pour ses besoins, Barra Jover développe une méthodologie qui permet de vérifier des hypothèses sur un changement diachronique en respectant le principe de cohérence grammaticale. Je retiens de cette méthodologie trois exigences. La première est que l'on doit étudier des idiolectes, et donc que les hypothèses proposées doivent être vérifiées au sein d'une grammaire cohérente. La seconde est qu'il faut définir le domaine de cohérence grammaticale exigible en fonction de l'hypothèse à vérifier. Et la troisième est que l'hypothèse doit être formulable sous forme de conditionnel.

Barra Jover (2007:110-111) décrit ainsi son étude de 2002 :

- a) L'hypothèse « B apparaît parce que A » devient le conditionnel expérimental « si un locuteur X produit B, alors ce même locuteur X doit produire A ». [...]
- b) Le conditionnel est testé sur chaque locuteur.

Son hypothèse est diachronique, et il propose donc que le domaine de cohérence grammaticale exigible est l'idiolecte ponctuel entier d'un locuteur, c'est-à-dire dans son cas un texte entier produit par un auteur donné. Ne lui reste plus alors qu'à parcourir son corpus et vérifier la

validité du conditionnel au sein de chaque texte de son corpus. Si c'est le cas, il peut conclure provisoirement que, dans la diachronie de la langue étudiée, « B apparaît parce que A ».

Grâce à la formulation de l'hypothèse de la diglossie en termes de zones zribi-hertziennes, on peut appliquer directement cette méthode. Tout d'abord, l'hypothèse de la diglossie se reformule par les conditionnels suivants :

- (2) a. Si une proposition contient une variante Z1, alors elle ne présente que des variantes Z1 ou Z2, donc aucune variante Z3.
- b. Si une proposition contient une variante Z2, alors elle peut présenter des variantes Z1, Z2, ou Z3, mais pas à la fois des variantes Z1 et des variantes Z3.
- c. Si une proposition contient une variante Z3, alors elle ne présente que des variantes Z2 et Z3, donc aucune variante Z1.

C'est-à-dire :

- (3) Une proposition ne présente pas à la fois des variantes Z1 et des variantes Z3.

Ensuite, on cherche bien à montrer une cohérence grammaticale, comme annoncé à la section précédente, ce qui est reformulé ainsi :

- (4)

Une proposition grammaticalement cohérente est une proposition qui ne présente pas à la fois une variante Z1 et une variante Z3.
--

Et enfin, cette exigence de cohérence grammaticale devra être constatée sur l'ensemble des productions d'un locuteur donné pour pouvoir conclure qu'il est diglosse.

Pour appliquer cette méthode, on peut s'y prendre de deux façons. La première possibilité, que je présente ci-après et que j'applique à la section 9.2, est de s'en remettre aux jugements de grammaticalité. La seconde, présentée à la section 9.1.2.3, consiste à s'en remettre à l'étude d'un corpus mono-locuteur.

9.1.2.2 Auto-observation du schéma diglossique

Pour se baser sur l'auto-observation pour vérifier le schéma diglossique décrit à l'encadré (4), il faut forger des exemples de mélanges de variantes prédits comme possibles. Ce sont les mélanges Z1-Z2 d'une part et Z2-Z3 d'autre part. Ils devront être jugés grammaticaux. Puis, il faut forger des exemples de mélanges de variantes prédits comme impossibles. Ce sont les mélanges Z1-Z3. On prédit alors qu'ils seront jugés agrammaticaux.

L'article de Zribi-Hertz (2006) se propose, entre autres choses, d'investir cette méthode. Zribi-Hertz y soumet par exemple les énoncés en (5), tirés de diverses études de syntaxe du français.

- (5) a. *Ce foutu médecin m'a encore charcuté la jambe, le salaud.* (Ruwet 1982)
b. *Il a fait casser la croûte à sa famille.* (Kayne 1975)
c. *Qu'ait téléphoné ton ami me surprend.* (Kayne et Pollock 2001)
d. *Quelle communication veux-tu qu'il ait ratée ? Il a assisté à toutes les séances.* (Obenauer 1994)

Dans ces exemples, Zribi-Hertz dénonce le mélange de variantes syntaxiques Z1 et d'un vocabulaire Z3. Elle en conclut que ces exemples sont fictifs, qu'ils n'existent pas. Puis, elle forge une série d'exemples pour montrer avec plus d'évidence les contrastes observables entre les mélanges Z1-Z2 ou Z2-Z3 d'une part et Z1-(Z2)-Z3 d'autre part. Je re-

produis ces exemples ci-dessous, avec leur typographie originale, qui sont les exemples (20) à (23) de Zribi-Hertz (2006:16).

(6) Syntaxe Z1, lexique Z2-Z3 :

- a. # *Que ne ferait-il* pour réussir, *ce salopard* ?
- b. # *Il fallait absolument que Jean pût* *casser la croute* au soleil.

(7) Syntaxe Z2, lexique Z2-Z3 :

- a. Qu'est qu'il *ne* ferait pas pour réussir, *ce salopard* .
- b. Il fallait absolument que Jean *puisse* *casser la croute* au soleil.

(8) Syntaxe Z3, lexique Z1-Z2 :

- a. # *En raison de* la chaleur, Pierre *a pas* fermé l'œil de la nuit.
- b. # *Pierre il* a fait ça *afin que* Marie guérisse.

(9) Syntaxe Z2, lexique Z2 :

- a. *À cause de* la chaleur, Pierre *n'* a pas fermé l'œil de la nuit.
- b. *Pierre a* fait ça *pour que* Marie guérisse.

Selon les jugements proposés par Zribi-Hertz, les prédictions sont respectées. Pourtant, il est difficile, il me semble, de juger ces exemples comme *agrammaticaux*. Zribi-Hertz elle-même ne les note que comme *inadéquats* (à travers le symbole #). Ceci est peut-être dû au fait que les mélanges proposés sont syntactico-lexicaux. Et puisque la cohérence recherchée est grammaticale, il serait préférable de forger des exemples mélangeant des variantes morphosyntaxiques.

Zribi-Hertz (2006:1-2) propose d'ailleurs un tel mélange. Elle remarque que « l'occurrence des clitiques nominatifs [. . .] est solidaire de l'effacement du *ne* de négation », ce qu'elle illustre par la paire (10) (son exemple (3)).

(10) a. *Jean ne viendra pas.*

Chapitre 9 Quoi chercher ?

b. *Jean il viendra pas.*

Selon Zribi-Hertz (d'après les exemples (6) à (9)), la négation sans *n* est Z3, et la négation avec *n* est Z2. D'après ces mêmes exemples, et selon ses termes, la non-occurrence du clitique sujet est Z2 et son occurrence est Z3. Donc tout mélange de ces variantes doit être grammatical (ce sont des mélanges Z2-Z3). Pourtant, son propos cité concernant l'exemple (10) prétend que (10c) est agrammatical. Il faudrait requalifier la négation avec *n* comme variante Z1 et non Z2.

(10) c. **Jean il ne viendra pas.*

Cette contradiction montre une difficulté majeure de la méthode par auto-observation : les intuitions et les jugements de Zribi-Hertz ne lui permettent pas de décrire avec une cohérence assurée ce qui ressortit à son FCT et ce qui ressortit à son FD.

Et, pour finir, puisque, selon Zribi-Hertz, la variante qui ne fait pas apparaître le clitique nominatif est Z2 (cf. (9)), (10d) est également un mélange Z2-Z3, donc prédit comme grammatical. Qu'en diraient les intuitions de Zribi-Hertz ?

(10) d. *Jean viendra pas.*

La section 10.3.3 permettra de proposer une réponse claire, et ne reposant pas sur des jugements de grammaticalité discutables, à la question des mélanges possibles des variantes que l'on vient d'observer, et donc d'établir clairement, pour le locuteur étudié dans le chapitre 10, l'appartenance de ces quatre variantes à Z1, Z2 ou Z3. Pour vendre la mèche, on peut annoncer que le locuteur étudié produit les mélanges (10abd) et pas le mélange (10c). Donc, chez lui, la négation avec *n* est Z1, et celle

sans n est Z3, tandis que la non-occurrence du clitique sujet est Z2, et son occurrence Z3.

Pour autant, je ne rejeterai pas complètement cette méthode. Même si elle demandera toujours une confirmation par la méthode empirique non-guidée, elle n'en reste pas moins un indicateur intéressant, et elle est pour l'instant le seul moyen que je connaisse pour déterminer l'appartenance des variantes rares aux zones zribi-hertziennes. À ce titre, la section 10.4.3.1 propose un éclairage sur les limites et les limitations de la méthode empirique. C'est pourquoi je consacrerai la section 9.2 à l'illustration d'une telle étude par auto-observation.

9.1.2.3 Observation sur corpus du schéma diglossique

Les jugements de grammaticalité peuvent faire défaut, c'est-à-dire ne pas voir une agrammaticalité prédite, voire prétendre à l'agrammaticalité d'un énoncé dont la structure est pourtant productive, sous une forme de déni fréquent lorsqu'il s'agit d'une langue ultra-normée comme le français. C'est pourquoi il est nécessaire d'en passer par l'étude d'un corpus spontané. Ce sera l'objet du chapitre 10.

Le principe sera le suivant. Parmi les productions spontanées d'un locuteur unique, on repèrera sa variation morphosyntaxique, en particulier lorsque les variantes produites sont marquées sociolinguistiquement. Ensuite, on étudiera chaque proposition qui contient deux variantes. Et à partir de l'observation des mélanges possibles face aux mélanges potentiels mais non réalisés, on classera chaque variante dans les zones zribi-hertziennes de façon à ce qu'elles respectent la cohérence énoncée à l'encadré (4).

Chaque occasion manquée de mélange potentiel sera l'occasion de dénoncer des variantes incompatibles (c'est-à-dire que l'une est Z1, et

l'autre Z3). Enfin, si la double grammaire du locuteur étudié est bien conforme au schéma diglossique, alors aucun des mélanges manqués ne devra contredire les conclusions tirées de l'observation d'un autre mélange manqué : chaque variante observée doit être univoquement dénonçable comme Z1 (respectivement Z2 ou Z3) tout au long des constats de mélanges manqués.

Pour le dire autrement, on remplace les jugements d'agrammaticalité par les 'trous' manifestes d'un corpus donné.

9.1.2.4 Qualitatif et quantitatif

L'hypothèse concerne la variation individuelle, donc elle est à tester parmi les productions d'un seul locuteur. Ce qui veut dire que chaque locuteur peut avoir des variantes différentes et différemment réparties, voire que certains locuteurs ne sont pas diglosses. Mais, l'hypothèse prédit que les locuteurs qui produisent à la fois des variantes purement normées et purement hors-norme ont une répartition Z1-Z2-Z3 consistante (c'est-à-dire que ces zones sont nécessaires et suffisantes pour décrire sa compétence linguistique).

La méthode ne met pas à l'abri d'un futur contre-exemple. Ce qui renforce les convictions acquises après une étude de corpus n'est donc pas de l'ordre du quantitatif. Au contraire, l'intérêt de la méthode par induction est de proposer des conclusions, qui, même si elles seront toujours provisoires, sont de l'ordre du qualitatif. On ne cherche pas à savoir si un locuteur donné est plus ou moins diglosse, on cherche à savoir si, dans les limites de ce qu'on a pu observer, ce locuteur est provisoirement descriptible comme un diglosse ou pas, sans nuance intermédiaire. Multiplier les observations ne rendra pas le locuteur en question encore plus diglosse, mais pourra renforcer la conviction de l'observateur que sa conclusion

provisoire vaut certitude.

La méthode a enfin la propriété épistémologique intéressante d'offrir naturellement les moyens de réfuter les hypothèses qu'elle se propose de tester.

Mais quoi faire concrètement si on trouve un contre-exemple ? En particulier, si ce contre-exemple est isolé, est-ce qu'on sera prêt à abandonner l'hypothèse testée ?

Si l'on choisit de ne pas abandonner l'hypothèse testée, il faut proposer d'interpréter le contre-exemple. En somme, il faut lui trouver une excuse. C'est-à-dire qu'il faudra justifier le contre-exemple en-dehors de l'hypothèse. Même si le corpus étudié au chapitre 10 ne présente aucun contre-exemple, je proposerai un cas précis au chapitre 11, en fondant l'explication sur la notion d'interférence.

Cette notion, explicitement avancée par Zribi-Hertz (2006:10) pour prévenir l'apparition d'un contre-exemple, suggère qu'un locuteur peut être amené à insérer des éléments de FCT alors même qu'il a activé sa grammaire du FD, par exemple parce qu'il se trouve dans une situation contradictoire, où il se force à produire du FCT malgré l'informalité de la situation, ou bien où sa production perd son caractère spontané pour une raison quelconque. Si une telle explication est plausible et fondée, elle permet de justifier l'interférence entre les deux grammaires du locuteur, et ainsi de pardonner le contre-exemple en le considérant comme en dehors des données à observer pour tester les hypothèses (c'est-à-dire en dehors d'une production orale spontanée et non-guidée ni auto-surveillée).

Mais il reste clair que les contre-exemples qui ne trouveraient pas d'excuse externe seraient dévastateurs. Il n'y a pas de locuteur presque diglosse.

9.2 Une cohérence grammaticale auto-observée

Dans cette section je m'interroge sur quatre variables de mon idiolecte. Chacune des quatre variables possède une variante plus formelle que l'autre. Chacune des sections ci-dessous avancera dans l'exploration des mélanges que je juge comme possibles et des mélanges que je ne pense pas produire. Est-ce que les résultats seront cohérents avec ma prétention à être diglosse ? On verra que oui, sauf que je serai amené à préciser que mes jugements de grammaticalité ne sont pas parfaitement assurés.

9.2.1 /pa{z}/ à /pA/

Dans mon idiolecte, on trouve deux variables qui concernent l'item négatif **PAS**. D'une part, je peux prononcer la voyelle antérieure ou postérieure, comme en (11).

- (11) la variable *A* :¹
- a. variante *A-1* : pá bzwě
 - b. variante *A-2* : pà bzwě

D'autre part, je peux réaliser une consonne de liaison [z] à la fin du mot, comme en (12).

- (12) la variable *Z* :

1. La variante *A-2* est peu répandue géographiquement. Dans mon idiolecte, elle est extrêmement restreinte, peut-être même seulement aux deux lexèmes **PAS** et **BAS**. Mais elle reste saillante phonétiquement, par exemple dans la paire *tapáløj* ~ *tapàløj*, où le premier élément de la paire correspond indifféremment à *tape-à-l'oeil* ou *t'as pas l'oeil* et où le second renvoie obligatoirement à *t'as pas l'oeil*.

9.2 Une cohérence grammaticale auto-observée

- a. variante $Z-1$: $\text{pa-z } \tilde{\text{a}}\text{kor}$
- b. variante $Z-2$: $\text{pa-}\emptyset \tilde{\text{a}}\text{kor}$

Les combinaisons logiquement possibles de ces deux variables (en contexte de liaison devant voyelle) sont donc au nombre de 4 :

- (13)
- a. $A-1 \times Z-1$: $\text{páz } \tilde{\text{a}}\text{kor}$
 - b. $A-1 \times Z-2$: $\text{pá } \tilde{\text{a}}\text{kor}$
 - c. $A-2 \times Z-1$: $\text{pàz } \tilde{\text{a}}\text{kor}$
 - d. $A-2 \times Z-2$: $\text{pà } \tilde{\text{a}}\text{kor}$

On pourrait formuler cette variation en termes de forme phonologique sous-spécifiée de l'entrée lexicale de **PAS**. La variable A est alors un phonème caractérisé comme voyelle basse mais reste non spécifié quant à sa place d'articulation (noté par exemple $/A/$), et s'oppose au phonème spécifié comme antérieur (noté $/a/$), telle que la voyelle du *nom* **PAS**. La variable Z , la liaison facultative, déjà mille fois traitée, peut être résumée par la notation $/\{z\}/$. Ce qui donne la forme phonologique de **PAS** en (14).

- (14) la forme phonologique sous-spécifiée de **PAS** qui génère ses quatre formes de surface logiquement possibles :

$/\text{pA}\{z\}/$

Or, il se trouve que je ne réalise jamais la combinaison $A-2 \times Z-1$. C'est-à-dire que l'exemple (13c) est agrammatical dans mon idiolecte. Les conséquences pour la description de mon entrée lexicale de **PAS** sont importantes : il faut désormais générer la variante $Z-1$ seulement avec la variante $A-1$, et réciproquement, c'est-à-dire la variante $A-2$ sans la variante $Z-1$.

Chapitre 9 Quoi chercher ?

- (15) a. $A-1 \times Z-1$: páz ãkor
b. $A-1 \times Z-2$: pá ãkor
c. $A-2 \times Z-1$: *pàz ãkor
d. $A-2 \times Z-2$: pà ãkor

Si on veut conserver la forme sous-spécifiée de (14), il faut par exemple proposer une règle qui efface $\{/z\}/$ lorsqu'il se trouve derrière $/A/$, comme en (16).

- (16) pas de liaison facultative après A :
A-SANS-Z : $\{z\} \rightarrow \emptyset / A_ \#$

Cette solution, ou toute autre qui souhaite conserver la forme sous-jacente unique, doit se résoudre à introduire une certaine idiosyncrasie, comme cette règle A-SANS-Z qui ne semble pas unifier des phénomènes plus généraux de la langue.

Je proposerai donc de rendre compte de la variation sur PAS en imaginant que l'ensemble des données de surface provient des réalisations possibles de deux entrées lexicales distinctes. D'un côté, la forme $/pa\{z\}/$ génère *páz* et *pá*. De l'autre, $/pA/$ génère *pà* et *pá*. Ainsi, on a préféré décrire, pour la sous-grammaire générant PAS, deux sous-grammaires distinctes, faisant appel chacune à une entrée lexicale distincte et une variable propre. Le fait que les deux 'mini-sous-grammaires' (17a) et (17b) génèrent possiblement la même forme *pá* n'empêche pas de préférer la solution à deux grammaires simples de (17) à la solution à une grammaire unique plus complexe de (14) et (16).

- (17) mes deux PAS :
- a. PAS_{FCT} : $/pa\{z\}/$ (contenant la variable Z)
b. PAS_{FD} : $/pA/$ (contenant la variable A)

9.2 Une cohérence grammaticale auto-observée

Cette représentation génère les bonnes données, et prévoit que la variante de zone Z3 A-2 ne doit pas être mélangée avec la variante de zone Z1 Z-1, ou toute autre variante Z1. Et réciproquement, que la variante de zone Z1 Z-1 ne doit pas être mélangée avec la variante de zone Z3 A-2, ou toute autre variante Z3. Ces prédictions seront vérifiées dans la section sub-suivante.

9.2.2 Et plus

Un avantage certain de décrire deux entrées lexicales pour **PAS**, c'est que cela offre un modèle tout prêt pour **PLUS**, qui lui aussi peut apparaître sous une forme phonétique qui n'est pas compatible avec Z-1. En effet, la forme *py*, variante de *ply*, ne combine pas avec Z-1 :

- (18) a. $L-1 \times Z-1$: *plyz isi*
b. $L-1 \times Z-2$: *ply isi*
c. $L-2 \times Z-1$: **pyz isi*
d. $L-2 \times Z-2$: *py isi*

(19) mes deux **PLUS** :

- a. **PLUS_{FCT}** : /ply{z}/ (contenant la variable Z)
b. **PLUS_{FD}** : /p(l)y/ (contenant la variable L)

9.2.3 pas, plus... et ne

9.2.3.1 $Z \times Neg$

On vient de discuter de deux variables présentes sur **PAS**, et une sur **PLUS**. Il en est une troisième qui leur est très liée, c'est la négation du verbe (*Neg*). La variante avec *n* sera notée *Neg-1*, et la variante sans *n* *Neg-2*.

Chapitre 9 Quoi chercher ?

Une étude systématique a été menée par Ågren (1973) sur la combinaison de la variable *Neg* et de la variable *Z* de **PAS**. Il en retient les faits suivants :

- Le « niveau de langue » influence la réalisation de la variable *Z* de **PAS** : la variante **paz** est nettement plus fréquente dans le niveau « soigné » que dans le niveau « courant » (tableau 9.1).
- « Dans mon corpus j'ai pu maintes fois remarquer que [l']absence de *ne* s'accompagne de la NL [non-liaison] de *pas*. Afin de vérifier quantitativement ce phénomène, j'ai calculé le nombre de cas, qui, dans une langue plus soutenue, auraient dû contenir un *ne*. [...] Voici le résultat du sondage :
 - Cas de *ne* omis et comportant la L [liaison] : 4=5%
 - Cas de *ne* omis et comportant la NL : 72=95%M'appuyant sur cette enquête, je puis donc prétendre que la corrélation entre l'absence de *ne* et la NL de *pas* se trouve entièrement prouvée. » (Ågren, 1973:102-103)
- Soit, répété dans mes termes, la variante *Neg-2* de la négation implique la variante *Z-2* sur **PAS**. D'où les deux sous-grammaires de (20).

	paz	pa
soigné	101	151
intermédiaire	85	185
courant	37	258

TABLEAU 9.1: La variable *Z* de **PAS** et le « niveau de langue », Ågren (1973:102)

9.2 Une cohérence grammaticale auto-observée

- (20) a. le PAS de la négation avec *n* contient la variable *Z* :
 $Neg-1 = n \vee pa\{z\}$
b. le PAS de la négation sans *n* ne contient pas la variable *Z* :
 $Neg-2 = \vee pa$

Se déduit de ces faits l'appartenance de *Neg-2* à la zone *Z3*, puisqu'elle est incompatible avec *Z-1*, elle-même variante *Z1*.

Et puisqu'il s'agit de mon idiolecte, et de mes jugements sur ces mélanges potentiels, je remarque que j'exclus quasi-formellement de produire (21).

- (21) *ʒ-e paz ãkor d ide

9.2.3.2 La difficulté de boucler la boucle

Les mélanges étudiés ci-dessus prédisent que *Neg-1*, dans le cas où elle est compatible avec *A-2*, variante *Z3*, est une variante *Z2*, puisqu'elle est également compatible avec *Z-1*, variante *Z1*. Et si c'est le cas, *Neg-1* doit aussi être compatible avec *L-2*, pour respecter la cohérence de mes deux grammaires. Si au contraire *Neg-1* est incompatible avec *A-2*, alors elle sera une variante *Z1*, et sera automatiquement incompatible avec *L-2*.

Si aucune des deux situations cohérentes ci-dessus n'est respectée, alors mes grammaires du FCT et du FD ne sont pas cohérentes, et il n'y a plus lieu de décrire ma compétence linguistique en termes de double grammaire. L'hypothèse de la diglossie serait rendue caduque. La question est donc : est-ce que je peux produire les mélanges de (22) ?

- (22) a. ??ʒ-n-e pà kōpri
b. ??ʒ-n-e py d ide

Comme je l'ai indiqué à l'aide de doubles points d'interrogation, il

m'est difficile d'être parfaitement formel quant à l'agrammaticalité de ces deux mélanges, tout en reconnaissant que je serais pour le moins surpris que l'on me surprenne à les produire. Il est également possible que l'aspect morpho-phonologique plutôt que morphosyntaxique des variantes étudiées ici joue un rôle dans ma petite hésitation. Mais je n'ai aucune hypothèse concrète à proposer.

Malgré ce léger doute, je propose une synthèse des résultats ci-après, en admettant que je ne puisse en effet pas produire (22) spontanément.

9.2.4 Synthèse

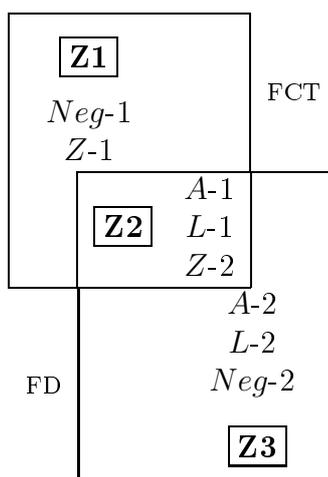


FIGURE 9.1: Les variantes de mon FCT et de mon FD, auto-observées

Dans le schéma 9.1, je synthétise les résultats obtenus pour observer que les zones zribi-hertziennes sont adéquates, c'est-à-dire nécessaires et suffisantes, pour décrire ma variation.

Et le tableau 9.2 propose une très succincte idée de ce que donnerait une modélisation de mes sous-grammaires du FCT et du FD telles qu'elles

9.2 Une cohérence grammaticale auto-observée

	FCT	FD
négation du verbe :	NE-V PAS/PLUS	V PAS/PLUS
PAS :	/pa{z}/	/pA/
PLUS :	/ply{z}/	/p(l)y/

TABLEAU 9.2: Les grammaires partielles de mon FCT et de mon FD

se dégagent de ces auto-observations.

Et puisque les jugements de grammaticalité posent les difficultés esquissées dans ce chapitre et à la section 2.4.2.2, et puisque la linguistique de corpus a depuis longtemps montré que les locuteurs sont peu fiables (par auto-conviction ou par déni), il est temps de passer à une étude de corpus.

Chapitre 10

Étude de corpus

C E CHAPITRE est une étude de corpus qui a pour but de soumettre concrètement l'hypothèse de l'exclusion des variantes Z1 et des variantes Z3 à des données empiriques, telle que cette hypothèse a été formulée au chapitre précédent, où elle a également été rapidement soumise à quelques données issues de l'auto-observation.

Dans un premier temps, je présente le corpus HH que j'ai constitué, ainsi que ses caractéristiques principales. Dans un second temps, on verra en détail les principaux faits de variation observables dans le corpus, et on en retiendra cinq variables.

Dans un troisième temps, on observera systématiquement les propositions produites par le locuteur étudié qui instancient deux variables au moins. On observera des absences de mélanges particuliers, et on en déduira l'appartenance de chaque variante à une des zones zribi-hertziennes.

Enfin, dans un quatrième temps, on montrera la compatibilité parfaite du modèle zribi-hertzien de la diglossie et du patron de variation du locuteur étudié, c'est-à-dire qu'il possède bien à la fois des variantes Z1, Z2 et Z3. Je serai conduit à mettre en avant la propriété qu'a la méthodologie utilisée de décrire des données négatives à partir de l'observation

de données positives, lesquelles données négatives sont bien la clef de l'hypothèse de la diglossie et de la formulation qui en a été faite dans ce travail. Et on conclura sur des remarques concernant les limites et les limitations de cette étude : le corpus étudié est trop petit pour pouvoir élargir les conclusions à des variables rares, et les variables étudiées ne permettent pas de définir clairement quelle est l'unité syntaxique dans laquelle les productions doivent rester grammaticalement cohérentes pour rentrer dans l'hypothèse de la diglossie.

10.1 Présentation du corpus HH

Le corpus présenté est une entrevue que j'ai menée personnellement avec un ami. Le thème principal de l'entrevue est la carrière rugbystique de mon interlocuteur, que je désignerai ci-après sous le nom de code UE. Aucun but autre que l'aspect sociologique de l'entretien ne lui a été dévoilé, et UE ne s'est pas douté que j'observerais en premier chef sa grammaire. Je pense très sincèrement avoir ainsi obtenu une production spontanée, et non-auto-surveillée.

Après avoir enregistré plusieurs entrevues avec des interlocuteurs différents, j'ai choisi de retenir HH pour plusieurs raisons. D'abord, UE y est spontané. Ensuite, HH est la plus longue des entrevues que j'ai menées (1h22mn), ce qui améliore l'informativité des conclusions à tirer sur la grammaire d'un seul locuteur. Enfin, malgré sa spontanéité, UE produit des variantes sociolinguistiquement variées, notamment des variantes classiques dans un discours que l'on percevra dans l'ensemble comme plutôt démotique, ce qui n'est pas le cas d'autres locuteurs que j'ai eus en entrevue¹. Un avantage de cette hétérogénéité, c'est qu'elle permet d'as-

1. Par exemple, sur 78 occurrences de *pa*, le locuteur IR ne produit jamais la va-

10.1 Présentation du corpus HH

sumer complètement l'hétérogénéité de la langue externe, même au sein de la production d'un même locuteur, et de donner à voir l'hétérogénéité de sa langue interne (c'est-à-dire sa double-grammaire, si tant est qu'elle présente un patron diglossique).

L'entretien a eu lieu en mars 2005, entre UE et moi-même uniquement, sur le thème du rugby. Dans le but de provoquer la production de variantes classiques, j'ai essayé de ponctuer mes interventions de variantes normées, par exemple en utilisant l'inversion dans les interrogatives. J'ai d'abord transcrit le corpus orthographiquement en alignant les énoncés sur le fichier son. Puis j'ai découpé les énoncés en propositions. Ensuite, j'ai anonymisé le corpus en remplaçant les prénoms par la lettre P suivie d'un chiffre, puis les noms de famille, avec la lettre N, les lieux, avec L, et les entités (par exemple les noms de clubs), avec E.

Ensuite, j'ai annoté les propositions en détail pour pouvoir automatiser les recherches : chaque proposition s'est vue affecter des valeurs, chaque fois que c'était pertinent, pour la personne du sujet, la négation du verbe, le type de proposition, le mot Qu- employé, le type de clivée, la présence d'un Topique, d'un Anti-Topique, ou d'un pronom fort détaché, le type de réalisation du sujet, et enfin la liaison facultative après l'item verbal *être* à la troisième personne.

Chaque exemple présenté ici a été retranscrit dans la transcription que j'ai adoptée pour cette thèse. Parfois, je me contente de reproduire la seule proposition de l'énoncé pertinente pour l'illustration de mon propos (notamment si cette proposition est indépendante), parfois je reproduis la proposition dans le contexte de l'énoncé dans son ensemble.

Les exemples sont identifiés par l'information entre parenthèses donnant l'identifiant du locuteur, puis le numéro de l'énoncé dans le corpus.

riante avec *n*.

Chapitre 10 Étude de corpus

Enfin, signalons que, souvent, j'ai laissé tomber certaines scories de la production orale spontanée (les répétitions, les abandons, les ruptures, les corrections, etc.) pour ne donner à voir que la proposition finalement obtenue et transmise par le locuteur (cf. section 3.3.2.3). Lorsque j'ai malgré tout donné à voir un énoncé interrompu, le lieu de la rupture syntaxique a été noté d'un +. L'exemple (23) illustre ces étapes, en anonymisant les noms propres immédiatement.

- (23) a. identification de l'énoncé :
[elaʒətõbsyɤlafɪʃdəuynafɪʃdəujavekomãisapel
avɛkN3dəsyP3N3kiʒuakikieʒwœɛdəL5dəL6]
- b. alignement et transcription orthographique :
début : 3mn36s2/10 ; fin : 3mn55s7/10 ; locuteur : UE
*et là je tombe sur l' affiche de où une affiche où y avait comment
il s' appelle avec N3 dessus P3 N3 qui joue à qui qui est un
joueur de L5 de L6*
- c. découpage en propositions :
1 : *et là je tombe sur l' affiche de où une affiche*
2 : *où y avait avec N3 dessus P3 N3*
3 : *comment il s' appelle*
4 : *qui joue à qui qui est un joueur de L5 de L6*
- d. annotation des propositions (exemple de la quatrième proposition) :
personne : 3sg ; type : subordonnée relative ; mot Qu- : *qui* ;
réalisation du sujet : pronom relatif (*qui*) ; liaison facultative
après être : non-réalisée ; pas de négation ; pas de clivée ; pas
de Topique ; pas d'Anti-Topique ; pas de pronom fort détaché.
- e. illustration de la variante non-réalisée de la liaison facultative
après être :

e la ʒ-tõb syr yn afif u i-ave P3 N3 ki e ã ʒuør d L6 (UE : 41)

Cet exemple illustre également la difficulté de choisir entre deux découpages possibles en propositions. En effet, on aurait pu choisir de ne voir que trois propositions, en comprenant que le locuteur a abandonné la relative en où au profit d'un syntagme prépositionnel en avec. Ce qui aurait donné (24).

- (24) a. 1 : et là je tombe sur l' affiche de où une affiche où y avait avec
 N3 dessus P3 N3
 2 : comment il s' appelle
 3 : qui joue à qui qui est un joueur de L5 de L6
 b. e la ʒ-tõb syr yn afif avec P3 N3 dsy ki e ã ʒuør d L6 (UE : 41)

J'ai retenu 627 énoncés de UE, découpés au total en 1418 propositions.

Au cours de ma présentation, je proposerai des chiffres. Pourtant, il ne sera pas question de tenir un quelconque propos quantitatif, mais bien un propos qualitatif, comme cela a été annoncé et justifié à la section 9.1.2.4. C'est pourquoi il n'apparaîtra aucune statistique.

10.2 Les variables retenues

Pour cette étude de corpus, il fallait retenir des variables présentant au moins deux caractéristiques. D'une part, le locuteur étudié doit les produire dans son discours, et d'autre part, elles doivent avoir une chance d'être interprétables en termes de double-grammaire.

Les variables qui tombent dans ce cadre sont la négation du verbe (*Neg*), la première personne du pluriel (*1pl*), le *il* explétif impersonnel (*Expl*), le sujet lexical et ses alternatives (le Topique, l'Anti-Topique, les

clivées en avoir) (*SujLex*), et la liaison facultative en *t* après le verbe être (*Etre*).

Des variantes comme la liaison de pluriel à droite du nom n'apparaissent pas chez le locuteur étudié, tout comme les interrogatives à inversion.

10.2.1 La négation du verbe

La négation du verbe (*Neg*) a été présentée longuement au chapitre 7. Je me contenterai donc de présenter des exemples du corpus pour confirmer que UE produit bien la variante avec *n* (*Neg-1*) et la variante sans *n* (*Neg-2*).

J'exclus de l'étude les négations des expressions figées *ne serait-ce que* et *n'empêche que* puisqu'elles n'ont pas d'équivalent sans *n*, et les négations sans *n* non verbales, puisqu'elles n'ont pas d'équivalent avec *n*.

Avec *n*, UE produit 17 propositions avec *pa*, et 5 avec un autre item négatif, soit 22 cas de *Neg-1*.

(25) mem le premier diviziõ n-atire pa l atãsiõ k L3 atire a set epok (UE : 198)

(26) a. L3 n-a zame vremã y d vitrin (UE : 172)

b. s-ki ratrap la betiz s-e zystmã l grup ki fe k l mek n-e ply tu søl
(UE : 540)

c. z-n-prefer ni + (UE : 646)

Sans *n*, UE produit 140 propositions avec *pa*, et 40 avec un autre item négatif, soit 180 cas de *Neg-2*.

(27) a. me o nivo d la vitrin sa marje pa (UE : 177)

b. i-a pa d rezõ (UE : 626)

c. e mem P13 ave pa syporte otã k mwa (UE : 339)

- (28) a. ʒ-m-suviẽ mem ply dy tu d s-ki s-e pase (UE : 207)
 b. il-ave di avã j-ore ʒame pãse k + (UE : 530)
 c. ʒ-konese okẽn arab (UE : 531)
 d. me biẽ u pa vwala sa osi s-e ni vre ni fo (UE : 623)

Enfin, les cas ambigus où *on* précède un mot à initiale vocalique, au nombre de 6, ont été exclus.

- (29) õ per ã kar d final ẽ matʃ k {õn,õ-n}-ore ʒame dy perdr (UE : 413)

10.2.2 La première personne du pluriel

La première personne du pluriel (*1pl*) est réalisée par UE avec le pronom *on* et le verbe à la troisième personne du singulier dans 103 cas (variante *1pl-2*).

- (30) e õn-a diskyte a-pø-pre yn ør (UE : 65)

Dans deux autres cas (variante *1pl-1*), réalisés au cours du même énoncé, la première personne du pluriel est réalisée avec l'accord en *+õ* du verbe, mais sans le sujet pronominal *nu* : une fois, le sujet est le syntagme nominal complexe *mwa e sertẽ*, l'autre fois, le sujet est réalisé avec le pronom relatif *ki*.

- (31) ʒ-se k i-a y enormemã d rsãtimã kã L3 a termine kõtr dez ãtrenør e k
 mwa e sertẽ n-kõprniõ pa e ki n-kõprniõ pa parsk õ{n,-n}-ave pøtetr
 pa le mem vy (UE : 469)

10.2.3 Le clitique sujet impersonnel

Dans le corpus, deux verbes impersonnels présentent deux variantes : l'une avec un clitique sujet explétif (variante *Expl-1*), l'autre sans (variante *Expl-2*). Il s'agit du verbe *falloir* (9 cas avec explétif, 6 cas sans) et de la locution existentielle *il y a* (1 cas avec explétif, 102 cas sans).

- (32) a. i-fo vwar k ue L3 ave ěn enorm eko dā la reziõ (UE : 167)
b. fo pa reflefir (UE : 463)
- (33) a. i-ave le rākōtr ěter-akademik k il-n-i-a pa la mētnā (UE : 263)
b. i-a okē privelez d l ě syr l otr (UE : 669)

10.2.4 Le sujet lexical

Lorsque le référent du sujet de la proposition doit être exprimé lexicalement, UE produit des constructions SV(O) (variantes *SujLex-1*) et des alternatives à ce SV(O) (variantes *SujLex-2*).² Ces alternatives sont le Topique et l'Anti-Topique, ainsi que les clivées en *avoir*.

UE produit 122 constructions SV(O), exemplifiées en (34). Un aspect intéressant de ces constructions SV(O) est le fait que l'éloignement entre le sujet et le verbe ne semble pas être un problème. La relative *ki õ fe yn kariat apre* en (34b), et en (34c) l'abandon du syntagme verbal *õ vuly par ābisiõ personel* mélangé à l'incise *s ki e pa dy tu kōdanabl 3-vø dir* n'empêchent pas le verbe d'apparaître sans clitique sujet (*sort* et *õ y*). Les autres exemples montrent la variété des sujets possibles dans ces constructions SV(O).

2. Je reprends la terminologie établie à la section 8.2.1 pour ce qui concernera la variable *SujLex*.

- (34) a. SN défini singulier :
 [l rygbi]_S ete partu prezã (UE : 92)
- b. SN quantifié de la forme *Q de N* :
 fẽ [buko d me koleg dajør ki õ fe yn karier apre]_S sort d s kartie
 (UE : 99)
- c. SN défini pluriel :
 L3 3-e vremã veki ãn aspe tre politik dy tryk me vremã s-e la u õ-
 vwa ãn aspe ekstremã politik pars-k [le dirizã]_S õ vuly par ãbisiõ
 personel s ki e pa dy tu kãdanabl 3-vø dir par ãbisiõ personel õ a
 traver l rygbi y yn ãbisiõ personel ki s-e mi o grã zur a la fẽ biẽ
 syr o momã de kriz tuzur ki õ ekøre sertẽ 3uør d L1 ki ete dã set
 ytopi (UE : 468)
- d. nom propre :
 [L2]_S pas a kẽz (UE : 113)
- e. nom propre dans une subordonnée :
 ue kã [L2]_S a arete s-ete kelkfoz kwa (UE : 124)
- f. pronom indéfini :
 [sertẽ ki ave dy talã e ki õ pa ete akseptẽ dã l grup]_S sõ parti dy
 L2 (UE : 134)
- g. SN indéfini pluriel :
 [plysiør suz-ẽdividualite]_S ã fõ yn ply grãd (UE : 497)
- h. SN indéfini singulier :
 [yn ekol d rygbi]_S s kre (UE : 153)
- i. SN quantifié universellement :
 [tu l kartie]_S e vny (UE : 335)
- j. pronom *tout* :
 [tut]_S e parti d la (UE : 165)
- k. SN indéfini non-spécifique :
 k [de fij]_S s-ẽteres o rygbi 3-truv sa biẽ (UE : 622)

Chapitre 10 Étude de corpus

Parmi ces constructions SV(O), j'ai choisi de compter les constructions VS, au nombre de 4, les constructions avec un pronom fort en position de sujet (4 cas), et les constructions à sujet ellipsé (6 cas). Ce qu'elles ont en commun et qui m'incite à les classer avec les constructions SV(O), c'est le fait qu'elles auraient pu avoir un clitique sujet ou être formulées par une clivée. Les exemples (34)' à (37)' proposent de telles alternatives.

- (35) a. bõ d L3 e sorti P7 N6 (UE : 160)
b. l ätrenmã k nuz-ave done N ete vremã tre tre tre pwëty (UE : 433)
- (36) a. e lyi e parti (UE : 130)
b. ø krwaje a ë dezëteresmã me äfë ø-mem n-l-ave pa (UE : 471)
- (37) e lyi e parti dõk a rãkõtre sõ per e parti zue a L3 (UE : 130)
- (34)' a. l rygbi il-ete partu prezã
b. i-a buko d me koleg dajør ki õ fe yn karier apre ki sort d s kartie
c. s-e la u õ-vwa ën aspe ekstremã politik pars-k le dirizã iz-õ y yn äbisiõ personel a traver l rygbi
- (35)' a. i-a P7 N6 ki e sorti d L3
b. l ätrenmã k N i-nuz-ave done il-ete vremã tre tre tre pwëty
- (36)' a. e (lyi) il-e parti
b. (ø) i-krwaje a ë dezëteresmã me äfë ø-mem il-n-l-ave pa
- (37)' e (lyi) il-e parti dõk il-a rãkõtre sõ per il-e parti zue a L3

En face de ce groupe, on trouve deux groupes, dont la propriété commune est d'être des alternatives au sujet lexical, comme je viens de l'illustrer. Le premier est celui des Topiques et des Anti-Topiques, le second est celui des clivées en *avoir*.

On trouve des Topiques et des Anti-Topiques de natures syntaxiques

différentes. Parmi les Topiques, on trouve des syntagmes nominaux, coïncidés avec un clitique animé (masculin ou féminin, voire *õ*), comme en (38), ou coïncidés avec le clitique neutre *s/sa*, comme en (39).

(38) SN Topique, clitique sujet animé (6 cas) :

- a. evidamã [l mek]_T il-e dã s kote zystmã mek fij (UE : 636)
- b. [l trez]_T s k il-a s-e k il-e alor ply fizik (UE : 697)
- c. même un SN quantifié :
e ã katr-vẽ-nøf [tu l rest]_T õ-lez-a r3wẽ (UE : 140)

(39) SN Topique, clitique sujet neutre (26 cas)

- a. un nom propre :
e dõk [L2]_T s-ete ã syperb ãn enorm kløb avek de problem (UE : 102)
- b. même dans une subordonnée :
s-ete yn grav kriz pur L2 parsk [L2 trez]_T s-ete ã tryk enorm (UE : 116)
- c. un SN pluriel défini :
e [le fãpiona yniversiter]_T s-ete yn grãd fet (UE : 267)
- d. un SN singulier défini :
kã Deleuze di [le rãkõtr]_T sa-s-fe pa avek de zã õ-s-ã-fu de zã (UE : 490)
- e. un grand SN défini :
me [l eksperias k mwa 3-e y ãn yniversiter par egzãpl a trez]_T s-e le ply bo suvnir d ma karier (UE : 566)
- f. un SN de groupe :
[l rest]_T s-ete de futbolør de sprintør (UE : 583)
- g. un SN indéfini (lecture générique) :
bõ apre õn-a di [. . .] [yn fij]_T sa-dwa pa s-abime sa-dwa pa tu sa (UE : 635)

Chapitre 10 Étude de corpus

Et on trouve des groupes infinitifs, introduits par la préposition *d*, comme en (40), et des propositions relatives nominales comme en (41).

(40) Groupe infinitif Topique, toujours introduit par *d* (3 cas)

- a. e [d arive ãn ãgløter kom sa]_T s-ete yn fraktyr (UE : 302)
- b. e [d le-vwar zue]_T s-ete ekstremã drol pask iz-ete dãz yn total liberte (UE : 584)

(41) Proposition relative nominale Topique (18 cas)

- a. pronom relatif sujet :
ãsyit dy L2 [s ki s-e pase]_T s-e k o L2 i-a y ã katr-vẽ-nøf yn disolysiö dy kløb ki pase a kěz a koz d problem politik (UE : 110)
- b. [s ki ratrap la betiz]_T s-e zystmã l grup ki fe k l mek finalmã n-e ply tu søl (UE : 540)
- c. pronom relatif objet :
me [s k i-fo rgarde]_T s-e s k i-fe kwa (UE : 549)

Un cas me semble inclassable, tout en paraissant bien être un cas de Topique :

(42) Inclassable (une proposition relative nominale manquée?) (1 cas)
[i-ave z-rprofe mwa si ty vø o kěz e ki fe part- inerã o kěz]_T s-e d tape de ku d pie (UE : 680)

Parmi les Anti-Topiques, on ne trouve pas de syntagmes nominaux coïncidés avec un clitique animé, mais seulement des syntagmes nominaux coïncidés avec le clitique neutre *s/sa*, comme en (43).

(43) SN Anti-Topique, clitique sujet neutre (5 cas)

- a. un SN singulier défini :

10.2 Les variables retenues

- s-ete P9 N7 [l ãtrenør]_{AT} (UE : 211)
- b. un nom propre :
e 3-m-syi di s-e pøtetr parsk s-e pa ã tryk tre tre vif [L18]_{AT} me
nõ (UE : 259)
- c. un SN pluriel défini :
e s-e sa [le vre rãkõtr]_{AT} (UE : 488)

On trouve également des groupes infinitifs, introduits par la préposition *d*, comme en (44), et une proposition relative nominale comme en (45).

- (44) Groupe infinitif Anti-Topique, toujours introduit par *d* (2 cas)
- a. e d la sa-s-e fe natyrelmã [d ale s-ěskrir o rygbi]_{AT}
- b. sa-m-ave fe bizar [d l-vwar]_{AT} (UE : 328)
- (45) Proposition relative nominale Anti-Topique (1 cas)
pronom relatif sujet :
s-ete pa grav [s k iz-õ fe]_{AT} (UE : 192)

Les autres constructions alternatives à SV(O) sont les clivées en *avwar*. On y trouve tout d'abord les clivées en *i-a*, comme en (46).

- (46) Clivée en *i-a* (20 cas)
- a. SN quantifié de la forme *Q de N* :
e i-a boku boku d zuør d mek ki sõ sorti d s kartie boku boku ki
õ zue apre (UE : 101)
- b. nom propre, *i-a* à l'imparfait :
e i-ave P6 N5 ki ete o lep d L9 (UE : 127)
- c. SN indéfini singulier, *i-a* au passé composé, dans une subordonnée complétive :
sa-ave pri de proporsiõ si biẽ k ty vva k a L3 i-a y yn ekol d rygbi

Chapitre 10 Étude de corpus

- ki gape tu (UE : 158)
- d. SN négatif de la forme *pas de N* :
i-a pa d vitrin ki s-e forme (UE : 181)
- e. SN indéfini singulier, cliticisé par *ã* :
e dõk l lãdmẽ i-ãn-a ã ki e vny m ferfe pur ale a l ãtrenmã (UE : 324)
- f. SN indéfini pluriel :
i-a de person ki tuf plys k d otr syrmã (UE : 506)
- g. pronom indéfini :
i-a kekfoz ki s-pas (UE : 596)

Ensuite, on trouve les clivées en *avwar* personnel, comme en (47).

- (47) Clivées en *avwar* personnel (3 cas)
- a. 2^e personne du singulier, SN quantifié de la forme *Q de N* :
t-a plẽ d zuør ki d kade a zynior õ ete rzete (UE : 133)
- b. 1^{re} personne du singulier, SN indéfini singulier :
z-e di si z-m-bles yn fwa z-rzu ply si z-e yn søl blesyr ki m-tõb dsy (UE : 381)
- c. 1^{re} personne du pluriel, SN indéfini singulier, *avwar* à l'imparfait :
õn-ave ã program d mysy a l ãter-sezõ ki nuz-ete done (UE : 430)

Enfin, on trouve un ensemble de clivées en *i-a*, avec la particularité qu'elles ont une deuxième partie formée d'une proposition indépendante et non d'une relative, ce que j'appellerai les clivées en *i-a* directes, comme on peut le voir en (48).

- (48) Clivées en *i-a* directes (3 cas)
- a. SN indéfini pluriel, cliticisé par *ã* :

- i-ãn-a iz-ete deza parti a L3 ãn ã avã (UE : 136)
- b. SN défini pluriel, *avwar* à l'imparfait :
i-ave le dirizã iz-õ eseje d fer ã tryk pur elarzir l kløb (UE : 150)
- c. SN indéfini pluriel, cliticisé par *ã* :
i-ãn-a i-kõdan mem l øvr ã mem tã k l mek kwa (UE : 552)

Toutes ces alternatives à SV(O) sont bien des variantes de ce dernier, puisqu'il est envisageable, dans chaque cas, de produire une construction SV(O) équivalente à ces alternatives. C'est ce que les exemples en (49) et (50) proposent.

(49) Constructions SV(O) équivalentes aux Topiques et Anti-Topiques observés :

- a. evidamã [l mek]_S e dã s kote zystmã mek fij
- b. bõ apre õn-a di [yn fij]_S dwa pa s-abime
- c. e [le-vwar zue]_S ete ekstremã drol pask iz-ete dãz yn total liberte
- d. [s ki ratrap la betiz]_S e zystmã l grup ki fe k l mek finalmã n-e ply tu søl
- e. [l ãtrenør]_S ete P9 N7
- f. [l-vwar]_S m-ave fe bizar
- g. [s k iz-õ fe]_S ete pa grav

(50) Constructions SV(O) équivalentes aux clivées en *avwar* :

- a. e [boku d zuør ki sõ sorti d s kartie]_S õ zue apre
- b. z-e di si [yn søl blesyr]_S m-tõb dsy z-rzu ply
- c. [sertẽ]_S kõdan l øvr ã mem tã k l mek

Dans cette perspective, les cas de Topiques et Anti-Topiques formés d'un groupe infinitif ou d'une proposition relative nominale ne représentent pas de véritables alternatives à SV(O) dans la mesure où, en ce

qui concerne UE et dans la limite du corpus étudié, ils ne s’opposent pas à des cas *effectivement observés* de SV(O) avec un groupe infinitif ou une proposition relative nominale en position de sujet. On ne les considèrera donc pas comme des instances de la variante *SujLex-2*. Ceci laisse un groupe de 66 cas d’alternative au sujet lexical.

Il me faut m’arrêter sur un dernier point important des variantes autour du sujet lexical. Le chapitre 8 a montré que les constructions de Topique et d’Anti-Topique et les clivées en *avoir* ne sont pas équivalentes pragmatiquement, ni entre elles, ni à la construction SV(O) d’Arrière-Plan. Mais elles représentent toutes des variantes possibles strictement équivalentes à une construction SV(O), dans la mesure où cette dernière serait la seule vraiment acceptée dans la norme pour remplir les fonctions pragmatiques attribuées à ces premières. Par ailleurs, les constructions SV(O) sont potentiellement des propositions d’Arrière-Plan, et dans ce cas elles sont à la fois des réalisations possibles de FCT et de FD. On pourra être amené à tenir compte de ce caractère de variante possiblement à la fois *SujLex-1* et *SujLex-2* lors de l’observation des mélanges de variables à la section 10.3.

10.2.5 Une liaison

Le corpus est très pauvre en liaisons facultatives. Une seule ressort assez nettement, c’est la liaison en *t* après l’item verbal *être*, que UE produit 7 fois (variante *Etre-1*), comme en (51), contre 63 fois où il ne la produit pas (variante *Etre-2*), comme en (52).

- (51) a. ʒ-syi ne l katorz novãbr si le detaj sõt ãportã mil-nøf-sã-swasãt-e-
õz (UE : 5)
b. i-ave P14 N11 ki et ã zuør dy L2 ki et ãternasional plẽ d fwa

(UE : 328)

- (52) a. s-e ã spor ki e ply difisil a zue k l trez (UE : 696)
 b. la diferã s ãtr l kãz e l trez el-e esãsielmã la z-tõb z liber (UE : 681)

10.3 Mélangeons

Les cinq variables retenues sont mélangeables, sauf *Expl* et *SujLex*, *Expl* et *Etre*, *Expl* et *1pl*. Pour illustrer ce que sont des mélanges de variables, je commencerai par le cas qui n'existe pas dans le corpus. On trouve dans le corpus 202 proposition avec une négation du verbe, c'est-à-dire 202 instanciations de la variable *Neg*, et 70 réalisations de l'item verbal *être* à la troisième personne du singulier devant un mot à initiale vocalique, c'est-à-dire 70 instanciations de la variable *Etre*. Et dans le corpus, aucune proposition n'instancie à la fois *Neg* et *Etre*. Et pour cause, la plupart des instanciations de *Neg* placent un item négatif généralement à initiale consonnantique, tels *pa* ou *ply*, immédiatement après le verbe, comme en (53).

- (53) s-e pa trez ãteresã (UE : 186)

Des cas tels (54a-d) sont par hasard totalement absents.

- (54) a. il-e apsolymã pa l mejør
 b. il-n-e apsolymã pa l mejør
 c. il-et apsolymã pa l mejør
 d. il-n-et apsolymã pa l mejør

Dans les six autres cas de mélanges possibles, on trouve des propositions qui instancient les deux variables à la fois. Chaque cas sera dé-

taillé ci-dessous, et à chaque fois on se demandera si les quatre instantiations possibles des deux variables, illustrées en (54), sont effectivement réalisées. Si ce n'est pas le cas, ce qu'on constatera à chaque fois, on conclura provisoirement, c'est-à-dire dans les limites du corpus, que certaines paires de variantes ne sont pas générables dans une même proposition, c'est-à-dire qu'il faut modéliser deux grammaires autonomes pour les produire.

Après avoir étudié chaque cas, on synthétisera les conclusions (section 10.4), et on vérifiera qu'elles sont bien consistantes, c'est-à-dire que les variantes trouvent toutes une place dans les zones zribi-hertziennes, lesquelles trois zones devront être nécessaires et suffisantes pour rendre compte du patron d'apparition des variantes étudiées sur ce corpus.

10.3.1 *Neg* × *1pl*

<i>Neg</i> × (6) <i>1pl</i>	<i>Neg-1</i> (22)	<i>Neg-2</i> (180)
<i>1pl-1</i> (2)	2	0
<i>1pl-2</i> (103)	0	4

TABLEAU 10.1: *Neg* × *1pl*

Le tableau 10.1 révèle que 6 propositions produites par UE instancient à la fois *Neg* et *1pl* (la négation et la première personne du pluriel). Par deux fois UE instancie *Neg-1* et *1pl-1*, en (55), et par 4 fois *Neg-2* et *1pl-2*, en (56).

- (55) ʒ-se k i-a y enormemã d rsãtimã kã L3 a termine kōtr dez ätrenør e k mwa e sertẽ n-kõprniõ pa e ki n-kõprniõ pa parsk õ{n,-n}-ave pøtetr

- pa le mem vy (UE : 469)
- (56) a. ð se pa trō (UE : 50)
- b. ð puve pa reste avek si pø d frik kwa (UE : 335)
- c. e mem si ð gaŋe pa boku a trez bẽ l arzã pase tre suvã o prmie
plã (UE : 574)
- d. ð gaŋe pa enormemã (UE : 574)

Aucune proposition n'instancie à la fois *Neg-1* et *1pl-2*, ni aucune *Neg-2* et *1pl-1*. La conclusion provisoire de ces observations est que *Neg-1* et *1pl-1* sont générées par la même grammaire, laquelle ne doit pas générer ni *Neg-2* ni *1pl-2*, qui sont générées par une autre grammaire. C'est-à-dire que *Neg-1* et *1pl-1* sont des variantes Z1, et *Neg-2* et *1pl-2* des variantes Z3.

À la lumière de cette conclusion, on est tenté de proposer de lever l'ambiguïté des exemples où ð précède un verbe à initiale vocalique. Puisque dans aucun des cas non-ambigus *1pl-2* n'apparaît avec *Neg-1*, on peut se décider à dire que ces cas dits ambigus sont aussi des mélanges de *1pl-2* et *Neg-2*. L'exemple proposé précédemment peut ainsi être réécrit sans ambiguïté comme (57).

- (57) ð-per ã kar d final ã matf k ðn-ore zame dy perdr (UE : 413)

10.3.2 *Neg* × *Expl*

L'observation détaillée des mélanges de *Neg* et *Expl* (la négation et l'explétif) indique qu'il est plus judicieux de séparer les cas où *Expl* est instanciée autour de la locution *il y a* et les cas où elle est instanciée autour du verbe *falloir*. On tentera une interprétation de la différence qui apparaît entre ces deux cas après les avoir détaillés ci-dessous.

<i>Neg</i> × (28) <i>(il)ya</i>	<i>Neg-1</i> (22)	<i>Neg-2</i> (180)
<i>(il)ya-1</i> (1)	1	0
<i>(il)ya-2</i> (100)	0	27

TABLEAU 10.2: *Neg* × *(il)ya*

Le tableau 10.2 révèle que 28 propositions produites par UEinstancient à la fois *Neg* et *(il)ya*. À une seule occasion, UE instancie *Neg-1* et *(il)ya-1*, en (58), et par 27 fois *Neg-2* et *(il)ya-2*, comme illustré en (59).

(58) i-ave le rākōtr ěter-akademik k il-n-i-a pa la mētnā (UE : 263)

(59) a. abityelmā ajør i-ave k de balō d fut (UE : 93)

b. i-ave ply tro d kløb vwazē (UE : 144)

c. i-ave pa d koeziō (UE : 243)

d. i-a pa d adekuasiō ātr l zeni e l om biē (UE : 546)

Aucune proposition n’instancie à la fois *Neg-1* et *(il)ya-2*, ni aucune *Neg-2* et *(il)ya-1*. La conclusion provisoire de ces observations est la même que pour *Neg* et *1pl*, c’est-à-dire que *Neg-1* et *(il)ya-1* sont des variantes Z1, et *Neg-2* et *(il)ya-2* des variantes Z3.

<i>Neg</i> × (4) <i>(il)faut</i>	<i>Neg-1</i> (22)	<i>Neg-2</i> (180)
<i>(il)faut-1</i> (9)	0	2
<i>(il)faut-2</i> (6)	0	2

TABLEAU 10.3: *Neg* × *(il)faut*

On voit dans le tableau 10.3 que 4 propositions produites par UE instancient à la fois *Neg* et *(il)faut*. Par deux fois, UE instancie *Neg-2* et *(il)faut-1*, en (60), et par 2 fois *Neg-2* et *(il)faut-2*, en (61).

- (60) a. i-fo pa krwar k s-e sĕkāt pursā d l ekip (UE : 536)
 b. s-e ni vre ni fo d dir le fij n-dwav pa el-dwav pa el-dwav pa i-fo
 pa s-e pa biĕ s-e ni vre ni fo (UE : 620)
- (61) a. fo pa reflejir (UE : 463)
 b. fo pa s-prādr pur de star (UE : 574)

(il)faut-1 et *(il)faut-2* sont instantiables avec *Neg-2*, donc doivent être toutes les deux générées par la grammaire qui génère *Neg-2*. Donc, puisque *Neg-2* est Z3, on peut conclure provisoirement qu'elles sont chacune soit Z2, soit Z3.

Pour ce qui est de l'instanciation de *Neg-1* avec *(il)faut*, le corpus induit des conclusions douloureuses. En effet, ces chiffres pris à la lettre, il faut conclure que *Neg-1* est incompatible avec *(il)faut*, ce qui revient à dire qu'il n'est pas de négation possible du verbe *falloir* en FCT, ce qui est facheux. Pour résoudre ce problème, je propose, entorse ponctuelle à la méthodologie de ce chapitre, de s'en remettre aux jugements de grammaticalité. En imaginant que l'absence d'instanciation commune de *Neg-1* et *(il)faut* est un hasard du corpus³, comme l'a été l'absence d'instanciation commune de *Neg* et *Etre*, on peut observer que nos jugements de grammaticalité révèlent qu'une instanciation commune de *Neg-1* et *(il)faut-1* est possible, alors que *Neg-1* n'est pas compatible avec *(il)faut-2*, ce qu'indique (62).

3. La plausibilité de ce hasard est confirmée par les chiffres : seules 2 instanciations de *Neg-2* sur les 180 produites le sont avec *(il)faut*. Il n'est donc pas surprenant qu'aucune des seulement 22 instanciations de *Neg-1* ne soit produite avec *(il)faut*.

(62) *(il)-n-fo pa reflejir

Selon ce scénario, si le corpus avait proposé des instanciations de *Neg-1* et *(il)faut*, elles auraient instancié *Neg-1* et *(il)faut-1*, et pas *Neg-1* et *(il)faut-2*.

Cette petite acrobatie permet la conclusion provisoire suivante : *(il)faut-1*, compatible avec *Neg-1* comme *Neg-2*, est une variante Z2, et *(il)faut-2* et *Neg-1*, incompatibles, sont respectivement une variante Z3 et une variante Z1.

On peut désormais revenir à la différence de comportement entre la locution *il y a* et le verbe *falloir* quant à la réalisation de l'explétif. Pourquoi est-ce que la variante produisant l'explétif est une fois une variante Z1, l'autre fois une variante Z2 ? Une première interprétation est d'imaginer que c'est le hasard du corpus que de ne pas fournir d'instanciation de *(il)ya-1* et *Neg-2*, laquelle instanciación amènerait à la conclusion que *(il)ya-1*, comme *(il)faut-1*, est une variante Z2. Mais on peut aussi remarquer que *(il)faut-1* est beaucoup plus répandue que *(il)ya-1* (9 cas sur 15, contre 1 cas sur 101). Ainsi, la diffusion lexicale de *Expl-1* (la production d'un explétif) n'est pas la même pour toutes les entrées lexicales : dans le FD de UE, *Expl-1* touche *falloir* et pas *il y a*.

10.3.3 *Neg* × *SujLex*

Neg et *SujLex* (la négation et le sujet exprimé lexicalement) se rencontrent à 24 reprises. On trouve 11 instanciaciones de *Neg-1* et *SujLex-1*. Les cas de SV(O) avec SN sujet sont les plus nombreux (8 cas) :

- (63) a. un nom propre :
L3 n-a zame vremã y d vitrin (UE : 172)
- b. un SN défini pluriel :

<i>Neg</i> × (24) <i>SujLex</i>	<i>Neg-1</i> (22)	<i>Neg-2</i> (180)
<i>SujLex-1</i> (117)	11	8
<i>SujLex-2</i> (66)	0	5

TABLEAU 10.4: *Neg* × *SujLex*

mem le primier diviziõ n-atire pa l atãsiõ k L3 atire a set epok

(UE : 198)

c. un SN défini singulier :

l diskur de matfo n-vo ply dã s tip la

(UE : 637)

Les 3 autres cas sont les suivants :

(64) a. un sujet ellipsé :

i-s-sõ rturte e õ rgarde k l kote politik dy tryk n-õ rtny k l kote politik dy tryk

(UE : 470)

b. un pronom fort :

ø krwaje a ã dezëteresmã me ãfë ø-mem n-l-ave pa

(UE : 471)

c. un pronom fort coordonné à un pronom indéfini :

z-se k i-a y enormemã d rsãtimã kã L3 a termine kõtr dez ãtrenør e k mwa e sertë n-kõprniõ pa e ki n-kõprniõ pa parsk õ{n,-n}-ave pøtetr pa le mem vy

(UE : 469)

On ne trouve pas de construction VS, ni de SN quantifiés ou indéfinis (en dehors du cas mixte de (64c)).

Une première remarque sera une remarque sur les chiffres : *Neg-1* apparaît dans l'ensemble seulement 22 fois sur 202. Et 11 de ces 22 fois sont des apparitions avec *SujLex-1*, qui sont toutes les apparitions conjointes

Chapitre 10 Étude de corpus

de *Neg-1* et *SujLex*. La corrélation empirique entre ces deux variantes est quantitativement frappante dans ce corpus.

Lorsque *Neg-2* rencontre *SujLex*, il s'agit de *SujLex-1* dans 8 cas, et de *SujLex-2* dans 5 cas. Je commence par détailler le groupe où *Neg-2* apparaît avec les alternatives au sujet lexical, qui sont les exemples suivants :

- (65) a. une clivée en *il y a*, négation sur le sujet clivé :
iz-õ rkryte plysiør zuør ki fe k i-a pa d vitrin ki s-e forme (UE : 181)
- b. un SN défini pluriel générique en Topique :
kã Deleuze di [le rãkõtr]_T sa-s-fe pa avek de zã õ-s-ã-fu de zã
(UE : 490)
- c. un SN défini singulier générique en Topique :
bõ apre õn-a di [. . .] [yn fij]_T sa-dwa pa s-abime sa-dwa pa tu sa
(UE : 635)
- d. une clivée en *il y a*, négation sur le sujet clivé (lui-même topicalisé) :
pur mwa l rygbi a kēz e l rygbi a trez i-ēn-a pa ě ki e mejør k l
otr
(UE : 670)

Puisque les alternatives au sujet lexical n'apparaissent pas avec *Neg-1*, la corrélation empirique entre *SujLex-2* et *Neg-2* est claire également.

Enfin, voici 5 des 8 cas où *Neg-2* apparaît avec *SujLex-1*.

- (66) a. pronom indéfini négatif :
person t-parle (UE : 305)
- b. nom propre :
e mem P13 ave pa syporte otã k mwa (UE : 339)
- c. SN défini pluriel :
me parã m-õ zame di fo rãtre a + (UE : 458)

- d. SN défini singulier générique :
 z-di pa k l arzã e riẽ (UE : 505)
- e. SN défini singulier :
 la fij dwa pa tuje a s spor (UE : 636)

Ces cas montrent principalement un point : puisque $SV(O)$ est compatible avec *Neg-2*, c'est qu'il est produit par la grammaire démotique de UE. Pourquoi est-ce que UE ne produit pas systématiquement une alternative à $SV(O)$ dans ces cas ? On peut proposer plusieurs interprétations. Tout d'abord, $SV(O)$, comme structure d'Arrière-Plan, c'est-à-dire à sujet non ou faiblement topical, n'est pas, dans les termes du chapitre 8, un équivalent pragmatique des alternatives *SujLex-2*. Il a donc sa place à leurs côtés. De ce point de vue, UE est un chaotique, tel que je l'ai discuté sous 8.4.3, qui produit à la fois $SV(O)$ et (T) pro-V (X) (AT) dans sa grammaire démotique. $SV(O)$ comme Arrière-Plan est donc une variante Z2. En revanche, lorsque $SV(O)$ et une alternative sont véritablement équivalents, le propos du chapitre 8 prédit qu'il s'agira d'une production de FCT.

Pour confirmer cela, il faut que toutes les constructions $SV(O)$ produites avec *Neg-2* soient des structures d'Arrière-Plan, puisque *Neg-2* est exclusivement démotique.

C'est ce que je vais suggérer en détaillant les exemples sous (66). Le cas de (66a) semble assez clair, puisque *person* n'a pas de référent, lequel ne peut donc être topical. En (66b), le sujet est un nom propre. Comme le remarque Lambrecht (1987:248), les noms propres ont la particularité d'avoir des propriétés les rapprochant des pronoms, ce qui suggère que les noms propres sont des sujets peu lexicaux, et (66b) serait à rapprocher d'une construction pro-V. (66c) a été produit dans le contexte discursif détaillé en (67).

- (67) a. **I grup_i** purkwa 3-evok **I grup_i** parsk bö 3-t-e di mwa d kel klas 3-sor 3-sor d klas proleter dök s-e yn kõsepsiõ socialist de tryk fě yn kõseption (UE : 448)
- b. spor ědividyel s-e marã 3-e fe dø spor (UE : 449)
- c. 3-e boku eme l rygbi me 3-e boku eme l tenis osi (UE : 450)
- d. fě 3-dire [**plě d sportif**]_S em plě d spor (UE : 452)
- e. me 3-e enormemã eme s kote ědividyalist dy tenis osi (UE : 453)
- f. me s-e l 3ø plys ki m-ěterese (UE : 454)
- g. o rygbi 3-syi ne deza tu-d-syt ĩ grup mwa (UE : 455)
- h. 3-e tuzur marfe par grup mwa kã 3-ete 3øn (UE : 456)
- i. 3-e pa y d edykasiõ rtire (UE : 457)
- j. 3-e tuzur ete dã + mwa [**me parã**]_S m-õ zame di fo rãtre a + ĩfě 3-e tuzur trene dã de grup kwa (UE : 458)
- k. e dök [**set espri d grup**]_{T_i} s_i-e ěn espri k 3-sor symã + ki e determine par tut ma vi e pyi par l eksperiãs k 3-e y dy rygbi ĩ grup tu-d-syt ĩfě i-a pa d foz kafe derier s-e 3yst + (UE : 459)

Le thème du discours est l'esprit de groupe du locuteur, annoncé par deux fois en (67a). Le fait que le locuteur est lui-même lié au thème du discours est marqué ici par l'emploi répété du pronom fort détaché *mwa* (67aghj), chose qu'il ne fait pas à d'autres moments de son discours. Le passage est ensuite construit ainsi. UE donne d'abord le contexte permettant d'aboutir ensuite à la conclusion en (67k) sur l'origine de son esprit de groupe. En donnant ce contexte, il dessine l'Arrière-Plan de son propos. À cette occasion, on remarque que les sujets dont le référent doit être exprimé lexicalement sont réalisés comme SV(O). C'est le cas de *plě d sportif* en (67d), et de *me parã* en (67j). Une fois ce contexte exposé, le locuteur reprend le thème de son discours, et puisque ce thème se trouve être le référent du sujet, il est réalisé explicitement comme Topique au

sein de sa proposition (*set espri d grup*, en (67k)). Donc, la construction SV(O) de (66c) est bien une structure d'Arrière-Plan.

Enfin, (67d) et (67e) sont également envisageables comme des sujets lexicaux faiblement topicaux.

- (68) a. ty-rākōtr ãn ãtrenør ty-e pa dāz ã rapor d puvwar t-e dā ã rapor
d kreasiõ d zarm kwa alor k n-sere-s k pur l oppoze avek ã bulo
ã bulo ty-va travaje pur gaje tõ pẽ t-e dā ã rapor_i trist avek tõ
patrõ (UE : 500)
[...]
- b. e l grup e pa + e i-s-pas de tryk ãfẽ sa-razut yn rifes ãkrwajabl
kwa kã s-e biẽ fe kã s-e biẽ fe d otã plys k s-e sãz arzã i-a pa d
arzã i-a pa tu sa (UE : 504)
- c. z-di pa k [l arzã]_s e riẽ me z-vø dir me vwala s-e ã rapor_i ki pure
etr ply rif k s k il-e (UE : 505)

Dans cet extrait, le thème du discours est le rapport avec un patron. Le sujet lexical / *arzã* en (68c) y est bien un élément de l'arrière-plan du discours, avec une faible topicalité, qui n'est là que pour justifier de la tristesse du rapport qu'on entretient avec un patron.

L'autre cas apparaît dans le contexte suivant :

- (69) evidamã l mek il-e dã s kote zystmãn mek *fij*_i [la *fij*]_{si} dwa pa tuje a
s spor (UE : 636)

Le sujet lexical *la fij* n'est pas référenciel : il reprend le nom nu *fij* qui est utilisé comme déterminatif, donc non actualisé, dans l'expression *s kote mek fij*. Puisqu'il n'est même pas référenciel, il n'y a même pas lieu de se demander si son référent est topical.

La conclusion provisoire que l'on tire de ce patron de mélanges est la

suiuante : *Neg-1* et *SujLex-2* s'excluent mutuellement, ce qui fait de la première une variante Z1 et de la seconde une variante Z3. Les sujets lexicaux topicaux, c'est-à-dire les cas de *SujLex-1* qui ne sont pas des structures d'Arrière-Plan, sont réalisés avec *Neg-1* exclusivement, donc il s'agit d'une variante Z1. Les sujets lexicaux non-topicaux sont réalisés soit avec *Neg-1* soit avec *Neg-2*, on a donc là une variante Z2.

10.3.4 *Etre* × *SujLex*

<i>Etre</i> × (15) <i>SujLex</i>	<i>Etre-1</i> (7)	<i>Etre-2</i> (63)
<i>SujLex-1</i> (117)	3	4
<i>SujLex-2</i> (66)	0	8

TABLEAU 10.5: *Etre* × *SujLex*

La variante *Etre-1* (la liaison en *t* réalisée après *être*) est réalisée exclusivement, et dans 3 cas, avec la variante *SujLex-1* (le sujet exprimé lexicalement sous la forme SV(O)) :

- (70) a. ʒ-syi ne l katorz novãbr si le detaj sõt ẽportã mil-nøf-sã-swasãt-e-õz (UE : 5)
 b. õ-dve s-rtruve ã dmi e ã dmi L26 et arive ã final (UE : 414)
 c. l espri y et ẽn espri tre tre trez ẽportã tre tre tre tre riʃ (UE : 588)

La variante *SujLex-1* est aussi compatible, dans 3 cas, avec *Etre-2* :

- (71) a. l rygbi a trez ete ãkor tre vivã dã l syd-uest (UE : 191)
 b. l ane d apre N7 ete ãtrenør d la pmier (UE : 220)

- c. o nivo d l ãtrenmã P9 N7 ete êkõpetã vremã êkõpetã (UE : 230)
 d. me L3 s-ete asosie o E4 a l epok (UE : 384)

Enfin, *SujLex-2* et *Etre-2* sont réalisables conjointement, ce qui se trouve à 8 reprises :

- (72) a. e dõk L2 s-ete ãn enorm kløb avek de problem (UE : 102)
 b. s-ete yn grav kriz pur L2 parsk L2 trez s-ete ã tryk enorm (UE : 116)
 c. e i-ave P6 N5 ki ete o lep d L9 (UE : 127)
 d. i-a d plyz-ã-plys d dirizã ki s-sõ êterese a set ekip ptit ekip zynior (UE : 148)
 e. alor l arive ãn ãglãter s-ete ã fok pur mwa tu zøn (UE : 299)
 f. l prmie matf s-ete o L2 (UE : 350)
 g. i-ave l ekip d L30 ki ete arive (UE : 580)
 h. la diferã s ãtr l kãz e l trez el-e esãsielmã la z-tõb z-liber (UE : 681)

La conclusion de ce patron est d'une part que *Etre-1* est une variante Z1 et *SujLex-2* une variante Z3, puisqu'elles s'excluent mutuellement, et d'autre part que *Etre-2* et *SujLex-1* sont des variantes Z2, puisqu'elles sont compatibles entre elles et que la première est compatible avec une variante Z3, et la seconde avec une variante Z1.

10.3.5 $1pl \times SujLex$

Il peut paraître curieux de vouloir observer le mélange de $1pl$ et *SujLex* (la première personne du pluriel et le sujet exprimé lexicalement), puisque la première personne du pluriel est rarement réalisée lexicalement. Cependant, ce n'est pas à exclure, comme l'exemple construit (73) le montre, et trois exemples du corpus méritent que l'on s'y attarde.

$1pl$ × (3) <i>SujLex</i>	$1pl-1$ (2)	$1pl-2$ (103)
<i>SujLex-1</i> (117)	1	0
<i>SujLex-2</i> (66)	0	2

TABLEAU 10.6: $1pl \times SujLex$

(73) *fabiẽ e mwa (nu-)partõ pur barselson*

On trouve tout d'abord un cas où $1pl-1$ et *SujLex-1* sont réalisés conjointement :

(74) *mwa e sertẽ n-kõprniõ pa* (UE : 469)

Et ensuite on trouve deux cas quiinstancient $1pl-2$ et *SujLex-2*⁴ :

(75) a. *e ã katr-vẽ-nøf tu l rest õ-lez-a r3wẽ* (UE : 140)

b. *o bu d kelk tã avek P13 õn-a deside d rãtre* (UE : 333)

Comme $1pl-1$ et *SujLex-2* s'excluent, la première est une variante Z1 et la seconde une variante Z3. Les corrélations entre $1pl-1$ et *SujLex-1* d'une part, et entre $1pl-2$ et *SujLex-2* d'autre part étant parfaites, *SujLex-1* est donc Z1 comme l'est $1pl-1$, et $1pl-2$ est Z3 comme l'est *SujLex-2*.

<i>1pl-2</i> × (5) <i>Etre</i>	<i>1pl-2</i> (103)
<i>Etre-1</i> (7)	0
<i>Etre-2</i> (63)	5

TABLEAU 10.7: *1pl-2* × *Etre*

10.3.6 *1pl-2* × *Etre*

1pl et *Etre* (la première personne du pluriel et la liaison en *t* après *être*) sont instanciables ensembles seulement dans le cas de *1pl-2*, puisque *1pl-1* (*nu-som*{*z*}) ne présente pas la liaison facultative en *t* de la troisième personne. Et les résultats révèlent un fait simple :

- (76) a. 3-se ply 3ysk u õn-e arive (UE : 226)
 b. õn-ete ale ã bys (UE : 282)
 c. la prmier ane õn-ete ãn elit d∅ (UE : 398)
 d. si õn-e arive ã final d kup s-e pa pur riẽ (UE : 423)
 e. le kẽzist dize k õn-ete ã spor d bagarør (UE : 668)

Si la variante *1pl-2* a la forme de l'indéfini (*õ* et le verbe à la forme de la troisième personne du singulier), elle n'est pas pour autant une réalisation de la troisième personne du singulier, mais bien de la première personne du pluriel, puisque, selon le tableau 10.7, elle est incompatible avec *Etre-1*. Il faut cependant nuancer cette conclusion, puisque les chiffres sont trop minces pour assurer qu'il ne s'agit pas d'un trou hasardeux du corpus.

4. La construction particulière de (75b), où le syntagme prépositionnel avec *X* réfère à *X et moi*, est discutée brièvement à la section 11.1

10.3.7 Le choix de la proposition comme unité de cohérence

Dans cette étude, j'ai choisi de vérifier la cohérence grammaticale du FCT et du FD de UE au sein de chacune des propositions qu'il produit. Ce n'est pas un choix théorique, mais un choix pratique fait a posteriori et guidé par l'observation des mélanges à l'échelle des énoncés. Ceci révèle clairement que UE produit assez facilement des variantes contradictoires au sein d'un même énoncé. Par exemple, en (77), UE instancie 7 fois *Neg*, parmi lesquelles une instanciation de *Neg-1* au milieu de 6 instanciations de *Neg-2*, et de même en (78) on trouve une fois *Neg-1* et une fois *Neg-2* dans un contexte strictement identique.

(77) s-e ni vre ni fo d dir le fij n-dwav pa el-dwav pa el-dwav pa i-fo pa s-e
pa biẽ s-e ni vre ni fo (UE : 620)

(78) si biẽ k vwala + a la fwa s problem ki n-m-ẽteres pa ki m-ẽteres ply
dy mwẽ ki a pri + (UE : 523)

De même, (79) instancie *1pl-1* et *1pl-2*.

(79) 3-se k i-a y enormemã d rsãtimã kã L3 a termine kõtr dez ãtrenør e k
mwa e sertẽ n-kõprniõ pa e ki n-kõprniõ pa parsk õn-ave pøtetr pa le
mem vy (UE : 469)

Et (80), lui, instancie *(il)ya-1* et *(il)ya-2* :

(80) i-ave le rãkõtr ẽter-akademik k il-n-i-a pa la mẽtnã (UE : 263)

La tentation de se restreindre à l'échelle de la proposition devient d'autant plus forte que chacune des propositions ci-dessus reste cohérente.

Ceci n'est pas très visible en observant une seule variable, puisque les variables étudiées n'apparaissent pas elles-mêmes plusieurs fois au sein de la même proposition. Mais cela devient frappant lorsqu'on regarde quelles autres variantes apparaissent dans chacune des propositions.

On voit alors que la proposition où *Neg-1* apparaît dans (77) instancie *SujLex-1*, et pas *SujLex-2*. De même, les deux propositions qui instancient *1pl-1* en (79) instancient *Neg-1*, et non *Neg-2*. On pourrait ajouter que la proposition qui instancie au contraire *1pl-2* instancie *Neg-2*, mais il faut s'en retenir puisque ce cas est ambigu entre les deux découpages indiqués en (81) (même si j'ai proposé un argument a posteriori à la section 10.3.1 pour privilégier cette interprétation).

(81) [õnave] \sim õn-ave ou bien õ-n-ave

Et enfin, l'instanciation de (*il*)*ya-1* dans la seconde proposition de (80) est couplée à l'apparition de *Neg-1* et non de *Neg-2*.

Constater l'incohérence fréquente des énoncés est douloureux au sein de l'hypothèse diglossique. Mais cette incohérence contraste très fortement avec la cohérence parfaite des propositions. En effet, la section 10.4.1 constatera que l'étude menée n'a repéré aucune proposition incohérente. C'est bien l'envie de trouver un domaine de cohérence grammaticale qui m'a poussé à restreindre cette idée de domaine à l'échelle de la proposition. Aucune intuition à ce sujet ne m'était venue avant de constater l'incohérence des énoncés. Et aucune position théorique a priori ne m'y a guidé. C'est presque simplement un hasard heureux qu'un tel domaine de cohérence existe. À la section 10.4.3.2 ci-dessous, je tenterai de cerner mieux cette problématique, et les questions théoriques qu'elle pose.

10.4 Synthèse

10.4.1 La cohérence du FCT et du FD de UE

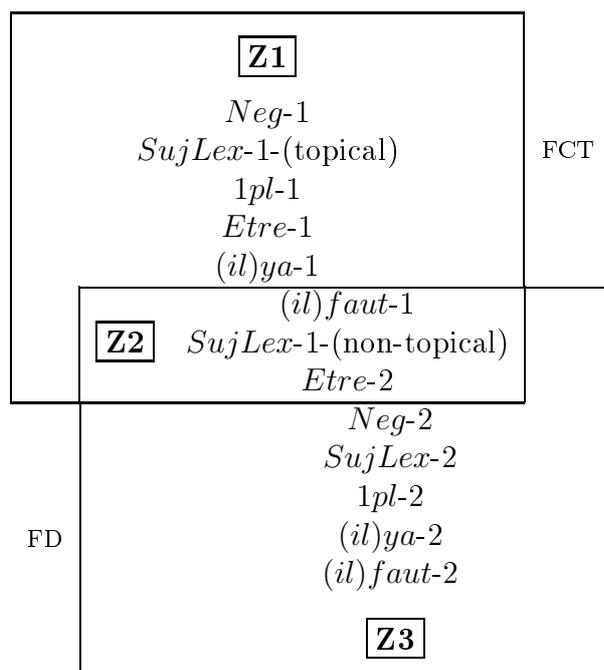


FIGURE 10.1: Le FCT et le FD de UE d'après le corpus HH

Le schéma 10.1 reprend l'ensemble des conclusions provisoires obtenues à la section précédente. On se rend compte que UE y a été consistant. C'est-à-dire :

- (82) La variation produite par UE autour des cinq variables étudiées et dans les limites du corpus HH suit un patron, observé au sein des propositions instanciant plusieurs variables, qui est compatible avec la représentation de l'hypothèse de la diglossie en termes de zones zribi-hertziennes.

Les mélanges observés, aussi peu nombreux soient-ils dans certains cas, sont chacun une occasion de contredire l'hypothèse. On remarquera, comme seule note quantifiée, que sur les 1418 propositions du corpus, on a pu observer 85 mélanges, parmi lesquels aucun n'a constitué un contre-exemple.

10.4.2 Des agrammaticalités révélées par l'étude

Si l'on assume jusqu'au bout la conclusion de cette étude, il faut conclure que les mélanges non-produits sont non-générables, donc qu'ils répondent à la définition d'un énoncé agrammatical. Ce qui donne les données suivantes à intégrer dans la modélisation du FCT comme du FD de UE :

- (83) a. *õ-n-se pa tro
b. *mwa e sertẽ kõprniõ pa
- (84) a. *i-ave le rãkõtr ãter-akademik k il-i-a pa la mëtnã
b. *n-i-ave pa d koeziõ
- (85) a. *n-fo pa reflefir
- (86) a. *yn fij sa-n-dwa pa s-abime
b. *l rygbi a kãz e l rygbi a trez il-n-i-ãn-a pa ã ki e mejør k l otr
- (87) a. *i-a d plyz-ã-plyz d dirizã ki s-sõt ãterese a set ekip
b. *la diferã s ãtr l kãz e l trez el-et esãsielmã la z-tõb z-liber

Chapitre 10 Étude de corpus

- (88) a. *mwa e sertẽ nu-n-kõprniõ pa
b. *filip e mwa a deside d rãtre
- (89) a. *si õn-et arive ã final d kup s-e pa pur riẽ

Cette conséquence n'est pas anodine puisque nombre des exemples ci-dessus ne sont pas facilement jugés agrammaticaux par les locuteurs. La méthodologie employée dans ce chapitre a donc ceci d'intéressant qu'elle propose des données négatives à partir de la pure observation de données positives, au-delà même des intuitions des locuteurs.

Ensuite, encore moins anodin, il faut supprimer de la modélisation du FCT de UE les variantes Z3, et de celle de son FD les variantes Z1. C'est-à-dire qu'un énoncé grammatical de FD peut être agrammatical en FCT, et inversement. Là également, la méthodologie propose des données négatives qui ne peuvent pas être jugées agrammaticales en dehors de l'acrobatie qui consiste à se demander si un énoncé est « bien formé dans un contexte formel/informel ».

- (90) Quelques données négatives du FCT de UE :
- a. *evidamã l mek il-e dã s kote zystmã mek fij (UE : 636)
 - b. *z-konese okẽn arab (UE : 531)
 - c. *e õn-a diskyte a-pø-pre yn ør (UE : 65)
 - d. *fo pa reflefir (UE : 463)
 - e. *i-a okẽ privilez d l ã syr l otr (UE : 669)
 - f. *e s-e sa le vre rãkõtr (UE : 488)
 - g. *i-a kekfoz ki s-pas (UE : 596)
- (91) Quelques données négatives du FD de UE :
- a. *mem le premier diviziõ n-atire pa l atãsiõ k L3 atire a set epok (UE : 198)

- b. *mwa e sertẽ n-kõprniõ pa (UE : 469)
- c. *(le rãkõtr ãter-akademik) k il-n-i-a pa la mẽtnã (UE : 263)
- d. *z-syi ne l katorz novãbr si le detaj sõt ãportã mil-nøf-sã-swasãt-e-õz (UE : 5)

10.4.3 Quelques questions ouvertes

10.4.3.1 HH, un ‘petit’ corpus

Pour des raisons de temps, mais aussi de contraintes sur le type de corpus (spontané, donc oral, et mono-locuteur), HH ne fait que 10000 mots environ. Ceci limite drastiquement la possibilité d’observer la variation à grande échelle. On voit bien que seules des variables structurales et génériques sont apparues en nombre suffisant pour permettre des conclusions qualitatives. Toutes les variables liées à un item lexical particulier sont redevables de la fréquence d’apparition de cet item. Or, en 10000 mots, très peu d’items lexicaux sont suffisamment fréquents. Les trois seuls items lexicaux qui ont pu jouer un rôle direct dans l’étude sont *être*, *il y a* et *falloir*.

Pour espérer étudier la variation de (92), liée à l’apparition du verbe *continuer*, il faut se placer à une échelle bien supérieure.

- (92) a. z-e kõtinye d zue
 b. z-e kõtinye a zue

Le corpus des déclarations de François Mitterrand⁵ est d’une taille qui permettrait une telle étude puisque, en 4 492 076 mots, la variable de (92) y apparaît 762 fois, 560 fois sous la variante (92a) et 202 fois sous

5. Ce corpus, parmi d’autres, est consultable et interrogeable grâce à l’outil KWIC développé par Achim Stein, et mis en ligne à l’adresse <http://inferno.philosophie.uni-stuttgart.de/twic/twic.html>.

la variante (92b)⁶. Hélas, bien que mono-locuteur et oral, ce corpus n'est pas spontané (même si les interventions ne sont pas lues, elles sont probablement sujettes à une auto-surveillance forte), et par ailleurs ne présente que peu de variation de registre⁷.

Ceci m'amène à formuler la difficulté suivante : est-ce qu'il est vraiment envisageable de constituer un corpus d'oral spontané mono-locuteur de l'ordre du million de mots ? Sinon, à supposer que l'on réunisse 100 corpus équivalents à HH, est-ce qu'un tel corpus multi-locuteur pourrait y pallier ? Et quelle méthodologie serait nécessaire pour le traiter sereinement ? Enfin, est-ce qu'il existe des outils qui permettraient de repérer automatiquement toutes les occurrences de chacune des variables étudiées ici, c'est-à-dire sans nécessiter un parcours manuel de l'ensemble du corpus, chose que je ne suis pas parvenu à éviter ? Il ne serait en effet pas envisageable de chercher à la main tous les Topiques ou tous les explétifs dans un million de mots.

10.4.3.2 Le domaine de portée des variables

Toutes les variables prises en compte pour cette étude sont liées au verbe. Non pas que ce soit un choix, mais simplement, en parcourant le corpus, aucune autre ne m'est apparue. On aurait pu imaginer trouver les deux variantes de (93), liées au domaine nominal.

- (93) a. le fosyr d mō frer
b. le fosyr a mō frer

6. La requête KWIC utilisée est :

lemma=continuer lemma=.*{0,3} lemma=(à|de) pos=VER.*

7. Par exemple, un coup d'œil rapide sur les 100 premiers résultats (sur 36533) de la recherche du patron V + pas n'a relevé aucune occurrence sans *ne*. Il ne serait d'ailleurs pas surprenant que les *ne* 'manquants' aient été ajoutés lors de la transcription, tout comme de nombreuses autres 'corrections'.

Quel patron de variation est-ce qu'on peut s'attendre à observer en mélangeant cette variable et une autre variable nominale, comme (94) ?

- (94) a. nyl fosyr
b. okyn fosyr

Intuitivement, (93a) est Z2, et (93b) est Z3. Et (94a) est Z1 et (94b) Z2. Quels mélanges parmi (95abcd) seraient empiriquement observables ? Lesquels ne le seraient pas ? Est-ce que le patron serait bien celui prévu et suggéré en (95) ?

- (95) a. nyl fosyr d mō frer
b. okyn fosyr d mō frer
c. okyn fosyr a mō frer
d. *nyl fosyr a mō frer

Et aussi, quoi prédire quand à l'association d'une variable nominale et d'une variable verbale ? Si le domaine syntaxique touché par les variables verbales est bien la proposition, est-ce que le domaine syntaxique touché par les variables nominales est le syntagme nominal ? C'est-à-dire, est-ce que l'emploi d'une variante Z1 (respectivement Z3) dans un SN est bien empiriquement incompatible avec une variante verbale Z3 (respectivement Z1) dans la même proposition ? Est-ce que cela dépendrait du statut argumental de ce SN dans la proposition ? De même, est-ce que deux SN distincts de la même proposition peuvent faire des choix de variantes contradictoires ?

Quid, donc, des exemples en (96) ?

- (96) a. ma mer n-em okyn fosyr a mō per
b. ma mer el-em nyl fosyr d mō per

- c. i-a nyl fosyr d mǔ per ki ple a ma mer
- d. le fosyr a mǔ per n-plez pa a ma mer
- e. dǎ lez ami a mǔ frer dø n-em nyl fosyr d mǔ per
- f. dǎ la mezǔ a mǔ frer nyl fosyr n-et otorize
- g. nyl fosyr n-dekor la mezǔ a mǔ frer

On pourrait imaginer, si des mélanges contradictoires sont possibles (mes intuitions personnelles vont dans ce sens pour certains des exemples ci-dessus, (96adfg) notamment), que cela suggère de décrire une grammaire du domaine nominale et une grammaire du domaine verbal. Or, cette concession contredit pleinement l'hypothèse de la cohérence *globale* de chacune des grammaires du FCT et du FD.

Ou bien, si l'on trouve un contraste entre les SN argumentaux et les SN hors de la portée du domaine verbale (c'est-à-dire si par exemple (96g) et (96f) contrastent), il faut redéfinir le domaine de portée de l'exigence de consistance des énoncés : il ne s'agirait plus de la proposition (avec ses adjoints) mais de la proposition sans ses adjoints d'une part, et des adjoints de la proposition d'autre part. Je m'empresse, pour finir, de rappeler que je n'avance ici aucune théorie suggérant de réponse à cette problématique.

10.4.4 Conclusion

Pour conclure ce chapitre, je dirai deux choses. D'abord, l'étude de corpus menée est, à l'intérieur de ses propres limites et limitations, très convaincante, et sa méthodologie est prometteuse. *UE y est bien un diglosse*. Ensuite, malgré cette conviction, et peut-être la sensation que toute étude comparable confirmerait les résultats, bien des questions restent ouvertes et invitent à élargir l'observation de patrons de variation

10.4 Synthèse

au domaine nominal puis à formuler une hypothèse, et proposer une méthodologie pour la tester, sur la portée de l'exigence de cohérence grammaticale des productions des locuteurs diglosses FCT-FD.

Chapitre 11

Clashes

L'HYPOTHÈSE diglossique et le modèle que j'en défends prédisent l'agrammaticalité d'énoncés (plus précisément de propositions) dont chaque partie, individuellement, est grammaticale. Cette agrammaticalité viendrait du fait que les parties de l'énoncé sont générées dans des grammaires distinctes, dans des zones hors de l'intersection de ces grammaires, c'est-à-dire les zones Z1 et Z3 du modèle.

Les deux chapitres précédents ont confirmé ces prédictions. Pourtant, on peut observer certaines données qui contreviennent à l'idée qu'un locuteur n'active qu'une grammaire à la fois. Dans ce très bref chapitre, on en verra un exemple. Je proposerai de trouver une interprétation qui permette d'y voir le résultat d'une interférence entre le FCT et le FD, en essayant de justifier qu'une telle interférence puisse intervenir dans le cas étudié.

11.1 Un flagrant délit de mélange interdit

(97) a. * *nous partons pour barcelone avec fabien pour 4 jours à partir de demain.*

=‘Fabien et moi partons pour Barcelone [...]’

Dans l'exemple (97)¹, on trouve au moins deux choix de variante associés à des valeurs sociolinguistiques. Tout d'abord, on trouve la variante *1pl-1*, réputée Z1, de la variable *1pl* (ce que confirme mon intuition sur l'idiolecte de la locutrice, que je connais personnellement, et ce qui a été conclu concernant UE au chapitre précédent). La variante *1pl-2* aurait donné (97b) dans ce cas.

(97) b. ð-par pur barsə̀lòn avek fabiẽ pur katr zur a partir d dmẽ

Ensuite, on trouve une autre variante marquée sociolinguistiquement, mais cette fois-ci comme Z3. Il s'agit de la construction où l'association du locuteur et d'un tiers exprimée lexicalement pour former la première personne du pluriel est obtenue au moyen du pronom de première personne du pluriel accompagné du groupe prépositionnel *avek x*. Appelons cette variable *X-et-moi*, et la variante réalisée en (97a) *X-et-moi-2*. Une autre variante, que je nomme *X-et-moi-1* et que je suggère de classer en Z2, est proposée en (97c).

(97) c. nu-partõ pur barsə̀lon fabiẽ e mwa pur katr zur a partir d dmẽ

Cette variante suggère un dernier mélange :

(97) d. ð-par pur barsə̀lon fabiẽ e mwa pur katr zur a partir d dmẽ

UE a d'ailleurs produit un énoncé équivalent, à ceci près que le groupe *avek x* se trouve avant le verbe :

1. Il s'agit d'un écrit informel attesté (un courriel entre deux jeunes cousins très proches) que je reproduis à la lettre.

(98) o bu d kelk tã avek P13 òn-a deside d rãtre

(UE : 333)

11.2 S'expliquer le contre-exemple

L'exemple attesté en (97a) présente donc un mélange Z1-Z3, le mélange de *1pl-1* et *X-et-moi-2*. Comment défendre la diglossie face à ce contre-exemple flagrant ? Je propose l'explication suivante :

1. La locutrice, dans une situation informelle, produit spontanément un énoncé de Z3 ou Z2-Z3, soit (97b) ou (97d). Ici, elle 'tombe sur' (97b).
2. Comme elle se trouve dans la situation où elle peut planifier son énoncé, puisqu'il s'agit d'un courriel, la locutrice, par réflexe associé à la production écrite, se lance dans une activité métalinguistique de correction normative. Elle repère donc la variante *1pl-2* qu'elle décide de corriger en *1pl-1*, et obtient (97a).
3. Son investigation normativisante s'arrête là, pour au moins deux raisons possibles. Elle peut simplement ne pas s'être investie plus avant dans la normativisation de son énoncé, par exemple parce que la situation, bien qu'écrite, n'en reste pas moins informelle, ce que suggère l'absence de majuscules. Ou bien, elle peut aussi n'avoir pas repéré l'autre variante Z3, *X-et-moi-2*, peu connue, donc nettement moins stigmatisée et moins aisément repérable et normativisable.
4. L'énoncé agrammatical (97a) est donc, selon ce scénario, le résultat d'une interférence entre une production informelle Z2-Z3 et un processus de normativisation provoqué par la possibilité de planifier, par réflexe, cette production écrite.

11.3 Un scénario reproductible

L'exemple suivant confirme le besoin d'un scénario plausible pour ce type de mélange. (99) a été produit dans des conditions finalement similaires à (97a). Il s'agit d'une anecdote publiée par une lectrice d'un site internet à caractère humoristique, dont les visiteurs souhaitent une forte proximité avec les anecdotes publiées. La tension entre le contexte informel et la pression de la publication écrite provoque à nouveau le mélange discuté.

(99) VDM numéro 155549 (<http://www.viedemerde.fr/155549>, je souligne) :

Aujourd'hui, nous avons fait l'amour avec mon copain, la musique en mode aléatoire. Tout se passait bien jusqu'au moment où Johnny Hallyday s'est mis à brailler dans les enceintes : "On est champions, on est tous ensemble..." Et pour achever ce moment romantique, mon copain a crié "ET BUUUT!". VDM

Non seulement l'interférence invoquée est donc plausible, mais elle semble même être un scénario dont la description est nécessaire.²

11.4 L'écrit est décidément particulier

D'après ce scénario, l'écrit est tout simplement à même de provoquer des interférences permanentes lorsqu'il est associé à un contexte sociolin-

2. Au moment de boucler la rédaction, je reçois ce mail :

- (i) *peux tu me l envoyer ton [rapport] car cela marche pas des masses le mail du club en ce moment, [X] est en vacances...*

Le lecteur appréciera je crois la truculence des interférences observables.

11.4 *L'écrit est décidément particulier*

guistique informel. Selon cette idée, il n'est plus envisageable d'étudier le FD sur des corpus écrits. En tout cas, tant que l'on pourra leur préférer des données orales spontanées, il faudra les éviter. Et, s'ils sont inévitables, par exemple lors d'une étude historique, l'étude de ces corpus écrits doit être accompagnée d'une théorie et d'une modélisation explicites du phénomène d'interférences, puisque l'écrit semble ne jamais pouvoir assurer la spontanéité de la production.

Et pour l'étude du FCT ? On pourrait continuer à se dire que, comme langue normée, le FCT est à décrire simplement en suivant les prescriptions de la norme. Donc un écrit bien normé convient comme base à sa description, tout comme une liste de prescriptions. Mais il me semble qu'il y a là un soucis également. Étudier le FCT, c'est bien étudier une langue vivante, une langue implémentée par des grammaires internes productives et acquises par les locuteurs. Pourquoi pourrait-on alors s'en remettre à des productions écrites, sujettes à des interférences qui en suppriment potentiellement la spontanéité, sans prendre les précautions que je viens d'évoquer à propos du FD ?

Les productions variables d'UE d'une part, et l'étude inductive que l'on en a faite d'autre part pour en dégager un FCT partiel, montrent qu'il serait envisageable, voire souhaitable, de replacer l'étude du FCT dans une méthodologie classique de description d'une langue vivante, quelle que soit la théorie que l'on défende de l'idée de langue vivante (générativiste ou autre). Et dans le cas particulier d'une théorie fondée sur l'idée de langue interne, il faudrait assumer que les productions écrites n'y offrent pas un accès idéal.

Les générativistes assument cette situation puisqu'ils proposent des jugements de grammaticalité, même pour étudier le FCT. Mais, puisque le FCT est sujet à une norme standard, ces jugements sont guidés par des motivations externes. En tout cas, c'est ce que laisse entendre Zribi-Hertz

(2006:4, typographie originale) :

« GS [=FCT] doit conventionnellement être activée dans la production de certains écrits : articles de journaux sérieux, textes savants, rapports professionnels, lettres de candidature, devoirs d'étudiants — textes dont le scripteur *contrôle la conformité à la norme standard*, ce qu'il ne fait pas nécessairement dans d'autres écrits tels que les lettres intimes, et les messages SMS ou courriel. Pour savoir si une forme est produite par GS, il suffit donc de répondre à la question : cette forme peut-elle s'employer dans un *écrit contrôlé* (càd. : contrôlé pour sa conformité) ? »

La contradiction qui est donnée à voir est la suivante. Étudier une langue, c'est, selon le programme générativiste, étudier la grammaire interne des locuteurs. Cette grammaire interne est explicitable, selon cette théorie, par les jugements de grammaticalité. Or, dans le cas de l'étude du FCT, ces jugements de grammaticalité semblent être remplacés par des jugements de conformité. Cette conformité est une norme externe, qui n'est pas établie par l'étude d'idiolectes, mais, en toute bonne norme, reflète une langue externe, dont Zribi-Hertz elle-même dénonce l'hétérogénéité qui la rend impropre à la modélisation de grammaires internes.

Enfin, et c'est ce que Zribi-Hertz reproche à certains générativistes, même en faisant l'effort de décrire le français à partir de jugements non-influencés par la norme, on peut être amené à émettre des jugements qui ne correspondent pas aux productions effectives des locuteurs, en particulier parce qu'il faudrait pour cela, comme on l'a montré, assumer correctement une structure interne de la compétence des francophones en termes de double-grammaire.

11.5 Définitions empiriques du FCT et du FD

Pour conclure cette partie, j'invite à considérer la possibilité qu'elle offre de définir empiriquement le FCT et le FD. À partir de l'observation de mélanges interdits au sein des productions d'un locuteur, on induit des variantes Z1 et des variantes Z3. Devant toute variante nouvellement observée, on prédit qu'elle doit être compatible soit seulement avec toutes les variantes déjà classées Z1, soit seulement avec toutes les variantes déjà classées Z3, soit avec toutes les variantes de ces deux ensembles. Un locuteur qui respecte ces prédictions sera dit diglosse FCT-FD, et, quelle que soit la méthode adoptée pour repérer un mélange interdit (les jugements de grammaticalité, ou l'étude de corpus spontané, ou autre), on pourra décrire sa compétence linguistique en ces termes :

- (100)

Le français classique tardif d'un locuteur est tout ce qui est compatible dans ses productions spontanées avec ses variantes exclusivement classiques, et son français démotique est tout ce qui est compatible dans ses productions spontanées avec ses variantes exclusivement démotiques.
--

Conclusion

ME VOICI à l'heure de conclure ce travail. Je commencerai par dresser un bilan sommaire du chemin parcouru. Ensuite, je listerai l'ensemble des hypothèses que cette thèse a reprises depuis la littérature ou générées elle-même, puis un bref aperçu des anomalies que ce travail engendre. Puis j'indiquerai quelques manques et quelques points ouverts au sortir de cette thèse. Je terminerai par deux mots de la fin. Le premier est la reproduction d'un article d'Alain Kihm paru dans *Libération* l'hiver dernier (Kihm, 2008), qui fait écho à l'extrait de film présenté à l'introduction, et qui constitue la deuxième entorse à ma volonté de ne pas mettre les pieds dans le débat extra-linguistique. Le second est une citation évoquant la frustration qui peut naître de la sensation d'inachevé au moment de boucler une thèse.

Bilan du programme de la thèse

Au programme de cette thèse, on trouvait le souhait de montrer la valeur de l'hypothèse de la diglossie au sein du débat qui anime les linguistes qui s'intéressent d'une façon générale aux données du français qui sortent de la norme officielle.

La première partie a permis de présenter une certaine idée de la langue. Il s'agit de la conception de la tradition générativiste autour de la notion de langue interne. Ensuite, cela a permis d'exposer les tenants et aboutissants de la problématique de la variation morphosyntaxique individuelle en français, puis de définir clairement le contenu de l'hypothèse de la diglossie au sein de cette problématique et face à d'autres hypothèses.

Concernant maintenant l'idée que le FCT et le FD ont bien chacun une grammaire distincte de l'autre, voici ce qu'on a pu voir. La morphologie du nombre (partie II), étudiée à la lumière d'une transcription donnant

la commune mesure entre les productions formelles et les productions informelles et à la lumière d'un modèle des statuts morphosyntaxiques possibles du nombre, n'est pas fondamentalement guidée par deux organisations autonomes. Au contraire, le nombre du FCT et le nombre du FD sont simplement le nombre du français en général.

Dans le cas de la négation et de la syntaxe de la structure informationnelle (partie III), il est au contraire nettement plus adéquat de parler de deux grammaires distinctes, chacune des deux étant plus simple à décrire, et plus facile à envisager en termes typologiques.

Ce contraste entre le cas de morphologie flexionnelle qu'est le nombre et les cas de syntaxe de la négation et de la SI se dessine à l'opposé de ce que propose Blanche-Benveniste dans l'ensemble de son œuvre (par exemple Blanche-Benveniste *et al.*, 1990, Blanche-Benveniste, 1997, 2003, 2004) : « [entre la grammaire de l'écrit et la grammaire de l'oral] on voit des différences [dans la morphologie] bien plus incontestables que dans la syntaxe » (Blanche-Benveniste, 2004:130).

Je crois que ce résultat opposé s'explique d'une part pour la morphologie par l'absence de considérations orthographiques de ma démarche, puisque l'orthographe ne représente pas une commune mesure entre les diverses productions, et d'autre part pour la syntaxe par le refus de décrire nécessairement ensemble tout ce que peut produire un même locuteur spontanément et oralement.

Enfin, le point majeur de mon argumentation est la situation induite du corpus étudié dans la partie IV : la diglossie est une réalité des données orales spontanées des locuteurs qui sont capables de produire des variantes de la norme et des variantes en dehors de celle-ci.

Hypothèses générées

Les hypothèses générales que cette thèse a générées ou reprises sont les suivantes :

- L'écrit n'est pas un objet linguistique banal, et son étude ne donne pas un accès simple à la compétence spontanée des locuteurs.
- La variation grammaticale en français est bipolaire (ce qui n'exclut pas une multipolarité à l'intérieur de chaque pôle).
- Les objets d'étude de cette variation — les (ensembles de) variantes grammaticales — sont alors à comprendre comme ressortissant au français classique tardif ou au français démotique tels que ceux-ci ont été définis ici.
- Ces objets d'étude de la description grammaticale sont individuels et à décrire à travers des idiolectes.
- Les locuteurs produisant du FCT et du FD sont diglosses, c'est-à-dire qu'ils ont deux grammaires internes, l'une produisant du FCT, l'autre produisant du FD.
- Lorsque deux idiolectes divergent, leur intercompréhension relève d'une accommodation.

Les hypothèses spécifiques concernant des points précis de grammaire sont les suivantes :

- Le nombre du FCT et le nombre du FD ont une grammaire équivalente.
- Ce nombre a une force grammaticale faible. En particulier, il a une morphologie presque entièrement dépendante, et il n'a pas d'autonomie syntaxique.
- La négation du verbe avec *n* et la négation du verbe sans *n* ne ressortissent pas nécessairement d'une structure syntaxique commune.

- La syntaxe associée à la structure informationnelle, dans le cas des sujets topicaux, n'est pas la même en FCT, qui est de type SV(O), et en FD, qui est de type presque exclusivement (T) pro-V (X) (AT). Enfin, les hypothèses concernant le patron de variation des productions spontanées des locuteurs sont les suivantes :
- Les locuteurs n'activent qu'une grammaire à la fois, soit celle du FCT, soit celle du FD.
- Le domaine de cohérence grammaticale des productions est la proposition.
- Les variantes produites par un locuteur diglosse se laissent qualifier de façon nécessaire et suffisante par le concept des zones zribihertziennes.

Anomalies générées

Ce travail génère bien sûr nombre d'anomalies, dont les anomalies importantes suivantes :

- Dans le cas de la négation et de la syntaxe associée à la SI, il peut être attrayant de voir une fracture typologique : négation verbale vs. négation de constituant, et SV(O) vs. (T) pro-V (X) (AT). En revanche, le foisonnement des variantes grammaticales dans d'autres cas, que je n'ai pas présentés ici, comme les différents types de propositions relatives, ne se prête pas aussi intuitivement à une bipartition.
- Souvent, les énoncés ne peuvent pas être identifiés comme du FCT ou du FD. Est-ce que les cas mis en valeur dans le corpus HH ne sont finalement pas marginaux ? Est-ce que Z2 ne serait pas en fait hypertrophiée, rendant l'existence de Z1 et Z3 plutôt anecdotique ?

Le lecteur a sûrement repéré d'autres anomalies. Il n'est pas question de les nier, les ignorer, ou les minimiser. La question est de savoir si ces anomalies empêchent ou non de considérer provisoirement la diglossie comme le meilleur modèle pour observer la variation grammaticale dans les productions des francophones contemporains.

Et l'on invite le lecteur intéressé par le débat à se poser cette question, voire à se motiver pour améliorer, ou au contraire contrarier et défier le modèle.

Manques

Quelques considérations sont absentes de mon travail, sans pour autant que cela y porte grand préjudice. D'autres manques peuvent être ressentis comme problématiques.

Les considérations sociolinguistiques se sont limitées à des intuitions qui me semblent fortes, mais qui n'ont pas été clairement corrélées par une étude sociolinguistique, voire sociologique. Est-ce que les notions de variété H et de variété L sont une réalité sociologique plausible de la communauté linguistique française contemporaine ?

Les cas de diglossie discutés par Ferguson (1959) l'amènent à décrire, au-delà de deux morphosyntaxes distinctes, également deux lexiques et deux phonologies. Si le lexique peut intuitivement se prêter à l'idée d'une répartition en zones zribi-hertziennes, Koch (1997) met en doute qu'il présente bien le patron décrit par Ferguson. Et concernant la phonologie, l'intuition qu'il y ait deux ensembles de faits à décrire me paraît peu convaincante.

Enfin, il serait nécessaire de se lancer pleinement dans une description globale de la grammaire du FD, qui émettrait notamment des hypothèses

sur la cohérence typologique, descriptive et théorique des variantes du FD dans leur ensemble. Ceci constituerait une contre-intuition nécessaire face à l'intuition assumée ici de l'idiosyncrasie de la répartition des variantes entre FCT et FD. À ce titre, on peut citer l'étude diachronique de Barra Jover (2004a) qui propose une telle corrélation entre les interrogatives et la négation du FCT d'une part, et entre les interrogatives et la négation du FD d'autre part. Si le FD est une réalité grammaticale cohérente, quel est son type ?

Ouvertures

Ensuite, quelques points abordés restent ouverts à ce stade. Tout d'abord, on n'a pas proposé de description satisfaisante des N-et-N du FCT. De même, l'enjeu du domaine de portée de l'exigence de cohérence demande à être clarifié et des positions concrètes et justifiées sont encore attendues : est-ce que la proposition est bien ce domaine de cohérence ? Si oui, pourquoi, et comment fonder cela théoriquement ? Si non, quel est ce domaine, s'il existe vraiment (l'étude du chapitre 10 semble bien montrer cette existence) ?

Un point de phonologie qui sollicite mon intuition pour démarquer le FCT et le FD est le cas de la prosodie de l'énoncé. Il me semble qu'une hypothèse corrélant l'opposition de type entre SV(O) et (T) pro-V (X) (AT) à des divergences prosodiques est prometteuse.

Bien des sujets de grammaire doivent pouvoir se prêter à la confrontation à l'hypothèse de la diglossie. Des classiques comme les interrogatives ou les relatives aux faits plus marginaux liés à des items lexicaux isolés, il y a du pain sur la planche. Non pas que ces sujets aient été absents du débat jusqu'à maintenant, mais qu'il soit peut-être intéressant de les

relire dans la perspective dessinée ici.

Dans une zone connexe entre la description linguistique et les attitudes extra-linguistiques (comme les politiques linguistiques), la problématique du Français Langue Étrangère ne manque pas d'être concernée par le débat sur la variation grammaticale du français.

Quel futur pour les données du FD ?

Il y a déjà plusieurs décennies que les données du FD ont droit de cité chez les linguistes. Et les linguistes de l'école diglossique ont déjà pris le parti de les décrire pour elles-mêmes, sans la tutelle à la fois descriptive et théorique du FCT. Si ce programme de mise hors tutelle se popularise, alors une grammaire du FD se dessinera d'elle-même, et ce FD pourra gagner le statut de langue autonome, ce qui fera toujours plus de lui un objet d'étude comparative dans les études synchroniques, et ce qui fera du couple FCT-FD un sujet d'étude diachronique toujours plus privilégié.

Pour aller encore plus loin, l'école diglossique représente un cadre programmatique permettant d'étudier sereinement les données démotiques. Ce programme est déjà mis en œuvre chez certains chercheurs. Et, sous l'impulsion d'Anne Zribi-Hertz et d'autres enseignants-chercheurs, les cours de grammaire française (notamment) du cursus de linguistique de l'UFR SDL de l'université Paris 8 réalisent ce programme depuis plusieurs années comme objet de transmission des connaissances.

Le mot de la fin

Les linguistes ont d'autres mots à dire (Kihm, 2008)

Alain Bentolila, professeur de linguistique à l'université

Paris-V, soutient que les jeunes des banlieues souffrent massivement d'un « déficit linguistique », manifesté par un « vocabulaire exsangue et une organisation approximative des phrases » (Le Monde, 21 décembre 2007). Selon lui, les causes de cette carence résident dans le fait que ces jeunes vivent en vase clos dans leurs quartiers ghettos et leurs bandes, partageant les mêmes goûts, les mêmes intérêts et le même désintérêt pour le reste du monde. Ainsi réduits à une « communication de proximité », ils n'auraient presque rien à se dire, car tout entre eux serait déjà connu. Deux ou trois mots, juxtaposés sans ordre et à peine articulés, suffiraient alors dans la plupart des cas à évoquer dans l'esprit de l'interlocuteur une information trop prévisible.

Alain Bentolila est linguiste. Un linguiste devrait savoir que la majorité des langues du monde (environ 5000) se parle au sein de groupes humains dont les membres se comptent en quelques centaines, souvent bien moins. Dans ces sociétés, dites traditionnelles, les individus se connaissent tous, partagent les mêmes croyances, les mêmes coutumes, les mêmes travaux, la même nourriture. Où trouverait-on connivence plus forte ? Qui aurait moins besoin de se parler pour se comprendre que ces villageois d'Amazonie, d'Afrique, d'Australie, de Nouvelle-Calédonie, etc. ? Si l'on suit le raisonnement de Bentolila, leurs langues devraient être de peu de mots et de syntaxes imprécises... comme on l'a longtemps cru, du reste. Pourtant — un linguiste ne peut l'ignorer — elles sont non seulement aussi riches que le français « cultivé », mais bien souvent d'une complexité dans les procédés de formation de leurs mots et de leurs phrases (morphologie et syntaxe) dont

même le latin et le grec ne nous donnent qu'une pâle idée. Cette absence de corrélation entre le nombre et la condition des locuteurs et le degré d'élaboration du système linguistique qu'ils partagent montre que la communication, au sens strict du terme, n'est pas la fonction première du langage. C'est bien plutôt le récit. La complexité du langage s'explique avant tout par ce besoin vital de conter des histoires aux enchaînements intriqués, pleines d'entités fictives. Comme outil de communication pratique, le langage est du reste bien peu efficace. Essayez donc d'expliquer par les seuls mots comment monter la chaîne d'une tronçonneuse !

Des récits communs, cela s'appelle de la culture. Il faut donc que ces jeunes aphones n'en aient pas... et c'est bien ce qui nous est donné à entendre. Ils ont de la religion, en revanche ! Ils « croient en le même Dieu ». Façon contournée de nous dire que ce sont tous des musulmans. A supposer que ce fût vrai, cela ne suffirait pas à leur donner une culture. Ou bien l'Islam n'en est pas une, mais je me garderai d'imputer à Alain Bentolila une opinion aussi outrancière ; ou bien ce n'est chez eux qu'un reste obscur de cet « ailleurs estompé et confus » qu'ils auraient pour unique origine, affirmation gratuite et certainement fausse, car les jeunes des banlieues savent pour la plupart fort bien d'où ils viennent. Mais l'important pour Alain Bentolila n'est pas qui ils sont mais ce qu'ils sont, à savoir des pauvres. Et les pauvres, on l'a dit avant lui, n'ont pas de (vraie) culture, donc pas de (bon) langage. Le développement absurde sur la « communication de proximité » n'est là que pour nous resservir cette rengaine qu'on voulait croire réfutée. Ici, la référence essen-

tielle reste Richard Hoggart sur la Culture du pauvre, sans parler des travaux de Pierre Bourdieu, qui nous manque toujours davantage. Dans les années 60 et 70, le grand linguiste américain William Labov découvrait chez les jeunes Noirs des quartiers ghettos de New York une culture narrative, d'une richesse insoupçonnée, faite de récits d'expérience, d'énigmes, de blagues ritualisées, de poèmes satiriques, le tout manifestant une virtuosité langagière que presque tous partagent. Insoupçonnée cette richesse, parce que d'une culture consciente d'être stigmatisée. On ne montre pas ce qu'on peut faire à qui, on le sait, l'a par avance jugé et condamné.

Je n'affirme pas que les jeunes de nos banlieues possèdent des richesses équivalentes ; j'affirme que, s'ils les ont, ils ne les montreront jamais aux semblables d'Alain Bentolila... et ils feront bien. Le point de vue de Bentolila participe de ce courant néoréactionnaire, anti-68, où, sous couvert d'opposition à une prétendue « pensée unique », les pires contre-vérités s'énoncent fièrement. Il démontre que les sciences du langage ont si peu pénétré le public que quelqu'un qui s'en réclame s'estime autorisé à soutenir une idée aussi absurde que des personnes, qu'aucune pathologie neurologique n'affecte, aient un déficit de langage. Comme si la faculté de langage n'était pas une propriété génétique de l'espèce humaine ; comme si tous les enfants, ayant acquis sans effort leur langue maternelle, ne se livraient pas à des jeux de langage que seule la maîtrise spontanée d'une syntaxe et d'une sémantique élaborées rend possibles.

Il n'est pas question de nier l'inégalité sociale. Ce pays est en situation de diglossie : il s'y pratique plusieurs langues qui,

pour être égales en valeur expressive aux yeux du linguiste sans préjugés, ne le sont pas aux yeux de beaucoup de ceux qui font les promotions — maîtres, professeurs, recruteurs, intellectuels, etc. Cette situation ne changera pas dans un avenir prévisible. Ce serait alors le rôle de l'école que d'avertir les élèves issus des banlieues (entre autres) que, s'il n'y a rien de mauvais à dire « la cité que je t'ai parlé » (correct en ancien français et dans bien des langues), il est recommandé, pour des raisons qu'on peut leur expliquer (ils ne sont pas idiots et connaissent le monde), d'écrire et dire devant certaines personnes « la cité dont je t'ai parlé ». Il s'agit de les rendre bilingues, de leur apprendre le français « standard » comme une autre langue. Si certains en profitent pour l'approfondir et l'illustrer, tant mieux. On peut espérer qu'ils vivront mieux de n'avoir pas été dénigrés dans leur plus intime, leur langue quotidienne.

Le second mot de la fin

Et le deuxième mot de la fin revient à Laurent R. qui m'a dit, il y a déjà quelques années : « Une thèse, on la termine pas, on l'inter

Bibliographie

- ABEILLÉ, Anne et GODARD, Danièle (1997). «The Syntax of French Negative Adverbs». In Danielle FORGET, Paul HIRSCHBÜHLER, France MARTINEAU et María-Luisa RIVERO (sld), *Negation and polarity : syntax and semantics*. Amsterdam studies in the theory and history of linguistic science : Current issues in linguistic theory, (pp. 1–27).
- ÅGREN, John (1973). *Étude sur quelques liaisons facultatives dans le français de conversation radiophonique – fréquences et facteurs*. KåWe Tryck : Uppsala, 150 pp.
- ALLIÈRES, Jacques (1996). *La formation de la langue française*. Paris : Presses Universitaires de France, troisième édition, 127 pp.
- ANDERSEN, Hanne Leth et NØLKE, Henning (sld) (2002). *Macro-syntaxe et macro-sémantique*. Berne : Lang.
- ARRIVÉ, Michel (1993). *Réformer l'orthographe ?* Paris : PUF.
- ARTIAGOITIA, Xavier (2002). «The functional structure of the Basque noun phrase». In Xavier ARTIAGOITIA, Patxi GOENAGA et Joseba A. LAKARRA (sld), *Erramu Boneta : Festschrift for Rudolf P. G. de Rijk*, Bilbao : Anejos de ASJU, (pp. 73–90).
- ATILF (). «Trésor de la Langue Française Informatisé». URL <http://atilf.atilf.fr/tlf.htm>
- BARRA JOVER, Mario (1999). «Vers une syntaxe guidée par les traits lexicaux : synchronie et diachronie de l'espagnol». Habilitation. Mémoire d'habilitation.

Bibliographie

- BARRA JOVER, Mario (2002). *Propiedades léxicas y evolución sintáctica. El desarrollo de los mecanismos de subordinación en español*. La Coruña : Toxosoutos.
- BARRA JOVER, Mario (2004a). «Interrogatives, négatives et évolution des traits formels du verbe en français parlé». In Barra Jover (2004b), (pp. 110–125).
- BARRA JOVER, Mario (sld) (2004b). *Le français parmi les langues romanes, Langue française*, tome 141. Paris : Larousse.
- BARRA JOVER, Mario (2007). «S'il ne restait que l'induction : corpus, hypothèses diachroniques et la nature de la description grammaticale». *Recherches Linguistiques de Vincennes*, tome 36, pp. 89–122.
- BARRA JOVER, Mario (2008). «Comment évolue un trait grammatical : le pluriel du français dans une perspective romane». *Romance Philology*.
URL <http://halshs.archives-ouvertes.fr/halshs-00266121/fr/>
- BERRENDONNER, Alain (1983a). «La variation polylectale en syntaxe – hypothèses théoriques générales». In Berrendonner (1983b), (pp. 7–109).
- BERRENDONNER, Alain (sld) (1983b). *Principes de grammaire polylectale*. Lyon : Presses Universitaires de Lyon.
- BERRENDONNER, Alain (1990). «Pour une macro-syntaxe». *Travaux de linguistique*, tome 21, pp. 25–36.
- BERRENDONNER, Alain (2002). «Morpho-syntaxe, pragma-syntaxe». In Andersen et Nølke (2002).
- BLANCHE-BENVENISTE, Claire (1983). «L'importance du “français parlé” pour la description du “français tout court”». *Recherches sur le français parlé*, tome 5, pp. 23–45.

Bibliographie

- BLANCHE-BENVENISTE, Claire (1997). *Approches de la langue parlée en français*. L'Essentiel. Paris : Ophrys.
- BLANCHE-BENVENISTE, Claire (sld) (2002a). *Choix de textes de français parlé*. Paris : Champion.
- BLANCHE-BENVENISTE, Claire (2002b). «Macro-syntaxe et micro-syntaxe : les *dispositifs* de la rection verbale». In Andersen et Nølke (2002).
- BLANCHE-BENVENISTE, Claire (2003). «La langue parlée». In Marina YAGUELLO (sld), *Le grand livre de la langue française*, Paris : Seuil, (pp. 317–344).
- BLANCHE-BENVENISTE, Claire (2004). «Le singulier et le pluriel en français parlé contemporain». *Bulletin de la Société de linguistique de Paris*, tome XCIX, pp. 129–154.
- BLANCHE-BENVENISTE, Claire, BILGER, M., ROUGET, C. et VAN DEN EYNDE, K. (sld) (1990). *Le français parlé : Études grammaticales*. Paris : CNRS Éditions.
- BLANCHE-BENVENISTE, Claire et CHERVEL, André (1978). *L'orthographe*. Les textes à l'appui. Paris : François Maspéro, deuxième édition, 260 pp.
- BLANCHE-BENVENISTE, Claire et JEANJEAN, Colette (1987). *Le français parlé : Édition et transcription*. Paris : Didier-Érudition.
- BÖHME-ECKERT, Gabriele (2004). «De l'ancien français au français moderne : l'évolution vers un type « à part » à l'époque du moyen français». In Barra Jover (2004b), (pp. 56–68).
- BOUCHARD, Denis (1997). «The Syntax of Sentential Negation in French and English». In Danielle FORGET, Paul HIRSCHBÜHLER, France MARTINEAU et María-Luisa RIVERO (sld), *Negation and polarity : syntax and semantics*. Amsterdam studies in the theory and history of linguistic science : Current issues in linguistic theory, (pp. 29–52).

Bibliographie

- BOUCHARD, Denis (2002). *Adjectives, number and interfaces : why languages vary*. North-Holland linguistic series : linguistic variations. Amsterdam : Elsevier, 458 pp.
- BOUCHARD, Denis (2003). «Les SN sans déterminant en anglais et en français». In Phillip MILLER et Anne ZRIBI-HERTZ (sld), *Essais sur la grammaire comparée du français et de l'anglais*, Paris : PUV, (pp. 55–95).
- BOUORA, Leïla (2003). *Étude des systèmes d'écriture des langues vocales et des langues signées : description et analyse comparatives de deux systèmes "idéographiques" et de Sign Writing*. Mémoire de DEA, Université Paris 8, Saint-Denis.
- CATACH, Nina (1998). *L'orthographe. Que sais-je?* Paris : PUF, 8^e édition, 128 pp.
- CHAPAT, François-Pierre (1999). *Avant que langage ne meure – Lexique du patois de Puisaye*. éditions Le Carrouge.
- CHAURAND, Jacques (sld) (1999). *Nouvelle histoire de la langue française*. Paris : Seuil, 808 pp.
- CHAURAND, Jacques (2006). *Histoire de la langue française*. Paris : Presses Universitaires de France, 11^e édition, 127 pp.
- CHÉRY, Henry (1933/1976). *Lexique du « Parler » de Mézilles vers la fin du XIX^e siècle*. Toucy (Yonne) : Association d'études, de recherches et de protection du vieux Toucy.
- Corpus GARS (2004). *Autour du corpus de référence du français parlé, Recherches sur le français parlé*, tome 18. PUP.
- COVENEY, Aidan (1996). *Variability in Spoken French : interrogation and negation*. Exeter : Elm Bank Publications.
- CSÉCSY, Madeleine (1971). «Les marques orales du nombre». In André RIGAULT (sld), *La grammaire du français parlé*, Paris : Hachette.

- CUXAC, Christian (2000). «Compositionnalité sublexicale morphémique-
iconique en langue des signes française». *Recherches Linguistiques de
Vincennes*, tome 29, pp. 55–72.
- DEJEAN, Yves (1983). «Diglossia revisited, French and Creole in Haiti». *Word*,
tome 34 (3), pp. 307–334.
- DOBROVIE-SORIN, Carmen et LACA, Brenda (2003). «Les noms sans
déterminant dans les langues romanes». In Danièle GODARD (sld),
Les langues romanes, problèmes de la phrase simple, Paris : CNRS
Éditions, (pp. 235–279).
- ESCHMANN, Jürgen (1976). *Die Numerusmarkierung des Substantivs im
gesprochenen Französisch*. Tübingen : Niemeyer, 108 pp.
- FARANDJIS, Stelio (1983). «Préface». In Frédéric FRANÇOIS (sld),
J'cause français, non ?, Paris : La Découverte/Maspero.
- FERGUSON, Charles A. (1959). «Diglossia». *Word*, tome 15 (2), pp.
325–340.
- FERNÁNDEZ, Mauro (1993). *Diglossia : a comprehensive bibliography
1960-1990*. Amsterdam : Benjamins, 472 pp.
- FISHMAN, Joshua A. (sld) (1971). *Advances in the sociology of language*,
tome I, Basic Concepts, Theories and Problems : Alternative Ap-
proaches. La Hague/Paris : Mouton.
- GADET, Françoise (1992). *Le français populaire*. Que sais-je? Paris :
PUF.
- GADET, Françoise (1997a). «La variation en syntaxe». *Langue française*,
tome 115.
- GADET, Françoise (1997b). *Le français ordinaire*. Paris : Armand Colin,
deuxième édition.
- GADET, Françoise (2003). «La variation». In Marina YAGUELLO (sld),
Le grand livre de la langue française, Paris : Seuil, (pp. 91–152).

Bibliographie

- GREVISSE, Maurice (1993). *Le bon usage : grammaire française*. Paris : Duculot, 13^e édition.
- GUILLAUME, Gustave (1919/1975). *Le Problème de l'article et sa solution dans la langue française*. Paris : A.-G. Nizet.
- HAIMAN, John (1985). *Natural syntax : iconicity and erosion*. Cambridge : Cambridge university press, 285 pp.
- HAWKINS, Peter (1983). «Diglossia revisited». *Language and Communication*, tome 3 (3), pp. 219–232.
- HIRSCHBÜLLER, Paul et LABELLE, Marie (1993). «Le statut de (ne) pas en français contemporain». *Recherches Linguistiques de Vincennes*, tome 22, pp. 31–58.
- HUCHON, Mireille (2002). *Histoire de la langue française*. Paris : Librairie générale française, 315 pp.
- KELLER, Rudi (1994). *On language change : The invisible hand in language*. Londres : Routledge.
- KIHM, Alain (2008). «Les linguistes ont d'autres mots à dire». *Libération*. 7 février 2008, rubrique « Rebonds ».
- KOCH, Peter (1997). «Diglossie in Frankreich?» In Winfried ENGLER (sld), *Frankreich an der freien Universität, Geschichte und Aktualität*, Stuttgart : Franz Steiner Verlag, (pp. 219–249).
- KOCH, Peter et OESTERREICHER, Wulf (1990). *Gesprochene Sprache in der Romania*. Numéro 31 in Romanistische Arbeitshefte. Tübingen : Niemeyer.
- KOCH, Peter et OESTERREICHER, Wulf (2001). «Langage parlé et langage écrit». In Günter HOLTUS, Michael METZELIN et Christian SCHMITT (sld), *Lexikon der romanistischen Linguistik*, Tübingen : Niemeyer, tome I.2, (pp. 584–627).

- KURODA, S.-Y. (1972). «The categorical and the thetic judgment.» *Foundations of Language*, tome 9, pp. 153–185.
- LADUSLAW, W. (1994). «Thetic and Categorical, Stage and Individual, Weak and Strong». In M. HARVEY et L. SANTELMANN (sld), *Proceedings of SALT IV*. Ithaca, NY : Cornell University DMLL, (pp. 220–229).
- LAMBRECHT, Knud (1981). *Topic, Antitopic and Verb-Agreement in Non-Standard French*. Pragmatics and Beyond. Amsterdam : John Benjamins, 113 pp.
- LAMBRECHT, Knud (1987). «On the status of SVO sentences in French discourse». In R. TOMLIN (sld), *Coherence and grounding in discourse*, John Benjamins, (pp. 217–261).
- LAMBRECHT, Knud (1994). *Information Structure and Sentence Form*. Cambridge : Cambridge university press.
- LAMBRECHT, Knud (2000). «Prédication seconde et structure informationnelle. La relative de perception comme construction présentative». *Langue française*, tome 127, pp. 49–66.
- LAMBRECHT, Knud (2002). «Topic, focus, and secondary predication. The French Presentational Relative Construction». In Claire BEYSSADE, Reineke BOK-BENNEMA, Frank DRIJKONINGEN et Paula MONACHESI (sld), *Proceedings of Going Romance 2000*. John Benjamins.
- LANGACKER, Ronald Wayne (1991). *Concept, image and symbol : the cognitive basis of grammar*. Berlin/New York : Mouton de Gruyter, 355 pp.
- Le Monde de l'Éducation (2004). «Orthographe, pourquoi ça fait mal». *Le Monde de l'éducation*, (325), pp. 26–40.
- LEJEUNE, Alain (2001). *L'parler d'cheu nous – lexique du patois de la Puisaye*. publié par l'auteur.

Bibliographie

- LÜDI, Georges (1990). «Diglossie et polyglossie». In Günter HOLTUS, Michael METZELIN et Christian SCHMITT (sld), *Lexikon der romanistischen Linguistik*, Tübingen : Niemeyer, tome IV, (pp. 307–334).
- MAIR, Victor H. (2003). «Préface de J. Marshall Unger *Ideogram : Chinese Characters and the Myth of Disembodied Meaning*». University of Hawaii Press.
- MASSOT, Benjamin (2002). «Schwa autour des groupes de consonnes obstruante-liquide en Poyaudin». Mémoire de maîtrise. Université Paris 8.
- MASSOT, Benjamin (2003). *Éléments linguistiques pour une vision diglossique du français contemporain*. Mémoire de DEA, Université Paris 8, Saint-Denis.
- MCHOMBO, Sam (2003). «On discontinuous constituents in Chichewa». In Patrick SAUZET et Anne ZRIBI-HERTZ (sld), *Typologie des langues d'Afrique et universaux de la Grammaire — Volume 1 : Approches transversales — Domaine bantou*. Paris : L'Harmattan.
- MILLER, Phillip, PULLUM, Geoff et ZWICKY, Arnold (1992). «Le principe d'inaccessibilité de la phonologie par la syntaxe : trois contre exemples apparents en français». *Linguisticæ Investigationes*, tome XVI (2), pp. 317–343.
- MILNER, Jean-Claude (1982a). *Ordres et raisons de langue*. Paris : Seuil.
- MILNER, Jean-Claude (1982b). *Une anaphore non-référentielle : Le système de la négation en français et l'opacité du sujet*, (pp. 186–223). In Milner (1982a).
- MULLER, Claude (1991). *La négation en français*. Genève : Droz.
- PICOCHÉ, Jacqueline et MARCHELLO-NIZIA, Christiane (1996). *Histoire de la langue française*. Paris : Nathan, quatrième édition, 396 pp.
- QUENEAU, Raymond (1965). *Batons, chiffres et lettres*. Paris, deuxième édition.

- RECOURCÉ, Gaëlle (1996). «Une double analyse de la particule *ne* en français». *Langages*, tome 122, pp. 62–78.
- RIALLAND, Annie (2005). «Phonological and phonetic aspects of whistled languages». *Phonology*, tome 22 (2), pp. 237–271.
- ROODENBURG, Jasper (2004). *Pour une approche scalaire de la déficience nominale : la position du français dans une théorie des noms nus*. Thèse de doctorat, Utrecht – Université Paris 8.
- RUBELLIN, François (sld) (1996). *Marivaux dramaturge : La double inconstance ; Le jeu de l'amour et du hasard*. Paris : Champion.
- SASSE, Hans-Jürgen (1987). «The thetic/categorical distinction revisited». *Linguistics*, tome 25 (3), pp. 511–580.
- SÖLL, Ludwig (1985). *Gesprochenes und geschriebenes Französisch*. Berlin : E. Schmidt, troisième édition.
- TOLLIS, Francis (1998). *La description du castillan au XVe siècle : Villena et Nebrija : sept études d'historiographie linguistique*. Paris/Montréal : L'Harmattan, 414 pp.
- WREN, Sebastian (). «The Brain and Reading».
<http://www.sedl.org>.
- WREN, Sebastian (2006). «The simple view of reading – $R = D \times C$ ».
<http://www.balancedreading.com/simple.html>,
traduction française sur
<http://appy.ecole.free.fr/3evoie/fran03.pdf>.
- ZANUTTINI, Raffaella (1997). *Negation and Clausal Structure*. Oxford : Oxford University Press.
- ZRIBI-HERTZ, Anne (2006). «Français standard et français dialectal : pour une reconnaissance générale de la diglossie francophone dans les descriptions syntaxiques du français». Ms.

Annexes

Annexe A

Présentation sommaire du poyaudin

Cette annexe présente en quelques mots le poyaudin, qui a été évoqué à quelques reprises.

Nomenclature

Il existe deux séries de synonymes, dont certains termes (indiqués en italique) sont privilégiés dans l'usage :

la région	<i>la Puisaye</i> [pɥize]	la Poyaude [pwajod]
les habitants	les Puisayens [pɥizajɛ̃]	<i>les Poyaudins</i> [pwajodɛ̃]
la langue	le puisayen	<i>le poyaudin</i>
le qualificatif	puisayen	<i>poyaudin</i>

La Puisaye

Le pays de Puisaye, région géologique naturelle, regroupe 5 cantons du sud-ouest de l'Yonne (Bléneau, Charny, Saint-Fargeau, Saint-Sauveur-en-

Puisaye, Toucy), un canton du nord-ouest de la Nièvre (Saint-Amand-en-Puisaye) et quelques communes du Loiret. On compte environ 31000 habitants.

Le poyaudin

Le poyaudin est un parler d'oïl, de la famille des parlers du centre. Les derniers locuteurs sont tous âgés et le poyaudin s'exprime principalement à travers les traces qu'on en trouve dans le français local, tant dans l'accent et la phonologie que dans le lexique et quelques tours grammaticaux. Le poyaudin n'est plus transmis comme langue maternelle.

C'est une langue très peu étudiée. On trouve quelques lexiques, qui sont parfois accompagnés de remarques sur la prononciation ou la conjugaison (Chéry, 1933/1976, Chapat, 1999, Lejeune, 2001). Massot (2002) traite un point de phonologie.

Il n'y a pas de tradition écrite et les graphies employées pour écrire en poyaudin ne se libèrent que très peu de l'orthographe standard du français. C'est pourquoi j'utilise le même principe de système de transcription que pour le français.

Mon principal informateur est mon grand-père maternel, 86 ans, né à Saint-Amand-en-Puisaye (58).