

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/270881784>

Syntaxe générative et acquisition : le sujet dans le développement du système linguistique du jeune enfant

Thesis · December 2009

DOI: 10.13140/2.1.2722.0806

CITATIONS

4

READS

142

1 author:



[Katerina Palasis](#)

Univ. Nice Sophia Antipolis, CNRS, BCL, UMR 7320

19 PUBLICATIONS 46 CITATIONS

[SEE PROFILE](#)

Université de Nice-Sophia Antipolis
UFR Lettres, Arts et Sciences Humaines
Département des Sciences du Langage
École Doctorale « Lettres, Sciences Humaines et Sociales » (ED86)

***SYNTAXE GÉNÉRATIVE ET ACQUISITION :
LE SUJET DANS LE DÉVELOPPEMENT
DU SYSTÈME LINGUISTIQUE DU JEUNE ENFANT***



« L'escargot i(l) dort. Un nuage il arrive... » (Alan, 2;8.18)

*Thèse de Doctorat
Présentée par Katérina PALASIS
Le 8 décembre 2009*

*Sous la direction de Jean-Philippe DALBERA
et Michèle OLIVIÉRI*

Jury :
DALBERA Jean-Philippe
EZEIZABARRENA SEGUROLA Maria-José
HUDELOT Christian
KAISER Georg A.
MORGENSTERN Aliyah
OLIVIÉRI Michèle

« La science ne poursuit jamais l'objectif illusoire de rendre ses réponses définitives ou même probables. Elle s'achemine plutôt vers le but infini encore qu'accessible de toujours découvrir des problèmes nouveaux, plus profonds et plus généraux, et de soumettre ses réponses, toujours provisoires, à des tests toujours renouvelés et toujours affinés. »

Popper (1935:287)

CONVENTIONS

LUS	:	Code correspondant au pseudonyme d'un locuteur (LUS = Lucie) ¹ .
2;7.9	:	Âge d'un locuteur, donné en années;mois.jours.
II.9b,132	:	Référence du fichier dans la base de données, suivie du numéro de la ligne sur laquelle se trouve la transcription ² .
xx	:	Syllabe inintelligible.
xxx	:	Ensemble de syllabes inintelligibles.
i(l)	:	La forme canonique du morphème est transcrite mais les parenthèses entourent les phonèmes non produits par le locuteur (i(l) = [i]).
[?]	:	La transcription de l'élément précédent est incertaine.
[: mot]	:	Lorsque la transcription d'un élément est différente du terme canonique, ce dernier est ajouté entre crochets.
0	:	Élément non réalisé.
[*]	:	L'élément précédent est agrammatical.
parce+que	:	Les éléments de part et d'autre du signe + forment une seule unité pour le logiciel d'étiquetage (pour les noms propres : Père_Noël).
[/]	:	Reprise d'un élément dans le cours d'un énoncé.
<> [/]	:	Reprise de plusieurs éléments dans le cours d'un énoncé.
<> ["]	:	Citation (répétition, lecture, paroles de chansons, etc.).
/syllabe	:	La syllabe est accentuée.
mot [!]	:	Le mot est accentué.
^	:	Pause entre deux syllabes.

¹ La liste complète des codes se trouve au Chapitre III, page 179.

² L'ensemble des transcriptions se trouve sur le CD inséré en Annexes, page 371.

RÉSUMÉ EN FRANÇAIS

La problématique de l'acquisition du sujet syntaxique chez l'enfant francophone est traitée dans le cadre d'une approche qui allie les formalisations de la grammaire générative à des méthodes empruntées à la dialectologie. Ainsi, l'analyse proposée est formulée dans le cadre des Principes et des Paramètres mais repose sur l'étude d'un corpus de données enfantines spontanées recueillies sur une période de sept mois auprès de vingt enfants entre deux ans et demi et quatre ans. Ces données se composent de transcriptions orthographiques accompagnées d'un étiquetage morphosyntaxique et d'un codage des agrammaticalités détectées. L'ensemble se présente sous la forme d'une base de données compatible avec la plateforme CHILDES (*Child Language Data Exchange System*)¹. Un total de 10674 sujets potentiels est ainsi analysé (clitiques nominatifs seuls, doublés ou triplés par des pronoms toniques ou des noms, éléments interrogatifs et relatifs, syntagmes nominaux utilisés seuls et cas d'absences de tout élément réalisé). Nos recherches permettent alors de proposer l'existence de deux grammaires du français : (i) un français « spontané », système linguistique initialement manipulé par tout locuteur francophone dans lequel les clitiques nominatifs sont des préfixes verbaux et (ii) un français « normé » dans lequel les clitiques nominatifs sont des arguments syntaxiques. Une fission du paramètre *pro-drop* en deux paramètres est également proposée afin de rendre compte des variations inter-linguistiques observées : le paramètre de la suffixation verbale distinctive et celui de la préfixation verbale.

Mots clés :

Acquisition, sujet, syntaxe générative, dialectologie, corpus, paramètre *pro-drop*, sujet nul.

¹ « Système d'échange de données linguistiques enfantines ». Cf. <http://childes.psy.cmu.edu/>.

TITRE ET RÉSUMÉ EN ANGLAIS

Generative Syntax and Acquisition: the Subject within the Development of the Young Child's Linguistic System

We address the matter of the acquisition of the syntactic subject by French-speaking children combining the generative framework with methods borrowed from dialectology. Indeed, our analysis is forwarded within the Principles and Parameters theory but nevertheless also rests on the scrutiny of a broad corpus of spontaneous data collected over seven months with twenty children between two years and a half and four years old. These data are transcribed orthographically and are labelled on two additional dependent tiers with morpho-syntactic and error coding. The database is hence compatible with CHILDES (Child Language Data Exchange System)¹. A total of 10674 potential subjects is analysed, to include nominative clitics uttered on their own, doubled, or tripled with tonic pronouns or nouns, interrogative and relative items, nominal phrases uttered on their own, as well as null elements. Our research hence allows us to argue in favour of the existence of two French grammars: (i) an initial “spontaneous” system manipulated by all the French speakers in which the nominative clitics are handled as verbal prefixes and (ii) a “normed” system in which the nominative clitics are syntactic arguments. A split *pro-drop* parameter, i.e. a distinctive suffix parameter and a prefix parameter, enables us to account for the observed interlinguistic variations.

Keywords:

Acquisition, subject, generative syntax, dialectology, corpus, pro-drop parameter, null subject.

¹ Cf. <http://childes.psy.cmu.edu/>.

LABORATOIRE DE RATTACHEMENT

Cette thèse a été préparée au sein du Laboratoire :

UMR 6039 « Bases, Corpus, Langage »
CNRS – Université de Nice-Sophia Antipolis
MSH de Nice
98 bd Edouard Herriot
06200 NICE

TABLE DES MATIÈRES

CONVENTIONS	3
RÉSUMÉ EN FRANÇAIS	4
TITRE ET RÉSUMÉ EN ANGLAIS	5
LABORATOIRE DE RATTACHEMENT	6
TABLE DES MATIÈRES	7
INTRODUCTION	14
CHAPITRE I :	
SUJET, DÉVELOPPEMENT ET SYSTÈME LINGUISTIQUE	19
I. L'ENFANT LOCUTEUR	20
1. Les développements cognitifs du jeune enfant	21
2. Les premiers développements langagiers	22
2.1 Du sujet énonciateur au sujet grammatical	23
2.2 Ce que nous apprend Léonard	23
II. LA NOTION DE SUJET	27
1. Une fonction à la croisée des chemins	28
2. La nature de ses éléments	30
2.1 Une grande diversité	30
2.2 La délicate question du « pronom »	32
2.3 Les pronoms personnels du français moderne	33
2.3.1 Les pronoms atones.....	34
2.3.2 Les pronoms toniques	37
2.4 Les autres constituants étudiés	38
3. Vers les questions de registres de langue et de normes	39
III. QUEL(S) SUJET(S) POUR QUEL(S) FRANÇAIS ?	41
1. La tradition des définitions contrastives.....	41
1.1 Écrit vs oral	41
1.2 Les français « avancé », « parlé courant » et « très évolué »	44
1.2.1 La dislocation	46
1.2.2 Le doublement.....	47
1.2.3 La quantification	49

2. Comment concilier grammaire interne et variation	50
2.1 La « diglossie francophone »	51
2.2 Instaurer une chronologie des événements	52
3. Comment définir le sujet dans ce cadre	55
3.1 L'importance du facteur prosodique	55
3.2 Quelques caractéristiques du français oral adulte	56

CHAPITRE II :

L'ACQUISITION DU LANGAGE

ET LA THÉORIE LINGUISTIQUE	58
---	-----------

I. LE CHOIX DE NOTRE CADRE THÉORIQUE

1. Le comportementalisme

1.1 L'état initial vu comme une coquille vide.....	60
1.2 Le mécanisme de l'apprentissage.....	61
1.3 <i>Quid</i> du langage ?.....	62
1.4 La psycholinguistique	64

2. L'innéisme

2.1 Chomsky vs Skinner.....	65
2.2 Les racines de la théorie générative	66
2.3 Les hypothèses fondamentales du générativisme	68

3. Une nouvelle perspective pluridisciplinaire.....

3.1 Pour l'innéisme	70
3.2 Le problème des données	73
3.3 Les apports complémentaires d'une approche dialectologique	75
3.4 Une première application : l'étude du « sujet nul » enfantin	78
3.4.1 L'origine du sujet nul enfantin.....	79
3.4.2 La nature des sujets enfantins	80

II. LE PROCESSUS ACQUISITIONNEL

1. Les Principes.....

1.1 Le Principe de Projection Étendu.....	85
1.2 Le Principe des Catégories Vides	87
1.3 Vers des propriétés moins spécifiquement linguistiques	88

2. Les Paramètres.....

2.1 Le Paramètre <i>pro-drop</i>	90
2.1.1 Ses origines	90
2.1.2 Ses limites et les amendements proposés.....	93
2.2 La fixation des paramètres	95
2.3 Les Paramètres dans le Programme Minimaliste : syntaxiques ?	97

III. LE PROGRAMME MINIMALISTE

1. Quelques nouveautés du modèle.....

2. Les différents types de traits

3. La distribution ordonnée des thème-rôles.....

4. Les projections fonctionnelles.....

5. La fission de la périphérie gauche	109
5.1 Une profusion de projections	109
5.2 Topique vs Focus.....	111
5.3 Les clitiques nominatifs ne sont pas toujours les bienvenus.....	113
6. La révision de <i>EPP</i> et du paramètre <i>pro-drop</i>	114
6.1 La formulation de Pollock (1997).....	115
6.2 Vers l'élimination de <i>pro</i> ?	115
IV. PLUSIEURS CANDIDATS, UNE SEULE FONCTION.....	117
1. Le doublement du clitique objet	118
1.1 Description du phénomène.....	118
1.2 Les problèmes théoriques.....	120
1.2.1 Le clitique porte le thème-rôle = doublement.....	121
1.2.2 Le clitique porte le thème-rôle = dislocation	122
1.2.3 Le syntagme prépositionnel porte le thème-rôle.....	124
2. Le débat sur le statut du clitique nominatif.....	126
2.1 Des langues standards aux dialectes du nord de l'Italie.....	127
2.2 Un débat binaire : syntaxe vs morphologie	129
2.2.1 La redondance des marques d'accord	130
2.2.2 L'indépendance syntaxique du clitique nominatif.....	131
2.2.3 Le lien entre interprétation et structure syntaxique.....	133

CHAPITRE III : LA CONSTITUTION

DE NOTRE BASE DE DONNÉES..... 138

I. RÉFLEXIONS ÉPISTÉMOLOGIQUES ET MÉTHODOLOGIQUES SUR LE CORPUS DE DONNÉES ENFANTINES

1. Vous avez dit « corpus » ?	140
2. Les différents types de données enfantines.....	142
2.1 Les données issues de la compréhension et du jugement	142
2.2 Les données issues de la production	145
3. Les bases de données spontanées francophones.....	149
3.1 Les différentes collectes.....	149
3.2 La diffusion de ces données	152
4. Les limites des bases de données existantes	153
4.1 Les problèmes liés à la collecte et à la transcription.....	154
4.1.1 La pluralité des sources.....	155
4.1.2 L'absence de support audio ou vidéo.....	157
4.2 Les problèmes liés à la variation linguistique	160
4.2.1 La variation individuelle	161
4.2.2 La variation collective.....	164
5. La mise en place de notre cahier des charges.....	168
II. LA COLLECTE DES DONNÉES.....	171
1. Quelques considérations liminaires d'ordre pratique	172

2. Les différentes démarches	174
2.1 L'information aux familles	175
2.2 Le respect de la vie privée.....	176
2.3 Le formulaire d'autorisation	177
3. La rencontre avec les enfants	178
3.1 L'identification de nos informateurs	178
3.2 La mise en place du calendrier des séances	180
3.3 Les moyens techniques	181
3.4 Le protocole de collecte	182
3.5 Les résultats de la collecte.....	187
III. VERS LA BASE DE DONNÉES	188
1. Le choix de la plateforme	188
2. La mise en place des fichiers	190
2.1 La capture des données brutes sur informatique	190
2.2 La création des fichiers CHAT	191
IV. LA TRANSCRIPTION DES DONNÉES	193
1. La segmentation en énoncés	194
1.1 Définir la notion d'« énoncé »	195
1.2 De la difficulté à transcrire les pauses.....	196
1.2.1 La parataxe	196
1.2.2 L'ordre des constituants	200
2. L'anonymisation des données	201
3. Les deux étapes de la transcription	203
3.1 La transcription à partir des fichiers vidéo.....	203
3.2 La transcription à partir des fichiers audio.....	205
V. LE CODAGE DES DONNÉES.....	208
1. Le codage des agrammaticalités	209
1.1 Le choix des occurrences	210
1.2 Notre grille de codes	211
1.3 L'importance du contexte	213
2. Le codage morphosyntaxique	215
3. En conclusion.....	218
<i>CHAPITRE IV : ANALYSE.....</i>	<i>219</i>
I. EXTRACTIONS PRÉLIMINAIRES DE DONNÉES.....	219
1. CLAN	219
2. Premières requêtes afin de composer notre corpus	220
2.1 Les locuteurs	221
2.2 Les phrases verbales.....	222
2.3 Clitiques nominatifs et verbes fléchis	225
3. Récapitulatif	226
4. Les différentes formes de sujets.....	227

II.	LES CLITIQUES NOMINATIFS.....	228
1.	Les clitiques utilisés seuls	228
1.1	L'ensemble du corpus	229
1.2	Le système enfantin.....	231
1.3	L'évolution des clitiques : ce que nous dit la diachronie	234
1.4	L'évolution des clitiques : ce que nous dit l'acquisition.....	236
2.	Les clitiques doublés	238
2.1	L'ensemble du corpus	239
2.2	Le système enfantin.....	240
3.	L'évolution des clitiques en termes de traits	241
3.1	Les traits [Personne] et [Locuteur].....	242
3.2	Le trait [Genre].....	243
3.3	Le trait [Nombre]	247
3.4	Le trait [Référentiel].....	249
3.5	Synthèse	251
4.	Ce que les clitiques nous disent sur leur statut morphosyntaxique.....	252
4.1	L'étude des clitiques préverbaux non nominatifs	253
4.1.1	La particule négative « ne »	253
4.1.2	Les clitiques « y » et « en ».....	254
4.1.2.1	Les codages de « y » et « en ».....	255
4.1.2.2	Le système enfantin	257
4.1.3	Les clitiques accusatifs et datifs.....	260
4.2	Un ajustement phonétique contraint des clitiques nominatifs.....	262
III.	LES ÉLÉMENTS QU-.....	270
1.	Les pronoms interrogatifs	271
1.1	Les différentes formes.....	272
1.2	Computations syntaxiques ou locutions figées ?	273
1.3	<i>Qui il ?</i>	276
1.4	La difficulté à extraire un sujet	277
2.	Les pronoms relatifs.....	277
2.1	La théorie	278
2.2	Les données	282
2.3	Commentaires	284
2.4	Vers une fission du paramètre <i>pro-drop</i>	289
IV.	LES CO-RÉFÉRENTS D'UN CLITIQUE.....	291
1.	Le doublement.....	291
1.1	Les pronoms toniques	291
1.2	La latéralisation du doublement des pronoms toniques	294
1.3	Les autres constituants	296
1.4	La latéralisation du doublement des autres constituants.....	298

2.	Le triplement	302
2.1	Un clitique, deux pronoms toniques.....	303
2.2	Un clitique, deux syntagmes	305
2.3	Le triplement et la théorie des copies.....	307
V.	LES SYNTAGMES NOMINAUX SEULS	310
1.	Définition	311
2.	L'extraction des données	312
3.	Les caractéristiques de cette structure	313
3.1	Une augmentation des occurrences au fil des séances	313
3.2	Les éléments indéfinis et quantifiés	313
4.	Une structure rare : pourquoi ?	316
5.	Vers une redistribution de la vérification des traits	317
VI.	L'ABSENCE DE TOUT ÉLÉMENT RÉALISÉ	322
1.	Les impératifs	322
1.1	La théorie	322
1.2	Les données.....	324
2.	Les « sujets nuls »	327
2.1	Définition et nouvelle appellation : le phénomène du clitique nul	327
2.2	L'extraction des énoncés.....	328
2.3	Quelles personnes verbales ?	331
2.4	Quels verbes ?	332
2.5	Les verbes dans l'ensemble du corpus	333
2.6	Notre hypothèse sur le phénomène du clitique nul	334
2.7	La disparition progressive des clitics nuls	338
VII.	SYNTHÈSE	340
1.	Les principales catégories de sujets potentiels	340
2.	La répartition de ces catégories dans le système infantin	343
	CHAPITRE V : CONCLUSIONS	345
I.	LES FRANÇAIS SPONTANÉ ET NORMÉ	345
II.	LE CLITIQUE EST D'ABORD MORPHOLOGIQUE	347
1.	Le préfixe ne marque pas l'accord de façon redondante	349
2.	Le préfixe n'est pas aléatoirement optionnel	350
3.	La proclise est un mécanisme productif	351
4.	Les statuts morphologique et syntaxique coexistent	352
5.	Focalisation et préfixation sont compatibles	355
6.	Les sujets silencieux	357
III.	LA FISSION DU PARAMÈTRE <i>PRO-DROP</i>	359
1.	Le paramètre de la suffixation verbale distinctive	360
2.	Le paramètre de la préfixation verbale	361

IV.	POUR TERMINER	363
1.	De nombreuses pistes à développer	363
2.	Une base de données destinée à s'agrandir.....	364
	ANNEXES	366
I.	LA CONSTITUTION DE LA BASE DE DONNÉES	367
1.	Le dossier d'information destiné aux parents	367
2.	Le formulaire d'autorisation parentale	370
3.	Les fichiers informatiques de la base de données	371
4.	Liste des codes %err	372
4.1	Tournures incorrectes : \$INC ou \$POS	372
4.2	Omissions d'éléments : \$LOS.....	372
4.3	Ajouts d'éléments : \$ADD ou \$RED	374
4.4	Substitutions d'éléments : \$SBL ou \$SUB	374
4.5	Liaisons : \$ALL	376
4.6	Phonologie : \$PHO	376
4.7	Lexique : \$LEX.....	377
II.	LES ÉNONCÉS.....	378
1.	Base de données complète.....	378
2.	Phrases verbales et non verbales	379
III.	LES SYNTAGMES NOMINAUX SEULS	380
1.	Statistiques.....	380
1.1	Tous locuteurs confondus	380
1.2	Enfants.....	381
2.	Liste exhaustive	382
IV.	LES VERBES	384
1.	Les verbes dans les énoncés à clitique nul.....	384
2.	Les verbes enfantins dans l'ensemble du corpus.....	385
	BIBLIOGRAPHIE	387

INTRODUCTION

La notion de sujet peut être étudiée selon un grand nombre de points de vue différents, tant au niveau intralinguistique (notamment en phonologie, morphologie, syntaxe, sémantique, prosodie, énonciation ou pragmatique) qu'au niveau extralinguistique (notamment dans le cadre de la psychologie). Par ailleurs, des approches théoriques différentes coexistent et/ou se succèdent dans l'histoire de chacune de ces disciplines. De plus, l'étude de l'acquisition du langage en général et celle du sujet en particulier représentent des problématiques extrêmement transversales dans la mesure où elles peuvent être traitées dans chacune des disciplines mentionnées ci-dessus. Néanmoins, la recherche scientifique, quel que soit son domaine d'application, exige une maîtrise préalable et approfondie d'un domaine d'étude et d'une théorie en particulier afin d'offrir des fondations de qualité à toute recherche nouvelle. Malheureusement, cette exigence présente également l'inconvénient, au tout début tout au moins, de donner un aspect parfois parcellaire à ces premières investigations ; mais ce n'est que pour mieux les compléter par la suite. Ainsi, les fondements de notre étude actuelle sur le sujet dans le développement du système linguistique du jeune enfant sont principalement syntaxiques et générativistes¹.

Depuis les années 1980, l'étude de l'acquisition de la syntaxe représente un domaine de recherches extrêmement florissant en grammaire générative. En effet, la théorie des Principes et des Paramètres mise en place principalement à partir de Chomsky (1981) afin de décrire à la fois les similitudes et les variations interlinguistiques offre un cadre formel fort convaincant aux chercheurs qui, par ailleurs, adhèrent à la conception innéiste de l'acquisition du langage. Grâce à ces paramètres, l'acquisition de caractéristiques pourtant variables d'une langue à une autre (telles

¹ Nous présenterons les justifications de ces choix au cours de notre travail.

que l'ordre relatif des constituants dans un syntagme ou la place du verbe fléchi dans la proposition) peut en effet être étudiée et formalisée de façon binaire. Dès les années 1980, un paramètre en particulier suscite alors un nombre important de travaux : il s'agit du paramètre *pro-drop* (ou paramètre du « sujet nul ») qui rend compte de la présence ou de l'absence du sujet, sous forme de pronom, dans les énoncés des différentes langues du monde. Ainsi, les exemples en (1) ci-dessous illustrent le fait que certaines langues ne requièrent pas la présence obligatoire d'un sujet, quelle que soit sa nature (*cf.* (a)), alors que les autres requièrent la présence minimale d'un pronom sous peine d'agrammaticalité (*cf.* (b)). Les premières sont qualifiées de langues *pro-drop* (ou langues à « sujet nul ») et, par opposition, les secondes sont des langues non *pro-drop* (ou langues à « sujet non nul »).

(1) La dichotomie instaurée par le paramètre *pro-drop* :

- | | | |
|-----------------------------------|---------------------|------------|
| (a) Langues <i>pro-drop</i> : | _ dhoulèvo | (grec) |
| | _ lavoro | (italien) |
| (b) Langues non <i>pro-drop</i> : | <u>je</u> travaille | (français) |
| | <u>I</u> work | (anglais) |

Les chercheurs en acquisition s'intéressent beaucoup à ce paramètre dans la mesure où un certain nombre de données enfantines dans différentes langues non *pro-drop* ne se conforment pas à la description illustrée en (1) ci-dessus. En effet, les corpus de français enfantin par exemple recueillis auprès d'enfants de moins de trois ans comportent tous des énoncés tels que ceux présentés en (2) ci-dessous dans lesquels le sujet obligatoire n'est pas réalisé. Ce type d'occurrences porte alors l'appellation de « sujet nul » enfantin :

(2) Le « sujet nul » enfantin¹ :

- | | | |
|------|--|--------|
| ALA: | 0 [*] veux ranger mon jardin . (I.3b, 292) | 0 = je |
| EMA: | 0 [*] a bobo dedans . (III.10b, 75) | 0 = il |

Dans Palasis (2005), nous nous sommes intéressée à ce phénomène et avons proposé une hypothèse basée sur l'étude d'un premier corpus enfantin de données

¹ Données extraites de notre présent corpus (Corpus n°2) dont les caractéristiques sont décrites aux Chapitres III et IV et dont les fichiers de transcriptions et de codages se trouvent en Annexes.

spontanées¹. Néanmoins, cette étape particulière du « sujet nul » ne représente qu'une seule des pièces du puzzle que constitue la problématique beaucoup plus vaste de l'acquisition du sujet syntaxique. Nous avons alors souhaité recueillir de nouvelles données auprès d'enfants différents afin de vérifier notre hypothèse initiale et afin d'élargir notre périmètre d'investigation à la totalité des sujets enfantins. Le présent travail de recherches propose ainsi une description précise, un classement et une analyse de 10674 sujets produits par vingt enfants francophones entre 2;6 et 4;0. L'étude se veut tant transversale que longitudinale dans la mesure où les données sont recueillies sur une période de sept mois et que la collecte dont nous utilisons les données à présent ne représente que la première des trois années que nous avons passées en compagnie de ce groupe d'enfants.

Néanmoins, une telle investigation ne peut être menée à bien si les contours de la problématique et le cadre théorique dans lequel cette problématique est traitée ne sont pas définis préalablement de façon précise. Le Chapitre I propose ainsi une description du moment à partir duquel il est possible de parler de « développement du sujet » chez un enfant ; il définit également les constituants potentiellement « sujets » en français et propose une réflexion autour de la notion de « système linguistique ». La problématique est alors définie comme celle de l'analyse des constituants (essentiellement représentés par des noms, pronoms clitiques et toniques) potentiellement sujets chez des enfants à partir de 2;9 environ. Par ailleurs, nous argumentons en faveur de l'existence de plusieurs grammaires du français : un français spontané qui est le système linguistique initialement manipulé par tout locuteur francophone et un français normé dont certains paramètres présentent une valeur différente par rapport au français spontané. Nous rejoignons ainsi notamment la notion de « diglossie » évoquée par Zribi-Hertz (2006). Depuis sa naissance à la fin des années 1950, la grammaire générative a, d'une part, énormément évolué et est, d'autre part, à l'origine de nombreuses confusions et controverses. Les objectifs du Chapitre II sont donc d'explicitier les fondamentaux initiaux de cette théorie linguistique ainsi que la formalisation déployée par l'actuel Programme Minimaliste. Néanmoins, nous détaillons également les raisons pour lesquelles nous favorisons une approche qui, tout en étant ancrée dans le cadre générativiste, utilise aussi des

¹ Corpus composé de 1072 phrases verbales recueillies auprès de dix-sept enfants entre 2;3 et 3;1 (Corpus n°1). Cf. Palasis (2005) pour l'intégralité de ces données ainsi que le détail de la démonstration. Cf. Chapitre II, Section I.3.4 du présent travail pour un résumé de cette hypothèse.

méthodes et des principes empruntés à la dialectologie, tels que la collecte de données sur le terrain afin de composer un corpus de travail et le regard porté sur ce corpus. Ces premiers jalons posés, nous nous tournons alors vers la question des données dans le Chapitre III dans la mesure où leur utilisation en tant que telle ne représente absolument pas un gage de qualité scientifique. En effet, comme le décrit Dalbera (2002), un corpus, « échantillon représentatif d'un système linguistique », est un outil qui se construit et qui est donc le fruit d'un travail de composition à partir d'un ensemble initial de données. Or, de nombreuses problématiques surgissent dès le choix qui doit être fait des locuteurs qui seront enregistrés. Ce troisième chapitre expose ainsi les réflexions d'ordres épistémologique et méthodologique menées autour des nombreuses questions afférentes à la construction de notre corpus infantin : du type de données (productions sollicitées ou spontanées par exemple) aux questions de représentativité liées aux variations linguistiques collectives et individuelles. Nous terminons le Chapitre III par la description des étapes de collecte, de transcription et de codage conduites afin de constituer notre propre base de données dont nous présentons une fiche signalétique complète au tout début du Chapitre IV. Cette base est compatible avec la plateforme CHILDES (*Child Language Data Exchange System*)¹ et se compose de 30542 énoncés transcrits orthographiquement et étiquetés selon deux types de codes : un codage morphosyntaxique (%mor) et une grille de codes qui permet d'extraire et de classer les agrammaticalités des énoncés (%err).

La problématique, le cadre théorique et la base de données sont ainsi en place au Chapitre IV. L'analyse de tous les éléments potentiellement sujets dans le corpus peut alors être effectuée. Sont ainsi passés en revue les clitiques nominatifs seuls, doublés ou triplés par des pronoms toniques ou des syntagmes nominaux, les éléments interrogatifs et relatifs, les syntagmes nominaux utilisés seuls, ainsi que les cas d'absences de tout élément réalisé, tels que les impératifs et les « sujets nuls ». Nos recherches nous permettent alors de proposer une description de l'apparition des clitiques nominatifs dans le système infantin en termes de traits ([Personne], [Locuteur], [Genre], [Nombre] et [Référentialité]) dont l'acquisition présente des similitudes extrêmement intéressantes avec la diachronie des systèmes linguistiques étudiés par Oliviéri (2009). Par ailleurs, plusieurs phénomènes tels que l'ajustement

¹ « Système d'échange de données linguistiques enfantines ». Cf. MacWhinney (2000a, 2000b) et le site internet de cette plateforme : <http://childes.psy.cmu.edu/>.

phonétique extrêmement contraint du clitique *il* dans certains contextes et l'émergence de la particule négative *ne* mettent au jour des indications quant à l'existence de deux clitiques distincts : un clitique morphologique, préfixe verbal, et un clitique phonologique, argument syntaxique du verbe. Ce corpus permet également de corroborer l'hypothèse sur le « sujet nul » enfantin initiée dans Palasis (2005) dans la mesure où nous confirmons qu'il s'agit de l'application d'un principe d'économie résultant de la configuration particulière des paradigmes verbaux enfantins mais que, en aucun cas, le jeune enfant ne considère son système linguistique comme une langue *pro-drop*. Nos données confirment également que la structure canonique du français normé (DP[^]V, *l'escargot dort* par exemple) n'est quasiment pas présente dans ce système spontané dans lequel les clitiques nominatifs tiennent une place prépondérante. Enfin, au Chapitre V, notre corpus permet de démêler des problématiques très discutées telles que les structures de « doublement » et de « dislocation » en français, notamment grâce à la fission que nous proposons du paramètre *pro-drop* en deux paramètres : le paramètre de la suffixation verbale distinctive et celui de la préfixation verbale.

CHAPITRE I :

SUJET, DÉVELOPPEMENT

ET SYSTÈME LINGUISTIQUE

Notre objectif étant d'étudier le sujet dans le développement du système linguistique du jeune enfant, il convient dans ce premier chapitre d'expliciter : (i) ce que nous entendons par « sujet », (ii) à quelle tranche d'âge nous nous référons pour traiter de son « développement » et (iii) de quel « système linguistique » nous parlons. Dans un premier temps, l'utilisation du terme « sujet » requiert des précisions dans la mesure où ce lexème est employé dans des domaines divers avec des significations différentes. Trois champs d'investigation sont alors particulièrement importants dans le cadre de notre étude : la psychologie, l'énonciation et la syntaxe. Ainsi, en psychologie, la notion de sujet représente celle de l'être individuel, de la personnalité. En revanche, dans le cadre énonciatif, la notion de sujet renvoie principalement à celle du sujet parlant, du locuteur et à la place que celui-ci occupe dans ses productions linguistiques, ses énoncés. Puis, lorsque ce terme est utilisé en syntaxe, il prend alors le sens de fonction grammaticale exercée au sein des propositions qui composent les énoncés. Néanmoins, notre formulation vise à montrer que l'existence de ces différentes acceptions ne signifie pas qu'elles ne présentent pas de liens entre elles : il faut en effet être « sujet individuel » avant de devenir « sujet parlant » afin de produire des « sujets syntaxiques ». Ainsi, nous postulons que, pour pouvoir étudier l'évolution du sujet grammatical dans les productions linguistiques du jeune enfant, il faut d'abord que nous nous arrêtons sur la thématique de son développement en tant qu'être individuel puis jeune locuteur.

Dans ce cadre, bien que notre problématique soit *in fine* syntaxique, celle-ci ne peut donc pas être abordée d'emblée dans la mesure où le développement syntaxique – et le développement linguistique en général – est subordonné à d'autres évolutions, notamment cognitives et énonciatives. Dans la première section de ce chapitre, nous proposons alors de rendre compte des étapes au cours desquelles l'enfant doit tout d'abord se découvrir et s'assumer en tant qu'être humain autonome afin de pouvoir par la suite s'assumer en tant que locuteur, c'est-à-dire se projeter dans ses énoncés. La première partie de ce chapitre est ainsi consacrée à exposer les connaissances en psychologie du développement et en acquisition de l'énonciation qui nous semblent représenter les préalables indispensables à l'étude du développement syntaxique chez le jeune enfant. La seconde section de ce chapitre traite d'un second préalable : la définition de la notion de sujet syntaxique. À cette fin, nous explicitons notamment les termes de nature et de fonction grammaticales et proposons un panorama des sujets potentiels en français. Ce terme de « français » nous amène enfin, dans la dernière section de ce chapitre, à reprendre les définitions qui sont généralement proposées de ce système linguistique afin de rendre compte des variations qui existent entre une utilisation plutôt « orale » de la langue et une utilisation plutôt « écrite » de cette même langue. Nous terminons ce tour d'horizon préliminaire en suggérant une approche différente qui prend en compte la dimension acquisitionnelle du langage et qui pose alors le problème non pas en termes de « variations » au sein d'un système linguistique unique mais de coexistence de deux systèmes linguistiques distincts.

I. L'ENFANT LOCUTEUR

À la suite d'auteurs tels que Piaget (1923), l'un des pères fondateurs de l'étude du développement infantin, on continue à postuler que le jeune enfant, jusqu'à l'âge de cinq-six ans, se trouve dans une période où l'égoïsme est prépondérant, c'est-à-dire que, schématiquement, son point de vue est globalement unique et plutôt autocentré. Dans ce cadre, il est notamment considéré que l'enfant, à ce stade, est incapable de se mettre à la place de quelqu'un d'autre et qu'il n'envisage donc pas que l'autre puisse également avoir un point de vue. À cette pensée égoïste, Piaget (1923) fait correspondre un langage « égoïste » dans lequel les productions verbales servent à « [...] accompagner, renforcer ou supplanter l'action » ; ainsi elles « [...] n'éveillent aucune réaction adaptée de la part des

interlocuteurs occasionnels »¹. D'après l'auteur, ce type de langage est largement majoritaire au cours du premier stade des productions enfantines entre deux et trois ans, puis il diminue progressivement entre trois et six ans pour représenter après sept ans moins d'un quart des productions de l'enfant. Cette analyse ne fait cependant pas l'unanimité dans la mesure où elle induit que le langage enfantin n'aurait dans un premier temps aucune fonction sociale. Ainsi, Vygotski (1934) notamment défend un point de vue opposé à celui de Piaget (1923) puisqu'il intègre à l'étude du développement du langage enfantin les dimensions culturelles et sociales de l'individu. L'auteur considère alors que le langage égocentrique a pour but la communication et l'extériorisation des pensées et qu'il a ainsi une fonction sociale. Nous n'approfondissons pas cette discussion dans le cadre de nos recherches actuelles ; néanmoins, nous considérons qu'il y a des connaissances issues de la recherche en psychologie du développement qu'il est intéressant d'intégrer dans une étude sur l'acquisition du langage.

1. Les développements cognitifs du jeune enfant

La troisième année de l'enfant est unanimement reconnue comme étant une période charnière pour son développement cognitif, et ce pour plusieurs raisons. Tout d'abord, il s'agit de la période pendant laquelle le jeune enfant prend conscience qu'il est un être humain autonome ; il accède ainsi à « la conscience de soi ». Cette étape représente l'aboutissement d'un processus progressif de différenciation qui permet à l'enfant de se détacher petit à petit des adultes qui l'entourent d'attentions, notamment sa mère. L'acquisition de la marche puis de la propreté représentent alors des jalons intermédiaires vers cette individuation. Même si tous les chercheurs ne sont pas d'accord sur le moment exact auquel l'enfant accède à cette conscience de soi, les travaux sur la reconnaissance de l'image spéculaire de Zazzo (1993) semblent néanmoins faire référence en la matière et situent cette étape entre dix-neuf et trente mois². Toujours d'après ces tests, le processus est complété lorsque, environ six mois plus tard, l'enfant parvient également à identifier le reflet d'une autre personne visible dans le miroir.

¹ Piaget (1923:23 et 36), cité par Bideaud *et al.* (1993:388).

² Le test consiste à apposer une marque colorée sur le visage de l'enfant puis à placer l'enfant devant un miroir. Après avoir vu la marque, si l'enfant touche celle-ci sur son propre visage, c'est qu'il a compris que le miroir ne renvoie que le reflet de lui-même. Par contre, s'il touche le miroir, c'est que l'enfant n'a pas encore compris qu'il ne s'agit que d'un reflet.

Cette période correspond également au moment où le jeune enfant parvient à se détacher de l'« ici et maintenant » pour accéder à une dimension plus symbolique de la pensée, ce que Piaget & Inhelder (1969) décrivent comme étant « [...] la capacité d'évoquer des objets ou situations non perçus actuellement en se servant de signes ou de symboles »¹. Cette évolution se traduit notamment par les jeux symboliques de l'enfant tels que l'invention de personnages et de situations, l'imitation différée d'événements réellement vécus ou encore le dessin. Ces autonomisations du jeune enfant, tant au niveau physique que psychique, représentent ainsi les conditions qui permettent alors à l'enfant de s'épanouir également en tant que jeune locuteur de sa ou ses langue(s) première(s).

2. Les premiers développements langagiers

Au cours de sa troisième année, le jeune enfant prend ainsi également son envol au niveau linguistique. De la cinquantaine de mots prononcés de façon plutôt isolée vers l'âge de dix-huit mois (ce que l'on appelle les holophrases), le lexique de l'enfant s'enrichit de plusieurs centaines d'items pendant cette période. Il s'agit de la phase de « l'explosion du lexique ». Cet âge est également celui de l'émergence de la syntaxe puisque l'enfant commence à combiner plusieurs éléments au sein d'un même énoncé. Ainsi, lors d'études sur l'acquisition de la syntaxe, la longueur moyenne des énoncés enfantins est très souvent calculée et le chiffre obtenu est généralement utilisé par les chercheurs pour évaluer à quel stade de son développement linguistique se trouve un enfant. Cette longueur moyenne des énoncés chez un sujet est appelée *MLU* (*Mean Length of Utterance*) dans la littérature anglophone et est généralement calculée en morphèmes. Lorsque celle-ci est calculée en mots (ce qui est plus rare), le sigle *MLU* est complété de l'indice *w* (pour *word*) : *MLU_w*. Cette notion, instaurée par Brown (1973), est très largement utilisée dans la mesure où elle sert à délimiter des étapes dans le développement syntaxique du jeune enfant. Ainsi, Brown (1973) propose cinq étapes dans l'évolution de la complexification syntaxique : (i) une *MLU* de 1,75 morphèmes vers l'âge de deux ans, (ii) 2,25 vers deux ans et demi, (iii) 2,75 vers trois ans, (iv) 3,50 vers trois ans et demi et (v) une *MLU* de 4 morphèmes vers quatre ans². Ces premières indications nous permettent de commencer à cerner la limite inférieure de

¹ Piaget & Inhelder (1969:75).

² Cf. Chapitre III pour une discussion sur la *MLU*.

la tranche d'âge pertinente pour notre étude ; cependant, elles demeurent encore insuffisantes. Il nous reste en effet à définir un moment, le plus précis possible et identifié grâce à des critères objectifs, à partir duquel nous pouvons étudier le développement du sujet syntaxique chez le jeune enfant. Nous nous tournons ainsi vers une autre approche du langage, celle de l'énonciation.

2.1 Du sujet énonciateur au sujet grammatical

Pour Benveniste (1958), « le langage n'est possible que parce que chaque locuteur se pose comme *sujet*, en renvoyant à lui-même comme *je* dans son discours »¹. L'enfant doit tout d'abord être capable de se projeter dans sa langue afin de manipuler les éléments de celle-ci :

« C'est dans et par le langage que l'homme se constitue comme *sujet* ; parce que le langage seul fonde en réalité, dans *sa* réalité qui est celle de l'être, le concept d' « ego ». [...]. Or nous tenons que cette « subjectivité », [...], n'est que l'émergence dans l'être d'une propriété fondamentale du langage. Est « ego » qui *dit* « ego ». »^{2,3}

L'émergence du sujet grammatical de première personne est donc subordonnée à la capacité de l'enfant à devenir sujet énonciateur. Nous affinons ainsi nos investigations et posons que le développement du sujet syntaxique peut être étudié à partir du moment où l'enfant maîtrise de façon stable cette capacité à se projeter dans le langage. En effet, à partir de cette étape, nous pouvons : (i) parler de *développement* dans le sens de « extension, épanouissement de quelque chose qui est déjà présent » et ainsi (ii) étudier les différentes formes de cet épanouissement. Nous nous tournons ainsi vers des recherches menées conjointement en énonciation et en acquisition, notamment les travaux de Morgenstern (2006) sur « l'émergence des facultés énonciatives » en français.

2.2 Ce que nous apprend Léonard

Les travaux de Morgenstern (2006) sont basés sur un corpus que l'auteur a collecté elle-même auprès d'un petit garçon francophone, prénommé Léonard. Les

¹ Benveniste (1958:260).

² Benveniste (1958:260-261).

³ Afin de rendre la lecture la plus fluide possible, nous citons les auteurs en langue française lorsque, bien sûr, le texte original est en français mais également lorsqu'il existe une traduction de l'ouvrage cité (cf. certains ouvrages de Chomsky par exemple). Lorsqu'il n'existe pas de traduction française, nous citons les propos originaux de l'auteur et proposons une traduction en bas de page.

transcriptions proviennent d'enregistrements vidéo effectués au domicile de Léonard, en compagnie de ses proches (mère, père, amis de la famille), au cours de quinze séances entre les âges de 1;8 et 3;3¹. L'auteur nous livre dix saynètes représentatives de l'évolution du discours de Léonard pendant cette période, notamment en ce qui concerne les désignations que fait cet enfant de lui-même en tant que sujet grammatical.

Il est tout d'abord remarqué que, dans un premier temps, les énoncés de Léonard ne comportent aucune forme de sujet grammatical renvoyant à lui-même. En effet, à 1;8, le nombre d'auto-désignations explicites est nul. Morgenstern (2006) baptise ainsi cette période la « préhistoire des auto-désignations »² tout en remarquant qu'il y a néanmoins des traces de la subjectivité de Léonard dans son discours grâce à un autre type de marqueur, moins explicite, tel que *non* ou *encore*, mais qui peut véhiculer « des valeurs très proches des énoncés explicitement à la première personne » et que « les travaux précédents sur la question de l'auto-désignation n'avaient pas envisagés »³. Ensuite, un total de neuf formes différentes d'auto-désignations se succèdent et/ou se mêlent :

(3) Les différentes formes d'auto-désignation chez Léonard⁴ :

(a) Absence de forme :	[sot] « (je) saute »	(1;10)
(b) Voyelle préverbale :	[ømũz] « je mange »	(2;0)
(c) <i>moi</i> :	[madɔ] « moi (je) donne »	(1;10)
(d) <i>il</i> :	[ladi] « il a dit »	(2;2)
(e) <i>Léonard</i> :	[leonaadi] « Léonard a dit »	(2;2)
(f) <i>Léonard + il</i> :	[leonaila] « Léonard il a »	(2;2)
(g) <i>je</i> :	ze pleurais	(2;11)
(h) <i>moi + il</i> :	moi aussi ze-, il aime le gromanoi aussi	(2;8)
(i) <i>moi + je</i> :	moi je trouve qu'i sont beaux	(3;0)

L'étude très précise du corpus et de son contexte de production permet alors à Morgenstern (2006) de cerner trois stades d'évolution dans la production de ces auto-désignations : (i) Période I (entre 1;10 et 2;1) pendant laquelle les formes utilisées ne

¹ Ce corpus fait partie d'un projet plus vaste « le projet Léonard » mentionné au Chapitre III. Cf. également le site Internet du projet : <http://anr-leonard.ens-lsh.fr/>.

² Morgenstern (2006:76).

³ Morgenstern (2006:78-79). Cf. également les travaux de Spitz (1962) sur des étapes antérieures du développement enfantin en rapport avec la notion de refus.

⁴ La transcription est phonétique jusqu'à 2;5 inclus, puis orthographique.

correspondent pas à des formes « adultes »¹ : il s'agit des formes illustrées de (a) à (c) ci-dessus ; (ii) Période II (entre 2;2 et 2;8) où toutes les formes sont employées, « comme s'il fallait en essayer beaucoup pour sélectionner les « bonnes » »² ; (iii) Période III (entre 2;9 et 3;3) pendant laquelle les formes attendues *je* et *moi je* deviennent majoritaires. L'auteur apporte ainsi une analyse fine de l'utilisation de chacune de ces formes en rapport avec les formes concomitantes. Il est notamment intéressant de noter que l'utilisation relativement importante de *il* comme marqueur d'auto-désignation au cours de la Période II³ est analysée comme étant un moyen pour l'enfant « de faire des récits sur lui-même en se plaçant du même point de vue que l'adulte » sans pour autant encore maîtriser « une forme marquant l'identification entre sujet de l'énoncé et sujet énonciateur »⁴. Néanmoins, l'auteur résume le développement de Léonard ainsi :

« Au terme du corpus, l'enfant est tout entier inclus dans ses prises de paroles avec la possibilité de parler de lui-même à travers le temps [...], et de donner ses opinions et ses sentiments qu'il articule à ceux des autres. »⁵

Afin de rendre compte de ce parcours, Morgenstern (2006) travaille sur une tranche d'âge bien précise qui s'étend de 1;8 à 3;3. Ce choix est délibéré dans la mesure où il s'appuie notamment sur les considérations d'ordre cognitif que nous avons mentionnées au début de ce chapitre. Ses observations sur le processus de mise en place de la subjectivité dans le langage enfantin confirment que cette période représente un tout dans l'émergence des facultés énonciatives :

« En prenant en compte les recherches dans le domaine de la psychologie, on peut trouver une confirmation du choix de la période d'analyse qui va de 18 mois à 3 ans en mettant en parallèle la constitution de la conscience de soi et l'apparition des auto-désignations. »⁶

¹ Nous revenons sur cette notion dans la section suivante.

² Morgenstern (2006:66).

³ Jusqu'à représenter 31 % de toutes les formes d'auto-désignation à 2;05.

⁴ Morgenstern (2006:144).

⁵ Morgenstern (2006:142).

⁶ Morgenstern (2006:42).

Ainsi, grâce aux recherches menées en psychologie du développement d'une part et à des travaux tels que ceux de Morgenstern (2006) en acquisition de l'énonciation d'autre part, nous pouvons à notre tour définir des contours d'âges précis afin de mener à bien des recherches dans un domaine encore différent qui est celui de l'acquisition de la syntaxe, par le biais d'un cadre théorique différent qui est celui de la grammaire générative.

La Période III de Léonard – entre 2;9 et 3;3 – est celle pendant laquelle l'enfant assoit sa subjectivité et maîtrise ainsi l'utilisation de la première personne grammaticale pour marquer concomitamment son rôle d'énonciateur et d'agent de l'action. Cette maîtrise est notamment marquée par la disparition totale de la troisième personne comme marqueur d'auto-désignation¹. Nous considérons donc que la phase de *l'émergence* proprement dite se termine et que nous pouvons alors parler de *développement*. Nous choisissons ainsi cette période comme point de départ pour nos propres recherches². Celles-ci utilisent donc, parmi d'autres fondations, l'état des lieux proposé par Morgenstern (2006) pour la production d'éléments sujets à la première personne avec, à 3;3, une utilisation dominante du marqueur *je* (66 %), suivi par l'emploi de *moi je* (20 %) et complété par une absence de forme (14 %). Par ailleurs, l'auteur note que l'apparition de la deuxième personne dans le système de Léonard se situe à 2;04 ; la forme de troisième personne étant utilisée dès 1;11 comme auto-désignation mais aussi pour désigner les autres. Ces observations peuvent d'ores et déjà être mises en relation avec les études acquisitionnelles génératives dans la mesure notamment où cette « absence de forme » fait partie des thématiques très largement traitées par la littérature générative sous le nom de « phénomène du sujet nul », notamment dans le cadre des Principes et des Paramètres³. Par ailleurs, la concomitance de deux items traditionnellement appelés « pronoms » – *moi* et *je* – montre déjà que la notion de sujet syntaxique n'est probablement pas si simple à cerner dès lors que l'on s'occupe d'oralité.

Pour le moment, nous avons fixé la frontière basse de la tranche d'âge pertinente pour notre problématique aux alentours de 2;9. Cette décision marque le début d'une longue aventure au cours de laquelle nous avons encore bon nombre d'autres

¹ Si l'on excepte les 3 % résiduels de *il* à 2;11 qui correspondent à une persistance liée à la description par l'enfant de sa propre image spéculaire (en l'occurrence, il s'agit des vidéos).

² Néanmoins, ces âges ne sont utilisés qu'à titre indicatif dans la mesure où chaque enfant possède ses propres parcours et rythme d'acquisition.

³ Cf. Chapitre II pour le détail.

décisions à prendre¹. Mais pour l'heure, tournons-vous vers notre deuxième thématique : la définition de la notion de sujet.

II. LA NOTION DE SUJET

Postuler qu'il y a syntaxe lors de la production des premiers énoncés enfantins revient à postuler que les éléments présents dans le continuum phonique présentent des relations fonctionnelles. Celles-ci sont diverses : certaines s'exercent au niveau de la phrase – ce sont les fonctions primaires, telles que le sujet ou le complément d'objet –, alors que d'autres s'exercent au niveau du syntagme – il s'agit alors des fonctions secondaires, telles que le complément du nom ou l'épithète. Toutes ces fonctions grammaticales font partie de notions qui appartiennent à notre fonds commun de connaissances dans la mesure où elles nous ont été décrites à des fins d'analyse grammaticale lors de nos premières années passées à l'école élémentaire. Nous pouvons ainsi trouver dans un cahier d'élève du Cours Élémentaire 1² : « le sujet est la personne, l'animal ou la chose qui fait l'action exprimée par le verbe ; par exemple : *les étoiles brillaient dans le ciel* ». Par ailleurs, la fonction sujet est considérée comme étant « la plus ancienne et aussi la plus stable : on la retrouve dans toutes les grammaires »³, notamment parce que celle-ci, à la différence des autres fonctions⁴, prend ses racines dans la réflexion sur les parties du discours chez les philosophes grecs de l'Antiquité, à commencer par Platon :

« Nous avons en effet, pour nous exprimer, par la voix, à propos de l'être, quelque chose comme deux genres de signes ; ce qui est le signe des actions nous l'appelons *rhêma* ; quant aux sujets qui les accomplissent, le signe vocal qui s'y applique est *onoma*. »⁵

A priori, il s'agit donc d'une notion simple et consensuelle :

¹ Les autres modalités de notre corpus sont définies au Chapitre III.

² Année scolaire 2003-2004, école élémentaire des Alpes Maritimes.

³ Gary-Prieur (1985:96).

⁴ Le terme « complément d'objet » par exemple apparaît chez les grammairiens à partir de la fin du XVII^e siècle (notamment chez Beauzée) en remplacement de la terminologie « cas régime ».

⁵ Platon [1920-1964]:261c-262a.

« Il existe un principe fondamental en grammaire traditionnelle, que l'on retrouve dans une grande partie de la théorie syntaxique moderne, selon lequel toute phrase déclarative simple est composée de deux constituants principaux obligatoires, un sujet et un prédicat [...]. [...] considérons la phrase *Jean assassina Pierre* [...]. Le sujet est *Jean*, le prédicat est *assassina Pierre* [...]. »¹

Néanmoins, les difficultés émergent rapidement et en nombre dès que l'on s'attarde davantage sur cette notion. Le nombre important d'articles et de colloques ayant pour thématique cette fonction grammaticale illustre d'ailleurs le fait qu'il y a matière à recherches². Ainsi, la problématique semble pouvoir se définir à partir de deux thématiques : d'une part, la notion de sujet ne semble pas pouvoir se contenter d'une définition strictement syntaxique et, d'autre part, toute analyse grammaticale en termes de fonction implique inévitablement une réflexion quant à la nature des éléments considérés.

1. Une fonction à la croisée des chemins

Dans le cadre de la grammaire générative, la notion de sujet est clairement définie dès Chomsky (1965) comme étant « intrinsèquement relationnelle » : « Nous disons [...] le *Sujet-de* la phrase (et non pas le Sujet) »³. La question subséquente est de définir le constituant avec lequel il est en relation. Cette définition n'est pas si évidente et la théorie générative a évolué sur cette question : ainsi on parle de sujet de la phrase dans les années 1960, puis cette formulation évolue dans les modèles suivants pour devenir sujet du syntagme verbal dans la théorie actuelle⁴. Nous pouvons ainsi dégager un critère stable qui caractérise cette notion : ce critère est bien relationnel dans la mesure où, pour qu'il y ait un sujet dans une phrase, il faut qu'il y ait également un verbe. Néanmoins, cette caractéristique s'applique également à d'autres fonctions primaires, telles que celle de complément d'objet ou de complément d'agent. Elle est donc insuffisante. La syntaxe prend également en charge la question de l'ordre des constituants. Les langues sont ainsi notamment inventoriées en fonction de l'ordre relatif qui est établi entre le sujet S, le verbe V et l'objet O. L'ordre canonique du français est notamment défini comme étant SVO

¹ Lyons (1970:256).

² Cf. par exemple Lagane (1969), Yaguello (1994), McCloskey (1997) ou Merle (2003).

³ Chomsky (1965:100).

⁴ Nous détaillons et complétons ces notions dans le Chapitre II.

avec un sujet placé à gauche du verbe mais, de nouveau, cette analyse n'est pas si simple et de nombreuses recherches débattent de cette thématique à travers la notion de déplacement de ce constituant par rapport au verbe. Une phrase telle que *est venue l'idée de lancer une gamme*¹ illustre par exemple ce propos. Par ailleurs, la théorie syntaxique ne se limite pas à cette définition relationnelle dans la mesure où elle utilise également des caractéristiques sémantiques, telles que les rôles sémantiques, des caractéristiques morphologiques, telles que les cas, ainsi que des caractéristiques pragmatiques, telles que la notion de topique.

Le type de définition que l'on trouve dans les manuels scolaires et que nous avons illustré en premier lieu montre d'ailleurs que la première caractéristique qui est donnée de la notion de sujet n'est pas syntaxique mais sémantique dans la mesure où, dans ce type d'approche, (i) la notion de sujet recouvre la notion d'agent du procès et (ii) le verbe est défini en termes d'action. Il est évident que les limites de cette approche sont visibles très rapidement, l'agent n'étant pas le seul rôle thématique possible et le verbe ne véhiculant pas systématiquement une action, comme le montre l'énoncé *ce bébé a chaud*. D'autres définitions évoquent alors des critères morphologiques pour définir le sujet. Ainsi pour des langues à flexions casuelles comme le latin ou l'allemand, le sujet peut être considéré comme étant le constituant qui porte le cas nominatif. Mais cette définition trouve également rapidement ses limites dans la mesure où, d'une part, le nominatif peut être associé à d'autres fonctions, telles que l'attribut du sujet et, d'autre part, les sujets peuvent porter un cas autre que le nominatif dans d'autres langues, comme l'ablatif, l'accusatif² ou l'ergatif³ entre autres. La seconde caractéristique morphologique envisagée est l'accord : « Ce mot [sujet] dénote la fonction assumée par le terme ou le membre qui confère au verbe ses catégories de personne, de nombre et éventuellement de genre »⁴. Mais cette approche ne couvre pas non plus tous les cas de figure. Que faire en effet de *voyager est formateur* ou *que tu viennes me ferait plaisir* ?⁵

¹ Interview télévisée sur TF1, 23/05/09.

² En latin, le sujet des subordonnées participiales est à l'ablatif et celui des subordonnées infinitives à l'accusatif. Cf. Touratier (2003).

³ Cf. Nash (1997) sur le basque par exemple.

⁴ Wagner & Pinchon (1962:28).

⁵ Nous repoussons le traitement de la question du topique au Chapitre II, Section III.5.2 dans la mesure où l'analyse qui en est proposée dans le cadre de la grammaire générative est liée aux projections fonctionnelles dont nous détaillons le développement dans le Chapitre II, Section III.5.1.

En bref, quel que soit le domaine envisagé ou le cadre théorique adopté, la notion de sujet est source de réflexions, discussions, voire controverses. Grevisse & Goosse (2008) préconisent ainsi que « Le plus sage est sans doute de considérer la notion de sujet comme une espèce de postulat, et de fournir seulement des moyens de l'identifier »¹. Pour ce qui est de la grammaire générative, Svenonius (2002) résume la situation ainsi :

“The subject occupies a precarious position in generative linguistics; on the one hand it is an indispensable concept at a descriptive level, and is accorded basic status as a primitive notion in some frameworks (along with other grammatical relations); on the other hand it has defeated all attempts at a cross-linguistically valid definition, and a substantial part of the field takes it to be no more than a descriptive level for an epiphenomenal collection of properties [...]. In sum, the notion of subject is a multifaceted one. In the Government-Binding, Principles and Parameters, and Minimalist frameworks, subjecthood is generally a descriptive label, without theoretical status.”²

Le sujet, en tant que fonction grammaticale, possède donc un statut paradoxal : il est extrêmement prégnant tout en restant relativement insaisissable. Voyons maintenant si la nature des éléments sujets peut nous aider à limiter nos investigations.

2. La nature de ses éléments

2.1 Une grande diversité

Les constituants traditionnellement décrits comme sujets en français sont, la plupart du temps, soit nominaux soit pronominaux, comme le montrent les exemples de Grevisse & Goosse (2008) ci-dessous :

¹ Grevisse & Goosse (2008:248).

² Svenonius (2002:3-5) : « Le sujet occupe une position instable en linguistique générative ; d'une part, il représente un concept indispensable au niveau descriptif et on lui accorde un statut basique en tant que notion primitive dans certains cadres de travail (aux côtés d'autres relations grammaticales) ; d'autre part, il a mis à mal toutes les tentatives d'une définition inter-linguistique valide et une bonne part des chercheurs le considère uniquement comme une étiquette sous laquelle on peut regrouper un ensemble secondaires de propriétés [...]. En résumé, la notion de sujet a plusieurs facettes. Dans les cadres GB, P&P et PM, elle est généralement descriptive et n'a pas de statut théorique. »

(4) Exemples de sujets du français¹ :

- (a) *Le professeur* écrit au tableau.
- (b) *Je* travaille.

Néanmoins, nos exemples de la section précédente ont déjà illustré le fait que d'autres entités, telles que les verbes à l'infinitif ou les propositions, peuvent également occuper cette fonction. Ainsi, l'étendue des possibilités en termes de nature de ce constituant est vaste et, par conséquent, pas plus définitoire que les critères précédemment mentionnés. Cependant, un trait se dégage tout de même de ces quelques exemples : dans tous ces cas de figure, un seul constituant peut être analysé comme le sujet dans ces phrases.

L'étude devient néanmoins rapidement moins consensuelle lorsque l'on est face à des énoncés oraux, tels que ceux de Léonard mentionnés précédemment ou ceux proposés en (5) ci-dessous :

(5) Autres exemples du français issus de notre corpus enfantin :

- (a) ANT: *et l(e) papillon i(l) vole* . (VII.29a, 323)
- (b) EVA: *elles sont là les cartes* . (VII.29b, 415)
- (c) EKT: *moi je veux dessiner* . (VI.26e, 221)
- (d) LUS: *nan regarde j' ai la souris moi* . (IX.34a, 1247)

En effet, dans ces exemples, tous les sujets sont représentés par deux éléments distincts : un élément traditionnellement appelé pronom personnel sujet, dit atone, clitique ou conjoint et un deuxième constituant qui, souvent, est soit un pronom dit tonique ou disjoint, soit un syntagme nominal. Ce type de configurations porte un grand nombre de noms différents dans les grammaires du français. Dans Grevisse & Goosse (2008) apparaissent les termes de « reprise »², « détachement »³ et divers types de « redondances »⁴ ; par ailleurs, on trouve également la notion d'« apposition »⁵, ou encore de « redoublement » ou de « dislocation »⁶. Les appellations sont donc variées, ce qui traduit une diversité dans les approches proposées. Globalement, trois possibilités sont envisagées quant à l'élément qui

¹ Grevisse & Goosse (2008:248).

² Grevisse & Goosse (2008:257).

³ Grevisse & Goosse (2008:258).

⁴ Grevisse & Goosse (2008:462-467).

⁵ Bonnard (1997:143).

⁶ Denis & Sancier-Chateau (1994:406).

exerce la fonction sujet : (i) des deux éléments, seul le pronom clitique est considéré comme sujet, l'autre élément est alors relégué en dehors de la valence verbale¹ ; (ii) l'élément clitique n'est qu'une simple « reprise formelle » du véritable sujet grammatical qui est donc représenté par l'autre constituant² ; (iii) les deux éléments forment ensemble le sujet³. Il n'y a donc pas davantage de consensus lorsque l'on s'intéresse à la nature du sujet ; cette question de l'analyse de la cooccurrence d'un pronom clitique et d'un syntagme co-référencés en français représente d'ailleurs un problème central dans notre propre problématique. La diversité des points de vue dans ce domaine provient essentiellement de l'existence de deux autres débats récurrents en linguistique française depuis plusieurs dizaines d'années : la définition du statut des pronoms clitiques et la caractérisation des différentes variétés du français. Commençons par nous arrêter sur cette notion de pronom clitique.

2.2 La délicate question du « pronom »

« PRONOM : [...] ; lat. *pronomen*, de *pro* « à la place de » et *nomen* « nom ». Mot grammatical qui sert à représenter un nom de sens précis déjà employé à un autre endroit du contexte ou qui joue le rôle d'un nom absent, généralement avec une nuance d'indétermination. »⁴

On enseigne aux enfants de l'école primaire à manipuler des termes issus de la grammaire traditionnelle, tels que *nom*, *pronom*, *verbe*, *adjectif*, etc., pour désigner des catégories grammaticales. À certains termes, comme *nom* ou *verbe*, il semble généralement possible, intuitivement tout au moins, d'attribuer des contours (même si ceux-ci peuvent bien sûr rester plus ou moins vagues en fonction de l'intuition et de la scolarisation de chacun). À l'opposé, il apparaît très vite que pour d'autres termes, tels qu'*adjectif*, la tâche se révèle nettement plus complexe, notamment à cause de l'hétérogénéité même de ces catégories grammaticales et/ou à cause de l'étymologie du terme. Ainsi, il semble que le terme *pronom* fasse partie de cette deuxième classe d'« étiquettes » dans la mesure où il s'agit d'un terme qui est tout à

¹ Cf. par exemple Kayne (1975).

² Cf. par exemple Gross (1975) qui qualifie les pronoms de « particules préverbaux » ou Creissels (1995) qui parle d'« indices pronominaux ».

³ Cf. par exemple Bauche (1928), cité par Grevisse & Goosse (2008:845). D'après l'auteur, cette configuration relève du « langage populaire » et implique qu'il n'y a aucune pause entre les deux éléments (ou aucune virgule lorsqu'il s'agit d'écrit). Kayne (2000) propose ce genre d'analyse pour les objets.

⁴ Robert *et al.* (2002:2092).

la fois très familier mais qui renvoie à une classe d'éléments très hétérogène. À tel point que Moignet (1965) par exemple soulève la question de la réalité linguistique d'une telle catégorie : ne serait-elle pas « une catégorie fourre-tout »¹ ? D'autre part, l'étymologie et la définition traditionnelle du terme peuvent être réductrices. « À la place d'un nom » en effet ne se vérifie pas toujours, loin de là. Ainsi, les travaux fondateurs de Benveniste (1946, 1956, 1958) dans le domaine de l'énonciation éclairent ce terme *pronom* différemment. Les pronoms personnels de première et deuxième personnes – *je* et *tu* – ne remplacent ainsi jamais un nom. Seuls les pronoms de troisième personne peuvent en effet remplir ce rôle². Que dire également du pronom explétif *il* que l'on trouve associé aux verbes météorologiques, tels que *il pleut*, *il neige*, et aux verbes défectifs, tels que *il faut* ? Ce pronom ne remplace absolument pas un nom, encore moins « un nom de sens précis ». Se pose alors un problème de terminologie sur lequel nous reviendrons à la suite de l'analyse que nous proposons de ces éléments³.

2.3 Les pronoms personnels du français moderne

Les grammaires posent l'existence de deux séries de pronoms personnels qui peuvent occuper la fonction sujet en français moderne : une série atone et une série tonique. Ces deux séries, issues du latin, sont détaillées dans le Tableau 1 ci-après.

¹ Cf. Moignet (1965:9) : « Le sentiment commun et la grammaire traditionnelle reconnaissent l'existence d'une catégorie grammaticale du pronom, mais, d'une part, ils y mettent tant de choses différentes, et, d'autre part, en excluent tant d'autres manifestement liées aux premières (nous pensons notamment à l'article [...]) qu'ils ont fait naître dans la pensée de certains l'idée que la catégorie *pronom* était artificielle, que le pronom, en tant que tel, ne correspondait à aucune réalité linguistique : quel dénominateur commun peut convenir à *je*, *celui-ci*, *leur*, *quoi*, *tout*, *dont*, *on*, etc. ? Le pronom serait une catégorie fourre-tout [...]. »

² Nous revenons sur le détail des travaux de Benveniste en matière de traits lors de notre analyse dans le Chapitre IV.

³ Cf. Chapitre V.

Personnes		Formes atones	Formes toniques
Singulier	1	<i>je</i>	<i>moi</i>
	2	<i>tu</i>	<i>toi</i>
	3 masculin	<i>il</i>	<i>lui</i>
	3 explétif	<i>il</i>	-
	3 féminin	<i>elle</i>	<i>elle</i>
	3 indéfini	<i>on</i>	-
Pluriel	1	<i>nous</i>	<i>nous</i>
	2	<i>vous</i>	<i>vous</i>
	3 masculin	<i>ils</i>	<i>eux</i>
	3 féminin	<i>elles</i>	<i>elles</i>

Tableau 1 : les pronoms personnels sujets en français moderne

D'après Moignet (1965), « La grande innovation [de l'ancien français], par rapport au latin, même tardif, consiste dans l'invention du pronom proprement dit, c'est-à-dire d'un existentiel différent du nom de la personne et non prédicatif »¹. Voyons donc précisément quand et comment ces constituants pronominaux émergent dans l'histoire du français.

2.3.1 Les pronoms atones

La série atone vient d' « une série de mots traditionnellement appelés « pronoms personnels » [...]. *Ego, tu, nos, uos* sont des noms des personnes [...] qui sont parfois apposés au verbe ; ils sont prédicatifs comme *lui* et n'en constituent nullement, au départ, le sujet »². Ces pronoms toniques, employés au nominatif, marquent alors une volonté de focaliser l'attention sur ce constituant mais sans pour autant que ce dernier soit considéré comme le sujet grammatical de l'énoncé. Pendant la période classique, Moignet (1965) les décrit d'ailleurs comme « hors phrase comme pouvait l'être un substantif au vocatif »³. Mais cela ne concerne que les formes de première et de deuxième personnes (singulier et pluriel) car le paradigme des pronoms nominatifs ne comporte pas encore de formes de troisième personne. Celles-ci sont intégrées, pendant la période postclassique, à la série des pronoms nominatifs par le biais du démonstratif *ille* « celui-là ». Cependant, Moignet (1965), se référant à des statistiques effectuées par Pei (1932) sur des documents du VIII^e siècle, précise que la présence de ces pronoms est largement minoritaire à cette

¹ Moignet (1965:108).

² Moignet (1965:40).

³ Moignet (1965:41).

époque¹ et que ce n'est qu'à partir de l'ancien français qu'elle se généralise peu à peu :

« L'équilibre de la présence et de l'absence est acquis en subordonnée dès la *Chanson de Roland*, alors qu'il ne sera atteint en principale qu'au début du XIII^e siècle, date à laquelle la présence du pronom en subordonnée est la règle presque générale [...]. S'il n'en est pas de même dans les indépendantes et principales, c'est qu'elles commencent avec prédilection par un régime, suivi immédiatement du verbe, ce qui entraîne, soit la postposition du pronom sujet, soit, bien plus souvent aux origines, son absence : en effet, tout se passe comme si, la personne restant assignée au verbe, il était jugé superflu de la signifier immédiatement après lui par un désigné de la même personne. »²

Le lien entre la présence croissante de ces pronoms et l'affaiblissement du caractère distinctif des suffixes verbaux est ainsi explicité : « L'histoire du pronom sujet est celle de l'abolition progressive de l'incidence purement interne dans le verbe »³ ; processus que l'auteur qualifie de « déflexivité »⁴. Pour l'auteur, c'est « l'avènement des déflexifs sujets qui a permis [...] la neutralisation des finales verbales »⁵. Cependant, l'hypothèse inverse (selon laquelle la perte des suffixes distinctifs du latin est à l'origine de la présence du pronom) est généralement qualifiée de plus « traditionnelle »⁶. Nous verrons au Chapitre II que ce lien entre pronoms nominatifs et suffixation verbale est une piste également largement exploitée dans le cadre de la grammaire générative. Une troisième voie consiste à poser que les deux phénomènes n'ont pas de lien direct mais qu'ils se produisent simplement de façon parallèle⁷. Néanmoins, quelle que soit la conjecture, la comparaison des paradigmes du verbe *aimer* au présent de l'indicatif en latin et en français moderne, présentée dans le Tableau 2 ci-dessous, illustre que la perte de la suffixation verbale distinctive en elle-même ne fait aucun doute : toutes les formes du latin sont distinctives alors que

¹ Moignet (1965:43) renvoie aux travaux de Pei (1932:381) sur les textes du recueil de Tardif (documents du VIII^e siècle du nord de la Gaule) qui mettent en évidence 375 cas d'omissions pour 63 emplois, soit un taux de présence d'environ 14 %.

² Moignet (1965:91-92).

³ Moignet (1965:96).

⁴ Faisant ainsi référence aux travaux de Guillaume (1964).

⁵ Moignet (1965:109-110).

⁶ Cf. par exemple Foulet (1930).

⁷ Cf. par exemple Adams (1987).

les verbes du premier groupe en français moderne ne possèdent plus que trois formes phoniques différentes. En effet, cette richesse morphologique du latin, encore perceptible en ancien français, disparaît à l'époque du moyen français.

Personnes		Latin	Français moderne
Singulier	1	amo	ɛm
	2	amas	ɛm
	3	amat	ɛm
Pluriel	1	amamus	ɛmɔ̃
	2	amatis	ɛme
	3	amant	ɛm

Tableau 2 : les paradigmes du présent de l'indicatif du verbe *aimer* en latin et en français moderne (API)

Par ailleurs, cette série d'hypothèses est controversée. En effet, Kaiser (2009) nous rappelle que, « dans la plupart des études diachroniques du français, au moins quatre développements concomitants sont perçus comme liés à la perte du sujet nul »¹. Il ne s'agirait donc pas d'une relation univoque entre un phénomène et un autre mais plutôt d'un faisceau de phénomènes concomitants qui se produisent en ancien français et en moyen français. Il s'agit ainsi non seulement de la perte de la richesse morphologique suffixale mais également de la perte concomitante du phénomène V2², de l'émergence des pronoms sujets *clitiques* et de la réalisation des pronoms explétifs³. En effet, concernant l'évolution phonologique de ces pronoms, jusqu'en ancien français, ceux-ci apparaissent sous une forme atone, uniquement lorsqu'ils sont postposés au verbe ; *a contrario*, leur utilisation préverbale est donc d'abord tonique. La série atone des pronoms personnels du français est donc enclitique avant de progressivement devenir proclitique à partir de l'ancien français⁴. C'est d'ailleurs dans ce cadre proclitique qu'apparaît le pronom atone *on* (issu du nom *hom*) qui, bien que traditionnellement considéré uniquement comme pronom indéfini, est à présent

¹ Kaiser (2009:138) : “According to most diachronic studies of French, there are at least four concomitant developments which are held to be connected with the loss of the null subject property in French”.

² Caractéristique selon laquelle les verbes fléchis doivent systématiquement occuper la deuxième position dans la proposition, comme en allemand moderne par exemple. Il est généralement postulé que le français perd cette caractéristique en ancien et moyen français mais il est également proposé que le français médiéval n'ait pas été une langue V2 (cf. Kaiser (2002)). Nous ne rentrerons pas dans ce débat qui dépasse les limites de l'étude du sujet enfantin en français moderne.

³ Pronoms non-référentiels, comme *il* dans *il pleut*. Mais cf. Zimmermann (2009) pour une discussion sur une telle corrélation.

⁴ Cf. Kaiser (2009:143-145) pour le détail et les références.

généralement intégré dans la série des pronoms personnels du français moderne¹. Le pronom explétif *il* vient également s'adjoindre à cette série à partir de l'ancien français². La liste des pronoms personnels atones du français moderne est ainsi complète à partir du moyen français. C'est d'ailleurs en moyen français que ces formes, devenues proclitiques, commencent à être utilisées conjointement à une seconde série de pronoms également issue du latin : les pronoms toniques.

2.3.2 *Les pronoms toniques*

En complément de la série des pronoms atones, l'ancien français dispose d'une seconde série de pronoms régimes qui, eux, sont toniques et prédicatifs ; leurs formes sont *moi, toi, lui, li, nos, vos, eus* et *eles*³. Mais, dès l'ancien français, « [...] la forme prédicative régime des personnels, [...], manifeste une tendance à s'étendre à certains secteurs de la fonction sujet »⁴. L'une des tournures émergentes concerne plus particulièrement notre étude. Il s'agit de la cooccurrence d'un pronom tonique et d'un pronom atone ; cette tournure est employée surtout à partir du moyen français (*cf.* (6) ci-dessous).

(6) Émergence de la cooccurrence d'un pronom tonique et d'un pronom atone⁵ :

- (a) Non fas *je moy*.
- (b) Mais, *moy, je* puis tout seul suffire.

Ainsi, il est généralement considéré que le système pronominal du moyen français correspond globalement à celui du français moderne :

« Aucun des faits de discours que nous avons examinés [...] ne témoigne que le système de langue des personnels ait été modifié depuis son institution en moyen français. [...]. C'est à peine si quelques innovations du français avancé – nous pensons surtout aux tours : *ton père il t'appelle, ma femme il est venu, nous on marche* – semblent vouloir amorcer une évolution. »⁶

¹ Cf. Moignet (1965:101): « Il apparaît donc qu'il est institué en langue comme signe de la personne animée indéterminée, qui précédemment ne figurait pas en système [...]. [...] il s'assimile en tout aux autres pronoms sujets ».

² Cf. Zimmermann (2009) par exemple pour le détail et les références.

³ Formes du XII^e siècle, données par Moignet (1965:49).

⁴ Moignet (1965:83).

⁵ Exemples cités par Moignet (1965:122) extraits de *Miracles de Nostre Dame*, III (1350:431) et *Moralité* (1426:625).

⁶ Moignet (1965:158).

Nous reviendrons sur cette notion de « français avancé » et nous attarderons sur les « tours » présentés par l'auteur dans la Section III de ce chapitre dans la mesure où ils sont présents dans une part importante de nos énoncés oraux modernes et que, comme mentionné précédemment, ils sont à la source de nombreux problèmes d'analyse.

2.4 Les autres constituants étudiés

Les autres éléments que nous souhaitons passer en revue sont de deux ordres. Il s'agit, d'une part, des pronoms non personnels que l'on peut trouver en position sujet et, d'autre part, des éléments qui relèvent d'autres catégories grammaticales que celles du nom et du pronom. En ce qui concerne la classe des pronoms, nous considérons plusieurs démonstratifs dans le cadre de notre étude, notamment les pronoms *ce*, qui est élide en *c'* devant voyelle, et *ça* « forme syncopée de *cela* »¹ dont les premières attestations remontent au XVI^e siècle². Ces formes sont également issues d'un pronom latin, *hoc* « ceci », qui renforcé d'une particule elle-même démonstrative *ecce* « voici » a donné la combinaison *ecce-hoc* réduite à *ce* pour la forme simple ; le pronom *ça*, forme syncopée de *cela*, est quant à lui issu de la combinaison de *ecce* et de *hac* « par ici »³. Mais, alors que le pronom *ce* est toujours atone en français moderne (*cf.* (a) et (d) ci-dessous), les utilisations différentes du pronom *ça*, également illustrées en (7) ci-dessous, montrent que cet élément appartient non seulement au paradigme des pronoms atones (*cf.* (b)) mais également à celui des pronoms toniques (*cf.* (c) et (d)). Ces faits sont confirmés par l'utilisation concomitante des deux formes, comme illustré dans l'énoncé (e) ci-dessous⁴. Par ailleurs, l'énoncé (b) illustre également l'utilisation de formes composées et toniques telles que *celui-là*.

¹ Robert *et al.* (2002:321).

² *Cf.* Grevisse & Goosse (2008:898) qui proposent qu'il y ait eu « une interférence avec *ça* adverbe ».

³ *Cf.* Dubois *et al.* (1964:133 et 109).

⁴ Auger (1994:295-297) par exemple arrive aux mêmes conclusions pour le français québécois.

(7) Les différents pronoms démonstratifs étudiés¹ :

- (a) ANT: *c' est un chat* . (XIII.48d, 96)
- (b) NOE: *et c(el)ui+là ça va où lui ?* (V.23a, 418)
- (c) CEL: *ça c' est Victor* . (XI.41d, 1198)
- (d) ZOE: *c' est les cadeaux ça* . (II.7, 1676)
- (e) ANT: *ça ça va ici* . (VI.24a, 464)

3. Vers les questions de registres de langue et de normes

Le tour d'horizon que nous avons effectué jusqu'à présent ne rend néanmoins pas encore compte de tous les cas de figure possibles concernant la description du sujet. En effet, les exemples en (8) ci-dessous indiquent qu'il faut considérer d'autres possibilités, notamment celles qui relèvent de l'absence de tout constituant réalisé.

(8) D'autres possibilités à prendre en compte² :

- (a) LUS: *Regardez ma p(e)tite [/] p(e)tite cabane* . (II.9b, 132)
- (b) XXX: *Défions-nous du sort, et prenons garde à nous*.³
- (c) CEL: *Faut pas la casser* . (II.9b, 913)
- (d) XXX: *Ne parviens pas à prendre sur moi de ne plus fumer*.⁴
- (e) MAX: *Sais pas* . (III.13d, 270)

Cependant, dans ce cadre, il convient de distinguer plusieurs types d'énoncés. En effet, certains sont considérés comme grammaticaux, quel que soit le registre de langue employé ; il s'agit des énoncés qui comportent un verbe à l'impératif (*cf.* (a) et (b) ci-dessus). *A contrario*, d'autres énoncés sont considérés comme grammaticaux uniquement lorsqu'ils relèvent d'un registre de langue particulier. Le premier registre en question est « la langue parlée, et surtout [...] la langue populaire »⁵ où le clitique explétif est « souvent omis »⁶ devant un verbe défectif comme *falloir* (*cf.* (c)). Il est alors intéressant de noter la différence de traitement qu'il y a entre ce type de verbes défectifs et les verbes météorologiques. En effet, ces derniers, bien qu'étant également défectifs, ne semblent pas autoriser l'absence du

¹ Ces énoncés sont extraits de notre corpus enfantin.

² Les énoncés introduits par le symbole XXX sont extraits de la littérature, les autres proviennent de notre corpus enfantin.

³ La Fontaine (1678).

⁴ Gide (1951), cité par Grevisse & Goosse (2008:253).

⁵ Grevisse & Goosse (2008:256).

⁶ Grevisse & Goosse (2008:256).

pronom explétif (**pleut*)¹. Le second registre est celui des « journaux intimes ou écrits similaires »² où « le pronom personnel de la 1^{ère} personne est omis assez fréquemment »³. Concernant les formes que peut prendre le sujet dans un énoncé, il semble donc qu'une variation importante soit liée au registre de langue employé. Cerner le registre dans lequel se situent les énoncés que nous étudions représenterait alors un préliminaire indispensable dans la mesure où une telle description permettrait de fixer les premiers contours de la définition du sujet dans ce registre particulier. Néanmoins, concernant la série d'exemples en (8) ci-dessus, une distinction est souvent établie entre les énoncés adultes et les énoncés enfantins (*cf.* (d) *vs* (e)), notamment dans le cadre de la grammaire générative. En effet, nous verrons au Chapitre II que l'hypothèse d'un « déficit » est souvent envisagée lorsqu'il s'agit d'énoncés enfantins, hypothèse qui n'est pas proposée dans le cadre de productions adultes. Une telle différenciation renvoie à plusieurs remarques importantes effectuées par Blanche-Benveniste (2000) sur la « tendance à toujours confondre fautif et parlé » et sur les *a priori* que les locuteurs ont vis-à-vis de cette notion de « faute », particulièrement vis-à-vis des enfants :

« Nos préjugés nous amènent à penser, souvent sans examen préalable, que certains parlent plus mal que d'autres. Nous remarquons plus facilement les fautes chez les jeunes enfants que chez les adultes [...]. »⁴

Ainsi, nous nous retrouvons face à une série de problématiques complexes. En effet, dans un premier temps, nous avons vu que la définition de la notion de sujet offre des possibilités nombreuses et variées, tant au niveau de la sémantique, que de la morphologie et de la syntaxe ; une définition unique et consensuelle de cette notion n'existe donc pas. Dans un second temps, il est constaté que les productions enfantines ne sont généralement pas considérées exactement de la même façon que les productions adultes, même lorsque les structures semblent similaires chez les deux catégories de locuteurs. Nous devons donc nous arrêter sur les deux

¹ Néanmoins, Haegeman (1997:134) évoque le fait que la phrase *A plu toute la journée* est acceptée par des locuteurs francophones comme pouvant faire partie du registre écrit dans les journaux intimes (le deuxième type de registre que nous considérons ci-après).

² Grevisse & Goosse (2008:253).

³ Grevisse & Goosse (2008:253) et *cf.* notamment les travaux de Haegeman (1997) pour une étude détaillée de ces omissions en français et en anglais.

⁴ Blanche-Benveniste (2000:37).

problématiques suivantes : la description des différents registres du français et le positionnement du français enfantin dans ce paysage.

III. QUEL(S) SUJET(S) POUR QUEL(S) FRANÇAIS ?

La notion de registre mentionnée ci-dessus fait l'objet de nombreuses formalisations dans la littérature, tant dans le cadre plutôt descriptif des grammaires traditionnelles, telles que celles de Wagner & Pinchon (1962), Bonnard (1997) ou Grevisse & Goosse (2008), que dans le cadre plus explicatif des études linguistiques. Dans un premier temps, nous nous tournons alors vers les différentes partitions que les linguistes ont posées ces dernières années afin de rendre compte des variations évoquées dans la section précédente.

1. La tradition des définitions contrastives

Quel que soit le cadre théorique dans lequel elles sont proposées, les descriptions qui sont faites des différentes variétés du français exploitent majoritairement une dichotomie¹. En effet, une opposition binaire est généralement posée entre un français formel, plutôt utilisé à l'écrit et un français moins formel, plutôt utilisé dans un registre oral ou apparenté, tel que celui des journaux intimes mentionné précédemment. Ainsi, Grevisse & Goosse (2008) par exemple distinguent le « français oral [qui] connaît de grandes diversités sociales et régionales, par exemple en matière de prononciation [et] l'écrit, diffusé par l'école surtout, [qui] présente une plus grande uniformité, grâce à l'orthographe notamment »². Mais existe-t-il des critères définitoires qui permettent de distinguer l'un de l'autre et l'un se définit-il par rapport à l'autre ?

1.1 Écrit vs oral

Il semble ainsi consensuel d'opposer deux types de français. La méthode généralement employée consiste alors à contraster les deux variétés. Lambrecht (1981) par exemple distingue le français « non standard » du français « standard »

¹ Dans ce chapitre, nous limitons notre discussion aux français pratiqués en France métropolitaine. Néanmoins, nous aborderons la question de la variation géographique au niveau international au Chapitre III dans le cadre de l'exploitation d'un corpus lorsque les données proviennent de pays francophones différents.

² Grevisse & Goosse (2008:13).

sur la base d'un certain nombre de traits distinctifs que nous reprenons en (9) ci-dessous¹.

(9) Les traits du français « non standard » d'après Lambrecht (1981) :

- (a) Élimination des déplacements dans les interrogations (l'inversion verbe-clitique et le mouvement de l'élément interrogatif n'ont pas lieu) :
où vas-tu ? > où tu vas ? > tu vas où ?
- (b) Élimination du clitique négatif *ne* :
il ne vient pas > il vient pas
- (c) Remplacement de *nous* par *on*.
- (d) Perte fréquente des « sujets grammaticaux » :
il faut > faut ; il s'agit de > s'agit de ; il y a > y a ; il paraît > paraît, etc.
- (e) Changements dans le système pronominal :
*cela > ça (ça me plaît) ; ce > ça (ça sera tout) ; il > ça (ça pleut) ;
ça générique (la mer ça me plaît).*
- (f) Érosion phonétique des pronoms clitiques :
tu as > t-as ; vous êtes > vs-êtes ; il me voit > i-m-voit ; etc.
- (g) Remplacement du pronom relatif *qui* par *que* :
*l'homme qui fait cela > l'homme que fait ça ;
c'est moi qui fais cela > c'est moi que je fais ça.*
- (h) Fréquence importante et baisse de l'aspect « marqué » des dislocations.

L'auteur qualifie toutefois cette partition de « représentation théorique créée pour les besoins de l'analyse » et admet qu'il est « impossible de tracer une frontière nette » entre les deux variétés². Néanmoins, un grand nombre de ces spécifications sont également retenues par d'autres linguistes. Blanche-Benveniste (2000) par exemple évalue ainsi le taux d'absence de *ne* à environ 95 % en français « parlé »³. Elle précise également que c'est la généralisation d'une tournure à toutes les classes sociales qui semble légitimer cette tournure dans son statut de critère distinctif. Ainsi, la particule négative *ne* est omise en français parlé « quels que soient les locuteurs »⁴. Blanche-Benveniste (2000) mentionne également le trait (a) ci-dessus en précisant que le français parlé privilégie la tournure interrogative *est-ce que* qui

¹ D'après Lambrecht (1981:6-7).

² Lambrecht (1981:5).

³ Blanche-Benveniste (2000:39).

⁴ Blanche-Benveniste (2000:39).

remplace l'inversion du sujet et du verbe dans la mesure où « bien des effets d'interrogation par post-position du sujet seraient ridicules », comme dans l'exemple *cours-je* ?¹ L'auteur rejoint également Lambrecht (1981) quant à l'utilisation du pronom *on* (cf. (c) ci-dessus). Néanmoins, elle mentionne d'autres caractéristiques, telles que l'utilisation de la tournure présentative *c'est...* de façon invariable, que le syntagme nominal soit au singulier ou au pluriel et qualifie l'ensemble de ces tournures de « fautes qui n'en sont plus » par opposition aux fautes « typantes » qui « agissent encore comme des marqueurs sociaux », telles que *aller au docteur* ou les régionalismes, tels que le passé surcomposé dans une proposition principale (*ça a eu payé*), qui est employé davantage dans le sud que dans le nord de la France². Les critères ainsi relevés définissent les contours et les tolérances de chaque variété. Il est intéressant de remarquer que l'impression générale selon laquelle les caractéristiques de ce français « non standard » ou « parlé » sont récentes est fautive. Ainsi, l'absence de la particule négative *ne* n'est ni moderne, ni typiquement enfantine. Blanche-Benveniste (2000) par exemple rapporte ainsi que le jeune roi Louis XIII « sautait souvent le *ne* de la négation, qu'il interrogeait avec *est-ce que* et qu'il omettait souvent certains subjonctifs »³. Il en va de même pour la tournure de la dislocation du sujet dont nous avons constaté l'émergence en moyen français. En effet, Blanche-Benveniste (2000) précise que :

« La tournure de dislocation, que les grammairiens appellent parfois « redondance syntaxique » (*mon père, il arrive*) a été attribuée tour à tour au manque de maturité enfantine, aux influences étrangères, aux régions, ou à une évolution récente de la langue. Il est pourtant facile de vérifier que tout le monde l'utilise et qu'elle est ancienne. Les grammairiens du XVII^e siècle discutaient déjà gravement de son statut. « Ne dites pas *Mon père, il est malade*, au lieu de *Mon père est malade* » écrivait le grammairien Chifflet en 1659, ce qui prouve qu'on le disait beaucoup. »⁴

¹ Blanche-Benveniste (2000:39).

² Blanche-Benveniste (2000:38-43).

³ Blanche-Benveniste (2000:44) d'après le journal tenu par le précepteur du roi. Cf. par exemple Foisil (1989).

⁴ Blanche-Benveniste (2000:37).

1.2 Les français « avancé », « parlé courant » et « très évolué »

Également dans ce cadre plus restreint de l'étude du sujet, Zribi-Hertz (1994) propose un découpage plus fin des différentes variétés du français. En effet, bien que son point de départ soit également une dichotomie, l'auteur instaure dans un deuxième temps une seconde partition au sein de l'une de ses variétés de base. La terminologie de Zribi-Hertz (1994) diffère également de celle que nous avons évoquée jusqu'à présent, ce qui suscite d'ailleurs certaines controverses¹. Dans un premier temps, l'auteur reprend donc une dichotomie du type « standard » vs « non standard » qu'elle reformule en termes de « français standard moderne » (FSM) vs « français avancé » (FA)². L'auteur définit ces deux variétés comme suit :

« Le français standard moderne [FSM] est la langue standard productive, c'est-à-dire débarrassée de ses archaïsmes, i.e. des tournures bien formées du français classique devenues improductives aujourd'hui. »³

« Le français avancé [FA] est celui qui échappe ou résiste, à des degrés divers, à la norme du 'bon usage' écrit, et qui révèle les mutations grammaticales non encore entérinées par l'Académie. »⁴

Dans un second temps, l'auteur distingue deux variétés à l'intérieur du français avancé : le « français parlé courant » (FPC) et le français très évolué (FTE), qu'elle définit comme suit :

« Le français parlé courant [FPC] est le moins avancé des dialectes avancés ; c'est la langue parlée familièrement, c'est-à-dire dans les échanges verbaux ordinaires, par les locuteurs qui maîtrisent par ailleurs le français standard. »⁵

Nous retrouvons dans cette variété la caractéristique (e) de Lambrecht (1981) sur les « changements dans le système pronominal » (FSM : *cela inquiète beaucoup Marie*

¹ Cf. Section 2 ci-après.

² En référence à Frei (1929:32) qui définit le français avancé comme étant « le français qui s'écarte de la norme traditionnelle ».

³ Zribi-Hertz (1994:136) : « Ainsi l'ellipse de *on* dans une structure coordonnée [*On retourna son portrait contre le mur et _ fit défense de prononcer son nom.* Sartre (1964:3)], est-elle jugée inacceptable par tous les grammairiens en français standard moderne. »

⁴ Zribi-Hertz (1994:136).

⁵ Zribi-Hertz (1994:137).

vs FPC : *ça inquiète beaucoup Marie* par exemple). La seconde sous-classe de français est définie comme suit par Zribi-Hertz (1994) :

« Le français très évolué [FTE] comprend toutes les formes perçues comme substandard ou dialectalement marquées par les locuteurs maîtrisant le français standard ; il inclut, notamment, le français dit populaire, [...] et le français considéré comme ‘enfantin’ [...]. »¹

C’est la première fois que nous croisons une référence au langage enfantin dans le cadre de ces typologies des variétés du français. Cette référence n’est toutefois pas directe dans la mesure où l’auteur apparente seulement une variété de français adulte à une façon que les enfants auraient de s’exprimer.

Par ailleurs, en (10) ci-dessous, nous schématisons les observations de l’auteur quant à l’utilisation des clitiques nominatifs dans ces quatre variétés de français :

(10) Les clitiques nominatifs et les strates du français, d’après la typologie de Zribi-Hertz (1994) :

- | | | |
|-----------|------------------------------|--|
| (a) FSM : | <i>Pierre/Il mange.</i> | Pas de cooccurrence |
| (b) FSM : | <i>Pierre, il mange.</i> | Cooccurrence par dislocation |
| (c) FPC : | <i>Pierre il mange.</i> | Cooccurrence par doublement |
| (d) FTE : | <i>Personne i(l) m’aime.</i> | Cooccurrence par doublement d’un élément quantifié |

Ainsi, nous croisons de nouveau le chemin de cette structure qui combine un clitique nominatif et un élément co-référencés. Dans le type (a) ci-dessus, on trouve soit un nom soit un pronom. Dans ce cadre, l’analyse du constituant sujet n’est pas problématique dans la mesure où un seul élément est présent ; lorsqu’il s’agit du clitique, ce constituant remplace alors vraiment un nom et son appellation de « pronom », utilisée au sens étymologique du terme, est pleinement justifiée. Le statut du pronom clitique dans ce contexte ne provoque donc aucun débat dans la mesure où il est consensuellement analysé comme un argument syntaxique du verbe. Cette structure, lorsqu’elle comprend un syntagme nominal, est généralement considérée comme relevant d’un français plutôt « standard » dans la mesure où il a été observé qu’elle n’est présente que dans 20 % des phrases orales adultes dans le

¹ Zribi-Hertz (1994:137).

cadre d'une conversation courante¹ et que sa représentation chez les jeunes enfants est nulle ou quasi-nulle². Le langage enfantin ne satisfait donc pas au critère syntaxique prototypique du français : SV(O), avec S = syntagme nominal. Par contre, les jeunes enfants utilisent très largement cette structure avec S = clitique. Une telle disparité devra donc trouver une explication dans le cours de notre analyse.

1.2.1 La dislocation

Globalement, le deuxième type d'énoncés présenté par Zribi-Hertz (1994) (et que nous reprenons en (11) ci-dessous) ne soulève pas non plus de polémiques quant au statut du clitique dans la mesure où une telle présentation typographique (avec une virgule entre le syntagme nominal et le clitique) laisse supposer qu'il y a une éventuelle pause légère entre les deux éléments à l'oral³.

(11) Cooccurrence par dislocation :

FSM : *Pierre, il mange.*

Il s'agit alors d'une dislocation et, dans ce contexte, le pronom clitique est également analysé en tant qu'argument syntaxique du verbe. Le syntagme nominal est donc « disloqué », c'est-à-dire qu'il n'est pas compris dans la valence du verbe ; en d'autres termes, il est considéré « hors phrase ». Nous avons vu que cette structure n'est ni particulièrement moderne ni spécifique à une variété de français. Quelques phrases extraites de la littérature française et reprises en (12) ci-dessous illustrent ce propos. Elles montrent également que la dislocation peut survenir à gauche ou à droite du noyau de la phrase.

¹ Cf. Blanche-Benveniste (2003:82-83). Dans ce cadre, les syntagmes nominaux portent alors une valeur plutôt générique comme dans les exemples produits par un mineur décrivant son travail : *le mineur doit nettoyer... le matériel se composait de...*

² Notre premier corpus exploité dans Palasis (2005) renvoie un taux de 0 % (cf. Chapitre II, Section I.3.4 pour plus de détails sur ce corpus). Notre nouveau corpus, dont les détails se trouvent au Chapitre III et en Annexe, renvoie un taux extrêmement bas (cf. Chapitre IV).

³ L'analyse de la dislocation n'est néanmoins pas complètement consensuelle. Cf. Chapitre II.

(12) Exemples de dislocations dans la littérature française :

- (a) *Toi, tu* vas rentrer dans ton appartement.¹
- (b) *Elle* me fit peur, *cette lettre*.²
- (c) *Mes deux frères et moi, nous* étions tout enfants.³
- (d) *Ils* approchaient de la rive, *les contrebandiers*.⁴

1.2.2 Le doublement

L'observation de l'énoncé (c), que nous reprenons en (13) ci-dessous, semble montrer une seule différence par rapport à l'énoncé précédent dans la mesure où ce troisième type d'énoncés, attribué au « français parlé courant » est présenté sans virgule entre le syntagme et le clitique :

(13) Cooccurrence par doublement :

FPC : *Pierre il mange.*

Si nous suivons le raisonnement posé pour l'énoncé analysé dans la section précédente, l'absence de la virgule devrait être corrélée à une absence de pause légère entre le syntagme et le noyau de la phrase. Cependant, cette différence prosodique entre les deux types d'énoncés n'est pas considérée comme définitoire par les phonéticiens eux-mêmes⁵. À ce titre, nous ne la retenons pas non plus dans le cadre de nos recherches⁶. Toutefois, contrairement aux apparences, il reste une différence entre les énoncés (b) et (c) présentés en (10) ci-dessus. En effet, leurs structures syntaxiques sont fondamentalement différentes dans la mesure où, pour le type (c), le statut du clitique change radicalement. Alors qu'en (a) et (b), le clitique est considéré comme un argument syntaxique, en (c), il ne ressortit plus à la syntaxe mais à la morphologie et est alors analysé comme un préfixe verbal. Ce statut morphologique du clitique laisse ainsi la voie libre à une analyse syntaxique du syntagme qui, subséquentement, est considéré comme l'argument sujet du verbe.

Néanmoins, cette hypothèse morphologique est loin de faire l'unanimité. Ainsi, certains linguistes ne reconnaissent que l'existence de la dislocation mentionnée

¹ Labiche (1844), cité par Grevisse & Goosse (2008:845).

² Vigny (1835), cité par Grevisse & Goosse (2008:258).

³ Hugo (1856), cité par Grevisse & Goosse (2008:846).

⁴ Loti (1897), cité par Grevisse & Goosse (2008:258).

⁵ Cf. par exemple Rossi (1999). Nous revenons sur cette notion de pause au Chapitre II, Section IV.1.2.2.

⁶ C'est la raison pour laquelle nos transcriptions ne comportent jamais de virgule.

précédemment¹ et d'autres, tels que Kayne (2000), admettent que « Bien qu'il soit écrit avec une virgule, il est possible que l'analyse de [l'exemple que nous reprenons en (a) ci-dessous] soit ambiguë entre une dislocation à gauche [...] et un doublement non disloqué du type [que nous reprenons en (b)] »².

(14) Les exemples de Kayne (2000) :

- (a) *Moi, je vois Marie.*
- (b) *Jean me connaît moi.*

Lorsque le doublement est reconnu aux côtés de la dislocation, certains critères sont alors proposés afin de distinguer les deux structures. Nous synthétisons les caractéristiques mises en avant par Cabredo Hofherr (2004) en (15) ci-dessous³ :

(15) Caractéristiques du doublement, d'après Cabredo Hofherr (2004) :

- (a) Le doublement ne véhicule pas de mise en relief de l'élément doublé alors que la structure avec dislocation véhicule ce type d'information⁴.
- (b) L'élément doublé n'a pas forcément un antécédent dans le contexte ; contrairement à la dislocation, il peut donc être indéfini ou quantifié⁵ :
Doublement : *Personne il a rien dit.*
Dislocation : **Personne, il a rien dit.*
- (c) Le doublement suit l'élément interrogatif alors que la dislocation le précède⁶ :
Doublement : *Pourquoi Jean il parle comme ça ?*
Dislocation : *Jean, pourquoi il parle comme ça ?*

Néanmoins, il nous semble qu'en ce qui concerne les variétés orales du français, la frontière entre les deux types de structures ne soit pas extrêmement nette. En effet,

¹ Cf. Brandi (1989:114) par exemple : « En français, [...], la cooccurrence de sujets lexicaux et de clitiques sujets n'est autorisée que dans des phrases avec une dislocation à gauche [*Jean, il parle vs *Jean il parle*]. »

² Kayne (2000:165 et 167).

³ Cf. Cabredo Hofherr (2004:105-106) qui reprend des exemples d'Auger (1994) et de Zribi-Hertz (1994).

⁴ D'après les travaux d'Ashby (1982) et Lambrecht (1984) qui indiquent que 70 % des syntagmes nominaux sont doublés en français parlé. Ce taux élevé exclut l'effet d'emphase. Kaiser (1994:133) suit également cette voie et note que l'emphase doit alors être marquée par un doublement supplémentaire (par exemple : *Jean lui il mange*, d'après Auger (1994)).

⁵ D'après des exemples de Zribi-Hertz (1994). Nous détaillons ces notions dans la section suivante.

⁶ Exemples de Hulk (1991).

Cabredo Hofherr (2004) reprend les propos de Blanche-Benveniste (2000) sur ce que cette dernière nomme « dislocation »¹, dans le cadre du « redoublement » :

(16) Comparaison des deux formulations :

- (a) Blanche-Benveniste (2000:37) : « La tournure de dislocation, que les grammairiens appellent parfois « redondance syntaxique » [...]. Les grammairiens du XVII^e siècle discutaient déjà gravement de son statut. « Ne dites pas *Mon père, il est malade*, au lieu de *Mon père est malade* » ».
- (b) Cabredo Hofherr (2004:105) : « [...] le redoublement systématique du sujet lexical : *Mon père il est malade.* »

1.2.3 La quantification

Le quatrième type d'énoncés proposé par Zribi-Hertz (1994) (que nous reprenons en (17) ci-dessous) fait de nouveau référence au doublement du clitique et du syntagme mais focalise la problématique sur un doublement bien particulier. En effet, il s'agit d'une configuration où le syntagme nominal est « quantifié » :

(17) Cooccurrence par doublement d'un élément quantifié :

FTE : *Personne i(l) m'aime.*

D'après Zribi-Hertz (1994), l'une des caractéristiques du « français très évolué » est de permettre une telle cooccurrence. Cette notion de quantification renvoie à la partition qui est établie au sein des différents déterminants du français quant à leur valeur quantifiante. Ainsi, sont considérés comme uniquement « quantifiants »², les articles indéfinis (*un, une, des*), les déterminants numériques (*un, deux, etc.*) et les déterminants indéfinis (*tout, tous, chaque, aucun, etc.*), auxquels s'ajoutent les pronoms indéfinis tels que *personne* ou *rien*. En ce qui concerne la cooccurrence d'un clitique nominatif et d'un syntagme quantifié, Zribi-Hertz (1994) établit alors une corrélation entre la variété de français employée et l'apparition de cette structure.

C'est ainsi que le statut des clitics nominatifs du français « parlé » ou « avancé » se retrouve au cœur de l'un des débats majeurs qui préoccupent les linguistes depuis plus de trente ans : ces clitics nominatifs sont-ils des arguments

¹ Terme que nous avons également cité dans la Section III.1.1 ci-dessus.

² Par opposition à « quantifiants et caractérisants » (déterminants possessifs, démonstratifs, interrogatifs et exclamatifs). Grevisse & Goosse (2008:738) remarquent que l'on peut trouver plusieurs termes différents : « quantifiants », « quantifieurs » ou « quantificateurs » selon les ouvrages.

syntaxiques ou des préfixes verbaux ? Nous reviendrons sur cette thématique complexe plusieurs fois dans le cours de nos travaux : dans le Chapitre II, afin d'exposer la problématique dans le cadre de la grammaire générative (notamment Section IV) puis dans le cours de notre analyse (Chapitre IV) et dans nos conclusions (Chapitre V) dans la mesure où nous reconsidérons ce débat à la lumière de nos données enfantines. Pour l'heure, nous reprenons quelques points qui prêtent à discussion dans le cadre de cette section sur les différentes variétés de français.

2. Comment concilier grammaire interne et variation

La typologie proposée par Zribi-Hertz (1994) fait l'objet de certaines critiques. En effet, De Cat (2007) par exemple, qui argumente en faveur d'une analyse syntaxique des clitiques, attaque la justification même d'une telle classification¹ :

“It seems to me that the motivation behind the definition of Advanced French stems essentially from the need to identify the variety/varieties of French for which a morphological analysis of subject clitics can be argued to hold.”²

L'auteur attribue une motivation similaire à la dichotomie proposée par Lambrecht (1981). Son raisonnement se base notamment sur une réflexion menée quant aux locuteurs potentiels d'une variété de langue :

“It seems reasonable to assume that whenever a group of speakers can be identified as using the variety in question, the identification of this variety is methodologically justified.”³

Ainsi, dans ce cadre, l'existence des variétés régionales est justifiée pour De Cat (2007). En revanche, elle précise que « la maîtrise du français standard moderne n'est pas un critère suffisant pour identifier les locuteurs du français informel »⁴. Suivant Côté (1999), De Cat (2007) énumère ainsi un certain nombre de critiques à

¹ De Cat (2007) signale que la notion de français « avancé » est également remise en question par Gadet (1998) et Côté (1999) dont elle reprend l'argumentation.

² De Cat (2007:9) : « Il me semble que la motivation derrière la définition du Français Avancé provient essentiellement du besoin d'identifier la(les) variété(s) de français compatible(s) avec une analyse morphologique des clitiques. »

³ De Cat (2007:7) : « Il semble raisonnable de supposer que, lorsqu'un groupe de locuteurs peut être identifié comme utilisant la variété en question, l'identification de cette variété est méthodologiquement justifiée. »

⁴ De Cat (2007:8).

l'égard de la classification de Zribi-Hertz (1994) notamment concernant le faible nombre d'exemples proposés et l'hétérogénéité des traits mis en avant afin de définir la notion de « français très évolué ».

2.1 La « diglossie francophone »

Zribi-Hertz elle-même revient sur cette problématique de l'hétérogénéité dans une réflexion plus récente sur ces questions de variations confrontées à la notion de « grammaire interne », telle que définie par les générativistes¹. Ainsi, Zribi-Hertz (2006) considère le conflit qu'il y a à vouloir décrire une grammaire interne unique du français alors que les phénomènes de variation chez un même locuteur peuvent être extrêmement divers. Dans le Tableau 3 ci-après, nous reprenons les « dimensions de la variation », telles que proposées par Zribi-Hertz (2006)².

Dimensions	Dichotomies	
Norme	standard	non standard
Medium	écrit	parlé
Diachronie	archaïque	productif, avancé
Sociologie	?bourgeois	populaire
Pragmatique	formel	informel
Géographique	?central	régional
Stylistique	soutenu, recherché, littéraire	vulgaire, relâché, châtié

Tableau 3 : les dimensions de la variation (Zribi-Hertz, 2006)

Dans le cadre de nos recherches sur l'acquisition, nous nous intéressons alors tout particulièrement à l'approche que l'auteur développe dans sa réflexion qui a pour point de départ l'enseignement du français tant langue maternelle que langue étrangère :

¹ La grammaire interne d'une langue (*I-language*) est l'ensemble des règles de cette langue ; elle comporte des principes universels et des paramètres dont les valeurs sont fixées par l'enfant au cours de son acquisition. Cf. Chomsky (1995a:6) et cf. notre Chapitre II pour le détail.

² Zribi-Hertz (2006:3).

« On admet volontiers qu'un locuteur bilingue a intériorisé deux grammaires différentes. On admettra de même assez facilement que des variétés régionales très éloignées d'une «même» langue – par exemple l'acadien et le francilien – sont générées par des grammaires distinctes. Mais lorsqu'il s'agit d'explicitier la compétence linguistique d'un francophone non bilingue, on fait l'hypothèse qu'il n'a intériorisé qu'une seule grammaire – *sa* grammaire idiolectale [...]. J'aimerais argumenter [...] que cette approche est incorrecte, et qu'il convient de distinguer crucialement la *variation* – dégagée de la *comparaison* des grammaires internes – de ce que je nommerai *diglossie*, qui concerne la structure même de la grammaire interne. »¹

Zribi-Hertz (2006) postule ainsi l'existence de deux grammaires internes chez les locuteurs francophones : une grammaire standard (GS) « qui incarne la norme du « bon usage » » et une grammaire dialectale (GD) « activée en situation de communication informelle »². Ces deux grammaires sont « sécantes » dans la mesure où, d'après l'auteur, certains énoncés sont générés par les deux grammaires à la fois. C'est ce que l'auteur intitule la « diglossie francophone »³. Les trois énoncés en (18) ci-dessous illustrent ce propos :

(18) La « diglossie francophone », d'après Zribi-Hertz (2006)⁴ :

- (a) GS : *à quelle heure le train arrive-t-il à Nice ?*
- (b) GD : *le train (il) arrive à Nice à quelle heure ?*
- (c) GS∩GD : *à quelle heure est-ce que le train arrive à Nice ?*

2.2 Instauration d'une chronologie des événements

Dans Palasis (2009), sur la base de l'observation de notre premier corpus complétée par les premières extractions effectuées sur notre deuxième ensemble de données, nous esquissons une hypothèse quant à l'existence de plusieurs grammaires chez les locuteurs francophones : une grammaire d'un français que nous appelons

¹ Zribi-Hertz (2006:4).

² Zribi-Hertz (2006:4).

³ Zribi-Hertz (2006:5).

⁴ Zribi-Hertz (2006:5).

« spontané » et une seconde grammaire d'un français « normé »¹. Ces deux grammaires sont activées l'une après l'autre au cours de l'enfance. Les clitiques nominatifs sont ainsi analysés comme des affixes morphologiques dans la première et comme des arguments syntaxiques dans la seconde. Nous postulons également que le français « parlé » adulte se trouve à l'intersection de ces deux systèmes et que le poids que prend une grammaire par rapport à l'autre dépend de facteurs éducatifs, sociaux, pragmatiques, etc. La conséquence pour les clitiques nominatifs est qu'ils possèdent alors un double statut chez les locuteurs du français : ils sont tantôt morphologiques, tantôt syntaxiques. Les objectifs de ce travail de recherches sont de tester cette hypothèse en la confrontant à l'ensemble de notre deuxième corpus et, si elle n'est pas falsifiée, d'en proposer une formalisation précise dans le cadre de la grammaire générative et notamment la théorie des Principes et des Paramètres.

Ainsi, il est important de rappeler que l'écrit ne précède pas l'oral ni au niveau individuel ni au niveau historique. L'écriture est certes un support qui permet de fixer la langue et de développer une réflexion à son sujet – à tel point que Auroux (1996) par exemple propose de situer le point de départ de la réflexion linguistique au moment de l'invention de l'écriture² – mais la langue écrite n'est pas la transposition graphique de la langue orale³ et la langue orale n'est aucunement le parent pauvre de la langue écrite. Il s'agit de deux systèmes différents qui privilégient chacun des structures particulières. Or, nous avons vu que la variété « parlée » du français est souvent décrite d'après la variété « normée ». Lambrecht (1981) par exemple, lorsqu'il décrit le français « non standard », utilise les termes d'« élimination » (de l'inversion, du clitique négatif), de « remplacement » (de *nous* par *on* ; *cela* par *ça* ; *qui* par *que*, etc.), de « perte » (des sujets grammaticaux), d'« érosion » (phonétique des pronoms). Ces termes suggèrent que le français non standard est analysé par rapport au français standard, ce qui implique que le français standard, au cours de cette démarche, reste la langue de référence lors de l'analyse. Si l'on ramène cette relation d'implication au niveau du locuteur, cela revient à supposer que le locuteur connaît le français standard et que, à partir de cette variété

¹ Nous explicitons et développons cette hypothèse, ainsi que notre définition du français « normé », dans le Chapitre V sur la base de l'observation de nos nouvelles données.

² Cf. Auroux (1996:60-61) : « [...] l'histoire nous montre incontestablement que dans les civilisations où il y eut apparition d'un savoir linguistique [...] celui-ci est indubitablement né *après* que l'écriture eut été une technique bien dominée. Pour l'histoire des connaissances linguistiques le *seuil de l'écriture* paraît donc fondamental. »

³ Sauf, bien sûr, dans le cas particulier d'une transcription.

de français, celui-ci applique un certain nombre de modifications afin de pratiquer le français non standard. Nous n'adoptons pas ce point de vue et considérons que, lorsque nous plaçons la problématique de la description des différentes variétés du français dans une perspective acquisitionnelle, il est beaucoup plus visible que ces descriptions rencontrent des problèmes.

Prenons par exemple la définition proposée par Jakubowicz & Rigaut (1997) qui travaillent dans le cadre de l'acquisition du français :

« Le premier [le français standard] correspond, *grosso modo*, au français écrit productif, le deuxième [le français parlé] est la langue utilisée familièrement, dans les échanges verbaux ordinaires, par les locuteurs qui maîtrisent par ailleurs le français standard. »¹

Or, le jeune enfant n'écrit pas ; pourtant il parle. Il ne démarre pas sa carrière de locuteur natif en produisant le pronom *nous*, la particule négative *ne*, etc. ; *a contrario*, il commence par utiliser le pronom *on* et par indiquer la négation grâce à l'adverbe négatif *pas*. Le jeune enfant ne maîtrise donc pas le « français standard ». C'est la raison pour laquelle nous suggérons une description des variétés du français qui est, en quelque sorte, inversée par rapport à ce qui est généralement proposé, notre critère de référence étant le parcours du locuteur et donc le développement de ses capacités et connaissances, et non le système standard de la langue.

Nos propos, tenus à partir de l'observation de données acquisitionnelles, rejoignent ceux de Zribi-Hertz (2006), élaborés dans un cadre didactique, tant au niveau de la coexistence de grammaires chez les locuteurs francophones qu'au niveau de la chronologie des événements puisque Zribi-Hertz (2006) postule que la grammaire dialectale est présente chez les enfants avant la grammaire standard ; nos objectifs également sont similaires :

« Pour enseigner le français langue maternelle, c'est-à-dire GS à des élèves locuteurs de GD, on devra s'attacher à expliciter les conditions d'activation de GS et GD [...]. »²

¹ Jakubowicz & Rigaut (1997:58).

² Zribi-Hertz (2006:20).

3. Comment définir le sujet dans ce cadre

Le prototype du sujet français est traditionnellement présenté dans les grammaires comme étant un syntagme nominal directement placé devant le verbe dont il est le sujet. Comme mentionné au tout début de la Section II.1, l'ordre canonique du français est donc considéré comme étant SV(O). Or, toutes les études qui portent sur le français oral indiquent que cette disposition ne fait pas partie de ce deuxième système¹.

3.1 L'importance du facteur prosodique

Blanche-Benveniste (2003) remarque que, d'une part, « la fonction sujet se réalise dans des formes grammaticales remarquablement variées » (ce que nous avons entraperçu dans la Section II ci-dessus) mais que, d'autre part, les enfants parviennent néanmoins à « identifier ces procédés » spontanément et rapidement². La prosodie est une explication envisagée. En effet, des études montrent que les adultes s'adressent d'une façon particulière à leur nouveau-né, en insistant notamment sur la mélodie des phrases. Ce type de langage fait partie de ce qui est appelé le L.A.E. (Langage Adressé aux Enfants). Veneziano (2000) par exemple le définit comme suit :

« [...] le LAE diffère par plusieurs des caractéristiques du langage utilisé entre locuteurs déjà compétents. [...] les mères s'adressent à leurs enfants dans un langage clairement articulé, produit avec un ton de la voix plus élevé et une intonation exagérée. Le LAE utilise un vocabulaire plus limité et est en général syntaxiquement plus simple : les énoncés de la mère contiennent moins de pronoms de troisième personne, sont plus courts, contiennent moins de formes verbales composées et moins de subordonnées. Du point de vue de son contenu, il porte le plus souvent sur le « ici et maintenant » [...]. »³

Jisa (2004) accorde également une grande importance à la prosodie dans le cadre de l'acquisition. Elle propose ainsi une illustration de Gadet (1990) dans laquelle

¹ Cf. par exemple François-Geiger (1974), Jeanjean (1980), Lambrecht (1987) ou Blanche-Benveniste (1994, 2003).

² Blanche-Benveniste (2003:88).

³ Veneziano (2000:246).

l'attribution des fonctions grammaticales aux syntagmes nominaux se fait grâce à l'intonation :

(19) Exemples de Gadet (1990) sur l'information apportée par la prosodie pour la phrase *Marie Nicole elle ne l'aime pas* :

- (a) *Marie_i / Nicole / elle_i ne l'aime pas.* « Marie n'aime pas Nicole »
- (b) *Marie / Nicole_i elle_i ne l'aime pas.* « Nicole n'aime pas Marie »

Jisa (2004) précise ainsi que ces structures ne peuvent pas être utilisées à l'écrit car « de telles distinctions prosodiques n'ont pas d'équivalents clairs en termes de ponctuation à l'écrit »¹.

3.2 Quelques caractéristiques du français oral adulte

Ainsi, Blanche-Benveniste (2003) et Jisa (2004) présentent des indications similaires quant à la notion de sujet en français oral adulte :

(20) Quelques caractéristiques des sujets en français oral adulte :

- (a) Les syntagmes nominaux préverbaux sont relativement peu fréquents puisqu'ils ne dépassent pas 20 % des énoncés produits².
- (b) L'introduction des nouveaux référents se fait souvent après le verbe fléchi, grâce à :
 - i) une tournure impersonnelle : *y a un homme*³, *ça m'étonnerait que...*⁴
 - ii) une fonction autre que celle de sujet : *hier j'ai vu un homme.*⁵
- (c) Le sujet grammatical est souvent :
 - i) un pronom clitique seul : *il m'a donné un bonbon*,⁶
 - ii) une « dislocation » : *ma sœur, elle était restée là-bas.*⁷

Dans le Tableau 4 ci-dessous, nous reprenons les taux proposés par Blanche-Benveniste (1994) pour l'un de ses corpus oraux adultes. Il s'agit d'un corpus

¹ Jisa (2004:143).

² Blanche-Benveniste (2003:83). Ce taux est élevé comparé au taux de 5,9 %, avancé précédemment (cf. Blanche-Benveniste (1994:90)).

³ Jisa (2004:144).

⁴ Blanche-Benveniste (2003:78).

⁵ Jisa (2004:144).

⁶ Jisa (2004:145). Tous clitiques confondus, Blanche-Benveniste (1994:87) rapporte un taux allant jusqu'à environ 80 % des énoncés.

⁷ Cf. Blanche-Benveniste (2003:77). Le taux le plus élevé de Blanche-Benveniste (1994) est de 12,8 % des énoncés.

élaboré à partir de l'enregistrement d'une stagiaire en pharmacie de vingt-cinq ans qui raconte son travail à une amie¹.

Types de sujets	Clitiques seuls	Dislocations	Syntagmes Nominaux
%	81,3	12,8	5,9

Tableau 4 : répartition des types de sujets oraux adultes

Dans ce corpus, les clitiques sont présents dans 94,1 % des énoncés. Serait-il alors simpliste de vouloir associer la notion de sujet oral à celle de clitique nominatif? Notre analyse (au Chapitre IV) dira si ce type de répartition apparaît également chez les jeunes enfants. Pour le moment, tournons-nous vers la grammaire générative et ses aspects théoriques quant aux traitements qu'elle propose du processus acquisitionnel et de la notion de sujet.

¹ Blanche-Benveniste (1994:80).

CHAPITRE II :

L'ACQUISITION DU LANGAGE

ET LA THÉORIE LINGUISTIQUE

Le regard que porte l'être humain sur ses semblables lorsque ceux-ci se trouvent encore dans leur enfance a connu d'énormes bouleversements au cours de notre histoire. En effet, de simple passage imposé et malheureusement obligatoire en vue de devenir, enfin, un être humain à part entière, l'enfance a progressivement évolué vers un statut de moins en moins négatif et de plus en plus digne d'attention(s)¹. Mais les modifications les plus profondes de ce regard remontent au tout récent XX^e siècle. La Déclaration des Droits de l'Enfant ne date-t-elle pas de 1959 ? Quoiqu'il en soit, la modification de ce statut s'est alors fait sentir à de nombreux niveaux. Ainsi, au cours de ces dernières décennies, on s'est préoccupé de l'enfance en tant que telle en médecine et en psychologie notamment². La reconnaissance d'un développement tant physique que psychique étant acquise, il est alors intéressant de s'interroger sur un troisième domaine central : celui de la cognition. Deux points au moins prêtent alors à discussion : le stade initial du développement cognitif de l'enfant et le moteur de ce développement ; ou en d'autres termes, ce que sait un enfant à la naissance et comment opère le mécanisme de l'apprentissage des connaissances. Ces questions nous intéressent au premier chef car elles se posent sous les mêmes termes pour le domaine du langage et de son acquisition. S'ajoute alors une troisième interrogation, et non des moindres : le langage est-il un domaine cognitif comme les autres ou représente-t-il un domaine particulier ? Certaines de ces questions ont été posées bien avant les années 1900 ; d'autres ont suscité des débats plus récents. Cependant, elles ont toutes en commun le fait qu'il a fallu attendre le

¹ Cf. Bideaud *et al.* (1993:3-23) par exemple pour un bref historique de ce statut.

² Avec l'avènement de la pédiatrie et de la psychologie du développement de l'enfant.

XX^e siècle et le développement des sciences du langage pour qu'elles soient considérées d'une façon globale et que des réponses soient apportées dans le cadre de théories linguistiques.

Nous avons choisi le cadre de la grammaire générative pour conduire nos travaux. Dans ce chapitre, nous exposerons donc tout d'abord les fondements de cette théorie dans la mesure où ils motivent notre choix. Cependant, nous discuterons également des limites de cette approche, notamment en ce qui concerne l'utilisation de données ; cette discussion nous amènera alors à proposer un cadre de travail certes centré sur le générativisme mais dont certains préceptes ne sont pas sans rappeler ceux de la dialectologie. Dans les deuxième et troisième parties du chapitre, nous détaillerons ensuite les outils théoriques mis en place par la grammaire générative pour rendre compte du processus acquisitionnel des langues chez les enfants et afin de décrire et d'expliquer les phénomènes linguistiques en général. Nous terminerons par les analyses et débats génératifs qui concernent une problématique majeure dans le cadre de l'analyse du sujet syntaxique : la présence concomitante, dans un même énoncé, d'un clitique nominatif et d'un syntagme qui possèdent tous deux la même référence, comme mentionné au Chapitre I, et que nous illustrons de nouveau en (1) ci-dessous.

(1) Un type problématique d'énoncés¹ :

QUE: *les bébés_i ça_i va à la crèche !* (VIII.33b, 198)

JUL: *ji' aime bien les grenouilles moi_i .* (XIII.48a, 71)

I. LE CHOIX DE NOTRE CADRE THÉORIQUE

Comme le rapportent notamment Bideaud *et al.* (1993), ce qui a été considéré dans l'enfance s'est longtemps restreint à « sa finalité, c'est-à-dire son intégration plus ou moins rapide dans le tissu social, et son évolution plus ou moins dirigée vers un état achevé qui corresponde aux critères d'une époque »². Cette perspective va généralement de pair avec une mortalité infantile longtemps élevée, ainsi qu'une espérance de vie bien plus courte que celle que nous connaissons au XXI^e siècle dans les pays développés. Néanmoins, cette association n'est pas systématique puisque ce

¹ Énoncés extraits de notre corpus dont le détail se trouve au Chapitre III et en Annexes.

² Bideaud *et al.* (1993:4-5).

point de vue a également cours au XX^e siècle dans des pays industrialisés tels que les États-Unis¹.

1. Le comportementalisme

1.1 L'état initial vu comme une coquille vide

De ce regard porté sur l'enfance, au mieux désintéressé, au pire négatif, découle un autre point de vue longtemps dominant concernant les connaissances éventuelles que pourrait avoir le très jeune enfant du monde qui l'entoure ; et ce point de vue se résume à « aucune ». En d'autres termes, l'enfant est comparé à une coquille vide. Les propos d'Aristote sur l'intellect² et ceux de Locke sur l'esprit³ sont ainsi souvent cités sur ce sujet. Locke compare d'ailleurs volontiers les enfants avec les « idiots », les « sauvages » et les « illettrés » :

« C'est une erreur, je le crains, d'attendre d'un enfant inculte ou d'un sauvage ces maximes abstraites et les fameux principes de sciences. »⁴

Globalement, l'impression qui se dégage au fil de l'histoire est, qu'à la naissance, l'enfant n'a aucune connaissance. Cette conception de la connaissance n'est néanmoins pas omniprésente. En effet, dans l'Antiquité par exemple, on trouve également le postulat inverse, notamment chez Platon qui, par le biais d'un dialogue entre Socrate et Menon à propos d'une réflexion géométrique menée à terme par un esclave qui n'avait aucune formation en la matière, écrit que « en lui [l'esclave] la connaissance était sa possession » et que « la vérité du monde se trouve logée en notre souffle vital »⁵.

¹ Cf. Watson (1928) par exemple : "After all, childhood is simply the preparation for adulthood, it is not a state within itself." « Après tout, l'enfance ne sert qu'à préparer l'état adulte, il ne s'agit pas d'un état à part entière. »

² Aristote [1994]:91, Livre III,4,430a : « Il doit en être comme d'une tablette où rien ne se trouve inscrit en entéléchie : c'est précisément ce qui arrive dans le cas de l'intellect. »

³ Locke (1690:164) : « Supposons que l'esprit soit, comme on dit, du papier blanc, vierge de tout caractère, sans aucune idée ». Jean-Michel Vienne, le traducteur de l'édition de 2001, note que l'expression *tabula rasa* que l'on retrouve citée dans la littérature provient en fait du premier traducteur (Coste) qui avait traduit le terme original *white paper* par « table rase » dans le texte avec l'équivalent latin *tabula rasa* en note de bas de page.

⁴ Locke (1690:89).

⁵ Platon [1999]:89-91. Cette référence à Platon est notamment présente de façon récurrente chez les générativistes dans le cadre de l'argument de « la pauvreté du stimulus ». Cf. Section I.3.1 ci-après.

Ce questionnement sur la préexistence de la connaissance va de pair avec une seconde interrogation. En effet, que l'on considère l'enfant comme une « feuille blanche » ou pas, se pose également la question du mode d'acquisition d'un certain nombre de connaissances.

1.2 Le mécanisme de l'apprentissage

Schématiquement, deux camps s'affrontent depuis fort longtemps : les empiristes et les rationalistes, en d'autres termes, les partisans de l'expérience opposés à ceux de l'innéisme. Reprenons la citation de Locke que nous complétons à présent :

« Supposons que l'esprit soit, comme on dit, du papier blanc, vierge de tout caractère, sans aucune idée. Comment se fait-il qu'il en soit pourvu ? [...] D'où puise-t-il ce qui fait le matériau de la raison et de la connaissance ? Je répondrai d'un seul mot : de l'*expérience* ; en elle, toute notre connaissance se fonde et trouve en dernière instance sa source [...].¹ »

Ces propos sont typiques de la position empiriste, c'est-à-dire du point de vue qui considère que l'apprentissage est le résultat d'une expérience. Les notions de stimulus, de réponse et de renforcement font alors partie intégrante de ce mécanisme. Issu notamment des travaux du physiologiste russe Ivan Pavlov portant sur les réflexes conditionnés chez les chiens², cette approche de l'apprentissage a été très largement développée puis renforcée dans le courant du XX^e siècle aux États-Unis par deux psychologues : John B. Watson puis Burrhus F. Skinner³. Leurs travaux, connus sous le nom de « comportementalisme » ou « behaviourisme » posent que le comportement humain est le résultat d'un conditionnement environnemental et que, par conséquent, le mécanisme de l'apprentissage relève du réflexe conditionné. La citation ci-dessous fait ainsi partie des plus diffusées de Watson (1924) :

¹ Locke (1690:164).

² Cf. Pavlov (1927).

³ Cf. par exemple Watson (1924, 1928) et Skinner (1938).

“Give me a dozen healthy infants, well-formed, and my own specified world to bring them up and I’ll guarantee to take any one at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant, chief and, yes, even beggarman and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors.”¹

Dans ce même ouvrage, Watson (1924) discute également la notion d’instinct et son observation de nombreux nouveau-nés le conduit d’ailleurs à la mettre en cause :

“Are we not ready to admit that the whole concept of instinct is thus academic and meaningless? Even from the earliest moment we find habit factors present—present even in many acts so apparently simple that we used to call them physiological reflexes. [...]. Actual observation thus makes it impossible for us any longer to entertain the concept of instinct.”²

La mention du terme « enfant », dans ce cadre, apparaît ainsi parfois en cooccurrence avec celui de « jeune animal ». D’ailleurs, Watson (1928) préconise une connaissance préalable « de la physiologie et de la psychologie animale » avant de passer à l’étude du comportement du jeune enfant³.

Le comportementalisme est donc le modèle prégnant dans les années 1930-1950 aux États-Unis en ce qui concerne le domaine de la psychologie. Mais qu’en est-il de la linguistique ?

1.3 *Quid du langage ?*

Ce modèle comportementaliste omnipotent a des répercussions en linguistique, tant au niveau des méthodes d’analyse utilisées par les linguistes américains de

¹ Watson (1924:82) : « Donnez-moi une douzaine d’enfants normaux, ainsi que mon propre environnement pour les élever et, prenant n’importe lequel d’entre eux au hasard et le formant, je vous garantis que je peux en faire un expert dans n’importe quel domaine de mon choix : médecin, avocat, artiste, marchand, patron et, oui, même mendiant ou voleur, quels que soient ses talents, penchants, tendances, capacités, vocations et origine raciale. »

² Watson (1924:104) : « Ne sommes-nous pas prêts à admettre que le concept d’instinct dans sa globalité est donc spéculatif et dénué de sens ? Dès les premiers instants de la vie, nous rencontrons jusqu’à présent des facteurs d’habitudes présents jusque dans de nombreux agissements d’apparence si simple que nous avons pour habitude de les considérer comme des réflexes physiologiques. L’observation concrète de ces faits nous empêche dorénavant d’entretenir le concept de l’instinct. »

³ Watson (1928:88) : “No one should attempt to make studies upon the infant until he has had considerable training in physiology and in animal psychology.”

l'époque qu'au niveau de la conception de l'acquisition du langage. Le distributionnalisme américain, avec notamment Bloomfield (1933), se fixe ainsi pour tâche d'étudier la langue de façon mécaniste en se basant uniquement sur des corpus de données produites (des réponses, en quelque sorte), sans tenir compte de l'aspect humain du langage. Le distributionnalisme est donc lié au comportementalisme dans la mesure où il rejette non seulement tout ce qui concerne le signifié des données mais également tout ce qui touche à une éventuelle introspection de l'être humain et donc de sa pensée (le mentalisme) ; seuls comptent les faits observables, en l'occurrence les signifiants produits et leur distribution afin de dégager les unités syntagmatiques puis morphémiques dans ces énoncés. Dans ce cadre, Bloomfield (1933) mentionne relativement brièvement la question de l'apprentissage du langage :

“Every child that is born into a group acquires these habits of speech and response in the first years of his life. [...]. Exactly how children learn to speak is not known; the process seems to be something like this. [...]. The child's speech is perfected by its results. [...]. In short, his more perfect attempts at speech are likely to be fortified by repetition, and his failures to be wiped out in confusion.”¹

Ainsi, nous retrouvons chez Bloomfield tous les ingrédients du comportementalisme appliqués à l'émergence d'une langue chez un enfant : le stimulus, la réponse et le renforcement. Dans ce cadre, il s'agit donc d'un mécanisme d'*apprentissage* de connaissances, rendu possible grâce à l'imitation que fait l'enfant d'informations qui lui sont apportées. Pour Bloomfield, ce mécanisme est incessant et concerne aussi bien la phonologie que les autres domaines de la linguistique, tels que la morphologie ou la syntaxe. Cependant, Bloomfield ne s'attarde pas davantage sur ces questions. En tout cas, dans ce cadre, l'apprentissage du langage n'est pas perçu comme un domaine particulier et l'apprentissage des connaissances linguistiques est donc considéré sur le même plan que les autres apprentissages. Quelques années plus tard, Skinner (1938, 1953, 1957), le continuateur de Watson, s'attarde davantage sur

¹ Bloomfield (1933:29-31) : « Tout enfant né au sein d'une communauté [linguistique] acquiert ses habitudes de production et de réponse au cours des premières années de sa vie. La façon exacte par laquelle les enfants apprennent à parler n'est pas connue mais le processus pourrait être décrit ainsi. [...]. Le langage infantin est amélioré par ses propres résultats. [...]. En bref, les meilleurs essais de l'enfant ont toutes les chances d'être renforcés par la répétition, alors que ses échecs sont oubliés. »

la question du langage dans le cadre comportementaliste mais développe les mêmes idées¹.

1.4 La psycholinguistique

Mais au cours des années 1950, ce modèle rigide de stimulus-réponse semble trouver ses limites au sein même du domaine de la psychologie. Ainsi se développe notamment le constructivisme issu des travaux de Piaget et naît un domaine de recherches interdisciplinaire : la psycholinguistique, avec notamment un intérêt nouveau porté à ce qui se passe entre le moment du stimulus et celui de la réponse, c'est-à-dire au traitement de l'information². Puis, suivant les progrès de l'informatique, entre en scène l'aspect automatique de ce traitement et le connexionnisme se développe alors dans les années 1980 avec par exemple le modèle de Rumelhart & McClelland (1986)³. Et même si Plunkett (1995) souligne que les connexionnistes actuels ne sont pas des « empiristes de la *tabula rasa* »⁴, les travaux connexionnistes consistent tout de même en l'élaboration de réseaux qui reproduisent les mécanismes de l'apprentissage basés uniquement sur les apports de l'environnement⁵. Ce point de vue se retrouve également dans la récente linguistique « cognitive et fonctionnelle » (*Cognitive-Functional Linguistics*) qui développe un modèle « basé sur l'usage » (*usage-based model*) de l'acquisition du langage⁶. L'apport de l'environnement demeure donc une constante primordiale dans tous ces modèles.

¹ Skinner (1957:5) : “What happens when a man speaks or responds to speech is clearly a question about human behavior and hence a question to be answered with the concepts and techniques of psychology as an experimental science of behavior.”. « Ce qui se passe lorsqu'un homme parle ou répond à un discours relève clairement du comportement humain et doit ainsi être analysé avec les concepts et les techniques de la psychologie en tant que science expérimentale du comportement. »

² Cf. Shannon (1949).

³ Il s'agit d'un algorithme d'apprentissage décrit sous la forme d'un réseau de type neuronal auquel sont présentés des stimuli. Ces stimuli renforcent les connexions en direction des réponses correctes et éliminent les connexions fausses. D'après les auteurs, ce modèle rend compte de façon correcte du mécanisme de l'apprentissage des formes régulières et irrégulières du prétérit chez les petits anglophones.

⁴ Plunkett (1995:71).

⁵ Cf. par exemple Plunkett (1995:71) : “To be sure, the newborn mind comes equipped with a sophisticated set of constraints for processing its environment.”

⁶ Cf. notamment Tomasello (2003:5) : “[...] language structure emerges from language use” et Tomasello (2000) pour l'application au domaine de la syntaxe.

2. L'innéisme

Les années 1950 marquent ainsi un tournant dans l'histoire de la psychologie avec de nouveaux points de vue et la naissance de la psycholinguistique. Mais, de nouveau, qu'en est-il de la linguistique ?

2.1 Chomsky vs Skinner

On peut, pour le moins, également parler de changement – certains parlent de révolution – face notamment à l'article de Chomsky (1959). Cette publication est en effet marquante pour plusieurs raisons. Tout d'abord, Chomsky (1959) y développe une recension radicale de l'ouvrage de Skinner (1957) dans la mesure où tant la forme que le fond de l'ouvrage de Skinner y sont fortement critiqués. Chomsky qualifie par exemple la démarche de Skinner d'« illusion d'une théorie scientifique rigoureuse » et de « mystification générale »¹. L'emploi des termes de « stimulus », « réponse », « renforcement » et même « comportement verbal » sont décrits comme trop larges et finalement vides de sens. Chomsky en conclut que Skinner n'apporte ainsi aucune preuve pour étayer l'hypothèse que « les procédures de renforcement mises en place par la communauté linguistique représentent le seul facteur responsable du maintien de la vitalité du comportement verbal »². La première caractéristique de cette recension repose donc sur la dureté de la critique. Mais Chomsky ne s'arrête pas là et c'est ce qui fait, à notre avis, l'une des grandes forces de cet article. À peine a-t-il ouvert cette brèche importante dans l'approche behaviouriste « toute puissante » de l'époque qu'il s'y engouffre :

“The magnitude of the failure of this attempt to account for verbal behaviour serves as a kind of measure of the importance of the factors omitted from consideration, and an indication of how little is really known about this remarkably complex phenomenon.”³

¹ Chomsky (1959:28-29).

² Chomsky (1959:36) : “It is also not easy to find any basis [...] to the claim that reinforcing contingencies set up by the verbal community are the single factor responsible for maintaining the strength of verbal behavior.”

³ Chomsky (1959:27) : « L'amplitude de l'échec de cette tentative de rendre compte du comportement verbal permet de mesurer toute l'importance des facteurs occultés et nous indique la faible quantité de choses que nous savons vraiment sur ce phénomène remarquablement complexe. »

Chomsky conteste donc avec véhémence le comportementalisme mais il ne laisse pas la place vide. Il pose dès les années 1950 les premiers jalons d'une hypothèse radicalement opposée : celle de l'innéisme.

2.2 Les racines de la théorie générative

Ainsi, dès la fin des années 1950, Chomsky introduit une perspective totalement anti-comportementaliste de l'acquisition du langage et développe concomitamment les thèses fondatrices d'une nouvelle théorie linguistique : la grammaire générative. Néanmoins, certaines de ses idées ne sont pas nouvelles et Chomsky attribue d'ailleurs explicitement les fondements de sa réflexion à différents philosophes et scientifiques qui l'ont précédé.

Certaines des pensées de Chomsky ne sont donc pas fondamentalement novatrices. Chomsky suit notamment les développements tout récents de l'éthologie dont les principes viennent à peine d'être appliqués à l'étude de l'homme par Tinbergen (1951). Dans ce cadre, Lenneberg (1967) notamment postule alors que le comportement humain n'est pas le résultat d'une influence de l'environnement mais que, au contraire, il y a quelque chose de sous-jacent à ce comportement¹. L'auteur remarque ainsi que le développement linguistique se fait de façon très similaire d'un enfant à un autre (en ce qui concerne l'ordre et l'âge des acquisitions notamment) et pose la question de la particularité du langage en rapport avec les autres comportements humains :

“The central and most interesting problem is whether the emergence of language is due to very general capabilities [...], or whether there might be some factors specific to speech and language [...].”²

L'auteur arrive à la conclusion que l'émergence du langage chez le jeune enfant est le résultat d'un processus spécifique et donc différent des autres développements³. Sont posés ici deux principes majeurs repris par Chomsky dans le cadre de la

¹ Ce qu'il appelle ses « potentialités ». Cf. Lenneberg (1967:104).

² Lenneberg (1967:103-104) : « Le problème central et le plus intéressant concerne l'émergence du langage : est-elle le fait de capacités générales [...] ou de facteurs spécifiques à la parole et au langage [...] ? »

³ Lenneberg (1967:105) : “[...] the specific history for speech control stands apart dramatically from histories of finger and hand control”. « Le fil particulier du développement langagier est totalement distinct de ceux qui concernent le développement moteur. »

grammaire générative : le fait que certaines capacités soient innées et le fait que le domaine du langage et de son acquisition relève d'un domaine cognitif particulier. Il est donc explicité que les facteurs externes à l'être humain, tels que l'environnement, sont nécessaires à l'acquisition du langage mais qu'ils ne sont pas suffisants¹. Chomsky emploie ainsi le qualificatif de « génétique » dès les années 1950 et notamment dans sa recension de 1959 pour décrire ces facteurs internes à l'homme :

“However, there appears to be good evidence that these responses to fairly complex “sign stimuli” are genetically determined and mature without learning.”²

Chomsky fait également référence à de nombreux philosophes du XVII^e siècle, « siècle de génie »³, à commencer par Descartes⁴. Descartes n'a pas beaucoup publié sur le langage mais Chomsky (1966) relève l'aspect créateur que Descartes (1637) attribue au langage⁵, ainsi que le fait qu'il le considère comme une faculté propre à l'homme et issue de sa « raison ». C'est ainsi que les termes de « rationaliste », de « cartésien », ou de « mentaliste »⁶ s'opposent également parfois à « empiriste » ou « naturaliste »⁷ dans le débat sur nos capacités linguistiques. Plus près de lui dans le temps, des philosophes allemands tels que Schlegel (1801) et Humboldt (1836) ont également marqué la réflexion de Chomsky. La formule de Humboldt (1836) « [Le langage] est donc appelé à faire, à partir de moyens finis, un usage infini, ce qui n'est possible que parce que c'est un seul et même dynamisme qui produit la pensée et le langage » a d'ailleurs été largement citée depuis⁸.

¹ Chomsky (1968:82) : “What seems to me important in ethology is its attempt to explore the innate properties that determine how knowledge is acquired and the character of this knowledge.” « Ce qui me semble important dans l'éthologie est la tentative qu'elle fait d'explorer les propriétés innées qui déterminent comment la connaissance est acquise et les caractéristiques de cette connaissance. »

² Chomsky (1959:37) : « Cependant, des indices forts semblent indiquer que les réponses à ces « stimuli linguistiques » relativement complexes sont déterminées génétiquement et qu'elles se développent sans apprentissage. »

³ Chomsky (1968:5).

⁴ Mais également Cordemoy (1666) et La Forge (1666).

⁵ Descartes (1637:vol 6,56-57), cité par Chomsky (1966:4) : « Car c'est une chose bien remarquable qu'il n'y a point d'hommes si hébétés et si stupides, sans excepter les insensés, qu'ils ne soient capables d'arranger ensemble diverses paroles, et d'en composer un discours par lequel ils fassent entendre leurs pensées ; et qu'au contraire il n'y a point d'autre animal, tant parfait et tant heureusement né qu'il puisse être, qui fasse le semblable. »

⁶ Chomsky (1965:13) : « [...] la théorie linguistique est mentaliste, au sens technique de ce mot, puisqu'elle s'attache à découvrir une réalité mentale sous-jacente au comportement effectif. »

⁷ Cf. Diderot (1751), cité par Chomsky (1965:7).

⁸ Humboldt (1836:122) repris notamment par Chomsky (1966:20).

2.3 Les hypothèses fondamentales du générativisme

Une des toute premières thèses fondamentales du générativisme réside donc dans l'aspect inné et spécifique du langage humain. C'est ce qui est appelé la « Faculté de langage » (FL), définie comme « une composante de l'esprit/cerveau humain, dédiée au langage »¹. Ces dernières années, Chomsky va même jusqu'à qualifier cette faculté – terme jusque là relativement abstrait – d'« organe du langage » – terme beaucoup plus concret qui vise à montrer que le langage fait partie intégrante des gènes de l'être humain. Puisque cette Faculté est présente chez tous les êtres humains, quels que soient leur époque et leur lieu de naissance, il faut que le langage ait également une forme initiale commune à tous. Ce stade initial est appelé « grammaire universelle » :

« Définissons la « grammaire universelle » (GU) comme le système des principes, des conditions et des règles qui sont des éléments ou des propriétés de toutes les langues humaines, pas simplement par accident, mais par nécessité – nécessité biologique et non logique, évidemment. Ainsi on peut considérer que GU exprime l'« essence du langage humain ». GU ne variera pas selon les individus. »²

Chaque locuteur est ainsi génétiquement équipé de cette grammaire universelle. Chaque locuteur a donc une connaissance interne du langage en général mais également une capacité qui lui permet de créer et de produire des énoncés sans cesse nouveaux dans sa (ou ses) langue(s). Deux notions s'ajoutent donc à celle de la grammaire universelle : la « compétence » et la « performance » du locuteur :

« Nous établissons donc une distinction fondamentale entre la *compétence* (la connaissance que le locuteur-auditeur a de sa langue) et la *performance* (l'emploi effectif de la langue dans des situations concrètes). »³

Chomsky pose ainsi une dichotomie qui n'est pas sans rappeler la partition saussurienne de langue vs parole⁴. L'auteur revendique d'ailleurs cette parenté mais déclare qu'« il est nécessaire de rejeter le concept saussurien de « langue », qui réduit

¹ Chomsky (2000:89).

² Chomsky (1975:40).

³ Chomsky (1965:4).

⁴ Cf. Saussure (1916).

celle-ci à un inventaire systématique d'éléments, pour revenir à la conception humboldtienne qui fait de la compétence sous-jacente un système de processus génératifs »¹. L'aspect dynamique de la création linguistique fait donc également partie des thèses fondatrices de cette théorie, à telle point que cette faculté à *générer* des énoncés a même donné son nom à la théorie :

« La grammaire d'une langue se propose d'être une description de la compétence intrinsèque du locuteur-auditeur idéal. Si la grammaire est, de plus, parfaitement explicite – en d'autres termes, si elle ne fait pas simplement confiance à la compréhension du lecteur intelligent, mais fournit une analyse explicite de l'activité qu'il déploie – nous pouvons, non sans redondance, l'appeler *grammaire générative*. »²

D'après l'auteur, ce qu'il manque aux grammaires en général est un aspect dynamique qui leur permettrait de rendre compte des régularités présentes dans le langage au lieu de se focaliser sur les irrégularités des différentes langues³. La grammaire générative ambitionne donc non seulement de décrire les langues mais également d'expliquer leur fonctionnement.

Ces notions représentent les concepts clés de la grammaire générative. Ce sont les fondamentaux que l'on trouve donc aussi bien lors de la mise en place de la théorie à la fin des années 1950 que de nos jours après cinquante années d'existence. Cependant, l'aspect de la théorie générative a énormément évolué au cours de ces années, tant en ce qui concerne la formalisation des mécanismes sous-jacents qui nous permettent de générer nos phrases que sur celui de l'innéisme et de l'acquisition du langage⁴.

3. Une nouvelle perspective pluridisciplinaire

Tomasello (2006) résume l'opposition entre connexionnistes et générativistes en ces termes :

¹ Chomsky (1965:14).

² Chomsky (1965:4).

³ Cf. Chomsky (1965:5) : « [...] elles [les grammaires traditionnelles] manquent à formuler bien des régularités fondamentales de la langue [...]. Cela est particulièrement clair au niveau de la syntaxe, où aucune grammaire, qu'elle soit traditionnelle ou structuraliste, ne parvient, par-delà la classification des exemples particuliers, à formuler des règles génératives de quelque extension. »

⁴ Nous revenons sur ces questions dans la deuxième partie de ce chapitre.

“It is easy to see in this historical sketch two distinct strands. One derives from researchers who take a formal approach to language and its acquisition – a more adult-centered approach emanating from Chomsky’s theory of generative grammar – and the other derives from researchers who take a more functional, usage-based approach to language and its acquisition – a potentially more child-centered approach with room for serious developmental change.”¹

Nous ne souhaitons pas poser le problème du choix de notre cadre théorique de cette manière binaire et oppositionnelle. En effet, malgré les divergences profondes de point de vue qu’il y a entre les connexionnistes et les générativistes, nous pensons qu’une position consensuelle est possible, notamment grâce à une approche originale qui associe les principes de la grammaire générative et les méthodes d’analyse de la dialectologie.

3.1 Pour l’innéisme

Dans un premier temps, nous retenons comme plus convaincants les arguments de la grammaire générative en faveur d’une conception innéiste du langage. Les apports de l’environnement sont certes fondamentaux mais à nos yeux insuffisants pour expliquer l’acquisition du langage au cours des premières années de la vie. Un dialogue entre le psycholinguiste Martin Braine et sa fille, rapporté par Pinker (1994), nous apparaît comme une illustration intéressante des limites de l’apport environnemental :

¹ Tomasello (2006:5-6) : « Deux positions distinctes apparaissent de façon évidente dans ce survol historique. L’une vient de chercheurs qui adoptent une approche formelle du langage et de son acquisition – une approche plus centrée sur l’adulte, issue de la grammaire générative de Chomsky – et l’autre vient de chercheurs dont l’approche est plus fonctionnelle et basée sur l’usage – une approche potentiellement plus centrée sur l’enfant, qui accorde de la place à un changement développemental important. »

« L'enfant : Veux autre cuillère, Papa.
 Le père : Tu veux dire que tu veux L'AUTRE CUILLÈRE.
 L'enfant : Oui, veux autre cuillère.
 Le père : Est-ce que tu peux dire « l'autre cuillère » ?
 L'enfant : Autre... cuillère.
 Le père : Dis... « L'autre ».
 L'enfant : L'autre.
 Le père : « Cuillère ».
 L'enfant : Cuillère.
 Le père : « L'autre cuillère ».
 L'enfant : L'autre cuillère. Maintenant veux autre cuillère. »¹

Ainsi, la « réponse » de la fille de Martin Braine reste la même et le processus de « renforcement » ne semble pas visible, malgré la répétition du « stimulus » et malgré le fait qu'il est notable que l'enfant peut également produire la forme requise. Dans le cadre de nos recherches, nous ne développerons pas cette problématique de la répétition. Toutefois, il est important de souligner que le regard porté sur ce phénomène a beaucoup évolué et que, comme le mentionnent Morgenstern & Prak-Derrington (2008), la répétition n'est absolument plus considérée comme « un simple automatisme passif » mais « comme un processus constructif [utilisé] par l'enfant dans son appropriation du langage »², dans la mesure notamment où elle permet au jeune enfant de développer ses compétences lexicales et pragmatiques³.

Néanmoins, cet argument en faveur d'une relative indépendance entre développement linguistique et environnement rejoint un autre argument développé en faveur d'une approche innéiste de l'acquisition du langage qui est celui dit de la pauvreté du stimulus qu'Obenauer (1990) explicite en ces termes : « l'argument de la « sous-détermination de la grammaire atteinte par les données linguistiques primaires » »⁴. Dans ce cadre, il est considéré que les données dont les enfants disposent sont qualitativement et quantitativement insuffisantes. En effet, Chomsky (1968) qualifie ces données de « qualitativement et quantitativement pauvres »⁵ dans

¹ Pinker (1994:279).

² Morgenstern & Prak-Derrington (2008:250).

³ Cf. Morgenstern & Prak-Derrington (2008) pour des exemples de dialogues enfant-adulte qui illustrent ces différents types d'acquisition.

⁴ Obenauer (1990:76).

⁵ Chomsky (1968:53) : “degenerate and restricted”.

le sens où elles sont partielles, fragmentées, imparfaites et parfois même ambiguës. Des études ont été menées par les défenseurs du générativisme¹, des explications ont été formulées², mais bien sûr des réticences existent³. Pour notre part, nous adhérons au fait que l'information donnée aux jeunes enfants présente effectivement des lacunes qualitatives et quantitatives importantes dans le sens où elles ne sont ni explicites ni univoques⁴. Nous pensons donc que leur traitement serait effectivement impossible sans la présence d'une faculté de langage innée. Cependant, nous souhaitons associer à cette position le fait que le langage destiné aux enfants présente néanmoins des caractéristiques notamment intonatives, syntaxiques et lexicales qui sont là pour faciliter la segmentation du continuum phonique⁵. De plus, l'enfant est immergé dans un « bain linguistique » qui lui envoie des données directes et indirectes pendant approximativement deux ans avant que celui-ci ne produise effectivement ses premières combinaisons de deux mots. La « rapidité » de l'acquisition peut donc sembler quelque peu relative. Néanmoins, dans le cadre de cette réflexion, il convient de distinguer le moment effectif où l'enfant commence à produire ses premiers énoncés du moment où il détermine les valeurs des paramètres de sa grammaire⁶.

Nous adhérons également au postulat selon lequel le langage peut être considéré comme un domaine cognitif particulier. Cette position renvoie à la notion d'indépendance et de spécialisation de chacune de nos capacités cognitives⁷. Il a ainsi été démontré que, malgré le rapprochement qui est souvent établi entre facultés mentales et développement du langage⁸, ce dernier peut néanmoins se révéler indépendant des autres facultés intellectuelles⁹. De plus, parler du langage comme d'un tout est de toute façon erroné puisque l'existence de différents types d'aphasies

¹ Cf. par exemple Lidz *et al.* (2003) qui mettent en avant l'insuffisance qualitative des données parentales.

² Cf. Obenauer (1990).

³ Cf. par exemple Bouchard (2001) qui qualifie cet argument « d'artefact ».

⁴ Dans Palasis (2005:139-141), nous mentionnons par exemple la contradiction que représentent les deux paradigmes de l'indicatif et de l'impératif présent qui, dans le premier cas seulement, nécessite la présence d'un sujet.

⁵ Cf. par exemple Veneziano (2000) et le LAE (Langage Adressé aux Enfants) mentionnés page 55.

⁶ Cf. Section II de ce chapitre pour le détail.

⁷ Cette position est également connue sous le nom de « modularité » mais nous n'utilisons pas ce terme, dû aux nombreux contresens qu'il a générés. Cf. Scheer (à paraître) qui propose un chapitre consacré à l'historique et aux contresens qui ont été faits sur cette notion.

⁸ Cf. Leonard (1996:287) : “[...] almost all children with moderate to severe retardation have language disorders.” « Quasiment tous les enfants modérément ou très retardés ont des troubles du langage ».

⁹ Cf. par exemple Smith & Tsimpli (1991) qui exposent le cas d'un jeune homme au QI limité mais aux capacités linguistiques extraordinaires.

par exemple met en évidence l'existence de sous-spécialisations dans des aires différentes du cerveau¹. Ces sous-spécialisations sont d'ailleurs représentées dans la théorie générative depuis les années 1960 par le biais d'une architecture de la grammaire qui se dessine en trois composantes : la syntaxe, la phonologie et la sémantique². De nombreux débats sont ouverts sur ces sujets et l'une des préoccupations relativement récentes concerne la transmission des informations entre ces composantes, d'où une littérature abondante sur la question des « interfaces »³.

Nous adoptons donc comme cadre théorique l'approche innéiste proposée par la grammaire générative. Néanmoins, nous retenons également certaines des critiques qui lui sont adressées en conséquence de quoi nous ne nous limitons pas exclusivement aux méthodes et préceptes qu'apporte cette théorie.

3.2 Le problème des données

Les propos de Tomasello (2006) présentés au début de la Section 3 permettent de cerner les contours de l'une des critiques souvent formulées à l'encontre de la grammaire générative : celle-ci ne fonde pas ses théories sur une observation détaillée de la réalité. Ce reproche renvoie notamment au fait que les générativistes n'utilisent pas traditionnellement des données en grande quantité afin de formuler leurs hypothèses mais que, au contraire, ils idéalisent les productions possibles en se fiant à des énoncés souvent imaginés pour la circonstance. Cette méthode relève d'une volonté délibérée d'abstraction dans la mesure où le but est de décrire et d'expliquer les mécanismes sous-jacents à nos productions. De plus, puisque celles-ci sont en nombre infini, il n'est pas considéré comme méthodologiquement pertinent d'en collecter un échantillon :

“Observed use of language may provide evidence but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline.”⁴

¹ Les aphasies sont des troubles du langage dus à une lésion du cerveau. Cf. par exemple Jakobson (1956) ou Pinker (1994:295-328).

² Cf. Chomsky (1965:31). Dans la terminologie actuelle, la phonologie est prise en charge par le système « articulatoire et perceptuel » (*articulatory-perceptual*) et la sémantique par le système « conceptuel et intentionnel » (*conceptual-intentional*). Cf. Chomsky (1995b:167-169).

³ Cf. par exemple les réflexions de Chomsky *et al.* (2002:151) sur les nouvelles perspectives en la matière.

⁴ Chomsky (1965:4) : « L'observation de l'utilisation du langage peut éventuellement apporter des indices mais ne peut absolument pas représenter l'objet d'étude de la linguistique, si celle-ci veut être une discipline sérieuse. »

Chomsky, comme Saussure en son temps, ne s'intéresse donc pas aux résultats de la production linguistique mais seulement à la « compétence » sous-jacente. Cette ligne de conduite méthodologique, qui focalise donc l'intérêt sur l'une des deux sous-parties du langage par le biais de productions idéalisées, prête alors à controverse :

“[...] generative grammar has produced many explanatory hypotheses of considerable depth, but is increasingly failing because its hypotheses are disconnected from verifiable linguistic data. Issues of frequency of usage are by design made external to matters of syntax, and as a result categorical judgments are overused where not appropriate, while a lack of concern for observational adequacy has meant that successive versions have tended to treat a shrinking subset of data increasingly removed from real usage.”¹

Puisqu'il critique la méthodologie de la grammaire générative, on pourrait alors penser que Manning (2003) œuvre en faveur des linguistes qui utilisent de nombreuses données. Or, ce n'est pas non plus le cas :

“On the other side, corpus linguistics [...] or “usage-based models of grammar” [...] has all the right rhetoric about being an objective, falsifiable, empirical science interested in the totality of language use, but is failing by largely restricting itself to surface facts of language, rather than utilizing sophisticated formal models of grammar, which make extensive use of *hidden structure* (things like phrase structure trees and other abstract representational levels).”²

¹ Manning (2003:296) : « [...] la grammaire générative a produit de nombreuses hypothèses explicatives d'une portée considérable, mais elle échoue de plus en plus parce que ses hypothèses sont sans rapport avec des données linguistiques vérifiables. Les questions de fréquence dans l'usage sont de fait rendues extérieures au domaine de la syntaxe avec pour conséquence d'entraîner une utilisation abusive et inappropriée de jugements catégoriques alors qu'un manque d'intérêt pour une adéquation avec les faits a signifié que les versions successives ont eu tendance à analyser un sous-ensemble de données de plus en plus rétréci et de plus en plus éloigné de l'usage réel. »

² Manning (2003:296) : « D'un autre côté, la linguistique de corpus [...] ou « modèles de grammaire basés sur l'usage » peut argumenter qu'elle est une science objective, falsifiable, empirique et intéressée par la totalité du langage mais elle échoue en se restreignant aux phénomènes de surface du langage, plutôt que d'utiliser des modèles de grammaire formels et sophistiqués qui font largement appel à une structure sous-jacente (qui apparaît dans les arbres syntagmatiques et autres niveaux représentationnels abstraits). »

Cette description des faits s'adapte parfaitement à la situation que connaît l'étude de l'acquisition de la syntaxe : d'une part, des générativistes qui manipulent une théorie extrêmement complexe avec des représentations de plus en plus abstraites mais relativement peu de données ; et d'autre part, des psycholinguistes qui, à l'inverse, adoptent des représentations plus concrètes mais recueillent des quantités de données considérables.

3.3 Les apports complémentaires d'une approche dialectologique

Bien qu'ayant choisi la grammaire générative pour cadre théorique pour nos recherches, cette question des données nous interpelle ainsi depuis nos tout premiers travaux¹. Nous la formulons ainsi : à quoi bon disposer d'un arsenal théorique extrêmement sophistiqué si les données à partir desquelles cet arsenal fonctionne ne sont pas complètement représentatives de la réalité ? Ainsi, le premier point méthodologique que nous promovons n'appartient pas à la tradition générative puisque nous pensons que, dans le domaine de l'acquisition, il est fondamental de travailler à partir d'un nombre important de données effectivement produites par de jeunes enfants. Nous pensons que nous nous rapprochons ainsi de ce que Tomasello (2006) appelle « une approche plus centrée sur l'enfant ». Cependant, il est évident que ce critère quantitatif à lui seul ne garantit absolument pas une quelconque qualité et encore moins une quelconque représentativité de ces données. De très nombreuses questions épistémologiques et méthodologiques se posent alors dans le sillage de ce choix. C'est pourquoi nous consacrons un chapitre entier de ce travail de recherches aux détails de la réflexion que nous avons menée lors de la constitution de notre propre base de données². Par ailleurs, cette ouverture plus grande sur la réalité linguistique et sa diversité nous entraîne à proposer une seconde extension au cadre génératif traditionnel. En effet, nous avons vu que l'approche générative n'est pas, par vocation, empirique, et que, de fait, elle ne s'est donc pas préoccupée des questionnements qu'entraînent, d'une part, la collecte et l'utilisation de données orales et, d'autre part, la variation. Il nous faut alors compléter notre cadre théorique par une seconde approche qui, elle, dispose d'un grand savoir-faire dans ces domaines afin que les données utilisées aient la plus grande valeur scientifique possible. Nous nous tournons ainsi vers la dialectologie en tant que : (i) discipline traditionnellement empirique, qui décrit les caractéristiques de systèmes linguistiques

¹ Cf. Palasis (2005:88-90).

² Cf. Chapitre III.

à partir de données orales collectées auprès d'informateurs sur le terrain¹, et (ii) discipline qui a pour vocation d'étudier la variation linguistique aux niveaux diachronique et géographique.

En effet, notre démarche est celle de l'investigateur de terrain qui recueille des données le plus représentatives possibles dans le but de les étudier en tant que système linguistique à part entière qui, comme tout système, évolue au fil du temps ; Cohen (1924) par exemple parle de succession de « petits systèmes »². La notion de « système à part entière » signifie que nous considérons le langage enfantin comme un dialecte, c'est-à-dire un système oral, organisé, dont les caractéristiques relèvent donc directement de la structure de ce système. Concrètement, dans le cadre de l'étude du phénomène du « sujet nul » enfantin par exemple³, nous n'adhérons pas aux approches classiques de la grammaire générative en termes d'acquisition. En effet, lorsqu'il y a une différence entre la production infantine et la langue cible, comme illustré en (2) ci-dessous, l'approche générativiste traditionnelle consiste à : (i) présenter les formes enfantines comme agrammaticales (et donc astérisquées) et (ii) postuler que ces agrammaticalités proviennent d'une déficience de l'enfant.

(2) Illustration du phénomène du « sujet nul » enfantin :

(a) Production infantine⁴ :

EMA: oui *o* [*] mange des bonbons ! (I.3d, 5520)

(b) Production cible :

XXX: oui *je* mange des bonbons !

Selon les hypothèses, cette déficience ressortit soit à la compétence de l'enfant (en termes générativistes) soit à sa performance. Ainsi, les travaux fondateurs de Hyams (1986) ou encore ceux de Rizzi (2002) postulent tous les deux un mauvais paramétrage initial enfantin⁵. Pour ce qui est du déficit de la performance, les sujets nuls enfantins sont attribués à des capacités limitées des enfants au niveau de la mémoire, de la computation et/ou de la production. Jakubowicz & Rigaut (1997), par exemple, évoquent une réduction phonologique plus ou moins importante des pronoms sujets et donc une déficience de la production.

¹ Cf. par exemple les travaux de Dalbera (1994) ou Oliviéri (1983).

² Cohen (1924:33).

³ La description de ce phénomène, ainsi que les points principaux développés lors de notre première analyse (Palasis (2005)) sont synthétisés dans la section suivante.

⁴ Extraite de notre corpus.

⁵ Nous revenons sur ces notions dans la suite de ce chapitre.

En ce qui concerne le premier point mentionné ci-dessus, la notion d'agrammaticalité renvoie bien évidemment à celle de norme déjà abordée au Chapitre I. Sur cette thématique, nous rejoignons notamment l'approche préconisée par Morgenstern (2006) lorsqu'elle précise que « [...] l'objectif n'est pas de comparer le langage de l'enfant à une norme adulte, mais d'analyser ses énoncés pour ce qu'ils sont ». Ainsi, lorsque nous cherchons une explication au phénomène d'omission mentionné ci-dessus, nous n'adhérons pas aux hypothèses qui postulent que la différence entre la production de l'enfant et celle de l'adulte provient d'un déficit infantin. Pour autant, nous ne nions pas l'existence de différences entre les deux systèmes. Cependant, nous considérons que ces différences ne représentent pas des « déficits » (ou toute autre caractéristique connotée négativement) qui empêcheraient les enfants de s'exprimer « correctement ». Nous ne regardons donc pas ces différences à un niveau qualitatif mais plutôt à un niveau quantitatif. En effet, le système infantin est réduit par rapport à celui de l'adulte, notamment en termes de nombre de verbes et en termes de nombre de personnes utilisées au sein des paradigmes verbaux. Nous postulons alors que c'est cette différence quantitative qui a des implications sur les productions enfantines¹ ; les sources des différences se trouvent donc dans la structure même du système infantin. Imaginons que nous soyons dialectologue et que nous souhaitions étudier un dialecte D. Nous savons par ailleurs que D n'est pas totalement inconnu de la communauté scientifique et que cette langue est donc déjà classée, grâce à des études antérieures, dans la catégorie des langues dont les productions doivent respecter le type (b) des exemples donnés en (2) ci-dessus. Nous allons sur le terrain, recueillons des données et les analysons. Nous constatons alors que notre corpus comporte environ un quart de phrases qui ne se conforme pas à (b) mais qui relève de (a). Face à cette constatation, il ne viendrait à l'idée d'aucun dialectologue de postuler que la source de (a) est un déficit de la compétence ou un déficit de la performance des locuteurs enregistrés, pour deux raisons : (i) il s'agit d'adultes donc toute déficience relèverait d'aspects pathologiques du langage et (ii) une telle hypothèse, ne se basant que sur les énoncés de type (a), ne rendrait pas compte du fonctionnement global du système.

Nous souhaitons également établir des rapprochements entre l'aspect diachronique du travail de dialectologie et l'aspect évolutif de l'acquisition. La dialectologie étudie la variation des systèmes linguistiques à la fois dans le temps et

¹ Cf. la section suivante pour le détail.

dans l'espace ; c'est ce qui la différencie de la géolinguistique notamment dont le point de vue est centré sur la diatopie et non pas sur la diatopie associée à la diachronie. Dans ce cadre, les dialectes forment un continuum dont chaque maillon géographique peut être perçu comme représentant une étape de la chaîne diachronique¹. Le parallèle que nous établissons entre les travaux en dialectologie et ceux en acquisition relève de l'aspect diachronique des deux disciplines dans la mesure où nous considérons que, dans les deux cas, l'objet d'étude est un continuum de systèmes linguistiques apparentés qui évoluent dans le temps à partir d'un système source. Ainsi, toute hypothèse acquisitionnelle doit être cohérente par rapport à l'étape qu'elle décrit mais également par rapport à l'étape antérieure et l'étape subséquente (puisque les systèmes sont apparentés). Cette notion de continuité présente alors une problématique intéressante en rapport avec les paramètres postulés par la grammaire générative et la fixation de leurs valeurs, problématique sur laquelle nous revenons dans la Section II ci-dessous sur le processus acquisitionnel.

Nous adoptons ainsi une approche qui ne se limite pas à la connaissance et à l'exploitation d'un champ théorique unique mais qui adopte un point de vue éclectique dans la mesure où elle lie étroitement une théorie syntaxique et acquisitionnelle – par le biais de la grammaire générative – à une approche empirique qui n'est pas sans rappeler celle de la dialectologie.

3.4 Une première application : l'étude du « sujet nul » enfantin

Dans ce cadre, nos recherches sur la question du sujet enfantin débutent par l'étude d'un phénomène bien précis dans le cours de l'acquisition du français : le phénomène du « sujet nul »². Celui-ci survient dès les premières combinaisons de mots, vers deux ans, et perdure jusqu'à l'âge approximatif de trois ans. Pendant cette période, le jeune enfant omet alors parfois le sujet dans ses phrases verbales, produisant ainsi des énoncés considérés comme agrammaticaux. Dans notre premier corpus³, le taux d'omission de ce constituant est de 22,4 %⁴. Quelques exemples issus de ce premier corpus sont donnés ci-dessous :

¹ Cf. Dalbera (2006).

² Ce phénomène concerne également d'autres langues, telles que l'anglais ou l'allemand.

³ Palasis (2005).

⁴ Notre taux est donc proche de celui de Rasetti (1996) par exemple (26,2 %), qui est établi à partir d'autres données francophones.

Sujets nuls	Prénoms & Âges	Énoncés cibles
_ <i>est là.</i>	Raphaël, 2;7	“ <u>e</u> lle est là”
_ <i>veux livre.</i>	Tess 2;7	“ <u>j</u> e veux le livre”
_ <i>fais ça.</i>	Thibault, 2;9	“ <u>j</u> e fais ça”
_ <i>sais pas.</i>	Maxence, 2;9	“ <u>j</u> e (ne) sais pas”

Tableau 5 : quelques exemples de « sujets nuls » enfantins (Corpus 1)

Dans le cadre de la grammaire générative, ce phénomène est déjà largement étudié et analysé en termes de « déficit », comme décrit et discuté dans la section précédente. Dans Palasis (2005), nous adoptons un point de vue différent et démarrons nos investigations par l'étude des 1072 phrases verbales de notre premier corpus recueilli auprès de dix-sept enfants entre 2;3 et 3;1. Plus précisément, nous portons notre attention sur les verbes qu'utilisent ces enfants dans les phrases avec sujet ainsi que dans celles qui n'en présentent pas et nous détaillons la structure de ces deux types d'énoncés.

3.4.1 L'origine du sujet nul enfantin

Dans un premier temps, le Tableau 6 ci-dessous reprend les formes verbales enfantines principales. Cette grille met notamment en avant le fait que ces enfants ne produisent que peu de verbes différents et que ces verbes sont *être* et *avoir* ainsi que des verbes qui appartiennent majoritairement au troisième groupe de conjugaison. Nous constatons également que les paradigmes utilisés sont réduits puisque ce système enfantin se compose quasiment exclusivement d'une seule opposition dans la mesure où Pe1 et Pe3 représentent à elles seules 95,9 % des formes constatées¹.

		<i>être</i>	<i>vouloir</i>	<i>faire</i>	<i>avoir</i>	<i>savoir</i>	<i>voir</i>	1 ^{er} groupe
Pers.	%	48,7	17,2	16,2	7,0	3,6	2,2	5,1
1sg	35,0	sɥi	vø	fɛ	e	se	vwa	mãʒ
3sg	60,9	e	-	fɛ	a	se	vwa	mãʒ

Tableau 6 : les formes verbales enfantines (Corpus 1)

L'architecture du système verbal enfantin est donc très différente de celle de l'adulte en termes de nombre de verbes et en termes de nombre de personnes grammaticales. L'information que véhicule chaque forme verbale à l'intérieur de ce système est par conséquent différente chez l'adulte et chez l'enfant : notre premier corpus met notamment en avant le fait que, dans de nombreux cas, la morphologie des formes verbales enfantines à elle seule apporte une information précise sur la personne grammaticale : en l'occurrence, les deux formes des verbes *être* et *avoir* sont

¹ Les occurrences restantes étant Pe2 (2,9 %) et Pe6 (1,2 %).

distinctives, ainsi que la forme unique du verbe *vouloir*. En ce qui concerne le phénomène du sujet nul enfantin, nous avançons donc l'hypothèse que l'absence d'environ un quart des sujets enfantins est une conséquence de cette richesse morphologique. En effet, la configuration décrite ci-dessus rappelle fortement celle du système verbal de langues comme l'italien, l'espagnol ou le grec qui ne demandent pas la présence obligatoire d'un sujet et qui présentent des formes verbales majoritairement distinctives.

Grâce à son environnement linguistique, le jeune enfant francophone reçoit donc des données qui l'informent que le français est une langue qui nécessite la présence d'un sujet. Cependant, l'architecture de son propre système n'est pas encore identique à celle de l'adulte. L'enfant produit donc une majorité de phrases avec un sujet mais, sous l'influence de la richesse morphologique de son système verbal, en omet quelques-uns. Nous pensons que cette hypothèse présente plusieurs avantages, notamment de rendre également compte de la disparition du phénomène et de pouvoir être étendue à d'autres langues dans lesquelles ce phénomène est également constaté, telles que l'anglais ou l'allemand. En effet, dans ce cadre, la disparition du phénomène peut être attribuée à l'utilisation progressive de formes supplémentaires dans le paradigme et de nouveaux verbes, en particulier des verbes du premier groupe qui ne présentent plus cette richesse morphologique. Le système de l'enfant s'étoffe ainsi petit à petit pour se rapprocher progressivement du système de l'adulte. La différence entre les deux systèmes linguistiques – identifiée comme étant le déclencheur du phénomène – s'atténue alors et le phénomène disparaît. Par ailleurs, cette hypothèse serait inter-linguistique, s'il était démontré que les jeunes enfants anglophones ou germanophones par exemple utilisent essentiellement, à ce stade, des verbes irréguliers aux Pe1 et Pe3¹.

3.4.2 La nature des sujets enfantins

Notre premier corpus permet également de cerner précisément la nature des éléments que le jeune enfant réalise et omet en fonction sujet. Ainsi, le Tableau 7 ci-dessous détaille tout d'abord la répartition des différents types de sujets produits.

¹ Ce qui donnerait les oppositions distinctives suivantes : anglais : *am* et *is*, *have* et *has* ; allemand : *bin* et *ist*, *habe* et *hat*.

Trois natures sont possibles – pronom clitique, pronom tonique et nom – et différentes combinaisons de ces trois types sont attestées¹.

Nature des sujets	Exemples	N	%
+PrC	<i>Elle est grande</i>	459	42.8
+PoC	-	0	0.0
+PrN	-	0	0.0
+PrN +C	<i>Le renard il a mangé</i>	16	1.5
+C +PoN	<i>Il est là la galette</i>	30	8.1
+PrP +C	<i>Lui i(l) s'appelle Raphaël</i>	87	2.8
+C +PoP	<i>Je peux jouer moi ?</i>	64	6.0
+PrP +C +PoP	<i>Moi j'ai un grand vélo moi</i>	6	0.5
Impératifs	<i>_ attends !</i>	170	15.9
Sous-total pour les phrases avec sujet		832	77.6
Sous-total pour les phrases sans sujet		240	22.4
Total corpus 1		1072	100.0

Tableau 7 : la nature des sujets enfantins réalisés (Corpus 1)

Les données ci-dessus fournissent deux indications importantes. La première concerne la nature du sujet enfantin ; la seconde concerne sa place. En effet, tous les types relevés dans ce corpus amènent à la même conclusion : toutes les phrases grammaticales enfantines comportent un pronom clitique, celui-ci pouvant être utilisé seul ou en association avec un pronom tonique ou un nom. Cette conclusion est notamment possible car ces données ne comportent par ailleurs aucune phrase construite avec un nom préverbal seul. De plus, l'absence de clitiques postverbaux nous indique également que, lorsque l'enfant produit un clitique, celui-ci est obligatoirement préverbal. Le phénomène de « l'inversion » verbe-sujet que l'on peut trouver dans certaines phrases interrogatives ne s'applique donc pas dans ce système. Les phrases enfantines qui ne présentent pas une omission du sujet montrent donc une caractéristique constante : la réalisation d'un clitique préverbal.

Il est maintenant intéressant de confronter cette observation aux phrases sans sujet issues du même corpus. Le Tableau 8 ci-dessous détaille ainsi également la répartition des phrases enfantines qui présentent un cas de « sujet nul ».

¹ + = réalisation, – = omission ; Pr = préverbal, Po = postverbal ; C = pronom clitique, N = nom, P = pronom tonique.

Nature des omissions	Exemples	N	%
-C -N	<i>_ veux un camion</i>	151	14.1
+PrP -C	<i>moi _ fais</i>	33	3.1
-C +PoP	<i>_ veux vert moi</i>	33	3.1
-C +PoN	<i>_ est fermé le manège</i>	15	1.4
+PrP -C +PoP	<i>et moi _ veux ça moi</i>	8	0.7
Sous-total pour les phrases sans sujet		240	22.4
Sous-total pour les phrases avec sujet		832	77.6
Total corpus 1		1072	100.0

Tableau 8 : la nature des sujets enfantins omis (Corpus 1)

Nous constatons que les données ci-dessus ne falsifient pas notre première observation dans la mesure où le Tableau 8 présente une distribution complémentaire parfaite avec le Tableau 7 en ce qui concerne la nature du sujet enfantin. En effet, toutes les phrases « agrammaticales » montrent également une caractéristique commune : aucune ne comporte un clitique préverbal.

Ces deux tableaux nous permettent ainsi de formuler avec des termes et un point de vue différents le phénomène étudié. En effet, lorsque l'on parle de « sujet » nul, on postule que le phénomène est celui d'une omission d'une fonction, en l'occurrence celle du sujet, et le point de vue est de fait syntaxique. En revanche, si l'on décrit le phénomène comme étant celui d'une omission d'un élément « clitique préverbal », nous disons simplement, à ce stade de nos recherches, que l'enfant omet parfois de réaliser, dans le continuum phonique que représente une phrase, un segment phonologiquement dépendant du verbe fléchi et précédant celui-ci¹. Les seules indications objectives que nous retenons pour le moment sont donc la dépendance au niveau phonologique ainsi que la précédence. Or, il est à noter que notre formulation ne relève alors plus directement du domaine de la syntaxe puisque nous ne faisons pas mention de la notion relationnelle de « sujet ». Ce constat nous ramène à un débat ancien dans la littérature générative sur le statut des ces clitiques préverbaux : doit-on les considérer comme des arguments du verbe, et donc comme des « pronoms » au sens étymologique du terme, ou doit-on les analyser comme des préfixes verbaux ? La distinction est importante puisque, dans le premier cas, le segment relève de la syntaxe et exerce la fonction de sujet, alors que, dans le deuxième cas, le segment relève de la morphologie et représente un affixe verbal, dont il resterait à définir le signifié en rapport notamment avec les suffixes verbaux dont l'analyse ne suscite par contre aucune polémique. L'un des objectifs de cette

¹ Toutes les études s'accordent à dire que les enfants ne produisent (quasiment) jamais ce type de clitiques devant un verbe non fléchi. Cf. par exemple Pierce (1992:94), Hamann *et al.* (1994:101) ou Palasis (2005:205).

thèse est justement de revenir sur cette question, notamment par le biais de l'analyse d'un nouveau corpus. Mais, pour l'heure, il nous faut tout d'abord expliciter comment la grammaire générative rend compte du processus acquisitionnel en général et de l'acquisition du sujet en particulier.

II. LE PROCESSUS ACQUISITIONNEL

Dès lors que l'on postule l'existence d'universaux linguistiques (comme décrit dans la Section I.2.3 ci-dessus), il faut se demander : (i) sous quelle forme ces universaux sont présents dans les langues du monde et chez l'être humain dès la naissance, (ii) sous quelle forme se présentent les divergences entre les langues du monde, et (iii) comment chaque enfant reconnaît et acquiert les spécificités de sa (ses) langue(s) première(s). Dans le cadre de la grammaire générative, l'ensemble de ces questionnements est pris en charge à partir des années 1980 par la « Théorie des Principes et des Paramètres ».

En effet, depuis les années 1970, de plus en plus de langues sont étudiées par les générativistes et celles-ci semblent présenter des divergences morphosyntaxiques importantes. Une illustration de ces différences en ce qui concerne le sujet syntaxique et la morphologie verbale peut être mise en avant très simplement grâce à la comparaison de quatre paradigmes de conjugaison au présent de l'indicatif. Le Tableau 9 ci-après présente ainsi le verbe *être* en latin et en chinois¹, ainsi que le verbe *parler* en français et en trentin².

latin		trentin		français		chinois	
∅	sum	∅	parlo	je	parle	wo ³	shi ⁰
∅	es	te	parli	tu	parles	ni ³	shi ⁰
∅	est	el	parla	il	parle	ta ¹	shi ⁰
∅	sumus	∅	parlem	nous	parlons	wo ³ men ⁰	shi ⁰
∅	estis	∅	parlé	vous	parlez	ni ³ men ⁰	shi ⁰
∅	sunt	i	parla	ils	parlent	ta ¹ men ⁰	shi ⁰

Tableau 9 : illustration de la variation inter-linguistique concernant le sujet et la flexion verbale

Ces données montrent que la présence d'un élément préverbal obligatoire est nulle en latin, partielle en trentin et totale en français et en chinois. Les formes verbales quant

¹ Données comparées extraites de Jacquesson (2008:32) ; les exposants notent les différents tons du chinois.

² Dialecte de l'Italie du nord. Données extraites de Brandi (1989:113).

à elles sont phonétiquement distinctives pour les six personnes en latin, cinq en trentin, trois en français et aucune en chinois. Par ailleurs, il faut noter que ces quatre paradigmes n'épuisent pas les possibilités offertes par les langues du monde : les paradigmes des dialectes romans à eux seuls par exemple montrent de nombreuses autres combinaisons¹. Face à cette diversité des faits, un paradoxe se dégage alors : comment concilier le postulat d'une Grammaire Universelle commune à tous les hommes, avec le fait que les quelques milliers de langues humaines effectivement parlées semblent présenter des divergences morphosyntaxiques fondamentales ? Une telle diversité linguistique ne semble donc pas plaider *a priori* en faveur d'une théorie innéiste du langage. Pour résorber ce paradoxe, il faut alors définir des propriétés communes à toutes les langues et limiter la latitude des variations possibles à un strict minimum. La théorie des Principes et des Paramètres, formalisée par Chomsky (1981), est née de cette nécessité de concilier théorie et réalité. Une dichotomie est alors posée entre ces propriétés communes – appelées principes – et les propriétés divergentes – les paramètres. Concernant l'innéisme et le processus d'acquisition, des réponses sont ainsi également apportées. En effet, puisque les principes décrivent les contraintes communes à toutes les langues, ceux-ci peuvent faire partie de notre patrimoine génétique et ainsi être disponibles chez tous les êtres humains à leur naissance ; les principes sont donc innés. Il en découle que l'acquisition du langage se limite à la découverte et à la maîtrise d'une sous-partie du système linguistique seulement : celle qui présente des variations et qui est donc prise en charge par les paramètres. La restriction apportée par ce modèle à la quantité d'informations qui doit être acquise par un enfant génère alors une explication quant à la relative rapidité du processus et à la relative pauvreté quantitative et qualitative du stimulus nécessaire. Voyons comment cette théorie s'applique à l'acquisition du sujet syntaxique.

1. Les Principes

Les Principes sont donc les contraintes qui pèsent sur toutes les langues du monde. Si la terminologie est nouvelle dans les années 1980, la notion de contrainte en elle-même ne l'est pas. Chomsky illustre par exemple le Principe de la Dépendance à la Structure (*Structure Dependence Principle*) dès Chomsky (1968) et Chomsky (1975). Ce principe stipule que toutes les opérations grammaticales

¹ Cf. Oliviéri (2009).

s'effectuent sur la base de la structure de la phrase en syntagmes et pas sur la base de l'ordre linéaire des parties du discours dans la phrase. L'illustration classique de ce principe est apportée par la formation des questions oui/non en anglais¹. En ce qui concerne l'acquisition de cette tournure, Chomsky (1975) arrive à la conclusion suivante :

« L'enfant analyse la phrase déclarative en syntagmes abstraits ; ensuite, il repère la première occurrence de "is" [ou de tout autre auxiliaire] qui suit le premier syntagme nominal ; puis il antépose cette occurrence de "is" pour former la question correspondante. [...] l'enfant emploie une « règle dépendante de la structure », qui fait intervenir une analyse en mots et en syntagmes abstraits. Les syntagmes sont « abstraits » au sens où leurs limites et leur appartenance catégorielle (syntagme nominal, syntagme verbal, etc.) ne sont pas nécessairement marquées physiquement. »²

Dans les années 1980, les Principes représentent donc les contraintes communes à toutes les langues du monde et ils appartiennent à la Grammaire Universelle dont chaque être humain dispose à la naissance. Dans le cadre de notre étude sur le sujet syntaxique, deux principes en particulier nous intéressent. Il s'agit du Principe de Projection Étendu et le Principe des Catégories Vides.

1.1 Le Principe de Projection Étendu

Comme son nom l'indique, le Principe de Projection Étendu (*Extended Projection Principle : EPP*) représente un amendement à un principe déjà existant : le Principe de Projection. Ce dernier prévoit que toutes les informations argumentales et thématiques contenues dans le lexique se retrouvent projetées dans la syntaxe (par le biais de la projection des quatre catégories lexicales que sont le nom, l'adjectif, le verbe et la préposition)³. Or, ce principe pose des problèmes notamment en ce qui concerne la fonction sujet. En effet, certains éléments étiquetés comme

¹ Nous détaillons cette démonstration dans Palasis (2005:27-28).

² Chomsky (1975:43). Cf. le protocole de sollicitation de Crain & Nakayama (1987) expliqué au chapitre III et qui entérine cette hypothèse.

³ Cf. Chomsky (1981:29) : "Representations at each syntactic level (i.e., LF, and D- and S-structure) are projected from the lexicon, in that they observe the subcategorization properties of lexical items." « Les représentations à chaque niveau syntaxique (LF, D- et S-structure) sont issues du lexique, dans la mesure où elles respectent les propriétés de sous-catégorisation des items lexicaux. »

« sujets » dans la syntaxe ne sont pas le reflet de la distribution de thêta-rôles dans le lexique. Ces « contrevenants » sont notamment représentés en français par le pronom explétif *il* dans une phrase comme *il semble que la théorie a évolué*. Le verbe *sembler* n'attribue en effet pas de thêta-rôle externe et, pourtant, la place du sujet syntaxique est remplie. Les verbes météorologiques sont également des exemples bien connus de ce cas de figure. Le Principe de Projection initial est alors étendu afin de corriger cette contradiction. Ce Principe de Projection Étendu est ainsi défini par Chomsky comme étant « la combinaison du Principe de Projection et de la condition qui rend les sujets obligatoires pour les propositions »¹. L'insertion d'un pronom explétif à S-Structure est ainsi expliquée par ce nouveau principe formalisé en 1982.

Cependant, depuis la fin des années 1970, un intérêt particulier et nouveau est porté à l'italien qui, comme chacun le sait, présente des divergences fondamentales en termes de sujet, notamment dans la mesure où cette langue peut présenter des phrases qui ne comportent aucun sujet réalisé phonétiquement, que celui-ci soit référentiel ou explétif. L'élément *pro* est alors également introduit dans le modèle génératif à cette époque. Il s'agit du pendant du pronom phonétiquement réalisé car, dans la typologie des catégories nominales, c'est un élément pronominal et non-anaphorique mais la différence se situe au niveau de la réalisation de surface puisque *pro* n'a pas d'enveloppe phonique. Il permet cependant de rendre compte des phrases telles que *parla* en italien ou *miláo* en grec dans le cadre du Principe de Projection Étendu dans la mesure où *pro* représente le sujet obligatoire². S'agissant d'un principe, l'information sur la présence obligatoire de projections destinées à accueillir le sujet syntaxique au cours de la dérivation n'a pas à être acquise au cours des premières années de l'enfant ; cette information est innée. Il reste donc à l'enfant à déterminer si sa langue première possède des sujets phonétiquement réalisés ou non. Dans la mesure où cette information représente une variation entre les langues, celle-ci n'est plus gérée par un principe mais par un paramètre ; elle doit donc être acquise par le jeune enfant.

Adapté à la terminologie du Programme Minimaliste, *EPP* est rebaptisé « trait-*EPP* » (*EPP-feature*). Il opère alors par le biais de la vérification et de l'effacement d'un trait, le trait fort D présent sur la projection fonctionnelle T³. Deux opérations

¹ Chomsky (1982:87).

² La représentation est : [_{IP} *pro*_i *miláo*_j [_{VP} *t*_i *t*_j]].

³ Nous revenons sur toutes ces notions dans la section suivante.

sont alors possibles : (i) soit le sujet est un argument du verbe et il se déplace, comme mentionné précédemment, de [Spec, VP] à [Spec, T] pour vérifier et effacer ce trait fort sur la tête T (il s'agit de l'opération *Move*) ; (ii) soit le sujet est un explétif et il est fusionné directement dans cette position sans qu'il y ait recours à un mouvement mais cette opération permet tout de même de vérifier et d'effacer le trait fort D (il s'agit de l'opération *Merge*). Plusieurs variantes de ce principe sont développées dans le cadre du Programme Minimaliste ; nous revenons sur ces différentes hypothèses dans la Section III.6 ci-dessous.

1.2 Le Principe des Catégories Vides

Un second principe formulé dans le cadre du Gouvernement et du Liage (*Government and Binding, GB*) relève directement de notre thématique sur le sujet et est lié à *EPP*. Il s'agit du Principe des Catégories Vides (*Empty Category Principle* ou *ECP*) selon lequel une catégorie vide doit être « proprement gouvernée »¹. Un élément A (une tête) gouverne proprement un élément B si : (i) A gouverne B (c'est-à-dire si A m-commande B et qu'il n'a pas de projections maximales entre A et B) et (ii) A thêta-marque B ou A est co-indicé avec B. Les catégories vides possèdent les mêmes propriétés que les catégories phonétiquement réalisées ; elles sont représentées soit par un élément non réalisé (*pro* par exemple), soit par la trace laissée par le déplacement d'un élément (*t*). Dans ce cadre, *pro* en italien ou en grec est proprement gouverné par AGR dans la syntaxe². L'idée sous-jacente relève de la richesse de l'information que véhicule le verbe fléchi quant aux traits de [Personne]. Ce principe sert également à réguler la latitude des mouvements possibles dans une phrase dans la mesure où seuls les mouvements qui laissent une trace proprement gouvernée sont licites. Cette notion de gouvernement rend alors notamment compte de l'asymétrie qui existe dans le déplacement des sujets et des objets. Pour le français par exemple, l'extraction de l'objet enchâssé en (3a) ci-dessous est licite alors que l'extraction du sujet en (3b) ne l'est pas puisque V gouverne son objet mais pas son sujet.

(3) *ECP* et l'asymétrie dans l'extraction des arguments internes vs externes :

- (a) *Quel enfant_i crois-tu que la maîtresse choisira t_i ?*
- (b) **Quel enfant_i crois-tu que t_i viendra ?*

¹ Cf. Chomsky (1981:250).

² AGR = morphologie verbale. Cf. Chomsky (1981:253-278) d'après les travaux de Rizzi (1982) sur l'italien.

En termes minimalistes, un tel principe doit être reformulé. En effet, la relation de gouvernement n'est pas une notion minimaliste dans la mesure où c'est la relation entre nœuds « sœurs »¹ qui détermine le déroulement des opérations ; le gouvernement, plus complexe, est donc abandonné. Nous verrons, lors de notre analyse des éléments *qu-* dans le Chapitre IV, comment la théorie a évolué à ce sujet.

1.3 Vers des propriétés moins spécifiquement linguistiques

L'utilisation de ces Principes vise ainsi à atteindre les deux objectifs initiaux de la grammaire générative que sont la description et l'explication de l'acquisition du langage. Cependant, à partir des années 2000, le Programme Minimaliste cherche à aller plus loin encore, « au-delà de ce caractère explicatif »². Plusieurs points de vue sont ainsi radicalement novateurs dans le Programme Minimaliste. Tout d'abord, Chomsky (2001) souhaite revenir sur la description de la Faculté de langage (FL) qui représente l'état initial de la grammaire chez l'être humain. En effet, la formulation présentée ci-dessus (en termes de principes communs et de paramètres particuliers) lui paraît incomplète. Ainsi, dans Chomsky (2001), l'auteur propose une nouvelle description de FL, répartissant notamment les caractéristiques de cette faculté en deux types : celles qui sont effectivement directement dépendantes de la capacité linguistique (celles qui correspondent donc à la Grammaire Universelle) et d'autres qui, *a contrario*, ne sont pas spécifiques au langage mais qui représentent des « propriétés générales des systèmes biologiques »³. L'acquisition du langage dépend ainsi non plus de deux mais de trois « facteurs » : (i) un état linguistique initial génétiquement déterminé, (ii) l'expérience de chaque locuteur, et (iii) des propriétés non spécifiquement linguistiques. Ce programme de recherches est donc « minimaliste » dans le sens où, à ce stade de la réflexion, on cherche à réduire le facteur (i) – spécialisé et par conséquent peu « rentable » – à un strict minimum et à faire ainsi passer des principes qui avaient été jusque là définis comme linguistiques, vers le domaine du facteur (iii), c'est-à-dire celui des propriétés plus générales^{4,5}. Comme le mentionne Chomsky (2007), le point de vue est donc novateur :

¹ Ce sont deux nœuds dominés par un même constituant.

² Titre de Chomsky (2001), “Beyond Explanatory Adequacy”.

³ Chomsky (2001:105).

⁴ La position la plus extrême étant alors de postuler qu'aucun principe ne dépend plus de ce facteur (i) ; il s'agit d'une hypothèse de travail nommée *Strong Minimalist Thesis (SMT)* (cf. par exemple Chomsky (2001:106)).

⁵ Chomsky explique cet intérêt pour ce facteur (iii) en renvoyant notamment aux pensées de Galilée (1632) sur « la perfection de la nature ». Cf. Galilée (1632:397) : “nature is perfect and simple, and creates nothing in vain” (citation mentionnée par Boeckx (2008:9)).

“Throughout the modern history of generative grammar, the problem of determining the character of FL has been approached “from top down”: How much must be attributed to UG to account for language acquisition? The MP seeks to approach the problem “from bottom up”: How little can be attributed to UG while still accounting for the variety of I-languages attained, relying on third factor principles?”¹

Dans le Programme Minimaliste, tant que l’hypothèse la plus forte concernant le facteur (i) n’est pas retenue, il est généralement postulé que GU se compose d’un ensemble de traits linguistiques F^2 . Deux options sont alors envisageables. La première consiste à postuler que l’ensemble de tous les traits possibles appartient à GU. Tous les traits sont ainsi disponibles initialement et intégralement, quelle que soit la langue acquise³ :

“[...] no particular acquisitional issue arises from considering the structure of the clause provided by UG as rich and articulated [...]. The obvious consequence from assuming [its] universality is that less is left for the child to acquire.”⁴

La seconde option consiste à poser qu’un choix doit être opéré au sein des projections possibles, en fonction de la langue acquise. Dans les deux cas, intervient alors la notion de paramétrage qui est directement liée à celle de la variation inter-linguistique.

2. Les Paramètres

Friedemann & Rizzi (2000) donnent une excellente vue d’ensemble de la grammaire du français dans le cadre des Principes et des Paramètres :

¹ Chomsky (2007:3) : « Tout au long de l’histoire récente de la grammaire générative, la question de déterminer la nature de FL a été abordée « de haut en bas » : quelles sont les limites maximales que l’on peut attribuer à GU afin de rendre compte de l’acquisition du langage ? Le PM cherche à approcher la question « de bas en haut » : quelles sont les contours minimaux qui peuvent être attribués à GU tout en rendant compte de la diversité des langues internes atteintes, en se basant sur les principes du troisième facteur ? »

² Cf. par exemple Chomsky (2000:100).

³ Cette thèse est soutenue par Cinque (1999) par exemple qui pense également que toutes les projections sont présentes dans toutes les phrases produites.

⁴ Cinque (1999:107) : « [...] considérer la structure de la proposition fournie par GU comme riche et articulée ne soulève aucun problème particulier [...]. La conséquence évidente qui découle de cette universalité supposée est que l’enfant a moins de choses à acquérir. »

“The grammar of French is UG with the headedness parameter fixed on the ‘head initial’ value, the null subject and V-2 parameters fixed negatively, the clitic parameters fixed positively [...] and so on.”¹

Ces quelques exemples illustrent le fait que les paramètres sont conçus comme des choix que l’enfant fait au cours de son processus d’acquisition dans un cadre très restreint dans la mesure où l’éventail des possibilités est limité et constant. En effet, les paramètres sont tous binaires. Ainsi, le Paramètre de Tête (*the Headedness Parameter*) indique que la tête d’un syntagme soit précède son complément², soit le suit³. Le Paramètre V-2 quant à lui fixe la place du verbe dans la proposition sur la deuxième position⁴ et le Paramètre qui gère les clitiques indique l’hôte susceptible de les accueillir⁵. Cependant, l’impression globale qui se dégage de la littérature générative sur ce modèle est que le paramètre qui a suscité et suscite encore le plus grand nombre de travaux est celui qui gère l’apparition phonique du sujet syntaxique dans une phrase : le Paramètre *pro-drop*⁶.

2.1 Le Paramètre *pro-drop*

2.1.1 Ses origines

Dès le début des années 1970, Perlmutter met en avant deux caractéristiques du français. En effet, Perlmutter (1971) constate que : (i) il est impossible d’extraire un sujet lorsque celui-ci se trouve dans une proposition enchâssée et (ii) les pronoms sujets sont obligatoires même lorsque la morphologie verbale permet de retrouver le pronom omis⁷ :

¹ Friedemann & Rizzi (2000:3) : « La grammaire du français correspond à GU avec le paramètre qui régit la position de la tête dans le syntagme fixé sur la valeur ‘tête devant son complément’, les paramètres du sujet nul et V-2 fixés négativement, les paramètres qui régissent les clitiques fixés positivement, etc. »

² C’est le cas du français où le verbe précède le complément par exemple.

³ Le turc et le coréen, entre autres, illustrent cette possibilité.

⁴ Ce paramètre est donc positif en allemand et négatif en français.

⁵ L’hôte des clitiques objets en français est le verbe par exemple.

⁶ Cf. Ayoun (2003:78-80) qui propose un inventaire total d’une trentaine de paramètres syntaxiques ainsi qu’une liste des chercheurs qui les ont étudiés.

⁷ Cf. Perlmutter (1971:99-122).

(4) Les contraintes mises en avant par Perlmutter (1971) pour le français :

(a) Impossibilité d'extraire un sujet enchâssé :

*Qui_i a-t-il dit [que t_i s'est évanoui] ?

*La speakerine_i [qu'il a dit [que t_i s'est évanouie]].

(b) Impossibilité d'omettre un pronom sujet :

*(Nous) avons travaillé toute la journée.

L'auteur corrèle ces deux caractéristiques dans la mesure où il leur donne une explication commune, sous la forme d'une « contrainte de surface » contenue dans la grammaire du français :

“Any sentence other than an Imperative in which there is an S [Sentence] that does not contain a subject in surface structure is ungrammatical.”¹

Sur la base de ces deux propriétés issues d'une même contrainte, Perlmutter (1971) établit une typologie binaire des langues : les langues qui possèdent cette contrainte dans leur grammaire, telles que le français ou l'anglais, et les langues qui ne la possèdent pas, telles que l'italien ou le serbo-croate. Cette corrélation, connue sous le nom de « généralisation de Perlmutter »², est ensuite étendue à plusieurs autres phénomènes puisque Chomsky (1981) décrit un « faisceau » de cinq caractéristiques³ :

¹ Perlmutter (1971:100) : « Toute phrase, à part les impératives, qui comprend une proposition sans sujet en structure de surface est agrammaticale. »

² *Perlmutter's generalization*. Cf. par exemple Chomsky (1981:174).

³ Chomsky (1981:240).

(5) Les cinq propriétés regroupées par Chomsky (1981)¹ :

(a) Absence du sujet :

Ho trovato il libro

« *(j') ai trouvé le livre »

(b) Inversion libre dans les phrases simples :

Ha mangiato Giovanni

« *a mangé Jean »

(c) Déplacement long du sujet wh- :

L'uomo [che mi domando [chi abbia visto]]

« *l'homme_i [que je me demande [t_i a vu qui]] »

(d) Pronoms anaphoriques nuls dans les propositions enchâssées :

Ecco la ragazza [che mi domando [chi crede [che possa VP]]]

« *c'est la fille [que je me demande [qui croit [que (elle) pourrait VP]]] »

(e) Violation apparente du filtre *[que-trace]² :

Chi credi [che partirà]

« *qui_i crois-tu [que t_i partira] »

Chomsky (1981) pose alors une corrélation forte entre ces cinq propriétés au sein de chaque langue puisqu'il suggère qu'elles relèvent toutes d'un seul paramétrage par le biais d'un paramètre unique baptisé *pro-drop parameter*, nom attribué d'après la première des caractéristiques, l'omission du pronom. Sont ainsi distinguées deux catégories de langues : les langues *pro-drop* comme l'italien ou le grec et les langues non *pro-drop* comme l'anglais ou le français, comme le montrent nos gloses agrammaticales pour le français en (5) ci-dessus. Étant globale, une telle analyse est qualifiée dans la littérature de « macro-paramétrique »³. En tout cas, ce paramètre est une réponse théorique aux principes issus du même modèle, notamment le principe de projection étendu (*EPP*) qui stipule que toute phrase doit avoir un sujet. Les langues non *pro-drop* ont un sujet réalisé (minimalement réalisé sous la forme d'un pronom explétif) et les langues *pro-drop* ont également un sujet (*pro*) mais ce dernier

¹ D'autres propriétés, telles que l'accord vers la droite dans les constructions avec copule (*sono io* « *suis je » suggéré par Giuseppe Longobardi, sont mentionnées par Chomsky (1981:280) uniquement sous forme de note et ne font donc pas partie de la liste principale des propriétés rapportées par l'auteur.

² Impossibilité d'extraire le pronom interrogatif de la subordonnée et de le déplacer dans la principale par-dessus un complémenteur (en l'occurrence « que »).

³ Cf. par exemple Manzini & Savoia (2002:194).

n'est pas phonétiquement réalisé ; *pro* est donc l'équivalent non réalisé d'un pronom :

- (6) Français : *il* aime *il* pleut
Grec : *pro* agapa *pro* vrèchi

2.1.2 Ses limites et les amendements proposés

Cependant, dès ses débuts, le paramètre *pro-drop* ne fait pas l'unanimité et de nombreux points sont débattus. Ainsi, la liste des propriétés qui permet d'établir une telle typologie des langues varie. Chomsky (1981) lui-même met en avant la possibilité que les propriétés détaillées ci-dessus puissent se réduire aux deux premières. Rizzi (1982) pour sa part fait également référence aux critères regroupés en (5) ci-dessus, mais en ajoutant une propriété additionnelle qui réfère à la richesse morphologique des paradigmes verbaux¹. L'appellation proposée par Chomsky n'est pas non plus systématiquement adoptée. Rizzi (1982), entre autres, privilégie le terme *null subject parameter*², et la littérature francophone utilise généralement la traduction de cette appellation : « Paramètre du sujet nul ». En ce qui nous concerne, à ce stade de la réflexion, nous conservons l'appellation originale de Chomsky – *pro-drop* – parce qu'elle n'induit qu'une référence à la nature de l'élément incriminé – son aspect pronominal – et que, par conséquent, celle-ci n'oblige pas à se prononcer quant à la fonction de cet élément au sein de la phrase³.

Plus généralement, la capacité même d'un paramètre binaire à rendre compte de la très grande diversité des langues du monde est mise en doute. Ainsi, une certaine « souplesse » est introduite à différents niveaux du paramétrage, à commencer par l'élément *pro* lui-même. En effet, celui-ci n'est plus considéré comme une entité unique mais comme un élément couvrant des réalités diverses : *pro* purement explétifs (du type *il y a*), *pro* « quasi-argumentaux » (pour les verbes météorologiques) et *pro* argumentaux⁴. La classification des langues n'est donc plus binaire mais quaternaire : langues qui ne possèdent aucun type de *pro* (comme

¹ Cf. Palasis (2005:33-48) pour le détail de ses travaux et une discussion sur chaque critère proposé.

² Rizzi (1982:173,fn1) : “‘Null subject’ is preferred here to the more usual label ‘Pro drop’ since in the analysis I will propose (as in many other recent accounts) no pronoun deletion rule is involved”. « L'étiquette « Sujet nul » est préférée ici à celle plus usuelle de « Pro drop » dans la mesure où dans l'analyse que je propose (parmi de nombreuses autres études récentes), aucune règle de suppression du pronom n'est impliquée. »

³ Rejoignant ainsi la formulation adoptée à la suite de notre première analyse mentionnée à la Section I.3.4

⁴ Cf. Rizzi (1986b) dont les arguments sont discutés par Cabredo Hofherr (2003).

l'anglais), celles qui ne possèdent que des *pro* explétifs (comme l'allemand), celles qui possèdent des *pro* explétifs et des *pro* quasi-argumentaux (comme le yiddish) et celles qui présentent les trois types de *pro* (comme l'italien). Ces distinctions induisent également une formulation plus fine du paramètre *pro-drop*. Rizzi (1997b) notamment propose une distinction entre : (i) le « licenciement » de *pro* qui autorise la présence de cette catégorie phonétiquement nulle (cette condition oppose ainsi l'italien au français par exemple) et (ii) son « identification » qui permet d'obtenir une interprétation référentielle ou non-référentielle pour les langues qui licencient *pro*. La notion de « *pro-drop* partiel » est également envisagée. Dans ce cadre, Deal (2005) par exemple relève que, dans certaines langues telles que le finnois ou l'hébreu, seules les personnes 1 et 2 peuvent être omises alors que, dans d'autres langues telles que le français, seule la personne 3 peut être omise. L'auteur postule ainsi l'existence de deux paramètres indépendants : un paramètre régissant l'omission des pronoms de personnes 1 et 2 (qu'elle nomme *1st/2nd person pro-drop*) et un autre paramètre (*3rd person pro-drop*) régissant les omissions de troisième personne. Par ailleurs, la littérature générative soulève également largement le problème d'un certain nombre de langues face à ce paramètre, notamment le chinois, le japonais et le coréen qui présentent des éléments non réalisés phonétiquement mais qui ne se conforment pas aux critères *pro-drop* stipulés ci-dessus¹. Toutes ces réflexions menées depuis la création du paramètre militent ainsi en faveur d'une approche plus micro-paramétrique de la problématique.

Ce type de traitement est également développé dans le cadre de l'étude des données dialectales, notamment en France² et en Italie³. Ainsi, d'un seul paramètre qui gère un faisceau de propriétés considérées comme corollaires, nous passons à plusieurs paramètres qui prennent en charge chacun un phénomène unique, très limité et indépendant des autres phénomènes qui étaient perçus comme solidaires. Manzini & Savoia (2002) par exemple expliquent ce point de vue par le fait que « les paramètres sont des choix lexicaux »⁴. Dans ce cadre, le paramètre *pro-drop* implique alors la lexicalisation de D, tête universelle. Les langues *pro-drop* lexicalisent D par le biais du verbe (comme en italien) ; les langues à clitiques le

¹ L'une des premières discussions est attribuée à Huang (1984). Cf. Neeleman & Szendrői (2007) par exemple pour un exposé récent de la discussion.

² Cf., entre autres, Dalbera (1994) et Oliviéri (2006).

³ Cf., entre autres, Brandi & Cordin (1989), Poletto (2000) et Manzini & Savoia (2002).

⁴ Cf. Manzini & Savoia (2002:194) : “The strictly local and independent character of the parameters derives from the fact that they effectively reduce to lexicalization choices.”

lexicalisent par une tête spécialisée (dans les dialectes du nord de l'Italie) ; alors que les langues non *pro-drop* le lexicalisent grâce à un DP (en anglais). Dans ce système, le français apparaît alors comme faisant partie d'un quatrième et dernier type : un type mixte où le trait D se réalise tantôt sur la tête D (comme les dialectes du nord de l'Italie), tantôt sur le DP (comme en anglais). Dans la Section III.6 ci-dessous, après avoir explicité les nouveautés du Programme Minimaliste, nous reviendrons plus en détail sur ces reformulations récentes à la fois du paramètre *pro-drop* et des principes attenants.

2.2 La fixation des paramètres

Quelle que soit la formulation adoptée concernant le paramètre *pro-drop* et les autres paramètres décrits et étudiés dans le cadre générativiste, la problématique acquisitionnelle reste néanmoins constante. En la restreignant à notre champ d'investigation, la thématique s'articule autour des questions suivantes : (i) quel est l'état initial d'un paramètre ? (ii) à quel moment l'enfant lui donne-t-il une valeur ? (iii) cette valeur peut-elle être fautive et donc modifiée ?

Concernant l'état initial des paramètres, il semble que toutes les hypothèses soient envisagées. Fodor (1998) par exemple postule que les paramètres connaissent un état initial au cours duquel leur valeur est « non spécifiée » et que l'enfant ne fait son choix que lorsqu'il a reçu des données non ambiguës qui lui permettent ainsi de déterminer à quelle catégorie sa langue appartient. *A contrario*, Chomsky (2001) par exemple suggère que chaque paramètre a une valeur initiale par défaut qui est sa valeur « non marquée » (*unmarked value*). Dans un premier temps, cette hypothèse nous semble plus délicate à définir dans la mesure où il faudrait objectiver la notion de « non marquée » pour chaque paramètre, ce qui ne nous apparaît pas comme une tâche aisée. En second lieu, cette hypothèse nous semble moins économique et universelle à appliquer que la première car elle implique la possibilité que la valeur initiale soit erronée, ce qui aurait pour conséquence de devoir corriger cette première programmation alors que, dans le cas de l'hypothèse de Fodor (1998), tous les enfants du monde possèdent le même état initial et tous les enfants du monde n'ont qu'une seule et même tâche à réaliser, celle de définir la valeur qui correspond à leur langue première. Cependant, ce débat peut ne jamais trouver d'issue car, en ce qui concerne la syntaxe, les enfants produisant leurs premiers énoncés vers l'âge de dix-huit mois, cela laisse une marge d'au moins dix-huit mois entre le moment où

l'enfant commence à recevoir des informations et le moment où il entame sa propre production syntaxique. Ainsi, à moins de recueillir des données par le biais de tests de compréhension¹, il nous paraît impossible de trancher sur cette question à un niveau empirique. Néanmoins, pour des raisons théoriques d'universalité, nous optons en faveur de l'hypothèse d'un paramètre dont la valeur initiale est non spécifiée ; l'enfant en fixe alors la valeur au cours des premiers mois de sa vie, avant de produire ses premières combinaisons syntaxiques.

Dans ce cadre, nous suivons alors notamment l'hypothèse de Wexler (1998) sur une « fixation précoce des paramètres » (*Very Early Parameter Setting*) et, concernant le point (iii), nous optons pour une fixation correcte de leurs valeurs. Les productions enfantines donnent des indications quant à certains paramètres. L'ordre des mots par exemple ne semble généralement pas enfreint par les jeunes enfants, que leur langue présente une structure plutôt Verbe-Objet (comme en français, en anglais ou en italien) ou une structure plutôt Objet-Verbe (comme en japonais ou en coréen). Ces indications privilégient ainsi l'hypothèse d'une fixation correcte et précoce du paramètre qui régit la place de la tête dans le syntagme. Nous avons vu que les données paraissent moins tranchées en ce qui concerne le paramètre *pro-drop* puisque les enfants francophones ou anglophones par exemple omettent certains de leurs sujets. Une conséquence de ce phénomène est que l'une des hypothèses majeures portant sur la fixation de ce paramètre – l'hypothèse de Hyams (1986) – postule que les enfants acquérant une langue non *pro-drop* démarrent leur acquisition en fixant le paramètre sur la mauvaise valeur². Or, cette fixation erronée n'est pas tenable face à l'étude du système enfantin dans sa totalité. L'auteur développe d'ailleurs par la suite d'autres hypothèses quant à cette problématique³. Il n'en demeure par moins qu'une telle conjecture a eu énormément de succès et que la notion de la valeur erronée fait ainsi partie des pistes envisagées dans ce domaine. Nous n'adoptons pas ce point de vue et nous postulons que le paramétrage est donc initialement non spécifié, qu'il est activé dans les premiers mois de la vie de l'enfant,

¹ Cf. Chapitre III.

² Dans le cadre des hypothèses sur le déficit de la compétence linguistique des enfants mentionnées précédemment.

³ Cf. par exemple Hyams (1992) sur l'uniformité morphologique.

que sa première valeur est correcte et qu'elle n'a donc pas à être modifiée par la suite¹.

2.3 Les Paramètres dans le Programme Minimaliste : syntaxiques ?

Dès Chomsky (1995b), on constate une volonté de restreindre fortement le domaine de la variation paramétrique aux traits présents sur les têtes fonctionnelles² :

“Language differences and typology should be reducible to choice of values of parameters. [...]. One proposal is that parameters are restricted to *formal features* with no interpretation at the interface. A still stronger one is that they are restricted to formal features of functional categories [...]. I will assume that something of the sort is correct [...].”³

Plus récemment :

“Parametric variation is restricted to the lexicon, and insofar as syntactic computation is concerned, to a narrow category of morphological properties, primarily inflectional.”⁴

Ainsi, dans la littérature de ces dernières années, il apparaît que la frontière entre syntaxe et morphologie soit de plus en plus difficile à définir et à positionner. Peut-on d'ailleurs parler de morphologie d'un côté et de syntaxe d'un autre côté ? Dans la mesure où, dans ce modèle, les computations syntaxiques sont régies par des traits et que ces traits relèvent de la morphologie, il semble que le terme de morphosyntaxe soit plus approprié. De plus, dans le cadre de notre étude, l'une des thématiques centrales reste la question du rapport entre la *morphologie* verbale et la réalisation du sujet *syntaxique*. Afin d'éclaircir ce débat, il faut à présent nous plonger dans les détails techniques de la théorie minimaliste actuelle.

¹ Cette hypothèse est donc cohérente par rapport à la notion de systèmes apparentés développée en parallèle avec la dialectologie. Cf. Section I.3.3.

² Toutes les notions théoriques de ce modèle sont explicitées dans la section suivante.

³ Chomsky (1995a:6) : « Les différences inter-linguistiques et la typologie des langues devraient être réductibles à un choix de valeur des paramètres. [...]. Il est proposé que les paramètres soient restreints à des traits formels, sans interprétation à l'interface. Une proposition plus forte est qu'ils sont restreints aux traits formels des catégories fonctionnelles. [...]. Je suppose que ce type d'hypothèse est correct. »

⁴ Chomsky (1999:2) : « La variation paramétrique est restreinte au lexique et, en ce qui concerne la computation syntaxique, [elle est restreinte] à une catégorie limitée de propriétés morphologiques, principalement flexionnelles ».

III. LE PROGRAMME MINIMALISTE

La grammaire générative recherche, depuis ses débuts, un certain niveau d'abstraction afin de s'ouvrir des horizons théoriques plus larges que ceux qui avaient été explorés jusqu'alors¹. Les représentations sont donc symboliques, ce qui différencie le générativisme du connexionnisme notamment, dont les représentations sont d'inspiration physiologique². La représentation hiérarchisée et arborescente de la structure de la phrase est donc l'une des marques de fabrique bien connues et constantes du générativisme. Ce qui a beaucoup évolué en revanche est l'unité minimale qui est représentée par chacune des branches de ces arbres : des matrices de traits insérées par le biais des Règles d'insertion lexicale dans la Théorie Standard dans les années 1960³, le modèle est passé aux projections des catégories majeures avec le modèle X-Barre de la Théorie Standard Étendue des années 1970⁴, à la projection additionnelle de catégories fonctionnelles dans le modèle du Gouvernement et du Liage des années 1980⁵, pour arriver à un foisonnement de projections dans le modèle actuel, dans la mesure où chaque projection correspond désormais à un trait et que le nombre de traits, notamment fonctionnels, a connu une inflation sans précédent ces dernières années⁶.

Le modèle actuel de la grammaire générative s'intitule donc le « Programme Minimaliste ». Chomsky insiste d'ailleurs beaucoup sur cette notion de Programme⁷. Le point de départ de ce modèle se situe dans les années 1990 avec notamment la publication de Chomsky (1995b) qui regroupe des textes rédigés initialement entre 1986 et 1993. Depuis, on compte de nombreuses publications de Chomsky lui-même⁸, un grand choix de manuels⁹, ainsi que de très nombreux articles et ouvrages

¹ Cf. notamment Chomsky (1957:15-28) et la description de « l'impasse » que représente le modèle de Markov basé sur la linéarité des morphèmes dans la phrase.

² Cf. les réseaux neuronaux mentionnés précédemment.

³ Cf. Chomsky (1965).

⁴ Cf. Chomsky (1975, 1977a). Les catégories majeures sont : le nom (N), l'adjectif (A), le verbe (V) et la préposition (P).

⁵ Cf. Chomsky (1981, 1982, 1986). Les catégories fonctionnelles sont : la flexion (I) et le complémenteur (C).

⁶ Cf. par exemple les quelques trente projections fonctionnelles proposées par Cinque (1999:106) pour rendre compte du mode, temps, aspect, voix, etc., dans une proposition.

⁷ Cf. par exemple Chomsky (2007) : « [...] there has been so much misunderstanding, even within professional circles, that it is perhaps worth reiterating that it is a *program*, not a *theory*, [...] » « Il y a eu tant de malentendus, même chez les professionnels, qu'il est peut-être nécessaire de répéter qu'il s'agit d'un *programme* et pas d'une *théorie*. »

⁸ Cf. notamment Chomsky (1999, 2000, 2001, 2004, 2005, 2007) et Chomsky *et al.* (2002).

⁹ Cf. Radford (1997), Ouhalla (1999), Adger (2003), Hornstein *et al.* (2005) pour n'en citer que quatre.

concernant des questions théoriques de plus en plus pointues et des langues de plus en plus diverses. Ce Programme suscite donc de nombreuses recherches depuis une quinzaine d'années. Par ailleurs, il s'agit d'un modèle dérivationnel, c'est-à-dire que la structure syntaxique finale d'une phrase est obtenue grâce à une succession plus ou moins importante d'opérations. Ces opérations sont déclenchées par la présence d'éléments appelés « traits ». Mais avant d'entrer dans le détail de l'utilisation de ces traits, passons en revue quelques nouveautés de ce modèle.

1. Quelques nouveautés du modèle

Comme son nom l'indique, ce modèle se veut profondément économique en termes d'items théoriques nécessaires à la description et à l'explication des phénomènes syntaxiques. Tous les outils de *GB* pour lesquels il est démontré qu'ils ne sont pas indispensables sont ainsi abandonnés. Les structures intermédiaires (D-Structure et S-Structure) deviennent obsolètes face à une dérivation unique ; les projections X-barre, également intermédiaires, ne sont réalisées que si nécessaires ; les opérations se réduisent à *Agree* (accord), *Merge* (fusion de deux constituants), et *Move* (déplacement, mais en dernier recours uniquement car syntaxiquement plus coûteux). Le Programme Minimaliste laisse également derrière lui la notion de trace et le principe des catégories vides (*ECP*) par le biais de la Condition d'Inclusivité, telle que formulée par Chomsky (1995a) :

“A ‘perfect language’ should meet the condition of inclusiveness: any structure formed by the computation [...] is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation apart from rearrangements of lexical properties (in particular, no indices, bar levels [...]).”¹

Cette condition implique que, lorsqu'un élément se déplace, celui-ci ne laisse plus une trace derrière lui (puisque'il s'agirait d'un nouvel objet) mais une copie de lui-même, comme illustré par Chomsky (1995b) :

¹ Cf. Chomsky (1995a:228) : « Une langue parfaite devrait remplir la condition d'inclusivité : toute structure formée par la computation [...] est constituée d'éléments déjà présents dans les items lexicaux sélectionnés pour N ; aucun objet nouveau n'est ajouté dans le courant de la computation sauf les réarrangements de propriétés lexicales (en particulier, pas d'indices, de niveaux X-barre [...]). »

- (7) John wondered [which picture of himself] Bill saw [which picture of himself]¹
« Jean se demandait [quelle photo de lui-même] Bill a vu [quelle photo de lui-même] »

Le constituant peut alors être interprété dans l'une ou l'autre de ses positions, d'où la double lecture de *himself* ci-dessus (qui peut référer à John ou à Bill)². Les éléments présents dans la dérivation syntaxique sont extraits un à un de la numération N. Cette numération est définie comme étant un « ensemble de paires (LI, *i*), où LI est un élément du lexique et *i* son exposant, c'est-à-dire le nombre de fois que LI est sélectionné »³. Le lexique comporte uniquement les informations idiosyncrasiques de chaque item c'est-à-dire celles dont la composante computationnelle a besoin pour produire un objet interprétable, en fin de dérivation, à la fois par la sémantique et la phonologie (les interfaces logique et phonique). Ces informations sont livrées sous formes de traits : matrices phonologiques, propriétés sémantiques et traits formels, tels que le genre, le nombre ou le cas⁴.

2. Les différents types de traits

Les traits formels sont morphosyntaxiques et classés en deux grandes catégories : les traits « interprétables » et les traits « ininterprétables ». Cette opposition repose sur leur interprétation sémantique. Une phrase est grammaticale lorsqu'à la fin de la dérivation, il ne subsiste que des traits interprétables. On dit alors que la dérivation « converge ». Les traits ininterprétables doivent donc être effacés dans le courant de la dérivation sans quoi celle-ci « capote » et la phrase est agrammaticale.

À l'intérieur de ces deux catégories, « interprétables » et « ininterprétables », les traits se regroupent en plusieurs sous-classes. Ainsi, les traits morphologiques de [Personne], de [Nombre] et de [Genre] forment une première sous-classe appelée traits « Phi » (ϕ). Ces traits sont sémantiquement interprétables lorsqu'ils se trouvent sur une catégorie lexicale adéquate. Le trait [Pluriel] par exemple est donc interprétable lorsqu'il se trouve associé à un nom. Par contre, ce même trait [Pluriel] est sémantiquement ininterprétable lorsqu'il se trouve sur une catégorie non lexicale

¹ Cf. Chomsky (1995a:206).

² Nous revenons plus en détail sur cette théorie de la copie dans le Chapitre IV.

³ Cf. Chomsky (1995a:225).

⁴ Cf. Chomsky (1995a:235-241).

telle que celle du temps verbal T¹. Une deuxième sous-classe de traits porte l'étiquette de traits « catégoriels ». Ceux-ci indiquent à quelle catégorie lexicale appartient la partie du discours qui le porte. Dans ce modèle, il y a toujours quatre catégories lexicales, dites catégories « majeures » : le nom [N], le verbe [V], l'adjectif [A] et la préposition [P]. Ainsi le trait [N] est un trait catégoriel, sémantiquement interprétable, lorsqu'il se trouve associé à la catégorie lexicale adéquate (en l'occurrence le nom) mais ce même trait [N] est ininterprétable lorsqu'il se trouve sur une autre projection. Ce trait [N] peut ainsi notamment être également présent sur les verbes. Sa présence indique alors que le verbe en question sélectionne un complément². Ce type de trait ininterprétable est donc un trait catégoriel « sélectionnel ». Son ininterprétabilité est indiquée par la présence d'un *u* pour *uninterpretable*. Un trait [N] sur un verbe est donc signalé ainsi : [*u*N].

Ces traits sélectionnels déterminent donc les associations possibles entre les différentes catégories ; ces associations sont appelées « fusions » (*Merge*). Un verbe transitif porte ainsi un trait catégoriel interprétable [V] et un trait sélectionnel ininterprétable [*u*N] qui indique que ce verbe doit fusionner avec un nom (son complément). L'opération de fusion permet alors d'effacer le trait ininterprétable [*u*N] présent sur le verbe par le biais de l'opération de « vérification » qui s'effectue grâce à la présence du trait interprétable de même nature (en l'occurrence [N]) sur le complément. L'opération de vérification est locale, c'est-à-dire qu'elle concerne des projections sœurs. Cependant, il est également possible qu'un trait ininterprétable ne se trouve pas, initialement, à proximité d'un homologue interprétable. Le premier doit alors se déplacer (*Move*) afin d'être vérifié. La condition qui permet l'effacement de ces traits ininterprétables est appelée « accord » (*Agree*) ce qui signifie que seuls des traits homologues peuvent entrer en relation, sans quoi la dérivation capote.

Ce modèle rend compte de la variation inter-linguistique grâce à une caractéristique supplémentaire des traits : leur force (indiquée par un astérisque après le trait). La présence d'un trait fort entraîne le déplacement du trait et de la forme phonique qui lui est associée (le trait est alors vérifié localement grâce à l'opération *Agree*) alors que la présence d'un trait faible induit uniquement le déplacement du

¹ Cf. Section 4 ci-après sur les projections non lexicales, dites fonctionnelles.

² Pour des besoins de simplicité à ce stade de l'explication, nous n'utilisons pas la projection D. Celle-ci est introduite plus tard, dans notre exemple de dérivation.

trait en laissant la forme phonique sur place. Par exemple, la différence de comportement des verbes en français et en anglais est ainsi formalisée par la présence d'un trait [T*] fort sur *v* en français mais faible en anglais, qui explique la place finale différente du verbe par rapport à l'adverbe dans ces deux langues¹.

3. La distribution ordonnée des thêta-rôles

La structure d'une phrase se construit de bas en haut en deux temps. La première étape structure la phrase en fonction des thêta-rôles associés au verbe de cette phrase. Le nombre de thêta-rôles détermine alors le nombre d'arguments dans la phrase. Un verbe assigne ainsi entre zéro et trois thêta-rôles :

- Les verbes météorologiques tels que *pleuvoir* ou *neiger* n'attribuent aucun thêta-rôle dans la mesure où ce qui est traditionnellement analysé comme étant leur sujet syntaxique (*il* dans *il pleut* par exemple) n'est pas un constituant référentiel, un argument du verbe, mais simplement un élément explétif dont la présence satisfait le Principe de Projection Étendu.
- D'autres verbes attribuent un seul thêta-rôle : le sujet ; ce sont les verbes traditionnellement appelés « intransitifs ». Néanmoins, cette partition peut être détaillée davantage dans la mesure où différents types de thêta-rôles sont possibles, notamment celui d'agent (il s'agit alors d'un verbe inergatif, comme le verbe « courir ») et celui de thème/expérienceur (il s'agit alors d'un verbe inaccusatif, comme le verbe « tomber »).
- Lorsqu'un verbe attribue deux thêta-rôles, il s'agit de verbes « transitifs » (comme le verbe « manger » qui requiert un sujet et un complément d'objet).
- Les verbes qui attribuent trois thêta-rôles sont les « ditransitifs » (comme le verbe « envoyer » qui requiert un sujet et deux compléments d'objet)².

Ces thêta-rôles sont distribués de façon ordonnée suivant un principe qui postule que la hiérarchisation des projections est universelle³. Le Programme Minimaliste ne fait ainsi plus appel aux notions jusqu'alors fondamentales de « spécifieur » et de

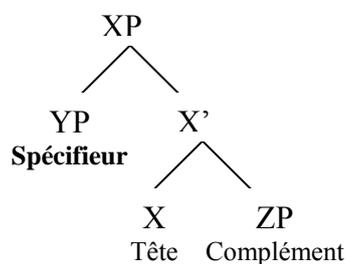
¹ Par exemple « Rémi joue_i souvent t_i au football » vs « Remi often plays football » où le trait [*u*T] du verbe a été vérifié dans les deux cas : de manière « visible » en français et de manière « furtive » en anglais.

² Les compléments circonstanciels sont des compléments « adjoints ». Ils ne font pas partie de la valence verbale et ne portent donc pas de thêta-rôles.

³ Cette hypothèse est connue sous le nom de *UTAH* (*Uniformity of Theta-Assignment Hypothesis*) et est proposée par Baker (1997).

« complément ». L'évolution de la notion de « spécifieur » nous intéresse alors au premier chef car, depuis les premiers développements de la Théorie X-Barre¹, cette position a toujours été celle du sujet syntaxique de la phrase. Cette position était alors définie comme celle d'une projection maximale (YP dans le schéma (8) ci-dessous) qui se combine avec une projection de niveau intermédiaire (X') pour former une nouvelle projection maximale (XP). Le spécifieur s'oppose alors à une autre position dans la hiérarchie syntaxique qui est celle du complément. En effet, cette dernière était également celle d'une projection maximale (ZP) mais, à la différence du spécifieur, elle se combine avec une tête (X) pour former une projection de niveau intermédiaire (X'). Cette différence structurelle entre spécifieur et complément permettait notamment de formaliser la notion classique de prédicat (X' dans le schéma ci-dessous) qui regroupe le verbe (X) et son (ou ses) objet(s) (ZP), sans y inclure son sujet (YP). Cette hiérarchisation des constituants et ce niveau intermédiaire (X') rendaient ainsi compte de la plus grande dépendance qui avait été constatée à travers les langues entre le verbe et son (ou ses) objet(s) (se(s) argument(s) « interne(s) ») et entre le verbe et son sujet (son argument « externe »)^{2,3}.

(8) La représentation du spécifieur dans le modèle X-Barre :



Dans le Programme Minimaliste, cette structure sur trois niveaux est remplacée par une structure simplifiée à deux niveaux, appelée *Bare phrase structure*⁴. La formalisation de la distinction entre la position de spécifieur et celle de complément change donc également : le complément devient le premier élément à être fusionné avec le verbe, alors que le spécifieur devient le second (dans le cas de deux arguments). Cependant, comme le mentionne Chomsky lui-même, les termes

¹ Cf. notamment Chomsky (1972) et Jackendoff (1977).

² D'après la terminologie de Williams (1980).

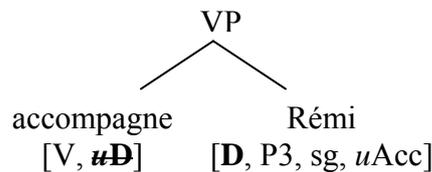
³ L'exemple des expressions idiomatiques qui comprennent toutes un verbe et son (ses) objet(s) est un argument en faveur de cette hiérarchisation. Cf. par exemple Pollock (1997:38).

⁴ Cf. par exemple Chomsky (1995a:246).

précédents continuent à être conservés par « convenance terminologique »¹ et le niveau intermédiaire continue à être projeté lorsque nécessaire. L'ordre de fusion est établi par le biais de la nature des différents thème-rôles. Ainsi, deux projections verbales sont distinguées : V et *v*. V assigne les rôles traditionnellement portés par les compléments (tels que thème ou bénéficiaire) alors que *v* est la tête qui assigne le rôle d'agent. Dans ce modèle, le sujet est donc souvent généré en spécifieur de *v*P mais il peut également apparaître en tant que complément de V (à l'intérieur du VP) lorsqu'il porte le thème-rôle de thème². Les projections présentes à ce stade et leur ordre sont donc : *v*P > VP > DP (ou PP) ; l'ordre linéaire final reflétant l'ordre des computations³. Ces premières computations sont illustrées en (9) ci-dessous.

(9) Exemple de dérivation avec la phrase « Juliette accompagne Rémi » :

(a) Fusion du verbe et de son complément et effacement de [*u*D]⁴ :



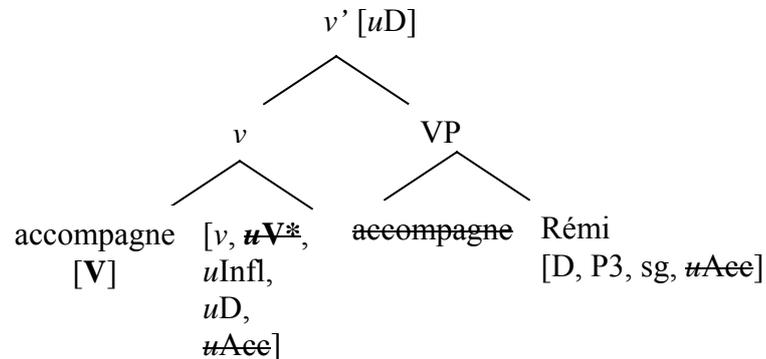
¹ Cf. par exemple Chomsky *et al.* (2002:134) : “So the whole notion of complement and specifier disappears except as a terminological convenience: you have the things that you merge first, the things that you merge second, and so on.” « Ainsi, les notions de complément et de spécifieur disparaissent sauf pour des convenances terminologiques : il y a les éléments qui sont fusionnés en premier, puis ceux qui sont fusionnés en second, etc. »

² C'est le cas des verbes inaccusatifs comme *arriver*. Cf. page 320 pour une illustration de cette possibilité.

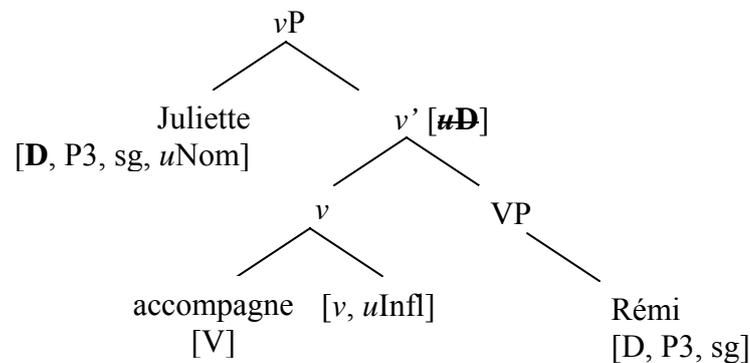
³ Cf. Baker (1985:375) sur le principe du miroir : l'ordre des morphèmes (par exemple : V-T-Pers dans *parl-er-ai*) est le reflet de l'ordre de réalisation des projections V-T-AgrS mais l'ordre inverse lorsque l'on regarde la représentation hiérarchique AgrS>T>V.

⁴ Depuis les travaux de Abney (1987), la projection fonctionnelle supplémentaire D (déterminant) est admise en tant que site principal des traits grammaticaux (les traits ϕ) des syntagmes nominaux. D héberge ainsi les déterminants et les pronoms et peut prendre comme complément un NP, soit la hiérarchie : DP > NP.

- (b) Fusion de VP et de v , adjonction de V à v afin d'effacer [μV^*] sur v ; le trait ininterprétable restant sur v est projeté sur v' pour une vérification ultérieure (v vérifie également le cas Accusatif de l'objet) :



- (c) Fusion de v' et de l'agent et vérification du trait [μD] sur v' :



À ce stade, et malgré les apparences, la dérivation n'est pas terminée dans la mesure où il reste des traits ininterprétables (notamment ceux qui concernent le Cas [μNom] et la flexion [$\mu Infl$]).

4. Les projections fonctionnelles

La seconde phase de construction de la phrase concerne les projections fonctionnelles qui sont donc celles qui n'assignent pas de thème-rôles. Les travaux de Pollock (1989) sont fondateurs dans ce domaine dans la mesure où l'auteur démontre qu'une seule projection (la projection IP dans le modèle GB) ne suffit pas à rendre compte de la diversité inter-linguistique de l'ordre des constituants (notamment des adverbes par rapport aux verbes). Ainsi, à partir des travaux de Pollock (1989), les notions de temps et d'accord sont distinguées grâce aux projections T et Agr. De

nombreux débats sur le nombre et l'ordre des projections suivent alors¹. Puis, dans le cadre du Programme Minimaliste, la pertinence même de la projection Agr est discutée par Chomsky (1995a) qui ne retient que la projection fonctionnelle T (pour la flexion) dans la mesure où les traits de T sont interprétables alors que ceux de Agr sont considérés comme ininterprétables et donc inaptes à vérifier d'autres traits eux-mêmes ininterprétables. T prend ainsi une place centrale dans ce modèle en tant que projection charnière entre le domaine thématique et le domaine fonctionnel et c'est le spécifieur de TP qui accueille le sujet de la phrase en fin de dérivation. Cinque (1999) développe la même ligne d'argumentation et rejoint la position de chercheurs tels que McCloskey (1997) ou Svenonius (2002) sur l'aspect pluriel de la notion de sujet, mentionné au chapitre précédent, et dont les propriétés sont ainsi réparties sur plusieurs têtes dans ce modèle :

“[...] if all there is to agreement is a morphological relation (with no LF relevance) between a DP in specifier position and the corresponding head, little justification remains for positing the existence of an independent (AgrP) projection [...]. [...] Subject properties are distributed over a sequence of derivationally linked positions (a theta-related one, a Case related one, and so on).”²

La dérivation se complexifie donc dans ce cadre théorique du fait d'une augmentation du nombre des projections fonctionnelles³. De nombreuses variantes existent au sein même de la grammaire générative. Néanmoins, afin de rendre compte des mécanismes principaux de vérification qui ont lieu au cours d'une dérivation, il nous faut choisir parmi les différents points de vue. Ainsi, les descriptions qui suivent et la dérivation présentée en (10) ci-dessous renvoient à ce qu'exposent notamment Adger (2003) et Hornstein *et al.* (2005).

Pour le français, au-dessus de *v*P, intervient alors la projection T (pour le Temps verbal). Ce trait étant fort, il provoque l'adjonction de *v* à T, ce qui permet de vérifier le trait ininterprétable [*u*Infl] de *v* et de lui accorder une valeur interprétable (telle

¹ T>Agr pour Pollock (1989) vs Agr>T pour Belletti (1990) par exemple.

² Cinque (1999:108) : « Si la seule chose qui reste à Agr est un lien morphologique entre un DP en position de spécifieur et une tête (sans conséquences pour LF), il reste peu d'arguments pour poser l'existence d'une projection indépendante (AgrP). [...] les propriétés du sujet sont réparties sur une série de positions reliées dans la dérivation (une liée au thème-rôle, une au Cas, etc. »

³ Dans le cadre de nos travaux, nous laissons de côté les projections liées à l'aspect et au mode verbal.

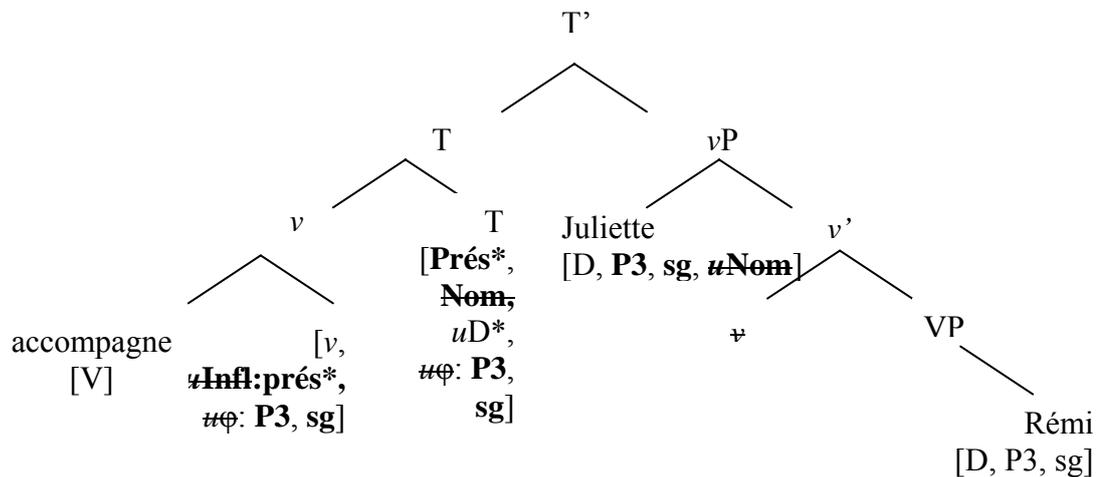
que « présent » par exemple). T porte également le trait [μ Nominatif] qui permet la vérification et l'effacement du trait casuel présent en spécifieur de ν P¹. Cette relation entre T et le spécifieur de ν P permet également d'attribuer une valeur aux traits phi, ininterprétables, présents sur T. Ces valeurs sont ensuite transmises à ν en vue de l'accord entre le sujet et le verbe (cf. (10d) ci-dessous). La tête T compte également un trait fort ininterprétable [μ D*] qui est le trait *EPP* qui force le déplacement du sujet en spécifieur de TP (cf. (e)). La hiérarchie des projections est donc : TP > ν P > VP > DP (ou PP). Les phénomènes d'accords et de mouvements déclenchés par les traits présents sur T sont donc nombreux et divers. Cette constatation nous interpelle donc dans la mesure où il nous semble qu'une telle multitude d'actions, concentrée sur une seule tête apparaît comme paradoxale dans un modèle qui prône la multiplication des têtes fonctionnelles. Deux problèmes sont notamment discutés : la redondance de certains de ces traits (Cas et *EPP* par exemple) et la nécessité de l'existence même des traits ininterprétables². Par ailleurs, les seuls niveaux intermédiaires qui demeurent nécessaires dans cette dérivation et par conséquent les seules positions de « spécifieurs » qui sont conservées sont celles qui ont un rapport avec la notion de « sujet ». De plus, les accords et effacements de traits ininterprétables ont tous lieu entre deux éléments « sœurs » (c'est-à-dire des éléments qui sont immédiatement dominés par le même constituant), sauf dans le cas de l'attribution de la valeur des traits phi à T puis à ν puisque la relation qui existe entre T et le DP en spécifieur de ν P est une relation similaire mais moins locale : la c-commande.

¹ Ces traits casuels (notamment nominatif et accusatif) sont les seuls traits à être ininterprétables à la fois sur le « vérifieur » et sur le « vérifié ». Cela vient du fait que ces cas sont « structurels » ; ils ne sont donc pas liés à un rôle thématique en particulier. Ils sont donc sémantiquement opaques à la différence des cas « inhérents » tels que le locatif qui, eux, sont sémantiquement interprétables.

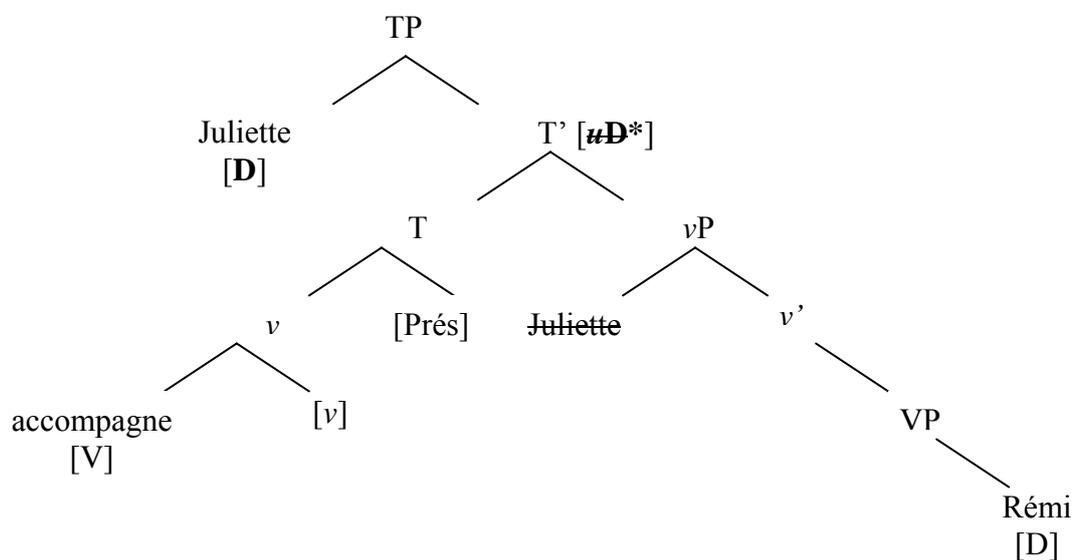
² Cf. Martin (1999:19) : “[...] insofar as we think that C_{HL} may be perfect or optimal in some serious sense, the existence of features not interpreted by interface systems is surprising”. « [...] dans la mesure où nous pensons que C_{HL} [composante computationnelle du langage humain] est parfaite ou tout au moins optimale, l'existence de traits non interprétables par les interfaces est surprenante. »

(10) Exemple de dérivation avec la phrase « Juliette accompagne Rémi » (suite) :

- (d) Fusion de vP et de T ; le trait $T[\text{présent}^*]$ est fort provoquant l'adjonction de v à T afin d'effacer $[u\text{Infl}]$ sur v et d'attribuer la valeur « présent » à v ; $[u\text{Nom}]$ sur l'agent est également vérifié grâce au trait $[\text{Nom}]$ de T et les traits ϕ de T s'accordent avec ceux du spécifieur de vP afin d'être transférés sur v :



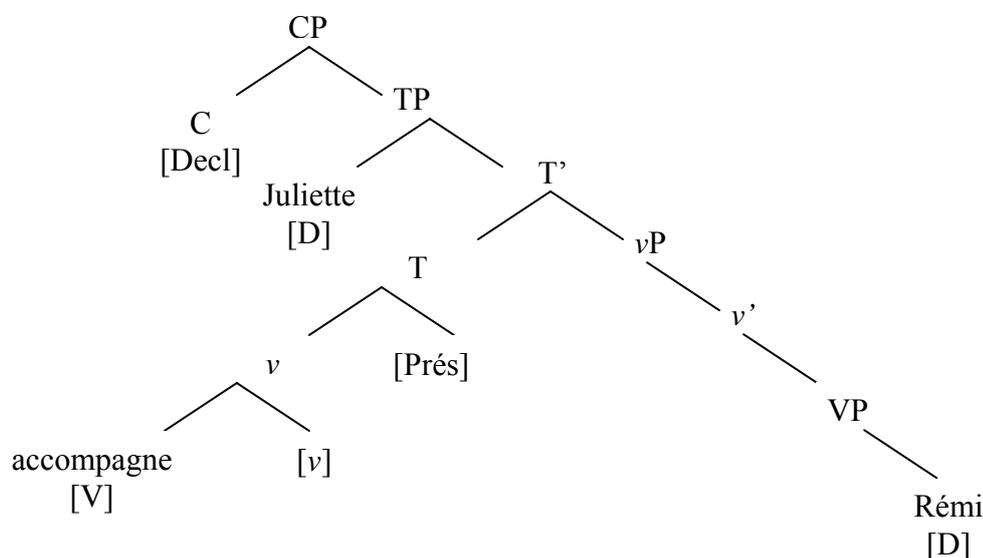
- (e) Il reste un trait fort ininterprétable sur T : $[uD^*]$ (trait *EPP*) qui se projette donc sur T' et force le déplacement du sujet en spécifieur de TP afin de l'effacer :



Au-dessus de TP se trouve la projection C du complémentateur (*cf.* (11) ci-dessous). Cette catégorie gère le type de la phrase (déclarative par exemple) et accueille notamment les conjonctions de subordination dans le cas de propositions enchâssées. L'ordre verbe-sujet présent dans certaines structures interrogatives est alors dû à la

présence d'un trait [μ Interrogatif] sur T qui déclenche le déplacement du verbe en C afin de vérifier et d'effacer ce trait.

(11) Exemple de dérivation avec la phrase « Juliette accompagne Rémi » (fin) :



Cette structure rend ainsi compte d'une phrase simple en français standard. Cependant, cette hiérarchie rencontre rapidement ses limites lorsqu'elle doit accommoder des énoncés moins simples ou moins standardisés. En effet, comment représenter par exemple une phrase telle que « Juliette elle accompagne Rémi » ? La partie la plus à gauche de la phrase a ainsi été abondamment étudiée ces dernières années afin de prendre en compte davantage de structures, notamment celles produites dans les variétés « non-standards » des langues, c'est-à-dire notre oral quotidien.

5. La fission de la périphérie gauche

Si les travaux de Pollock (1989) représentent la référence incontournable concernant la fission des projections fonctionnelles initialement regroupées sous IP, ce sont ceux de Rizzi (1997a) qui constituent la pierre angulaire du développement de CP.

5.1 Une profusion de projections

Lorsque l'on parle de « profusion de projections », on pense notamment aux travaux de Cinque (1999). Bien que largement repris par ailleurs, ils ne concernent cependant pas directement nos travaux sur le sujet dans la mesure où la hiérarchie

que développe cet auteur se réfère essentiellement à ses travaux sur les adverbes. Rizzi (1997a) quant à lui propose la fission reprise en (12) ci-dessous et qui nous concerne directement :

(12) Fission des projections fonctionnelles de CP d'après Rizzi (1997a) :

[ForceP [TopicP [FocusP [FinitenessP [...]]]]]

Cet ensemble se répartit en deux sous-groupes : le système force-finitude et le système topique-focus. Les projections Force et Finitude sont liées dans la mesure où la première porte le trait qui correspond au type de l'énoncé (déclaratif ou interrogatif par exemple) et la deuxième indique si le verbe est fléchi ou non. La projection Finitude sert ainsi de relais entre Force et l'ensemble des projections liées à la flexion verbale puisque, si le trait Finitude est positif, les projections Mode et Temps sont activées. *A contrario*, un trait Finitude négatif implique « un système plus rudimentaire de distinctions de temps »¹. Cette distinction est illustrée par l'auteur comme suit :

(13) Illustration du fonctionnement du système force-finitude :

(a) Credo *che* il tuo libro, loro lo *apprezzerebbero* molto.

[ForceP *che* ... [FinP \emptyset [TP [VP ...

« je pense que ton livre eux vont l'apprécier beaucoup »

(b) Credo, il tuo libro, *di apprezzarlo* molto.

[ForceP \emptyset ... [FinP *di* [VP ...

« je pense, ton livre, l'apprécier beaucoup »

Entre ForceP et FinitudeP, deux projections supplémentaires peuvent être insérées : TopiqueP et FocusP. Rizzi (1997a) définit le topique comme étant « un élément préposé habituellement séparé du reste de la proposition par une pause et se référant à une information ancienne, disponible dans le discours qui précède »². *A contrario*, le focus représente une information nouvelle dans le discours ; il n'est pas séparé de la proposition par une pause mais il est mis en relief par le biais d'une intonation emphatique (indiquée par les majuscules dans les transcriptions). L'auteur illustre cette distinction grâce aux exemples repris en (14) ci-dessous. L'opération qui consiste à déplacer un argument de sa position d'origine vers la périphérie renvoie à

¹ Rizzi (1997a:244).

² Rizzi (1997a:245).

ce que Chomsky (1977b) appelle « la topicalisation »¹. Depuis Cinque (1990) et l'étude de cette structure dans les langues romanes, cette configuration est distinguée des mouvements *qu-* et porte le nom de « dislocation à gauche du clitique » (*Clitic Left Dislocation* ou *CLLD*).

(14) Illustration de la distinction entre topique et focus :

- (a) Topique : Il tuo libro, lo ho letto.
« ton livre, je l'ai lu »
- (b) Focus : IL TUO LIBRO_i ho letto t_i (non il suo)
« TON LIVRE_I j'ai lu t_i (pas le sien)

5.2 Topique vs Focus

Bien que paraissant similaires, ces deux structures sont décrites comme syntaxiquement très différentes et ces divergences sont capitales pour notre analyse. Deux distinctions sautent aux yeux au regard des exemples en (14) ci-dessus : la présence d'un clitique avec le topique et la présence d'une trace avec le focus.

La présence d'un clitique « dont les traits sont identiques à ceux de l'objet [et qui] apparaît concomitamment avec l'objet » est la définition que propose par exemple Alexiadou (2006) de la structure appelée « doublement du clitique » (cf. (15a) ci-dessous) sur laquelle nous revenons plus longuement dans le cours de ce chapitre. Par ailleurs, alors que l'opération de focalisation illustrée en (14b) ci-dessus comprend une trace (et donc un déplacement de l'argument), celle de topicalisation en (14a) n'en comprend pas. Il s'agit d'un parti pris de la part de l'auteur dans la mesure où les théoriciens débattent à ce sujet, la question étant de savoir si la structure baptisée *CLLD* par Cinque (1990) est issue d'une adjonction d'un syntagme qui n'entre pas dans la valence du verbe et qui est donc considéré comme « hors phrase » ou s'il s'agit d'un déplacement de l'argument du verbe vers la périphérie gauche, comme l'illustre Alexiadou (2006) dont nous reprenons les exemples pour le grec en (15) ci-dessous.

¹ D'après des données en anglais comme par exemple *This book, we should give to John.*

(15) Débat sur l'origine du *CLLD* :

- (a) Doublement : To_i agorasa [to vivlio tu Jani]_i
 le_i ai acheté [le livre de Jean]_i
 « je l'ai acheté le livre de Jean »
- (b) *CLLD* : [To vivlio tu Jani] to agorasa (= adjonction)
 ou [To vivlio tu Jani]_i to agorasa t_i (= déplacement)

Une troisième caractéristique qui distingue les topiques des focus concerne leurs aspects quantifié et défini. En effet, il est admis (à différents degrés cependant) que « les quantifieurs (*personne, tous, etc.*) [...] ne peuvent pas être des topiques dans les constructions *CLLD* [...] »¹. Pour le français, Zribi-Hertz (1994) illustre cette incompatibilité par les exemples repris en (16) ci-dessous :

(16) Incompatibilité entre topiques et quantifieurs² :

- (a) Expression avec quantifieur : **Personne*, il n'a rien dit.
(b) Expression sans quantifieur : *Pierre*, il n'a rien dit.

La même incompatibilité avec cette position de topique est souvent mise en avant pour les éléments indéfinis, comme l'illustre par exemple De Cat (2007) dont les exemples sont repris en (17) ci-dessous :

(17) Incompatibilité entre topiques et indéfinis³ :

- (a) **Quelqu'un*, il est venu.
(b) **Un enfant*, il est arrivé et il t'a posé une question.

Les agrammaticalités illustrées ci-dessus sont expliquées par le fait que l'on n'arrive pas à attribuer une référence précise à de tels constituants, ce qui est incompatible avec la position de topique qui ne contient que des informations déjà connues de l'interlocuteur⁴. En d'autres termes, puisque l'information est connue, le constituant ne peut pas être indéfini mais doit être défini. *A contrario*, Rizzi (1997a) définit les focus comme des éléments « quantitatifs » et subsume ainsi les distinctions précédemment proposées par Cinque (1990) sous cette caractéristique :

¹ Rizzi (1997a:250).

² Zribi-Hertz (1994:134).

³ De Cat (2007:82).

⁴ Cf. la définition de Rizzi (1997a:245) donnée précédemment.

(18) La quantification distingue focus et topiques¹ :

- (a) Focus : NESSUNO ho visto t.
 « personne j'ai vu »
 TUTTO ho fatto t.
 « tout j'ai fait »
- (b) Topiques : *Nessuno, lo ho visto.
 *Tutto, lo ho fatto.

Or, comme le souligne Rizzi (1997a), et comme le montrent nos gloses agrammaticales ci-dessus, « certaines [...] langues (par exemple le français) ne semblent pas utiliser une position structurale de focus »². Nous comprenons alors, d'après De Cat (2007) par exemple que la position d'un focus en français est la « position canonique du sujet », c'est-à-dire le spécifieur de TP :

“If an XP coindexed with a(n adjacent) subject clitic allows for a focus interpretation of that XP, it implies that the XP in question is not dislocated and hence it occupies the canonical subject position [...].”³

5.3 Les clitiques nominatifs ne sont pas toujours les bienvenus

Cette distinction entre topiques et focus est centrale dans notre problématique puisque, d'après ce qui vient d'être dit, la structure syntaxique nécessaire pour véhiculer ces deux types d'informations (ancienne et donc définie *vs* nouvelle) ne peut pas être identique, ni en ce qui concerne la place du DP, ni en ce qui concerne la place du clitique nominatif, lorsque ce dernier est réalisé concomitamment au DP. De Cat (2007) distingue ainsi les deux structures, reprises en (19) ci-dessous, qui rendent compte de la place que chaque élément pourrait occuper en fonction de l'information véhiculée :

¹ D'après Rizzi (1997a:250).

² Rizzi (1997a:246).

³ De Cat (2007:22) : « Si un XP co-indicé avec un clitique sujet (adjacent) autorise une interprétation en tant que focus, cela implique que le XP en question n'est pas disloqué et qu'il occupe alors la position canonique du sujet [...]. »

(19) Les structures syntaxiques des topiques et des focus en français oral¹ :

- (a) XP en focus : [CP ... [TP XP_i [T' clitique_i+T ...]]]
(b) XP en topique : [CP XP_i [TP clitique_i [T' ...]]]

Or, d'après l'auteur, « dès qu'un XP est associé avec un clitique sujet en français oral, cet XP est disloqué et interprété en tant que topique »². La structure représentée en (19a) ci-dessus est donc impossible pour De Cat (2007) dans la mesure où « le français oral n'autorise pas le doublement du sujet »³. Pourquoi une telle affirmation ? Nous croisons de nouveau le chemin des éléments « perturbateurs » que sont les clitics nominatifs. Le positionnement de De Cat (2007) est celui des chercheurs qui postulent que les clitics nominatifs du français ne peuvent pas avoir un statut affixal et donc morphologique (ils ne peuvent donc pas être représentés en T comme dans l'exemple (a) ci-dessus). En revanche, leur statut uniquement argumental oblige à les cantonner à des positions argumentales et donc syntaxiques (comme celle du spécifieur de TP en (19b)). Conséquemment, lorsqu'un DP apparaît dans le même énoncé qu'un clitique nominatif (co-indiqué au DP), le DP ne peut occuper cette même position et est donc relégué à une position « hors phrase », non-argumentale : celle du topique dans la périphérie gauche de la phrase (avec la question ouverte de savoir si le topique est adjoint ou déplacé, comme mentionné précédemment). Cette proposition ne fait cependant pas l'unanimité d'où l'existence d'un débat récurrent en grammaire générative sur le statut de ces clitics nominatifs en français oral, depuis la publication des premiers travaux sur les clitics français par Kayne (1975). Nous clôturerons ce chapitre théorique en exposant le détail et les arguments de chacun concernant ce débat. Pour l'heure, revenons aux outils du Programme Minimaliste et voyons si et comment celui-ci intègre une catégorie vide telle que *pro* et le principe qui a motivé sa conception : *EPP*.

6. La révision de *EPP* et du paramètre *pro-drop*

Nous avons vu que des deux types de catégories vides créées dans le cadre *GB*, celle des traces est abandonnée dans le Programme Minimaliste. La question se pose alors quant à la légitimité de *pro* dans ce nouveau cadre et la littérature illustre que son statut soulève en effet de nombreux problèmes encore non résolus.

¹ De Cat (2007:22).

² De Cat (2007:33).

³ De Cat (2007:19).

6.1 La formulation de Pollock (1997)

Pollock (1997) reprend les notions développées par Rizzi (1982) sur la richesse morphologique de la flexion verbale en italien qui distingue les traits [Personne], [Locuteur] et [Pluriel]. Ainsi, « la simple adjonction du verbe à Infl° suffit à satisfaire au principe de projection étendu et aux exigences casuelles du temps fini »¹. Dans ce cadre, l'élément *pro* reçoit uniquement le thème-rôle attribué par le verbe. L'auteur reformule ainsi le « paramètre du sujet nul » :

(20) Formulation de Pollock (1997) du « paramètre du sujet nul »² :

- (a) « L a un accord verbal fini [\pm pronominal].
- (b) Un accord verbal fini est [$+$ pronominal] ssi il distingue morphologiquement [\pm Personne], [\pm Locuteur] et [\pm Pluriel].
- (c) Un accord verbal fini pronominal est [$+$ N].
- (d) Un accord verbal [$+$ N] est [$+$ cas] ».

Dans ce cadre, lorsque la morphologie verbale perd son aspect « pronominal », elle perd également son trait [$+$ N] et le déplacement du verbe en I ne peut plus satisfaire *EPP*.

6.2 Vers l'élimination de *pro* ?

Comme le mentionnent les auteurs, l'hypothèse d'Alexiadou & Anagnostopoulou (1998, 1999) se rapproche de celle proposée par Pollock (1997) ci-dessus. Travaillant principalement à partir du grec moderne mais étendant leur hypothèse à d'autres langues, Alexiadou & Anagnostopoulou (1998, 1999) proposent que le trait D de AgrS³ soit universellement fort et que la variation entre langues à « sujet nul » et langues à « sujet non-nul » provienne du mode de vérification de ce trait fort. Les auteurs postulent ainsi deux modes de vérification différents. Les langues à « sujet non-nul » vérifient ce trait comme proposé dans Chomsky (1995a), c'est-à-dire par déplacement ou fusion d'une projection maximale (XP) en spécifieur de AgrS, alors que les langues à « sujet nul » adoptent un second mode de vérification qui opère, non par déplacement ou fusion d'une projection maximale mais par déplacement ou fusion d'une tête, en l'occurrence de la tête verbale (V°). En effet, les auteurs proposent que, dans ces langues, la morphologie verbale inclut

¹ Pollock (1997:170).

² Pollock (1997:173).

³ Ce que nous avons décrit sous le nom de T.

un élément nominal qui porte le trait [+D], des traits phi interprétables et potentiellement un trait [+Cas]. Ils postulent ainsi que la flexion verbale d'une langue comme le grec moderne « joue exactement le même rôle que les pronoms en anglais » et que « ainsi, la montée de V en AgrS suffit à effacer le trait *EPP* »¹. Dans ce cadre, Alexiadou & Anagnostopoulou (1998, 1999) n'ont donc plus besoin d'avoir recours à *pro*. Cette hypothèse forte entraîne ainsi des conséquences importantes concernant la notion et la place du sujet. Dans les langues où V suffit à vérifier *EPP*, le spécifieur de TP n'est pas projeté. La conséquence directe est que, lorsqu'il y a un élément nominal préverbal, celui-ci ne peut pas occuper cette position argumentale mais doit occuper une position non argumentale, plus élevée dans la structure². Cette approche soulève alors la question du thème-rôle externe attribué par le verbe. Les auteurs résolvent ce problème en mentionnant que leur hypothèse pour les langues à « sujet nul » est à rapprocher des cas d'explétifs dans les langues à « sujet non nul » qui ne reçoivent donc pas de thème-rôle. Alexiadou & Anagnostopoulou (1998, 1999) proposent ainsi de se passer du *pro* explétif. Comme le souligne notamment Roberts (2007), cette approche est théoriquement intéressante si l'on considère, d'une part, l'abandon qui est fait par ailleurs d'autres types de catégories vides telles que les traces et, d'autre part, le cadre « économique » du Programme Minimaliste. D'autres recherches vont dans ce sens, bien qu'adoptant des stratégies diverses. Nous avons mentionné les travaux de Manzini & Savoia (2002) sur la lexicalisation de D par le verbe dans les langues à « sujet nul », proposition qui permet également de ne pas avoir recours à *pro*³.

Ce qui ressort de ce tour d'horizon rapide et ce qui, nous pensons, est important dans le cadre de nos recherches, est le fait que la théorie soit passée d'une macro-paramétrisation binaire et relativement rigide à une approche beaucoup plus nuancée (et donc micro-paramétrique) reflétant ainsi davantage la réalité linguistique. Cela vient en partie de l'évolution des outils théoriques utilisés : tout d'abord formulée à partir d'un faisceau de propriétés, la variation est à présent décrite en termes d'unités plus petites que sont les traits. Notons que, dans ce cadre, morphologie et syntaxe sont ainsi étroitement liées et que, comme mentionné dès le Chapitre I, lorsque l'on s'intéresse à la question du « sujet nul », celle de la richesse de la morphologie

¹ Cf. Alexiadou & Anagnostopoulou (1999:104-105).

² Cf. notamment les projections de topique et de focus proposées par Rizzi (1997a).

³ Cf. également Manzini & Savoia (1997).

verbale n'est jamais vraiment très loin. Dans Palasis (2005), nous établissons également un lien entre l'absence du clitique préverbal et la présence d'une suffixation verbale riche chez les jeunes francophones. Grâce aux outils performants du Programme Minimaliste et à de nouvelles données plus nombreuses, nous proposerons dans le Chapitre V une approche théorique micro-paramétrique du développement du sujet syntaxique chez le jeune enfant, approche qui permettra également de rendre compte de la « diachronie » de la production linguistique au cours de la vie d'un locuteur. Mais, avant de présenter notre nouveau corpus et son analyse, il nous reste un point théorique important à détailler. Il s'agit du phénomène du « doublement du clitique » que nous avons commencé à évoquer lors de la distinction entre les notions de topique et de focus et que nous avons abordé grâce à la définition de Alexiadou (2006) : « structure [...] dans laquelle un clitique dont les traits sont identiques à ceux de l'objet apparaît concomitamment avec l'objet ». Notons que cette définition est limitée à l'objet mais que nous avons néanmoins déjà également mentionné la question du doublement du sujet par le biais des recherches de De Cat (2007). Les deux chercheurs s'accordent néanmoins sur un point : ni Alexiadou (2006) pour l'objet, ni De Cat (2007) pour le sujet ne considèrent que le français dispose d'une telle structure. Nous revenons alors à notre phrase proposée en fin de Section III.4 ci-dessus, « Juliette elle accompagne Rémi », et nous nous demandons quelles approches théoriques sont développées afin de rendre compte de ce genre de structures en français.

IV. PLUSIEURS CANDIDATS, UNE SEULE FONCTION

Tout va bien dans le meilleur des mondes théoriques, tant que, pour une fonction syntaxique, un thème-rôle ou encore un cas, une phrase verbale ne comporte qu'un seul élément susceptible de porter cette fonction, ce rôle ou ce cas. Les exemples de Grevisse & Goosse (2008) pour le français standard, mentionnés au Chapitre I et que nous reprenons en (21a) ci-dessous, illustrent cette adéquation entre théorie linguistique et énoncés produits. Cependant, l'analyse prend une toute autre tournure lorsque l'on observe des énoncés issus du français oral adulte (*cf.* (21b)) ou du français infantin (*cf.* (21c)).

(21) Exemples de sujets :

(a) En français standard¹ :

Le professeur écrit au tableau.

Je travaille.

(b) En français oral adulte² :

KAT: *toi aussi tu fais un escargot Alan ?* (I.1a, 67)

KAT: *j(e) crois que Tom i(l) veut faire celui+là aussi .* (I.3c, 192)

(c) En français enfantin³ :

JUL: *nan moi j(e) vais pas à la cantine !.* (VII.30b, 1166)

ANT: *mon petit chien il [ʔ] <il est> [ʔ] il est mon amou(r) mon petit chien .*
(I.4, 73)

En effet, des divergences entre les sujets de (a) ci-dessus et ceux de (b) et de (c) sont perceptibles à plusieurs niveaux. En termes syntaxiques, la présence de deux éléments (ou plus) soulève de nombreuses interrogations quant à l'attribution du thème-rôle et du cas ; en termes phonologiques, les formes [ʒə] ou [il] du français standard présentent à l'oral des variations non négligeables qui, dans le courant de notre analyse, apparaîtront d'ailleurs comme révélatrices de mécanismes morphosyntaxiques sous-jacents.

Dans le cours de la réflexion générative, l'intérêt des linguistes s'est porté dans un premier temps sur la question du doublement de l'objet avant de s'étendre, dans un deuxième temps, sur celui du sujet. Nous reprenons donc la chronologie de cette réflexion afin d'en comprendre les tenants et les aboutissants et afin d'exposer, en fin de chapitre, la nature du débat qui anime les chercheurs concernant le statut du clitique nominatif en français oral.

1. Le doublement du clitique objet

1.1 Description du phénomène

À la fin des années 1970, des occurrences problématiques apparaissent dans la littérature générative par le biais notamment des thèses largement citées de Strozer

¹ Grevisse & Goosse (2008:248).

² Données extraites de notre corpus.

³ Données extraites de notre corpus.

(1976) et de Rivas (1977) sur les objets dans certains dialectes de l'espagnol. Fontana (1993) par exemple résume la situation ainsi :

« [...] en espagnol moderne, les clitiques objets apparaissent souvent avec un NP coréférentiel sur un modèle qui est devenu connu sous le nom de « doublement de clitique ». En effet, dans aucun dialecte de l'espagnol moderne, les pronoms personnels objets [toniques] ne peuvent apparaître seuls dans la phrase ; ils doivent être « doublés » par un clitique co-indicé »¹.

Certains de ses exemples repris en (22) ci-dessous illustrent les différentes cooccurrences possibles en espagnol moderne (cf. (a) et (b)), les apparitions d'éléments non co-indicés (cf. (c) et (d)), ainsi que l'agrammaticalité que présente l'apparition du pronom tonique seul (cf. (e)) :

(22) Exemples d'objets indirects en espagnol :

- (a) *Le_i mandé una carta a ella_i.*
lui ai envoyé une carte à elle
« je lui ai envoyé une carte à elle »
- (b) *Le_i mandé una carta a Marta_i.*
lui ai envoyé une carte à Marta
« je lui ai envoyé une carte à Marta »
- (c) *Mandé una carta a Marta.*
ai envoyé une carte à Marta
« j'ai envoyé une carte à Marta »
- (d) *Le mandé una carta.*
lui ai envoyé une carte
« je lui ai envoyé une carte »
- (e) **Mandé una carta a ella.*
ai envoyé une carte à elle
« j'ai envoyé une carte à elle »

Par ailleurs, ce phénomène de doublement de l'objet ne se limite pas aux dialectes de l'espagnol puisque, approximativement à la même époque, il est également étudié

¹ Fontana (1993:43-44).

non seulement dans d'autres langues romanes, telles que le roumain¹, mais également dans des langues très différentes, telles que l'hébreu² ou le grec³. Il est donc intéressant de noter que, dès cette époque, le français n'est généralement pas considéré comme une langue qui autorise cette structure du doublement de l'objet. Les exemples astérisqués de Jaeggli (1982, 1986), repris en (23) ci-dessous, illustrent ce point de vue. Néanmoins, certains chercheurs distinguent les compléments d'objets directs des compléments indirects, comme le montre l'exemple non astérisqué de Borer (1984) en (23c) ci-dessous. La section suivante présente les justifications théoriques de chacun pour ces différents cas de figure.

(23) Le doublement des objets en français :

- (a) *Marie *le* voit *Jean*⁴
- (b) *Jean *lui* a donné des bonbons à *Marie*⁵
- (c) Paul la *lui* présentera à *Juan*⁶

1.2 Les problèmes théoriques

Les doublements constatés dans ces différentes langues soulèvent au moins trois questions théoriques importantes concernant : (i) la position initiale du clitique dans la structure et son éventuel mouvement, (ii) le statut syntaxique du clitique et (iii) le statut syntaxique du syntagme co-indicé. La question (i) se limite aux clitics objets ; nous la mentionnons donc sans nous y attarder. Par contre, les questions (ii) et (iii), initialement soulevées pour le doublement des objets sont, par la suite, étendues aux sujets et nous intéressent donc plus particulièrement.

Les travaux de Kayne (1975) sont largement reconnus comme étant les travaux fondateurs dans l'analyse générative des pronoms du français. Concernant les pronoms objets, Kayne (1975) pose que ceux-ci sont générés exactement de la même façon que leurs contreparties nominales – c'est-à-dire à droite du verbe, dans une position argumentale, donc apte à recevoir un thème-rôle et un Cas – et qu'ils subissent ensuite un déplacement syntaxique vers la gauche du verbe et une cliticisation phonologique. Cette thèse en faveur du mouvement des clitics objets

¹ Cf. Borer (1984).

² Cf. Borer (1984).

³ Cf. Philippaki-Warbuton (1987).

⁴ Jaeggli (1986:18).

⁵ Jaeggli (1982:12-13).

⁶ Borer (1984:40).

est également défendue à la même époque pour le grec par exemple¹. Dans ce cadre, ces éléments sont alors considérés comme de véritables pronoms – au sens étymologique du terme – c'est-à-dire qu'ils remplacent un nom et qu'ils représentent ainsi un argument à part entière du verbe. Cette analyse des pronoms clitiques rencontre néanmoins des difficultés lorsqu'elle est confrontée aux doublements mentionnés précédemment. Le problème théorique vient alors du fait que, dans ce type de phrases, on trouve *deux* éléments susceptibles de recevoir un seul et unique thème-rôle et un seul et unique Cas : un syntagme nominal ou un pronom tonique *et* un pronom clitique. Un débat théorique s'engage alors (et dure depuis) sur le statut de chacun de ces candidats. En tout état de cause, *a priori*, un seul des deux constituants peut conserver son statut argumental ; l'autre devant alors endosser un statut différent, à définir. De très nombreux travaux voient ainsi le jour et plusieurs hypothèses sont avancées dans le cadre de l'étude de l'objet.

1.2.1 Le clitique porte le thème-rôle = doublement

Conférer un statut argumental au clitique est la conséquence logique de la thèse du déplacement du clitique telle que préconisée par Kayne (1975). En effet, lorsque l'on postule le mouvement d'un pronom, cela implique que celui-ci a été généré dans la même position qu'un syntagme nominal. Dans ce cas, le pronom est donc un argument du verbe : il en reçoit ses attributs thématique et casuel puis se déplace. Cette hypothèse est défendue par exemple par Philippaki-Warburton (1987) pour le grec :

(24) Le doublement de l'objet direct en grec² :

ton_i ksero e_i to Jani
 le-acc connais-1sg le-acc Jean-acc
 « je le connais (le) Jean »

Ainsi, lorsqu'il y a doublement, comme en (24) ci-dessus, l'auteur postule que le clitique reçoit le thème-rôle « thème » et le Cas « accusatif » puis se déplace. Le syntagme nominal *to Jani* est alors analysé comme occupant une position non argumentale dans laquelle il ne reçoit donc pas de thème-rôle. Le détail ci-dessus montre néanmoins qu'il se voit attribuer un Cas, le même que le clitique, car ces deux éléments restent tout de même co-référents, ce qui permet d'employer le terme

¹ Cf. Philippaki-Warburton (1977).

² Philippaki-Warburton (1987:304-305).

« doublement ». Nous avons donc ici un thème-rôle et un dédoublement du Cas, ce qui est incompatible avec la théorie de la vérification dans le cadre plus récent du Programme Minimaliste, dans la mesure où tout trait ininterprétable (en l'occurrence le trait casuel [*u*Accusatif] sur *v*) est effacé dès la première vérification ; il ne peut donc pas vérifier un second trait [*u*Accusatif] dans le courant de la même dérivation.

1.2.2 Le clitique porte le thème-rôle = dislocation

La structure syntaxique décrite ci-dessus s'apparente alors très fortement à celle de la « dislocation » qui est également définie comme étant la cooccurrence d'un clitique en position argumentale et d'un syntagme en position non-argumentale. Les définitions de ces deux structures sont pour le moment identiques. La question est donc de savoir s'il est possible de cerner un ou plusieurs critères objectifs qui permettraient de distinguer un doublement d'une dislocation. Les exemples très diffusés de Jaeggli (1986), repris en (25) ci-dessous, visent à montrer que oui.

(25) Exemples français de Jaeggli (1986)¹ :

- (a) Je l'ai vu, *l'assassin*.
- (b) Je *lui* ai raconté mes secrets, *le traître*.

L'intonation est le premier critère mis en avant par Jaeggli (1986) pour poser l'existence de deux structures syntaxiques différentes. En effet, l'auteur postule qu'il doit y avoir une « pause intonative importante » entre l'élément disloqué et le reste de la phrase. Celle-ci est généralement transcrite par une virgule. Ainsi, sans pause, l'auteur postule que ses exemples seraient agrammaticaux pour le français car, d'après lui non plus, cette langue n'autorise pas le doublement. Dans le cadre plus récent du Programme Minimaliste, De Cat (2007) propose une analyse structurelle similaire à celle de Jaeggli (1986) dans la mesure où elle postule également (i) que les éléments disloqués n'ont pas le statut d'arguments et (ii) qu'ils sont le résultat d'une opération d'adjonction. Le point de vue des deux auteurs diverge néanmoins quant à la question de la prosodie. En effet, De Cat (2007), à la suite de Rossi (1999) notamment, déclare qu'« il n'y a aucune pause obligatoire entre le noyau de la phrase et un élément disloqué à droite »². Sa conclusion est identique pour les éléments disloqués à gauche et, suivant les travaux de Deshaies *et al.* (1993) sur le français québécois, De Cat (2007) ne retient pas non plus le critère de la pause à gauche,

¹ Jaeggli (1986:33-34).

² De Cat (2007:35).

postulant qu'elle relève davantage d'une « impression due à l'allongement de la dernière syllabe du constituant disloqué » et qu'« elle n'est pas en fait confirmée par des analyses acoustiques précises »¹. En ce qui concerne notre base de données, devant cette absence de critère prosodique catégorique quant à la réalisation ou non d'une pause, nous décidons de ne pas prendre partie au moment de la transcription de nos données et nous n'insérons donc pas de virgule lorsqu'un syntagme nominal et un clitique coréférentiel se présentent dans une même phrase, comme le montrent nos exemples en (26) ci-dessous, extraits parmi les 186 occurrences de clitics objets codés comme « doublés »² :

(26) Transcription des constituants objets coréférentiels dans notre base de données :

- (a) ANT: je *le* voulais *le papillon* (XIII.48d, 139)
- (b) KAT: oh non on peut pas *l'* ouvrir *ça* (I.2, 543)
- (c) NOE: +, en plus *les chaussons* on peut *les* mettre (IV.16, 729)
- (d) KAT: *celles+là* je *les* garde ? (XI.42, 1958)
- (e) JUL: *les grosses* on *les* met dans la boîte *les grosses* (XIII.48d, 51)
- (f) KAT: et *c(e)lui+là* tu *le* connais *c(e)lui+là* ? (III.10a, 1201)

Bien que cette structure ne soit pas le cas de figure dominant³, ces exemples illustrent tout de même de façon indiscutable le fait qu'un clitique accusatif peut avoir la même référence qu'un syntagme nominal ou un pronom tonique dans une phrase en français oral, que ce soit chez des jeunes enfants (*cf.* (a), (c) et (e) ci-dessus) ou chez un adulte (*cf.* (b), (d), et (f)), que le syntagme se trouve à droite du verbe, à sa gauche ou dans les deux positions concomitamment. Il en va de même pour les clitics nominatifs doublés, comme l'illustrent nos exemples en (27) ci-dessous :

¹ De Cat (2007:47).

² Tous locuteurs confondus (adultes et enfants). Requête CLAN : freq +t%mor +s"pro:cli:d:acc|*" +u @ > Pro_Acc_Doubl_freq et kwal +t%mor +s"pro:cli:d:acc|*" @ > Pro_Acc_Doubl_kwal.

³ Les clitics accusatifs employés seuls sont au nombre beaucoup plus élevé de 1089 (requête CLAN : freq +t%mor +s"pro:cli:acc|*" +u @ > Pro_Acc_freq).

(27) Transcription des constituants sujets coréférentiels dans notre base de données¹ :

- (a) EMA: *elle est là ma photo* (I.1a, 171)
- (b) KAT: *comment i(l) s' appelle l' ours ?* (I.2, 104)
- (c) TOM: *là les lapins i(ls) sont contents* (X.39, 1259)
- (d) KAT: *et toi tu t' appelles Alan !* (I.1a, 37)
- (e) ANT: *mon sac il est super mon sac* (X.39, 1063)
- (f) KAT: *alors et toi qu'est+c(e)+que tu fais Julien ?* (I.1a": line 274)

1.2.3 *Le syntagme prépositionnel porte le thème-rôle*

Le deuxième critère avancé par Jaeggli (1986) pour distinguer un doublement d'une dislocation concerne la présence d'une préposition : ses exemples de dislocations en (25) ci-dessus n'en comportent effectivement pas² alors que l'exemple de doublement proposé en (28) ci-dessous contient un complément d'objet introduit par une préposition :

(28) Doublement du complément d'objet direct³ :

Lo vimos a él.

Le-acc vîmes à lui

« Nous le vîmes lui »

La présence et le rôle de cette préposition émanent d'une hypothèse souvent citée dans la littérature sous le nom de « généralisation de Kayne »⁴ qui postule que le doublement n'est possible que lorsque l'élément qui n'est pas le clitique est un syntagme prépositionnel. Dans le cas contraire, la coréférence est qualifiée de dislocation. Les exemples en (23) et (28) ci-dessus indiquent qu'une partition est posée dans le traitement des compléments d'objets dans les différentes langues ; le fait central dans la formalisation théorique du doublement étant alors la présence de cette préposition. En effet, celle-ci est présente dans les compléments d'objets directs de l'espagnol, du roumain, ainsi que dans les compléments d'objets indirects du

¹ Tous locuteurs confondus, les clitiques nominatifs doublés sont au nombre de 5326 (requête CLAN : freq +t%mor +s"pro:cli:d:nom|*" +u @ > Pro_Nom_Doubl_freq). Les non doublés sont au nombre de 12020 (requête : freq +t%mor +s"pro:cli:nom|*" +u @ > Pro_Nom_freq).

² En tant que locuteur francophone, nous émettons tout de même quelques réserves sur la grammaticalité de sa phrase (b) dans la mesure où il est généralement considéré que les dislocations de compléments d'objets indirects sont introduites par une préposition. Cf. par exemple De Cat (2007:140).

³ Exemple valable pour tous les dialectes de l'espagnol, extrait d'Anagnostopoulou (2006:535).

⁴ À la suite de Jaeggli (1982:20). Cf. Kayne (1975).

français¹. Dans ce cadre, il est postulé que : (i) le clitique est directement généré à gauche du verbe, (ii) les deux candidats reçoivent chacun un Cas mais dans des conditions différentes et (iii) seul le syntagme prépositionnel reçoit un thème-rôle. Le clitique est alors injecté dans la structure directement aux côtés du verbe. Ainsi, il n'occupe pas la position traditionnelle de l'objet, ni même une position argumentale quelle qu'elle soit, et ne reçoit donc pas de thème-rôle. Ce dernier est donc disponible pour le syntagme prépositionnel. Par contre, dans cette position, le clitique « absorbe » le Cas accusatif attribué par le verbe ce qui le rend indisponible pour le nom ou pronom tonique co-indicé. Le syntagme reçoit néanmoins le même type de Cas que le clitique mais par l'intermédiaire d'un autre élément – la préposition (*a* en espagnol, *pe* en roumain, par exemple) – d'où la nécessaire présence de cette dernière dans ce type de structure. La première différence par rapport aux deux autres hypothèses ci-dessus est donc la distribution du thème-rôle (clitique *vs* syntagme).

Dans cette hypothèse, il existe une seconde distinction capitale qui est à la source d'un vaste débat au sein des générativistes : le statut non argumental du clitique. Le point de vue sur le statut du clitique est donc différent avec cette hypothèse puisque Borer (1984) par exemple le considère clairement comme un affixe flexionnel du verbe :

“We assume clitics to be affixes, the output of an inflectional rule which inserts number, gender and person features [its inherent features] and associates them with an already existing Case feature on a lexical head [the non inherent feature from V]. Clitics are a spell-out of Case features in the sense that once the Case feature is associated with the inserted number, gender and person features, it is given an independent phonological representation and can no longer be transferred to a complement of the head.”²

¹ Cette analyse pour le français est néanmoins contestée. Jaeggli (1982) par exemple propose que le *à* français est un « marqueur de Cas », pas une véritable préposition. Cf. également Oliviéri (1994:121-126) pour une réflexion sur l'existence de plusieurs types de prépositions.

² Borer (1984:41) : « Nous supposons que les clitics sont des affixes résultant d'une règle flexionnelle qui insère les traits de nombre, genre et personne [les traits inhérents du clitique] et qui les associe avec un trait Cas déjà existant sur une tête lexicale. Les clitics sont une épellation des traits Cas dans la mesure où une fois que le trait Cas est associé aux traits nombre, genre et personne, il reçoit une représentation phonologique indépendante et ne peut plus être transféré vers un complément de la tête. »

La réponse théorique consiste donc à : (i) répartir la distribution du thème-rôle et du Cas traditionnellement attribués par un même verbe à un même constituant, à plusieurs émetteurs et plusieurs récepteurs, et (ii) attribuer au clitique non plus un statut purement syntaxique mais plutôt morphosyntaxique. Un autre point de vue concurrent consiste à analyser le syntagme prépositionnel comme un adjectif. Celui-ci n'entre ainsi pas dans la valence verbale et seul le clitique conserve le statut d'argument avec le thème-rôle et le Cas.

Ces hypothèses font ainsi la prédiction unanime que le doublement d'un argument par un clitique ne peut survenir que dans des configurations qui comportent un argument prépositionnel. Le doublement est donc limité à certaines langues et/ou à certains Cas au sein d'une même langue, comme le montrent nos exemples en (23) ci-dessus pour le français. Ainsi, Borer (1986) considère le doublement de l'objet direct comme « impossible » en français¹. Nos énoncés en (26) ci-dessus, transcrits tels quels, seraient donc astérisqués par de nombreux linguistes dans la mesure où il s'agit d'objets directs et qu'ils ne présentent pas la préposition nécessaire à l'attribution du double Cas. Dans ce cadre, leur structure ne peut donc pas être analysée en tant que « doublement de l'objet ». Plusieurs possibilités s'offrent alors aux chercheurs : (i) considérer que ces exemples du français oral relèvent d'un autre type de structure et définir lequel, (ii) considérer que ces exemples du français oral sont bien des doublements et postuler que la « généralisation de Kayne » ne couvre pas tous les cas de figure.

Ces dissensions entre théorie générative et données issues initialement de dialectes de l'espagnol sont ainsi à la source de nombreuses interrogations, travaux et débats qui perdurent depuis la fin des années 1970. C'est également l'étude de dialectes qui éveille l'intérêt des chercheurs pour ces questions de doublement et de dislocation dans le cadre de l'étude du sujet.

2. Le débat sur le statut du clitique nominatif

Nous avons vu comment la grammaire générative des années 1980 réussit à prendre en charge une distinction majeure entre les langues du monde concernant la réalisation de leurs sujets : *EPP* rend leur présence obligatoire et le paramètre *pro-*

¹ Borer (1986:7) : “It is, presumably, the absence of such a preposition in French which renders clitic doubling impossible in that language.”

drop régit une dichotomie stricte entre deux types possibles de langues. Les diverses études menées sur les objets dans les dialectes de l'espagnol montrent cependant, à partir de la fin des années 1970, que la réalité est souvent beaucoup moins tranchée que la théorie de cette époque. Pour les sujets, les contrevenants à cette élégante dichotomie du *pro-drop* sont des dialectes du nord de l'Italie, notamment le trentin et le florentin pris comme exemples par Rizzi (1986a) dans son article majeur sur le statut des clitiques sujets dans les langues romanes.

2.1 Des langues standards aux dialectes du nord de l'Italie

Commençons par reprendre les exemples de Rizzi (1986a)¹ en (29) ci-dessous : ils comprennent une langue standard non *pro-drop*, une langue standard *pro-drop* et le trentin.

(29) L'insertion des dialectes du nord de l'Italie dans le paysage *pro-drop* :

(a) Français : langue standard non *pro-drop* :

Jean mange **Mange*

(b) Italien : langue standard *pro-drop* :

Gianni mangia *Mangia*

(c) Trentin : ?

**El Gianni magna* **Magna* (*El Gianni*) *el magna*

Ainsi, le trentin, entre autres dialectes du nord de l'Italie, illustre le fait que la représentation classique des langues *pro-drop* (cf. (30) ci-dessous) dans laquelle « *pro* est licencié et interprété en tant que pronom référentiel défini par le biais de INFL qui contient une spécification AGR riche »² ne semble pas respectée puisque le verbe fléchi ne peut survenir sans le clitique dans ce dialecte, comme le montrent les exemples en (29c) ci-dessus.

(30) La représentation classique de la dichotomie *pro-drop* :

(a) *pro-drop* : [_{IP} *pro* [_I AGR^{riche} [_{VP} *mangia*]]]

(b) non *pro-drop* : [_{IP} *Jean* [_I AGR^{pauvre} [_{VP} *mange*]]]

Or, l'analyse de Rizzi (1986a) montre que, au contraire, ce schéma s'étend également à ce type de dialectes ; « la seule différence avec l'italien standard est que AGR fort

¹ Avec une modification de notre part : nous traduisons directement ses exemples anglais en français.

² Rizzi (1986a:392).

est réalisé phonétiquement non seulement sur la morphologie verbale visible, mais aussi sur sa position syntaxique abstraite », comme illustré en (31) ci-dessous ; phénomène identifié en dialectologie comme la « réduplication de l'accord ».

(31) La représentation des dialectes du nord de l'Italie :

[_{IP} *El Gianni* [_I *el* [_{VP} *magna*]]]

Le débat surgit alors lorsque l'on compare une structure du trentin comme celle présentée en (31) avec un énoncé du français oral du type présenté en (32) ci-dessous :

(32) L'une des sources du débat :

Jean il mange.

En effet, face à une telle similitude (présence concomitante d'un syntagme nominal et d'un clitique co-indicés), la tentation est grande de supposer que le français aussi réalise la structure présentée en (31) ci-dessus. Or, Rizzi (1986a) argumente contre un tel rapprochement, notamment sur la base du comportement des sujets français lorsque ceux-ci contiennent un quantifieur comme le pronom indéfini *personne*. En effet, pour l'auteur, un énoncé français devient agrammatical dès lors que l'on associe la quantification à la présence concomitante d'un syntagme nominal sujet et d'un clitique, comme illustré en (33a) ci-dessous¹. Il explique cette agrammaticalité par le fait qu'en français, une telle structure impliquerait la présence concomitante des deux éléments (*personne* et *il*), ce qui est impossible car le syntagme nominal serait alors disloqué à gauche, dans la position du topique, comme indiqué en (33b), ce qui est incompatible avec la caractéristique des topiques de véhiculer une information ancienne et donc identifiable².

¹ Notons que cet énoncé comporte non seulement un quantifieur mais également la particule négative *ne* sur laquelle nous revenons dans le cours du Chapitre IV.

² Cf. Section III.5.2.

(33) L'autre source du débat :

(a) **Personne il ne mange.*

(b) *[_{TOP} *personne* [_{IP} *il ne* [_I AGR^{pauvre} [_{VP} *mange*]]]]

Rizzi (1986a) conclut ainsi que les clitiques du trentin et du français « diffèrent de part leur statut catégoriel et leur position structurelle »¹. Toutefois, la position et les arguments de Rizzi (1986a) ne font pas l'unanimité et un débat s'instaure et dure depuis autour de cette problématique.

2.2 Un débat binaire : syntaxe vs morphologie

À de rares exceptions près², le débat amorcé par l'étude des dialectes du nord de l'Italie est généralement posé de façon binaire : soit tous les clitiques nominatifs du français oral sont considérés comme des arguments syntaxiques soit ils sont tous analysés comme des affixes verbaux. Cette discussion est ainsi traditionnellement limitée non seulement au français oral mais également au français adulte. L'un des objectifs de ce travail de recherches est donc de reconsidérer les arguments généralement proposés dans le cadre de cette discussion, à la lumière de nos données enfantines, dans le but de proposer un point de vue nouveau sur la problématique, point de vue qui, nous l'espérons, permettra de faire avancer le débat. Lors de notre analyse, nous allons donc confronter notre corpus aux arguments majeurs à propos desquels les linguistes débattent depuis les années 1980. À cette fin, nous terminons notre tour d'horizon théorique par la description de ces arguments. La littérature étant extrêmement abondante sur ce sujet, nous choisissons d'exposer ce débat par le biais d'une synthèse proposée par De Cat (2005) qui, comme nous l'avons vu précédemment, argumente en faveur du statut argumental des clitiques nominatifs du français oral. Dans son article, l'auteur reprend ainsi les arguments présentés par la partie adverse³ et « teste les conséquences théoriques et empiriques » de l'analyse morphologique de ces clitiques, en concluant notamment que « l'analyse morphologique des clitiques sujets du français n'est pas tenable, tout au moins dans sa forme actuelle, dans le cadre d'une approche dérivationnelle de la grammaire »⁴. La démonstration de De Cat (2005) s'articule autour de trois points principaux : les

¹ Rizzi (1986a:414). Le revers de cette analyse, noté également par l'auteur, est que « la notion même de « clitique sujet » définit alors une classe syntaxique douteuse ».

² Cf. Cabredo Hofherr (2004:108) qui propose que « la syntaxe des clitiques sujets varie selon le nombre et la personne ».

³ Cf. notamment Auger (1994) et Miller & Monachesi (2003).

⁴ De Cat (2005:1217).

marques d'accord dans la phrase, l'indépendance syntaxique du clitique nominatif par rapport au verbe fléchi et le lien entre l'interprétation que l'interlocuteur a du syntagme nominal co-indicé et sa place dans la hiérarchie structurale¹.

2.2.1 La redondance des marques d'accord

Nous avons vu au Chapitre I que le français oral présente un taux élevé de structures dans lesquelles un syntagme nominal et un clitique co-indicés apparaissent de façon concomitante². Dans le cadre d'une analyse morphologique de ces clitiqes, il est postulé que ces derniers sont des morphèmes flexionnels au même titre que les suffixes verbaux dont le statut morphologique (et non syntaxique) ne soulève aucune contestation. Dans le cadre du Programme Minimaliste, les items lexicaux présents dans la numération portent dès ce stade leurs affixes flexionnels (signalés en termes de traits), la dérivation permettant notamment ensuite de vérifier leurs traits ininterprétables et d'attribuer une valeur aux traits phi présents sur v par l'intermédiaire de l'accord qui s'opère entre le syntagme nominal, T et v . Dans ce système, lorsque le clitique est considéré comme préfixe, il est injecté dans la structure concomitamment au verbe en V en tant qu'unité syntaxique représentée par la concaténation morphologique préfixe+radical+suffixe, le tout se déplaçant alors d'un seul bloc. De Cat (2005) soulève une première objection quant à une telle proposition dans la mesure où la présence d'un préfixe ajoutée à celle d'un suffixe « génère un certain niveau de redondance aléatoire dans le paradigme morphologique de l'accord verbal en français »³. Effectivement, lorsque l'on considère ses exemples, repris en (34) ci-dessous, on constate que : en (a), les traits de [Personne] et de [Nombre] sont marqués pré- et post-verbaleme nt de façon redondante ; en (b) et (c), la présence de ce clitique n'est cependant obligatoire que dans certains cas.

¹ Ce dernier critère a déjà été partiellement mentionné dans le cadre de la distinction entre topique et focus.

² Cf. par exemple Blanche-Benveniste (1994).

³ De Cat (2005:1197).

(34) L'accord est aléatoirement redondant :

- | | | | |
|-----|-----------|---------------|--------------------------|
| (a) | <i>Je</i> | <i>vais.</i> | |
| | [1, sg] | aller-[1, sg] | |
| (b) | Eux | \emptyset | <i>pourront</i> t'aider. |
| | | \emptyset | pouvoir-[3, pl] |
| (c) | *Moi | \emptyset | <i>pourrai</i> t'aider. |
| | | \emptyset | pouvoir-[1, sg] |

Par ailleurs, l'utilisation de la proclise (plutôt que l'enclise habituellement utilisée pour marquer l'accord en français) constitue pour l'auteur une singularité additionnelle de l'analyse morphologique qui, en fin de compte, aurait ainsi à justifier d'un certain nombre de particularités. Dans le cadre de nos recherches, il sera donc intéressant d'étudier la présence vs l'absence des clitiques nominatifs dans les énoncés enfantins afin d'établir si l'utilisation de ces éléments par les enfants revêt un caractère aléatoire et/ou redondant en termes d'accord.

2.2.2 L'indépendance syntaxique du clitique nominatif

S'il entre dans la dérivation syntaxique en tant que préfixe verbal, le clitique nominatif n'a aucune autonomie syntaxique. Cela signifie que, en tant que préfixe, il ne peut pas être manipulé par des computations syntaxiques et qu'il ne peut donc pas se retrouver ni dans une position postverbale ni séparé du verbe par un constituant qui, lui, aurait été manipulé par la syntaxe. Or, les clitiques nominatifs peuvent occuper une position postverbale (*cf.* (35a) ci-dessous) ou être séparés du verbe fléchi par d'autres clitiques, tels que la particule négative *ne* ou des clitiques qui portent d'autres cas, tels que l'accusatif, le datif ou l'ablatif (*cf.* (35b)) :

(35) Les différentes places du clitique nominatif par rapport au verbe fléchi¹ :

- | | | |
|-----|-----------------|---|
| (a) | Subséquente : | <i>Peut-il</i> avoir une petite bouchée ? |
| (b) | Non-adjacente : | Je <i>ne t'en</i> veux pas. |

Pour De Cat (2005), ces deux types de configurations ne peuvent donc pas survenir dans le cadre d'une analyse préfixale du clitique nominatif. Le débat sur le statut de ce clitique entraîne alors au moins deux autres. D'une part, il s'agit en effet de déterminer si l'ordre verbe-clitique présent dans les énoncés interrogatifs est un phénomène productif en français oral et, d'autre part, il faut statuer sur l'origine

¹ Exemples de De Cat (2005:1200).

(morphologique ou syntaxique) des éléments qui peuvent se trouver entre le clitique nominatif et le verbe fléchi. Concernant les énoncés interrogatifs, Auger (1994), se basant sur les études de Lambrecht (1981) et Zribi-Hertz (1994) notamment, postule que le phénomène d'inversion n'est pas appliqué en « français européen familier »¹, ce qui corrobore les données de De Cat (2005) pour le français de France (contrairement au français de Belgique par exemple) pour lequel seuls 2 % des questions comportent cette inversion². La problématique soulevée par l'exemple en (a) ci-dessus semble donc ne pas s'appliquer au cas du français que nous étudions. Nous vérifierons néanmoins avec les données enfantines si des phénomènes d'inversions apparaissent et dans quelles proportions.

La seconde problématique, illustrée par l'énoncé (35b) ci-dessus, nous semble devoir être scindée en plusieurs questions dans la mesure notamment où les différents clitiques mentionnés par De Cat (2005) font l'objet de traitements différents dans la littérature générative. En effet, une première distinction peut être établie entre la particule *ne* qui n'est pas un constituant qui relève de la valence verbale et l'ensemble des autres clitiques. Néanmoins, dans le cadre d'une analyse morphologique des clitiques nominatifs, il peut simplement être postulé que le clitique *ne* (quel que soit son statut) n'apparaît pas de façon productive en français oral ; c'est la thèse que soutient Auger (1994). *A contrario*, les données de De Cat (2005) établissent que « la particule *ne* est utilisée en français spontané en Belgique, au Canada et en France »³. Par ailleurs, suivant des exemples extraits des travaux de Rowlett (1998) sur la négation en français, l'auteur souligne que la présence de *ne* est parfois même obligatoire (dans des énoncés tels que *rien du tout n'a été fait*⁴). De nouveau donc, nous comparerons ces résultats (contradictaires, en ce qui concerne ce clitique) avec nos propres données enfantines afin d'établir le rôle précis que pourrait jouer cette particule dans le débat sur le statut des clitiques nominatifs. En ce qui concerne les autres clitiques (accusatifs, datifs, etc.), la problématique consiste à déterminer leur statut (morphologique ou syntaxique). Les deux hypothèses sont envisagées par les chercheurs. Ainsi, De Cat (2005) par exemple argumente en

¹ Cf. Auger (1994:67).

² Cf. De Cat (2005:1199) pour l'ensemble de ses données de Belgique, de Canada et de France.

³ Cf. De Cat (2005:1201) et notre Chapitre III pour le détail des données enfantines qu'utilise cet auteur (*The York and Cat Corpora*).

⁴ Néanmoins, comme le fait remarquer Aliyah Morgenstern (communication personnelle), la structure plus fréquente *rien (n') a été fait* ne permet pas d'établir la présence ou l'absence de *ne* du fait de la liaison.

faveur de leur statut syntaxique alors que Kaiser (1994) par exemple favorise l'approche morphologique de ces éléments. Dans le cadre de ce débat, De Cat (2005) soulève notamment la question de la montée des clitiques objets¹. En effet, ces clitiques n'apparaissent pas toujours dans la périphérie des verbes dont ils sont les arguments : ils « montent » donc dans la structure au cours de la dérivation et se retrouvent cliticisés à l'auxiliaire fléchi, comme illustré en (36) ci-dessous. Dans ce cadre, les autres positions intermédiaires ne sont pas licites.

(36) Illustration du processus de « montée du clitique »² :

Il [*les_i*] a [**les_i*] fait [**les_i*] lire [**les_i*].

A contrario, Kaiser (1994) propose que les clitiques objets soient analysés comme des affixes verbaux dans la mesure notamment où, lorsque ces clitiques sont doublés par un syntagme nominal, l'intonation de l'énoncé ne comprend pas de pause avant le syntagme (*cf.* (37) ci-dessous). Les clitiques entrent alors dans la numération déjà préfixés au verbe (en V°) et ils ne peuvent donc pas être manipulés isolément par la syntaxe. Nous étudierons donc également ces phénomènes dans le cadre des productions enfantines.

(37) L'absence de pause avant le syntagme nominal objet doublé par un clitique³ :

- (a) Marie l'a vu Jean hier.
- (b) Marie le lui a donné le livre à Jean.

2.2.3 Le lien entre interprétation et structure syntaxique

Cette dernière problématique concerne la présence d'un clitique nominatif doublé d'un syntagme et l'interprétation de ce syntagme. Ces questions relèvent alors de deux points différents mais liés : la structure syntaxique de l'énoncé et l'interprétation des éléments d'après leur place dans cette structure. Les deux interprétations en question sont celles de focus et de topique, telles que décrites précédemment⁴. Les deux structures syntaxiques correspondantes sont reprises en (38) ci-dessous ; celles-ci montrent également où se placent le syntagme (XP) et le clitique.

¹ Ce processus est notamment détaillé par Kayne (1989).

² Exemple extrait de De Cat (2005:1203).

³ *Cf.* Kaiser (1994:135).

⁴ *Cf.* Section III.5.2.

(38) Les structures syntaxiques des topiques et des focus en français oral¹ :

- (a) XP en focus : [CP ... [TP XP_i [T' clitique_i+T ...]]]
- (b) XP en topique : [CP XP_i [TP clitique_i [T' ...]]]

Nous avons déjà mentionné la position de De Cat (2005) sur cette problématique². Nous pouvons la reformuler ainsi : (i) la structure (a) ci-dessus n'existe pas en français oral parce qu'un syntagme en spécifieur de TP est obligatoirement un focus, (ii) un focus ne peut pas être repris par un clitique parce que la présence du clitique force à interpréter le syntagme comme un topique. Ce qui est donc postulé dans ce raisonnement est l'incompatibilité entre la présence d'un clitique pléonastique et une interprétation du syntagme en tant que focus. L'auteur réfère notamment à un test de jugement³ effectué auprès de quatorze locuteurs francophones (de Belgique, de Canada et de France) auxquels les deux structures sont présentées afin de véhiculer une information qui doit être mise en avant. Le contexte ainsi que les deux structures sont repris en (39) ci-dessous :

(39) Test de jugement effectué par De Cat (2005) :

- (a) Contexte : La voiture bleue est foutue.
- (b) Réponse 1 : Non, la voiture ROUGE elle est foutue.
- (c) Réponse 2 : Non, la voiture ROUGE est foutue.

Dans ce cadre, les locuteurs interrogés doivent indiquer quelle structure ils utiliseraient afin d'insister sur l'adjectif *rouge* : seuls 4,7 % des locuteurs acceptent la réponse 1. Il est donc conclu que la focalisation et la présence d'un clitique co-référent sont incompatibles et que la présence de ce clitique force donc la topicalisation du syntagme (ce qui, dans ce contexte, est impossible puisque l'adjectif *rouge* apporte une information nouvelle). À ce stade, nous aimerions simplement apporter un commentaire quant au type de test effectué. En effet, dans le protocole mentionné ci-dessus, les locuteurs choisissent à partir d'un inventaire fermé de réponses. Or, dans le cadre de l'étude d'une langue « spontanée », nous pensons que des réponses plus représentatives du système étudié auraient été obtenues si aucun modèle n'avait été fourni aux locuteurs. Ainsi, en nous basant sur le même contexte que De Cat (2005), nous avons demandé à quatre locuteurs d'âges différents de

¹ D'après De Cat (2007:22).

² Cf. Section III.5.3.

³ Cf. Chapitre III sur ce type de tests.

fournir une réponse à la situation évoquée, avec pour seule instruction d'insister sur le fait que c'est la voiture rouge et non la voiture bleue qui « est foutue ». Le résultat de notre enquête, repris en (40) ci-dessous, montre que la stratégie de focalisation n'est pas celle imaginée en (39c) ci-dessus mais qu'elle est tout autre puisque, dans tous les cas, il s'agit de structures dites clivées ou à présentatif :

(40) Réponses spontanées au même contexte :

- (a) Non, c'est la voiture ROUGE qui est foutue.
- (b) Non, c'est la ROUGE.

La question que nous posons face à ces résultats concerne alors la stratégie de focalisation et nous nous demandons s'il peut être affirmé que la focalisation en français oral est véhiculée par la structure syntaxique dans laquelle un syntagme apparaît en spécifieur de TP. En ce qui concerne le corpus enfantin, il nous faudra donc vérifier : (i) si ce type de structure véhicule (uniquement) la focalisation et (ii) comment la focalisation est véhiculée par ailleurs. Cela nous permettra de faire le point par rapport à la première partie de l'argumentation de De Cat (2005) ((i) « la structure (a) en (38) ci-dessus n'existe pas en français oral parce qu'un syntagme en spécifieur de TP est obligatoirement un focus »).

La seconde partie de l'argumentation de l'auteur ((ii) « un focus ne peut pas être repris par un clitique parce que la présence du clitique force à interpréter le syntagme comme un topique ») relève de l'incompatibilité à utiliser certains quantifieurs et certains indéfinis en position de topique. Nous reprenons les exemples emblématiques de cette démonstration en (41) et (42).

(41) Incompatibilité entre topiques et quantifieurs¹ :

- (a) Expression avec quantifieur : **Personne*, il n'a rien dit.
- (b) Expression sans quantifieur : *Pierre*, il n'a rien dit.

(42) Incompatibilité entre topiques et indéfinis² :

- (a) **Quelqu'un*, il est venu.
- (b) **Un enfant*, il est arrivé et il t'a posé une question.

Les énoncés ci-dessus sont transcrits avec une virgule entre le syntagme et le clitique pléonastique ; ce signe indique que ces structures sont envisagées en tant que topicalisations (avec le syntagme en position non-argumentale et le clitique en position argumentale). Mais dans la littérature on trouve également les versions suivantes :

(43) Autres présentations de l'association syntagme + clitique pléonastique :

- (a) *Personne* il a rien dit³.
- (b) En campagne, quand *quelqu'un* il-dansait, le curé il-le-montrait presque du doigt.⁴
- (c) *Tout le monde* il se baignait, là-bas⁵.

Les exemples de ce type sont généralement proposés par les partisans de l'hypothèse morphologique du clitique. En effet, les éléments *personne*, *quelqu'un*, *tout le monde*, etc. ne peuvent pas représenter des topiques (c'est-à-dire une information ancienne, donc connue et définie). Ainsi, si ces éléments peuvent être associés à un clitique, c'est que le clitique est, dans ce cas, un affixe verbal puisque le syntagme occupe la position argumentale. L'un des problèmes de cette démonstration concerne les exemples sur lesquels elle repose. En effet, il s'agit souvent des mêmes occurrences que l'on retrouve d'une démonstration à une autre. Elles semblent donc rares et cette rareté ne joue pas en faveur de cette hypothèse. Nous examinerons donc également notre corpus à la recherche d'éventuelles cooccurrences entre éléments quantifiés ou indéfinis et clitiques nominatifs.

¹ Zribi-Hertz (1994:134).

² De Cat (2007:82).

³ Zribi-Hertz (1994:137).

⁴ Cabredo Hofherr (2004) d'après Auger (1993).

⁵ Blanche-Benveniste (2000:42).

Nous avons annoncé vouloir associer théorie et empirie. Une partie importante de la théorie générative a été décrite dans ce chapitre. Tournons-nous donc à présent vers la partie empirique de nos recherches, c'est-à-dire, dans un premier temps, vers la réflexion que nous avons menée afin de constituer notre propre base de données enfantines (Chapitre III) avant de nous plonger dans les données proprement dites (Chapitre IV).

CHAPITRE III : LA CONSTITUTION DE NOTRE BASE DE DONNÉES

Toute humaine qu'elle est, la linguistique moderne n'en est pas moins une science. Ce statut scientifique de la linguistique implique qu'elle doit se soumettre à une méthodologie rigoureuse. Comme le physicien ou le biologiste, le chercheur en linguistique observe des phénomènes puis il les décrit et les analyse. Les recherches dans tous ces domaines, que ces sciences expérimentales soient dites exactes ou humaines, procèdent ainsi de données. L'acquisition du langage par exemple peut être étudiée à partir d'un grand nombre de langues différentes à travers le monde. Elle peut tout aussi bien porter sur l'acquisition de la langue première d'un jeune enfant, que sur un éventuel bilinguisme, ou sur l'apprentissage plus tardif d'une seconde ou d'une troisième langue. La recherche peut se focaliser sur des sujets considérés comme normaux ou sur des personnes atteintes de pathologies et elle peut être soit longitudinale¹ soit transversale². De plus, chacune de ces thématiques peut être examinée sous plusieurs angles théoriques possibles³. Il s'ensuit logiquement que les recherches en acquisition du langage sont abondantes et diverses. Une caractéristique rapproche néanmoins toutes ces perspectives. En effet, quelle que soit la langue, le stade de l'acquisition, son degré de « normalité », le point de vue ou le cadre théorique choisi, les chercheurs ont ceci en commun qu'ils ne peuvent émettre une hypothèse qu'à partir d'un minimum de données. L'outil de travail, prérequis de toute recherche, est donc un échantillon du système linguistique à analyser, c'est-à-dire un corpus. Mais cette définition du terme corpus est-elle suffisante ? Qu'attend-on d'un tel outil ? La première partie dans ce chapitre apporte des éléments de

¹ C'est-à-dire qu'elle s'intéresse au suivi développemental d'un ou de plusieurs enfants.

² Étude comparative d'un phénomène chez plusieurs enfants.

³ Au Chapitre II, nous avons vu la distinction qui existe entre connexionnistes et générativistes par exemple.

réponse à ce type de questions, qui sont tant épistémologiques que pratiques, grâce à une réflexion approfondie sur la notion de corpus et à un état des lieux critique des corpus enfantins préexistants à notre étude. Ces observations nous incitent alors à constituer notre propre base de données enfantines. Les étapes de son élaboration sont détaillées dans la deuxième et la troisième partie de ce chapitre ; celles-ci reprennent en effet les étapes de la collecte des données tout d'abord puis celles qui ont jalonné le moment où ce recueil d'informations est devenu un outil informatisé, c'est-à-dire une base de données. Puis, la quatrième partie détaille l'une des étapes fondamentales de ce travail, celle de la transcription de ces données. Elle nous permet ainsi d'explicitier notre méthodologie et les difficultés que nous avons rencontrées. La cinquième partie complète ce chapitre en détaillant les deux processus de codage que nous avons utilisés : le codage des agrammaticalités relevées dans les productions enfantines et le codage morphosyntaxique de la totalité des énoncés. À ce stade, la base de données est alors prête à être manipulée à des fins de statistiques et d'analyses.

I. RÉFLEXIONS ÉPISTÉMOLOGIQUES ET MÉTHODOLOGIQUES SUR LE CORPUS DE DONNÉES ENFANTINES¹

« L'analyse ne vaut que ce que vaut le corpus »². Cette phrase sans appel donne le ton : un mauvais échantillon, que ce soit en linguistique ou dans n'importe quelle autre science, préfigure une mauvaise théorie. Prenons un exemple tout à la fois extrême et léger pour illustrer ce propos. Imaginons un océanographe qui souhaiterait étudier la faune marine dans la rade de Villefranche, qui y prélèverait une citerne d'eau et qui n'y trouverait aucun poisson³. Dans ce cas, il est tout aussi flagrant que l'échantillon est mauvais et que les conclusions seront erronées. Mais qu'est-ce qui nous permet d'en être aussi sûre et qu'en est-il pour les échantillons linguistiques ? La première section ci-après ambitionne donc de définir avec précision les notions d'échantillon et de corpus en linguistique et de cerner ce que nous en attendons afin de proposer un travail de recherches scientifiquement rigoureux. Ce n'est qu'une fois

¹ Nous tenons à remercier Sophie Wauquier, directrice de l'UMR 7023 « Structures formelles du langage », qui en avril 2006 avait très gentiment commenté une version antérieure, première ébauche de cette partie.

² Dalbera (2002:94).

³ Pour des méthodes plus scientifiques, cf. l'Observatoire Océanologique de Villefranche sur mer à l'adresse Internet <http://biodev.obs-vlfr.fr/>.

que ces notions fondamentales sont clairement définies qu'il devient intéressant de découvrir les bases de données existantes, de façon de plus en plus ciblée (dans les Sections I.2 et I.3), afin de déterminer (en Section I.4) que ces bases de données ne correspondent pas à nos attentes, ce qui nous amène (en Section I.5) à définir notre propre cahier des charges.

1. Vous avez dit « corpus » ?

L'exemple de l'océanographe ci-dessus montre que, pour que l'analyse d'un échantillon ait une quelconque validité scientifique, il faut que, dans un premier temps, le chercheur s'intéresse aux caractéristiques de son échantillon. Toujours dans le cas de notre océanographe, il est évident que son échantillon présente le problème majeur de ne pas être représentatif de son objet d'étude. Cette constatation permet donc de supposer que l'une des caractéristiques *sine qua non* d'un échantillon est d'être le plus représentatif possible d'un système complet, sans quoi les observations et les hypothèses qui en découlent ne valent que pour cet échantillon, n'ont donc aucune perspective de généralisation et ne présentent ainsi qu'un intérêt scientifique extrêmement limité. Mais comment un chercheur sait-il qu'un échantillon de données est représentatif ? Pour être certain de cette représentativité, il faudrait déjà bien connaître le système en question, ce qui *a priori* n'est pas le cas ; sinon pourquoi l'étudier ? Le problème semble alors circulaire. Une partition établie dans le domaine de la linguistique par Dalbera (2002) aide à y voir plus clair. En effet, l'auteur distingue « deux phases dans une étude linguistique : la phase d'analyse d'un ensemble fini de données et la phase de confrontation des résultats de cette analyse [...] à la réalité ». Il s'ensuit qu'il existe au moins deux types d'échantillons linguistiques. Le premier est celui qui est collecté de façon aléatoire au cours de la première phase de la recherche. Découvrant son objet d'étude, le chercheur espère alors que ses données seront représentatives de l'ensemble. Le corpus peut alors être défini comme un « sous-ensemble d'éléments susceptibles de représenter le système linguistique complet étudié ». Mais, comme le précise Dalbera (2002), il s'agit alors tout autant d'un « pari » que d'une « hypothèse » scientifique. Les distributionnalistes américains dans les années 1930 par exemple se sont trouvés dans ce cas de figure lorsqu'ils ont commencé à étudier les langues amérindiennes, encore mal connues à l'époque¹. Le second type d'échantillon est celui qui est

¹ Leonard Bloomfield par exemple.

collecté par le chercheur qui a déjà une connaissance de son objet d'étude. C'est le cas de figure qui correspond à l'état actuel des connaissances dans de nombreux domaines et pour de nombreuses thématiques en linguistique, dont l'acquisition du langage¹. L'aspect totalement aléatoire des corpus est donc dépassé et son caractère représentatif n'est plus un pari mais une exigence scientifique. À ce stade de l'investigation, le corpus se définit alors comme un « échantillon représentatif du système linguistique étudié ou du phénomène étudié ». Cette condition en entraîne alors une autre : que le corpus ne soit plus un matériau brut à partir duquel le travail de recherches peut démarrer mais un outil construit résultant de choix humains. Il reste alors l'immense tâche de définir quels sont les choix à effectuer. Les réflexions développées ci-après, avec notamment l'exemple de l'application de filtres², ainsi que les détails de notre analyse³, sont là pour cerner, argumenter et codifier nos propres choix.

À ce premier type de questionnement, s'ajoute le fait que, alors qu'en sciences exactes, les données utilisées résultent exclusivement d'expériences réalisées en laboratoire, en linguistique, les sources des données sont plus diverses. En effet, en fonction du domaine d'application au sein de la linguistique, les données peuvent soit également être collectées en laboratoire (le cadre est alors plutôt neurolinguistique ou psycholinguistique⁴) soit collectées à l'extérieur d'un laboratoire, sur « le terrain » (il peut alors notamment s'agir de dialectologie, de sociolinguistique ou d'acquisition du langage⁵). L'étude du langage implique ainsi que le linguiste recueille souvent ses données dans un cadre non scientifique. Le « terrain » peut alors être un environnement privé tel que le domicile, un environnement plus public tel qu'une école, ou un cadre totalement ouvert à tous comme la rue. Conserver un cadre de travail scientifique à partir de données collectées dans un environnement non scientifique représente ainsi un défi supplémentaire pour la linguistique de terrain.

Mais pour l'heure, procédons comme n'importe quel scientifique dans n'importe quelle discipline : focalisons-nous précisément sur la thématique de notre recherche – l'acquisition du sujet syntaxique – et sur la langue concernée par notre étude – le français. Dans un premier temps, il convient alors de se demander quel type de

¹ Cf. Section 2.2 ci-après pour un bref historique de la discipline.

² Cf. les données filtrées *a posteriori* présentées dans le Tableau 14 page 163.

³ Cf. Chapitre IV.

⁴ Cf. Section 2.1.

⁵ Cf. Section 2.2.

données apporte des informations susceptibles d'éclairer ce domaine spécifique d'étude. Les différentes catégories de corpus enfantins destinées à l'étude de l'acquisition de la syntaxe sont ainsi passées en revue dans la Section 2 afin d'identifier celle(s) qui répond(ent) à nos besoins.

2. Les différents types de données enfantines

Il existe trois principaux types de données qui permettent d'étudier l'acquisition de la syntaxe : le chercheur peut analyser soit les *productions* linguistiques enfantines, soit la *compréhension* que les enfants ont de certaines structures, soit leurs *jugements* à propos de la véracité d'une affirmation ou de la grammaticalité d'une occurrence. Le chercheur peut également choisir de combiner deux de ces méthodes ou même les trois. Son choix dépend de plusieurs facteurs, notamment de son cadre théorique, de l'objet de son étude et de la tranche d'âge concernée.

2.1 Les données issues de la compréhension et du jugement

L'avantage indéniable des tests de compréhension par rapport aux deux autres types de tests est qu'ils peuvent être pratiqués très tôt dans le développement linguistique du jeune enfant. En effet, *de facto*, ces tests ne nécessitent pas que l'enfant soit en âge de produire des énoncés syntaxiquement complexes ; ils permettent donc de collecter des informations dans une tranche d'âge inaccessible aux deux autres types de données. L'émergence de ces énoncés complexes étant généralement située vers l'âge de dix-huit mois, cela signifie que les tests de compréhension permettent d'évaluer le développement linguistique chez des enfants très jeunes. Mais en quoi consistent exactement ces tests et que permettent-ils d'évaluer ?

D'une manière générale, pratiquer des tests de compréhension chez des enfants qui n'expriment pas encore des énoncés complexes présuppose que les productions linguistiques du jeune enfant (la performance, au sens de Chomsky) ne sont le reflet que d'une partie de sa compétence grammaticale sous-jacente et que, en conséquence, l'enfant sait plus de choses que ce qu'il peut effectivement exprimer. Destinés à évaluer des enfants très jeunes (à partir de douze mois environ), les différents tests de compréhension reposent généralement sur un même type de protocole qui consiste à solliciter le regard de l'enfant. Le principe est le suivant : pour chaque stimulus linguistique proposé à l'enfant (un son, un mot ou une phrase),

au moins deux stimuli visuels le sont aussi, grâce à des dessins ou des photographies, par le biais d'écrans de télévision¹. Il est supposé que, si l'enfant comprend le stimulus sonore, celui-ci doit porter son regard plus longtemps sur la représentation qui y correspond plutôt que sur celle qui n'y correspond pas. Il s'agit donc essentiellement d'une étude de la compréhension des contrastes, qu'ils soient phonologiques, morphologiques, syntaxiques ou lexicaux. Par exemple, ce type de test permet à Hirsh-Pasek & Golinkoff (1996a) de montrer que les enfants de quatorze mois sont sensibles à l'ordre des mots dans une phrase et à Hirsh-Pasek *et al.* (1995) de mettre en avant la compréhension des Principes A et B de la théorie du Liage de Chomsky (1981) par des enfants de trois ans^{2,3}. Ce type de test est ainsi généralement utilisé jusqu'à l'âge de quatre ans, âge auquel le chercheur peut également recourir aux autres types de données que nous allons mentionner ci-après.

Un des principaux inconvénients de ce genre de protocole est qu'il semble impossible d'en tirer des données longitudinales, c'est-à-dire de suivre l'évolution d'un phénomène linguistique particulier chez un ou plusieurs enfants. D'une part, ce type de test nécessite une mise en place très rigoureuse à de nombreux niveaux pour que les résultats soient scientifiquement recevables. L'environnement qui entoure l'enfant notamment est capital. En effet, qu'il soit sonore, visuel, tactile ou olfactif, celui-ci ne doit lui apporter aucune distraction pendant les tests. De plus, les stimuli visuels proposés doivent apparaître exactement à la même fraction de seconde et à une distance identique afin de ne pas influencer la réaction de l'enfant. Par conséquent, ces tests sont toujours réalisés au sein de laboratoires d'études expérimentales avec un personnel formé à ce genre de pratiques et du matériel spécifique afin de contrôler au maximum les conditions de leur déroulement. Ces nécessités présentent ainsi l'inconvénient d'imposer un environnement non familier et des personnes étrangères aux jeunes sujets, même si tout est fait pour que les enfants se sentent le mieux possible. Tous les enfants ne se plient donc pas à ces protocoles rigoureux et il n'est pas toujours aisé de mener une séance de tests à son

¹ Il s'agit d'un protocole mis en place par Golinkoff *et al.* (1987) à partir des travaux de Spelke (1979) et appelé le *intermodal preferential looking paradigm*.

² Cf. Hirsh-Pasek & Golinkoff (1996b) et Gerken & Shady (1996) pour d'autres exemples d'études et pour de plus amples détails sur les protocoles utilisés.

³ Le Principe A stipule que les anaphores (les pronoms réfléchis et les réciproques) doivent être c-commandés par un élément co-indicé qui se trouve dans la même proposition (exemples : *Rémi_i se_i lave vs *Rémi_i se_k lave*). Le Principe B stipule que les pronoms ne doivent pas être c-commandés par un élément co-indicé qui se trouve dans la même proposition (exemples : *Juliette_i la_k lave vs *Juliette_i la_i lave*).

terme¹. D'autre part, il semble encore plus délicat de demander à un même enfant de renouveler un type identique d'expérience pour des raisons évidentes de lassitude qui peuvent soit faire que l'enfant ne mène pas la nouvelle séance à son terme soit qu'il regarde n'importe quel stimulus visuel afin de terminer la séance au plus vite parce que cela ne l'amuse plus. Ce premier type de données n'est donc pas adapté à une étude comme la nôtre dans la mesure où nous avons notamment besoin de données longitudinales. Néanmoins, des conclusions issues de certains de ces tests, comme celles de Hirsh-Pasek *et al.* (1995) sur le liage des pronoms, peuvent sans doute apporter des indications intéressantes et complémentaires à une étude longitudinale sur l'acquisition des pronoms.

Les tests de jugement² présentent à peu près les mêmes inconvénients que les tests de compréhension. En effet, ils exigent le même type d'environnement neutre afin de ne pas fausser les résultats et ils semblent donc également peu adaptés à une étude longitudinale. Néanmoins, comme les tests de compréhension, ils peuvent apporter un éclairage complémentaire à d'autres données. Les travaux de Crain & McKee (1985) par exemple portent sur le Principe C de la théorie du Liage³ et peuvent ainsi également apporter des informations supplémentaires.

Le troisième type de données enfantines, celles issues de la production, apparaissent alors comme les seules qui puissent nous éclairer de façon adéquate sur le développement du sujet enfantin. En effet, ce type de données, contrairement aux deux précédents, peut être recueilli dans un cadre autre que celui d'un laboratoire de recherches. Il permet donc de collecter des données auprès d'un ou de plusieurs enfants de façon longitudinale sans se heurter à l'aspect intrusif mentionné plus haut pour les tests de compréhension et de jugement dans la mesure où le suivi peut être effectué dans un cadre qui est familier aux enfants comme le domicile, la crèche ou l'école⁴. De plus, le phénomène que nous souhaitons étudier implique que nous nous intéressions en priorité aux productions enfantines syntaxiquement complexes dans

¹ Hirsh-Pasek & Golinkoff (1996b:118) par exemple mentionnent un taux d'échec allant de 15 à 50 %, en fonction de l'âge des enfants.

² Cf. Gordon (1996) sur les jugements de véricité et McDaniel & Cairns (1996) sur les jugements de grammaticalité.

³ Le Principe C stipule que les expressions référentielles (les noms propres et les noms communs) doivent être libres, c'est-à-dire qu'elles ne doivent pas être c-commandées par un élément co-indicé (exemples : *Stéphane_i a mangé des frites quand il_i était dans la cuisine* vs **Il_i a mangé des frites quand Stéphane_i était dans la cuisine*).

⁴ Nous avons vu au Chapitre I que c'est cette méthode qu'a notamment adoptée Morgenstern (2006) avec Léonard.

la mesure où la notion de sujet est une notion fonctionnelle, définie en référence à un autre constituant capital de la phrase : le verbe¹. Donc, si l'on postule qu'un énoncé comporte un sujet, cela implique qu'il comporte alors au moins également un verbe et il s'agit de fait d'une structure syntaxiquement articulée. Notre première décision porte donc sur le choix du type de données à utiliser – les données issues de la production – dans la mesure où la tranche d'âge que nous souhaitons étudier le permet, que nous ne travaillons pas dans le cadre d'un laboratoire de recherches expérimentales et que c'est un type de collecte moins intrusif que les tests de compréhension et de jugement.

2.2 Les données issues de la production

Il existe deux principaux types de productions orales enfantines : les productions spontanées et les productions induites. L'induction est l'approche la plus récente et date des années 1950. Elle consiste à solliciter la compétence linguistique du jeune enfant grâce à des tests préalablement et précisément formatés en fonction du domaine de recherches. Comme les tests de compréhension et de jugement, ils se déroulent donc en général en laboratoire de recherche. L'un des premiers protocoles de sollicitation est le test du *wug* mis en place par Jean Berko, considérée comme la pionnière dans ce domaine². Le *wug* est un animal imaginaire (qui ressemble à un oiseau) dont on présente des dessins à l'enfant. L'un des premiers objectifs de ce test est d'évaluer la compétence en morphologie de jeunes enfants anglophones ; il permet notamment aux chercheurs de solliciter des formes associant un radical nominal ou verbal à un suffixe du pluriel ou du passé. Pour le pluriel par exemple, le chercheur montre un puis deux dessins de *wug* à l'enfant en lui disant : « Ceci est un *wug*. Maintenant, il y en a deux. Il y a deux _ . » Il est donc demandé à l'enfant de produire le pluriel du nom *wug* afin de tester sa capacité à combiner un radical nominal et le suffixe du pluriel *-s*. Le fait que le nom *wug* n'existe pas garantit alors aux chercheurs que l'enfant n'a jamais entendu la forme *wugs* auparavant et que donc, en cas de production, celle-ci ne peut absolument pas être le résultat d'une imitation mais uniquement d'une computation de sa part.

Dans le domaine de la syntaxe, de nombreux protocoles de sollicitation sont exploités. Crain & Nakayama (1987) par exemple sollicitent la production de trente

¹ Cf. Chapitre I sur la notion de sujet.

² Cf. Berko (1958).

enfants anglophones entre trois et cinq ans afin de déterminer si la règle que les enfants appliquent pour former les questions oui/non repose sur la linéarité des constituants (l'enfant apprendrait alors à formuler ces questions en inversant le sujet et le premier verbe des phrases déclaratives) ou sur la connaissance d'une structure sous-jacente (l'enfant emploierait alors une règle de dépendance à la structure qui fait intervenir une analyse plus abstraite en syntagmes¹, en faisant notamment intervenir la notion d'enchâssement indiquée par les crochets dans les exemples ci-dessous).

(1) Exemples de productions sollicitées par Crain & Nakayama (1987) :

(a) Phrase déclarative produite par le chercheur à l'enfant :

The boy [who is happy] can see Mickey Mouse.

« Le garçon qui est heureux peut voir Mickey »

(b) Le chercheur sollicite l'enfant : il demande à celui-ci de demander à une tierce personne (souvent une figurine) où se trouve le garçon :

Ask Jabba if the boy [who is happy] can see Mickey Mouse.

« Demande à Jabba si le garçon qui est heureux peut voir Mickey »

(c) Deux productions sont alors possibles :

Linéarité : **Is_i the boy [who e_i happy] can see Mickey Mouse?*

Structure : *Can_i the boy [who is happy] e_i see Mickey Mouse?*

Crain & Nakayama (1987) observent ainsi qu'aucun enfant du groupe ne déplace le premier verbe qui se trouve dans la proposition relative au cours de ce test. Les chercheurs établissent ainsi que les enfants connaissent de façon implicite la règle de la dépendance à la structure. Dans le cadre de la grammaire générative, une telle règle appartient à la Grammaire Universelle, c'est-à-dire à l'ensemble des principes innés et universels dont chaque enfant, en tant qu'être humain, bénéficie à sa naissance.

Ce type de sollicitation présente l'avantage pour le chercheur de travailler sur un nombre conséquent de structures similaires et ciblées, structures qui ne seraient produites spontanément qu'à un rythme beaucoup plus sporadique par de jeunes enfants. Cependant, lorsqu'un chercheur sollicite une structure syntaxique en particulier, de fait, il se focalise sur celle-ci. Une conséquence d'une telle

¹ Cf. Chomsky (1975:43).

investigation est alors que le chercheur ne recherche pas forcément un éventuel parallèle avec d'autres structures et que, donc, il ne vise pas une possible généralisation dans le système linguistique de l'enfant¹. De plus, ce type de sollicitation s'effectue souvent de façon ponctuelle sur un groupe d'enfants. En effet, bien que présentés sous forme de jeux, ces tests, comme les tests de compréhension et de jugement, peuvent parfois devenir fatigants ou ennuyeux pour certains enfants. De plus, il est difficile de proposer le même type de test plusieurs fois à un même enfant sans envisager que celui-ci n'y décèle certaines régularités. Dans ce cas, ses réponses ne seraient plus uniquement le reflet de sa compétence linguistique mais comporteraient également une part de conditionnement. Cette méthode n'est donc pas généralement propice à un suivi longitudinal d'enfants ; elle n'est donc pas non plus particulièrement adaptée à notre problématique.

Nous avons donc vu que les premiers tests de sollicitation remontent aux années 1950. Cependant, l'intérêt porté par l'homme à l'acquisition du langage et à des thématiques qui lui ont été apparentées, telles que la naissance du langage et des différentes langues ou le changement linguistique², est bien sûr beaucoup plus ancien que le XX^e siècle. La légende du Prince égyptien Psammétique, qui date des années 400 avant J.C.³, montre à quel point l'acquisition de la langue première et la question de l'origine des langues sont des phénomènes qui, de tout temps, ont été étroitement liés et ont fasciné les hommes. Ce double intérêt, après avoir donné naissance à des hypothèses théologiques et philosophiques, engendre grâce à l'avènement des sciences des propositions plus cartésiennes. Ainsi, certains comparatistes du XIX^e siècle, tels que l'helléniste Egger (1879), s'intéressent à la question de l'acquisition du langage. Ils travaillent alors sur un type de données unique : les données enfantines spontanées. Les linguistes travaillent ainsi à partir de carnets de notes, sortes de cahiers de bord comme peuvent en tenir de nombreux parents, dans lesquels un adulte garde une trace écrite de ce que l'enfant dit de façon spontanée au cours de ses activités quotidiennes, le plus souvent dans le cercle familial. À

¹ Cf. notre analyse au Chapitre IV et le rapprochement que nous établissons entre les différentes élisions dans notre corpus.

² Cf. par exemple Passy (1890), phonéticien français et l'un des initiateurs de l'Alphabet Phonétique International (API) : « Tous les changements majeurs qui touchent à la prononciation et que nous avons été capables d'étudier prennent leur source dans le langage infantin. »

³ Cette légende est rapportée par Hérodote ([1930]:64-66) : Psammétique, voulant savoir quel peuple était le plus ancien (les Égyptiens ou les Phrygiens) proposa d'élever deux enfants dans le silence absolu afin de savoir quelle langue ils parleraient en premier. La légende dit que leur premier mot fut phrygien (*bekos* « pain »). Les Égyptiens reconnurent ainsi que les Phrygiens étaient plus anciens qu'eux.

l'époque, ces cahiers de notes rendent alors souvent compte uniquement du développement lexical de l'enfant et ne sont pas tenus par des personnes formées à la transcription particulièrement difficile des données enfantines¹. De plus, de telles notes rendent souvent compte des éléments les plus saillants dans le discours enfantin, c'est-à-dire généralement des tournures typiquement enfantines et des agrammaticalités ; les données sont donc partielles. Il faut attendre les années 1930 et le travail du linguiste belge Grégoire (1937, 1947) pour disposer de données spontanées longitudinales, notées en phonétique par un linguiste et accompagnées d'indications sur le contexte de production. La recherche sur l'acquisition du langage se spécialise alors et se diversifie à la fois ; à l'intérêt initial porté à l'acquisition du lexique s'ajoutent en effet l'étude de l'acquisition de la phonologie (avec notamment Jakobson (1941, 1960)), et celle de la syntaxe (avec entre autres la naissance de la grammaire générative à la fin des années 1950 aux États-Unis²). Les progrès technologiques du XX^e siècle permettent ensuite de passer de recueils manuscrits à des enregistrements audio puis vidéo.

Ce type de données permet ainsi de suivre le développement d'un ou de plusieurs enfants dans un cadre familial. Il semble donc correspondre tout particulièrement au type de données que nous souhaitons exploiter pour notre étude. Nous optons par conséquent pour une étude de données spontanées mais, comme mentionné auparavant, sans pour autant fermer la porte aux autres types de collectes dans la mesure où nous pensons que ces dernières peuvent éventuellement être exploitées dans un deuxième temps pour compléter des données spontanées déjà étudiées.

Nous avons donc étudié avec soin les différents types de données qu'un chercheur en acquisition de la syntaxe peut utiliser afin de mener à bien une étude et nous avons choisi un type de données bien particulier, adapté à notre type de recherches. Notre deuxième étape consiste à présent à considérer les bases de données déjà constituées et composées de productions spontanées, en nous focalisant sur les seules données francophones. Notre démarche sera similaire ; elle consiste à prendre connaissance des collectes qui existent déjà afin de choisir au mieux celles qui sont le plus adaptées à notre travail.

¹ Cf. Sections I.4.1 et IV de ce chapitre.

² Cf. Chomsky (1959) et Chapitre II pour le détail.

3. Les bases de données spontanées francophones

À notre connaissance, cinq principaux ensembles de données spontanées produites par des enfants francophones de la tranche d'âge qui nous intéresse ont été collectés par des chercheurs en acquisition afin d'étudier le développement de la syntaxe dans le cadre de la grammaire générative. Nous devons donc à présent passer en revue les caractéristiques de ces ensembles afin de déterminer si nous pouvons travailler à partir de l'un ou plusieurs d'entre eux.

3.1 Les différentes collectes

Toutes les personnes qui se sont un tant soit peu intéressées à la littérature générative traitant de l'acquisition connaissent bien Philippe, Nathalie, Daniel et Grégoire. La raison de cette relative célébrité auprès des chercheurs réside dans le fait que, jusqu'à ces dernières années, les seuls corpus de données enfantines francophones spontanées utilisés pour la recherche étaient constitués des données produites par ces quatre enfants. En effet, Philippe¹ (peut-être le plus cité de tous) a été enregistré par Madeleine Léveillé² en 1971 à Paris ; Nathalie et Daniel ont été enregistrés par Patsy Lightbown³ en 1974/1975 à Montréal ; et les productions de Grégoire ont été collectées par Christian Champaud⁴ en 1988 à Paris. Les caractéristiques principales de ces trois bases de données sont détaillées dans le Tableau 10 ci-après :

Pseudonymes	Âges	Séances	Intervalles (semaines)	Auteurs	Dates	Lieux
Philippe	2;1 à 2;6	12	1 ⁵	Léveillé	1971	Paris
Nathalie	1;9 à 2;3	7	3 ½ à 4 ⁶	Lightbown	74/75	Montréal
Daniel	1;8 à 1;11	5	3 ½ à 4	Lightbown	1975	Montréal
Grégoire	1;9 à 2;3	7	2 à 7	Champaud	1988	Paris

Tableau 10 : les enfants « stars » de la littérature générative

Pendant des années donc, notamment pendant les années 1980 et 1990 (voire même au-delà), qui représentent la période durant laquelle le modèle des Principes et des

¹ Les quatre prénoms sont des pseudonymes. Nous revenons en détail sur la question de l'anonymisation dans les Sections II.2.2, II.3.1 et IV.3.1.

² Du Laboratoire de Psychologie C.N.R.S. Paris. Il s'agit d'enregistrements audio. Cf. Suppes *et al.* (1973).

³ Il s'agit également d'enregistrements audio. Cf. Lightbown (1977).

⁴ Du Laboratoire de Psychologie C.N.R.S. Paris. Les données de Grégoire proviennent d'enregistrements vidéo.

⁵ Avec une interruption de 3 mois entre les séances n° 10 et n° 11.

⁶ Avec une interruption de 5 semaines.

Paramètres suscite un foisonnement des recherches génératives en acquisition, les données transcrites par et accessibles à des chercheurs sont restées limitées. Cette offre relativement restreinte de données francophones dans une période où la recherche connaît un véritable essor représente ainsi la raison de la « célébrité » de ces quatre enfants dans la mesure où, de fait, ce sont leurs productions qui apparaissent de façon quasi-systématique dans la littérature générative sur l'acquisition du français. Les travaux de Crisma (1992) sur les questions *wh*-¹ et ceux de Côté (2001) sur les clitics sujets, par exemple, reposent exclusivement sur les productions de Philippe ; Friedemann (1993) utilise les données de Philippe et de Grégoire pour son étude sur la position du sujet ; et Pierce (1992), pour son étude comparative sur l'acquisition de la syntaxe principalement du français et de l'anglais, Déprez & Pierce (1993), pour la négation, et Ferdinand (1996), sur le développement des catégories fonctionnelles, fondent tous leurs hypothèses sur les productions conjointes de ces quatre enfants.

Plus récemment, d'autres collectes sont venues compléter ces premiers ensembles. Il s'agit notamment de données recueillies en Suisse par l'Université de Genève, sous la direction d'Uli Frauenfelder et de Luigi Rizzi. Un premier enfant (Augustin) est ainsi suivi pendant dix mois. Les données de cet enfant ont principalement servi de base aux travaux de Hamann *et al.* (1994, 1996) sur l'acquisition du système pronominal du français. Deux autres enfants suisses appartenant à une même fratrie (Marie, la sœur aînée et Louis, son frère) ont également été enregistrés en 2002 par leurs parents. Un quatrième enfant (Jean) a été enregistré par Lucienne Rasetti à Genève en 1996. Les détails correspondants à ce groupe d'enfants se trouvent dans le Tableau 11 ci-après :

¹ Ce sont les questions du type *que, qui, où*, etc. (nommées questions *wh*- d'après les termes anglais *what, who, where*, etc.).

Pseudonymes	Âges	Séances	Intervalles (semaines)	Auteurs	Dates	Lieux
Augustin	2;0 à 2;9	10	3 à 4 ¹	Univ. de Genève	1993	Neuchâtel
Marie	1;8 à 2;3	17	2	Parents	2002	Genève
Louis	1;9 à 2;3	12	2	Parents	2002	Genève
Jean	1;7 à 2;0	4	4	Rasetti	1996	Genève

Tableau 11 : les données suisses

Nous retrouvons les données de ces quatre enfants, combinées à celles de Philippe, Daniel et Nathalie dans les travaux de Rasetti (2003) par exemple sur les catégories grammaticales omises par les jeunes enfants.

De Cat (2002, 2007) exploite deux autres sources de données pour ses travaux sur la dislocation en français. Le *York Corpus* constitue sa première source. Ces données ont été collectées sous la direction de Bernadette Plunkett dans le cadre d'une recherche sur l'acquisition des questions *wh-* auprès de trois enfants (Léa, Max et Anne) entre 1997 et 1999. Sa seconde source, le *Cat Corpus*, est temporairement composée des transcriptions des données audio provenant d'un seul enfant (Tom) sur les quatre enregistrés. Les détails correspondants se trouvent dans le Tableau 12 ci-dessous :

Pseudo.	Âges	Fichiers	Intervalles (semaines)	Auteurs	Dates	Lieux
Léa	2;8 à 4;3	36	2	Grand-mère	97/99	Liège
Max	1;9 à 3;2	37	2	Étudiante	97/99	Montréal
Anne	1;10 à 3;5	35	2	Étudiante/ parents	97/99	Paris
Tom	2;1 à 2;10	16	2 à 3	Grand-mère	98/01	Bruxelles
Chloé				Mère		Bruxelles
Lisette	2;8		2 ½ h ²	C. De Cat	2000	Montréal
Pauline			3 h ³	C. De Cat	2000	Montréal

Tableau 12 : les données de B. Plunkett et C. De Cat

Très récemment, une quatrième base de données francophones a été diffusée par le NIMH⁴. Il s'agit d'un ensemble de conversations spontanées mère-enfant, qui comporte les enregistrements de quatre enfants entre douze et quarante-huit mois

¹ Avec deux interruptions : 2 mois (entre les séances n° 7 et n° 8) et 2 ½ mois (entre les séances n° 8 et n° 9). Cf. Hamann *et al.* (1994).

² Il s'agit ici d'un total de 2 ½ heures d'enregistrements sur une période de quatre semaines.

³ 3 heures d'enregistrements sur la même période de quatre semaines

⁴ *National Institute of Mental Health.*

collectés sous la direction d’Harriet Jisa du Laboratoire « Dynamique du Langage » (UMR 5596) à Lyon. Les détails se trouvent dans le Tableau 13 ci-après :

Pseudonymes	Âges	Fichiers	Intervalles (semaines)	Dates
Anaïs	1;0 à 3;0	41	2	2002/2004
Marie	1;0 à 4;0	46	2	2002/2004
Nathan	1;0 à 3;0	43	2	2002/2004
Tim	0;11 à 3;00	43	2	2002/2004

Tableau 13 : le corpus de Lyon

Le « Projet Léonard », mentionné au Chapitre I, constitue également une pièce maîtresse du paysage acquisitionnel du français. Il s’agit « d’étudier l’apparition et le fonctionnement des outils grammaticaux chez l’enfant [de douze mois à trois ans] de les comparer à leur fonctionnement en langue adulte et à leur évolution en diachronie et de les situer par rapport au développement moteur, cognitif, psychique de l’enfant »¹. Ce projet, coordonné par Aliyah Morgenstern de l’Université Sorbonne Nouvelle Paris 3 comprend une base de données déjà constituée ; il s’agit des productions de Léonard entre vingt et trente-neuf mois². D’autres ensembles (neuf enfants nés entre 2003 et 2006) sont en cours de constitution.

L’offre en données semble donc à présent assez conséquente. Pourquoi alors Philippe, Nathalie, Daniel et Grégoire restent-ils les « stars » dans la littérature spécialisée ?

3.2 La diffusion de ces données

Un facteur très important qui rend possible l’utilisation d’une base de données que l’on n’a pas élaborée personnellement est sa diffusion. À cet effet, les données enfantines disposent d’un site sur Internet qui leur est entièrement dédié. Il s’agit de CHILDES, acronyme de *Child Language Data Exchange System*³ qui permet le partage de productions enfantines. Les chercheurs qui le souhaitent peuvent ainsi y déposer leurs données et/ou en télécharger gratuitement. Cette base de données, créée en 1984 par Brian MacWhinney et Catherine Snow, entre autres chercheurs, n’a cessé de croître⁴ et elle comprend à présent un grand nombre de données

¹ Cf. le site Internet du projet : <http://anr-leonard.ens-lsh.fr/>.

² Nous détaillons ce corpus au Chapitre I.

³ « Système d’échange de données linguistiques enfantines ». Cf. <http://childes.psy.cmu.edu/>.

⁴ Cf. MacWhinney (2000a) pour l’historique complet.

enfantines¹ ; celles-ci y sont cependant très majoritairement anglophones. En 2000 (date de la dernière édition papier de l'ouvrage de MacWhinney), le site comptait quarante-deux ensembles pour l'anglais contre seulement quatre pour le français. À titre de comparaison, l'espagnol en comptait neuf, le néerlandais sept, l'allemand et l'hébreu quatre, le portugais trois, et l'italien deux ; d'autres langues, telles que le cantonais, l'estonien, le grec, le japonais, etc., y figurent également par l'intermédiaire d'un ou deux fichiers. Les quatre bases francophones présentes en 2000 étaient celles de Christian Champaud (les données de Grégoire), de Madeleine Léveillé (les données de Philippe), et deux autres que nous n'avons pas encore citées : celle de Jean Rondal et celle d'Helga Feider et de Madeleine Saint-Pierre (nommé « corpus Montréal »). Ces deux derniers ensembles n'ont pas été exploités par les chercheurs en acquisition dans le cadre générativiste car, pour le « corpus Montréal » les données ne correspondent pas à la tranche d'âge souhaitée (l'enfant est plus âgé : 5;6 à 11;6) et pour le « corpus Rondal », il s'agit d'interactions mère-enfant en français avec un peu d'anglais. En revanche, c'est probablement cette très large diffusion grâce à Internet qui explique la présence récurrente des productions de Philippe et de Grégoire dans de nombreux travaux sur l'acquisition de la syntaxe. Depuis 2000 cependant, deux importants corpus sont venus grossir les rangs des données francophones puisque le « corpus de York » et celui de Lyon sont disponibles sur le site depuis peu.

Cependant, CHILDES n'est bien sûr pas le seul site où l'on peut trouver des données enfantines. Des extraits des données du Projet Léonard par exemple sont consultables et téléchargeables sur la plateforme CLAPI². En conséquence, malgré l'absence de diffusion de certaines des données francophones citées précédemment, il semble que l'éventail des données spontanées francophones partagées reste assez large pour le chercheur qui souhaite utiliser un ou plusieurs ensembles existants. Il convient à présent d'étudier ces quelques bases de données en détail afin d'y pratiquer une sélection.

4. Les limites des bases de données existantes

Il nous semble que le regard que porte De Cat (2002) sur les corpus préexistants à ses recherches sur la dislocation marque un tournant dans l'approche que les

¹ La liste exhaustive est disponible sur Internet et dans MacWhinney (2000b).

² Corpus de Langue Parlée en Interaction. Cf. <http://clapi.univ-lyon2.fr>.

chercheurs peuvent avoir face à l'utilisation des données enfantines. En effet, comme mentionné dans la Section 3.1, de nombreux chercheurs avaient jusque là utilisé des données préexistantes à leurs propres travaux, provenant d'un ou de plusieurs enfants. Cette solution continue à être souvent adoptée car elle présente l'avantage indéniable de permettre de commencer le travail directement par l'étape du choix des énoncés à analyser¹, sans avoir à organiser une collecte de données (qui peut durer jusqu'à plusieurs années), à effectuer les transcriptions (travail extrêmement long et délicat²) ni l'éventuel étiquetage morphosyntaxique des occurrences. Le gain de temps est donc considérable. Cependant, l'utilisation de données telles que celles de Philippe, Nathalie, Daniel ou Grégoire présente également à nos yeux un certain nombre d'inconvénients majeurs.

4.1 Les problèmes liés à la collecte et à la transcription

Il est unanimement admis que toute transcription représente *de facto* une première interprétation des données produites³. En effet, une transcription est toujours le résultat d'un travail de conversion, par un transcripateur, de données originales, brutes, orales – appelées données primaires⁴ – en données écrites, qui portent alors le nom de données secondaires⁵. Les réflexions sur les difficultés de cette tâche ne manquent d'ailleurs pas. Celle de Snyder (2007) résume bien la situation en ce qui concerne la transcription du langage enfantin :

“Unfortunately for researchers, children’s speech is seldom perfectly clear. A major challenge in transcription is therefore to render the child’s utterances in a way that is both theoretically neutral and maximally informative. To make matters worse, these two goals often conflict.”⁶

¹ Cf. Section I du Chapitre IV sur l'extraction préliminaire des données.

² Cf. Baude (2006:30) : « Selon les estimations courantes, un minimum de trente minutes de travail est nécessaire pour transcrire une minute d'enregistrement. ». L'estimation de Demuth (1996:17) est similaire.

³ Cf. notamment Ochs (1979) et Dalbera-Stefanaggi (2001:9-41) sur cette thématique.

⁴ Cf. Baude (2006:45) : « Les données primaires sont constituées par les enregistrements [...]. Elles comprennent aussi les autres objets recueillis dans le contexte de l'action, comme les documents lus ou écrits [...], les objets manipulés, les images consultées, etc. »

⁵ Cf. Baude (2006:45) : « Les données secondaires sont constituées par la série de descriptions, transcriptions, annotations qui viennent enrichir les données primaires et qui sont souvent fournies après coup et sur la base de données. Elles comprennent aussi les métadonnées, les conventions de transcription, les autorisations des participants, etc. »

⁶ Snyder (2007:53-54) : « Malheureusement pour les chercheurs, le discours enfantin est rarement limpide. Un des principaux défis lors de sa transcription est donc de restituer les énoncés enfantins

Lorsqu'un chercheur entre en contact avec des données par le biais de leurs transcriptions, sa connaissance de ces productions enfantines est donc indirecte, médiata. Ce fait ne constitue pas en lui-même un obstacle rédhibitoire à l'exploitation de corpus déjà transcrits car, évidemment, la qualité du travail de transcription effectué par les différents transcripateurs n'est absolument pas remise en cause dans la mesure où les données mentionnées ci-dessus ont toutes été collectées et transcrites par des professionnels et une des destinations des bases de données est le partage par la communauté scientifique, les corpus n'étant bien sûr pas à usage unique. Cependant, nous pensons que ce choix méthodologique peut entraîner un certain nombre de faiblesses dans le corpus dans certains cas.

4.1.1 La pluralité des sources

Tout transcripateur doit pratiquer des choix tout au long de son travail. Ces choix, bien qu'étant limités au maximum par un souci d'objectivité scientifique, ont cependant chacun des conséquences. À nos yeux par exemple, il n'y a rien de plus difficile à transcrire que la longueur d'une pause dans un énoncé ; correspond-elle à une virgule, un point-virgule, deux points, un point, ou encore à un autre symbole relevant du système de transcription utilisé¹ ? Et quelle garantie y a-t-il que le même type de pause sera restitué de la même façon tout au long du travail de transcription ? Dans le cadre de notre étude sur le sujet, cette question est cruciale² :

« La transcription est une pratique qui, loin de se limiter à un exercice technique de reproduction, intègre de nombreux enjeux théoriques et interprétatifs. »³

Chaque transcripateur effectue donc des choix en tant qu'être humain qui perçoit et décode un continuum phonique. Ses choix sont également guidés par d'autres facteurs que sa perception seule. Le type de transcription choisi (orthographe standard, adaptée, Alphabet Phonétique International, etc.) impose les limites de son propre système ; le cadre théorique dans lequel travaille le transcripateur peut également influencer, etc. « Comme le signalait déjà E. Ochs en 1976, la transcription

d'une façon qui soit à la fois théoriquement neutre et maximale informative. Pour rendre les choses plus difficiles encore, ces deux objectifs sont souvent contradictoires. »

¹ Cf. Section IV ci-après sur nos conventions de transcription.

² Cf. Section IV ci-après sur la transcription et Chapitre I sur les implications théoriques de la distinction entre « redoublement » et « dislocation ».

³ Baude (2006:73).

engage toujours une théorie »¹. Rien ne permet donc d'établir que le chercheur-utilisateur d'un corpus préexistant aurait transcrit les productions originales de la même façon que le chercheur-transcripteur, qui lui-même n'est peut-être déjà pas le chercheur qui était sur le terrain, au contact avec les enfants. Que penser alors de données, regroupées en tant que « corpus » à des fins d'analyse, mais qui ont été transcrites par plusieurs personnes, à des époques distinctes et dans des pays différents² ? Cette diversité dans les sources introduit incontestablement de la variation au niveau du traitement des données primaires et, par voie de conséquence, une certaine hétérogénéité non souhaitée et non souhaitable dans les données secondaires. *A contrario*, une expérience, si elle veut être crédible aux yeux du monde scientifique, se doit de faire la chasse à toutes les sources qui sont susceptibles d'introduire une telle variation. En sciences exactes, ces variations, qualifiées de « non expérimentales » ou « parasites » sont traquées. Par exemple, au sein d'une même expérience, une science exacte ne comparera pas des données relevées avec des appareils différents. La question de la pluralité des sources a déjà été également abordée et traitée dans le domaine expérimental. Jakubowicz (1996) déclare par exemple que :

“The question arises whether or not it is legitimate or interesting to use the results obtained by these [different] investigators for comparative purposes [...]. The common sense indicates that because results like these arise from different testing conditions, they must be cautiously interpreted but not disregarded.”³

En effet, afin d'être analysé correctement, un phénomène doit avant tout être isolé et identifié sans quoi le chercheur risque de proposer une hypothèse faussée par la prise en compte d'éléments interférents. La linguistique, malgré son statut de science « humaine » se doit de ne pas négliger l'existence et les implications de ces variables afin de se fixer des protocoles scientifiques aussi rigoureux⁴. L'utilisation de données provenant d'un seul transcripteur ou tout au moins d'un seul groupe de transcriptions

¹ Baude (2006:30).

² Cf. par exemple les données regroupées dans le Tableau 10 : les enfants « stars » de la littérature générative, en page 149.

³ Jakubowicz (1996:283) : « La question se pose de savoir s'il est légitime ou intéressant de comparer des résultats obtenus par ces [différents] chercheurs [...]. Le bon sens suggère que, parce que de tels résultats proviennent de conditions de tests différentes, ils doivent être interprétés avec prudence mais pas ignorés. »

⁴ Cette remarque s'applique bien sûr à d'autres domaines de la linguistique. Cf. par exemple Oliviéri (à paraître) sur la transcription de certaines formes dialectales de clitiques.

travaillant selon un protocole de transcription identique semble donc représenter un premier critère à retenir en vue d'annuler les variations possibles dues à ce type d'hétérogénéité. Un fait, cependant, rendrait cette restriction superflue. Il s'agit de la possibilité pour le chercheur-utilisateur de consulter à tout moment, en plus des transcriptions, les données originales, par le biais d'éventuels enregistrements audio et/ou vidéo.

4.1.2 L'absence de support audio ou vidéo

Malheureusement, les transcriptions disponibles ne sont pas systématiquement accompagnées des données primaires correspondantes. Un second inconvénient vient alors s'ajouter lorsqu'un chercheur utilise des transcriptions préexistantes, provenant de différentes sources et pour lesquelles il n'a pas accès aux productions originales : le chercheur-utilisateur n'entend pas, ne voit pas et, *a fortiori*, ne connaît pas les enfants enregistrés. Il n'a pas non plus vécu la situation pendant laquelle les données ont été produites. En l'absence de données sonores et/ou visuelles, l'accès à de nombreuses informations, telles que la description de la scène, les personnes en présence, les relations de l'enfant avec ces personnes, son état d'esprit, son humeur du moment, ses gestes, son intonation, etc., n'est disponible qu'à travers d'éventuelles notes situationnelles prises par le chercheur présent lors de l'enregistrement et éventuellement reportées lors de la transcription. Il semble que ce contact indirect avec les données peut, dans un premier temps, desservir la qualité de la transcription (si les personnes qui collectent et transcrivent sont différentes) puis, dans un deuxième temps, de l'analyse dans la mesure où, aux deux niveaux, des erreurs d'interprétation peuvent se glisser dans la transcription et/ou l'analyse et ainsi nuire à la qualité et à la validité du travail final. Ainsi, De Cat (2002) n'utilise pas les corpus de CHILDES car les transcriptions présentes sur le site à cette date ne sont pas accompagnées des données sonores correspondantes et ne comprennent pas les détails prosodiques nécessaires à son étude sur la dislocation. Il nous semble que le point de vue de De Cat (2002) est donc novateur et qu'il permet de progresser d'un grand pas dans la rigueur nécessaire à l'analyse de données. L'auteur insiste en effet sur l'importance de la prosodie ainsi que sur la nécessité de tenir compte des éléments situationnels et pragmatiques lors de l'analyse de productions enfantines. Elle illustre ainsi les effets qu'induit une connaissance très détaillée du moment de la production grâce à deux exemples repris en (2) ci-dessous.

(2) Exemples de De Cat (2002)¹ :

- (a) Papa, i(l) tient pas. (Anne, 2;2)
- (b) ça, c'est moi. (Max, 2;0)

Dans la phrase d'Anne, en (2a) ci-dessus, le nom *Papa*, sans notes situationnelles, sans informations sur la prosodie et tel qu'il est transcrit (suivi d'une virgule), aurait toutes les chances d'être analysé comme un constituant disloqué, ayant le même référent que le pronom subséquent *il*. Seule la connaissance du moment de production et/ou la notation très précise de l'intonation permet de savoir que le nom *Papa* est en réalité dans cette phrase un vocatif. En effet, l'enfant parle non pas de son père lorsqu'elle utilise le pronom *il* mais d'un réveil. La ponctuation utilisée lors de la transcription ne rend donc pas clairement compte du cas vocatif de *Papa*. Une transcription plus explicite aurait donc été souhaitable comme par exemple *Papa ! I(l) tient pas* qui semble davantage rendre compte du cas vocatif de *Papa* et de la non-coréférence du nom et du pronom. La présence de notes situationnelles indiquant que l'enfant parle d'un réveil aurait également levé toute ambiguïté. La phrase de Max, en (2b) ci-dessus, présente également un « piège ». Sans notes situationnelles très précises, le pronom *ça* peut également être interprété comme étant une dislocation ayant le même référent que le pronom subséquent. En effet, on peut imaginer que l'enfant enregistré est en train de montrer une photographie ou un dessin le représentant. Seule la connaissance des détails du moment de l'énonciation permet de savoir que la peinture que l'enfant montre au moment où il parle ne le dépeint pas et que le pronom *ça* représente en réalité une dislocation d'objet non prononcé. En effet, la scène est la suivante : l'enfant montre des dessins et explique à son interlocuteur qui les a réalisés. Une phrase plus complète et explicite, d'après contexte, aurait alors été *ça, c'est moi qui l'ai fait*, montrant effectivement que *ça* représente un dessin (objet du verbe) et non l'enfant (sujet du verbe). C'est pourquoi De Cat (2002) n'exploite que des données avec lesquelles elle déclare avoir une certaine proximité. En l'occurrence, il s'agit du « Corpus de York », mentionné en page 151, pour lequel elle a participé, en tant qu'assistante, à la vérification, à l'étiquetage et au codage des transcriptions, grâce au fait que les séances avec les enfants avaient été filmées. Comme mentionné dans la Section I.3.2, les

¹ De Cat (2002:27-28). Les transcriptions sont reprises telles que fournies par l'auteur (les mots, les majuscules et la ponctuation sont identiques à l'original).

transcriptions de ce corpus sont maintenant disponible sur CHILDES, accompagnées des vidéos correspondantes.

Ces deux exemples illustrent à quel point l'utilisation de transcriptions non accompagnées d'un support audio ou vidéo peut avoir des conséquences néfastes. En effet, si certaines occurrences – et personne ne connaît leur proportion – sont mal interprétées, ces erreurs, bien qu'involontaires, se répercutent dans les statistiques du chercheur et donc inévitablement dans son analyse qui, de fait, est alors erronée. Le regard porté sur les transcriptions isolées semble ainsi avoir changé ces dernières années ; soit que certains chercheurs préfèrent à présent utiliser leurs propres données (comme dans le cas de De Cat (2002)), soit que certains auteurs mettent en garde les chercheurs dans des ouvrages sur la méthodologie de la collecte des données spontanées. Ainsi, sur les deux chapitres consacrés à ce type de données dans McDaniel *et al.* (1996), Demuth (1996) et Stromswold (1996) invitent les chercheurs à compléter leurs enregistrements par des notes contextuelles ou des enregistrements vidéo et mentionnent l'énorme inconvénient que représente la connaissance médiata que propose CHILDES des données orales à cette époque :

“Because transcripts of spontaneous speech provide only limited information about the phonetic form or nonlinguistic context of utterances, the problem of interpreting children's utterances is magnified if the person analyzing the transcripts does not have access to videotapes or audiotapes of the transcribed sessions (e.g., when researchers analyze CHILDES transcripts that they did not collect).”¹

Un choix fondamental qui s'impose donc à tout chercheur qui souhaite utiliser une base de données est de n'étudier que des transcriptions doublées de données *orales*, que la collecte soit préexistante à son étude ou que celle-ci ait été effectuée par le chercheur lui-même. La présence de moyens vidéo, comme l'a été en son temps l'introduction de moyens audio, représente ainsi un énorme progrès pour l'étude du langage en général. En effet, la vidéo autorise un accès différé et répété aux données

¹ Stromswold (1996:26) : « Parce que les transcriptions de productions spontanées n'apportent que peu de renseignements phonétiques ou contextuels, le problème de l'interprétation de productions enfantines est amplifié si la personne qui analyse les transcriptions n'a pas accès à des enregistrements vidéo ou audio des séances transcrites (par exemple, lorsque des chercheurs analysent des transcriptions de CHILDES qu'ils n'ont pas collectées eux-mêmes) ».

orales en situation. Des données complètes peuvent ainsi être partagées par un grand nombre de chercheurs, libre ensuite à chacun de proposer sa propre interprétation de tel geste articulatoire ou corporel, ou de telle ligne prosodique. En tout cas, l'étude de données uniquement par le biais de leur transcription est donc à proscrire absolument, sous peine de proposer une analyse complètement faussée. C'est la raison pour laquelle nous n'utilisons pas ce type de transcriptions.

Nous pourrions alors nous servir des corpus de York et/ou de Lyon mis à disposition sur le site de CHILDES ces dernières années. Mais, pour ces données également, nous souhaitons émettre des réserves. En ce qui concerne le corpus de Lyon, il s'agit tout simplement du fait que la tranche d'âge des quatre enfants enregistrés ne correspond que partiellement à celle que nous souhaitons étudier. En effet, comme mentionné dans le Tableau 13, les transcriptions des données d'Anaïs, de Marie et de Tim vont jusqu'à leurs trois ans. Or, comme précisé au cours du Chapitre I, nous avons besoin de données qui couvrent une période plus tardive afin d'étudier le développement du sujet enfantin. En effet, trois ans se trouve être l'âge, en général admis dans la littérature, auquel l'enfant quitte le stade du « sujet nul enfantin »¹. Il s'agit donc d'une période charnière et des données rendant compte du stade subséquent sont nécessaires afin d'étudier comparativement le stade du « sujet nul » et le stade suivant, afin de mettre en évidence les raisons pour lesquelles l'enfant passe d'un stade à l'autre. Seules les données de Marie, le quatrième enfant du corpus de Lyon, couvrent une période plus longue puisqu'elles vont jusqu'à ses quatre ans. Le corpus de Marie correspond donc bien à nos besoins en ce qui concerne la tranche d'âge de l'enfant (entre un et quatre ans) et les transcriptions sont accompagnées des vidéos correspondantes. Pourtant, nous avons choisi de ne pas exploiter ces données non plus.

4.2 Les problèmes liés à la variation linguistique

Chaque être humain possède des caractéristiques distinctes. Chacun d'entre nous a en effet sa propre personnalité et est ancré dans une certaine époque, avec un environnement lié à sa culture, à son lieu d'habitation, etc. Toutes ces caractéristiques se répercutent bien sûr dans le langage.

¹ Cf. Chapitre II pour les détails sur cette étape du développement et notre première analyse proposée initialement dans Palasis (2005).

4.2.1 *La variation individuelle*

L'inconvénient que nous souhaitons mettre en lumière à présent est que, si nous avions travaillé à partir des données de Marie, nous aurions formulé des hypothèses sur le développement du système linguistique en général à partir des données d'un seul enfant. Or, nous trouvons ce choix méthodologique hasardeux dans la mesure où chaque personne, qu'il s'agisse d'un enfant, d'un adolescent ou d'un adulte, possède une façon propre de s'exprimer et que l'aspect individuel du discours enfantin, dès le plus jeune âge, est un fait reconnu. De Boysson-Bardies (1996) déclare notamment que :

« Ces dernières années, il était dans le vent de tenir pour négligeables, sinon inintéressants, tous les effets que pouvaient engendrer les différences de langues ou de personnalités sur les modes d'acquisition. »¹

Or, « Le « jeu des possibles » chez l'être humain est souvent bien plus puissant que ne le pensent les chercheurs »². Chaque enfant adopte ainsi un style plutôt qu'un autre au cours de sa deuxième année³ et cette diversité continue à être visible par la suite dans le choix des structures syntaxiques et du lexique. Alors comment compenser toutes ces individualités lorsque l'on cherche à mettre au jour un processus acquisitionnel général ? Utiliser des corpus de grande taille, composés de productions issues de nombreux enfants apparaît alors comme une solution logique. Cohen (1924) par exemple préconise déjà cette approche :

« Un fait observé chez l'enfant n'est bien utilisable [...] que si la part originale de l'individu peut y être délimitée, ce qui ne se réalise bien que par des comparaisons nombreuses. »⁴

C'est d'ailleurs cette voie que choisissent Jakubowicz & Rigaut (1997) par exemple, qui présentent « une étude transversale des productions linguistiques de 12 enfants francophones âgés de 2;0 à 2;7 ». Plus récemment, les travaux de Chevrot *et al.*

¹ De Boysson-Bardies (1996:173 *et sqq.*).

² Cf. également Oliviéri (à paraître) sur l'importance de la variation individuelle.

³ Celui-ci peut être plutôt référentiel ou expressif et dépend de plusieurs facteurs. Cf. De Boysson-Bardies (1996:173-201).

⁴ Cohen (1924:34).

(2009)¹ sur la liaison en français font appel à deux cents enfants entre 2;4 et 6;1. Ces chiffres montrent clairement que l'utilisation de données issues d'enfants plus nombreux fait d'ores et déjà partie des habitudes des chercheurs lorsqu'ils travaillent dans le domaine expérimental. Par contre, cette pratique n'est pas appliquée à l'étude des productions spontanées dans le cadre de la grammaire générative. Friedemann & Rizzi (2000) utilisent par exemple le terme de « grand corpus » lorsqu'ils mentionnent les travaux de Rasetti (1996) qui portent sur six enfants. Demuth (1996), dans son travail consacré à la méthodologie de la collecte des productions spontanées, ne promeut pas non plus les grands corpus, même si elle mentionne qu'il est préférable de travailler avec plus qu'un seul enfant :

“It is [...] preferable to collect spontaneous production data from more than one child. Given a target of three children, it may be advisable to start a study with four.”²

L'auteur fait ainsi référence au corpus anglais de Brown (1973) composé des productions de trois enfants (Adam, Eve et Sarah), très largement diffusé dans la littérature sur l'acquisition de la syntaxe de l'anglais. Elle précise que ce corpus « propose un bel échantillon de variation » et que cette diversité « est vitale à la construction d'une théorie cohérente de l'acquisition »³. De nouveau, il est à noter que cette attitude, plutôt quantitativement minimaliste, préconisée dans le cadre de la collecte de productions spontanées est totalement à l'opposé de celle proposée dans le cadre expérimental. Le but commun à ces deux méthodes de collecte étant le même – déceler les comportements représentatifs de la compétence des enfants par le biais de l'étude de leur performance linguistique –, considérer la variation individuelle comme étant un facteur positif et souhaitable semble paradoxal. C'est pourquoi les données de Jakubowicz & Rigaut (1997, 2000) sont composées, non seulement de productions spontanées, mais également de données recueillies « au cours d'une tâche expérimentale de production induite ». Les auteurs estiment en effet que ces deux situations différentes de collecte permettent de déterminer si les

¹ Jean-Pierre Chevrot du Laboratoire LIDILEM à Grenoble ; Céline Dugua du LLL/CORAL à Orléans et Michel Fayol du LAPSCO à Clermont-Ferrand.

² Demuth (1996:7) : « Il est [...] préférable de recueillir des données de production spontanée chez plus qu'un seul enfant. Avec un objectif de trois enfants, on peut recommander de démarrer l'étude avec quatre ».

³ Demuth (1996:7) : “Brown's (1973) study of three children provides a nice sample of variation, where Eve is much precocious than either Adam or Sarah. Such diversity is vital to constructing a coherent theory of acquisition”.

omissions du sujet qu'elles étudient¹ relèvent de la compétence linguistique des enfants ou uniquement du niveau de la performance². Ce souhait de prendre en compte les possibles interférences dues aux éventuelles limites computationnelles ou mémorielles des jeunes enfants est intéressant. En effet, dans cette optique de travail, les facteurs liés à la performance représentent alors également des variables parasites. Cependant, Jakubowicz & Rigaut (1997) reconnaissent elles-mêmes qu'il manque « une caractérisation précise des facteurs de performance ». Et, dans la mesure où ceux-ci sont encore mal connus et cernés, il paraît difficile de pouvoir éliminer toutes les conséquences. Quoi qu'il en soit, cette interrogation ne remet pas en cause le fait qu'une analyse linguistique qui tient compte d'un grand nombre de sujets a forcément une légitimité plus grande pour formuler des généralisations qu'une analyse qui se fie à un nombre d'enfants peu élevé.

Notre étude du système verbal de dix-sept enfants entre 2;3 et 3;1³ avait ainsi mis en avant ces individualités dès la petite enfance. Le verbe *adorer* par exemple avait été utilisé massivement par un seul enfant (Malcolm) dans un certain type de phrase, comme le montrent les totaux du Tableau 14 ci-après.

Verbes	Totaux	Mat	Jul	Tess	Thi	Cla	Max	Raph	Noé	Alis	Aliz	Mal	Tho	Emi
<i>être</i>	81	1	2	1	11		4	46	1	3	10		2	
<i>adorer</i>	14											14		
<i>avoir</i>	14				1		2	7	1	1	1	1		
<i>faire</i>	13		1	1	4			6			1			

Tableau 14 : extrait du classement, par ordre décroissant de fréquence, des verbes dans les phrases [+Red, +Pro]⁴

Cette utilisation importante de *adorer* avait placé ce verbe en deuxième position en ce qui concerne sa fréquence d'utilisation dans ce type particulier de phrase, après le verbe *être* (81 occurrences sur un total de 182 phrases) et à égalité avec le verbe *avoir* (14 occurrences). Cependant, les autres colonnes du tableau montrent que ce verbe *adorer* n'avait été produit que par un seul enfant du groupe. Comptabiliser ce verbe comme étant l'un des verbes les plus souvent prononcés par ces jeunes enfants n'aurait donc pas été représentatif de l'ensemble des productions du groupe et ce, malgré son nombre élevé d'occurrences. Il avait ainsi été écarté de l'étude par

¹ Cf. plus particulièrement Jakubowicz & Rigaut (1997).

² Elles suggèrent ainsi que les omissions du sujet relèvent de la performance et qu'elles sont la conséquence d'une opération de réduction phonologique.

³ Cf. Palasis (2005:118-127).

⁴ Phrases comprenant un sujet composé d'un DP redoublé ou disloqué et d'un pronom, comme dans l'exemple *Il est là nounours* (Alizée, 2;6).

application d'un filtre *a posteriori* selon lequel nous avons décidé de comptabiliser uniquement les verbes produits par au moins trois enfants différents. Par contre, si notre étude avait porté sur les productions d'un seul enfant et que cet enfant avait été Malcolm, notre étude n'aurait pas été représentative et, pire, nous n'en aurions rien su. C'est pourquoi nous souhaitons fonder nos recherches sur des productions d'un échantillon le plus vaste possible d'enfants afin de proposer une analyse qui soit légitimement généralisable à la classe d'âge étudiée. Cependant, il convient d'être très prudente car plus un chercheur augmente le nombre de ses informateurs, plus il accroît en parallèle les risques d'introduire dans ses données d'autres variations non souhaitées.

4.2.2 La variation collective

Traditionnellement, en plus de la variation individuelle, quatre facteurs permettent de rendre compte des variations qui existent entre les différents systèmes linguistiques. Cette variation peut en effet être de type diachronique (intergénérationnelle par exemple), géographique (entre le français de France et celui du Québec par exemple), diastratique (pour le langage enfantin, entre catégories socioprofessionnelles des parents) et diaphasique (par exemple, le français de l'enfant face à un inconnu comparé à celui que l'enfant utilise lorsqu'il s'adresse à un membre de sa famille ou, inversement, le langage qu'utilise un adulte lorsqu'il s'adresse à un autre adulte et celui qu'il emploie lorsqu'il s'adresse à un jeune enfant¹). Si un ensemble de données cumule plusieurs types possibles de variations, il s'ensuit très logiquement que les différences effectivement constatées ne peuvent pas être identifiées et étudiées de façon univoque. Nous illustrons notre propos par une tentative de comparaison entre deux corpus enfantins.

Crisma (1992) utilise les données de Philippe (recueillies en 1971) pour son étude sur l'acquisition des questions *wh-* en français. Quelques-unes de ces questions sont reprises en (3) ci-dessous.

¹ Il s'agit alors du L.A.E. (Langage Adressé aux Enfants) décrit notamment au Chapitre I.

(3) Quelques questions *wh-* de Philippe¹ :

- (a) *Où* il est le fil ?
- (b) *Où* elle est la petite aiguille ?
- (c) *Où* elle est maman ?
- (d) *Où* elle est la voiture grosse voiture ?

Crisma (1992) souhaite ainsi tester l'hypothèse de « la troncation de la structure » initialement formulée par Rizzi (1992a) pour l'anglais selon laquelle, d'après le principe de la hiérarchie structurelle CP>IP>VP, un enfant qui omet un sujet dans un énoncé (le phénomène du « sujet nul enfantin », le sujet se trouvant en IP) omet alors également tous les éléments situés au-dessus du sujet dans la structure syntaxique, ce qui inclut la projection CP qui accueille les éléments *wh-* lorsqu'ils sont antéposés. Autrement dit, la structure est « tronquée » au niveau de la projection du sujet et cette troncation implique qu'il ne peut rien y avoir au-dessus de ce niveau de troncation. Crisma (1992) étudie donc les questions *wh-* formulées par Philippe en français et arrive à la conclusion que « la cooccurrence d'un élément *wh-* [antéposé] et d'un sujet nul est impossible »².

Dans l'optique de tester l'hypothèse de Rizzi et la conclusion de Crisma, nous avons comparé les questions de Philippe à des questions issues de notre premier corpus³. Il s'agit des phrases de Raphaël qui, sur un total de cinq cents phrases avait produit trente questions *wh-* dont quelques-unes du type *où* sont reprises en (4) ci-après.

¹ Crisma (1992:117).

² Crisma (1992:117) : “As the data clearly show, the co-occurrence of a *wh-* element and a null subject is impossible”.

³ Cf. Palasis (2005:112 et 210).

(4) Quelques questions *wh-* de Raphaël¹ :

- (a) Il est *où* ?
- (b) *Où* il est la toute ?
- (c) Elle est *où* ?
- (d) Je le mets *où* celui-là ?

Dans la grammaire générative², les éléments *wh-* sont des constituants qui sont sujets à déplacement. En effet, ceux-ci sont générés dans une position postverbale, comme le sont les adjoints dans les phrases non-interrogatives (par exemple : « il lit *dans la classe/où* »). Puis, ces éléments *wh-* peuvent être déplacés vers le début de la phrase, en spécifieur de CP (« *où* il lit ? »). L'inversion éventuelle du sujet et du verbe (« *où* lit-il ? ») correspond à un déplacement additionnel, celui du verbe fléchi qui va de I en C, s'intercalant ainsi entre l'élément interrogatif et le sujet. Les productions de Philippe et de Raphaël (qui ont à peu près le même âge) diffèrent de par leur structure et illustrent ainsi les deux options possibles. En effet, Philippe déplace systématiquement l'élément *wh-* de sa position initiale vers sa position antéposée, alors que Raphaël, sur un total de trente questions *wh-*, en pose seulement sept avec l'élément *wh-* déplacé (dont six avec *pourquoi*). *Où, quoi, comment* chez Raphaël restent majoritairement *in situ*. Tenter de falsifier l'hypothèse de certains chercheurs sur l'incompatibilité du déplacement des éléments *wh-* et du sujet nul enfantin en se basant sur les productions de ces deux enfants est donc impossible. Ces deux corpus en effet ne sont pas comparables dans la mesure où les deux enfants utilisent des stratégies différentes pour poser leurs questions (déplacement *vs* non déplacement des éléments *wh-*). On ne peut donc pas étudier une éventuelle incompatibilité entre les éléments *wh-* déplacés et le sujet nul chez Raphaël car ses trois questions avec un sujet nul comportent toutes un élément *wh-* resté *in situ*³. Ses productions n'infirment donc pas l'hypothèse de l'incompatibilité d'un élément *wh-* déplacé et d'un sujet nul mais elles ne la confirment pas non plus. La comparaison est donc vaine parce que plusieurs hypothèses sont possibles. En effet, dans la mesure où plus de trente ans séparent ces deux corpus, la variation diachronique peut être évoquée avec l'hypothèse d'une évolution dans la structure des questions dans le français oral (déplacement systématique de l'élément *wh-* dans la génération de Philippe *vs*

¹ Palasis (2005:210).

² Pour la démonstration qui suit, nous utilisons volontairement une terminologie simplifiée qui émane du modèle Gouvernement et Liage des années 1980. Elle suffit en effet pour illustrer ce propos.

³ Ses trois questions sont *est où cachée ?* et *est pourquoi ?* (prononcée deux fois).

l'élément *wh- in situ* pour la génération de Raphaël). La différence peut également être imputée à une variation diastratique et/ou diaphasique. Seule la piste de la variation géographique paraît peu vraisemblable dans la mesure où les enfants résident tous les deux en France métropolitaine et que la différence de structure ne peut pas raisonnablement être attribuée à une variation géographique entre le français parisien de Philippe et le français du sud de Raphaël. L'accumulation de variables parasites brouille donc les pistes et le fait que nous ne comparions que deux enfants ajoute à cette confusion car la représentativité dans les deux cas n'est pas garantie. Il faut donc isoler chacune de ces sources de variation afin de les annuler.

Par contre, dans d'autres cas de figure, la variation géographique apparaît également comme un facteur non négligeable. Il est largement établi par exemple que les langues parlées au Canada, en Belgique, en Suisse et en France, bien que portant le même nom, représentent en réalité plusieurs variétés de français dans la mesure où des variations sont attestées à tous les niveaux (phonologique, morphologique, syntaxique et lexical). Dans son étude sur le développement des catégories fonctionnelles, Ferdinand (1996), au moment de détailler les données à partir desquelles elle travaille (celles de Nathalie, Daniel, Grégoire et Philippe) établit bien une partition entre le « dialecte québécois » (Nathalie et Daniel résident à Montréal) et le « français européen » (qui, d'après l'auteur, est le français parlé par les parents de ces deux enfants). Par contre, l'auteur soutient que Nathalie et Daniel, « ayant un environnement européen », ne parlent pas ce « dialecte québécois »¹. Peut-on raisonnablement établir que ces enfants ne sont pas du tout influencés par la langue du pays dans lequel ils vivent ? Et peut-on parler de « français européen » ? Auger (1994) par exemple consacre une partie de sa thèse à la particule interrogative « -tu », typique du français québécois, comme dans l'exemple *elle part-tu* ?². Une étude de la distribution de l'inversion du clitique et du verbe dans les phrases interrogatives, menée par De Cat (2005) auprès d'adultes francophones, illustre également les disparités syntaxiques qui existent entre les différents pays. Les pourcentages repris dans le Tableau 15 ci-dessous pour différents types de tournures interrogatives (sans inversion, avec inversion, et avec utilisation de la locution « est-

¹ Ferdinand (1996:34) : "They [Nathalie and Daniel] are from families living in Montreal who have a European background, which means that they do not speak the Quebecois dialect, but European French". « Ils sont de familles qui résident à Montréal et qui viennent d'Europe, ce qui signifie qu'ils ne parlent pas le dialecte québécois mais le français de France ».

² Cf. Auger (1994:68).

ce que ») peuvent en effet être très différents entre la Belgique, le Canada et la France.

	Pas d'inversion	Inversion	Est-ce que	Total
Questions <i>wh-</i> (avec mouvement <i>wh-</i>)				
Belgique	8 % (59)	40 % (308)	53 % (408)	775
Canada	27 % (200)	1 % (4)	72 % (536)	740
France	25 % (229)	2 % (18)	73 % (656)	903
Questions <i>oui/non</i>				
Belgique	96 % (1900)	2 % (45)	2 % (38)	1983
Canada	61 % (1616)	21 % (567)	17 % (462)	2645
France	98 % (2133)	0 % (0)	2 % (45)	2178

Tableau 15 : distribution de différentes tournures interrogatives en Belgique, au Canada et en France (De Cat, 2005:1199)

Que penser alors d'une étude sur le français enfantin dont le corpus serait composé de données provenant à la fois du Canada, de France, de Belgique et éventuellement de Suisse ? Comment, dans ces conditions, dissocier les mécanismes propres à l'acquisition du langage de ceux en relation avec l'aire géographique ? Ces questions sont délicates et sans réponses toutes faites. C'est la raison pour laquelle nous ne souhaitons pas non plus travailler à partir des données du Corpus de York¹ car Léa, Max et Anne sont respectivement Belge, Canadien et Français.

5. La mise en place de notre cahier des charges

Ce tour d'horizon des différents corpus a ainsi permis de mettre en avant la notion de variable dans le cadre d'un recueil de données enfantines. Traditionnellement, cette notion fait partie intégrante de la méthodologie des sciences exactes (en biologie par exemple), et de certaines branches des sciences humaines (en psychologie expérimentale par exemple). Toute expérience ayant pour but l'analyse d'un phénomène, celle-ci doit permettre, dans un premier temps, d'identifier ce phénomène correctement. Il doit donc être isolé. Cependant, certaines variations au cours d'une expérience peuvent soit masquer complètement le phénomène étudié soit simplement brouiller les pistes. Ces variations non souhaitées et néfastes sont donc qualifiées de « non expérimentales » ou « parasites ». D'autres variations au contraire font partie intégrante de l'expérience et sont donc non seulement souhaitées mais nécessaires. L'étude de phénomènes dans le domaine de l'acquisition du langage, bien qu'utilisant des données, ne semble pas généralement

¹ Cf. les trois premiers enfants cités dans le Tableau 12, en page 151.

tenir compte de ces préceptes méthodologiques, peut-être parce que l'utilisation de ces données n'est pas considéré sous un aspect suffisamment expérimental. Conséquemment, certaines données souvent utilisées en acquisition présentent des variations multiples.

Deux principaux types de variations ont ainsi été identifiés dans ces corpus : des variations d'ordre méthodologique, liées à la collecte et à la transcription des données, et des variations d'ordre linguistique, en relation avec la personnalité des locuteurs et leur environnement. Toutes ces sources de variations ne sont cependant pas à mettre sur le même plan. Les variations d'ordre méthodologique sont à proscrire. Il faut ainsi éviter d'utiliser des données qui ont été collectées et transcrites dans des conditions disparates. Les variations dans ce domaine sont nombreuses (époque, pays, cadre théorique, objet d'étude, type d'enquête, moyens techniques, personnes en contact avec les enfants, protocole de recueil, système de transcription, etc.) ; cependant elles semblent assez aisément identifiables et donc annulables. Il convient ainsi d'utiliser des données qui proviennent d'une ou de plusieurs collectes qui présentent des caractéristiques méthodologiques homogènes. Les critères qui ne doivent pas comprendre de variabilité sont donc : la région géographique, l'époque, l'ensemble du protocole de collecte¹ ainsi que celui de la transcription². Cette première série de restrictions éradique du même coup les variations linguistiques d'ordre diachronique, géographique et diaphasique puisque toutes les données sont collectées au même moment, au même endroit, et par une même personne. D'autres variations méthodologiques, telles que l'objet d'étude et le cadre théorique de l'analyse, ne devraient pas avoir d'incidences négatives sur la qualité du corpus à la condition que les protocoles de recueil et de transcription soient suffisamment rigoureux et objectivés. Les corpus doivent en effet pouvoir être réutilisés en toute confiance par des chercheurs même lorsqu'ils travaillent dans des environnements différents.

A contrario, d'autres types bien précis de variations sont nécessaires. En effet, il apparaît excessivement aléatoire de prétendre formuler une généralisation à partir de données linguistiques provenant d'un groupe trop restreint de sujets ; soit que le(s) sujet(s) ai(en)t un façon très personnelle de s'exprimer (le cas du verbe *adorer* chez

¹ La collecte comprend de nombreuses étapes. Cf. Section II pour le détail.

² Cf. Section IV pour le détail.

Malcolm par exemple), dans ce cas il s'agit de variation individuelle ; soit que le(s) sujet(s) ne soi(en)t représentatif(s) que d'une certaine catégorie sociale (le cas de Philippe par exemple), dans ce cas il s'agit de variation diastratique. Ces deux types de variations doivent donc être intégrés dans les données, mais avec pour objectif d'annuler ensuite leurs effets parasites par l'application de filtres. Et la meilleure façon de les intégrer concomitamment est de construire un corpus à partir d'un nombre le plus élevé possible de sujets. Mais plusieurs questions surviennent. Combien d'enfants faut-il enregistrer ? Comment les sélectionner ? Ces interrogations nous ramènent alors à la thématique de la représentativité du corpus, abordée dans la Section 1 ci-dessus. Solliciter ses propres enfants et ceux de son entourage par exemple n'apporte aucune garantie de représentativité. La problématique se résume alors ainsi : quelle est la meilleure façon de composer un échantillon d'enfants suffisamment complet pour être représentatif de la population dans une région donnée à une époque donnée ? Une difficulté supplémentaire est introduite par l'aspect longitudinal de notre étude. En effet, il faut également que les enfants sélectionnés soient relativement facilement accessibles, dans des conditions similaires, et sur une période de temps assez longue¹. Nous avons néanmoins trouvé une réponse satisfaisante à notre interrogation. Nous avons ainsi considéré comme extrêmement intéressant de recourir à un groupe déjà constitué mais correspondant à la tranche d'âge souhaitée, et souvent perçu comme représentatif de la population en général : une classe entière d'enfants en école maternelle. Un tel groupe, en effet, combine, d'une part, l'homogénéité recherchée aux niveaux diachronique et géographique avec, d'autre part, la représentativité souhaitée, sous la forme d'une somme de personnalités individuelles et d'environnements sociaux relativement différents.

La première partie de ce chapitre (I. Réflexions épistémologiques et méthodologiques sur le corpus de données enfantines) a permis de développer une réflexion approfondie sur la base de données en acquisition du langage. Ces considérations ont notamment mené à la mise en place d'un cahier des charges précis qui, à son tour, a conduit à choisir comme échantillon représentatif pour notre étude l'ensemble des enfants d'une classe de maternelle. La seconde partie de ce chapitre (II. La collecte des données) retrace à présent notre cheminement, depuis ce moment

¹ Cette remarque entraîne une autre sur la question de la fréquence des séances d'enregistrement. Ce critère est abordé dans la partie suivante consacrée à la constitution de la base de données.

où nous avons choisi cette catégorie particulière d'informateurs, jusqu'au moment où nous avons effectivement mis en place puis réalisé notre collecte de données auprès d'un groupe de jeunes enfants dans une école maternelle.

II. LA COLLECTE DES DONNÉES

Dans de nombreux pays, l'inscription d'un enfant à l'école devient obligatoire lorsque l'enfant atteint l'année de ses six ans. En France, l'établissement qui accueille ces enfants est l'école élémentaire. Cependant, ce pays se distingue par l'existence d'une autre école qui prend en charge les enfants bien avant ce moment ; il s'agit bien sûr de l'école maternelle. Elle accueille en France la quasi-totalité des enfants de trois ans pour un cursus non obligatoire de trois ans composé de la Petite Section (trois-quatre ans), de la Moyenne Section (quatre-cinq ans) et de la Grande Section (cinq-six ans). Les objectifs sont les suivants :

« L'école maternelle a pour finalité d'aider chaque enfant, selon des démarches adaptées, à devenir autonome et à s'approprier des connaissances et des compétences afin de réussir au cours préparatoire les apprentissages fondamentaux. L'objectif essentiel de l'école maternelle est l'acquisition d'un langage oral riche, organisé et compréhensible par l'autre. À l'école maternelle, l'enfant établit des relations avec d'autres enfants et avec des adultes. Il exerce ses capacités motrices, sensorielles, affectives, relationnelles et intellectuelles ; il devient progressivement un élève. Il découvre l'univers de l'écrit. »¹

Cependant, dans certains établissements et sous certaines conditions de maturité physique et psychologique, les enfants peuvent également être scolarisés dès qu'ils atteignent l'âge de deux ans ; il s'agit alors de la Toute Petite Section (deux-trois ans). Enviée par beaucoup, décriée par d'autres², c'est cette section particulière qui nous intéresse en premier lieu³. Mais le parcours est encore complexe jusqu'à

¹ Bulletin Officiel de l'Éducation Nationale, Hors Série n° 3 du 19 juin 2008, consultable sur Internet à l'adresse http://www.education.gouv.fr/bo/2008/hs3/programme_maternelle.htm.

² Cf. par exemple le rapport de Bentolila (2007) commandé par le Ministre de l'Éducation nationale Xavier Darcos.

³ Le travail de collecte décrit dans cette thèse ne concerne que notre première campagne de collecte. Depuis, nous avons pu suivre et enregistrer ces mêmes enfants jusqu'à leur entrée à l'école élémentaire.

l'achèvement de la collecte en elle-même et les questions sont diverses. Nous décidons donc de mettre à profit notre précédente expérience en matière de collecte de données auprès de jeunes enfants¹ en commençant par définir, grâce à cette première expérience, les modalités pratiques que nous souhaitons mettre en place pour notre nouvelle collecte (cf. Section 1 ci-après). Ces modalités ainsi définies, nous nous tournons vers la deuxième étape qui est essentiellement composée de démarches administratives et informatives (cf. Section 2) qui ont pour but de nous autoriser à travailler avec des jeunes enfants au sein d'un établissement scolaire public. Il reste alors à organiser le déroulement des séances et à mettre en place un protocole de collecte (cf. Section 3).

1. Quelques considérations liminaires d'ordre pratique

Grâce à notre cahier des charges précis et à notre précédente expérience, nous avons donc en tête un certain nombre de détails pratiques concernant l'organisation de nos séances d'enregistrements. Il faut par exemple que les séances se déroulent impérativement le matin car l'après-midi de nombreux enfants de cet âge font la sieste et pratiquer une séance à leur réveil n'est pas indiqué². Nous projetons donc d'organiser nos séances le matin, aux heures habituelles de classe (entre 8h30 et 11h30). Le jour de la semaine a également son importance. Il semble plus favorable de cibler le lundi et/ou le jeudi car les enfants sont reposés et ils ont des souvenirs récents à raconter sur leurs occupations en dehors de l'école³.

Les activités en maternelle sont organisées sous forme d'ateliers, c'est-à-dire que la classe est répartie en petits groupes autour d'activités différentes (dessin, peinture, jeux de construction, jeux de société, etc.). Ces ateliers durent en général approximativement une demi-heure, ce qui correspond à la durée pendant laquelle on peut solliciter l'attention d'un enfant de deux-trois ans. Pour nos séances d'enregistrement, nous avons donc souhaité reproduire ce schéma de rotation afin de respecter à la fois le rythme de l'enfant et celui du reste de la classe, les enfants ne quittant/commençant ainsi pas un atelier déjà en cours dans la classe. Nous avons également souhaité enregistrer les enfants par petits groupes d'environ quatre

¹ Cf. Palasis (2005).

² Dans Palasis (2005), nous avons tenté une telle séance mais, après la sieste, il ne reste que peu de temps (au mieux une heure) et les enfants sont souvent encore endormis et très peu bavards.

³ Malheureusement, pour des raisons d'emploi du temps, nos premières séances n'ont pu avoir lieu que le vendredi.

enfants. Ce choix présente en effet plusieurs avantages. Il permet tout d'abord d'enregistrer un nombre d'enfants plus important que lors de séances individuelles. Ainsi, quatre rotations dans la matinée, comprenant chacune quatre ou cinq enfants, permettent d'enregistrer à peu près tous les enfants d'une même classe. Par ailleurs, cette organisation présente alors l'avantage d'inclure au moins deux types éventuels de situations discursives : des échanges entre l'adulte présent et les enfants ainsi que des échanges entre enfants. Le lieu dans lequel se pratiquent les enregistrements est également important. Lors de notre première expérience, nous avons enregistré les enfants tout au long de la matinée, c'est-à-dire aussi bien en classe (pendant les ateliers) que dans la cour (pendant leur récréation) et lors de leurs déplacements dans l'école (pour aller en gymnastique ou en bibliothèque). Ces essais s'étaient révélés totalement infructueux à cause du bruit environnant. Il est donc préférable de n'enregistrer les enfants qu'en intérieur, en excluant leurs déplacements, et, si possible, dans une pièce séparée du reste de la classe.

Notre cahier des charges et notre expérience nous permettent également d'être sûre du type d'enregistrements que nous voulons pratiquer. Un support audio bien sûr est indispensable en vue d'une transcription. Cependant, nous avons vu que c'est insuffisant pour plusieurs raisons. Comme mentionné précédemment, la vidéo permet par exemple un accès différé et répété aux données orales en situation. Elle est donc extrêmement utile à toutes les personnes qui souhaitent utiliser les données et qui ne sont pas présentes lors des enregistrements. Mais, dans un premier temps, nous avons également vu qu'elle est extrêmement précieuse dès l'étape de la transcription¹. Nous optons donc pour des enregistrements audio et vidéo concomitants.

À ce stade de la réflexion, un point important reste en suspens. Il s'agit de la question des intervalles entre les différentes séances d'enregistrements afin de suivre de façon efficace le développement syntaxique des jeunes enfants. Les intervalles des séances des enregistrements mentionnés en I.3.1 s'échelonnent entre une et sept semaines. Cependant, une majorité de chercheurs adopte le même intervalle de deux ou trois semaines, que nous nous proposons donc également de suivre. Une question corollaire concerne alors la régularité de ce rythme et l'éventuelle importance de cette régularité. Nous nous demandons par exemple si un calendrier irrégulier peut nuire à la qualité des données et conséquemment à la qualité des analyses, auquel cas

¹ Cf. IV.3.1 ci-après pour le détail de notre travail de transcription à partir de la vidéo.

un rythme irrégulier constituerait une variable parasite. En d'autres termes, sachant que l'acquisition n'est pas un processus linéaire et constant, est-il pertinent de pratiquer des enregistrements à des intervalles toujours identiques pour décrire au mieux son évolution ? Cette question, sans réponse précise pour le moment, en trouvera peut-être une dans le cours de notre analyse. Pour le moment, nous optons pour l'organisation d'un calendrier le plus régulier possible, à définir en accord avec l'école, sachant que de nombreux faits doivent être conciliés afin d'enregistrer de jeunes enfants régulièrement dans un environnement scolaire, notamment la prise en compte des activités déjà prévues par l'enseignant pour l'ensemble de la classe, les vacances scolaires et les absences relativement fréquentes chez des enfants de cette tranche d'âge.

Ces quelques considérations d'ordre pratique représentent les premiers jalons de la mise en place de notre collecte de données. En fait, cette définition précise des conditions dans lesquelles nous souhaitons enregistrer représente l'étape préliminaire à toute sollicitation d'enregistrement auprès de l'Éducation nationale afin que la demande soit sérieuse, concise et claire.

2. Les différentes démarches

Nous avons donc approché l'Inspection académique de notre département puis la Direction d'une école maternelle qui possède une Toute Petite Section. Grâce à leurs avis favorables, ainsi qu'à l'approbation de l'enseignante de la section concernée, il était alors possible de rencontrer, d'informer, et de solliciter l'autorisation écrite des parents des vingt-trois tout-petits de l'école maternelle approchée. Aucune des trois tâches que nous venons de mentionner n'est anodine. En effet, chacune fait surgir de nombreuses questions tant pratiques que juridiques. C'est pourquoi, afin de procéder à la collecte de nos données dans le respect des personnes enquêtées et dans le respect du cadre juridique français, nous avons notamment suivi les recommandations de Baude (2006) sur la question de l'information à apporter aux sujets enquêtés (ou le cas échéant à leurs représentants légaux, ce qui est notre cas) avant de procéder à quelque collecte que ce soit et sur la question plus générale des droits et des devoirs du chercheur qui fixe des productions verbales sur un support. Nous retrouvons ici la notion de « construction » du corpus, sous un angle légèrement différent :

« Un corpus oral, en effet, n'est pas une simple collection d'enregistrements de la parole humaine, c'est un objet « construit » : le traitement des données [...] permet non seulement de les conserver, mais les fait passer à un statut nouveau, matière de recherche et de valorisation. »¹

Il est donc indispensable que les personnes enquêtées soient informées en détail des tenants et des aboutissants d'un tel travail de collecte. Nous avons ainsi notamment travaillé sur deux notions cruciales : celle de consentement éclairé et celle d'anonymisation.

2.1 L'information aux familles

Avec l'aide précieuse de la directrice de l'école et de l'enseignante de la section, une réunion d'information a donc été mise en place le 29 septembre 2006 afin que les parents nous rencontrent et soient renseignés, de vive voix, sur notre projet de recherches dans tous ses détails.

« On parle souvent de formulaires d'autorisation à soumettre aux informateurs ; il est cependant important de faire dépendre cette autorisation de l'information préalable donnée aux personnes concernées : sans *information*, la *demande d'autorisation* n'a pas d'objet ni de sens. C'est pourquoi on parle de *consentement éclairé* (*informed consent*), dans le sens où l'acceptation de l'enregistrement est étroitement dépendante de la compréhension des finalités pour lesquelles il est effectué. »²

Un exposé oral a ainsi été présenté aux familles, exposé qui précisait, dans une première partie, les détails pratiques de notre activité, tels que le cadre de nos recherches, nos motivations, nos objectifs, ainsi que toutes les modalités pratiques de nos séances³. Le devenir de ces enregistrements a également été explicité : à court terme (avec la transcription, l'étiquetage morphosyntaxique et le codage des données), et à plus long terme (avec la possibilité d'utilisation des données sous leurs formes audio, vidéo et transcrite dans des communications scientifiques et le partage

¹ Baude (2006:11).

² Baude (2006:60).

³ Il a notamment été décidé, après concertation avec l'enseignante, que le rythme des enregistrements serait d'une séance toutes les deux ou trois semaines.

avec la communauté scientifique internationale¹, grâce notamment au support que représente le système CHILDES²). Nous avons également insisté sur le fait que rendre possible et participer activement à la constitution d'une telle base de données représente une contribution scientifique importante. En effet, comme le mentionne Baude (2006) :

« La France qui était en avance pour la mise au point des corpus de langue écrite (en particulier pour FRANTEXT qui est à la source du *Trésor de la Langue Française*), a pris un grand retard dans la constitution des corpus de langue parlée. »³

2.2 Le respect de la vie privée

Une autre partie importante de notre intervention au cours de cette réunion d'information auprès des parents a été consacrée à la question de l'anonymisation des données. En effet, cette problématique est également cruciale lors de la constitution d'une base de données car elle recouvre à la fois des aspects humains extrêmement importants (notamment le souci des familles de préserver leur enfant et leur vie privée) et des aspects juridiques incontournables. L'anonymisation est définie comme étant :

« [l'] opération par laquelle se trouve supprimé d'un ensemble de données recueillies auprès d'un individu ou d'un groupe tout lien permettant l'identification de ces derniers. »⁴

À cet effet, en ce qui concerne l'environnement scolaire tout d'abord, nous avons donc certifié aux parents que le nom, la localisation exacte de l'école maternelle où se déroule l'enquête, ainsi que les noms de famille du personnel enseignant ne seraient jamais mentionnés dans nos travaux. Il est donc prévu d'identifier la base de données comme provenant d'une enquête auprès d'enfants de la Toute Petite Section d'une école maternelle des Alpes-Maritimes dans le sud de la France pendant l'année scolaire 2006-2007. Ainsi, seuls les prénoms des membres du corps enseignant

¹ Cf. Baude (2006:36) : « La diffusion des résultats de la recherche fait partie des missions des chercheurs. » (cf. art. L 111-1 du code de la recherche et JO du 16.07.1982, p. 2273 *et sqq.*).

² L'adresse de ce site Internet a été communiquée aux parents, l'édition la plus récente des deux volumes publiés sur CHILDES ont circulé dans la salle de réunion et il a été précisé que seuls des chercheurs abonnés et ayant adhéré à une charte déontologique peuvent accéder aux données audio et vidéo.

³ Baude (2006:26).

⁴ Baude (2006:193).

(Claire, directrice, Sophie, enseignante et Christine, assistante), les dates des enregistrements et une localisation géographique approximative (puisque linguistiquement pertinente) seront divulgués tout au long de nos travaux. En ce qui concerne l'identité même des enfants, nous nous sommes engagée auprès de l'Inspection académique ainsi qu'auprès des parents à anonymiser les données recueillies. Néanmoins, ce processus complexe nécessite une réflexion appropriée à différents niveaux : humain, linguistique, juridique et purement technique. Dans la mesure où nous souhaitons ensuite partager notre corpus avec la communauté scientifique et que nous avons mis en avant l'importance que revêt la présence de données audio et vidéo concomitantes à la transcription écrite, la question de l'anonymisation de nos données concerne ces trois supports différents. En ce qui concerne les enregistrements audio et la piste son de la vidéo, nous étudions la faisabilité de « bipper » (ou tout au moins de rendre non identifiables) toutes les occurrences anonymisées dans les transcriptions¹. Pour ce qui est de la vidéo, nous avons sollicité l'autorisation des parents de diffuser les images sans flouter les visages des enfants. Par contre, ceux-ci portent des tabliers sur lesquels leurs prénoms sont parfois visibles. Un floutage de ces prénoms (ou toute autre manipulation adéquate) doit donc être réalisé.

2.3 Le formulaire d'autorisation

Au terme de notre réunion d'information, nous avons distribué à chaque parent un petit dossier comprenant nos coordonnées en cas de questions supplémentaires, ainsi qu'un récapitulatif des informations données au cours de la réunion (trois pages), et deux exemplaires d'un formulaire d'autorisation parentale (afin que les parents nous en retournent un et puissent garder l'autre pour mémoire)². Ce formulaire d'autorisation, pièce indispensable au démarrage de notre travail de collecte, comprend notamment l'état civil du représentant légal de l'enfant, les noms, prénom et date de naissance de l'enfant concerné, ainsi qu'une série d'alinéas qui reprend les points importants mentionnés pendant la réunion.

Le formulaire se termine par trois rubriques supplémentaires destinées à nous renseigner sur la fratrie de chaque enfant, les langues utilisées par les parents à leur domicile, et leur désir éventuel d'être informés tout au long de la durée du travail de

¹ Cf. IV.2 sur les modalités de l'anonymisation lors de la transcription.

² Cf. Annexes, pages 367-370.

l'avancement de la collecte et des recherches. Nous n'avons pas souhaité détailler davantage notre questionnaire (questions socioprofessionnelles par exemple) de peur que les familles ne perçoivent cette enquête comme trop intrusive et que ce sentiment ne provoque un refus de participation. Au terme de la réunion, les parents ont donc quitté l'école avec le dossier d'information et un délai de réflexion ; notre objectif étant de récupérer tous les formulaires (mentionnant soit l'autorisation, soit le refus) avant les vacances de la Toussaint (le 24 octobre 2006) afin de démarrer les séances le plus tôt possible après le retour de ces vacances (le 6 novembre).

3. La rencontre avec les enfants

Sur les vingt-trois familles sollicitées lors de notre réunion, vingt-deux nous ont autorisée à enregistrer leur enfant, et quinze d'entre elles nous ont communiqué leurs coordonnées afin de rester informées du déroulement du travail. Ce résultat constitue pour nous un premier retour très positif et il nous fait penser que le contact avec les parents a dû être clair et rassurant, comme nous le souhaitions.

3.1 L'identification de nos informateurs

Après dépouillement de ces formulaires, nous obtenons les caractéristiques détaillées dans le Tableau 16 ci-dessous pour nos vingt-deux petits informateurs. Ils sont classés du plus âgé au moins âgé et les données sont d'ores et déjà anonymisées grâce à l'utilisation de pseudonymes et de codes. La date de naissance de chacun nous permet d'indiquer sa tranche d'âge entre la date de la première séance d'enregistrements et la dernière¹. Les autres données correspondent aux réponses apportées par les parents à notre petit questionnaire. Ainsi, les cases vides du tableau correspondent à des rubriques qui n'ont pas été renseignées par certains parents et les langues (qui, dans le formulaire, avaient été demandées par ordre décroissant de fréquence) sont indiquées dans l'ordre fourni par les parents². Pour la fratrie, le premier chiffre indique le rang de l'enfant dans sa fratrie et le deuxième chiffre indique le nombre total d'enfants.

¹ L'âge des enfants à chaque séance est indiqué dans les fichiers de transcriptions.

² Dans le tableau, fr = français, an = anglais, it = italien, ru = russe, ar = arabe, als = alsacien, all = allemand, pt = portugais.

Pseudonyme	Code	Date de naissance	Âge première séance	Âge dernière séance	Langue(s) mère	Langue(s) père	Rang fratrie
Quentin	QUE	17/06/03	3;4.24	4;0.1	fr	fr	1/1
Laura	LAU	26/06/03	3;4.15	3;11.23		an	1/1
Léa	LEA	27/06/03	3;4.14	3;11.22			2/2
Eva	EVA	05/08/03	3;3.5	3;10.13	fr	fr	1/2
Hector	EKT	10/08/03	3;3.0	3;10.8	fr	fr	1/1
Clara	CLA	14/09/03	3;1.27	3;9.4	fr		2/2
Téo	TEO	09/10/03	3;1.1	3;8.9	fr	fr	2/2
Isia	ISI	05/11/03	3;0.5	3;7.13	fr	fr, it	2/2
Eloïse	ELO	18/11/03	2;11.23	3;7.0	ru	fr	1/1
Tom	TOM	01/12/03	2;11.9	3;6.17	fr	fr	2/2
Julien	JUL	02/12/03	2;11.8	3;6.16	fr	fr	2/2
Célia	CEL	04/12/03	2;11.6	3;6.14	ar, fr	ar, fr	2/2
Lucie	LUS	19/01/04	2;9.22	3;4.30	fr	fr	2/2
Noémie	NOE	26/01/04	2;9.15	3;4.23	fr		6/6
Antoine	ANT	28/01/04	2;9.13	3;4.21	fr, als	fr, all, an	2/2
Claire	KLE	02/02/04	2;9.8	3;4.16			
Romane	ROM	12/02/04	2;8.29	3;4.6			
Alan	ALA	23/02/04	2;8.18	3;3.26		it, an	2/2
Emma	EMA	10/03/04	2;8.0	3;3.8	fr	fr	1/1
Lola	LOL	27/03/04	2;7.14	3;2.22	fr, pt	fr, pt	2/2
Maxime	MAX	10/05/04	2;6.0	3;1.8	fr	fr	3/3
Zoé	ZOE	05/06/04	2;5.5	3;0.13	fr	fr	1/1

Tableau 16 : les informateurs de notre base de données 2006-2007

En ce qui concerne les langues pratiquées à la maison, le contact avec les enfants nous a permis de cerner plus précisément la place qu'occupe chacune d'entre elles dans la vie de chaque enfant¹. Ainsi, nous avons tout d'abord déterminé qu'aucun enfant n'était, au moins en partie, non francophone. Dans ce groupe, par exemple, il n'y a aucun enfant dont les deux parents sont exclusivement anglophones et qui parlent donc uniquement anglais à leur enfant². Nous avons donc supposé pour Laura par exemple que sa mère était francophone car sa maîtrise de la langue ne présente aucune information nous indiquant le contraire. Il en va de même pour les autres cas non renseignés dans le tableau ci-dessus. Néanmoins, nous avons décidé de différencier deux enfants du reste du groupe et, *in fine*, de ne pas les intégrer à notre analyse. Il s'agit de Lola (pour laquelle les deux parents déclarent parler français et

¹ Le multilinguisme (« connaissance multiple de langues chez un même individu », cf. Hagège (1996:11), que l'auteur distingue du plurilinguisme « coexistence d'une pluralité de langues dans un espace géographique ou politique donné ») est une thématique complexe que nous n'abordons pas dans ce travail. Nous nous limitons à cerner si un enfant peut raisonnablement être intégré dans une étude sur l'acquisition du français ou non.

² Ce qui avait été le cas dans Palasis (2005). Les données des enfants n'avaient alors pas été prises en compte car l'anglais était clairement leur langue dominante.

portugais) et Eloïse (dont la mère est russophone). Lola, en effet, parle très peu et produit beaucoup de *ola* ! (« bonjour » en portugais). Eloïse, quant à elle, s'exprime en français avec un très léger accent russe dans certains cas. Nous avons néanmoins transcrit leurs données mais en mentionnant ce qui nous semble représenter une forme de bilinguisme, par le biais d'un en-tête optionnel dans chaque fichier de transcription lorsque l'enfant participe à la séance, mais en gardant le français en première position¹ :

(5) Identification d'une forme de bilinguisme dans les fichiers de transcription :

- (a) Pour Lola : @Language of LOL : fr, pt
- (b) Pour Eloïse : @Language of ELO : fr, ru

Les autres enfants ne présentant aucune information contraire, nous avons décidé de considérer le français comme leur langue dominante. Cette information est inscrite au début de chaque fichier de transcription sous la forme d'un en-tête obligatoire qui indique la langue transcrite dans le fichier : @ Languages : fr.

3.2 La mise en place du calendrier des séances

Après avoir obtenu les autorisations, nous avons ensuite pu fixer un calendrier pour nos séances en collaboration avec la directrice et l'enseignante de la section. Treize séances ont ainsi été programmées sur l'année scolaire, entre novembre 2006 et juin 2007, à des intervalles de deux, trois ou quatre semaines, en fonction des activités des enfants et du rythme scolaire. Deux séances ont ainsi été planifiées pour novembre (le 10 et le 24) ; une seule pour décembre (le 15) car les enfants sont très pris par les préparatifs de la fête de fin d'année et ils sont assez agités et fatigués à cette période ; une en janvier (le 18) après les vacances ; deux en février (le 2 et le 19) et en mars (le 12 et le 26) ; une seule en avril (le 6) pour cause de vacances scolaires ; deux en mai (le 10 et le 24) et en juin (le 4 et le 18). Et bien que l'année scolaire se termine le 5 juillet, nous avons arrêté là notre calendrier car, par la suite, les enfants sont de nouveau très occupés par les préparatifs de fin d'année (scolaire, cette fois) et nous nous sommes réservé quelques jours de travail afin d'avoir le temps, nous aussi, d'organiser un petit goûter de fin d'année pour les enfants, leurs parents et le personnel de l'école, goûter au cours duquel une vidéo a été projetée, montrant une sélection de moments sympathiques et anecdotiques des séances

¹ Cf. Section III ci-après pour les autres détails concernant le formatage des fichiers.

d'enregistrements de l'année écoulée. Apportons à présent quelques précisions sur les moyens techniques utilisés pendant nos séances.

3.3 Les moyens techniques

Afin de collecter des données audio et vidéo de qualité, nous avons prévu d'utiliser deux appareils d'enregistrement de façon concomitante tout au long de nos séances. Le premier appareil est un caméscope¹, équipé de cassettes vidéo digitales (DV) d'une durée de soixante minutes, et posé sur un trépied à environ un mètre cinquante de notre table de travail et à hauteur des enfants, comme le montre l'Image 1 ci-dessous :



Image 1 : notre salle de travail

Étant seule avec les enfants lors des séances et ne pouvant donc pas manipuler la caméra tout en nous occupant des enfants, nous avons pratiqué l'enregistrement vidéo en continu, notre seule préoccupation technique étant alors de changer les cassettes. La vidéo est une source d'informations précieuse. D'une part, l'utilisation de cette caméra permet une prise de vue de l'ensemble des scènes ; les vidéos servent ainsi à déterminer quels sont les enfants présents, leurs activités, leurs tours de parole, etc., ce qui représente une foule de détails que l'on n'a pas toujours le temps de noter sur le fait. D'autre part, la caméra vidéo représente également une source son grâce à son micro stéréo intégré.

Le second appareil est un enregistreur vocal digital², équipé d'un micro de conférence stéréo³. Cet enregistreur de petite taille⁴ repose sur notre table de travail, le plus près possible des enfants à enregistrer. Il permet alors une prise du son plus

¹ Sony DCR-TRV22E.

² Olympus WS-310M.

³ Sony ECM-DS70P.

⁴ 9,5cm x 3,5cm x 1cm.

proche que celle effectuée grâce à la caméra et donc de meilleure qualité. Contrairement à la prise des images faite en continu, cette prise du son a été découpée en pistes en fonction de l'entrée et de la sortie des différents enfants. Chaque piste représente alors un fichier dans la mémoire de l'appareil, les données étant ainsi stockées sous forme de fichiers WMA¹.

3.4 Le protocole de collecte

Deux types de productions enfantines fondamentalement différents ont été décrits dans la partie I de ce chapitre : la production sollicitée et la production spontanée. Nous nous trouvons alors face à un paradoxe. Comment définir de façon cohérente notre protocole de collecte de données *spontanées* dans la mesure où nous avons été en contact avec de très jeunes enfants, pendant des laps de temps définis et que, afin de ne pas repartir bredouille, nous avons alors inévitablement employé une forme ou une autre de *sollicitation* ? En effet, la plupart des enfants avait entre deux ans et demi et trois ans lors de nos premières séances d'enregistrements et ils ne nous connaissaient pas. Or, nous savons que les enfants de cet âge peuvent se révéler spontanément très peu bavards. De plus, notre première expérience de collecte nous avait enseigné que, si l'on ne sollicite pas explicitement des échanges verbaux avec de jeunes enfants, leurs productions peuvent alors être extrêmement sporadiques. Nous avons donc dû concilier au mieux les caractéristiques du discours enfantin spontané et nos impératifs d'ordre pratique. Nous avons donc dû faire en sorte que, dans un premier temps, les enfants se sentent à l'aise afin que, dans un deuxième temps, ils soient motivés pour s'exprimer de façon fréquente et sereine.

De toute évidence, les premiers contacts avec les enfants représentent des moments importants. Nous avons donc choisi de nous mêler progressivement et discrètement à leur environnement scolaire. Notre première visite et une partie de notre deuxième matinée ont ainsi eu lieu dans leurs salles de travail habituelles (classe et bibliothèque) et non pas encore dans la salle attenante à la classe, prévue pour les enregistrements et qui, d'habitude, est leur dortoir de l'après-midi. Lors de notre première arrivée, l'enseignante nous a présentée aux enfants et nous nous sommes assise avec eux et avons participé à leurs différentes activités (dessin, lecture puis jeux de construction) afin que nous fassions connaissance et qu'ils s'habituent à la présence de notre matériel d'enregistrement. Ce premier contact établi, nous avons

¹ Windows Media Audio.

constaté que les enfants étaient à l'aise et que la relation avec eux était bonne. Nous avons alors décidé de procéder à nos séances en séparant un groupe de quatre ou cinq enfants du reste de la classe et de les enregistrer dans la salle prévue à cet effet. Cette organisation permet alors de collecter des données vidéo et audio de meilleure qualité grâce à un cadrage plus serré et à un environnement sonore plus calme. Dans la mesure du possible, nous avons alors essayé d'aligner nos changements sur les changements d'activités en classe mais, dans la pratique, la participation aux séances s'est faite sur la base du volontariat des enfants donc les entrées et les sorties ont pu être assez irrégulières et désorganisées. Au fil de nos séances, nous avons néanmoins constitué une fiche de suivi des enregistrements et des absences qui nous a permis de solliciter les enfants le plus régulièrement possible. Dans cette fiche, nous avons ainsi noté chaque participation (X) et chaque absence (abs) aux treize séances :

Pseudonymes	Séances												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Quentin			X	X	X			X	abs				
Laura	abs	abs		X	X	abs	X			X			abs
Léa		X	X						X				
Eva	X	X	X	X			X						X
Hector		X	X	X	X	X		X	X	X			
Clara	X	X			X		X				abs	abs	
Téo		X	X	abs			X	X		X	X	X	X
Isia	X	abs	abs	X	X		X					X	
Eloïse	X	X	abs	X	abs	abs			X	X	abs	X	X
Tom	X	X	X	X	X	X	X	X	X	X		X	X
Julien	X	abs		abs	abs		X			X	X	X	X
Célia	abs	X		X					X		X	X	
Lucie	X	X	X	X	X	X	X	X	X	X	X	X	X
Noémie	abs	X	abs	X	X	X	X	X	X	X	X	X	X
Antoine	X	abs	X	abs	X	X	X	X	X	X	X	X	X
Claire ¹	-	-	-	-	X	X	abs	X		abs	X	abs	abs
Romane	X	abs		X	X	X	X		X	X	X	X	
Alan	X	X	X	abs	X	X	X	X	X	X	X	X	X
Emma	X	abs	X	abs	X	X	X	X	X	X	X	X	X
Lola	X	X	X				X		X	X			X
Maxime	X	abs	X	X				X		X	abs	X	X
Zoé	X	X	X	X	X	X	X	X	abs		X		

Tableau 17 : suivi des enregistrements et des absences

Nous constatons alors des disparités importantes de suivi. Elles sont principalement dues à deux facteurs. Tout d'abord, l'attitude des enfants face à cette nouvelle

¹ Claire a rejoint la section en janvier 2007.

activité a été très variée et variable : certains enfants se sont tout de suite montrés très intéressés (Antoine, Alan, Lucie et Tom notamment), alors que d'autres n'ont montré qu'un intérêt limité ou ont mis plusieurs mois avant de se porter volontaires assidus (Julien notamment). Le second facteur réside dans nos choix personnels ; ils sont de deux ordres. Dans un premier temps, nous avons essayé de privilégier le suivi le plus régulier possible des enfants les plus jeunes. L'objectif est atteint pour Emma, Alan, Romane, Antoine, Noémie, Lucie et Tom qui ont tous moins de trois ans lors de la première séance. Pour d'autres enfants de cette tranche d'âge en revanche, le suivi est moins important : Zoé est souvent présente mais participe très peu, Maxime et Célia restent assez réservés pendant toute l'année ; Lola produit peu d'énoncés ; et Claire n'arrive dans cette classe qu'en janvier 2007 puis est souvent absente. Dans un deuxième temps, nous avons pratiqué un second choix : celui de ne pas laisser certains enfants ensemble au cours d'une même séance pour des raisons de « chahut ». Nous avons donc essayé de constituer des groupes différents d'une séance à une autre afin de provoquer des situations nouvelles que nous espérions intéressantes mais l'expérience nous a montré qu'il fallait éviter certaines « combinaisons » qui se révélaient stériles à cause du bruit, des déplacements excessifs ou des disputes.

En début de section, nous nous sommes interrogée sur le paradoxe éventuel de la collecte *organisée* de données *spontanées*. Cependant, nous qualifions bien nos données de « spontanées » et non d'« induites » car ce que nous sollicitons est la performance linguistique de l'enfant en général et non pas une ou plusieurs structures linguistiques en particulier. D'autre part, les données sont recueillies dans un environnement relativement familier à l'enfant. Finalement, pour rester dans ce cadre de données spontanées, il ne reste donc qu'à proposer aux enfants des situations ou des activités qui *a priori* leur sont déjà familières. Nous avons ainsi choisi trois types différents d'activités afin de provoquer des situations et des interactions variées : l'entretien, la « lecture » et le jeu. À ces trois genres majoritaires, s'ajoute un exercice un petit peu plus particulier, qui au fil de nos séances est devenu notre rituel ; il s'agit du « bonjour à la caméra ». En effet, afin que le caméscope ne les intimide pas et afin qu'ils sachent exactement à quoi il sert, les enfants sont invités à s'en approcher en début de séance (puis dans le cours de la séance pour ceux qui le souhaitent). Ils peuvent ainsi se voir en gros plan grâce à l'écran de contrôle qui est tourné vers eux (*cf.* Image 2 ci-après).

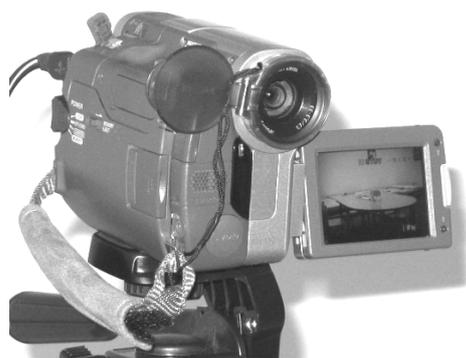


Image 2 : le caméscope et son écran de contrôle

Les enfants disent ainsi souvent bonjour, font des grimaces, parfois s'installent avec des livres, etc. L'utilisation de la caméra fait alors partie intégrante de nos séances et les enfants la considèrent comme un objet ludique. La même démarche a été appliquée avec l'enregistreur audio. Celui-ci est posé sur la table et il est expliqué aux enfants que c'est un appareil qui enregistre la voix et qu'ils peuvent parler et chanter, ce qu'ils ont fait¹.

L'entretien est un mode de collecte qui correspond particulièrement bien à notre type de données puisqu'il est défini comme suit :

« L'entretien est composé de questions ouvertes, l'objectif étant principalement de recueillir une quantité importante de données linguistiques. »²

Ce mode de sollicitation est donc largement utilisé dans un « souci de collecter les productions les plus naturelles possibles »³. Nous avons questionné les enfants principalement sur leurs activités du week-end ou du mercredi précédant chaque séance⁴. Les enfants sont alors plus ou moins bavards en fonction de leur humeur, de la vivacité de leurs souvenirs et de l'intérêt qu'ils y portent.

La « lecture » est notre deuxième type d'activité. Le terme est utilisé entre guillemets pour plusieurs raisons. La première raison évidente est que les enfants de cet âge ne lisent pas encore le texte des livres. D'autre part, nous utilisons ce terme de façon très générique dans la mesure où il inclut l'utilisation de trois types de

¹ Cf. par exemple la séance X.40e du 10 mai 2007 pendant laquelle Julien et Téo prennent le micro et chantent « Petit Papa Noël ».

² Baude (2006:48).

³ Baude (2006:48).

⁴ Cf. par exemple la séance XIII.46a avec Lucie, Eva, Emma et Lola.

supports : des livres pour enfants, des prospectus et un album de photographies. Les livres sont soit des ouvrages qui se trouvent habituellement dans la classe et que les enfants ont parfois déjà parcourus avec l'enseignante, soit des livres de notre bibliothèque personnelle. Nous « lisons » les histoires, c'est-à-dire que les enfants racontent ce qu'ils voient dans le livre ou me demandent parfois de leur lire des extraits ou de leur expliquer les événements¹. Les prospectus sont ceux que l'on peut trouver dans les boîtes aux lettres. Ils avaient suscité beaucoup d'enthousiasme auprès des enfants lors notre première collecte en 2003-2004, notamment dans la période qui avait précédé Noël. Par contre, les enfants de ce groupe n'ont pas montré un intérêt particulier pour ce genre de support² ; il n'a donc pas été souvent utilisé lors cette collecte. L'album, quant à lui, est un recueil de photographies composé par l'enseignante, dans lequel chaque enfant a inséré une photographie le représentant en compagnie de ses parents, de sa fratrie, de ses grands-parents, avec ses animaux de compagnie, etc. Les enfants pouvaient ainsi feuilleter cet album et décrire ce qu'ils voyaient sur leur photographie et sur celles de leurs camarades³.

La partie jeu de nos séances est essentiellement composée de l'utilisation de deux jeux de société qui ont eu un énorme succès auprès des enfants tout au long de l'année : nous les avons appelés « le jeu des animaux » et « le jeu des arbres ». « Le jeu des animaux » est un jeu de loto que nous avons fait découvrir aux enfants à partir de notre troisième séance. Chaque enfant doit compléter une planche avec six cartes représentant des animaux familiers qu'il pioche au fur et à mesure⁴. Le « jeu des arbres » est bâti sur le même principe ; il s'agit pour chaque joueur de remplir le feuillage de son arbre avec des jetons colorés, choisis en fonction de la couleur donnée par un dé. Alan et Antoine notamment ont été de grands fans de cette activité⁵. L'utilisation de ces jeux de société lors de nos séances a plusieurs buts. Dans un premier temps, leur utilisation permet de provoquer une situation qui impose un tour de rôle. Ainsi, alternativement et successivement, chaque enfant du groupe devient l'acteur principal de la scène, ce qui nous permet alors d'enregistrer des productions dans des conditions sonores plus calmes que lorsque tous les enfants s'expriment en même temps. Dans un second temps, cette activité nous permet de

¹ Cf. par exemple I.2 qui comprend notre première séance de lecture notamment avec Tom.

² Cf. la séance III.10b avec Téo.

³ Cf. par exemple II.5 qui est notre première séance de « lecture » avec cet album.

⁴ Cf. par exemple la séance III.10b.

⁵ Cf. V.20 pour notre première séance avec ce jeu.

solliciter tous les enfants à tour de rôle mais d'une façon similaire. En général, il leur est ainsi demandé d'explicitier ce qu'ils doivent faire pour jouer ou d'expliquer à leurs camarades ce que, eux, doivent faire à tel ou tel moment du jeu (*cf.* l'exemple (6) ci-dessous).

(6) Exemple de sollicitation dans les activités de jeu¹ :

KAT: qu+est+c(e)+qu' i(l) faut faire ?

NOE: fermer les yeux !

KAT: très bien et après qu+est+c(e)+qu' i(l) faut faire ?

NOE: prendre [*] les cartes.

%err: prendre = prendre \$MOR \$SUB \$REG \$INF

KAT: très bien et après ?

NOE: après <c' est tout> [/] après on ferme les yeux.

KAT: okay.

KAT: alors Noémie puisque tu sais bien comment ça marche tu vas montrer aux copines.

KAT: vas+y montre nous comment tu fais ?

KAT: qu+est+c(e)+que tu fais ?

NOE: 0.

%com: takes a card from the box.

Nous avons également parfois joué à des jeux de dominos mais ce type d'activité a rencontré nettement moins de succès auprès des enfants² et nous avons aussi réalisé des puzzles de façon très occasionnelle³.

3.5 Les résultats de la collecte

Au cours des treize matinées passées en compagnie de nos vingt-deux petits informateurs, nous avons donc enregistré nos rencontres sur deux supports différents : des cassettes vidéo de type DV et des fichiers audio de format WMA. Nous avons ainsi collecté près de vingt heures d'enregistrements vidéo, doublées de vingt heures d'enregistrements audio. La phase de collecte – première étape de la constitution d'une base de données – s'achève donc et nous avons maintenant en main des données primaires, relativement nombreuses, mais qui, en l'état, ne sont

¹ Séance XIII.32b ; NOE = Noémie, 3;2.0.

² *Cf.* par exemple la séance XIII.46c avec Lucie, Eva, Emma et Lola.

³ *Cf.* par exemple la séance XI.41d avec Célia, Zoé et Emma.

pas encore exploitables. Nous devons donc transformer ce matériau brut en un outil de travail qui soit le plus performant possible. Notre carnet de bord, cahier dans lequel toutes les informations relatives à la date des séances, l'absence des enfants, leurs entrées et sorties pendant les enregistrements, leurs âges, les activités pratiquées, etc. ont été consignées au fur et à mesure, et qui nous a accompagné tout au long de nos enregistrements, représente alors un allié très utile.

III. VERS LA BASE DE DONNÉES

Comme l'indique Baude (2006), une base de données se définit comme suit :

« On entend par base de données un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique et individuellement accessibles par des moyens électroniques ou par tout autre moyen. »¹

Un tel outil de travail est donc composé de deux parties : un contenu (les données) et un contenant (les moyens électroniques). Nous disposons pour le moment de données ; il reste alors à choisir la plateforme par l'intermédiaire de laquelle nous souhaitons les rendre accessibles.

1. Le choix de la plateforme

Ce choix est important car il a des répercussions tant en amont qu'en aval de la diffusion des données dans la mesure où les différentes plateformes possèdent généralement chacune un fonctionnement propre. Elles utilisent notamment des logiciels et des conventions de transcription et de codage différents et elles s'adressent à des publics distincts. Un inventaire des corpus francophones réalisé en octobre 2006 par Paul Cappeau et Magali Sejjido dans le cadre d'un projet de la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF) fait état de cette diversité². La base de données THESOC³ par exemple s'adresse ainsi

¹ Alinéa 2 de l'article L112-3 du Code de Propriété Intellectuelle (CPI), mentionné par Baude (2006:119).

² Cet inventaire est consultable à <http://crdo.fr/docs/inventaire-fr>.

³ Thesaurus Occitan du laboratoire Bases, Corpus, Langage (UMR 6039) à Nice-Sophia Antipolis, consultable à <http://thesaurus.unice.fr>.

plutôt à des dialectologues occitanistes, alors que CorpAix¹ cible l'étude morphosyntaxique du français. Parmi les bases de données enfantines, de telles différences existent aussi. Les plateformes CHILDES (*Child Language Data Exchange System*) et CLAPI (Corpus de Langue Parlée en Interaction) mentionnées précédemment illustrent d'ailleurs cette variété : CHILDES, comme son acronyme l'indique, est une plateforme entièrement dédiée aux données enfantines alors que CLAPI contient non seulement des productions enfantines mais également des interactions entre adultes dans des contextes variés ; CHILDES utilise un logiciel de transcription baptisé CHAT (*Codes for the Human Analysis of Transcripts*) alors que CLAPI utilise une convention nommée ICOR (Interaction Corpus²), etc. Nous devons donc définir notre place dans ce paysage afin de déterminer la plateforme dont nous décidons d'adopter le format.

Nos données ont été collectées dans le but d'étudier le développement syntaxique infantin ; elles relèvent donc clairement du domaine de l'acquisition. Néanmoins, du fait de notre présence active lors des séances, nos enregistrements constituent également un corpus d'interactions verbales adulte-enfants. Le type de données seul ne permet donc pas de trancher. En revanche, notre domaine d'application – l'étude de la morphosyntaxe dans le cadre de la grammaire générative – fait pencher la balance en faveur de CHILDES, pour deux raisons. La première concerne les logiciels proposés et leur accessibilité. En effet, le logiciel CHAT³ est spécialement conçu pour la transcription du langage infantin ainsi que pour son codage sur plusieurs niveaux (dont un niveau morphosyntaxique). Un second logiciel – CLAN (*Computerized Language Analysis*)⁴ – permet ensuite d'effectuer de nombreuses analyses à partir de ces fichiers CHAT. De plus, ces logiciels, ainsi que leurs manuels, sont facilement accessibles à tout moment par le biais d'Internet⁵. La seconde raison qui nous fait adopter la plateforme CHILDES est que la majorité de la littérature publiée dans notre domaine de recherches utilise des données provenant de cette base. Nous pensons donc que ce type de transcription est en adéquation avec nos travaux et que nos données y auraient éventuellement une

¹ Corpus du Groupe Aixoise de Recherches en Syntaxe (GARS) de l'Université de Provence, maintenant intégré aux données de l'équipe DELIC (DEscription Linguistique sur Corpus). Cf. <http://www.up.univ-mrs.fr/delic/>.

² Cf. http://icar.univ-lyon2.fr/documents/ICAR_Conventions_ICOR_2007.doc.

³ Sur lequel nous revenons en III.5.2 ci-après.

⁴ Les modalités de CLAN sont détaillées au début du Chapitre IV.

⁵ Cf. <http://childes.psy.cmu.edu/>.

place en vue de compléter les corpus francophones déjà présents sur ce site. Il convient alors de se former à l'utilisation du logiciel CHAT afin d'aborder l'étape de la transcription.

2. La mise en place des fichiers

À ce stade, nous avons donc nos deux éléments en main : d'une part, des données primaires et, d'autre part, un logiciel. Il reste alors à opérer la fusion entre les deux.

2.1 La capture des données brutes sur informatique

Dans un premier temps, les données vidéo et audio recueillies sont copiées et sauvegardées sur un support informatique. Cette opération permet également de convertir les deux formats d'origine afin de les rendre compatibles avec le logiciel CHAT. Les plages vidéo sont ainsi converties au format Mpeg1¹ et les pistes audio, originellement enregistrées en WMA, sont converties en MP3². Puis nous subdivisons les fichiers vidéo d'une durée initiale de soixante minutes afin d'obtenir des plages de durées plus ou moins équivalentes ou bien qui forment un tout par rapport à la présence d'un groupe précis d'enfants. Cette subdivision permet alors d'aligner la durée des fichiers vidéo sur celle des fichiers audio. Ces différents fichiers sont alors tous nommés suivant le même modèle qui comporte :

- Un chiffre romain qui indique le numéro de la séance (de I à XIII).
- Un chiffre arabe qui correspond au numéro de la piste audio (de 1 à 51).
- Une lettre en minuscule : lorsqu'une piste audio dure plus de quinze minutes, celle-ci est subdivisée en plusieurs sections (a, b, c, etc.) afin que la taille des fichiers vidéo et la longueur des fichiers de transcription ne soient pas trop importantes. Par exemple, la piste 3 de notre première séance (I.3) est ainsi divisée en quatre parties dans la mesure où celle-ci dure quarante-deux minutes. Nous obtenons ainsi les fichiers I.3a, I.3b, I.3c et I.3d.

Les données vidéo et audio sont à présent prêtes à être visionnées sur informatique par le biais des commandes appropriées dans le logiciel CHAT. À terme, cette compatibilité permettra également d'aligner ces fichiers sur les transcriptions grâce à

¹ MPEG (*Moving Picture Experts Group*) : Nom d'une famille de standards utilisés pour l'encodage d'informations audio et vidéo dans un format numérique compressé.

² MP3 (abréviation de *Mpeg-1 Audio Layer 3*) est un format de compression audio numérique.

l'insertion dans la transcription de balises actives qui permettront de lire une transcription sélectionnée tout en visionnant et/ou en écoutant l'enregistrement correspondant¹.

2.2 La création des fichiers CHAT

Avec pour base les références des fichiers vidéo et audio (par exemple, I.4), les cent vingt et un fichiers informatiques correspondants sont créés dans le logiciel de transcription. Tous ces fichiers CHAT doivent présenter un format identique et doivent comporter un certain nombre d'en-têtes, signalés par une arobase, comme dans l'exemple (7) ci-après. Certains de ces en-têtes sont obligatoires ; ils indiquent notamment le début et la fin des différents fichiers, la langue transcrite, ainsi que la liste des participants et quelques caractéristiques pertinentes telles que leurs âges. D'autres en-têtes sont facultatifs et sont donc librement sélectionnés par le responsable de la base de données. Nous avons choisi d'intégrer les en-têtes suivants à tous nos fichiers CHAT :

- le code du transcripteur,
- la date de la séance,
- la localisation géographique,
- un bref résumé des activités de la séance,
- un éventuel commentaire,
- la description de la salle,
- les références de la piste audio et de la cassette vidéo originales,
- la durée de la séance.

Tous les fichiers ont donc le même format avec les mêmes en-têtes présentés dans le même ordre, comme le montre l'exemple (7) ci-dessous. Cette organisation permet ainsi un accès facilité aux informations concernant chaque fichier.

¹ Mais, comme mentionné précédemment, certains problèmes d'anonymisation doivent être réglés auparavant.

(7) Exemple d'en-têtes de format CHAT (fichier VIII.32b) :

```
@Begin
@Languages: fr
@Participants: ZOE Zoé Target_Child1, NOE Noémie Target_Child, KLE
Claire Target_Child, ANT Antoine Target_Child, KAT Katérina Investigator
@ID: fr|palasis|ZOE|2;09.21|female|||Target_Child||2
@ID: fr|palasis|NOE|3;02.00|female|||Target_Child||
@ID: fr|palasis|KLE|3;01.24|female|||Target_Child||
@ID: fr|palasis|ANT|3;01.29|male|||Target_Child||
@ID: fr|palasis|KAT|||||Investigator||
@Coder: KAT
@Date: 26-MAR-2007
@Location: South of France
@Activities: Playing the card game.
@Comment: This file is the second of a series of two (32a and 32b) for the same
session.
@Room Layout: Room next to the classroom where the children have their
afternoon nap. A table and chairs are added for our sessions.
@Tape Location: A32 (audio) and 8 (video)
@Time Duration: 00:08:00
@End
```

À cette série d'en-têtes réguliers s'ajoutent les deux en-têtes ponctuels qui correspondent à la mention du bilinguisme de Lola et d'Éloïse lorsque ces enfants participent aux séances³. Les fichiers, ainsi formatés, sont prêts à recevoir les transcriptions.

Le logiciel CHAT est un programme qui gère plusieurs niveaux d'écritures (*tiers*). Le niveau principal est la ligne sur laquelle se trouvent les transcriptions et elle commence toujours par un astérisque suivi du code du locuteur en majuscules

¹ Tous les participants sont codés comme suit : un code de trois lettres, leur prénom, leur rôle lors de la séance (les rôles que nous utilisons sont : *target child* « enfant informateur », *investigator* « chercheur », *teacher* « enseignante » et *adult* « adulte »).

² La présence de plusieurs barres verticales successives indique que tous les champs disponibles n'ont pas été renseignés.

³ La description de ces en-têtes se trouve en page 180.

(par exemple : *KAT)¹. Ce niveau principal peut ensuite être relié à un ou plusieurs autres niveaux secondaires, dépendants de cette ligne de transcription. Ces niveaux sont reconnaissables au signe initial % suivi d'un code de trois lettres minuscules. Comme pour les en-têtes des fichiers, la liste de ces niveaux dépendants est vaste² et leur choix est laissé au chercheur afin qu'il adapte au mieux l'outil informatique à son domaine de recherche. La constitution de notre base de données passe donc à présent par deux étapes : la transcription de ce niveau principal (détaillée dans la partie IV ci-après) et la création des niveaux dépendants pertinents pour notre étude (détaillée en V).

IV. LA TRANSCRIPTION DES DONNÉES

Les linguistes de terrain sont particulièrement conscients de la difficulté que représente la transcription d'une production orale en une occurrence écrite. En effet, convertir un continuum phonique, quel qu'il soit, en signes graphiques ne va pas de soi, ne serait-ce que parce que l'oral et l'écrit constituent de fait deux types différents d'expression. Les outils de l'un (les sons, les pauses et l'intonation notamment) ne sont pas ceux de l'autre (les graphèmes et la ponctuation). Rendre compte du premier par le biais des outils du second semble donc de toute façon représenter un défi en soi. Par ailleurs, les chercheurs en acquisition savent bien que les énoncés des jeunes enfants représentent des productions orales particulièrement difficiles à transcrire. Ces dernières sont en effet souvent brèves, faiblement articulées et, de plus, présentent régulièrement des occurrences que l'on ne rencontre pas dans la langue cible³. Au mieux, un chercheur peut espérer pallier cette brièveté et cette faiblesse articulatoire par du matériel d'enregistrement sophistiqué, espérant ainsi fixer les moindres gestes articulatoires des enfants. Pour le reste, le transcripteur doit pouvoir rendre compte non seulement de l'oralité en général mais également des idiosyncrasies enfantines. L'outil utilisé se doit donc d'être extrêmement performant.

Comme mentionné précédemment, nous avons choisi le système de transcription de la plateforme CHILDES, baptisé CHAT (*Codes for the Human Analysis of Transcripts*). Ces codes représentent un système extrêmement large et divers dans la

¹ Cependant, lorsque nous utilisons des exemples issus de nos transcriptions, nous enlevons l'astérisque initial car sa présence pourrait présenter une ambiguïté dans la mesure où, en linguistique, ce symbole indique généralement une agrammaticalité.

² MacWhinney (2000a:195) propose vingt-six possibilités.

³ Cf. par exemple certains « sujets » de Léonard, mentionnés au Chapitre I.

mesure où ils ont été continuellement enrichis depuis leur création en 1984 afin de pouvoir être utilisés dans des domaines aussi différents que l'acquisition considérée comme typique, l'acquisition présentant une pathologie, l'alternance codique, l'analyse conversationnelle, la langue des signes, etc. Chaque responsable d'une future base de données est donc tenu, dans un premier temps, de décider quels signes sont pertinents d'après son cadre de recherches.

1. La segmentation en énoncés

À l'intérieur de chaque fichier CHAT, à la suite des en-têtes détaillés précédemment, sont insérées les lignes de transcription qui représentent les lignes dites principales. L'instruction suivante fait partie des toute premières recommandations données dans le manuel CHAT :

“Each main line should code one and only one utterance. When a speaker produces several utterances in a row, code each with a new main line.”¹

Chaque ligne de transcription correspond donc à un énoncé et un seul, débute par le code du locuteur précédé d'un astérisque et suivi de deux points (par exemple pour l'enfant Quentin : *QUE:). La majuscule ne doit pas être utilisée en début d'énoncé (sauf pour les noms propres) et chaque ligne se termine par un signe de ponctuation à choisir dans une liste assez vaste de possibilités². Nous limitons ce choix aux codes repris dans le Tableau 18 ci-après :

Codes	Significations	
Basiques	.	assertion
	?	question
	!	exclamation
Spéciaux	+!?	question avec exclamation
	+...	assertion incomplète, mais non interrompue
	+...?	question incomplète, mais non interrompue
	+/.	assertion interrompue par un interlocuteur
	+/?	question interrompue par un interlocuteur
	+//.	assertion interrompue par le locuteur
	+//?	question interrompue par le locuteur

Tableau 18 : codes de fin d'énoncé

¹ Cf. MacWhinney (2000a:18) : « Chaque ligne principale devrait coder un et un seul énoncé. Lorsqu'un locuteur produit plusieurs énoncés à la suite, il faut coder chacun des énoncés sur une nouvelle ligne principale. »

² Cf. MacWhinney (2000a:60-69) pour la liste complète.

De plus, les énoncés incomplets (et donc terminés par un code « spécial ») peuvent être suivis par un énoncé débutant par le code « +, » qui indique que cet énoncé complète un énoncé précédent du même locuteur. L'éventail des codes proposé par le logiciel est donc large et semble pouvoir rendre compte des nombreuses possibilités de délimitations lors d'échanges verbaux, comme le montrent les exemples repris en (8) ci-dessous.

(8) Exemples de lignes de transcription dans CHAT¹ :

- (a) KAT: comment tu t' appelles ?
ANT: Antoine.
KAT: et toi tu t' appelles Alan !
- (b) ANT: là [?] je fais un escargot +...
ANT: +, un escargot +...

1.1 Définir la notion d'« énoncé »

Cependant, dans un nombre non négligeable de cas, avant même de pouvoir décider du signe de ponctuation à utiliser à la fin de chaque énoncé, nous nous heurtons à une autre question, beaucoup moins subtile mais néanmoins capitale : à quels critères pouvons-nous nous fier pour segmenter de façon constante le continuum phonique de l'oral en *énoncés* ? Nous trouvons ainsi que l'instruction du manuel, mentionnée ci-dessus, d'apparence pourtant anodine, est en réalité très difficile à mettre en pratique face à certaines productions orales, que celles-ci soient enfantines ou adultes. Nous commençons donc par reprendre la définition du terme « énoncé » :

« Le mot *énoncé* désigne toute suite finie de mots d'une langue émise par un ou plusieurs locuteurs. La clôture de l'énoncé est assurée par une période de silence avant et après la suite de mots, silences réalisés par les sujets parlants. »²

Cependant, alors que les codes graphiques délimitent les productions écrites de façon univoque rendant leur lecture généralement non problématique, les contours d'un énoncé oral se révèlent nettement plus délicats à cerner et donc à transcrire de façon

¹ Séance I.1a. Nous rappelons que nous ne reprenons pas les astérisques dans le corps de la thèse afin de ne pas provoquer d'ambiguïté quant à la grammaticalité des énoncés.

² Dubois *et al.* (1994:180).

univoque. Bien sûr, il est généralement considéré que l'oral comporte des indicateurs tels que le « silence » mentionné dans la définition ci-dessus ou l'intonation qui est supposée baisser en fin d'énoncé déclaratif et monter en fin d'énoncé interrogatif. Néanmoins, l'expérience montre que ces indices ne sont pas si faciles à déceler dans une production orale. En effet, l'intonation ne suit pas toujours ces schémas canoniques et la rapidité avec laquelle les éléments d'une production s'enchaînent prive souvent le transcripteur de repères identifiables et *constants* qui lui assureraient une transcription homogène, la notion de « silence » notamment étant très aléatoire.

1.2 De la difficulté à transcrire les pauses

De plus, le discours oral quotidien développe des caractéristiques qui ajoutent à cette difficulté. L'une de ces principales caractéristiques est la structure syntaxique de ce type de discours. En effet, ce dernier est largement construit par parataxe, c'est-à-dire par simple juxtaposition des propositions, s'opposant ainsi à une construction par hypotaxe qui utilise les mécanismes de la coordination et/ou de la subordination.

1.2.1 La parataxe

La difficulté entraînée par la présence de la juxtaposition réside alors dans le fait que, lorsqu'il y a parataxe, il n'y a pas de dépendance syntaxique entre les propositions ; il y a simplement une « pause ». À l'écrit, la pause dans un discours déclaratif peut être transcrite de trois manières différentes : la ponctuation dispose en effet du point, du point-virgule et de la virgule, chaque signe correspondant à une pause de moins en moins marquée :

(9) La juxtaposition en français standard :

- (a) Mon père vient me chercher. Ma mère vient me chercher.
- (b) Mon père vient me chercher ; ma mère vient me chercher.
- (c) Mon père vient me chercher, ma mère vient me chercher.

CHAT, en plus des codes de fin d'énoncé mentionnés dans le Tableau 18 ci-dessus, possède des codes qui permettent de marquer les pauses à l'intérieur des énoncés. De nouveau, l'éventail de choix est assez large, comme l'indique le Tableau 19 ci-dessous :

Codes	Significations
,	pause légère
„	question tags
#	pauses de plus en plus marquées
##	
###	
#d	pause non attendue
^	pause entre deux syllabes d'un même mot

Tableau 19 : pauses à l'intérieur d'un énoncé

Tout d'abord, il nous semble qu'il serait extrêmement difficile pour un transcrip-teur d'utiliser tous les symboles ci-dessus de façon constante au sein d'une transcription. En effet, même en laissant de côté les signes symbolisant les question tags, les pauses non attendues et les pauses à l'intérieur d'un mot, il reste tout de même quatre longueurs différentes de pauses et nous supposons qu'une oreille humaine, même exercée, aurait beaucoup de peine à attribuer ces quatre codes de façon constante et univoque lors d'une transcription. Admettons donc, pour le moment, que nous soyons en mesure de distinguer deux types de pauses à l'intérieur d'un énoncé et que nous conservions ainsi deux de ces symboles (, et #) et appliquons l'instruction du manuel CHAT sur la segmentation des énoncés aux différents types de juxtaposition présentés en (9). Nous obtenons les trois transcriptions proposées en (10) ci-dessous :

(10) Transcriptions correspondantes :

- (a) KAT: mon père vient me chercher.
KAT: ma mère vient me chercher.
- (b) KAT: mon père vient me chercher # ma mère vient me chercher.
- (c) KAT: mon père vient me chercher, ma mère vient me chercher.

Nous constatons alors que la transcription en (a) diverge fortement de celles en (b) et (c) ci-dessus dans la mesure où (a) comporte maintenant deux énoncés alors que (b) et (c) n'en comportent qu'un seul. L'interprétation d'une longueur de pause a donc des répercussions importantes puisque ces transcriptions présentent à présent une variation importante au niveau de la longueur alors que les trois lignes sources, sous leur formes graphiques, ne semblent pas si éloignées les unes des autres. La transcription de (a) est à présent subdivisée en deux énoncés alors que (b) et (c) continuent à être considérés comme deux énoncés de même longueur. Cette différence de ponctuation, qui peut *a priori* sembler tout à fait anodine, entraîne ainsi des répercussions statistiques importantes, particulièrement dans le domaine de

l'acquisition. En effet, nous avons mentionné au Chapitre I que, dans ce domaine particulier, la longueur moyenne des énoncés est très souvent calculée et le chiffre obtenu¹ est généralement utilisé par les chercheurs pour évaluer à quel stade de son développement linguistique se trouve un enfant.

Nous revenons alors à la question que nous formulons initialement en page 195 mais que nous précisons à présent : comment segmenter, de façon sûre et constante, le continuum phonique de l'oral en *énoncés*, vu que l'oral utilise beaucoup de parataxe ? L'exemple mentionné en (9) et transcrit en (10) ci-dessus provient en fait d'une production d'Alan dans nos enregistrements. Au vu de l'intonation et des pauses produites par l'enfant, nous choisissons en fin de compte de le transcrire sous la forme de *deux* énoncés, comme indiqué en (11) ci-dessous :

(11) Transcription choisie² :

ALA: mon papa i(l) vient me chercher +...

ALA: +, ma maman i(l) [*] vient me chercher.

Néanmoins, nous ne pouvons pas garantir que nous aurions transcrit cette production exactement de la même façon si celle-ci avait été prononcée par l'enfant ou transcrite par nos soins à un autre moment. En effet, nous avons beaucoup hésité dans de nombreux cas au sujet de la segmentation à adopter face à la parataxe. Quelques exemples en (12) ci-dessous illustrent d'autres cas similaires. En (a) et en (b), nous constatons l'absence commune de la conjonction de subordination *que* mais un traitement différent quant à la segmentation. Les deux énoncés en (c) auraient pu être transcrits comme un seul avec une pause dans la mesure ils ont un équivalent très proche avec le pronom relatif *qui* à la place du pronom *il* et l'énoncé en (d) aurait pu être segmenté en deux énoncés comme en (e). (f) et (g), bien que similaires, ont également reçu un traitement différent et (h), bien qu'assez long, n'a pas été segmenté.

¹ *MLU (Mean Length of Utterance)* ou LME pour Longueur Moyenne des Énoncés.

² Séance I.2.

(12) Quelques autres exemples de parataxe extraits de notre corpus :

- (a) KAT: tu vois tu sais. (I.4)
- (b) KAT: tu es sûr ?
KAT: hier tu as mis le maillot+de+bain ? (VI.26e)
- (c) KAT: attends on écoute Antoine.
KAT: i(l) nous raconte l' histoire de Titi. (VIII.32a)
- (d) NOE: elle ferme les yeux elle prend une carte ! (VIII.32b)
- (e) MAX: il était allé sur l' eau.
MAX: il a de l' eau sur la bouche. (XIII.47b)
- (f) KAT: tu veux t' asseoir là +..?
KAT: +, comme ça tu me montres ce+que tu fais ? (I.1a)
- (g) KAT: on pose celui+là ?
KAT: comme ça i(l) tombe pas # d+accord ? (I.2)
- (h) KAT: nan on attend Lucie attends elle a pas fini regarde c' est le début de son livre. (VII.27)

A contrario, il est également intéressant de noter que les productions qui contiennent une marque de coordination ne sont pas forcément plus faciles à segmenter :

(13) Exemple de coordination segmentée :

- EMA: sa petite sœur elle [/] elle était dans son jardin.
- EMA: mais tout+à+coup y+a un grand méchant loup. (XIII.46e)

Il n'y a donc rien de plus difficile, à notre avis, que de juger de la qualité et de la longueur d'une pause que l'on peut trouver entre deux éléments, surtout quand ceux-ci sont juxtaposés. Nous limitons donc au strict minimum la variété des signes qui transcrivent la pause et retenons ainsi le point et ses équivalents interrogatif et exclamatif en fin d'énoncé. Quant aux pauses plus brèves à l'intérieur des énoncés, nous ne conservons qu'un seul signe (#) qui, dans notre transcription, marque une pause assez forte (comme en (12g) ci-dessus) ; la possibilité d'utiliser ce signe doublé (##) ou même triplé (###) est ainsi écartée. Nous n'utilisons pas non le signe de la virgule au sein d'un énoncé. Le choix est donc délibéré lorsque les éventuelles pauses légères, très difficiles à déceler avec une précision constante, ne sont pas transcrites. Cette dernière omission a une seconde motivation. Elle permet également de ne pas trancher, au niveau de la transcription, sur l'éventuelle existence/différence

entre les « dislocations » et les « redoublements »¹. Car n'oublions pas que toute transcription est une première interprétation des données ; en conséquence de quoi nous pensons qu'en limitant l'utilisation de la ponctuation aux cas indubitables, nous limitons également notre intervention sur les données.²

1.2.2 L'ordre des constituants

Une deuxième caractéristique des productions orales ajoute à ce sentiment de difficulté de segmenter les énoncés de façon constante. Il s'agit de l'ordre des constituants qui diffère, parfois radicalement, entre un énoncé écrit et une production orale. Les transcriptions en (14) ci-dessous illustrent ce propos et indiquent que, dans ce type de cas également, aucun signe n'est transcrit à l'intérieur des énoncés.

(14) L'ordre des constituants est également différent à l'oral :

- (a) NOE: oh ça va pas ici j(e) crois. (X.40b)
- (b) KAT: et i(ls) sont où là ? (XI.43c)
LUS: euh dans le nid je crois.
- (c) ELO: un perroquet j' ai trouvé ! (XII.44a)
- (d) NOE: ça c' est mon grand frère on dirait. (XII.44c)
- (e) EMA: non c' est moi c' est moi elle a dit. (XII.44d)
- (f) KAT: ou une cocotte+minute aussi ça s' appelle. (XII.44d)

Nous pensons donc qu'il est extrêmement difficile (voire impossible ?) de proposer une segmentation indiscutable et constante des productions orales, et en particulier des productions enfantines. Il semble que la parataxe et l'ordre non strict des constituants, deux particularités des productions orales, sont pour beaucoup dans cette difficulté. Par conséquent, il est également extrêmement délicat de prétendre proposer une transcription homogène et constante des différentes longueurs de pauses tout au long d'un travail de transcription. Dans ces conditions, l'utilisation de la longueur moyenne des énoncés comme critère d'évaluation du développement syntaxique des enfants nous semble présenter un certain aspect aléatoire. C'est la raison pour laquelle nous ne focalisons pas sur ce critère d'évaluation dans notre étude.

¹ Dont les définitions respectives suscitent bien des débats, comme mentionné aux Chapitres I et II.

² Morgenstern (2006:57) par exemple évoque également cette problématique et choisit quant à elle le tour de parole comme unité de comptage, regroupant ainsi plusieurs énoncés en une seule unité.

2. L'anonymisation des données

L'anonymisation des données recueillies fait partie des engagements que nous avons pris dès le départ auprès de l'Inspection académique et des parents de nos petits informateurs. Nous avons vu que l'une de nos premières tâches avait d'ailleurs consisté en l'attribution de pseudonymes à ces enfants¹. Cependant, il est évident que ce processus ne suffit pas à lui-seul à anonymiser l'ensemble des informations contenues dans une base de données. Il convient donc, lors de la transcription, d'anonymiser toutes les données nominatives produites au cours des interactions enregistrées. Il faut également veiller à ce que cette anonymisation soit faite de façon cohérente et homogène afin que ce processus ne nuise pas à la compréhension des énoncés et que les données transformées puissent tout de même être reconnues par le logiciel d'étiquetage morphosyntaxique à des fins d'analyse. C'est la raison pour laquelle deux moyens distincts sont mis en place pour anonymiser les éléments nominatifs présents dans les données primaires, le moyen choisi dépendant de la fréquence du type d'item.

Lorsque l'enfant produit un patronyme (quel qu'il soit, le sien, celui d'un camarade, d'un membre de sa famille, etc.), celui-ci est transcrit par le code Xx². Ce code Xx est suivi de l'indication [= family name] qui précise au lecteur que l'élément non transcrit est un patronyme. Cette indication permet de récupérer au mieux les informations perdues par le processus d'anonymisation, comme le montre l'exemple (15) ci-dessous :

(15) Exemple d'anonymisation patronymique (VIII.33c) :

EKT: merci madame Xx [= family name] ouh ouh ouh !

TOM: mais <j(e) suis pas> [/] j(e) suis pas une madame Xx [= family name]
hein moi.

Le même procédé est appliqué à la plupart des toponymes ; les noms de villes sont ainsi codés Xx [= town]. Cependant, afin de ne perdre que le strict minimum d'informations, des commentaires sont ajoutés lorsque l'enfant parle de distances qu'il a parcourues, comme dans l'exemple (16) ci-dessous. Les commentaires sont

¹ Cf. Tableau 16, page 179.

² Le code xx correspond par ailleurs dans le corpus à la transcription d'un mot inintelligible ; nous ajoutons la majuscule pour indiquer au lecteur (et au logiciel de codage morpho-syntaxique) qu'il s'agit d'un nom propre.

ainsi toujours insérés sous la ligne principale de transcriptions, sous la forme d'une ligne dépendante, intitulée %com.

(16) Exemple d'anonymisation géographique (XII.44a) :

KAT: vous avez fait beaucoup beaucoup beaucoup de route ?

LUS: oui.

KAT: oui ?

LUS: oui mais c' était à Xx [= town].

%com: town about 40km away.

Il arrive qu'un enfant parle de plusieurs villes. Dans ce cas, elles sont numérotées [= town1], [= town2], etc. Ainsi, ce codage ne nuit pas à la compréhension des productions. Certains toponymes par contre ne sont pas anonymisés dans la mesure où ils représentent des villes où beaucoup d'enfants sont allés, comme Paris. Dans ce cas, la conservation du nom de la ville ne donne aucune indication permettant de reconnaître l'enfant, comme le montre l'exemple ci-dessous :

(17) Exemple de non anonymisation géographique (XIII.48a) :

ANT: ben <j' ai> [/] j' étais [= ? j' ai été] à Paris.

Cette convention de réécriture Xx [= family name] ou [= town] est choisie dans la mesure où les enfants ne produisent que très peu de patronymes et de toponymes au cours de nos entretiens. En effet, nous pensons qu'un tel codage ne pourrait pas être appliqué de façon efficace à un type d'élément récurrent dans les productions car il surchargerait la ligne de transcription et serait fastidieux pour le lecteur. C'est pourquoi nous choisissons un autre moyen d'anonymisation pour les prénoms des enfants et ceux de leur entourage, beaucoup plus nombreux dans les énoncés. Les prénoms réels sont ainsi systématiquement remplacés dans les transcriptions et dans tous les documents par des prénoms pseudonymes attribués une fois pour toutes¹. Cette stratégie possède plusieurs avantages par rapport à l'anonymisation des patronymes et des toponymes. En effet, elle n'implique pas l'utilisation d'une indication subséquente supplémentaire sur la ligne de transcription, ce qui permet de rendre la transcription plus fluide et plus facile à suivre. L'utilisation de

¹ Au cours de notre première séance, certains enregistrements audio n'ont pas d'équivalent vidéo. Lorsqu'une voix n'est pas identifiée avec certitude, le locuteur est alors codé sous la forme UNI (= *unidentified child* « enfant non identifié »).

pseudonymes permet également d'éviter la perte d'informations inhérentes au codage en Xx telles que le sexe de l'enfant.

3. Les deux étapes de la transcription

La transcription de nos données primaires s'est déroulée en deux étapes du fait que nous disposons de deux sources d'enregistrement : le caméscope et l'enregistreur audio. Ainsi, dans un premier temps, nous avons toujours transcrit les données primaires issues du support vidéo avant de, dans un deuxième temps, vérifier et compléter ces premières transcriptions grâce aux données présentes sur le support audio. L'existence de ces deux sources de données enregistrées de façon concomitante mais depuis des points légèrement différents s'est révélée extrêmement précieuse et intéressante.

3.1 La transcription à partir des fichiers vidéo

À ce stade, nos données primaires sont donc prêtes à être visionnées par le biais du logiciel CHAT et tous les en-têtes du fichier informatique correspondant sont saisis et renseignés grâce aux informations notées dans notre carnet de bord. La transcription peut donc débiter. Nous procédons toujours de la même manière. D'abord, nous regardons un court passage de la vidéo afin de nous faire une première idée de la scène, des tours de parole et des gestes effectués par chacun lors de ses occurrences. Puis, le même extrait est visionné de nouveau afin de transcrire les occurrences proprement dites. L'extrait est visionné autant de fois que nécessaire, jusqu'à ce que la transcription nous paraisse rendre compte au mieux de ces données primaires. La réécoute systématique des données apparaît alors comme indispensable dans la mesure où différents items ne sont finalement transcrits qu'après plusieurs écoutes. Ces items sont de deux ordres. Il s'agit soit d'occurrences dont la compréhension ne se réalise qu'après les avoir entendues plusieurs fois, soit de reprises dans le discours de l'enfant qui ne sont effectivement perçues qu'après avoir écouté l'énoncé plusieurs fois. Ce deuxième cas de figure est illustré en (18) ci-dessous. Le symbole [/] indique ainsi que l'élément qui précède est repris ; les crochets < > précisent l'étendue de la reprise lorsque celle-ci concerne plus d'un item.

(18) Exemples de reprises perçues lors de réécoutes :

- (a) Première écoute : ALA: nan c' est à moi !
Deuxième écoute : ALA: nan c' est <à m(oi)> [/] à moi !
Troisième écoute : ALA: nan <c' est à m(oi)> [/] c' est à moi !
- (b) EMA: <moi je> [/] moi j' ai un tee+shirt de princesse !
EMA: <moi je> [/] moi j' ai un tee+shirt de [/] de [/] de princesse !

Il est ainsi intéressant de constater qu'il y a une émergence progressive des termes repris dans ces différentes transcriptions, probablement liée au fait que l'information véhiculée par ces reprises est redondante et qu'elle n'est donc pas retenue initialement.

C'est au cours de cette étape que nous insérons également deux types de données secondaires. Un premier type d'informations englobe tous les commentaires que nous faisons de la scène. À cette fin, nous utilisons deux des nombreuses lignes dépendantes proposées par le logiciel : la ligne %com (commentaires), déjà mentionnée, et la ligne %add (*addressee* « interlocuteur »). Les commentaires concernent généralement les gestes que fait le locuteur lors de sa production (*cf.* (19a) ci-dessous), les déplacements de l'interlocuteur (*cf.* (19b)), ou l'identité de l'interlocuteur sous forme de code (*cf.* (19c)).

(19) Exemples d'insertions de données secondaires (VI.26d) :

- (a) EKT: <et moi> [/] <et moi> [/] et moi j' ai pas raconté une histoire de ça.
%com: pointing at Romane's book.¹
- (b) EKT: une fois tu sors.
%com: Claire runs back to the door but does not leave the room.²
- (c) KAT: nan tu restes là ?
%add: KLE³

Le deuxième type d'informations qui est saisi au cours de cette étape concerne les agrammaticalités du corpus. Ainsi, la ligne principale de transcription peut inclure, si le transcripteur le décide, la mention que celui-ci a détecté une agrammaticalité. Le symbole [*] est alors placé après l'item agrammatical, comme le montre l'exemple (20a) ci-dessous où Alan utilise l'auxiliaire *être* à la place de l'auxiliaire *avoir*.

¹ « montrant le livre de Romane ».

² « Claire retourne à la porte en courant mais ne quitte pas la pièce ».

³ « Claire ».

Lorsque l'agrammaticalité provient de l'absence d'un item, celle-ci est codée grâce à l'ensemble des symboles 0 [*], comme dans l'exemple (20b) où Zoé omet le clitique *je*.

(20) La détection des agrammaticalités dans le corpus :

(a) ALA: et moi je suis [*] fait un escargot. (I.1a)

(b) ZOE: regarde moi aussi 0 [*] sais le faire. (I.3a)

Lorsqu'un transcripateur décide de mentionner les agrammaticalités dans un corpus, ces symboles représentent la première étape d'un codage qui en comporte deux. En effet, ces symboles sont ensuite reliés à une ligne dépendante, intitulée %err (erreur), sur laquelle le transcripateur décrit l'agrammaticalité¹.

3.2 La transcription à partir des fichiers audio

La seconde étape de la transcription débute alors, grâce à l'écoute des enregistrements audio. Nous nous sommes très rapidement aperçue de la très grande utilité de cette deuxième source. En effet, le micro de l'enregistreur vocal était plus proche des enfants que le micro du caméscope. Ces fichiers audio sont donc venus compléter la source son du caméscope. Ils ont notamment permis la transcription d'éléments qui n'avaient pas été identifiés à partir de la bande son de la vidéo et qui, donc, n'avaient pas été transcrits dans un premier temps. De plus, l'option « défilement lent » de l'enregistreur audio a permis d'identifier des éléments supplémentaires, incompris jusque là, malgré des écoutes répétées en vitesse normale. Quelques exemples en (21) ci-dessous illustrent l'aspect non négligeable de l'apport complémentaire de cette deuxième source de données (le symbole xxx sur chaque première ligne indique la présence de plusieurs syllabes restées jusque là inintelligibles).

¹ Cf. Section V.1 de ce chapitre.

(21) Exemples de la complémentarité de la deuxième source son :

- (a) NOE: xxx ma place. (V.23c)
NOE: *c' est à moi* ma place.
- (b) NOE: xxx arrête. (V23d)
NOE: *ah ben oui on* arrête.
- (c) EMA: parc(e)+que xxx. (VI.24b)
EMA: parc(e)+que *il est [*] pas besoin*.
- (d) NOE: et xxx les trois petits cochons xxx. (VII.29a)
NOE: et *maint(e)nant* les trois petits cochons *i(ls) chantent*.

Mais il ne s'agit pas du seul apport de cette source audio. Nous y avons trouvé un second intérêt extrêmement important. En effet, cette source audio a également permis de *corriger* des transcriptions faites à partir de la piste son du caméscope et qui, de fait, se sont révélées inexactes face à la confrontation avec cette deuxième source sonore enregistrée plus près des enfants. Quelques exemples en (22) ci-dessous illustrent que, de nouveau, cet apport est loin d'être insignifiant. L'exemple (23) reprend une correction particulièrement intéressante qui montre bien que, lorsque la transcription ne semble pas avoir de sens, c'est qu'une erreur s'y est probablement glissée. Bien sûr parfois, complément et correction se superposent.

(22) Quelques exemples de corrections :

- (a) ANT: nan nan t(u) as tout fini. (V.22)
ANT: *là là* t(u) as tout fini ?
- (b) ANT: moi j' ai tout fini ! (V.22)
ANT: *regarde* j' ai tout fini !
- (c) ANT: on va prendre le goûter ! (V.22)
ANT: *non là 0 [*] faut* prendre le goûter !
- (d) NOE: c' est toi c' est bleu. (V.23c)
NOE: *cette fois* c' est bleu.
- (e) KAT: toi tu penses quoi ? (VI.24b)
KAT: *qu+est+c(e)+que* tu penses *toi* ?
- (f) TEO: oh encore ! (VII.27b)
TEO: oh *un coq* !
- (g) TEO: +, je vais voir si xxx. (VII.27b)
TEO: +, je *veux* voir si *y+a le clown*.

- (h) EMA: nan pas encore. (VII.27b)
 EMA: nan *il est* pas encore.
- (i) LUS: quand il a tout bu je crois que maintenant i(l) va +... (VIII.31a)
 LUS: quand il a tout bu je crois que *maman elle va arriver*.
- (j) NOE: moi aussi j' ai xxx dessin animé. (XI.41a)
 NOE: moi aussi *je l' ai comme ça en* dessin animé.
- (k) LUS: le chien !
 ANT: où il est le p(e)tit chien ?
 LUS: *Lucien !*
 ANT: où il est *le tee+shirt ?*
- (l) EMA: regarde elle fait quoi le garçon xxx ?
 EMA: regarde elle *le [*]* fait quoi *là 0 [*]* le garçon *dans son xx ?*¹

(23) Complément et correction :

- (a) CLA: moi [/] moi je mange <chez ma m(amie)> [/] chez ma mamie Rolande et [/] et elle a pas un chat mais [/] mais [/] mais elle avait plus d(e) pizza xxx plus d(e) pizza elle est malade elle est morte.
 KAT: xxx ?
 CLA: c' est le chat.
- (b) CLA: moi [/] moi je mange <chez ma m(amie)> [/] chez ma mamie Rolande et [/] et [/] et elle a pas un chat mais [/] mais [/] mais elle avait *Pépita maintenant Pépita elle est plus* elle est malade elle est morte.
 KAT: *qui c' est Pépita ?*
 CLA: c' est le chat.

Ces quelques exemples illustrent l'apport très intéressant de cette confrontation entre nos deux sources son placées à deux endroits légèrement différents dans une même pièce. Ils établissent ainsi que ces deux sources sont complémentaires. En effet, le caméscope seul aurait fourni à la fois du son et de l'image et par conséquent une foule d'informations visuelles extrêmement importantes. Cependant, les corrections ci-dessus montrent que cette source son n'était pas complètement fiable. *A contrario*, l'enregistreur vocal seul aurait fourni des données audio de meilleure qualité mais il aurait manqué énormément d'informations quant à l'identification certaine des

¹ Si il n'y avait pas eu la correction, la première phrase aurait été codée avec un pronom elle [*] (= il), supposé coréférencé avec « le garçon ».

locuteurs, leurs gestes, leur(s) interlocuteur(s), etc. Notre expérience établit ainsi une réelle complémentarité dans l'utilisation de ces deux médias, en particulier dans le cadre d'un travail de collecte auprès de jeunes enfants. En effet, la fugacité de certaines de leurs occurrences ne permet pas de les fixer correctement sur le support vidéo, plus éloigné de la table de travail que le support audio. De plus, il semble y avoir une certaine distorsion au niveau du support le plus éloigné. Par conséquent, si nous n'avions travaillé qu'avec la vidéo, nous savons désormais que notre corpus aurait contenu un certain nombre d'erreurs de transcription, totalement involontaires et dont nous n'aurions même pas eu connaissance. Nous en concluons, qu'au niveau de la méthodologie de la collecte de données dans ce type d'environnement, il est plus sûr de travailler à partir de deux sources son concurrentes afin de les confronter et de pallier les éventuelles distorsions inhérentes dès qu'une petite distance sépare le sujet du micro qui l'enregistre.

Chacune de nos transcriptions est ainsi systématiquement confrontée à l'écoute des fichiers son correspondants, en progressant par petites séquences. Une fois arrivée à la fin du fichier complet, nous établissons une dernière vérification de la transcription, dans son intégralité cette fois, pour être bien sûre que les enchaînements entre les séquences sont corrects et qu'aucun passage n'est oublié. Nous corroborons alors les différentes estimations fournies sur le temps de travail nécessaire pour transcrire des données orales puisque notre propre estimation est du même ordre que celles mentionnées précédemment¹ : il nous a fallu approximativement une moyenne de trente minutes de travail pour transcrire et vérifier une minute de données, ce qui, pour ce corpus de vingt heures, représente environ six cents heures de travail.

V. LE CODAGE DES DONNÉES

Afin d'être manipulées, les données doivent ensuite être codées. Dans le cadre de notre travail de recherches sur l'acquisition du sujet syntaxique et dans le cadre de notre approche du langage enfantin en tant que système linguistique à part entière, nous avons effectué un certain nombre de choix concernant le codage de nos transcriptions. Dans un premier temps, il a ainsi été décidé de ne pas se focaliser uniquement sur les éléments sujets dans les énoncés. L'ensemble des données a ainsi

¹ Cf. les estimations de Baude (2006:30) et de Demuth (1996:17) données en page 154.

été traité afin de pouvoir établir d'éventuelles généralisations ou de possibles parallèles entre différents phénomènes. Nous verrons dans le chapitre suivant que cette stratégie s'est avérée payante. Dans un deuxième temps, nous avons restreint l'éventail des lignes dépendantes à la ligne de transcription principale aux seules lignes jugées pertinentes pour notre étude. En effet, comme pour les autres options du logiciel, les possibilités sont nombreuses afin de permettre au plus grand nombre de chercheurs travaillant dans des domaines les plus divers de coder leurs données de façon adéquate. Nous avons ainsi limité les lignes dépendantes régulières à deux : une ligne intitulée %err (erreur) qui rend compte des agrammaticalités signalées sur la ligne principale par le code [*], et une ligne intitulée %mor qui permet un codage morphosyntaxique des occurrences. Ces choix permettent alors de travailler à partir de deux sources d'informations différentes et complémentaires : un panorama complet du système linguistique enfantin grâce à l'étiquetage morphosyntaxique et un point de vue ciblé sur les divergences entre le système enfantin et le système cible de l'adulte. En de très rares occasions, nous avons également utilisé une troisième ligne de codage. Il s'agit de la ligne %pho qui permet une transcription phonétique par le biais des symboles de l'Alphabet Phonétique International. Cette ligne est insérée lorsque la transcription orthographique de l'occurrence est trop hasardeuse, comme illustré dans l'exemple (24) ci-dessous :

(24) Exemple de codage %pho (API) :

ROM: yy yy yy [*] se cache là ? (IX.34a.mor, 945)
 %pho: ke se le

Cependant, aucune de ces étapes de codage n'est générée de façon automatique. Chaque domaine possède sa codification propre et le chercheur est libre d'utiliser des codes existants ou d'en créer de nouveaux en fonction de ses besoins.

1. Le codage des agrammaticalités

Il convient bien sûr dans un premier temps de définir avec précision ce que nous entendons par « agrammaticalité ». Ainsi, sont identifiées comme agrammaticales toutes les occurrences (enfantines *et* adultes) qui présentent une divergence avec la langue cible correspondante, c'est-à-dire avec le français *oral* quotidien¹. L'éventail des agrammaticalités, toutes signalées par le symbole [*] sur la ligne principale de

¹ Cf. Chapitre I pour quelques caractéristiques de ce système.

transcription, semble donc *a priori* très large. Néanmoins, nous avons posé deux limites à ce codage.

1.1 Le choix des occurrences

Tout d'abord, nous ne signalons comme agrammaticales que les occurrences clairement perçues. Ainsi, lorsque la transcription est incertaine et que celle-ci comprend une agrammaticalité, nous choisissons de ne pas la relever en tant que telle mais uniquement d'identifier l'incertitude sous la forme du code [?], comme indiqué dans l'exemple (25) ci-dessous. Dans ce cas, les occurrences subséquentes, indiquées entre crochets, proposent la forme grammaticale qui pourrait être la forme cible correspondante. Lors d'études statistiques, c'est cette deuxième forme qui est comptabilisée.

(25) Exemple :

TOM: <c' écrit> [?] [: c' est écrit] quoi là ? (I.2)

Par ailleurs, en cas de doute entre une transcription grammaticale et une version agrammaticale d'une occurrence, la version grammaticale est préférée afin de ne pas comptabiliser d'éventuelles fausses agrammaticalités. Les exemples en (26) ci-dessous illustrent ce précepte. Il est notamment appliqué en ce qui concerne la présence possible mais non certaine : du clitique objet entre un clitique sujet et un verbe qui commence par une voyelle (*cf.* (a) ci-après)¹, de la segmentation correcte entre un nom et l'élément qui le précède lorsqu'il y a liaison (*cf.* (b)), de la forme correcte du pronom et du verbe en cas d'homophonie entre les formes du singulier et du pluriel (*cf.* (c)), de l'accord correct entre le verbe de la proposition relative avec l'antécédent (*cf.* (d))², l'accord correct entre le participe passé et le sujet ou le complément d'objet antéposé au verbe (*cf.* (e)), le temps/mode correct lorsque deux formes sont homophones (en l'occurrence le présent de l'indicatif et du subjonctif, *cf.* (f)). Par contre, par analogie avec les autres tournures du français oral où on la sait absente, l'absence de la particule négative dans les cas de liaison est privilégiée, comme illustré en (g) ci-dessous.

¹ Lors de l'analyse au Chapitre IV, des informations supplémentaires permettent néanmoins de trancher sur cette question.

² Bien que, lorsque les formes de 1sg et 3sg sont phonétiquement différentes pour un verbe, comme dans le cas de « vais » et « va », on observe que c'est la forme du 3sg qui est privilégiée par les enfants.

(26) Exemples :

- (a) MAX: il l' enlève. (IV.19a)
ELO: elle l' a. (IV.19f)
- (b) LUS: et la souris elle s' est accrochée sous un *arbre*. (X.38a)
- (c) ANT: et les moutons i(ls) *restent* là. (VII.29a)
LUS: tous les deux petits i(ls) *regardent* par la fenêtre
- (d) LUS: non non c' est moi qui *mets*. (IV.19f)
MAX: euh c' est toi qui l' *as*. (X.40f)
- (e) TOM: elle est *finie* ! (VIII.33c)
ANT: voilà je l' ai *cassée* ! (IX.37d)
- (f) JUL: 0 [*] faut pas qu' on les *voie*. (X.40b)
- (g) EMA: oh *on a* pas donné le xx.

L'homophonie représente également un handicap pour la transcription des clitiques sujets comme illustré dans les cas repris en (27) ci-dessous. En effet, en (a) et (b), la forme verbale est clairement identifiée comme étant agrammaticale alors que la forme du clitique est ambivalente. Dans ces cas, la forme préférée (d'après le postulat indiqué ci-dessus) est suivie de l'autre transcription possible, insérée sous forme d'équation, entre crochets et précédée d'un point d'interrogation ([= ?]).

(27) Exemples :

- (a) EMA: alors les [/] tous les petits chats i(ls) [= ? i(l)] boit [*] ici. (X.38b)
- (b) KAT: et qu+est+c(e)+qu' i(ls) font après ? (X.38b)
EMA: après i(ls) [= ? i(l)] s' en va [*] !

La transcription des deux alternatives présente l'intérêt de pouvoir extraire et travailler soit uniquement sur la première forme, soit sur les deux.

1.2 Notre grille de codes

Ces types d'occurrences étant écartés, toutes les autres agrammaticalités du corpus sont étiquetées en tant que telles afin de les extraire facilement à des fins d'analyse. Ainsi, dès lors que le symbole [*] est inséré sur la ligne principale de transcription, cela implique l'ajout d'un niveau subséquent dépendant dans lequel l'agrammaticalité est identifiée puis décrite par le chercheur. Dans un premier temps, le terme agrammatical est donc repris sur la ligne %err et est mis en équation avec le terme grammatical correspondant :

(28) LUS: ah ben moi j(e) va [*] écouter les cloches pour Pâques ! (IX.34a)

%err: va = vais

Dans un deuxième temps, comme pour les autres options du logiciel, CHAT dispose d'une liste variée de codes¹, tous identifiés par le symbole \$. Ces codes sont hiérarchisés entre eux et permettent de décrire de façon plus ou moins précise les agrammaticalités relevées. Un premier code permet ainsi de distinguer les agrammaticalités par domaine. Parmi les codes disponibles, nous exploitons les éléments suivants : phonologie (\$PHO), morphologie (\$MOR), syntaxe (\$SYN), lexique (\$LEX), et allomorphie (\$ALL). Un second niveau de codage permet de concaténer ce premier code avec un second qui rend compte du type d'erreur. Nous utilisons les possibilités suivantes : ajout (\$ADD), perte (\$LOS), substitution (\$SUB), amalgame (\$BLE), différents types d'inversions (\$EX, \$SH), anticipation (\$ANT), et persévération (\$PER). Puis un troisième niveau de codage détaille l'agrammaticalité en fonction du domaine d'occurrence. Si l'agrammaticalité est phonologique, des codes permettent alors d'indiquer si elle concerne, entre autres, un élément vocalique (\$VOW) ou consonantique (\$CON) ; les codes morphologiques distinguent notamment les préfixes (\$PRE) des suffixes (\$SUF), la flexion (\$NFL), l'accord (\$AGA), ou la surgénéralisation (\$REG). Les codes syntaxiques quant à eux incluent des codes de position (\$POS) et de fonction (\$SUBJ). Un possible quatrième niveau de codage détaille précisément le type d'élément impliqué dans l'agrammaticalité. C'est à ce niveau que peuvent être indiqués la personne et le nombre (\$1S, \$1P, etc.), le cas (\$NOM, \$ACC, etc.), la catégorie (\$N, \$PRO, etc.), et bien d'autres détails à choisir parmi une liste extrêmement fournie². La concaténation de ces différents niveaux de codage produit alors des codes très précis qui permettent d'extraire les données par domaine, par mécanisme ou par de nombreuses autres entrées telles que la catégorie grammaticale, le cas, la personne, etc. ; quelques exemples de ces combinaisons sont reproduits en (29) ci-dessous³ :

¹ Cf. MacWhinney (2000a:94-96 et 114-116).

² Cf. MacWhinney (2000a:114-116) pour la liste complète.

³ La liste complète des combinaisons utilisées se trouve en Annexes, pages 372-377.

(29) Quelques exemples de codages sur la ligne %err :

(a) Omission du clitique sujet de Personne 1 :

UNI: moi 0 [*] veux le jaune. (I.1a)

%err: 0 = je \$SYN \$LOS \$PRO \$SUBJ \$IS

(b) Omission du déterminant indéfini :

NOE: oh la [: y+a] 0 [*] oiseau dedans ! (IV.18a)

%err: 0 = un \$SYN \$LOS \$DET \$INDEF

(c) Mauvais genre sur le clitique :

ZOE: xxx moi il [*] a pris ça à moi Noémie. (II.8)

%err: il = elle \$MOR \$SUB \$AGA \$PRO

(d) Mauvaise personne sur le verbe :

LUS: ah ben moi j(e) va [*] écouter les cloches pour Pâques ! (IX.34a)

%err: va = vais \$MOR \$SUB \$AGC \$V

(e) Présence d'une liaison non attendue : \$ALL \$ADD

NOE: les mamans zarrivent [*] (extrait de l'énoncé ; V.23e)

%err: zarrivent = arrivent \$ALL \$ADD

(f) Ajout de consonne par persévération :

ZOE: xxx i(l) va arriver le Papa_Nonël [*] [: Noël] dans le liv(r)e.

%err: Nonël = Noël \$PHO \$ADD \$PER \$CON (II.8)

(g) Substitution de l'auxiliaire :

ALA: et moi je suis [*] fait un escargot. (I.1a)

%err: suis = ai \$LEX \$SUB \$AUX

1.3 L'importance du contexte

Ce type de codage par hiérarchie de codes permet donc de proposer une identification très fine des agrammaticalités. Comme indiqué ci-dessus, celles-ci sont formulées en termes d'omission, d'ajout ou de substitution. Concernant l'omission des clitiques sujets, nous retrouvons alors la dichotomie traditionnellement formulée dans la littérature sous les formes *null subjects* « sujets nuls » opposés aux *root infinitives* ou *optional infinitives* « propositions principales infinitives »¹ ; les premières correspondant à une omission du sujet en présence d'un verbe fléchi alors que les secondes renvoient à des omissions dans des propositions non fléchies. Il convient cependant de tenir compte du contexte lors du codage de ces propositions

¹ Il s'agit de la terminologie proposée par Rizzi (1992a,b) et Wexler (1998), respectivement.

infinitives. En effet, il est entendu qu'il en existe un certain nombre qui sont grammaticales, comme le montrent les échanges proposés en (30) ci-dessous :

(30) Propositions indépendantes infinitives grammaticales :

(a) KAT: qu+est+c(e)+que vous voulez faire ? (IV.19e)

EKT: *eah jouer.*

(b) KAT: qu+est+c(e)+qu' on doit faire ici ? (XIII.46e)

LUS: *dormir.*

Dans ces cas, il s'agit d'ellipses grammaticales dans la mesure où la partie omise dans la réponse correspond à des éléments produits dans la question (en l'occurrence la modalité : « je veux (jouer) », « on doit (dormir) »)¹. A *contrario*, les énoncés repris en (31) ci-après sont étiquetés comme agrammaticaux, l'agrammaticalité étant identifiée comme provenant d'une absence de flexion (\$SYN \$LOS \$INFL), qu'il s'agisse d'un infinitif ou d'un participe². Cette partition rejoint ainsi celle proposée par Ezeizabarrena (2003) dans le cadre de son étude sur les propositions infinitives en basque enfantin dans la mesure où l'auteur distingue les propositions infinitives grammaticales dans la langue cible (qu'elle nomme *Optional infinitives*³) des propositions agrammaticales dans la langue cible (qu'elle intitule *Root infinitives*).

(31) Propositions indépendantes infinitives agrammaticales :

(a) KAT: explique à Célia le jeu. (IV.19e)

EKT: *hm 0 [*] fermer les yeux.*

%err: 0 = il faut \$SYN \$LOS \$INFL

(b) EKT: *0 [*] déchiré toi.* (IV.19e)

%err: \$SYN \$LOS \$INFL

D'après notre codage, les énoncés en (31) ci-dessus ne ressortissent donc pas directement à l'absence d'un sujet mais plutôt à l'absence d'une flexion verbale pertinente dans ces contextes, par opposition aux contextes qui permettent l'absence de flexion, comme en (30) ci-dessus. L'absence du sujet est alors considérée comme un phénomène corollaire, qui découle donc de l'absence de la flexion. L'absence du sujet dans une proposition infinitive n'est donc pas considérée ici comme une agrammaticalité en soi ; elle est la conséquence d'une autre agrammaticalité. Dans

¹ Cf. par exemple Labelle (2000).

² Sachant que pour les verbes du premier groupe, la distinction n'est pas toujours possible.

³ Il s'agit notamment de l'omission grammaticale des auxiliaires dans certains contextes.

Palasis (2005), nous n'avions déjà pas tenu compte de ce type de phrases lors de notre analyse du sujet nul enfantin. Cette exclusion ne relevait pas d'une décision arbitraire mais provenait du fait que ce type d'occurrences ne représentait qu'une toute petite partie des sujets nuls dans notre premier corpus (1,4 % des phrases du corpus, soit 6,2 % des phrases sans sujet)¹. De plus, ces taux étaient compatibles avec d'autres études menées sur ce thème dans la mesure où il est généralement observé que la production d'infinitives agrammaticales a lieu jusqu'à approximativement l'âge de deux ans. Nous n'avions donc pas analysé la présence de ces infinitives. En revanche, leur quasi-absence avait été interprétée comme corroborant à la fois notre hypothèse sur le phénomène du sujet nul enfantin, et l'hypothèse de Wexler (1998) sur l'existence de langues dans lesquelles les enfants produisent de telles infinitives (comme le français) et des langues dans lesquelles leur taux est extrêmement bas ou inexistant (comme l'italien). Dans la mesure où notre premier corpus avait mis en avant le fait que le système linguistique du jeune francophone présente des similitudes avec le système de langues dites « à sujet nul » comme l'italien, il était donc logique que l'on ne trouve pas ou quasiment pas d'infinitives dans le système linguistique des enfants observés².

2. Le codage morphosyntaxique

À la différence du codage précédent, le codage morphosyntaxique balaie les données dans leur totalité. Ainsi, lors de cette étape, toutes les occurrences enfantines et adultes présentes dans la base de données sont étiquetées afin de les rendre informatiquement manipulables par le tableau de commandes CLAN³. Cette étape complète et termine donc la phase de construction de notre base de données. Il s'agit d'ajouter une nouvelle ligne dépendante (%mor), placée sous chaque ligne de transcription. Celle-ci propose alors une description morphosyntaxique de chaque mot graphique situé sur la ligne principale. C'est pour cette raison par exemple que les éléments suivis d'une apostrophe sont séparés du mot graphique suivant par un espace lors de la transcription, afin que ces éléments (« j' », « l' », etc.) soient pris en compte séparément par le codage morphosyntaxique. C'est également dans cette optique de codage que certains mots graphiques sont regroupés par le signe + sur la

¹ Cf. Palasis (2005:108-110) et Palasis & Oliviéri (2007:128-130) pour le détail.

² Mais cf. l'étude d'Ezeizabarrena (2001) sur le basque qui ne confirme pas cette dichotomie dans la mesure où le basque enfantin comprend un nombre certain de propositions infinitives alors qu'il s'agit d'une langue *pro-drop*.

³ Cf. Chapitre IV.

ligne de transcription. Nous transcrivons notamment les locutions interrogatives « qu'est-ce que, qu'est-ce qui, etc. » en les concaténant sous la forme « qu+est+ce+que » afin d'étiqueter l'ensemble en tant qu'une seule entrée. Ce procédé est également appliqué à quelques noms tels que cache+cache, sac+à+dos, maillot+de+bain, cochon+d+inde, appareil+photo, etc. ; quelques locutions, telles que en+train+de, à+partir+de, quand+même, à+cause+de, en+bas, etc. ; ainsi que pour les formes qui comportent le présentatif [ja] (« il y a ») transcrites y+a, y+en+a, y+avait. Les noms propres composés sont quant à eux transcrits avec un tiret, comme dans « Petit_Ours_Brun ». D'autres symboles utilisés lors de la transcription trouvent également leur raison d'être dans cette ligne dépendante. Il s'agit notamment du codage [:] qui suit une occurrence non standard et qui reprend, entre ces crochets, la forme standard afin que cette dernière soit ensuite la forme étiquetée par le codage morphosyntaxique. « Père_Nonël » est ainsi suivi de ce type de parenthèses dans lesquelles est indiqué le terme correspondant [: Noël]. Tous ces protocoles utilisés lors de la transcription permettent ainsi de délimiter les éléments que l'on souhaite être identifiés comme unités à étiqueter sur cette ligne %mor.

Chaque unité ainsi délimitée reçoit un code. De nouveau, et comme pour les autres options du logiciel, les possibilités de codage sont hiérarchisées et diverses. Elles peuvent donc être plus ou moins affinées et elles sont également laissées au libre arbitre du chercheur. Le site Internet de CHILDES propose des « grammaires » pré-formatées pour certaines langues. Pour le français, les codes %mor ont été principalement créés par Christophe Parisse¹. Tous ces codes présentent un format similaire : ils débutent tous par l'indication de la catégorie grammaticale de l'item, suivie d'un nombre varié de précisions. Pour notre base de données, nous utilisons les fichiers de Christophe Parisse auxquels nous apportons deux séries de modifications. Les premières concernent simplement l'ajout de fichiers destinés à recevoir le lexique spécifique à notre base de données. Les secondes impliquent des modifications plus profondes. En effet, dans le cadre de nos recherches, nous affinons notamment l'analyse grammaticale initiale des pronoms (pro), en insérant un certain nombre d'informations supplémentaires. Tout d'abord, nous distinguons les pronoms potentiellement sujets des autres en les marquant tous d'une étiquette

¹ Du Laboratoire MoDyCo (Modèles, Dynamiques, Corpus), UMR 7114 CNRS, Nanterre. Ces fichiers sont disponibles sur <http://childes.psy.cmu.edu/morgrams/french.zip>.

qui indique leur cas (nominatif, accusatif, etc.)¹. Puis, au sein de chaque cas, plusieurs sous-classes sont codées. Pour les pronoms sujets, dont les codes pour la Personne 1 sont repris dans le Tableau 20 ci-après, nous différencions ainsi : les pronoms clitiques (pro:cli), les toniques (pro:ton) et les démonstratifs (pro:dem), ceux qui sont utilisés seuls et ceux qui sont doublés (pro:cli:d), les doublements à gauche (dg) et à droite (dd). Ainsi pour la seule Personne 1, sept possibilités de codage sont envisagées :

Formes	Codes	Exemples
<i>j(e)</i>	pro:cli:nom je&1S	ROM: <i>j'</i> ai perdu l(e) chien . (VII.29b, 1851)
	pro:cli:d:nom je&1S	NOE: attends <i>je</i> l' ai moi . (VIII.32b, 328)
<i>moi</i>	pro:ton:nom moi&1S	TEO: <i>moi</i> 0 [*] veux ça les cartes . (VIII.33a, 65)
	pro:ton:dg:nom moi&1S	EVA: <i>moi j(e)</i> suis jaune ! (VII.29b, 2010)
	pro:ton:dd:nom moi&1S	MAX: j' ai perdu <i>moi</i> . (VIII.31b, 1046)
	pro:ton:dg:nom moi&1S pro:ton:dd:nom moi&1S	ALA: <i>moi j'</i> ai pas encore fini <i>moi</i> . (I.3a, 28)

Tableau 20 : les codes possibles pour un pronom sujet de Personne 1

Les exemples (32) ci-après illustrent quelques possibilités de codage en contexte et montrent les codes de quelques autres catégories grammaticales, telles que les verbes qui, eux aussi, sont subdivisés en différentes sous-classes (verbes lexicaux, auxiliaires et modaux notamment).

(32) Exemples de codage %mor (I.1a) :

- (a) KAT: comment tu t' appelles ?
 %mor: adv:int|comment
 pro:cli:nom|tu&2S
 pro:cli:refl|te&2S
 v|appeler&PRES&_2SV ?
 ANT: Antoine .
 %mor: n:prop|Antoine

¹ Un code spécifique est également prévu pour les occurrences de pronoms dans un contexte non verbal, tel que TOM: *et ben moi non plus* . (VIII.33b, 602) (codage : pro|moi&SING).

(b) KAT: et toi tu t' appelles Alan !

```
%mor: conj|et
      pro:ton:dg:nom|toi&2S
      pro:cli:d:nom|tu&2S
      pro:cli:refl|te&2S
      v|appeler&PRES&_2SV
      n:prop|Alan !
```

Pour certaines langues, ces fichiers de codes pré-formatés permettent de générer les lignes %mor de façon automatique. C'est le cas pour l'anglais, le néerlandais, l'allemand, le japonais, et l'espagnol¹. Ce n'est pas le cas pour le français ; l'étiquetage est donc semi-automatique dans la mesure où le logiciel présente au chercheur la liste des possibilités d'étiquetage pour chaque occurrence et que la sélection et la validation finales sont effectuées manuellement.

3. En conclusion

Cette étape d'étiquetage clôt ainsi la constitution de notre base de données. L'exemple (33) ci-dessous est une illustration de la combinaison de la ligne principale de transcription, suivie des lignes de code %mor, elles-mêmes suivies de la ligne de code %err qui détaille l'agrammaticalité signalée dès la ligne de transcription.²

(33) Combinaison de CHAT, %mor et %err (I.1a) :

```
ALA: et moi je suis [*] fait un escargot .
%mor: conj|et, pro:ton:dg:nom|moi&1S, pro:cli:d:nom|je&1S,
      v:aux|être&PRES&1SV [*], v:mdllex|faire&PP&_MASC&_SING,
      det|un&MASC&SING, n|escargot&_MASC .
%err: suis = ai $LEX $SUB $AUX
```

Le soin apporté au recueil de nos données, à leur transcription ainsi qu'à leur étiquetage est un premier pas en direction d'un travail de recherches qui se veut de qualité. Cet outil de travail est dorénavant prêt à être sondé dans le détail. Pour cela, CHILDES propose alors un deuxième type de logiciel, baptisé CLAN (*Computerized Language Analysis*) que nous découvrons et utilisons dans le chapitre suivant.

¹ Cf. MacWhinney (2000a:156).

² L'ensemble des fichiers de transcriptions et des codages se trouvent en Annexes, page 371.

I. EXTRACTIONS PRÉLIMINAIRES DE DONNÉES

1. CLAN

CLAN, acronyme de *Computerized Language Analysis*¹, est un programme informatique utilisé en complément de CHAT, créé par Leonid Spektor et complété au fil des années par de nombreux informaticiens et linguistes². Cet outil permet de formuler des requêtes d'extraction au sein de données transcrites et codées au format CHAT. Comme pour le codage, les possibilités de travail sont extrêmement variées dans la mesure où le chercheur compose lui-même chacune de ses requêtes en sélectionnant les items pertinents. Très schématiquement, ceux-ci se divisent en quatre catégories principales : (i) le programme adéquat, tel que le calcul de la fréquence d'un item (FREQ) ou l'extraction de combinaisons d'éléments (COMBO) ; (ii) le type de ligne sur lequel le programme doit fonctionner : ligne de transcription principale ou ligne dépendante, telle que %mor ; (iii) le(s) locuteur(s) étudié(s), et (iv) le(s) fichier(s) sources correspondant à une, plusieurs ou la totalité des séances d'enregistrement. D'autres options permettent également d'obtenir différentes présentations, de diriger les résultats vers certains fichiers, etc. La compilation de ces différentes catégories est illustrée en (1) ci-dessous.

¹ « Analyse informatisée du langage ».

² Cf. MacWhinney (2000a:5-6) pour un historique complet de ce programme.

(1) Exemple de requête CLAN : extraction des pronoms nominatifs par enfant, par séance et par ordre décroissant de fréquence :

```
freq +t*XXX -t* +t%mor +s"pro*:nom*|*" +o +u @ >proXXX1.all
```

- (a) freq : utilisation de la commande `FREQ` qui génère une liste d'occurrences, chacune accompagnée de sa fréquence et calcule le ratio type-token¹.
- (b) `+t*XXX` : sélectionne les lignes correspondant au locuteur XXX.
- (c) `-t*` : exclut les lignes principales (c'est-à-dire les lignes `*XXX`).
- (d) `+t%mor` : parmi les lignes dépendantes du locuteur choisi, sélectionne uniquement les lignes `%mor`.
- (e) `+s"pro*:nom*|*"` : recherche tous les pronoms (codés pro) en ne retenant que la contrainte du cas nominatif (codé nom) et englobant ainsi toutes les autres possibilités, telles que clitique ou tonique, doublé ou non, etc., indiquées par les astérisques.
- (f) `+o` : classe les occurrences par ordre décroissant de fréquence (par défaut, `FREQ` classe les éléments par ordre alphabétique).
- (g) `+u` : lorsque l'on travaille sur plusieurs fichiers à la fois, cette option permet de renvoyer un total global (par défaut, les totaux sont donnés par fichier).
- (h) `@` : les fichiers input sont sélectionnés dans le tableau de commandes CLAN afin de générer des recherches séance par séance (pour la séance I, les huit fichiers I.1a, I.1b, I.2, I.3a, I.3b, I.3c, I.3d, I.4 sont ainsi sélectionnés).
- (i) `>proXXX1.all` : envoie le résultat de la requête vers un fichier dont le nom est spécifié après le signe `>`.

Ces commandes présentant une relative complexité, nous avons mené une série de tests sur des extraits de fichiers afin de vérifier que les requêtes formulées extrayaient bien le type attendu de données. Dans un premier temps, ces requêtes nous permettent de connaître les caractéristiques globales de notre base de données et de notre corpus.

2. Premières requêtes afin de composer notre corpus

Nos premières requêtes portent sur le nombre total d'énoncés que comporte notre base de données. Après avoir extrait les totaux concernant tous les locuteurs

¹ Correspond au nombre total de mots différents divisé par le nombre total de mots afin de chiffrer la diversité lexicale.

enregistrés (soit trois adultes et vingt-deux enfants auxquels nous avons ajouté un locuteur « non identifié » afin de coder les énoncés pour lesquels nous ne sommes pas sûre de la provenance), nous devons opérer un premier tri afin de composer notre corpus de travail.

2.1 Les locuteurs

Comme mentionné précédemment, les productions de deux enfants (Éloïse et Lola) montrent de forts signes d'interférences avec une autre langue (le russe pour Éloïse et le portugais pour Lola). Ces deux enfants sont donc écartées de notre étude. Nous laissons également de côté les quelques énoncés dont le locuteur n'est pas clairement identifié. Nos informateurs pour cette étude se montent donc à vingt enfants et trois adultes. Cette première sélection nous laisse ainsi avec un total de 29504 énoncés, dont 16462 énoncés (soit 55,8 %) sont produits par les enfants. Ces premiers chiffres sont récapitulés dans le Tableau 21 ci-après¹.

Intitulés	Nombres d'énoncés
Totalité de la base de données	30542
Données écartées	1038
Nombre d'énoncés restants	29504
Sous-total pour les 20 enfants	16462
Sous-total pour les 3 adultes	13042

Tableau 21 : synthèse des premières requêtes

Ce premier taux de 55,8 % pour les productions enfantines est intéressant dans la mesure où il ne reflète pas du tout la proportion du nombre d'enfants comparé au nombre d'adultes présents lors des séances. En effet, dans l'immense majorité des cas, les productions adultes sont issues d'une seule personne (puisque KAT = 12981 énoncés, SOF = 33 et TAA = 28). Par contre, nous pensons que ce taux illustre la nécessité qu'a l'adulte de solliciter sans cesse des enfants de cet âge afin d'obtenir des productions orales. C'est une remarque que nous avons déjà formulée lors de notre première enquête² et que nous réitérons, d'autant plus que, lors de notre deuxième année d'enquête auprès de ces mêmes enfants, nous constatons que, dès la première séance en novembre 2007, les productions sont beaucoup plus spontanées et fluides. Nous faisons de nouveau cette remarque lorsque nous retrouvons les enfants en novembre 2008 pour notre troisième année d'enregistrements et nous notons que les interactions entre enfants sont alors beaucoup plus nombreuses et

¹ Le tableau comportant le détail par locuteur et par séance se trouvent en Annexes, page 378.

² Cf. Palasis (2005).

suivies. Les nombres d'énoncés de KAT pour le corpus actuel ainsi que leurs taux relatifs sont détaillés dans le Tableau 22 ci-dessous.

Séances	1	2	3	4	5	6	7	8	9	10	11	12	13
Tous loc.	1590	2333	2176	2365	2265	2046	2476	2027	2235	3169	2564	1494	2764
KAT	803	1131	822	1094	1031	965	1154	732	792	1432	1118	705	1202
% KAT	50,5	48,5	37,8	46,3	45,5	47,2	46,6	36,1	35,4	45,2	43,6	47,2	43,5

Tableau 22 : évolution des énoncés de KAT par rapport à l'ensemble du corpus

Ces chiffres permettent de constater une baisse globale des tours de parole de KAT au fur et à mesure des séances. Cependant, cette tendance peut être attribuée à différents facteurs (une mise en confiance des enfants tout autant qu'un développement de leur désir de communiquer), aussi laissons-nous cette thématique de côté dans le cadre de cette étude.

2.2 Les phrases verbales

Dans le Chapitre I, nous avons vu qu'une condition *sine qua non* de la présence d'un sujet est la présence d'un verbe. Il convient donc d'effectuer une deuxième sélection au sein de notre base de données afin de comptabiliser les phrases verbales uniquement. Les verbes sont codés sous quatre formes différentes : les verbes fléchis (par exemple : v:poss|avoir&pres&1sv pour la forme lexicale de *ai*), les infinitifs (par exemple : v:mdllex|aller&inf pour la forme lexicale de *aller*), les participes passés (par exemple : v:pp|acheter&_masc&_sing pour la forme *acheté*) et les participes présents (par exemple : v:prog|attendre pour la forme *attendant*). À ce stade, nous extrayons ainsi tous les énoncés qui comportent au moins une occurrence de l'une de ces quatre formes¹. Les quelques exemples repris en (2) ci-dessous illustrent le type d'occurrence écarté. Il est intéressant de noter que ce type d'énoncé n'est pas forcément mono-morphématique.

¹ Requête : mlu -t*ELO -t*LOL -t*UNI -t* +t%mor +s"v*|*" +u @ >WithVerb.

(2) Quelques exemples de phrases non verbales :

- (a) TOM: non . (I.2, 168)
- (b) KAT: et toi Alan ? (I.1a, 135)
- (c) ANT: le jardin de mon papa et ma maman . (I.3a, 262)
- (d) LEA: une p(e)tite photo bébé Léa (II.5, 262)
- (e) ANT: euh ben [/] ben [/] ben c(el)ui+là ! (XIII.49c, 1045)

Cette nouvelle requête fournit un total de 19568 énoncés, soit 66,32 % de tous les énoncés retenus à la suite de notre premier tri. La répartition entre enfants et adultes montre que chaque groupe produit environ la moitié de ces occurrences, comme le détaille le Tableau 23 ci-dessous¹ :

Types	Nombres	%
Totalité des énoncés	29504	100
Énoncés avec verbe(s)	19568	66,32
Sous-total pour les 20 enfants	9748	49,82
Sous-total pour les 3 adultes	9820	50,18

Tableau 23 : les énoncés verbaux dans notre base de données

Globalement, la proportion de phrases verbales est plus élevée chez les adultes que chez les enfants. Ainsi, le taux le plus élevé se trouve chez KAT (75,36 %) et le taux le plus faible est celui de KLE (46,43 %) mais l'un des adultes montre un taux de 50 % et ZOE a un taux avoisinant celui de KAT (74,08 %). Tous ces chiffres sont donc à manipuler avec précaution et, dans le cadre de notre étude, nous nous gardons également de proposer des hypothèses sur cette thématique. Il faudrait très probablement analyser la progression de ces taux d'une séance à l'autre, pour chaque locuteur, tout en considérant le contexte de l'énonciation. Par ailleurs, les différents symboles de ponctuation utilisés lors de la transcription permettent d'extraire les phrases selon leur type : déclaratif (. +... +/. +//.), interrogatif (? +..? +/? +//? +!?) ou exclamatif (!). Les taux repris dans le Tableau 24 ci-dessous indiquent clairement le sens des interactions, les adultes utilisant majoritairement l'interrogation afin de solliciter les enfants et les enfants privilégiant la forme déclarative en réponse à ces sollicitations.

¹ Le tableau avec le détail des taux par locuteur se trouve en Annexes, page 379.

	Tous types		Déclaratif		Interrogatif		Exclamatif	
	N	%	N	%	N	%	N	%
Adultes	9818	100,00	4584	46,69	4839	49,29	395	4,02
Enfants	9745	100,00	6468	66,37	1141	11,71	2136	21,92
Total	19563	100,00	11052	56,49	5980	30,57	2531	12,94

Tableau 24 : répartition des énoncés verbaux par type

Concernant le nombre de verbes dans la base de données, les requêtes menées avec le programme *FREQ*¹ livrent les chiffres repris dans le Tableau 25 ci-après. De nouveau, les pourcentages montrent que les items se répartissent à peu près équitablement entre les adultes et les enfants. Cependant, le détail et le total des formes différentes utilisées par chaque groupe de locuteurs indiquent que ces formes ne se superposent pas complètement et que la diversité lexicale n'est que légèrement supérieure chez les adultes.

Locuteurs	Nombres	%	Formes ≠	Diversité
Adultes	14320	51,20	913	0,064
Enfants	13650	48,80	818	0,060
Total	27970	100,00	1250	0,045

Tableau 25 : répartition adultes/enfants des verbes du corpus

En ce qui concerne notre investigation sur le sujet, il faut affiner ce classement dans la mesure où il n'est pas attendu que tous les verbes fassent surface en compagnie d'un sujet. En effet, d'une façon générale, les infinitifs et les participes ne possèdent pas leur sujet propre lorsqu'ils sont accompagnés d'un auxiliaire (par exemple : *j'ai dessiné*) ou d'un modal (par exemple : *je veux dessiner*)². Le nombre de verbes fléchis dans le corpus peut ainsi apporter une idée approximative du nombre de sujets potentiels ; le Tableau 26 ci-dessous synthétise les résultats des requêtes correspondantes³.

¹ (i) `freq -t*ELO -t*LOL -t*UNI -t* +t%mor +s"v*|" +u @ > Verbs_Freq_All` ; (ii) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v*|" +u @ > Verbs_Freq_Adults` ; (iii) `freq -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA -t* +t%mor +s"v*|" +u @ > Verbs_Freq_Kids`.

² Des tournures exclamatives du type *Moi, dessiner !* sont cependant également possibles.

³ Six requêtes sont nécessaires par groupe de locuteurs. Les requêtes pour le groupe des adultes est : (i) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v*|" +u @ > Verbs_Freq_Finite_Adults` ; (ii) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v:|" +u @ > Verbs_Freq_pp1_Adults` ; (iii) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v:pp|" +u @ > Verbs_Freq_pp2_Adults` ; (iv) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v:inf|" +u @ > Verbs_Freq_Inf1_Adults` ; (v) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v:|" +u @ > Verbs_Freq_Inf2_Adults` ; (vi) `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"v:progl|" +u @ > Verbs_Freq_Prog_Adults`.

Types	Tous	%	Adultes	%	Enfants	%
Fléchis	22899	81,87	11872	82,91	11027	80,78
Infinitifs	3351	11,98	1806	12,61	1545	11,32
Participes	1720	6,15	642	4,48	1078	7,90
Totaux	27970	100,00	14320	100,00	13650	100,00

Tableau 26 : répartition des verbes d'après la flexion

2.3 Clitiques nominatifs et verbes fléchis

Il n'y a pas de controverse sur le fait que l'hôte unique et obligatoire d'un clitique nominatif est un verbe fléchi en français. Ce fait est observé de façon incontestable chez l'adulte mais il est également attesté que le jeune enfant, au cours de son acquisition, ne produit pas d'énoncés dans lesquels un clitique nominatif s'appuierait sur un hôte autre qu'un verbe fléchi. Les différentes études présentent en effet des résultats consensuels à ce sujet avec, pour chaque corpus, un nombre non significatif d'exceptions¹. Ce nouveau corpus ne déroge pas à la règle puisque seuls quatre énoncés présentent cette configuration² :

(3) Liste exhaustive des énoncés où le clitique est extrait sans verbe fléchi :

- (a) EVA: ben moi je 0 [*] prendre le li(vre) [/] un livre . (IV.19e, 798)
%err: 0 = vais \$SYN \$LOS \$V \$MDL
- (b) ZOE: et je 0 [*] le refaire là . (V.23a, 393)
%err: 0 = veux \$SYN \$LOS \$V \$MDL
- (c) NOE: là [/] euh je 0 [*] faire du manège avec Carnaval . (VI.26c, 313)
%err: \$SYN \$LOS \$V \$MDL
- (d) EMA: moi quelle bonne idée moi j(e) 0 [*] faire des crêpes quelle bonne idée ! (XI.41b, 839)
%err: 0 = vais \$SYN \$LOS \$V \$MDL

Les occurrences indiquent que cette configuration ne provient ni d'un enfant en particulier, ni d'une séance spécifique, ni d'un âge précis ; de plus, son taux d'apparition est infinitésimal³. Ces combinaisons possèdent néanmoins une caractéristique commune puisque ces quatre occurrences présentent la même agrammaticalité, comme l'indiquent l'astérisque sur chaque ligne principale et le code %err sur chaque ligne dépendante. Cette agrammaticalité est codée comme

¹ Cf. Kayne (1975) sur le français en général et Pierce (1992), Hamann *et al.* (1994) ou Palasis (2005) sur l'acquisition du français.

² Requête : combo -t*ELO -t*LOL -t*UNI +t%mor +s"pro:cli*:nom*|^!v*|*v*" +x +u @ > Cli_NoFlexion.

³ Sur les 17440 clitiques nominatifs du corpus, le taux est de 0,02 %.

étant l'absence d'un verbe de modalité ; en l'occurrence, il s'agit du verbe *aller* en (3a) et (3d), de *vouloir* en (3b) et le contexte ne permet pas proposer le verbe manquant exact en (3c). Une requête portant sur cette agrammaticalité indique qu'elle n'apparaît dans le corpus que dans ces quatre occurrences¹. Nous notons également que ces quatre productions sont enfantines, que tous les clitiques sont de P1 et qu'ils sont tout de même combinés à un verbe mais que ce dernier n'est pas fléchi (le verbe *(re)faire* dans trois cas sur quatre). Les données de ce corpus confirment donc bien le fait que l'association d'un clitique nominatif avec un item autre qu'un verbe n'est pas une option possible et ce dès le système enfantin². De plus, les agrammaticalités rarissimes quant à l'attachement du clitique nominatif à un verbe non fléchi établissent qu'il y a un lien extrêmement étroit entre les clitiques nominatifs et la flexion verbale.

3. Récapitulatif

Pour l'heure, nous pouvons établir deux fiches signalétiques de nos données : celle de notre base complète, destinée à renseigner la communauté scientifique sur les contours de la teneur de cette base, et celle de notre corpus de travail, spécifiant le nombre de locuteurs et d'énoncés sur lesquels nous basons notre étude sur le développement du sujet enfantin. Ces deux fiches se trouvent dans le Tableau 27 et le Tableau 28 ci-après.

¹ Requête : `combo -t*ELO -t*LOL -t*UNI +t%err +s"$SYN^$LOS^$V^$MDL" +u @> NoModal.`

² Si cette association relève d'un paramétrage, celui-ci est donc fixé correctement à ce stade.

Langue	français
Lieu de collecte	Sud de la France
Dates de collecte	De Novembre 2006 à Juin 2007
Suivi longitudinal	7 mois
Nombre de séances	13
Intervalles entre deux séances	Min. :10 jours ; Max. : 40 jours
Nombre d'enfants	22
Âges à la première séance	De 2;5.5 à 3;4.24
Âges à la dernière séance	De 3;0.13 à 4;0.1
Nombre d'adultes	3
Enregistrements	Audio et vidéo
Transcription	Orthographique
Codage	Morphosyntaxique et agrammaticalités
Nombre total d'énoncés	30542
Répartition énoncés enfantins/adultes	17500/13042

Tableau 27 : fiche signalétique de notre base de données

Langue	français
Lieu de collecte	Sud de la France
Dates de collecte	De Novembre 2006 à Juin 2007
Suivi longitudinal	7 mois
Nombre de séances	13
Intervalles entre deux séances	Min. :10 jours ; Max. : 40 jours
Nombre d'enfants retenus	20
Âges à la première séance	De 2;5.5 à 3;4.24
Âges à la dernière séance	De 3;0.13 à 4;0.1
Nombre d'adultes	3
Enregistrements	Audio et vidéo
Transcription	Orthographique
Codage	Morphosyntaxique et agrammaticalités
Nombre d'énoncés verbaux	19568
Répartition énoncés enfantins/adultes	9748/9820

Tableau 28 : fiche signalétique du corpus de cette étude

Nous pouvons à présent nous tourner vers l'extraction et les statistiques spécifiques à notre thématique.

4. Les différentes formes de sujets

Nous l'avons vu dans le Chapitre I, le sujet peut prendre un nombre certain de formes différentes. À partir de notre premier corpus recueilli en 2003-2004, nous avons établi une première typologie des sujets grammaticaux et agrammaticaux produits par les dix-sept enfants enregistrés alors. Nous rappelons ce classement ainsi que les taux correspondants dans les tableaux ci-dessous dans la mesure où ces classifications représentent le point de départ de notre étude actuelle.

Types ¹	Exemples	N	%
+PrC	<i>Elle est grande</i>	459	42,8
+PoC		0	0,0
+PrN		0	0,0
+PrN +C	<i>Le renard il a mangé</i>	16	1,5
+C +PoN	<i>Il est là la galette</i>	30	8,1
+PrP +C	<i>Lui i(l) s'appelle Raphaël</i>	87	2,8
+C +PoP	<i>Je peux jouer moi ?</i>	64	6,0
+PrP +C +PoP	<i>Moi j'ai un grand vélo moi</i>	6	0,5
Impératif	<i>attends !</i>	170	15,9
Sous-total (phrases grammaticales)		832	77,6
Sous-total (phrases agrammaticales)		240	22,4
Totalité du corpus		1072	100,0

Tableau 29 : typologie des sujets grammaticaux (corpus 2003-2004)

Types	Exemples	N	%
-C -N	* <i>veux un camion</i>	151	14,1
+PrP -C	* <i>moi fais</i>	33	3,1
-C +PoP	* <i>veux vert moi</i>	33	3,1
-C +PoN	* <i>est fermé le manège</i>	15	1,4
+PrP -C +PoP	* <i>et moi veux ça moi</i>	8	0,7
Sous-total (phrases agrammaticales)		240	22,4
Sous-total (phrases grammaticales)		832	77,6
Totalité du corpus		1072	100,0

Tableau 30 : typologie des sujets agrammaticaux (corpus 2003-2004)

Ces deux tableaux montrent que la partition entre sujets grammaticaux et sujets agrammaticaux chez ces jeunes enfants est corrélée à la présence vs l'absence d'un clitique nominatif. Ce lien entre grammaticalité et présence d'un clitique nominatif est rendue possible par l'absence totale dans ce corpus de phrases grammaticales comportant un syntagme nominal utilisé seul comme sujet (*cf.* ligne 3 du Tableau 29 ci-dessus). Cette généralisation peut ensuite être étendue à tous les types de phrases agrammaticales qui eux ne comportent aucun clitique nominatif. Voyons si notre nouveau corpus apporte le même type de configuration.

II. LES CLITIQUES NOMINATIFS

1. Les clitiques utilisés seuls

Notre système de codage %mor des éléments traditionnellement étiquetés comme « pronoms » se développe sur plusieurs niveaux hiérarchisés. Nos codes se présentent ainsi sous forme de cascades d'étiquettes, ce qui permet des extractions

¹ +=présent, -=absent, Pr=préverbal, Po=postverbal, C=clitique, N=nom, P=pronom tonique.

très ciblées¹. Cette hiérarchie permet alors d’extraire les données qui correspondent à la requête « pronom + clitique + seul + nominatif ». Par défaut, avant la transcription des données, treize codes (correspondant aux formes *je, tu, il, il explétif, elle, on, ce, ça, nous, vous, vous politesse, ils, elles*) avaient été établis. La liste exhaustive des codes correspondant aux formes de notre base de données montre que nous avons ajouté un code afin d’adapter le codage aux formes effectivement produites et afin de maintenir un codage fin. Il s’agit du code qui correspond au *il explétif* utilisé de façon postverbale². La grille de codage comporte alors quatorze formes, comme le détaille le Tableau 31 ci-dessous :

1	pro:cli:nom je&1S	8	pro:cli:nom ce&3S
2	pro:cli:nom tu&2S	9	pro:cli:nom ça&3S
3	pro:cli:nom il&3S	10	pro:cli:nom nous&1P
4	pro:cli:nom:imprs il&3S&IMPRS	11	pro:cli:nom vous&2P
5	pro:cli:nom:imprs:post il&3S&IMPRS	12	pro:cli:nom:pol vous&2P
6	pro:cli:nom elle&3S	13	pro:cli:nom ils&3P
7	pro:cli:nom on&3S	14	pro:cli:nom elles&3P

Tableau 31 : liste exhaustive des clitiques nominatifs non doublés

Plusieurs programmes permettent de travailler sur ces codes. Dans un premier temps, nous utilisons le programme *FREQ* afin d’extraire une liste exhaustive des items présents dans le corpus ainsi qu’un décompte global pour chacun des éléments.

1.1 L’ensemble du corpus

La première requête, formulée pour l’ensemble du corpus³, renvoie un nombre total de 12244 clitiques répartis en douze formes différentes. Il est donc intéressant de noter que certaines formes prévues par défaut ne sont utilisées ni par les enfants ni par les adultes. Puis, les requêtes formulées, d’une part pour l’ensemble des enfants⁴, et d’autre part pour l’ensemble des adultes⁵, donnent une première vue d’ensemble de la répartition des différents clitiques. Ces résultats sont présentés par ordre décroissant de fréquence dans le Tableau 32 ci-après :

¹ Cf. Chapitre III, Section V.2 pour le détail.

² Il est donc entendu que, sauf indication contraire, les clitiques sont préverbaux.

³ Requête globale : `freq -t*ELO -t*LOL -t*UNI -t* +t%mor +s"pro:cli:nom*|*" +u @ > Cli_Freq_All.`

⁴ Requête enfants : `freq -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA -t* +t%mor +s"pro:cli:nom*|*" +u @ > Cli_Freq_Kids.`

⁵ Requête adultes : `freq +t*KAT +t*SOF +t*TAA -t* +t%mor +s"pro:cli:nom*|*" +u @ > Cli_Freq_Adults.`

Formes	Tous	Enfants	%	Adultes	%
pro:cli:nom ce&3s	3294	1803	54,74	1491	45,26
pro:cli:nom tu&2s	2313	483	20,88	1830	79,12
pro:cli:nom je&1s	1821	1361	74,74	460	25,26
pro:cli:nom on&3s	1643	597	36,34	1046	63,66
pro:cli:nom il&3s	1427	987	69,17	440	30,83
pro:cli:nom:imprs il&3s&imprs	485	174	35,88	311	64,12
pro:cli:nom vous&2p	368	14	3,80	354	96,20
pro:cli:nom elle&3s	357	207	57,98	150	42,02
pro:cli:nom ça&3s	314	93	29,62	221	70,38
pro:cli:nom ils&3p	190	100	52,63	90	47,37
pro:cli:nom elles&3p	28	8	28,57	20	71,43
pro:cli:nom:imprs:post il&3s&imprs	4	4	100,00	0	0,00
Totaux	12244	5831	47,62	6413	52,38

Tableau 32 : classement et répartition enfants/adultes des clitiques non doublés

Des quatorze formes initiales, ce sont les deux formes *nous* et *vous politesse* qui ne se trouvent pas dans ce corpus. Par ailleurs, la forme ajoutée *a posteriori* pour rendre compte de l'utilisation du *il explétif* postverbal correspond à des productions enfantines, ce qui montre une légère différence par rapport à notre premier corpus qui ne comprend aucun clitique postverbal. Une recherche ciblée de cette forme permet de connaître les contextes de ces quatre items qui sont reproduits ci-dessous¹ :

(4) Les quatre clitiques postverbaux du corpus en contexte :

- (a) EKT: euh: 0 [*] s(e) passe t il ? (IV.19d, 33)
- (b) QUE: qu+est+c(e)+qui 0 [*] passe t il [*] ? (IV.19d, 38)
- (c) EKT: que se passe t il ? (IV.19d, 52)
- (d) QUE: qu+est+c(e)+que [*] s(e) passe t il ? (IV.19d, 57)

La mise en contexte des occurrences indique que les quatre clitiques proviennent de la même scène. Il s'agit d'un dialogue entre Hector (EKT, 3;5.8) et Quentin (QUE, 3;7.1). Durant cet échange, les deux garçons se retrouvent seuls dans la salle de travail pendant un court instant et ils sont très surpris de cette situation. Ils entament alors la conversation reprise en (4) ci-dessus. Un point important à souligner est le ton qu'emploient les deux enfants. En effet, il est évident, d'après leur intonation, que les garçons, en formulant ces questions qui présentent des clitiques nominatifs postverbaux, imitent une intonation d'adultes qui seraient en train d'entretenir une conversation extrêmement « sérieuse ». Le ton n'est donc pas celui de leurs interactions spontanées habituelles. Par ailleurs, trois des quatre énoncés présentent

¹ Requête : combo +t%mor +s"pro:cli:nom:imprs:post|il&3s&imprs" @ >Cli_IExpPost.

une agrammaticalité liée au fait que la structure visée présente une double difficulté : (i) utilisation d'un verbe à la fois impersonnel et pronominal (*se passer*) et (ii) tournure interrogative. Une telle configuration, qui implique le déplacement du verbe et de l'objet sous forme interrogative et qui utilise un clitique non référentiel n'est, de toute évidence, pas maîtrisée par les deux garçons. En effet, même si Hector produit un énoncé correct (en (4c) ci-dessus), l'intonation semble tout de même indiquer que la forme relève davantage de l'imitation que de la computation. Le corpus indique donc que, globalement, les enfants ne manipulent pas de clitics nominatifs postverbaux et que les adultes enregistrés ne les utilisent pas non plus dans ce cadre. Il peut ainsi être déduit que les locuteurs n'utilisent pas l'ordre V-S dans leurs tournures interrogatives, qu'il s'agisse de questions oui/non (*cf.* (5a) ci-dessous) ou de questions *qu-* (*cf.* (5b)) :

(5) Quelques exemples de structures interrogatives :

- (a) KAT: tu connais le panda ? (I.2, 1269)
- EVA: et moi je peux ? (IV.19d, 533)
- (b) KAT: où il est ? (I.2, 1536)
- ROM: pourquoi i(l) pleut ? (IV.19e, 505)

Les autres chiffres de la première colonne du Tableau 32 ci-dessus indiquent que les clitics les plus fréquents sont pour la plupart des clitics singuliers (*ce, tu, je, on, il, il explétif*) et que les formes les moins présentes sont les deux formes de la troisième personne du pluriel (*ils, elles*). Les taux en gras quant à eux montrent une complémentarité logique entre une utilisation plus importante du *je* chez les enfants (74,74 % des occurrences de *je*) vs du *tu* (79,12 %) et *vous* (96,20 %) chez les adultes. Comme pour le type de phrase mentionné précédemment, ces taux flèchent clairement le sens des interactions au cours desquelles l'adulte sollicite l'enfant puis l'enfant parle de lui-même (*je*) ou de son environnement (*ce, il*) sans établir, encore à ce stade, de retour vers l'autre. Intéressons-nous à présent à la répartition des clitics au sein du système enfantin uniquement.

1.2 Le système enfantin

Si nous nous reportons au nombre de verbes fléchis enfantins extrait précédemment (soit 11027), le chiffre de 5831 clitics, mentionné dans le Tableau 32 ci-dessus, indique que ces éléments sont présents à la gauche de 52,88 %

de ces verbes. Cette catégorie d'éléments, à elle seule, représente donc la majorité des sujets enfantins. Par ailleurs, l'étude de la corrélation entre cette classe d'éléments et le type des phrases n'établit pas de lien particulier entre ces deux variables¹.

(6) Quelques exemples de clitiques non doublés :

- (a) JUL: *c'* est jaune . (I.1a, 270)
- (b) EKT: nan *je* veux mes épées ! (VIII.33c, 211)
- (c) TOM: oh *i(l)* mange encore ça ! (I.2, 374)
- (d) EVA: allez *on* va ranger la classe . (I.3c, 1010)
- (e) ANT: eh eh *t(u)* aimes les glaces ? (IX.37c, 509)

Le Tableau 33 ci-après présente le détail de la répartition de ces clitiques, en termes de nombres et de taux décroissants au sein du système enfantin.

Formes	N	%
pro:cli:nom ce&3s	1803	30,92
pro:cli:nom je&1s	1361	23,34
pro:cli:nom il&3s	987	16,93
pro:cli:nom on&3s	597	10,24
pro:cli:nom tu&2s	483	8,28
pro:cli:nom elle&3s	207	3,55
pro:cli:nom:imprs il&3s&imprs	174	2,98
pro:cli:nom ils&3p	100	1,71
pro:cli:nom ça&3s	93	1,59
pro:cli:nom vous&2p	14	0,24
pro:cli:nom elles&3p	8	0,14
pro:cli:nom:imprs:post il&3s&imprs	4	0,07
Totaux	5831	100,00

Tableau 33 : classement par ordre décroissant de fréquence des clitiques nominatifs non doublés enfantins

Un regroupement de ces clitiques d'après leurs traits [Personne] et [Nombre], présenté dans le Tableau 34 ci-dessous, livre une vue d'ensemble de ce système :

¹ La requête : `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:nom*|^v*|^*^*?" +x +u @ >SentenceCliVInterr_Children` renvoie un total de 592 énoncés, soit 51,88 % des interrogatives enfantines.

Formes	[Pers]	N [Pers]	% [Pers]	[Nbre]	N [Nbre]	% [Nbre]
pro:cli:nom je&1s	1	1361	23,34	sg	5709	97,91
pro:cli:nom tu&2s	2	483	8,28			
pro:cli:nom il&3s	3	3865	66,28			
pro:cli:nom:imprs il&3s&imprs						
pro:cli:nom:imprs:post il&3s&i.						
pro:cli:nom elle&3s						
pro:cli:nom on&3s						
pro:cli:nom ce&3s						
pro:cli:nom ça&3s						
pro:cli:nom vous&2p	5	14	0,24	pl	122	2,09
pro:cli:nom ils&3p	6	108	1,85			
pro:cli:nom elles&3p						
Totaux		5831	100,00		5831	100,00

Tableau 34 : répartition des clitiques nominatifs non doublés dans le système enfantin

Ces chiffres mettent en avant les caractéristiques suivantes : (i) l'écrasante majorité des clitiques possède le trait [Singulier] (97,91 %), et (ii) les clitiques [Singulier] sont majoritairement de Personne 3 (66,28 %) et de Personne 1 (23,34 %) puisque la Personne 2 n'est présente que dans 8,28 % des cas. D'une part, ces taux sont identiques à ceux de Palasis (2008) concernant l'absence totale du clitique *nous* et ils sont très similaires sur la distribution du trait [Singulier] vs [Pluriel]¹. Cependant, ils présentent un écart d'environ 6 % en ce qui concerne l'utilisation de la Personne 2². Les enfants du corpus actuel sont légèrement plus âgés que ceux de nos premiers enregistrements. Cet écart pourrait donc provenir de cette différence d'âge. Cette imputation impliquerait alors qu'il y aurait une émergence progressive de l'utilisation du *tu* avec l'âge. Cette hypothèse est compatible avec les données de Léonard recueillies par Morgenstern (2006)³ et elle présente des analogies extrêmement intéressantes avec l'étude diachronique proposée par Olivieri (2009) sur l'émergence des clitiques nominatifs dans les dialectes occitans parlés aux frontières de la zone occitane, là où l'on observe un continuum entre les systèmes linguistiques à « sujet nul » et les systèmes à « sujet non nul », ou les systèmes à « sujet nul partiel » comme les dialectes bien étudiés du nord de l'Italie⁴.

¹ Palasis (2008:246) : [Sg] = 98,8 % et [Pl] = 1,2 %.

² Palasis (2008:246) : [P2] = 2,9 %.

³ Cf. Morgenstern (2006:65,105) : *tu* apparaît chez Léonard à 2;4.

⁴ Cf. Chapitre II, Section II.2.1 sur ces notions.

1.3 L'évolution des clitiques : ce que nous dit la diachronie

Suivant l'une des hypothèses développées par Dalbera (2006) selon laquelle la variation diatopique est le reflet de la variation diachronique, Oliiviéri (2009) propose une étude des clitiques nominatifs dans des dialectes parlés aux frontières nord et est de l'Occitanie¹. Au fil du temps, ces dialectes ont perdu une partie de leurs suffixes verbaux distinctifs en termes de [Personne]. Oliiviéri (2009) résume la situation ainsi :

“Either new endings can be recreated in order to replace the lost ones and preserve the post-verbal information, e.g. in the west of the Alpes-Maritimes or the Person information previously conveyed by distinctive verbal suffixes can be brought back into the system thanks to preverbal elements built from the Latin personal nouns or pronouns TU, ILLU, ILLA, etc., e.g. in the east of the Alpes-Maritimes.”²

Comme le montre l'étude d'Oliiviéri (2009), ces systèmes linguistiques ont évolué de façons très diverses afin de pallier cette perte d'information. Conséquemment, le fil de ces évolutions est extrêmement délicat à reconstruire. En ce qui concerne l'ordre d'apparition des clitiques préverbaux, plusieurs hypothèses existent mais ne se superposent pas³. Oliiviéri (2009) apporte un éclairage nouveau à ces questions en considérant non pas le trait [Personne] porté par chaque clitique mais en focalisant l'attention sur la fonction distinctive de chaque clitique dans le système linguistique. Par exemple, afin de distinguer la Personne 2 de la Personne 3, certains dialectes voient d'abord émerger un clitique pour P2 alors d'autres dialectes utilisent d'abord un clitique pour P3. Avec une telle contradiction ($P2 > P3$ ou $P3 > P2$), il paraît impossible de poser une régularité. Cette impossibilité disparaît lorsque l'on subsume les deux évolutions sous une même fonction qui est celle de distinguer deux formes verbales devenues homophones. Que l'on marque l'une ou l'autre n'a aucune

¹ Les données utilisées par Oliiviéri (2009) proviennent de la base de données THESOC (cf. <http://thesaurus.unice.fr>) et de Manzini & Savoia (2005).

² Oliiviéri (2009:39) : « Soit de nouveaux suffixes peuvent être recréés afin de remplacer ceux qui ont été perdus et afin de préserver l'information postverbale, comme dans l'ouest des Alpes-Maritimes, soit l'information sur la Personne qui était précédemment véhiculée par les suffixes verbaux distinctifs peut être ramenée dans le système grâce à des éléments préverbaux construits à partir des noms ou pronoms latins TU, ILLU, ILLA, etc., comme dans l'est des Alpes-Maritimes. » Cf. Dalbera (1994) pour le détail sur les dialectes de l'ouest.

³ Cf., entre autres, les travaux de Renzi & Vanelli (1982), Heap (2002), Cabredo Hofherr (2004) et Kaiser & Hinzelin (à paraître).

importance ; ce qui compte, c'est que ces deux formes soient de nouveau différentes dans le paradigme. Oliviéri (2009) se base sur quatre traits pour formaliser une évolution commune à tous les systèmes linguistiques : (i) le trait [Personne] qui renvoie à la « corrélation de personnalité » de Benveniste (1946) et qui oppose P1 et P2 à P3 ; (ii) le trait [Locuteur] qui renvoie à la « corrélation de subjectivité » de Benveniste (1946) et qui oppose P1 à P2 ; (iii) le trait [Genre] qui distingue les différents clitiques de P3 et (iv) le trait [Nombre] qui distingue P3 de P6 par exemple. Oliviéri (2009) propose une émergence progressive de ces traits dans les systèmes linguistiques afin d'expliquer les ordres d'apparition apparemment divers des clitiques. Nous reprenons ce schéma en Figure 1 ci-dessous :

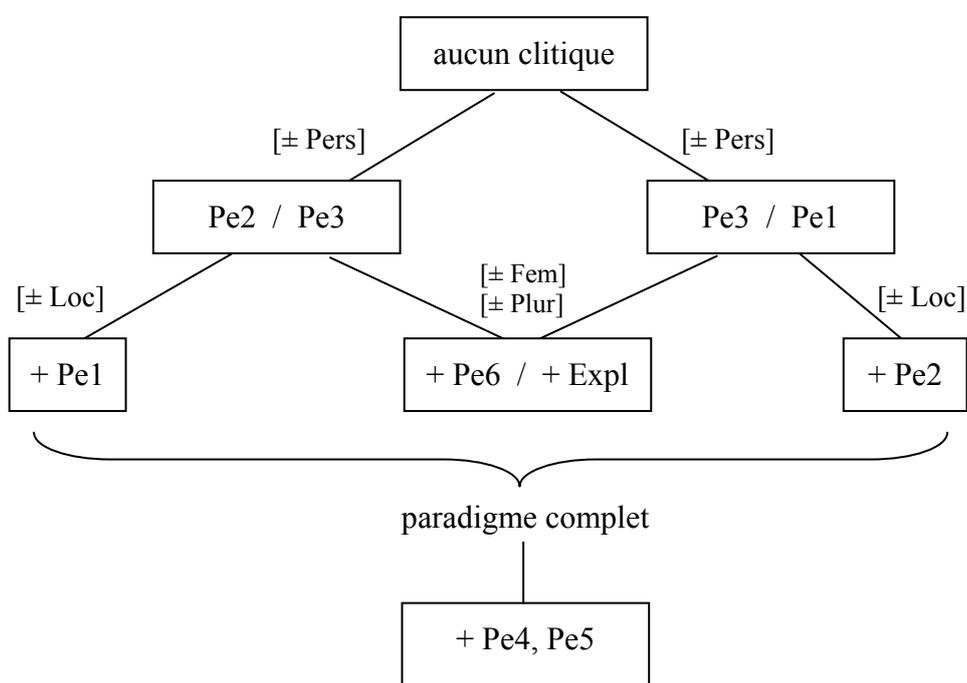


Figure 1 : émergence des traits dans les systèmes linguistiques (Oliviéri, 2009)

Dans ce cadre, les dialectes, avec leurs paradigmes de clitiques nominatifs variés tant au niveau des formes que des personnes réalisées, instancient au niveau diatopique les différentes étapes de l'évolution diachronique des clitiques que l'on trouve entre le stade 0 représenté par le latin (sans aucun clitique) et le stade le plus abouti représenté par le français standard (avec le paradigme complet). L'hypothèse que nous formulons d'après nos premiers enregistrements et que nous allons vérifier ici est que l'ordre d'implémentation de ces traits est également valable pour une autre « diachronie », celle qui touche au développement du système linguistique infantin.

1.4 L'évolution des clitiques : ce que nous dit l'acquisition

Dans Palasis (2005), les données recueillies sur cinq séances avaient permis de formuler les hypothèses suivantes : (i) le système enfantin initial possède une opposition unique P1 vs P3, basée sur le trait [Locuteur], (ii) les autres traits tels que [Genre] et [Nombre] ne sont pas pertinents dans la mesure où, initialement, il n'y a opposition ni entre [Masculin] et [Féminin] ni entre [Singulier] et [Pluriel] ; ces traits, par défaut, ont donc la valeur [+Masculin] et [+Singulier]¹, (iii) ces traits, ainsi que le trait [Référentiel]², deviennent actifs progressivement, au fur et à mesure que le système enfantin s'étoffe. Globalement, ces premières hypothèses sur le développement des clitiques chez l'enfant sont en adéquation avec l'hypothèse d'Oliviéri (2009) mentionnée dans la section précédente. Il convient cependant d'affiner ces hypothèses acquisitionnelles grâce à notre nouveau corpus dans la mesure où il comporte un plus grand nombre d'enfants et de séances. Nous procédons donc à l'extraction du compte des clitiques par enfant, en groupant les fichiers par séance afin d'obtenir une étude tant longitudinale que transversale de la production de ces éléments³.

Le Tableau 35 ci-après récapitule les âges auxquels ces différents traits sont utilisés pour la première fois dans le corpus par les vingt enfants. Les quatre premières colonnes renvoient aux traits [Personne] et [Locuteur]. Elles rendent compte des deux cas de figure constatés : (i) l'enfant, lors de sa première séance, utilise déjà P1, P2 et P3 concomitamment ; l'âge de l'enfant à cette date est alors indiqué dans la quatrième colonne et nos données ne permettent pas d'étudier l'émergence de ces traits, les enfants étant déjà trop âgés, ou (ii) l'enfant, lors de sa première séance, n'utilise pas les trois personnes concomitamment ; les trois premières colonnes rendent alors compte de l'émergence des traits. Quentin, par exemple, utilise *je*, *tu*, *il* et *ce* lors de sa première séance lorsqu'il a 3;5.29 ; les trois premières colonnes sont donc vides. Par contre, Eva, lors de sa première séance à 3;3.5, utilise *tu*, *il*, *ce* et *on*, donc uniquement P2 et P3 ; *je* apparaît lors de la séance suivante à 3;3.19 qui compte également les clitiques *on*, *ce* et *vous* ; les troisième et quatrième colonnes illustrent cette évolution. Les trois colonnes suivantes rendent

¹ Dans le cadre du débat sur la valeur initiale des traits, il peut également être postulé que ces traits n'ont, par défaut, pas de valeur initiale ou que leur valeur première est une valeur non-marquée dans le système.

² Cf. Section 3.4 ci-après pour le détail.

³ Requête sur le modèle : `freq +t*ALA +t%mor +s"pro*:nom*|*" +u.`

compte de l'émergence d'une opposition aux niveaux du [Genre] (avec les clitiques *elle* et *elles*), du [Nombre] (*vous, ils* et *elles*¹) et de la [Référentialité] (*il explétif*). La dernière colonne indique l'âge de l'enfant lors de sa dernière séance dans le cas où tous les traits n'ont pas été utilisés. Ainsi, Léa à 3;9.10, n'a produit aucun clitique pluriel non doublé lors de ses trois séances d'enregistrement qui comptent un total de 41 clitiques de ce type.

	[Ø] P3	[Pers] P1+P3	[Pers] P2+P3	[Loc] P1+P2+P3	[Genre]	[Nbre]	[Réf]	Age fin
QUE				3;5.29	3;9.9	3;7.1	3;9.9	
LAU				3;6.23	3;10.15	3;10.15	3;8.17	
LEA				3;4.28	3;4.28		3;4.28	3;9.10
EVA			3;3.5	3;3.19	3;5.13	3;3.19	3;3.19	
EKT				3;3.14	3;3.14	?3;3.14	3;3.14	
CLA		3;1.27			3;2.10	3;2.10		3;5.29
TEO				3;1.15	3;1.15	3;5.3	3;7.15	
ISI		3;2.13	3;2.28	3;4.7			3;2.13	3;6.30
TOM	2;11.9			2;11.23	2;11.9	2;11.23	2;11.9	
JUL	2;11.8			3;3.10	3;3.10		3;5.8	3;6.16
CEL			2;11.20	3;4.2	3;4.2	2;11.20	3;5.20	
LUS				2;9.22	2;10.5	2;10.5	2;10.5	
NOE				2;9.29	2;11.23	2;9.29	2;11.23	
ANT				2;9.13	2;9.13	3;0.22	2;9.13	
KLE		3;0.0			3;0.17	3;0.17		3;3.22
ROM				2;8.29	2;11.6	3;2.29	2;11.6	
ALA				2;8.18	2;9.23	2;9.23	2;8.18	
EMA				2;8.0	2;9.5	2;11.9	2;9.5	
MAX		2;6.0		2;8.8	2;10.16			3;1.8
ZOE				2;5.5	2;5.19	2;5.19	2;7.28	

Tableau 35 : âges auxquels émergent les différentes oppositions (première version)

Les commentaires que nous pouvons faire, à ce stade de la recherche, sont que : (i) lorsqu'il n'y a qu'une seule personne dans le système d'un enfant, il s'agit de la P3 (*cf.* Tom et Julien) ; (ii) lorsque deux clitiques émergent concomitamment, P3 est également toujours présente, opposée soit à P1 (*cf.* Clara, Isia, Claire et Max) soit à P2 (*cf.* Eva, Isia, Célia) ; les données d'Isia présentant la particularité d'opposer les personnes par paire (P1 vs P3 puis P2 vs P3) avant d'intégrer les trois personnes concomitamment ; (iii) les autres oppositions émergent ensuite progressivement mais certains enfants (Léa, Clara, Isia, Julien, Claire et Maxime) terminent cette campagne d'enregistrements sans avoir utilisé toutes les oppositions. Nous ne commentons pas ces résultats davantage car les données présentées dans le Tableau 35 sont, pour le

¹ Nous n'est pas mentionné car il est totalement absent du corpus.

moment, parcellaires. En effet, comme le montrent le Tableau 29 et le Tableau 30, celles-ci ne rendent compte que d'un seul type d'élément. Complétons donc ce premier état des lieux avec les autres clitiques du corpus que nous appelons les clitiques doublés.

2. Les clitiques doublés

Les clitiques peuvent être doublés à leur gauche, à leur droite ou de part et d'autre concomitamment. Ces différentes possibilités sont illustrées en (7) ci-dessous :

(7) Les différentes configurations du doublement du clitique nominatif :

- (a) ALA: *mon papa i(l)* vient me chercher +... (I.2, 475)
- (b) EMA: *elle* est là *ma photo* . (I.1a, 171)
- (c) TOM: ben *moi j'* ai fini ça . (I.2, 1336)
- (d) LEA: *on* voit Lucie *nous* . (IX.37d, 258)
- (e) ANT: *moi je* fais la fourmi *moi* . (I.1a, 150)

Néanmoins, dans notre base, tous les clitiques doublés présentent un corps de code en commun, comme le montre la liste exhaustive des clitiques de cette catégorie présentée dans le Tableau 36 :

1	pro:cli:d:nom je&1S	8	pro:cli:d:nom ça&3S
2	pro:cli:d:nom tu&2S	9	pro:cli:d:nom nous&1P
3	pro:cli:d:nom:post tu&2S	10	pro:cli:d:nom vous&2P
4	pro:cli:d:nom il&3S	11	pro:cli:d:nom:pol vous&2P&POL
5	pro:cli:d:nom elle&3S	12	pro:cli:d:nom ils&3P
6	pro:cli:d:nom on&3S	13	pro:cli:d:nom elles&3P
7	pro:cli:d:nom ce&3S		

Tableau 36 : liste exhaustive des clitiques doublés

Cette liste de treize codes n'est pas le pendant exact de la liste des clitiques non doublés pour deux raisons : (i) les formes impersonnelles ne sont jamais doublées et (ii) nos données ont conduit à ajouter un code afin de rendre compte de la présence du clitique *tu* en position postverbale, mais uniquement sous sa forme doublée. Cette forme du clitique survient en effet deux fois dans ce corpus :

(8) Les deux occurrences de « pro:cli:d:nom:post|tu&2S » en contexte :

(a) ALA: *poisson où es tu ?* (VI.24a, 147)

ALA: t(u) es caché ?

ALA: ah je <t' ai> [=? te] trouvé: !

(b) CLA: *où es tu Maman_Ours ?* (VII.30b, 1027)

KAT: qu+est+c(e)+qu' i(l) fait le petit ours ?

ISI: i(l) [/] i(l) met ses chaussures .

Dans le premier cas, Alan (2;11.27) vient de vider la boîte de cartes sur la table et cherche la carte du poisson en s'adressant directement à elle. Dans le deuxième cas, Clara (3;5.29) et Isia regardent un livre et le commentent. La différence principale entre les occurrences du *il* explétif postverbal mentionnées précédemment¹ et les deux clitiqes postverbaux présents dans les énoncés en (8) ci-dessus est que ces derniers apparaissent dans des énoncés grammaticaux. Néanmoins, le très faible nombre d'occurrences et le fait que ces deux clitiqes ne sont pas produits par le même enfant ne permettent pas de trancher entre l'hypothèse de la répétition et celle de la computation. Le constat demeure donc identique à celui formulé pour les clitiqes non doublés : globalement, les jeunes enfants ne manipulent pas de clitiqes nominatifs postverbaux et les adultes non plus.

2.1 L'ensemble du corpus

Ces deux clitiqes postverbaux font partie d'un total de 5196 clitiqes doublés². Les clitiqes doublés sont extraits grâce à la requête directe détaillée ci-après :

(9) Requête directe pour l'extraction des clitiqes nominatifs doublés :

```
freq -t*ELO -t*LOL -t*UNI +t%mor +s"pro:cli:d:nom*|*" +u @ >  
CliD_Freq_All
```

Cette requête indique que onze formes différentes seulement ont été transcrites sur les treize formes possibles. Très logiquement, les deux formes non utilisées pour les clitiqes doublés sont les mêmes que pour les clitiqes non doublés puisqu'il s'agit

¹ Cf. Section II.1.1.

² Leur taux d'apparition est donc de 0,038 %.

par un clitique (quasiment toujours) préverbal. Deux champs d’investigation nous semblent alors intéressants : la proximité syntaxique (l’adjacence) et la proximité phonologique (l’ajustement) des clitics nominatifs et des verbes fléchis. Nous reviendrons sur ces deux thématiques dans la Section II.4 ci-dessous. Pour l’heure, comme nous disposons à présent de statistiques complètes, tant sur les clitics non doublés que sur les clitics doublés, leur évolution en termes de traits, esquissée dans le Tableau 35 pour la première catégorie, peut être complétée.

3. L’évolution des clitics en termes de traits

Le Tableau 38 présente la répartition des clitics doublés au sein du système enfantin. Il permet de constater que cette répartition est quasiment identique à celle des clitics non doublés en termes de [Nombre] et qu’aucun pourcentage ne varie drastiquement en termes de [Personne].

Formes	[Pers]	N/[Pers]	%/[Pers]	[Nbre]	N/[Nbre]	%/[Nbre]
pro:cli:d:nom je&1s	1	1069	36,89	sg	2821	97,34
pro:cli:d:nom tu&2s	2	98	3,38			
pro:cli:d:nom post tu&2s						
pro:cli:d:nom il&3s	3	1654	57,07			
pro:cli:d:nom elle&3s						
pro:cli:d:nom on&3s						
pro:cli:d:nom ce&3s						
pro:cli:d:nom ça&3s	5	6	0,21	pl	77	2,66
pro:cli:d:nom vous&2p						
pro:cli:d:nom ils&3p						
pro:cli:d:nom elles&3p	6	71	2,45			
Totaux		2898	100,00		2898	100,00

Tableau 38 : répartition des clitics nominatifs doublés dans le système enfantin

Nous pouvons à présent étudier l’introduction de chacune de ces formes dans le système de chaque enfant afin de compléter les données du Tableau 35 ; nous intégrons donc les âges d’émergence de ces clitics doublés à nos premières constatations, ce qui modifie légèrement les résultats. La synthèse est présentée dans le Tableau 39 ci-après.

	[Ø] P3	[Pers] P1+P3	[Pers] P2+P3	[Loc] P1+P2+P3	[Genre]	[Nbre]	[Réf]	Age fin
QUE				3;5.29	3;5.29	3;7.1	3;9.9	
LAU				3;6.23	3;8.17	3;10.15	3;8.17	
LEA				3;4.28	3;4.28		3;4.28	3;9.10
EVA				3;3.5	3;3.19	3;3.19	3;3.19	
EKT				3;3.14	3;3.14	3;3.14	3;3.14	
CLA		3;1.27			3;2.10	3;2.10		3;5.29
TEO				3;1.15	3;1.15	3;5.3	3;7.15	
ISI				3;2.13	3;2.13		3;2.13	3;6.30
TOM		2;11.9		2;11.23 (2)	2;11.9	2;11.9	2;11.9	
JUL	(2;11.8)			3;3.10	3;3.10		3;5.8	3;6.16
CEL				2;11.20	2;11.20	2;11.20	3;5.20	
LUS				2;9.22	2;10.5	2;10.5	2;10.5	
NOE				2;9.29	2;9.29	2;9.29	2;11.23	
ANT				2;9.13	2;9.13	3;0.22	2;9.13	
KLE		3;0.0			3;0.17	3;0.17		3;3.22
ROM				2;8.29	2;11.6	3;2.29	2;11.6	
ALA				2;8.18	2;9.23	2;9.23	2;8.18	
EMA				2;8.0	2;8.0	2;11.9	2;9.5	
MAX		2;6.0		2;8.8 (3)	2;7.5			3;1.8
ZOE				2;5.5	2;5.19	2;5.5	2;7.28	

Tableau 39 : âges auxquels émergent les différentes oppositions (deuxième version)

3.1 Les traits [Personne] et [Locuteur]

Le Tableau 39 ci-dessus illustre que l'utilisation de la P3 seule subsiste chez un enfant seulement (Julien, 2;11.8). Cependant, au cours de cette séance, Julien ne produit qu'un seul énoncé verbal sur un total de douze énoncés (JUL : *c' est jaune*, I.1a). L'absence d'opposition peut donc provenir d'une absence absolue des autres personnes dans le système de Julien à cet âge mais elle peut tout aussi bien être imputée au fait que nous ne disposons que d'un énoncé verbal unique. Ce fait est donc noté dans le Tableau 39 mais les parenthèses marquent la réserve quant à la fiabilité des données. Une opposition initiale P1 vs P3, impliquant donc la valeur ± du trait [Personne] est observée chez quatre enfants dont Tom qui marque la P3 grâce à des clitiques doublés et non doublés et la P1 uniquement grâce à la forme doublée du clitique. La seconde opposition possible basée sur le trait [Personne] (P2 vs P3) disparaît lorsque l'on étudie à la fois les clitiques non doublés et les clitiques doublés dans la mesure où Eva, Isia et Célia utilisent des clitiques non doublés de P2 et de P3 mais également des clitiques doublés de P1. Cette opposition n'est donc finalement pas attestée en tant qu'opposition initiale unique et la colonne correspondante se

vide. Par contre, les données de Tom et de Maxime indiquent que P2 apparaît après P1 et P3 (trois occurrences de *tu* sur un total de 129 clitiques pour Tom au cours de sa deuxième séance et deux occurrences sur un total de 86 clitiques pour Maxime au cours de sa troisième séance) alors que Clara et Claire n'introduisent pas du tout P2 dans leurs productions au cours des enregistrements subséquents (sur un total de 38 énoncés pour Clara et 98 énoncés pour Claire). Les autres enfants manipulent les trois personnes dès leur première séance ; leurs données n'apportent donc pas d'informations supplémentaires quant à l'ordre d'apparition des traits [Personne] et [Locuteur]. L'ensemble de ces observations confirment néanmoins les remarques formulées dans Palasis (2005) en acquisition et les hypothèses proposées par Oliuéri (2009) en diachronie sur : (i) une émergence progressive et séquentielle des traits morphologiques portés par les clitiques ; (ii) l'existence d'une opposition initiale basée sur la valeur \pm du trait [Personne] qui, chez les enfants, différencie les énoncés de P1 et de P3 et qui, dans les dialectes, marque la distinction entre les énoncés de P1 et de P3 ou les énoncés de P2 et de P3 ; (iii) l'introduction subséquente de la valeur \pm du trait [Locuteur] lorsqu'il s'agit de distinguer les trois personnes utilisées concomitamment au sein d'un même système linguistique. La séquence [Personne] > [Locuteur] peut donc être supposée.

3.2 Le trait [Genre]

L'utilisation de la valeur \pm pour le trait [Genre] est déjà en vigueur lors de la première séance pour la moitié des enfants. Pour l'autre moitié (Laura, Eva, Clara, Julien, Lucie, Claire, Romane, Alan, Maxime et Zoé), cette opposition apparaît donc après celles de la [Personne] et du [Locuteur]¹ avec toutefois une exception à cet ordre : celle de Tom qui produit deux occurrences de *elle* (sur un total de 46 clitiques) lors de sa première séance, avant de produire les trois occurrences du clitique *tu* au cours de sa deuxième séance. Globalement, la séquence suivante peut donc être proposée : [Personne] > [Locuteur] > [Genre].

Cette absence initiale d'opposition pour le trait [Genre] dans la série des clitiques chez différents enfants soulève un certain nombre de questions. En premier lieu, nous nous demandons notamment si cette absence est limitée à la série des clitiques ou si cette opposition est totalement absente du système enfantin. L'étude

¹ Au cours de leur deuxième séance, sauf pour Alan qui commence à produire des clitiques féminins au cours de sa troisième séance (cinq occurrences de *elle* sur un total de 117 clitiques).

de la première séance de Laura, Eva, Clara, Julien, Lucie, Claire, Romane, Alan, Maxime et Zoé met en lumière plusieurs phénomènes¹. Tout d'abord, deux enfants (Julien et Romane) n'utilisent pas du tout l'opposition de genre lors de leur première séance². Romane produit en effet seize énoncés (équivalant à 60 morphèmes) dans lesquels seuls quatre noms masculins apparaissent ; elle utilise ainsi beaucoup le pronom *ça* comme le montrent les exemples en (10a) ci-dessous. Julien quant à lui produit douze énoncés quasiment tous mono-morphématiques (dix-sept morphèmes en tout) dont huit adjectifs épiciènes et deux noms masculins (*cf.* (10b)).

(10) Deux cas d'absence totale d'opposition de [Genre] :

- (a) ROM: je veux faire ça . (I.3b, 412)
 ROM: qu'est-ce que c' est ça ? (I.3b, 523)
- (b) JUL: marron ! (I.1a, 239)
 JUL: un soleil . (I.1a, 470)

La nature des énoncés de Romane et de Julien montre ainsi que l'utilisation des noms n'est pas prégnante dans leur discours. Ce fait a alors pour conséquence que l'opposition du trait [Genre] peut rester inactive dans le système de ces deux enfants dans la mesure où, lorsqu'il n'y a pas de noms, il n'y a pas non plus de déterminants ni d'adjectifs, qui sont, à part les pronoms, les principales catégories grammaticales sur lesquelles on s'attend à constater l'opposition de [Genre]. Chez Romane, cette opposition est présente lors de sa deuxième séance sous plusieurs formes³ : le clitique *elle* et les déterminants *la* et *ma*. Chez Julien, l'opposition est également visible lors de sa deuxième séance sous les formes *une*, *la*, *ma*, *sa* et *elle*⁴. L'étude des occurrences de Romane et de Julien ne répond cependant pas à une question importante. En effet, à ce stade, leurs énoncés ne permettent pas d'établir l'origine de l'absence de cette opposition. En d'autres termes, nous aimerions déterminer s'il s'agit simplement d'un hasard que ces deux enfants n'utilisent pas le trait [Féminin] ou si les données permettent de cerner un réel mécanisme sous-jacent. Nous nous tournons donc vers la première séance des huit enfants dont les

¹ Requêtes : combo +t*EVA +t*CLA +t*JUL +t*LUS +t*ROM +t*ALA +t*MAX +t*ZOE +t%mor +s**|*FEM**" +u @ > FEM1 (séance 1) ; combo +t*LAU +t%mor +s**|*FEM**" +u @ > FEM1_LAU (séance 4) ; combo +t*KLE +t%mor +s**|*FEM**" +u @ > FEM1_KLE (séance 5).

² Requêtes : mlu +t*XXX +u @ et kwal +t*XXX +u @.

³ Deux mois séparent les deux séances ; ceci est dû à l'absence de Romane à une séance intermédiaire et au fait que tous les enfants ne sont pas tous enregistrés à toutes les séances.

⁴ Quatre mois séparent les deux séances pour les mêmes raisons que Romane.

productions illustrent la séquence [Personne] > [Locuteur] > [Genre] pour les clitiques.

Chez ces enfants, sur un total de 51 morphèmes [Féminin], 24 sont effectivement phonétiquement marqués pour ce trait (*cf.* (11) ci-dessous). Ces marques se trouvent essentiellement sous la forme de déterminants définis, indéfinis ou possessifs (*cf.* (11a)) et deux transcriptions incertaines comprennent un adjectif (*cf.* (11b)). Les autres occurrences extraites comprennent également un morphème [Féminin] ; cependant la neutralisation du genre au pluriel implique que le morphème n'est pas phonétiquement marqué pour le genre dans ces cas (*cf.* (12)) ; ils ne représentent donc pas une opposition explicite avec le trait [Masculin] comme dans le cas précédent des déterminants *la vs le* ou *ma vs mon*.

(11) Exemples d'occurrences avec [Féminin] phonétiquement marqué :

- (a) LAU: ben moi j' ai *la* rose ! (IV.19b, 1022)
LAU: *une* souris . (IV.19b, 1381) (x2)
LUS: et j' ai fini *ma* ligne ! (I.3d, 612)
EVA: alors i(l) faut pas *toutes* les mêmes couleurs . (I.3c, 859)
- (b) ZOE: c' est *verte* [?] et ça c' est bleu . (I.3b, 121) (x2)

(12) Exemples d'occurrences avec [Féminin] phonétiquement non marqué :

- (a) LAU: pour aller à l' école . (IV.19a, 630)
- (b) LAU: i(l) met *ses* chaussures . (IV.19c, 379)
- (c) ALA: *deux* cornes . (I.1b, 76)

Ces quelques exemples indiquent que l'opposition de [Genre] est présente dans le système de ces huit enfants, particulièrement dans le cas des déterminants, même si c'est le trait [Masculin] qui domine largement¹. Cependant, ces premières constatations ne permettent toujours pas de trancher entre hasard et mécanisme linguistique sous-jacent. Un phénomène cependant permettrait d'étayer la deuxième hypothèse. En effet, si l'opposition de [Genre] était dans un premier temps inactive et émergeait dans un second temps dans le système enfantin, on s'attendrait à trouver, au cours de la première étape, des formes qui devraient être marquées pour le [Féminin] mais qui ne le sont pas. En d'autres termes, le corpus devrait comprendre

¹ 303 occurrences de MASC pour la même séance des mêmes enfants, ce qui représente 85,60 % des marques de [Genre] dans le corpus alors que le taux de noms masculins dans le lexique est de 56 % (*cf.* Roché (1992)).

des clitiques [Masculin] à la place de clitiques [Féminin] dans une proportion décroissante. Comme souvent en acquisition, ce sont donc les agrammaticalités enfantines qui peuvent mettre au jour un mécanisme sous-jacent.

Le codage %err de la base de données comprend une série de codes qui se rapporte à une substitution du [Genre]. Chaque code étant le reflet d'au moins une agrammaticalité effective dans la base, la liste reprise en (13) ci-dessous indique, d'une part, qu'une telle substitution fait partie des productions enfantines et, d'autre part, que cette substitution touche à la fois les pronoms, les noms, les déterminants, les adjectifs et les participes.

(13) Liste des codes %err ayant pour objet la substitution du trait [Genre] :

- (a) Pronom : \$MOR \$SUB \$AGA \$PRO
- (b) Nom : \$MOR \$SUB \$AGA \$N
- (c) Déterminant : \$MOR \$SUB \$AGA \$DET
- (d) Adjectif : \$MOR \$SUB \$AGA \$ADJ
- (e) Participe : \$MOR \$SUB \$AGA \$PART

Appliquée à la première séance des huit enfants qui ne produisent pas de clitiques [Féminin] à ce stade, la requête correspondante¹ fait émerger l'occurrence reprise en (14a) ci-dessous dont la comparaison avec l'énoncé (14b) est intéressante.

(14) L'absence du trait [Féminin] sur les clitiques d'Alan :

- (a) ALA: +, ma maman *i(l)* [*] vient me chercher . (I.2, 481)
- (b) ALA: mon papa *i(l)* vient me chercher +... (I.2, 475)

En effet, cette (quasi) paire minimale permet de constater que l'opposition de [Genre] est présente chez Alan sur le déterminant mais qu'elle n'apparaît pas sur le clitique. Sur les séances suivantes, Alan ne reproduit ce type d'erreur que trois fois (séances 2, 5 et 12), ce qui représente des proportions beaucoup trop basses pour établir une quelconque évolution. En revanche, si nous regardons l'ensemble des substitutions de [Genre] chez tous les enfants, nous constatons que, dans la quasi-totalité des cas, la substitution s'opère dans ce même sens, c'est-à-dire que la forme masculine apparaît à la place de la forme féminine attendue et que l'opposition de

¹ Requête du type combo +t*XXX +t%err +s"\$MOR^\$SUB^\$AGA^\$PRO" +u > Sub_Gender_Pro_I appliquée aux locuteurs et séances adéquates.

[Genre] des clitiques semble être maîtrisée plus tardivement que celles ayant un rapport avec le nom¹ :

(15) Quelques autres exemples de substitution du trait [Genre] :

- (a) ANT: euh la caréma [: caméra] [*] on *le* [*] tourne ! (V.22, 954)
- (b) MAX: oh *le* [*] grenouille *il* [*] est [?] là ! (IV.19c, 368)
- (c) LUS: ah non *i(ls)* [*] sont pas longues . (X.38c, 1120)

Vu leur nombre extrêmement limité, les occurrences d'Alan, présentées en (14) ci-dessus, ne constituent pas des preuves irréfutables de la séquence [Personne] > [Locuteur] > [Genre] proposée concernant l'activation des traits morphologiques dans le système enfantin. Néanmoins, il peut être considéré que l'absence initiale d'opposition pour le [Genre] chez dix enfants et les agrammaticalités constatées (qui privilégient largement la forme masculine à la forme féminine) constituent des indices en faveur d'une telle hypothèse ou que, tout au moins, ces éléments ne la falsifient pas.

3.3 Le trait [Nombre]

Sur les vingt enfants du corpus, cinq utilisent l'opposition de [Nombre] sur les clitiques dès la première séance et quatre ne l'utilisent pas du tout. Nous nous tournons donc vers les données des onze enfants qui introduisent l'opposition de [Nombre] au cours des enregistrements. Il s'agit de Quentin, Laura, Eva, Clara, Téo, Lucie, Antoine, Claire, Romane, Alan et Emma. La proportion d'enfants qui fait émerger ce trait après ceux de [Personne] et de [Locuteur] est donc similaire à celle constatée pour l'émergence du trait [Genre]. Néanmoins, les deux listes d'enfants ne se superposent pas exactement. Voyons si nous pouvons situer ce nouveau trait dans la séquence [Personne] > [Locuteur] > [Genre].

Les âges reportés dans le Tableau 39 montrent que l'émergence de l'opposition de [Nombre] n'est pas fortement corrélée à l'émergence d'un autre trait. Néanmoins, comme pour le trait [Genre], les données livrent des tendances à exploiter. Tout d'abord, l'émergence de [Nombre] n'est pas obligatoirement concomitante ni avec celle de [Genre] ni avec celle de [Référence], même s'il semble y avoir un lien plus

¹ Une étude approfondie de l'évolution de cette substitution pour chacune des catégories grammaticales chez chaque enfant pourra être entreprise ultérieurement par le biais de la ligne %err mais une telle investigation dépasse le cadre de la recherche actuelle.

étroit entre les deux premiers. En effet, [Nombre] et [Genre] apparaissent dans la même séance chez cinq enfants alors que [Nombre] et [Référence] ne sont concomitants que chez deux enfants. [Genre] et [Nombre] sont donc parfois concomitants et parfois subséquents ; la séquence [Genre] > [Nombre] est donc possible mais non certaine. Une règle absolue peut néanmoins être tirée de ces données : chez aucun des onze enfants étudiés dans ce cadre, [Nombre] n'apparaît avant [Genre].

Comme pour le trait [Genre], le reste du corpus de ces onze enfants est examiné afin de déterminer si l'opposition de [Nombre] est absente uniquement dans la série des clitiques ou si celle-ci est également absente dans le reste du système enfantin. La requête correspondante¹ indique globalement, comme pour l'opposition de [Genre], que le [Nombre] est présent chez les enfants, morphologiquement visible notamment sur les déterminants (*cf.* (16a) ci-dessous), certains quantifieurs (*cf.* (16b)) et certains noms (*cf.* (16c)). Seules les données de Quentin ne montrent pas une utilisation du trait [Pluriel] sur ce type d'éléments antérieure à son application aux clitiques dans la mesure où les deux émergent au cours de la même séance.

(16) Exemples d'occurrences avec [Pluriel] morphologiquement marqué :

- (a) ALA: *deux* cornes . (I.1b, 76)
 LUS: i(l) mange *des* noisettes . (I.2, 1394)
 EVA: qui+c+est+qui veut *les* étoiles ? (I.3a, 644)
- (b) EMA: regarde y+a *beaucoup de* bébés . (I.2, 1962)
 EVA: alors i(l) faut pas *toutes* les mêmes couleurs . (I.3c, 859)
- (c) LAU: ferme les *yeux* Hector ! (IV.19b, 491)
 LAU: +, c' est un *animaux* [*] . (IV.19c, 202)

Voyons maintenant si les enfants produisent des énoncés dans lesquels le trait [Singulier] est utilisé à la place du trait [Pluriel]. La liste des codes des agrammaticalités effectivement réalisées dans la base couvre plusieurs possibilités, reprises en (17) ci-dessous accompagnées d'exemples.

¹ Requête sur le modèle : `combo +t*XXX +t%mor +s"*|*PL*" +u @ > PL1.`

- (17) Liste des codes %err ayant pour objet la substitution du trait [Nombre] :
- (a) Syntagme nominal : \$MOR \$SUB \$NUM \$N
LUS: c' est *ta meilleure couleur* [*] le rouge et le bleu ? (IX.34b, 1426)
 - (b) Adjectif : \$MOR \$SUB \$NUM \$ADJ
ISI: trois *petit* [*] ours . (VII.30a, 400)
 - (c) Clitique et verbe : \$MOR \$SUB \$NUM \$PRO \$V
ANT: <*il est*> [*] (= ils sont) beaux ! (III.13b, 407)
 - (d) Déterminant par rapport à nom : \$MOR \$SUB \$DET \$NUM
NOE: moi je l' ai *L_Aristochats* [*] ! (IX.37a, 179)
 - (e) Nom par rapport à déterminant : \$MOR \$SUB \$AGC \$N
LAU: +, c' est un *animaux* [*] . (IV.19c, 202)
 - (f) Clitique et verbe par rapport à SN : \$MOR \$SUB \$AGC \$PRO \$V
CEL: tu sais 0 [*] les [/] les cloches i(l) [/] <*il est venu*> [*] chez moi xx
pour [/] pour [*] des Barbies xx . (IX.37c, 70)

Globalement le schéma est donc le même que pour l'opposition de [Genre] : la substitution s'opère quasiment toujours dans le même sens. En l'occurrence, les enfants produisent des éléments qui ne portent pas le trait [Pluriel] là où la forme attendue devait porter un tel trait. Les deux codes \$PRO \$V, en (17c) et (17f) ci-dessus, sont intéressants dans la mesure où ils indiquent que, si l'opposition n'est pas marquée sur l'un de ces deux éléments, elle n'est pas marquée sur l'autre non plus. Dans ces cas, l'accord se fait donc entre le clitique et le verbe, indépendamment soit du contexte (*cf.* (17c)) soit du syntagme nominal co-indiqué (*cf.* (17f)). Cette étude permet ainsi d'insérer le trait [Nombre] après le trait [Genre], ou au plus tôt concomitamment à ce dernier, ce qui donne la séquence suivante : [Personne] > [Locuteur] > [Genre] > [Nombre].

3.4 Le trait [Référentiel]

Le trait [Référentiel] renvoie à une opposition qui concerne le contenu sémantique des clitiques et notamment du clitique *il*. En effet, celui-ci peut être [+Référentiel] et ainsi renvoyer à un signifié connu par ailleurs soit grâce au contexte soit grâce à un élément co-indiqué, comme dans l'énoncé *il_i est pas là l'oiseau_i* (ROM, IV.19f). Mais le clitique *il* peut également être explétif, c'est-à-dire qu'il ne possède

pas de contenu sémantique et sa réalisation morphologique est alors exigée par le trait *EPP*¹, comme illustré en (18) ci-dessous :

(18) Quelques exemples d'utilisations du clitique explétif *il* :

- (a) TEO: *i(l)* faut dormir ! (XI.43a, 86)
- (b) EKT: *i(l)* pleut des cornes [*] . (VI.26b, 1005)
- (c) TOM: *i(l)* fait chaud . (I.2, 768)
- (d) ANT: regarde *i(l)* manque un bout . (XI.42, 2256)
- (e) LUS: moi *i(l)* m(e) reste le papillon le p(e)tit poussin et l' aigle . (XI.43b, 2296)
- (f) EVA: regardez c(e)+qu' *i(l)* y+a ici . (XIII.47a, 240)

L'utilisation de ce clitique en particulier dans un système linguistique est donc un indice important lors du classement des langues. Aux Chapitres I et II, nous avons en effet passé en revue différentes corrélations de propriétés des systèmes linguistiques et les discussions subséquentes concernant le Paramètre *pro-drop*. Certains concluent actuellement que « seule la corrélation entre le sujet nul et l'absence de sujets explétifs peut être maintenue sans exception »². Dans un premier temps, nous étudions donc la réalisation de l'explétif *il* dans le système enfantin afin de déterminer si l'opposition de [Référentialité] peut être intégrée à notre séquence³.

Sur les vingt enfants, cinq utilisent l'opposition [\pm Référentiel] dès la première séance et trois enfants ne l'utilisent pas du tout. Il nous reste donc les données de douze enfants pour répondre à nos interrogations. De nouveau, les informations quant à la séquence des traits n'est pas tranchée. Cependant, une tendance se dégage dans la mesure où six enfants sur douze (Quentin, Téo, Julien, Célia, Noémie, Zoé) font émerger cette opposition dans leur système en dernier, après avoir déjà utilisé les oppositions de [Genre] et de [Nombre]. Les autres enfants utilisent cette opposition soit en même temps que celles de [Genre] et de [Nombre] (Eva, Tom, Lucie), soit uniquement avec celle de [Genre] (Laura, Romane), soit entre les deux (pour Emma seulement). Un seul code %err est susceptible de renvoyer une substitution agrammaticale de ce clitique explétif : il s'agit du code \$MOR \$SUB

¹ Cf. Chapitre II, Section II.1.1.

² Cf. Haider (2001:285) cité par Kaiser (2009:150).

³ L'étude de l'absence de cet élément se fera dans la Section VI.2.

\$PRO \$SUBJ¹. Ce type d'agrammaticalité est présent dix fois dans le corpus, comme illustré en (19) ci-dessous :

(19) Extraction de \$MOR \$SUB \$PRO \$SUBJ :

- (a) il = je : TOM: i(l) [*] parle moi . (II.9b, 723) (x3)
- (b) il= c' : MAX: il [*] est où ça ? (X.40d, 1372)
- (c) je = tu: MAX: je [*] peux donner la boîte ? (IV.19b, 1319)
- (d) tu = je : QUE: tu [*] peux prendre ça Hector ? (VIII.33a, 735)
- (e) c' = elle: EKT: c' [*] est ici Célia . (II.9a, 105)
- (f) ça = on : ANT: toi un chat et nous ça [*] dirait des garçons . (IX.37c, 1139) (x2)
- (g) tu ? = il: LUS: tu [*] [?] pleut ! (II.5, 1794)

Seuls (19f) et (19g) impliquent des clitiques [-Référentiel] mais (19g) est noté en tant que transcription incertaine. Dans ce cas de figure, contrairement aux autres traits, la substitution de [Référentialité] ne semble donc pas attestée.

3.5 Synthèse

Le corpus donne des indications qui nous permettent de situer l'émergence de l'utilisation du trait [Référentialité] dans la séquence existante. En effet, chez les douze enfants étudiés, cette opposition n'apparaît qu'une seule fois avant celle du [Nombre] et elle n'apparaît jamais avant celle du [Genre]. Dans un premier temps, cette observation étaye donc l'hypothèse sur la séquence [Genre] > [Nombre]. Dans un deuxième temps, elle permet de proposer la séquence complète détaillée en (20) ci-dessous :

(20) Émergence des clitiques nominatifs d'après leurs traits morphologiques :

[Personne] > [Locuteur] > [Genre] > [Nombre] > [Référentialité]
 je/il/c'/ça/on > tu > elle > ils/elles/vous > il expl.

L'observation des clitiques à travers les traits morphologiques qui les composent permet ainsi de proposer un modèle global d'émergence des clitiques nominatifs dans le système enfantin². Il est extrêmement intéressant de noter l'identité qui existe entre ce modèle acquisitionnel et le modèle diachronique proposé par Oliiviéri

¹ Requête correspondante : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"\$MOR^\$SUB^\$PRO^\$SUBJ" +u @ > Sub_Expl.

² L'émergence des clitiques objets pourrait également être étudiée d'après ce modèle.

(2009). En effet, si la séquence est la même chez l'enfant et dans les langues en général, cela signifie que nous touchons non pas à un mécanisme de paramétrage (qui montrerait des différences d'une langue à une autre) mais à un phénomène bien plus profond : cette séquence relèverait alors d'un principe issu de notre grammaire commune, la Grammaire Universelle.

4. Ce que les clitiques nous disent sur leur statut morphosyntaxique

Dans le cadre du débat sur le statut des clitiques nominatifs, l'un des critères repris dans la littérature est celui de Zwicky & Pullum (1983) sur la nature des chaînes d'éléments clitiques. Ainsi, concernant l'étude du statut de la négation *n't* en anglais, les auteurs proposent que « les clitiques peuvent s'attacher à du matériel qui contient déjà des clitiques, mais les affixes ne le peuvent pas »¹. En d'autres termes et dans le cadre de nos recherches, les clitiques nominatifs du français ne peuvent être considérés comme des affixes (et donc ressortir à la morphologie) que si, entre ces clitiques et les verbes fléchis, nous ne trouvons aucun élément argumental (c'est-à-dire qui relève de la syntaxe). De Cat (2005) utilise cet argument pour démontrer que les clitiques nominatifs du français ne peuvent pas être considérés comme des affixes morphologiques. À cette fin, l'auteur étudie les éléments *ne*, *en*, *y* et les clitiques accusatifs et datifs.

Tout d'abord, il convient d'attirer l'attention sur le fait que les critères proposés par Zwicky & Pullum (1983) sont principalement établis sur la base de l'étude de la morphosyntaxe de l'anglais. Ainsi, depuis 1983, des recherches dans d'autres langues ont montré que ce critère pouvait être discuté, notamment pour les langues romanes. Heggie & Ordóñez (2005) par exemple proposent des contre-exemples au postulat de Zwicky & Pullum (1983) provenant de l'espagnol des Caraïbes et du portugais². Ainsi les exemples en (21) ci-dessous montrent que le suffixe *-n* marquant la Personne 6 s'attache à un clitique réfléchi (*cf.* (21a)) et à un clitique datif (*cf.* (21b)), tous deux considérés comme des arguments syntaxiques et l'exemple (22) illustre que l'affixe du futur *-ei* s'attache au clitique argumental *me*.

¹ Zwicky & Pullum (1983:504) : "Clitics can attach to material already containing clitics, but affixes cannot".

² Les exemples de Heggie & Ordóñez (2005) proviennent de Harris (1995) pour l'espagnol et de Cunha & Cintra (1984) pour le portugais.

(21) Exemples en espagnol des Caraïbes :

- (a) *Váya-se-n*
aller -refl-P6
« ils s'en vont »
- (b) *De-le-n un libro*
donner-eux-P6 un livre
« ils leur donnent un livre »

(22) Exemple en portugais :

- Calar-me-ei*
se taire-reflP1-futur
« je me tairai »

Il semble donc que l'argument de Zwicky & Pullum (1983) puisse être discuté pour les langues romanes. Néanmoins, les gloses ci-dessus illustrent bien le fait qu'en français les clitiques nominatifs ne s'adjoignent pas toujours directement à leur hôte phonologique, le verbe fléchi, dans la mesure où un certain nombre de constituants peuvent venir s'intercaler entre les deux, comme le souligne De Cat (2005). Voyons donc ce que nous apprennent nos données enfantines sur ces éléments interférents du français que sont *ne*, *en*, *y* et les clitiques accusatifs et datifs.

4.1 L'étude des clitiques préverbaux non nominatifs

4.1.1 La particule négative « *ne* »

Le corpus montre que la particule négative *ne* n'est quasiment pas présente dans le discours enfantin à ce stade. En effet, celle-ci apparaît seize fois dans l'ensemble des données, ce qui est extrêmement peu au regard du nombre d'énoncés négatifs présentés dans le Tableau 40 ci-après :

Types de négations	<i>pas</i>	<i>plus</i>	<i>jamais</i>	Totaux	%
<i>X</i>	1164	126	5	1295	98,8
<i>ne X</i>	7	9	0	16	1,2
Totaux	1171	135	5	1311	100,0

Tableau 40 : la négation dans le corpus enfantin

Les seize énoncés qui comprennent la particule *ne* sont produits par quatre enfants différents. Les seize autres enfants n'utilisent donc jamais cette partie de la négation au cours de nos enregistrements. De plus, parmi ces seize énoncés, dix sont produits

par le même enfant (Emma) dont huit sont prononcés à l'occasion d'une histoire qu'Emma raconte en chantonnant (cf. (23a) ci-dessous) ; l'unique occurrence produite par Hector survient également dans le cadre des paroles d'une chanson et l'énoncé ne comprend pas de clitique nominatif (cf. (23b)), ce qui représente une deuxième caractéristique peu fréquente chez ces jeunes enfants¹. Lucie produit trois *ne* dont un dans un énoncé qui ne comporte pas non plus de clitique nominatif (cf. (23c)) et Eva utilise *ne* deux fois, dont une occurrence qui accompagne un verbe à l'impératif (cf. (23d)).

(23) La particule négative *ne* dans le corpus (16 occurrences) :

- (a) EMA: il ne chante plus . (XI.41d, 1404)
EMA: il ne saute plus . (XI.41d, 1411)
EMA: il ne danse plus . (XI.41d, 1418)
- (b) EKT: le facteur n' est pas passé . (IX.36, 2126)
- (c) LUS: ta photo n' est pas là ? (II.5, 145)
- (d) EVA: ne te cache pas avec la photo ! (XIII.46e, 1086)

La particule négative *ne* est donc non seulement très rare chez les jeunes enfants mais, de plus, elle survient neuf fois sur seize dans des énoncés qui ne représentent pas du véritable discours spontané. Par ailleurs, douze de ces occurrences surviennent au cours de nos trois dernières séances, ce qui, à notre avis, remet l'étude de l'utilisation productive de *ne* à l'analyse des données enregistrées les années suivantes auprès des mêmes enfants. Pour l'heure, nous concluons donc que la particule négative *ne* n'est pas productive dans la grammaire des enfants de cet âge. Par ailleurs, les données montrent que, lorsque la particule *ne* est produite, il arrive qu'aucun clitique nominatif ne soit présent. Ainsi, à ce stade, la présence concomitante d'un clitique nominatif et du clitique négatif dans treize énoncés en tout ne peut pas être retenue comme un argument dans le débat sur le statut des clitiques nominatifs, ni dans un sens ni dans un autre.

4.1.2 Les clitiques « y » et « en »

L'étude de ces deux clitiques invariables est souvent menée conjointement. Cela est dû en partie à leurs origines. En effet, *y* et *en*, souvent caractérisés d'« adverbes pronominaux » ou de « pronoms adverbiaux », sont tous les deux issus d'adverbes

¹ Ce type de sujet est détaillé en Section V ci-après.

latins : *ibi* « là » et *inde* « de là »¹. Par ailleurs, les grammaires les regroupent également sur la base du fait qu'il est traditionnellement postulé que ces pronoms remplacent des syntagmes prépositionnels, comme illustré en (24) ci-dessous : *y* est analysé comme tenant la place d'un syntagme introduit par *à*² (*cf.* (a)) et *en* comme tenant la place d'un syntagme introduit par *de* (*cf.* (b)).

(24) L'analyse traditionnelle de *y* et *en*³ :

- (a) Vas-tu à *Paris* ? Oui, j'*y* vais.
 (b) Vient-il *de la ville* ? Oui, il *en* vient.

Néanmoins, cette description est réductrice pour au moins deux raisons. D'une part, comme le soulignent Denis & Sancier-Chateau (1994), « *y* et *en* pronominalisent des compléments prépositionnels de statuts très divers »⁴ et, d'autre part, comme le développe Oliviéri (1994), ces pronoms ne remplacent pas uniquement des syntagmes *prépositionnels*.

4.1.2.1 *Les codages de « y » et « en »*

En ce qui concerne le pronom *en*, Oliviéri (1994) distingue quatre types de pronoms, comme illustré en (25) ci-dessous : le *en* « génitif » qui représente un syntagme nominal dans un syntagme nominal (*cf.* (a)), le *en* « ablatif » qui est un argument du verbe (*cf.* (b)), le *en* « circonstant » qui est un syntagme prépositionnel adjoint (*cf.* (c)) et le *en* « quantitatif » qui représente un nom (*cf.* (d))⁵ :

¹ *Cf.* Grevisse & Goosse (2008:868) qui signalent également que les étymologistes sont partagés au sujet de *y* dans la mesure où l'adverbe latin *hic* « ici » est aussi proposé comme origine pour *y*.

² Mais d'autres prépositions sont également possibles, telles que *sur*, *dans*, *chez*, etc.

³ Exemples extraits de Grevisse & Goosse (2008:869).

⁴ *Cf.* Denis & Sancier-Chateau (1994:31-36). En effet, les fonctions de ces pronoms sont diverses : complément circonstanciel, complément d'objet, complément du nom, du pronom, de l'adjectif.

⁵ *Cf.* Oliviéri (1994:212-244) pour le détail. Le *en* « ablatif » est traditionnellement analysé comme un *en* « prépositionnel ». L'auteur ne retient pas cette hypothèse dans la mesure où elle postule que la préposition n'est pas une « vraie » préposition mais un « simple marqueur de cas » ; c'est donc le verbe qui assigne le cas ablatif directement au pronom (lorsqu'il s'agit d'un argument non-animé uniquement).

(25) Les quatre types de *en* :

(a) « Génitif » :

Paul apprécie le style *de cet auteur*. Paul *en* apprécie le style.

(b) « Ablatif » :

Paul parle *de Marie*. Paul *en* parle.

(c) « Circonstant » :

Il est mort *de la peste*. Il *en* est mort.

(d) « Quantitatif » :

Je veux *une pomme*. J'*en* veux *une*.

Nous conservons cette partition et appliquons conséquemment les quatre codes détaillés en (26) ci-dessous :

(26) Les codes du clitique *en* :

(a) « Génitif » : pro:cli:gen|en

(b) « Ablatif » : pro:cli:abl|en

(c) « Circonstant » : pro:cli:circ|en

(d) « Quantitatif » : pro:cli:quant|en

En ce qui concerne le pronom *y*, trois cas sont distingués, comme illustré en (27) ci-dessous : le *y* « datif » qui représente un syntagme nominal [-Individu]¹, argument interne du verbe (*cf.* (a))², le *y* « locatif sous-catégorisé » qui représente un syntagme prépositionnel sous-catégorisé par le verbe (*cf.* (b)) et le *y* « locatif adjoint », adjoint à V' (*cf.* (c))³.

¹ Cf. Oliiviéri (1994) qui utilise le trait [Individu] à la place du trait classique [Animé] afin de rendre compte d'occurrences telles que *il parle au mur* ; *il lui parle* où *le mur* est bien [-Animé] mais [+Individu].

² Cf. Oliiviéri (1994:245-251) qui fait écho à son analyse du pronom *lui*, contrepartie animée de *y* datif (par exemple : *Marie répondra à Paul*, *Marie lui répondra*) où *lui* reprend un syntagme nominal datif animé dans la mesure où, comme précédemment, la préposition n'est pas analysée comme une « vraie » préposition mais comme un « simple marqueur du Cas Datif » (*cf.* Oliiviéri (1994:202-205)).

³ Cf. Oliiviéri (1994:248) : les syntagmes prépositionnels adjoints à I' ne sont pas cliticisables (exemple : *Jean mange dans une assiette*, **Jean y mange*).

(27) Les deux types de *y* :

(a) « Datif » :

Ça ressemble à *un livre*. Ça *y* ressemble.

(b) « Locatif sous-catégorisé » :

Je vais à *la plage*. J'*y* vais.

Elle pense à *ses vacances*. Elle *y* pense.

(c) « Locatif adjoint » :

Il mange *dans ce restaurant*. Il *y* mange.

Les codes correspondant à ces trois pronoms *y* sont détaillés en (28) ci-dessous :

(28) Les codes du clitique *y* :

(a) « Datif » : pro:cli:dat|y&3S

(b) « Locatif » : pro:cli:loc|y

(c) « Locatif adjoint » : pro:cli:local|y

Face à la relative complexité des traits qui composent ces deux clitiques, nous pouvons nous attendre à ce que les jeunes enfants ne les emploient pas encore dans des proportions élevées et d'éventuelles agrammaticalités sont également prévisibles.

4.1.2.2 *Le système enfantin*

Le clitique *en* est présent 87 fois dans le corpus, ce qui est relativement rare si l'on rapporte ce chiffre au nombre total de verbes fléchis (11027) puisque le taux de présence devant un verbe fléchi est donc de 0,8 %¹. Ces occurrences se répartissent sur deux catégories de *en* uniquement² : le *en* ablatif (quinze réalisations en tout, cf. (29a) ci-dessous) et le *en* quantitatif (72 occurrences, cf. (29b)), parmi lesquelles nous distinguons les clitiques doublés grâce au code supplémentaire pro:cli:d:quant|en (cf. (29c)). Ainsi, il apparaît que le *en* génitif et le *en* circonstant ne font pas du tout partie du système enfantin à ce stade.

¹ Par ailleurs, il n'y a que sept occurrences de *en* devant un infinitif.

² Requêtes : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:gen|en" +u @ > En_Gen ; combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:quant|en" +u @ > En_Quant ; combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:d:quant|en" +u @ > En_Quant_D ; combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:abl|en" +u @ > En_Abl.

(29) Les différents types de *en* dans le corpus :

- (a) EVA: ben j' *en* ai pas besoin . (XIII.46c, 663)
- (b) TOM: et ben moi j' *en* ai pas beaucoup . (XIII.48d, 489)
- (c) ANT: j' *en* ai deux *de papillons* . (XIII.48d, 254)

Les codes %err de la base de données permettent de constater que les enfants omettent parfois ces clitiques de leurs énoncés, comme illustré en (30) ci-dessous pour les deux types de *en*¹. Néanmoins, ces omissions restent relativement rares puisqu'elles atteignent le chiffre global de onze (une pour l'ablatif, cf. (a) ci-dessous et dix pour le quantitatif, cf. (b)) :

(30) Les omissions de *en* :

- (a) « Ablatif » :
EMA: et maintenant il 0 [*] a besoin . (VI.24b, 913)
- (b) « Quantitatif » :
NOE: mais ma maman elle 0 [*] a . (X.40a, 728)

La quasi-totalité des clitiques *y* apparaissent dans les locutions *il y a* prononcées [ja], *il y en a* prononcées [jãna], *il y avait* prononcées [jave] et *ça y est* prononcées [saje]. Les nombres d'occurrences pour ces quatre tournures sont repris dans le Tableau 41 ci-après ; le total s'élève 350.

Tournures (API)	ja	jãna	jave	saje
Nombres	286	31	28	5

Tableau 41 : les tournures qui comprennent « y »

Malgré une présence qui semble relativement élevée de ces tournures, il ne peut pas être garanti que *y* soit utilisé comme un constituant autonome dans ces contextes. En effet, la prononciation de cet élément ne nous permet pas d'affirmer que ces tournures sont le résultat d'une computation de la part des enfants. Par ailleurs, il ne peut pas non plus être déterminé avec certitude qu'un clitique nominatif précède *y* dans ces cas. Nous ne retenons donc pas ces occurrences dans le cadre d'une argumentation sur les clitiques qui interviennent entre un clitique nominatif et un verbe fléchi. L'extraction de *y* devant verbe fléchi renvoie alors un total de seulement

¹ Requêtes : `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"$SYN^$LOS^$PRO^$ABL" +u @ > NoClitic_En_Abl ; combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"$SYN^$LOS^$PRO^$QUANT" +u @ > NoClitic_En_Quant.`

trois occurrences, reprises en (31) ci-dessous avec les codes morphosyntaxiques correspondant à la catégorie de chaque clitique y^1 :

(31) Les occurrences de y devant verbe fléchi :

- (a) ISI: j' y arrive pas . (IV.19c, 790)
 %mor: pro:cli:loc|y
- (b) JUL: c' est un oiseau (g)ris mais il y est pas l' oiseau (g)ris . (X.40d, 919)
 %mor: pro:cli:loc|y
- (c) EVA: ça veut dire que j' y arrive pas . (XIII.47b, 254)
 %mor: pro:cli:loc|y

À ce stade de leur développement, il est ainsi démontré que les enfants utilisent ce clitique très rarement entre un clitique nominatif et un verbe fléchi et que, lorsqu'il fait surface, il est uniquement employé en tant que pronom locatif sous-catégorisé. Par ailleurs, les codes %err correspondant aux différents types de y renvoient seulement quatre omissions de ce clitique devant un verbe fléchi (plus une devant un infinitif), toutes comprenant le y locatif sous-catégorisé². Ces occurrences, reprises en (32) ci-dessous, montrent que, sauf dans le cas de Julien qui réalise effectivement l'énoncé (31b) ci-dessus, les enfants qui produisent ces énoncés en sont au stade de la « tentative » puisque le y n'apparaît pas. L'étude des données enregistrées au cours de la seconde année de maternelle sera donc intéressante de ce point de vue également.

(32) Les omissions de y devant verbe fléchi :

- (a) EKT: j' 0 [*] arrive pas ! (IV.19c, 274)
- (b) EKT: j' 0 [*] arrive pas . (IV.19c, 283)
- (c) LUS: oui j' 0 [*] arrive ! (VI.24b, 2007)
- (d) JUL: et en plus il 0 [*] a un loup dedans . (XII.45, 1286)

Par ailleurs, d'autres données illustrent le fait que le clitique y (voire le clitique *en*, cf. (33d) ci-dessous) peut être interprété par certains enfants, même plus âgés, comme étant un préfixe verbal. Les exemples en (33) ci-dessous indiquent en effet que y ne se déplace pas devant l'auxiliaire fléchi (cf. (a) et (b)) et qu'il s'intercale

¹ Requête : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:dat|y&3S+pro:cli:loc|y" +u @ >Cli_y_Dat&Loc. L'extraction de y avec un verbe non fléchi renvoie un taux similaire, extrêmement bas.

² Requête : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"\$SYN^\$LOS^\$PRO^\$LOC+\$ADV" +u @ > No_Clitic_Loc&Adv.

même entre le préfixe *re-* et le verbe (*cf.* (c)). Cependant, ce type d'interprétation n'est pas surprenant si l'on considère la diachronie de verbes tels que *s'enfuir* ou *s'envoler* qui proviennent de l'agglutination complète du pronom *en* aux verbes *fuir* et *voler*¹. Les formes *y-aller* et *en-aller* pourraient alors représenter l'aboutissement de l'application du même mécanisme de fusion².

(33) Les verbes *y-aller* et *en-aller* :

- (a) On était *y-allés* (REM, 6;9.11)
- (b) Je suis pas *y-allée* (JUL, 11;5.8)
- (c) Il a dit qu'il aimerait bien *re-y-aller* (REM, 9;4.7)
- (d) Tu peux m' *en-aller* chercher une s'il te plaît ? (REM, 7;6.5)

De nouveau donc, si l'argument de Zwicky & Pullum (1983) devait être retenu, la présence peu fréquente de ces clitiques *en* et *y*, ainsi que leur utilisation par les enfants, jeunes et un peu moins jeunes, ne représentent pas des arguments forts en faveur du statut argumental des clitiques nominatifs. Dans le cas présent, au regard des exemples (33) ci-dessus, l'argumentation serait peut-être même en défaveur de cette hypothèse.

4.1.3 Les clitiques accusatifs et datifs

L'omission des clitiques accusatifs et datifs est un phénomène bien attesté chez les jeunes enfants³. Ainsi, il est notamment démontré que l'apparition de ce type de clitiques survient après celle des clitiques nominatifs. En effet, alors que les clitiques nominatifs apparaissent dès les premières combinaisons de deux mots vers dix-huit mois, les clitiques non nominatifs ne semblent émerger de façon stable qu'entre trois et quatre ans. Avant cette période, les enfants préfèrent ainsi produire un DP plein ou omettre l'argument⁴. Par ailleurs, concernant le langage adulte, Vendryes (1920) note une « [...] tendance, naturelle en français, à rattacher au verbe par un lien étroit le pronom sujet. Les cas où l'on peut disjoindre les deux deviennent de plus en plus

¹ Cf. par exemple Grevisse & Goosse (2008:878).

² Grevisse & Goosse (2008:880) indiquent d'ailleurs que cette agglutination est également attestée dans la littérature sous la forme *Il s'est en allé* chez Chateaubriand, Hugo, Flaubert, etc.

³ Cf. par exemple pour le français Kaiser (1994), Hamann *et al.* (1996) et Jakubowicz & Rigaut (2000).

⁴ Un rapprochement serait alors à étudier entre l'ordre d'acquisition de ces clitiques et l'ordre d'apparition des morphèmes flexionnels sur le verbe dans une langue telle que le basque où la réalisation de la projection AgrS (accord avec le sujet) est attestée avant celle de la projection AgrO (accord avec l'objet). Cf. par exemple Ezeizabarrena (2003:92-94) sur ces différentes étapes acquisitionnelles.

rare ; *je le dis, tu le sais* sont remplacés dans la langue parlée par *je dis ça, tu sais ça [...]* »¹.

Les codes %mor de notre base de données permettent d'extraire les clitiques accusatifs, datifs et réfléchis selon la même procédure que les clitiques nominatifs. Les requêtes adéquates² extraient ainsi un total de 1105 clitiques objets (524 accusatifs, 211 datifs et 370 réfléchis). Les données de ce corpus sont donc en adéquation avec les études précédentes dans la mesure où nous constatons que le nombre de clitiques non nominatifs est bien moindre que celui des clitiques nominatifs. En effet, comparés aux 8729 clitiques nominatifs, les premiers apparaissent huit fois moins souvent. Leur présence diminue encore lorsque nous restreignons leur contexte d'apparition à la production concomitante d'un clitique nominatif³. Dans ce cadre, le corpus comprend alors 701 occurrences de ces combinaisons que nous illustrons en (34) ci-dessous :

(34) Quelques exemples de clitiques accusatifs, datifs et réfléchis dans le corpus :

(a) Accusatif :

ALA: tu *le* refais . (I.3b, 809)

(b) Datif :

EVA: et elle *me* prête celui+là elle ? (II.9d, 974)

(c) Réfléchi :

NOE: i(l) *s'* appelle Christian celui+là . (II.5, 1004)

Dans la mesure où l'argument de Zwicky & Pullum (1983) est discuté et où il existe des hypothèses sur le statut affixal des clitiques objets⁴, nous concluons que cet argument sur la présence de ces clitiques objets n'est pas non plus décisif dans le cadre du débat sur le statut des clitiques nominatifs. De plus, dans un nombre non négligeable de cas, il est impossible d'être absolument sûr que le clitique accusatif attendu est bien réalisé, notamment lorsque le clitique *il* est utilisé devant un verbe dont l'initiale est une voyelle, comme le montrent les exemples en (35) ci-dessous :

¹ Vendryes (1920:107).

² Sur le modèle : freq -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:*acc|*" +u @ > Pro_Acc_All_Freq

³ Nous écartons ainsi notamment les clitiques produits avec un verbe non fléchi. Requête : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:*nom*|^^(pro:cli:*dat|*+pro:cli:*acc|*+pro:cli:*refl|)*" +u @ > Pro_ADR_Nom_Combo.

⁴ Cf. notamment Kaiser (1994).

(35) Réalisation incertaine de certains clitiques accusatifs :

(a) MAX: il l' [?] enlève . (IV.19a, 617)

(b) EMA: il l' [?] attrape hou ! (VII.27b, 1310)

Le doute sur la présence possible mais non certaine de ces clitiques objets est alimenté par le fait que les jeunes enfants omettent parfois ces clitiques. Les codes %err correspondants renvoient en effet un total de 68 omissions attestées dans ce corpus¹ :

(36) Omissions des clitiques accusatifs, datifs et réfléchis :

(a) JUL: j' 0 [*] ai pas moi . (X.40d, 1566)

(b) ANT: nan Maxime tu 0 [*] donne ça . (I.3a, 553)

(c) EKT: moi j' 0 [*] étais baigné . (VI.26e, 503)

À ce stade donc, il nous semble que les arguments avancés par De Cat (2005) en faveur du statut argumental des clitiques nominatifs ne sont pas déterminants dans le cadre du système linguistique du jeune enfant. Poursuivons néanmoins notre investigation en analysant l'aspect phonétique des événements lorsqu'un clitique nominatif et un verbe fléchi sont et ne sont pas adjacents.

4.2 Un ajustement phonétique contraint des clitiques nominatifs

Les données montrent un modèle d'ajustement phonétique des clitiques nominatifs très contraint, facilement repérable notamment grâce aux deux formes que prend le clitique *il*. En effet, à de très rares exceptions près (sur lesquelles nous revenons ci-après), les enfants utilisent la forme pleine *il* devant voyelle et la forme élidée *i(l)* devant consonne, comme illustré en (37) ci-dessous² :

¹ Requête : `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"$SYN^$LOS^$PRO^($OBJ+$REFL+$DAT)" +u @ > Pro_ADR_Los.`

² Cette constatation repose sur l'étude des 1688 clitiques *il* du corpus.

(37) L'ajustement phonétique du clitique *il* :

(a) *i(l)* + consonne :

LUS: *i(l)* veut manger . (I.3d, 1077)

(b) *il* + voyelle :

ANT: *il* est à moi . (I.3b, 68)

(c) Les deux :

NOE: *i(l)* pleure parc(e)+que [*] *il* est très malheureux . (X.40b, 194)

Ces faits ne sont pas surprenants au regard du modèle consonne-voyelle, prégnant dès le babillage chez les petits francophones¹. Ils sont néanmoins intéressants lorsqu'on les confronte au mécanisme d'ajustement phonétique qui est appliqué aux autres catégories du lexique infantin. En effet, nous nous apercevons alors que l'élision n'est pas systématique, comme le montrent les exemples en (38) ci-dessous pour les prépositions, les conjonctions et les déterminants.

(38) L'ajustement phonétique n'est pas systématique (41 occurrences) :

(a) Prépositions :

EMA: *de* [*] une carotte et *de* [*] une tomate . (IX.34a, 1317)

LUS: moi je me souviens *de* [*] autre chose . (XIII.46a, 652)

(b) Conjonctions :

TOM: parc(e)+*que* [*] on a pas à manger . (VI.26a, 545)

ANT: oui *que* [*] un chat . (XI.43b, 2043)

(c) Déterminants :

QUE: et ben moi il est ici *le* [*] oiseau . (III.13e, 714)

MAX: mais où il est *ce* [*] oiseau ! (IV.19c, 834)

Seules onze formes de *il* ne sont pas élidées devant consonne dans le corpus, dont neuf occurrences sont produites par Emma, toutes sur le même modèle (*cf.* (39c) ci-dessous).

¹ *Cf.* par exemple De Boysson-Bardies (1996:72-81) : les enfants francophones produisent entre 65 et 75 % de dissyllabes consonne-voyelle-consonne-voyelle au cours de leur première année.

(39) Les formes non élidées de *il* :

- (a) LUS: et après oh *il* dit <qui est là> ["] ! (VIII.31a, 947)
- (b) EVA: lui *i(l)* veut prendre la poule *i(l)* mange *il* n' a pas gagné pour [/] pour [/] pour les œufs . (XIII.46a, 1097)
- (c) EMA: *il* ne chante plus . (XI.41d, 1404-1472)

Ces énoncés sont extrêmement intéressants dans la mesure où le mécanisme d'ajustement ne s'applique pas, dix fois sur onze, dans des énoncés que nous avons déjà mis en valeur pour une autre raison. En effet, deux phénomènes notables surviennent ainsi concomitamment dans les énoncés d'Emma et d'Eva ci-dessus : l'absence de l'ajustement phonétique du clitique devant consonne et la présence de la particule négative *ne*. Dans ce contexte, *il* prend donc une forme inhabituelle par rapport au reste du corpus. Cela n'est pas simplement dû au fait que le clitique n'est plus strictement adjacent au verbe puisque les données montrent par ailleurs que le clitique nominatif *il* présente systématiquement sa forme élidée lorsqu'il est séparé du verbe par un clitique accusatif, un clitique datif, la combinaison des deux, et devant un verbe préfixé, comme dans les exemples en (40), (41), (42), (43) et (44) ci-dessous :

(40) Élisation systématique de *i(l)* devant un clitique accusatif :

- (a) NOE: euh <et mon p(apa)> [/] mais mon papa *i(l)* *me* laisse sur le poney . (II.5, 1205)
- (b) TOM: *i(l)* *m'* a griffé ! (II.9a, 1520)
- (c) ALA: eh *i(l)* *m'* a poussé ! (VI.24b, 952)

(41) Élisation systématique devant un clitique datif :

- (a) NOE: xxx *i(l)* *nous* fait mal . (II.7, 460)
- (b) TEO: *i(l)* *m'* a prêté +... (III.10a, 547)
- (c) EMA: *i(l)* *t'* en manque un . (III.10b, 964)
- (d) EVA: moi *i(l)* *me* manque plus que la tortue et j(e) vais gagner . (VII.29b, 1732)

(42) Élisation systématique devant un clitique réfléchi :

- (a) TOM: *i(l)* s' appelle Lucien . (II.5, 602)
- (b) EKT: *i(l)* se met les chaussons . (IV.19c, 99)

(43) Élisation systématique devant deux clitiques :

- (a) EKT: *i(l) me le* [?] passe Tom ! (II.9b, 267)
- (b) ALA: après *i(l) m(e) le* redonne . (X.40a, 1131)
- (c) ANT: *i(l) m(e) l'* a piqué ! (XI.42, 1240)

(44) Élisation systématique devant un verbe préfixé :

- (a) LUS: *i(l) l(e) remet* . (V.23e, 406)¹
- (b) ZOE: et *i(l) referme* . (V.23e, 302)

Tout d'abord, ces observations permettent de trancher face à l'incertitude des transcriptions mentionnées en (35) ci-dessus et reprises en (45a) ci-dessous puisque, d'une part, les enfants produisent des clitiques accusatifs, datifs et réfléchis avec le clitique *il* et que, d'autre part, ils produisent le clitique accusatif de troisième personne avec des clitiques autres que *il* (cf. (45b)). Il n'y a donc aucune raison apparente pour qu'ils n'associent pas les deux et que la forme du clitique nominatif ne soit pas la forme élidée attendue (cf. (45c)).

(45) Réalisation du clitique accusatif *l'* :

- (a) MAX: *il l'* [?] enlève . (IV.19a, 617)
EMA: *il l'* [?] attrape hou ! (VII.27b, 1310)
- (b) MAX: oh moi *je l'* a [*] trouvé . (III.13d, 523)
EMA: regarde et ça *je l'* enlève . (III.10b, 340)
- (c) MAX: *i(l) l'* [ʔ] enlève . (IV.19a, 617)
EMA: *i(l) l'* [ʔ] attrape hou ! (VII.27b, 1310)

Dans un deuxième temps, les données présentées en (40)-(44) ci-dessus indiquent que la non adjacence n'est pas suffisante pour expliquer les formes de *il* puisque les deux cas de figure se présentent lorsque *il* n'est pas au contact direct du verbe : élisation devant les clitiques objets vs non élisation devant la particule négative *ne*. La partie du processus d'ajustement mentionné en (37) ci-dessus qui concerne la consonne est donc reformulé en (46) ci-dessous :

¹ Dans ce cas, l'énoncé LUS: *i(l) l(e) met* . (V.23e, 268) permet d'attester qu'il y a bien préfixation du verbe *mettre*.

(46) Récapitulatif du processus d'ajustement de *il* devant consonne :

(a) *il + ne > il*

EMA: *il* ne saute plus . (XI.41d, 1411)

EVA: lui *i(l)* veut prendre la poule *i(l)* mange *il* n' a pas gagné pour [l]
pour [l] pour les œufs . (XIII.46a, 1097)

(b) *il + C≠ne > i*

i. NOE: parc(e)+que *i(l)* pleure plus le petit lapin xxx . (VI.26a, 443)

EMA: nan sinon <la maî(tresse)> [l] ma maîtresse *i(l)* s(e)ra pas
contente .

ii. EMA: oui *i(l)* saute . (VI.24b, 572)

EMA: comment *i(l)* s' appelle lui ? (XIII.46e, 1376)

EMA: <si le loup était *i(l)* nous mangerait> ["] ! (VII.27b, 1080)

LUS: ben non *i(l)* m' a pas mangée en plus . (III.13c, 209)

Dans la mesure où ce modèle d'ajustement est très contraint dans le corpus, il est envisageable qu'il ne soit pas le fruit d'un simple hasard mais que, bien au contraire, il représente la partie visible d'un mécanisme réel mais sous-jacent qui n'est pas lié à l'aspect négatif de l'énoncé, comme l'illustrent les exemples en (bi) ci-dessus. Il convient donc à présent de déterminer la source de la dichotomie présentée en (46a) vs (46b) ci-dessus. Plusieurs pistes ont d'ores et déjà été éliminées. La variation n'est pas liée : (i) à la négation *pas* ou *plus*, (ii) à la consonne initiale du verbe, (iii) à la non adjacence stricte entre *il* et le verbe. Il reste alors plusieurs hypothèses : (i) le clitique *ne* est différent des autres clitiques étudiés, (ii) les deux réalisations phonétiques *il* et *i(l)* qui, jusqu'à présent ont été considérées comme dérivant d'un même objet phonologique sous-jacent /il/, sont en fait deux entités différentes, (iii) une conjonction des deux premières hypothèses.

Concernant une éventuelle distinction entre *ne* et les autres clitiques préverbaux du corpus, nous savons déjà que la phonologie ne peut pas être évoquée puisque, dans tous les cas, il s'agit d'éléments atones, monosyllabiques et de type consonne-voyelle. Le signifiant seul ne semble donc pas en cause. Par contre, il apparaît dans la littérature qu'une différence morphosyntaxique entre les deux types d'éléments puisse être envisagée. En effet, De Cat (2005) par exemple propose que la particule négative *ne* ne puisse pas être analysée en tant qu'affixe verbal dans la mesure où la portée de ce marqueur négatif semble déterminée par la structure syntaxique de

l'énoncé¹. La littérature sur la question de la négation en français est abondante et divergente. En effet, la problématique est complexe du fait que la négation peut être marquée par un morphème discontinu (*ne...pas*), par l'élément *pas* utilisé seul ou par l'élément *ne* utilisé seul. Depuis Pollock (1989), une projection fonctionnelle NegP est néanmoins généralement admise dont la tête est la particule *ne* (réalisée ou non). Cependant, sa place dans la structure, le nombre de projections, ainsi que les traits qu'elle contient varient d'une hypothèse à une autre². Dans le cadre de nos recherches actuelles, nous ne pouvons donc pas discuter les différentes approches et nous nous contentons d'exposer l'une des hypothèses les plus récentes proposée dans le cadre minimaliste. Schapansky (2009) propose ainsi que la négation puisse être de deux types : « contradictoire » c'est-à-dire qu'elle nie une proposition, ou « contraire » c'est-à-dire qu'elle affaiblit une assertion³. Au niveau de la proposition, le premier type de négation est réalisé soit par *pas* utilisé seul soit par le morphème discontinu *ne...pas*. Les éléments du type *pas* (*plus, personne, rien, etc.*) possèdent un trait négatif [+N] mais ce sont également des quantifieurs ; ils possèdent donc un trait [+Q]. *A contrario*, le trait Q de *ne* est négatif et ininterprétable : [+N, -uQ]. L'élément *pas* se déplace donc au cours de la dérivation afin de vérifier ce trait ininterprétable. Dans ce cadre, il est postulé que « *pas* est généré en tant qu'adjoint à vP, [_{vP} <pas> [_{vP}]]⁴ puis qu'il vérifie le trait ininterprétable de la tête *ne* (qui peut être phonétiquement vide) lorsqu'il arrive en spécifieur de NegP : [_{NegP} pas_[Q] [(ne)_[#Q] [_{vP} <pas> [_{vP}]]]] ». Schapansky (2009) ne détaille pas le reste de la dérivation mais il faut également que le sujet et le verbe se déplacent en TP (qui se trouve au dessus de NegP) et l'auteur n'explique pas non plus l'opération qui permet à *ne* de se retrouver cliticisé entre le sujet et le verbe. Retenons pour le moment que *pas* est adjoint à vP, qu'il existe une projection NegP entre TP et vP et que *ne*, lorsqu'il est réalisé, n'est pas considéré comme un préfixe du verbe. *A contrario*, nous avons vu au Chapitre II que Kaiser (1994) par exemple propose que les clitiques objets puissent être analysés comme des affixes verbaux⁵. Dans le cas de la négation, nous avons alors affaire à une projection syntaxique et, dans les cas des clitiques objets, à des affixes

¹ Cf. De Cat (2005:1201-1203) pour la démonstration.

² Cf., entre autres, Zanuttini (1997), Rowlett (1998) et Schapansky (2009).

³ Cette partition est notamment illustrée par la différence qu'il y a entre *non humain* et *inhumain* (le premier terme étant contradictoire alors que le second est contraire).

⁴ L'opération d'adjonction (*Adjoin*) ajoute simplement un constituant à un syntagme existant ; elle est différente de l'opération de fusion (*Merge*) dans la mesure où l'adjonction ne crée pas un nouvel objet syntaxique contrairement à la fusion.

⁵ Cf. Chapitre II, Section IV.2.2.2.

morphologiques. La différence de comportement du clitique *il* relèverait alors de la structure syntaxique de l'énoncé dans la mesure où, en présence de *ne*, la structure syntaxique présenterait une tête réalisée supplémentaire.

Cette première hypothèse a des implications directes sur la deuxième hypothèse formulée ci-dessus. En effet, si *ne* relève d'une projection syntaxique supplémentaire, il paraît alors difficile dans ce cadre de postuler que le clitique nominatif soit un affixe du verbe. L'hypothèse attenante à l'hypothèse (i) est donc que, dans ce type de structures, le clitique nominatif relève alors également d'une projection syntaxique (au même titre que n'importe quel DP généré en spécifieur de *vP* ; *il* est alors considéré comme un pronom au sens étymologique du terme et est une tête D). Le second volet attenant à l'hypothèse (i) est que *il* peut également être considéré comme un affixe du verbe, tant qu'aucune tête intermédiaire ne survient dans la structure. L'hypothèse (i) sur la présence d'éléments de types différents entraîne alors l'hypothèse (ii) sur la présence de deux *il* différents, un *il* syntaxique qui ne montre pas de phénomènes d'ajustement par rapport à l'élément suivant (puisque généré dans une projection syntaxique différente) et un deuxième *il* qui, lui, manifeste ces phénomènes d'ajustement de façon stricte, dans la mesure où il est intégré à une chaîne de clitics qui sont tous des préfixes du verbe fléchi. L'hypothèse (iii), qui réunit (i) et (ii), est donc viable dans la mesure où la particule *ne* n'est quasiment pas insérée à ce stade du développement syntaxique et que notre corpus met en avant une réelle dichotomie quant aux ajustements phonétiques différents que subissent les clitics *il* dans les deux contextes identifiés ci-dessus. La question suivante consiste à se demander s'il y a un lien de cause à effet entre la présence de la particule négative *ne* et le statut du clitique *il*. Deux ordres sont possibles : (i) la présence d'une tête *ne* empêche *il* d'être préfixe / oblige *il* à être argument ; (ii) la présence du préfixe *il* empêche la réalisation de la tête *ne*.

Les seize occurrences de *ne* montrent que la particule négative est toujours strictement adjacente au verbe, sauf dans un cas. Il s'agit de l'énoncé à l'impératif produit par Eva à la treizième séance, repris en (47a) ci-dessous, dans lequel il n'y a donc pas de clitique nominatif. Par ailleurs, les enfants ne produisent que très peu d'énoncés avec *il, pas* et un clitique non nominatif (*cf.* (47b)).

(47) Cooccurrence de la négation et des clitiques non nominatifs :

- (a) EVA: *ne te cache pas avec la photo !* (XIII.46e, 1086)
- (b) JUL: *c' est un oiseau (g)ris mais il y est pas l' oiseau (g)ris .* (X.40d, 919)
LUS: *ben non i(l) m' a pas mangée en plus .* (III.13c, 209)

Un lien de cause à effet peut alors être postulé entre la présence de *ne* et le statut de *il* :

(48) La présence de la particule *ne* joue sur le statut de *il* :

- (a) Les clitiques nominatifs et non nominatifs entrent dans la numération en tant qu'affixes du verbe (exemple : N = {i-m(e)-donne, *pro*_[agent], *pro*_[bénéficiaire], un, livre}).
- (b) Ce processus morphologique semble entravé par la présence du marqueur négatif *pas* mais ne le rend pas impossible dans la mesure où *pas* n'est ni clitique ni préverbal en fin de dérivation (N = {i-m(e)-donne, *pro*_[agent], *pro*_[bénéficiaire], pas, un, livre}).
- (c) Le processus morphologique décrit en (a) est totalement bloqué lorsque la numération comprend le marqueur négatif clitique *ne* (*N = {i-m(e)-donne, *pro*_[agent], *pro*_[bénéficiaire], ne, pas, un, livre}).

Par rapport à la présence vs l'absence phonétique de la particule *ne*, deux hypothèses sont alors possibles lorsque *ne* est phonétiquement absent : (i) la tête Neg est présente dans la structure syntaxique mais phonétiquement vide ; celle-ci porte le trait [-uQ] qui est vérifié par le trait [+Q] de *pas* dans le spécifieur de NegP ; (ii) la tête Neg n'est pas projetée dans la syntaxe ; ceci ne fait pas capoter la dérivation en termes de vérification de traits puisque les traits de *pas* sont interprétables. Les placements corrects et incorrects de *pas* dans les énoncés enfantins seraient alors à étudier afin de rendre compte de la position et de la portée de cet élément chez les jeunes enfants et afin de déterminer ses éventuels déplacements¹. Les interactions entre mode impératif et négation présentent également des domaines d'investigation riches et complexes que nous réservons pour des recherches ultérieures, ainsi que l'étude précise de l'ajustement des autres clitiques nominatifs. Pour le moment, nous retenons les éléments suivants récapitulés en (49) ci-dessous :

¹ Le codage %err dispose d'une série de codes identifiant le placement agrammatical de la négation et d'autres constituants.

(49) Récapitulatif :

- (a) Une corrélation est établie entre la réalisation de la particule négative *ne* et les deux formes que prend le clitique nominatif *il* (*il* devant *ne* vs *i(l)* devant tout autre élément consonantique, quelle que soit la distance entre *i(l)* et le verbe fléchi).
- (b) Cette partition permet de postuler l'existence de deux objets distincts : un *i(l)* affixe verbal dont l'ajustement est fortement contraint (*il* devant voyelle vs *i(l)* devant consonne) et un *il* argument syntaxique qui ne connaît pas ce mécanisme d'ajustement.
- (c) Des deux objets, c'est l'affixe morphologique qui apparaît le premier dans le système enfantin. L'apparition du deuxième objet, l'argument syntaxique, semble être déclenchée (entre autres ?) par la réalisation phonétique/la projection de la particule négative *ne*.

Les éléments regroupés par De Cat (2005) sous le terme générique de clitiques intercalés ne sont donc pas à considérer globalement. Nous avons montré qu'il convient de distinguer la particule négative des autres clitiques et que, si ces derniers peuvent encore être considérés comme des préfixes verbaux, le raisonnement ne peut pas être suivi pour *ne* dans la mesure où sa projection (ou réalisation) n'a pas les mêmes conséquences que la réalisation des autres clitiques. Cette constatation nous amène à lui accorder un statut syntaxique différent.

Ces conclusions sur l'utilisation des clitiques nominatifs dans le système enfantin représentent la première pièce de notre puzzle. Tournons-nous à présent vers un second type d'éléments, également potentiellement sujets : les éléments « *qu-* ».

III. LES ÉLÉMENTS *QU-*

L'appellation « éléments *qu-* » réunit les pronoms interrogatifs et les pronoms relatifs. Le rapprochement entre ces deux catégories est effectué sur la base d'une proximité morphologique entre certains interrogatifs et relatifs (notamment *que* et *qui*) et sur le fait qu'il est postulé que ces éléments sont générés dans leur position argumentale puis qu'ils sont copiés dans une position non-argumentale, dans la périphérie gauche de la proposition, afin de vérifier les traits inhérents soit à l'interrogation soit à la relativisation présents sur la tête C. Ces mécanismes sont illustrés en (50) ci-dessous :

(50) Les éléments *qu-* :

- (a) Interrogation : [CP qui [TP ~~qui~~ fait ça]] ?
- (b) Relativisation : [DP la personne [CP qu_{+x} [TP ~~x~~ fait ça]]]

Néanmoins, les deux exemples proposés ci-dessus montrent également que les faits ne se réduisent pas à cette simple description, surtout en ce qui concerne la relativisation qui implique non seulement le déplacement d'un argument mais également la subordination et des questions, largement débattues, d'ordre morphologique concernant la fusion du complémenteur et de l'élément déplacé.

1. Les pronoms interrogatifs

L'activité qui a eu le plus de succès auprès des enfants au cours de nos séances d'enregistrements est celle que nous appelons « le jeu des animaux »¹. Cette activité permet, entre autres, de proposer un contexte propice à la formulation d'énoncés interrogatifs. En effet, à tour de rôle, chaque enfant pioche une carte et, si celle-ci ne figure pas sur son propre carton, l'enfant doit demander qui possède l'image correspondante. Par exemple, les questions de Tom, reprises en (51) ci-dessous, émergent après plusieurs sollicitations de notre part :

(51) Exemple de contexte propice à l'interrogation (VIII.33c, 1098-1125) :

- KAT: tu as pioché la vache .
KAT: est+c(e)+qu' elle est à toi la vache ?
TOM: nan .
KAT: alors i(l) faut demander +...
TOM: c' est à qui ?
KAT: voilà .
TOM: c' est à qui la vache ?

Les interrogations de Tom ci-dessus illustrent un premier phénomène enfantin quant au constituant interrogé dans les énoncés. Ainsi, le corpus met en avant le fait que les enfants ne font pas souvent porter l'interrogation sur le sujet grammatical. En effet, sur un total de 1141 énoncés interrogatifs verbaux, seuls 99 portent sur ce constituant (soit 8,7 % des interrogations verbales enfantines). La majorité des interrogations

¹ Cf. Chapitre III sur notre protocole de collecte.

enfantines est donc composée soit de questions oui/non (cf. (52a) ci-dessous), soit de questions qui portent sur l'objet ou le circonstant (cf. (52b) et (52c)) :

(52) La majorité des interrogations enfantines :

- (a) ZOE: y+en+a plus ? (I.3a, 762)
- (b) NOE: et ben le monsieur qu+est+c(e)+qu' i(l) fait ? (II.7, 950)
- (c) EVA: et pourquoi Lola elle veut pas jouer ? (III.10a, 801)

1.1 Les différentes formes

En ce qui concerne les interrogations qui portent sur le sujet, le corpus montre que les enfants utilisent un total de quatre formes différentes : *qui*, *c'est qui qu(i)*, *qui c'est qui* et *qu'est-c(e) qui*, comme illustré en (53) ci-dessous. Les transcriptions de ces éléments en (a) et (b) diffèrent de celles présentées en (c) et (d) dans la mesure où les deux dernières comportent le signe + qui n'apparaît pas dans les deux premières. Ainsi, lors du codage morphosyntaxique, nous avons décidé d'étiqueter individuellement chaque constituant, tant qu'il n'y a pas de déplacement visible de l'élément interrogatif *qu-* par rapport à la copule *est* (d'où l'absence de signes +). Par contre, nous avons transcrit les éléments de type (c) et (d) en tant que locutions interrogatives sujets en vue d'un étiquetage global de ces composantes en tant que locutions ; la locution *qui+c+est+qui* reçoit ainsi par exemple le code « pro:lo:int:nom|qui+c+est+qui ».

(53) Les différentes formes de pronoms interrogatifs sujets :

- (a) EKT: *qui* veut jouer ? (IV.19b, 201)
- (b) TOM: *c'est qui qui* l'a le cochon ? (VII.29c, 331)
LUS: *c'est qui qu'* a l(e) chat ? (V.23a, 538)
- (c) EVA: *qui+c+est+qui* veut les étoiles ? (I.3a, 644)
EVA: *qui+c+est+qui* a le papillon ? (IV.19d": line 880)
- (d) MAX: *qu+est+c(e)+qui* se cache ici ? (XIII.47b, 52)

Les requêtes sur ces quatre constituants¹ font émerger un total de 99 formes qui se répartissent comme indiqué dans le Tableau 42 ci-après ; les enfants montrent ainsi

¹ combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:int|qui^v*|" +u @ > Pro_Int_QUI+V ; combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:cli:nom|ce*^v:exist|^pro:int|qui^pro:rel|qu*" +u @ > Pro_Int_CEstQuiQu ; freq -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +s"qui+c+est+qui" +u @ > Pro_Int_QUICEstQui ; combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:lo:int:acc|qu\+est\+ce\+qui" +u @ > Pro_Int_QuEstCeQui_combo.

une préférence pour les deux formes intermédiaires *c'est qui qu(i)* et *qui+c+est+qui*. Deux autres faits sont également intéressants : le pronom relatif *qui* dans la troisième forme n'est jamais éliminé (comme indiqué par le second exemple en (53c) ci-dessus) et la dernière forme n'est jamais produite avec le pronom interrogatif *qui* au lieu de *qu(e)*.

Formes	Totaux	%
<i>qui</i>	17	17,18
<i>c'est qui qu(i)</i>	36	36,36
<i>qui+c+est+qui</i>	32	32,32
<i>qu+est+c(e)+qui</i>	14	14,14
Totaux	99	100,00

Tableau 42 : répartition des formes interrogatives

1.2 Computations syntaxiques ou locutions figées ?

L'étude longitudinale de l'utilisation de ces formes ne montre pas de modèle développemental particulier, c'est-à-dire qu'aucune forme ne progresse ni ne régresse significativement par rapport aux autres au fil des séances. Par contre, l'étude de la répartition de ces formes entre les différents enfants, illustrée par la Figure 2 ci-après, permet d'autres constatations.

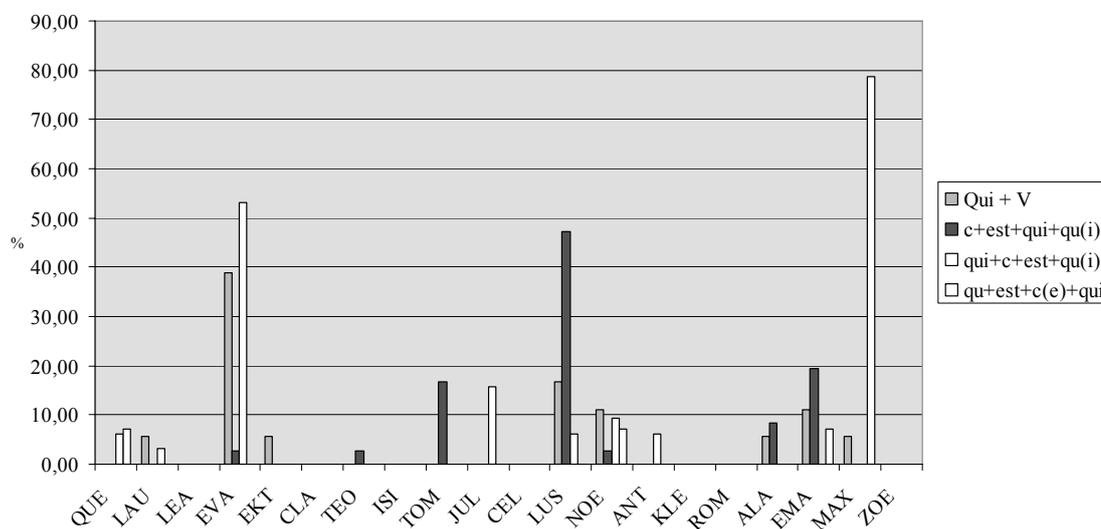


Figure 2 : répartition des interrogations sujets

Tout d'abord, sept enfants sur vingt (Léa, Clara, Isia, Célia, Claire, Romane et Zoé) n'utilisent jamais aucune de ces quatre formes. Pourtant, ils forment par ailleurs des interrogations mais, lorsqu'ils doivent identifier un possesseur, ils choisissent d'autres stratégies, comme illustré en (54) ci-dessous :

(54) D'autres types d'interrogations sur le possesseur :

- (a) CEL: à *qui* c' est ? (IV.19f, 1376)
- (b) ROM: c' est à *qui* ça ? (V.23a, 2097)
- (c) ROM: alors c' est *pour qui* ? (V.23d, 66)

Que ces tournures impliquent un déplacement *qu-* ou non (*cf.* (54a) *vs* (54b) et (54c)), celles-ci représentent toutes des structures syntaxiques plus simples que les constructions présentées en (53b-d) dans la mesure où les formes en (54) s'articulent toutes autour de la copule *est*, unique élément verbal de l'énoncé. L'élément interrogatif n'est donc pas le sujet syntaxique dans la question, contrairement aux exemples en (53b-d) dans lesquels la même forme *est* est enchâssée à l'intérieur de l'élément interrogatif *qui*, dans cette configuration, est le sujet d'un autre verbe dans l'énoncé. Une problématique récurrente dans ce domaine concerne alors l'analyse qui doit être faite des locutions interrogatives du type *qui+c+est+qui*. En effet, deux hypothèses sont possibles : soit l'on considère cette forme comme un tout figé, c'est-à-dire qu'elle représente une unité qui ne nécessite aucune opération syntaxique; soit cette forme est considérée comme le résultat d'une computation, auquel cas l'enfant doit dans un premier temps procéder à des computations syntaxiques sur ses constituants avant d'enchâsser l'ensemble dans le reste de la phrase. En ce qui concerne les différentes formes utilisées, et si ces formes sont bien le résultat d'une computation syntaxique, celles-ci peuvent être classées par ordre croissant de complexité comme suit¹ :

(55) Classement par ordre croissant de complexité syntaxique :

- (a) *qui*
- (b) *c+est+qui+qu(i)*
- (c) *qui+c+est+qu(i)*
- (d) *qu+est+c(e)+qui*

Les formes vont ainsi de l'élément simple *qui*, généré en [Spec, *v*P] puis déplacé en [Spec, TP] comme n'importe quel sujet non interrogatif et enfin déplacé en [Spec, CP] afin de vérifier le trait interrogatif de C. La présence d'un élément relatif dans les autres tournures implique que celles-ci sont complexes. Traditionnellement, la forme (55b) représente la forme de base, puis (55c) implique le déplacement de

¹ *Cf.* par exemple Strick (2007) sur la complexité du calcul syntaxique qui détermine l'acquisition des interrogatives.

l'élément interrogatif *qu-*, puis (55d) ajoute l'antéposition du verbe par rapport au pronom *ce*. Cependant, l'analyse développée ci-dessus ne fait pas l'unanimité. En effet, Munaro & Pollock (2005) par exemple proposent une interprétation différente de ce type de structure. Nous reprenons leur exemple de dérivation du type (d) ci-dessus en (56) ci-dessous. Les auteurs postulent que, dans les langues romanes, la périphérie gauche contient deux positions pour les éléments *qu-* simples (les éléments du type *que, qui*) : une position basse pour les quantifieurs (Op1P) où l'élément *qu-* vérifie son trait existentiel et une position haute pour les opérateurs (Op2P) où l'élément *qu-* vérifie son trait restrictif. Par ailleurs, les auteurs proposent que l'élément *ce* dans ce cadre ne doit pas être confondu avec le clitique nominatif *ce* que l'on trouve dans *c'est bien* par exemple « et dont la version non clitique est *cela* ». Le *ce* présent dans les structures *qu+est+ce* est qualifié de « fossile vivant » par les auteurs, c'est-à-dire qu'il s'agit du prédicat démonstratif issu de l'ancien et moyen français que l'on trouvait par exemple dans *ce suis-je* et que l'on trouve encore en français moderne dans *sur ce*. Dans ce cadre, la présence postverbale de *ce* ne peut donc pas être qualifiée d'inversion dans la mesure où *ce* n'est pas le sujet de la copule *est* mais le prédicat d'une petite proposition enchâssée dans la proposition qui contient la copule, comme illustré en (56a) ci-dessous :

(56) Dérivation de *Qu'est-ce qui tombe ?*¹

- (a) L'opérateur disjonctif *que* est fusionné dans la position sujet d'une petite proposition dont le prédicat est *ce* ; sa partie quantifieur est générée en [Spec, vP] et se déplace :

[ForceP [CopP est [SC **que ce**]] F° [Op1P **O_i** que] [IP **t_i** (i) [vP **t_i** tombe]]

- (b) CopP présente un site de cliticisation pour *que* qui est donc attiré dans cette position :

[ForceP [CLP **que_i** [CopP est [SC **t_i ce**]]] F° [Op1P **O_i** que] [IP **t_i** (i) [vP **t_i** tombe]]

- (c) Fusion avec Op2P et nouveau déplacement de *que* :

[Op2P **que_i Op2°** [ForceP [CLP **t_i** [CopP est [SC **t_i ce**]]]]] F° [Op1P **O_i** que] [IP **t_i** (i) [vP **t_i** tombe]]

¹ Munaro & Pollock (2005:562-563).

(d) Résultat (après cliticisation de *i* sur le complémenteur *que* à PF) :

[_{Op2P} **que**_i _{Op2°} [_{ForceP} [_{CLP} _{t_i} [_{CopP} **est** [_{SC} _{t_i} **ce**]]]]] F° [_{Op1P} _{O_i} **que**] [_{IP} _{t_i} (**i**) [_{VP} _{t_i} **tombe**]]

Une telle dérivation impliquerait que les enfants gèrent l'existence des deux *ce* mentionnés par les auteurs. On s'attendrait également à trouver des erreurs dans lesquelles les enfants ne feraient intervenir aucun déplacement ; on aurait alors la structure de départ : [_{CopP} **est** [_{SC} *que ce*]], ce qui n'est pas le cas dans notre corpus (mais la raison est peut-être l'âge « avancé » des enfants). Par ailleurs, rappelons le faible taux de réalisation de cette structure (quatorze fois en tout) et le fait qu'elle n'alterne pas dans le corpus avec la forme *qui+est+ce+qui*. De plus, la Figure 2 ci-dessus montre que Maxime à lui seul produit quasiment 80 % de ces énoncés, soit un total de onze formes qui ne se répartissent que sur trois types de phrases différents, comme détaillé en (57) ci-dessous. Ces différents éléments ne jouent pas en faveur de l'hypothèse de la computation syntaxique de cette forme mais davantage en faveur de l'utilisation figée d'une locution.

(57) Les onze énoncés *qu+est+c(e)+qui* de Maxime :

- (a) MAX: *qu+est+c(e)+qui* se cache ici ? (x6)
- (b) MAX: *qu+est+c(e)+qui* y+a écrit ? (x1)
- (c) MAX: *qu+est+c(e)+qui* la [: y+a] ici ? (x4)

1.3 *Qui il ?*

Une autre remarque concerne une série d'énoncés produits par Noémie et transcrits sur le modèle suivant :

(58) NOE: *qui (i)l* [*] a ça ? (VI.26a)

Cette transcription évoque le fait que l'enfant produit le pronom interrogatif *qui* suivi du clitique *il*. Cependant, il ne s'agit que de l'une des hypothèses possibles ; en effet, le son [l] pourrait également représenter le clitique objet *le*. Néanmoins, la première hypothèse n'est pas à écarter définitivement dans la mesure où nous verrons dans la section suivante qu'un phénomène similaire survient chez deux enfants, dont Noémie, avec les pronoms relatifs :

(59) NOE: et [/] et ça c' est le xx *qui i(l)* va tout les manger et [/] et là [/] là c'est des fourmis que xx +... (II.7, 535)

1.4 La difficulté à extraire un sujet

Dans le cadre de cette étude, nous constatons donc que l'interrogation qui porte sur le sujet de l'énoncé n'est pas une tournure extrêmement fréquente dans le corpus (99 formes sur un total de 1141 énoncés interrogatifs verbaux, soit 8,7 % des interrogations verbales enfantines) et, qu'à ce stade de leur développement, seuls 65 % des enfants l'ont intégrée sous au moins l'une des quatre formes détaillées ci-dessus. Il est donc intéressant de noter que ces faits font écho à une thématique beaucoup plus vaste développée dans la littérature générative sur l'asymétrie qui est largement constatée par ailleurs concernant l'extraction grammaticale des objets et celle des sujets qui, elle, est agrammaticale. Nous reprenons l'exemple de Rizzi & Schlonsky (2005) afin d'illustrer cette asymétrie mais sans développer davantage cette thématique¹.

(60) Asymétrie dans l'extraction des sujets et des objets :

- (a) Objet : Qui_i crois-tu que Paul va aider ~~qui~~_i ?
- (b) Sujet : *Qui_i crois-tu que ~~qui~~_i va gagner ?

2. Les pronoms relatifs

Continuons notre panorama avec une autre forme de sujet : les pronoms qui introduisent les propositions subordonnées relatives. L'étude de la relativisation représente un domaine de recherches à part entière². Dans le cadre de nos travaux, nous décrivons donc l'utilisation que font les enfants de ce constituant en nous focalisant, à ce stade de nos recherches, sur les informations qui sont susceptibles de nous renseigner sur le statut des clitiques nominatifs³.

En français, le pronom relatif peut prendre deux formes différentes lorsqu'il est sujet : la forme pleine *qui* (cf. (61a) ci-dessous) et sa forme élidée *qu'* (cf. (61b)) ; la forme *que* est donc exclue dans cette fonction (cf. (61c)) :

¹ Cf. Rizzi & Schlonsky (2005:1).

² Cf. par exemple Vergnaud (1973), Godard (1988) et Kayne (1994), pour le français en général, et Hudelot (1980), Labelle (1988), Guasti & Shlonsky (1995), Perez-Leroux (1995) et Labelle (1996) pour l'acquisition de cette structure.

³ Nous laissons donc de côté de nombreuses problématiques liées notamment aux questions de déplacement (opérateur *vs* tête interne par exemple, cf. Vergnaud (1973)) et à la présence de prépositions (dans la mesure où ces dernières relèvent de pronoms relatifs non nominatifs).

(61) Les pronoms relatifs sujets en français :

- (a) TOM: nan i(ls) tombent les livres *qui* sont là . (I.2, 155)
- (b) JUL: un [/] un aut(r)e *qu'* était par+terre . (X.40b, 370)
- (c) XXX: *i(ls) tombent les livres *que* sont là .

2.1 La théorie

Dans le cadre théorique de la grammaire générative, nous avons déjà évoqué les propositions déclaratives qui portent le trait [Decl], les exclamatives [Excl] et les interrogatives [Int]. Les propositions relatives représentent un quatrième ensemble de propositions mais d'un type légèrement différent dans la mesure où, avec la relativisation, nous abordons les questions de subordination et d'antécédent. Comme les autres propositions, les relatives sont un CP et elles portent donc un trait en C qui indique leur type. Suivant Chomsky (1995a) qui met en avant le caractère prédicatif des relatives, ainsi que les travaux de Rizzi (1990) et de Suñer (1998), nous adoptons le terme [prédicatif] pour caractériser ces propositions qui sont enchâssées dans un DP. Ce trait [Pred] est vérifié grâce à la présence de son homologue sur le NP antécédent « par l'intermédiaire du pronom relatif ou de l'opérateur nul »¹ (par exemple : [DP les [NP [+Pred] livres [CP [+Pred] qui tombent]]]). Par ailleurs, les propositions relatives sont obtenues par le biais d'un mouvement A', c'est-à-dire qu'un élément se déplace de sa position initiale dans laquelle il reçoit son thème-rôle (position argumentale A) vers le spécifieur de CP (position non argumentale). Lorsque la langue dispose de pronoms relatifs, ceux-ci sont donc réalisés en spécifieur de CP ; lorsque la langue n'en dispose pas, c'est un opérateur abstrait qui occupe cette position. Nous adaptons la représentation de Suñer (1998) reprise en (62) ci-dessous afin d'illustrer cette dérivation :

(62) Structure des propositions relatives² :

[DP D [NP N_k [CP {prorel_k / Op_k} [C] [TP {prorel_{k} / Op_k}]]]]}

Les formes des pronoms relatifs varient dans les langues du monde. Les relatives et les interrogatives sont néanmoins souvent traitées similairement dans la mesure où, dans certains cas, les pronoms relatifs sont des éléments *qu-*, comme dans les exemples anglais en (63a) et (63b) ci-dessous. Néanmoins, l'exemple (63c) illustre

¹ Suñer (1998:344).

² D'après Suñer (1998:345).

que toutes les relatives ne sont pas obligatoirement introduites par un élément *qu-* en anglais mais qu'un complémenteur seul peut suffire.

(63) Pronoms relatifs et interrogatifs :

- (a) The child *who* saw the film.
« l'enfant qui a vu le film »
- (b) *Who* saw the film ?
« qui a vu le film ? »
- (c) The books *that* fall.
« les livres que tombent »

Les gloses des exemples (a) et (b) ci-dessus semblent également indiquer une identité entre l'élément *qui* dans la relative et l'interrogative en français. Or, depuis Kayne (1976), il est généralement admis que la forme *qui* du français, lorsque celle-ci introduit une proposition relative, n'est pas un élément *qu-* mais l'allomorphe du complémenteur *que* lorsque la position sujet subséquente est vide, ce qui est le cas lorsque le pronom relatif exerce la fonction sujet dans la proposition enchâssée puisque l'argument s'est déplacé de [Spec, TP] où il a vérifié le trait *EPP* vers [Spec, CP], comme l'illustrent les dérivations en (64) ci-dessous, respectivement pour le français et l'anglais. Dans le cadre *GB*, Rizzi (1990) propose une analyse similaire puisque la forme *qui* du français est perçue comme le résultat morphologique de l'accord qui s'opère entre C (dans lequel se trouve le complémenteur *que*) et son spécifieur d'une part et entre C et l'antécédent, par l'intermédiaire de son complément, d'autre part. D'après l'auteur, *qui* est donc le résultat de cet accord simultané, ce qui explique également l'agrammaticalité de la forme *que* en (61c) ci-dessus¹.

(64) Exemples de dérivations :

- (a) [DP les [NP livres_k [CP {qui_k [C]} [TP ~~qui_k~~ tombent]]]]²
- (b) [DP the [NP books_k [CP Op_k [that] [TP Op_k fall]]]]

À ce stade et dans le cadre du programme minimaliste, il reste à justifier le dernier déplacement du pronom relatif lorsque celui-ci se déplace du spécifieur de TP vers celui de CP. En effet, dans ce cadre théorique, tout déplacement vers un spécifieur

¹ Cf. Rizzi (1990:56-57), reformulé en termes minimalistes dans Rizzi (2001).

² L'introduction de *qui* à ce niveau est une simplification : il s'agit d'une matrice de traits qui, après opération morphologique (indiquée par les accolades), prend la forme *qui*.

sert la cause de la vérification. Il doit donc rester un trait fort encore non vérifié en C qui justifie le déplacement de l'élément sujet de la relative vers le spécifieur de cette tête, trait qui n'est pas présent dans les déclaratives par exemple. Suñer (1998) propose que le trait présent sur le C des relatives soit le trait [Pronominal]. Dans un premier temps, sa présence est donc conditionnée par le caractère prédicatif de la proposition. Dans un deuxième temps, la force de ce trait justifie le déplacement ou non d'un pronom relatif dans cette position : lorsque le trait est fort, le déplacement est obligatoire ; lorsque le trait est faible, le pronom relatif reste *in situ* et c'est un opérateur abstrait qui occupe la position du spécifieur de CP (puisque'il faut tout de même qu'il y ait un lien entre la proposition et son antécédent)¹. Cette proposition reprend donc en termes minimalistes la notion d'accord évoquée précédemment. Dans ce cadre, le complémenteur d'une relative dispose donc de deux traits : un trait [Pred] et un trait fort [Pron] qui force le déplacement du sujet de la relative vers le spécifieur de CP.

D'autres analyses de *qui* existent cependant. En effet, Taraldsen (2001) par exemple propose que *qui* représente la fusion du complémenteur *que* et d'un explétif *i* généré en spécifieur de IP. Cet explétif ne doit cependant pas être assimilé à l'explétif *il* dans la mesure où un certain nombre de propriétés diffèrent pour ces deux éléments : leurs formes phoniques (*i* ne prend jamais la forme *il*), leur trait [Nombre] (le verbe d'une relative peut être au pluriel alors qu'avec *il* explétif le verbe est toujours au singulier) et leur trait [Défini] (seul *i* peut avoir un antécédent défini). Rizzi & Schlonsky (2005) développent alors cette approche dans le cadre minimaliste en reformulant notamment les principes de gouvernement des catégories vides et d'extraction des sujets proposés dans Rizzi (1990) sur la base de la nouvelle structure de la périphérie gauche proposée dans Rizzi (1997a). Le principe *EPP* (qui pose que toute proposition doit avoir un sujet) est ainsi reformulé sous l'appellation « Critère du Sujet » (*Subject Criterion*). Ce critère suggère la présence d'une projection fonctionnelle supplémentaire pour le sujet (*Subject P*) qui se trouve alors plus haut que la projection de T et qui « attire un élément nominal dans son spécifieur et détermine l'articulation sujet-prédicat »². Lorsque le sujet atteint cette position, il est alors « gelé sur place » et ne se déplace donc plus dans la mesure où il remplit sa tâche au niveau interprétatif. La structure simplifiée est CP > SubjP >

¹ Cf. Suñer (1998:348) dont nous reprenons une version simplifiée de la proposition.

² Rizzi & Schlonsky (2005:4) d'après Cinque (1999).

TP > vP > VP (cf. (65) ci-dessous qui détaille une dérivation). Ces postulats soulèvent alors des interrogations concernant les déplacements des sujets en spécifieur de CP. En effet, dans ce cadre, le sujet ne pourrait plus effectuer son dernier déplacement vers le spécifieur de CP puisqu'il serait gelé en spécifieur de SubjP (cf. (c) ci-dessous). Renvoyant à la structure proposée dans Rizzi (1997a)¹, les auteurs utilisent alors la projection de la tête Fin qui, dans ce type de dérivation, contient l'explétif *-i*, « réalisation nominale de la tête Fin ». C'est la fusion de Fin (avec ce contenant nominal) et de Subj (cf. (d)) qui satisfait le Critère du Sujet (dans une relation de tête à tête) et qui laisse alors la voie libre à l'opérateur pour se déplacer vers le spécifieur de CP (cf. (e)).

(65) Dérivation d'une relative proposée par Rizzi & Schlosky (2005) :

« l'homme qui va partir »

(a) Le sujet de la relative (Op) a atteint sa position en [Spec, TP] :

[Op_i T [va [t'_i partir t_i]]]

(b) Le tout est fusionné avec la tête Subj :

[Subj [Op_i T [va [t'_i partir t_i]]]]]

(c) *Si Op se déplace en [Spec, SubjP], il est gelé sur place :

*[Op_i Subj [t''_i T [va [t'_i partir t_i]]]]]

(d) (b) est donc fusionné avec la tête Fin et Op se déplace en [Spec, Fin] :

[Op_i Fin [Subj [t''_i T [va [t'_i partir t_i]]]]]]]

(e) Ce qui permet à Op de se déplacer en [Spec, CP] :

[l'homme [Op_i que [t'''_i Fin [Subj [t''_i T [va [t'_i partir t_i]]]]]]]]]

La relativisation implique donc une certaine complexité syntaxique, notamment au niveau de la vérification de traits supplémentaires et de la forme morphologique que prend le complémenteur. Il est donc attendu que ce type de structure ne soit pas présent dans des proportions importantes chez les jeunes enfants et/ou qu'il présente certaines agrammaticalités.

¹ Cf. Rizzi (1997a:257) : ForceP>TopicP>FocusP>TopicP>FinP>IP. Cf. notre Chapitre II pour le détail.

2.2 Les données

Les requêtes adéquates¹ extraient un total de 278 pronoms relatifs *qu(i)* de notre corpus². Ce premier chiffre confirme donc l'utilisation parcimonieuse de cette structure³. De plus, peu de ces énoncés comportent ce que nous appelons une « vraie » proposition relative. En effet, les énoncés repris en (66) ci-dessous indiquent que 67 % des structures sont introduites par un présentatif (*c'est* ou *y+a*, cf. (a) et (b) ci-dessous) et que 24 % d'entre elles comprennent uniquement un syntagme nominal étendu (cf. (c)) ; seuls 9 % des énoncés montrent donc une véritable dépendance syntaxique entre une proposition relative et une proposition principale (cf. (d))^{4,5}.

(66) Les différents types de propositions relatives dans le corpus :

(a) « c'est X qu(i) » (171 occurrences, 61,60 %) :

LUS: c' est moi *qui* pioche ! (X.38b, 2167)

(b) « y+a X qu(i) » (15 occurrences, 5,40 %) :

NOE: y+a Tom *qui* me l' a pris ! (XIII.48d, 515)

(c) « N qu(i) V » (67 occurrences, 24 %) :

ISI: le chien *qui* se baigne . (VII.30a, 651)

(d) « vraies relatives » (25 occurrences, 9 %) :

NOE: j' avais le cochon *qui* va là . (V.23a, 1396)

Sur les vingt enfants du corpus, quatre ne produisent aucune proposition avec un pronom relatif sujet (Laura, Léa, Claire et Zoé) et, chez les seize enfants qui utilisent cette structure, celle-ci émerge souvent lorsque les enfants atteignent leurs trois ans (Lucie 3;1.24, Noémie 3;0.7, Alan 2;11.27, Emma 3;0.2 et Maxime 3;0.0) ou plus tard (Eva 3;10.13 , Téo 3;8.9, Julien 3;6.16 et Célia 3;5.20). Ainsi, seul Tom produit une « vraie » relative dès sa première séance à 2;11.9 (cf. (61a) ci-dessus), ce qui ne permet pas de connaître le moment de l'émergence de cette structure chez cet enfant.

¹ combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:rel|qui" +u @ > Pro_Rel_QUI et combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:rel|qu" +u @ > Pro_Rel_Qu.

² Labelle (1990:102) propose que [ki] puisse en fait être *qu'i(l)*. Nous écartons cette hypothèse dans la mesure où les données montrent également des formes [ki] devant voyelle (par exemple : c'est moi qui est [*] bébé (TOM, II.5, 440).

³ Si nous rapportons ce chiffre aux 11027 verbes fléchis du corpus, nous obtenons un taux de 2,5 %.

⁴ Dans la mesure où les tournures *c'est/y'a X qui* sont généralement analysées comme des présentatifs complexes et non pas comme des phrases complexes.

⁵ L'étude de Hudelot (1980) sur des enfants un peu plus âgés (entre cinq et sept ans) met en avant une répartition similaire avec notamment 60 % de présentatifs complexes et 20 % de syntagmes nominaux étendus.

Il est également intéressant de noter que ces propositions sont rattachées à des syntagmes qui, en fin de dérivation, se retrouvent à l'extrémité droite de l'énoncé, et ce quelle que soit la fonction du DP dans la phrase, comme le montrent les exemples en (67) ci-dessous :

(67) La place des DP étendus dans les énoncés enfantins :

- (a) NOE: j' avais [_{DP}le cochon qui va là] . (V.23a, 1396)
- (b) MAX: elle est où [_{DP}la pièce qui va là encore] ? (X.40c, 767)
- (c) EVA: il est marqué [_{DP} le libou [*] [: hibou] qui frappe les enfants] . (XIII.46e, 938)

Par ailleurs, trois codes %err, indiquant un ajout, une omission ou une substitution en rapport avec un pronom relatif, permettent d'obtenir une vue d'ensemble des agrammaticalités enfantines dans ce domaine. Ces trois possibilités sont illustrées en (68) ci-dessous : l'ajout englobe les énoncés dans lesquels les enfants produisent un complémenteur fléchi ou non et un clitique nominatif (*cf.* (a)), l'omission indique que ni un complémenteur ni un clitique ne sont phonétiquement réalisés (*cf.* (b)) et la substitution marque la présence d'un clitique sans complémenteur (*cf.* (c)).

(68) Les agrammaticalités syntaxiques portant sur les pronoms relatifs sujets :

- (a) Ajout d'un pronom : \$SYN \$ADD \$PRO \$REL (9 occurrences)
 - EMA: un petit renard <qui il> [/] qui il [*] a faim . (XI.41a, 498)
 - CEL: i(l) trouve sa maman <qu' elle> [*] est cachée . (IX.37d, 194)
- (b) Perte du pronom relatif : \$SYN \$LOS \$PRO \$REL (19 occurrences)
 - ALA: c' est pas moi 0 [*] ai mal . (III.13d, 559)
- (c) Substitution du pronom relatif : \$SYN \$SUB \$PRO \$REL (16 occurrences)
 - ANT: c' est moi j' [*] ai fini . (I.3a, 91)

La base de données comprend également des codes %err morphologiques qui mettent en avant les difficultés que peuvent éprouver certains enfants en ce qui concerne l'accord du verbe enchâssé :

(69) Les agrammaticalités morphologiques dans les relatives :

(a) Problèmes d'accord sujet-verbe : \$MOR \$SUB \$AGC \$V¹

QUE: c' est moi qui a [*] le papillon . (IV.19d, 938)

EVA: fermez les yeux ceux qui veut [*] prendre une carte . (XIII.46e, 1174)

(b) Forme non attendue du verbe : \$ALL \$ADD²

TOM: les [/] les chats qui zont [*] [: ont] xxx +... (VIII.33b, 525)

2.3 Commentaires

Chez ces enfants, nous constatons ainsi que l'agrammaticalité peut porter sur le complémenteur et/ou le clitique nominatif. La combinaison de ces deux variables binaires ([±complémenteur] et [±clitique]) et de la possible flexion du complémenteur (mais uniquement lorsque ce dernier est présent) produit un total théorique de six résultats potentiels, dont un seul est considéré comme grammatical en français standard, comme l'indique la dernière colonne du Tableau 43 ci-après³.

Comp	Flex	Clit	Exemple	Occ	Gram
+	+	+	le xx <i>qui i(l)</i> va tout les manger .(NOE, II.7)	2	*
+	+	-	un livre <i>qui saute</i> ! (LUS, IX.34b)	278	✓
+	-	+	ça c' est moi <i>que je l' ai</i> . (NOE, VI.26a)	7	*/✓
+	-	-	non attesté (le livre <i>que tombe</i>)	0	*
-	0	+	c' est que moi <i>0 j' ai</i> allé à la montagne (ZOE, 41d)	16	*
-	0	-	c' est moi <i>0 l' ai</i> pris avant ! (ANT, XI.42)	19	*

Tableau 43 : combinaisons possibles en rapport avec le pronom relatif sujet

L'immense majorité des relatives étudiées est donc grammaticale. Néanmoins, les agrammaticalités mises au jour dans ce corpus apportent des informations intéressantes par rapport aux études existantes. Tout d'abord, toutes les combinaisons potentielles ne font pas surface ; notamment, aucune agrammaticalité du type *le livre que tombe* n'apparaît. Ces données, recueillies chez des enfants à partir de 2;5.5, permettent ainsi d'étendre la constatation faite par Labelle (1988) chez des enfants entre trois et six ans : aucun enfant ne produit de structure relative qui comporte un complémenteur non fléchi seul. L'explication proposée par l'auteur dans le cadre *GB* est la suivante. Dans ce type de structure, Labelle (1990) propose que le

¹ 32 occurrences en tout.

² 3 occurrences dans le cadre d'une relative, produites par trois enfants différents, plus une occurrence dans une principale (NOE : les mamans zarrivent [*], V.23e, 812).

³ La forme *que* + *clitique* est attestée à l'oral chez les adultes mais n'est pas acceptée en français standard.

complémenteur *que* serait suivi d'une catégorie vide ([le livre [que [*e* tombe]]]). Mais ce type de catégorie, pour être licite, doit être « proprement gouverné » (d'après le Principe des Catégories Vides ou *ECP*). Or *que* ne remplit pas les conditions pour gouverner proprement *e*¹. Il peut donc être supposé, comme le fait Labelle (1990), que « la grammaire [de ces enfants] respecte le principe des catégories vides, c'est-à-dire que les enfants savent qu'il est impossible d'avoir une catégorie vide non gouvernée »². En termes minimalistes, une telle configuration ne convergerait pas car aucun élément sujet n'est disponible pour vérifier le trait [Pron] fort en C (d'après les termes de Suñer (1998)) ou la tête Fin ne porte pas l'élément nominal adéquat afin de satisfaire le Critère du Sujet (d'après Rizzi & Schlonsky (2005)). La distinction avec les langues à « sujet nul » est donc visible dans ce cadre puisque ces dernières ont *pro* qui satisfait la vérification (par exemple : *el científico que pro ganó el premio Nobel*³). Si nous retournons à l'utilisation des traits [Pred] et [Pron] en C, cela signifie que, dès lors que l'enfant francophone établit le caractère prédicatif de la proposition, ce qui se traduit par le trait [Pred] et l'utilisation d'un complémenteur en C, le trait [Pron] est également obligatoirement activé (ou *-i* est intégré à Fin). Cette corrélation est notamment visible lorsque nous comparons les propositions relatives qui ne montrent donc jamais d'absence totale de cet accord, aux propositions indépendantes qui, elles, présentent des cas d'absence de sujets et des propositions subordonnées non prédicatives qui illustrent également cette absence⁴ :

(70) ANT: *pa(r)c(e)+que 0* [*] voulais pas qu' i(l) prend [*] le jaune il a dit le vert . (XIII.49c, 434)

L'absence d'accord n'est donc pas conditionnée par la présence ou l'absence du complémenteur en lui-même mais par le trait présent sur ce complémenteur (ou Fin). La formalisation la plus simple à ce stade (mais peut-être trop simple) consisterait alors à postuler que les deux traits [Pred] et [Pron] n'en font qu'un.

De part et d'autre de cette agrammaticalité non attestée, le Tableau 43 répertorie deux formes de propositions : celles qui présentent un complémenteur et celles qui n'en présentent pas. Les énoncés sans complémenteur sont repris en (71) ci-dessous :

¹ Dans la mesure où il ne lui attribue aucun thème-rôle et qu'il n'est pas co-indiqué avec lui.

² Labelle (1990:102).

³ « le scientifique que a gagné le Prix Nobel ». Exemple extrait de Suñer (1998:342).

⁴ Nous étudions ces omissions dans la Section VI.2 ci-après.

(71) Les énoncés sans complémenteur :

- (a) c' est que moi *0 j'* ai allé à la montagne .
- (b) c' est moi *0 l'* ai pris avant !

Pour le cas de figure en (a) ci-dessus, il peut être suggéré que l'enfant n'établit pas de lien prédicatif entre les deux pronoms (en l'occurrence, *moi* et *je*) et que l'énoncé se compose en fait de deux propositions déclaratives indépendantes. L'énoncé (a) ne présente ainsi aucune agrammaticalité et c'est uniquement la dernière proposition de l'énoncé (b) ci-dessus qui est à considérer comme agrammaticale, mais dans un autre cadre que celui de la relativisation puisqu'il s'agit alors du phénomène du « sujet nul » sur lequel nous reviendrons. Il reste ainsi deux types d'agrammaticalités à analyser dans la première partie du Tableau 43, dont chaque type est repris en (72) ci-dessous :

(72) Les énoncés avec complémenteur et clitique nominatif :

- (a) ça c' est moi *que je l'* ai .
- (b) le xx *qui i(l)* va tout les manger .

Ce type d'occurrences est rare dans le corpus (neuf en tout). Néanmoins, les agrammaticalités qu'elles contiennent sont intéressantes à examiner. Tout d'abord, nous remarquons que le type illustré en (a) ci-dessus apparaît sept fois, dont cinq sont produites par la même enfant (Célia), qui par ailleurs n'utilise quasiment pas la relativisation ; les deux autres occurrences de (a) sont produites par Noémie qui n'utilise pas non plus cette structure très souvent (quinze occurrences en tout, ce qui est plus rare que la relativisation chez Tom avec 28 occurrences ou Lucie avec le nombre le plus élevé de 69 occurrences par exemple). Noémie est également l'une des enfants qui produit les occurrences du type (b) ci-dessus, l'autre enfant étant Emma (35 relatives en tout). En termes de traits, la première agrammaticalité provient du fait que le complémenteur, bien que présent, ne se trouve pas dans une relation spécifieur-tête qui lui permettrait de vérifier son trait [Pron]. Cela est dû au fait que le sujet de la relative fait surface en tant que clitique en spécifieur de TP. Il faut donc expliquer pourquoi le dernier déplacement attendu dans ce type de dérivation (vers le spécifieur de CP) n'intervient pas. La morphologie du clitique montre que celui-ci ne présente pas de déficience en termes de traits puisqu'il est spécifié pour la [Personne], le [Genre], le [Nombre] et le [Cas]. La matrice de traits n'est donc pas en cause. Précédemment, nous avons associé les deux traits [Pred] et

[Pron] de C. *A priori*, ce n'est donc pas le trait [Pron] en lui-même qui fait défaut. La question qui se pose face aux énoncés de type (71a) ci-dessus relève davantage du choix même du type de proposition. En effet, si nous postulons que les enfants (et les locuteurs en général) qui produisent ce type de structures considèrent qu'il s'agit de déclaratives et non de prédicatives, la dérivation s'explique alors par le fait que le trait [Pron] n'est pas présent puisque, en amont, le trait [Pred] lui-même n'est pas activé. Le dernier mouvement vers le spécifieur de CP n'a donc pas lieu d'être. Cette stratégie ne présente alors qu'une seule des caractéristiques des relatives : la subordination, traduite par la présence d'un complémenteur en C. Labelle (1996), dont les données comprennent également ce type de propositions, propose que la relation de prédication entre la proposition et l'antécédent soit alors établie plus tard, après LF. C'est à ce moment seulement que « la proposition entière est alors interprétée au niveau pragmatique comme caractérisant la tête nom [...] »¹. Sans entrer dans le détail de cette hypothèse qui relève de la pragmatique, nous constatons que l'idée d'une interprétation globale et plus lâche de la proposition, en dehors d'une utilisation du trait syntaxique [Pred] est une piste envisageable. Ces subordonnées se seraient donc pas syntaxiquement dérivées par les locuteurs en tant que prédicatives mais uniquement en tant que déclaratives enchâssées. Au niveau acquisitionnel, il serait alors cohérent d'expliquer la présence de ce type de structures par leur plus grande simplicité syntaxique au regard de la difficulté supplémentaire que représente l'utilisation du trait [Pred] (dans la mesure où ce dernier implique également le trait [Pron]). Notons que ce type d'énoncés et l'analyse qui en est faite ne falsifie pas, à ce stade, l'hypothèse que [Pred] et [Pron] ne fasse qu'un. Si nous reprenons l'analyse dans les termes de Rizzi & Schlonksy (2005), ce type de dérivation trouve également une explication. En effet, ce type d'énoncé semble correspondre à la structure présentée en (65) ci-dessus et reprise en (73b) ci-dessous. Il manquerait alors la projection Fin qui contient l'explétif *-i* : le sujet est gelé en spécifieur de SubjP et la fusion avec le CP fait émerger le complémenteur *que* sous sa forme non fléchie. De nouveau, cette dérivation montre que la subordonnée n'a aucun caractère prédicatif ; l'opérateur n'a donc pas lieu d'être et le sujet fait surface sous forme d'un clitique nominatif.

¹ Labelle (1996:72).

(73) La structure *que* + *clitique nominatif* :

- (a) ça c' est moi *que* je l' ai .
- (b) *Si Op se déplace en [Spec, SubjP], il est gelé sur place :
*[Op_i Subj [t'_i T [va [t'_i partir t_i]]]]
- (c) Fusion avec C :
[l'homme [que [Op_i Subj [t'_i T [va [t'_i partir t_i]]]]]]]]

Il reste les deux énoncés produits par Noémie et Emma (respectivement 2;9.29 et 3;2.14) dans lesquels les enfants produisent à la fois un complémenteur fléchi et un clitique nominatif. Ces deux phrases complètes sont reprises en (74) ci-dessous :

(74) La structure *qui* + *clitique nominatif* :

- (a) NOE: et [/] et ça c' est le xx *qui* *i(l)* va tout les manger et [/] et là [/] là c'est des fourmis que xx +... (II.7, 535)
- (b) EMA: un petit renard <*qui* *il*> [/] *qui* *il* a faim . (XI.41a, 498)

Ces deux relatives présentent l'agrammaticalité commune de proposer deux formes concomitantes de sujets. En effet, le clitique nominatif *il* se trouve en spécifieur de TP et le pronom relatif *qui* vérifie le trait [Pron] de C en spécifieur de CP (ou Fin porte l'explétif *-i*). Une telle dérivation n'est en théorie pas réalisable dans la mesure où deux copies des mêmes traits pronominaux sont réalisés concomitamment (ou deux éléments satisfont le Critère du Sujet). D'ailleurs le corpus de Labelle (1988) ne comporte pas de telles réalisations. Néanmoins, ces deux occurrences présentent un intérêt certain dans le cadre de notre thèse sur le statut morphosyntaxique des clitiques nominatifs en français. Que ce soit dans les langues à « sujet nul » ou les langues à « sujet non nul », cette redondance est agrammaticale :

(75) Structures agrammaticales dans tous les cas :

- (a) *Un petit renard *qui* *il* a faim.
- (b) *El científico *quien* ganó el premio Nobel.¹

Elle est néanmoins maintenant attestée chez deux enfants francophones. Une hypothèse quant à cette cooccurrence d'un pronom relatif fléchi et d'un clitique nominatif consisterait alors à envisager que le clitique n'est pas considéré par l'enfant comme un pronom au sens étymologique du terme et qu'il ne remplit donc

¹ Exemple extrait de Suñer (1998:342).

pas la fonction attendue du sujet syntaxique. Dans ce cadre, l'enfant analyserait alors ce clitique comme un affixe du verbe et procéderait à la dérivation développée en (76) ci-dessous, d'après le modèle de Rizzi & Schlosky (2005) proposé en (65) ci-dessus :

(76) Dérivation sur le modèle de Rizzi & Schlosky (2005) :

« un renard qui il-a faim »

(a) Le sujet de la relative (Op) a atteint sa position en [Spec, TP] :

[Op_i T [t_i il-a faim]]

(b) Le tout est fusionné avec la tête Subj :

[Subj [Op_i T [t_i il-a faim]]]

(c) *Si Op se déplace en [Spec, SubjP], il est gelé sur place :

*[Op_i Subj [t''_i T [t_i il-a faim]]]

(d) (b) est donc fusionné avec la tête Fin et Op se déplace en [Spec, Fin] :

[Op_i Fin [Subj [t''_i T [t_i il-a faim]]]]

(e) Ce qui permet à Op de se déplacer en [Spec, CP] :

[un renard [Op_i que [t'''_i Fin [Subj [t''_i T [t_i il-a faim]]]]]]]

Dans le mesure où l'enfant analyserait le verbe comme étant préfixé (*il-a*), la configuration semble alors similaire à celle des langues « à sujet nul » dans laquelle le verbe porte une flexion morphologiquement riche. Or, cette analyse se heurte à l'agrammaticalité illustrée en (75) ci-dessus pour l'espagnol et que nous reprenons en (77) ci-dessous qui montre que, dans ce type de langues, « les relatives peuvent uniquement être introduites par l'équivalent du complémenteur [que], les pronoms relatifs réalisés [étant] complètement bannis dans ce contexte »¹.

(77) *El científico *quien* ganó el premio Nobel.²

✓El científico *que* ganó el premio Nobel.

2.4 Vers une fission du paramètre *pro-drop*

Produisant un pronom relatif, Emma ne semble donc pas considérer sa langue comme une langue « à sujet nul »³ ; par contre, si la démonstration en (76) ci-dessus est correcte, l'enfant considère tout de même l'élément clitique *il* comme un préfixe

¹ Suñer (1998:342).

² Exemples extrait de Suñer (1998:342).

³ Le taux relativement bas de « sujets nuls » dans ce corpus (3,16 %) illustre d'ailleurs ce propos. Cf. Section VI.2 ci-après pour le détail.

verbal et non comme un argument syntaxique. Il semble donc que l'enfant se trouve dans une situation « intermédiaire », entre un système à « sujet non nul » et un système dans lequel la morphologie verbale est néanmoins riche, avec la différence importante par rapport aux langues à « sujet nul » que cette richesse morphologique n'est pas suffixale mais préfixale. Or, un paramètre binaire, tel que le Paramètre *pro-drop*, ne semble pas pouvoir rendre compte d'une telle situation. Il nous semble alors qu'il faille distinguer et rendre compte de ces deux phénomènes séparément : d'une part, la morphologie verbale *suffixale*, qui est riche dans les langues à « sujet nul » et plutôt pauvre dans les autres langues et d'autre part, la morphologie verbale *préfixale*, qui ne semble pas implémentée dans les langues à « sujet nul » (puisque la suffixation est déjà distinctive) mais qui vient combler les lacunes de la morphologie suffixale pauvre des langues à « sujet non nul »¹. Nous nous dirigeons ainsi vers une formalisation des faits observés sous la forme, non pas d'un paramètre unique, mais de *deux* paramètres, chacun régissant un « côté » du verbe et intervenant de façon séquentielle selon le principe du miroir décrit dans le Chapitre II².

Dans la section précédente, nous avons vu que des énoncés qui semblent être bâtis sur un modèle similaire sont produits par Noémie dans le cadre d'interrogations. Ce type d'occurrences est repris en (78) ci-dessous.

(78)NOE: *qui (i)l [*] a ça ?* (VI.26a)

L'hypothèse esquissée ci-dessus, qui consiste à attribuer un statut morphologique au clitique nominatif, serait alors également à même de rendre compte de la cooccurrence entre le pronom interrogatif, argument sujet, et l'éventuel clitique nominatif, préfixe verbal. Mais avant de développer cette hypothèse³, nous devons continuer notre tour d'horizon des formes que peut prendre le sujet chez nos petits informateurs. Au cours de la Section II, les clitiques nominatifs ont été décrits et nous avons constaté que 2898 d'entre eux sont des clitiques qui possèdent la même référence qu'un autre élément dans l'énoncé ; nous les avons qualifiés de « clitiques doublés ». Voyons à présent ce que sont ces doubles.

¹ Nous revenons sur ce point dans le Chapitre V, page 361.

² Cf. Chapitre II, Section III.3.

³ Ce que nous ferons au Chapitre V.

IV. LES CO-RÉFÉRENTS D'UN CLITIQUE

1. Le doublement

Sur la ligne morphosyntaxique %mor de la base de données, un code est attribué à chaque mot graphique. L'unité codée n'est donc pas le syntagme mais chaque élément qui est susceptible de le composer. Les syntagmes qui doublent les clitiques sont alors identifiés d'après la nature de leur tête, comme illustré en (79) ci-dessous : pronominale (cf. (a) et (b) ci-après), nominale (cf. (c) et (d)), adjectivale¹ (cf. (e) et (f)) ou adverbiale (cf. (g)).

(79) Les différentes catégories grammaticales des doubles :

- (a) JUL: *moi j(e) veux chanter une musique* . (XII.45, 634)
- (b) LUS: *et après i(ls) s' en vont tous* . (VI.24a, 794)
- (c) EMA: *hou i(l) fait le loup* . (VII.27b, 1589)
- (d) ALA: *Matteo il est pas là +...* (VI.24a, 1233)
- (e) ANT: *après i(l) est revenu l' autre* . (VI.24a, 1633)
- (f) TEO: *oh où il est le jaune ?* (XI.43b, 162)
- (g) EKT: *nan là c' est Nathan* . (VI.26e, 844)

L'extraction de chaque catégorie relève alors d'un type de requête différent dans CLAN et, à l'instar des pronoms clitiques, seuls les pronoms toniques possèdent un code qui permet de les identifier en tant qu'éléments nominatifs. En revanche, les autres catégories grammaticales ne sont pas étiquetées d'après leur fonction. Nous commençons donc notre investigation par ces pronoms toniques.

1.1 Les pronoms toniques

Les pronoms toniques de la base de données incluent les pronoms traditionnellement appelés pronoms « disjoints » mais aussi quelques pronoms démonstratifs du type *celui+là*. Tous ces pronoms présentent un code qui spécifie qu'ils doublent un clitique, soit à sa gauche (dg), soit à sa droite (dd), comme le détaille le Tableau 44 ci-après. L'extraction de ces items est donc directe².

¹ Il peut également être considéré qu'il s'agit de syntagmes nominaux dont la tête N est vide.

² Requête : freq -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:ton:d*:nom*|" +s"pro:dem:d*:nom*|" +u @>TonD_Freq_Children.

1	pro:ton:dg:nom moi&1S	15	pro:dem:dg:nom ceux+là
2	pro:ton:dd:nom moi&1S	16	pro:dem:dd:nom ceux+là
3	pro:ton:dg:nom toi&2S	17	pro:dem:dg:nom celles+là
4	pro:ton:dd:nom toi&2S	18	pro:dem:dd:nom celles+là
5	pro:ton:dg:nom lui&3S	19	pro:ton:dg:nom nous&1P
6	pro:ton:dd:nom lui&3S	20	pro:ton:dd:nom nous&1P
7	pro:ton:dg:nom elle&3S	21	pro:ton:dg:nom vous&2P
8	pro:ton:dd:nom elle&3S	22	pro:ton:dg:nom vous&2P
9	pro:ton:dg:nom ça&3S	23	pro:ton:dg:nom:pol vous&2P&POL
10	pro:ton:dd:nom ça&3S	24	pro:ton:dd:nom:pol vous&2P&POL
11	pro:dem:dg:nom celui+là	25	pro:ton:dg:nom eux&3P
12	pro:dem:dd:nom celui+là	26	pro:ton:dd:nom eux&3P
13	pro:dem:dg:nom celle+là	27	pro:ton:dg:nom elles&3P
14	pro:dem:dd:nom celle+là	28	pro:ton:dd:nom elles&3P

Tableau 44 : liste des pronoms doublant un clitique

Dix-neuf formes différentes sont produites par les enfants donnant un total de 1972 pronoms toniques doublant un clitique. Sur les 2898 clitiques doublés, 68,05 % sont donc doublés par ce type de pronom, indiquant que la combinaison <pronom tonique + clitique> ou <clitique + pronom tonique> est la configuration dominante lors d'un doublement. Le Tableau 45 ci-après indique que ces toniques sont quasiment tous de P1 et P3 (96,55 % du total), ce qui est en adéquation avec les informations recueillies sur les personnes des clitiques doublés. L'information nouvelle qu'apporte ce tableau concerne le placement du pronom tonique par rapport au clitique. En effet, les pourcentages ci-dessous indiquent que le doublement se fait quasiment toujours à gauche (82,71 % des occurrences), particulièrement avec les pronoms qui incluent le locuteur (*moi* : 89,61 % et *nous* : 88,00 %) et les pronoms pluriel (*vous* et *eux* : 100 %) qui présentent les taux les plus élevés.

Formes	Totaux	%	Gauche	%	Droite	%
pro:ton:d*:nom moi&1s	1174	59,53	1052	89,61	122	10,39
pro:ton:d*:nom toi&2s	38	1,93	30	78,95	8	21,05
pro:ton:d*:nom lui&3s	86	4,36	63	73,26	23	26,74
pro:ton:d*:nom elle&3s	9	0,46	5	55,56	4	44,44
pro:ton:d*:nom ça&3s	611	30,98	441	72,18	170	27,82
pro:dem:d*:nom celui+là	23	1,17	13	56,52	10	43,48
pro:dem:d*:nom celle+là	1	0,05	0	0,00	1	100,00
pro:ton:d*:nom nous&1p	25	1,27	22	88,00	3	12,00
pro:ton:d*:nom vous&2p	2	0,10	2	100,00	0	0,00
pro:ton:d*:nom eux&3p	3	0,15	3	100,00	0	0,00
pro:ton:d*:nom elles&3P	0	0,00	0	0,00	0	0,00
pro:dem:d*:nom ceux+là	0	0,00	0	0,00	0	0,00
pro:dem:d*:nom celles+là	0	0,00	0	0,00	0	0,00
Totaux	1972	100,00	1631	82,71	341	17,29

Tableau 45 : répartition des pronoms toniques enfantins doublant un clitique

Ces données montrent également que les enfants utilisent le pronom *nous* uniquement sous sa forme tonique. L'extraction des énoncés correspondants indique par ailleurs que ce pronom [Pluriel] double toujours le même clitique [Singulier] *on*¹ :

(80) Exemples de l'utilisation du pronom tonique *nous* :

- (a) NOE: *nous on* est toutes seules hein ! (IX.36, 2494)
- (b) LEA: *on* voit Lucie *nous* . (IX.37d, 258)

Ces combinaisons confirment dès cet âge : (i) le statut ambivalent de *on* qui est sémantiquement [Pluriel] mais morphologiquement [Singulier] et (ii) le fait que, en cas de valeurs différentes pour un même trait, c'est bien la valeur du clitique que l'on retrouve sur le verbe fléchi, pas celle de l'élément qui double le clitique.

Les données présentées dans le Tableau 45 ci-dessus représentent des moyennes établies à partir de l'ensemble des données des vingt enfants sur les treize séances. Nous devons donc à présent vérifier que ces moyennes ne masquent pas des disparités individuelles importantes et nous devons déterminer si le modèle de doublement dégagé ci-dessus, majoritairement à gauche, présente une évolution au fil des séances ou si ce taux est stable du début à la fin des enregistrements.

¹ Sauf dans deux énoncés produits par Alan au cours de la même séance dans lesquels les codes associent à tort *nous* à *ça* : ANT: et on [/] on [/] toi on dirait un chat et *nous ça* [*] dirait des garçons : en paraphrasant par « on nous prendrait pour des garçons », *nous* est en fait complément d'objet direct et a pour attribut *pour des garçons*.

1.2 La latéralisation du doublement des pronoms toniques

La Figure 3 ci-après exploite les données de six enfants choisis car ce sont les enfants les plus représentés au fil des séances (Lucie par exemple a été enregistrée à chacune des treize séances). Cette figure utilise les taux de doublement à gauche par rapport à l'ensemble des doublements de cette catégorie pour chacun des enfants au cours de dix séances. Les lignes montrent parfois de grandes disparités d'une séance à une autre mais elles ne présentent pas de modèle développemental particulier : on ne constate ni croissance ni décroissance particulière du taux sur cette tranche d'âge dans la mesure où les derniers taux se trouvent globalement dans la même fourchette que les premiers.

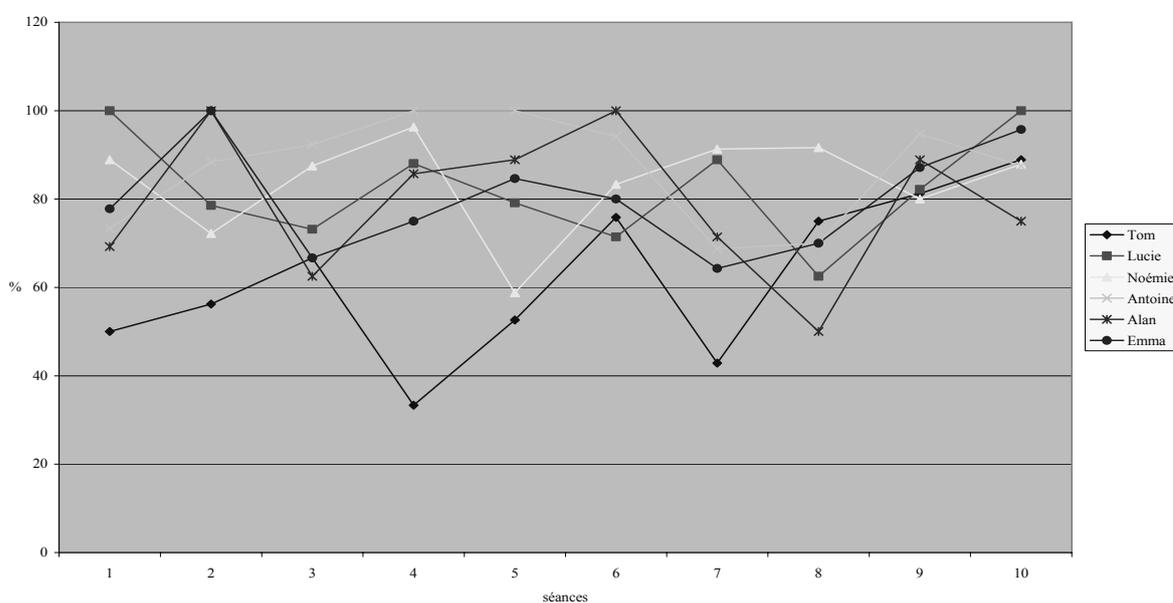


Figure 3 : évolution du doublement à gauche (pronoms toniques)

Dans cette tranche d'âge, la latéralisation du doublement n'est donc pas corrélée à un critère développemental. En effet, si cela avait été le cas, nous aurions constaté une évolution de ces taux.

Une seconde piste que nous explorons concernant cette latéralisation est le type d'énoncé concerné. Nous cherchons à déterminer si la latéralisation est conditionnée par l'aspect déclaratif, exclamatif ou interrogatif de l'énoncé. Les résultats des requêtes correspondantes¹ sont présentés dans le Tableau 46 ci-dessous :

¹ Requêtes sur le modèle : `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro:*.dg:nom|^*!pro:*.dd:nom|^*\" +u @ > TonDG&Decl_ChildrenAll.`

Toniques	Decl		Excl		Int		Totaux	
	N	%	N	%	N	%	N	%
Gauche	1144	86,80	292	80,00	77	44,25	1513	81,48
Droite	118	8,95	58	15,89	86	49,43	262	14,11
Gauche&Droite	56	4,25	15	4,11	11	6,32	82	4,42
Totaux	1318	100,00	365	100,00	174	100,00	1857	100,00

Tableau 46 : latéralisation et type d'énoncé (pronoms toniques)

Cette investigation permet alors d'affiner les résultats précédents dans la mesure où elle met au jour une troisième catégorie de doublement non encore détaillée : celle qui double le clitique concomitamment à gauche et à droite comme le montrent les exemples en (81) ci-dessous et pour laquelle nous adoptons le terme de « triplement ».

(81) Triplement du clitique nominatif :

- (a) NOE: *et moi j' ai un chat moi* . (IV.18a, 46)
- (b) QUE: *et ça c' est un fort ça !* (VIII.33a, 916)
- (c) EMA: *oh ça c' est quoi ça ?* (XIII.46e, 224)

Une étude approfondie de cette structure est proposée en Section IV.2 ci-dessous. Pour l'heure, en ce qui concerne une éventuelle corrélation entre latéralisation et type d'énoncé, les données montrent que le triplement est présent dans les trois cas de figure, dans une proportion relativement identique pour les trois catégories. Aucune corrélation ne peut donc être établie dans le cadre du triplement. Par contre, des tendances plus prononcées sont révélées pour le doublement. En effet, 86,80 % des énoncés déclaratifs et 80 % des énoncés exclamatifs comprenant un double tonique relèvent d'un doublement à gauche alors que seules 44,25 % des interrogations utilisent ce modèle. Il semble donc y avoir une partition entre les énoncés non-interrogatifs, majoritairement latéralisés à gauche et les énoncés interrogatifs dont la latéralisation est moins tranchée mais dont la forme la plus courante est la latéralisation à droite (49,43 %). Ces tendances sont illustrées en (82) ci-dessous.

(82) Latéralisation et types d'énoncés :

(a) Déclaratif et Gauche :

EKT: et ben *moi j'* ai fait du vélo . (X.39, 2022)

(b) Exclamatif et Gauche :

ZOE: *ça c'* est un lapin ! (I.2, 1307)

(c) Interrogatif et Droite :

EVA: et *elle* me prête celui+là *elle* ? (II.9d, 974)

Ces chiffres et conclusions sont néanmoins parcellaires dans la mesure où ils demandent à être complétés par ceux qui concernent les doubles autres que les pronoms toniques vers lesquels nous nous tournons à présent.

1.3 Les autres constituants

Il reste à nous pencher sur les 926 occurrences qui contiennent les clitiques doublés par un constituant qui n'est pas un pronom tonique, soit 31,95 % des doubles. Or, coder chaque nom, pronom indéfini, adjectif ou adverbe selon sa fonction dans chaque énoncé aurait considérablement allongé le travail d'étiquetage. En effet, dans la mesure où le nombre des possibilités initiales de codage pour ces items lexicaux est limité dans le dictionnaire original conçu par Christophe Parisse, le codage de ces éléments est conservé tel quel (contrairement au codage très ciblé que nous avons développé pour les pronoms toniques). Dans notre base de données, ces constituants ne disposent donc pas d'une étiquette indiquant leur fonction sur la ligne %mor¹ ; seule leur nature est indiquée, accompagnée la plupart du temps, lorsqu'il s'agit de noms communs, par leur genre et leur nombre, comme illustré en (83) ci-dessous. Globalement, il s'agit de syntagmes nominaux (*cf.* (a)-(c)) et d'adverbes (*cf.* (d)) :

¹ CLAN offre néanmoins la possibilité de coder les constituants d'après leur fonction mais ce codage nécessite l'implémentation d'une ligne dépendante supplémentaire (%syn) qui, elle, ne comporte pas de codes morphologiques.

(83) Étiquetage des items lexicaux :

- (a) JUL: ça existe pas *les chaussures* qui volent ! (XIII.48b, 805)
%mor: n|chaussure&_FEM-_PL
- (b) ALA: non *Maxime* tu joues pas ! (I.3a, 360)
%mor: n:prop|Maxime
- (c) TEO: ah il est là *le jaune* . (XI.43b, 168)
%mor: adj|jaune
- (d) EMA: *là* c' est t(r)ès beau . (I.3d, 1847)
%mor: adv:place|là

Dans ce cadre, l'extraction de ce type d'éléments d'après sa fonction n'est donc pas possible par le biais d'une requête directe. Dans un premier temps, il faut alors procéder indirectement, ce que montre le détail de la requête en (84) ci-après qui spécifie que l'on extrait les énoncés qui comprennent un ou plusieurs clitiques doublés, sauf lorsque ceux-ci sont accompagnés par un pronom tonique doublant ou par un pronom démonstratif doublant. Dans un second temps, le résultat de cette requête est filtré manuellement afin de déterminer la nature de chaque constituant ainsi que sa latéralisation.

(84) Extraction des clitiques doublés par autre chose qu'un pronom tonique :

```
combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor  
+s"pro:cli:d:nom*|^!(pro:ton:d*:nom*|^+pro:dem:d*:nom*)" +u +x @ >  
CliD&DP_Combo_Children
```

La requête détaillée ci-dessus livre un total de 948 combinaisons dont 840 impliquent un syntagme nominal (soit 89 % des occurrences). Ce total est légèrement supérieur au chiffre annoncé de 926, ce qui s'explique par le phénomène de triplement mentionné dans la section précédente. En effet, dans ce cadre, certains clitiques sont doublés non pas par un seul constituant mais par deux éléments concomitamment. La répartition de tous ces éléments d'après leur latéralisation est indiquée dans le Tableau 47 ci-après :

Types	Totaux		Gauche		Droite		G&D		Autres	
	N	%	N	%	N	%	N	%	N	%
Syntagmes Nominaux	840	100	412	49,05	414	49,29	14	1,67	0	0,00
Adverbes	105	100	94	89,52	11	10,48	0	0,00	0	0,00
Autres ¹	3	100	0	0,00	0	0,0	0	0,0	3	100
Totaux	948	100	506	53,38	425	44,83	14	1,48	3	0,32

Tableau 47 : latéralisation des doubles enfantins (autres que pronoms toniques)

Contrairement aux pronoms toniques fortement présents à gauche du clitique (82,71 %), l'ensemble de ces autres constituants ne montre pas une latéralisation prononcée. Le taux de doublement à gauche illustre néanmoins une légère prédominance de ce côté (dans 53,38 % des cas) mais cette tendance s'efface lorsque l'on considère les différents types de formes séparément. En effet, les syntagmes nominaux livrent des taux quasiment identiques de part et d'autre alors que les adverbes affichent une nette préférence pour la latéralisation à gauche du clitique (89,52 % des cas). Néanmoins, comme précédemment, ces taux sont établis à partir des données des vingt enfants sur l'ensemble des séances. Ainsi, il convient également de vérifier si une évolution survient au fil des séances et si cette latéralisation est corrélée à une variable quelconque.

1.4 La latéralisation du doublement des autres constituants

Comme le montre la Figure 4 ci-après, l'étude de l'utilisation des syntagmes nominaux chez les mêmes six enfants que précédemment ne semble pas livrer de modèle développemental particulier. En effet, les données varient beaucoup d'une séance à une autre et d'un enfant à un autre et la proportion des syntagmes nominaux présents à gauche d'un clitique ne croît ni ne décroît d'une façon particulière à l'intérieur de cette catégorie. La latéralisation de ce constituant ne semble donc pas liée à un critère développemental dans cette tranche d'âge.

¹ Il s'agit de propositions entières comme dans l'exemple : ANT: et qui+c'est+qui va gagner c' est moi . (VII.29c, 232).

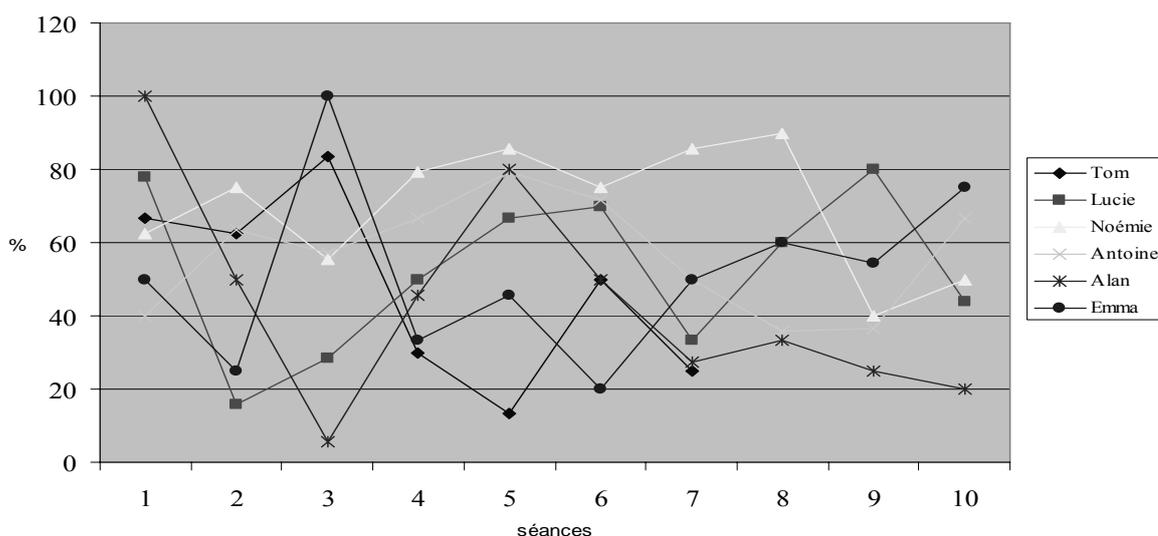


Figure 4 : évolution du doublement à gauche (syntagmes nominaux)

Nous nous tournons donc vers une autre variable : le type d'énoncé. Toutes les occurrences comportant un syntagme nominal sont ainsi examinées à travers ce second filtre. Les 840 énoncés sont ainsi répartis d'après le type de l'énoncé (déclaratif, exclamatif ou interrogatif) et la latéralisation du doublement, comme l'illustrent les entrées du Tableau 48 ci-après :

SN	Decl		Excl		Int		Totaux	
	N	%	N	%	N	%	N	%
Gauche	339	63,25	46	46,00	27	13,24	412	49,05
Droite	186	34,70	53	53,00	175	85,78	414	49,29
Gauche & Droite	11	2,05	1	1,00	2	0,98	14	1,67
Totaux	536	100,00	100	100,00	204	100,00	840	100,00

Tableau 48 : latéralisation et type d'énoncé (syntagmes nominaux)

Les pourcentages en gras mettent alors en avant deux tendances : la majorité des énoncés déclaratifs est produite avec un syntagme nominal latéralisé à gauche (63,25 % des déclaratives) et la quasi-totalité des interrogatives (85,78 %) comprend un syntagme latéralisé à droite. Ces deux types s'opposent donc et l'hypothèse d'une corrélation entre le type d'énoncé et la place du double peut ainsi être évoquée de nouveau. Par contre, contrairement aux énoncés avec un pronom tonique, les faits sont moins tranchés en ce qui concerne les exclamatives qui ne présentent qu'une légère dominance du côté droit (à 53 %). Néanmoins, les faits du Tableau 48 ci-dessus peuvent être rapprochés des constatations issues des chiffres du Tableau 46 sur les pronoms toniques. En effet, dans les deux cas de figure, le type déclaratif s'oppose nettement au type interrogatif et les formes exclamatives présentent

globalement une tendance plus proche des déclaratives que des interrogatives. Déclaratives et exclamatives sont donc subsumées sous une seule catégorie « non-interrogative » afin de synthétiser les observations précédentes. Les moyennes obtenues sont présentées dans le Tableau 49 ci-dessous qui étaye alors l'hypothèse de la corrélation entre la latéralisation du double et le type de l'énoncé : à gauche pour la non-interrogation, à droite pour l'interrogation.

Pronoms & SN	Non-int		Int		Totaux	
	N	%	N	%	N	%
Gauche	1821	78,53	104	27,51	1925	71,38
Droite	415	17,90	261	69,05	676	25,06
Gauche & Droite	83	3,57	13	3,44	96	3,56
Totaux	2319	100,00	378	100,00	2697	100,00

Tableau 49 : latéralisation et type d'énoncé (synthèse)

La place du double est donc liée à au moins l'une des spécificités de l'interrogation (comme par exemple la présence du trait [qu-] en C ou la présence d'un élément qu-). Néanmoins, il serait également intéressant de chercher la raison pour laquelle 27,51 % des interrogations échappent à ce modèle. Étudier le détail des structures des interrogatives nous mènerait cependant au-delà du cadre de notre présente recherche dans la mesure où la thématique de l'interrogation en français représente à elle-seule un domaine de recherches à part entière¹ et que les âges des enfants les plus jeunes dans notre base de données se situent aux alentours de l'âge de 2;6, âge auquel il est estimé que l'acquisition de l'interrogation est déjà réalisée. Néanmoins, si nous limitons nos investigations aux clitics nominatifs et à leurs doubles dans notre corpus, nous confirmons qu'un nombre conséquent de possibilités est déjà effectivement présent dans la tranche d'âge 2;5-4;0 (cf. (85), (86) et (87) ci-dessous). La série en (86) illustre notamment les deux cas de figure qui permettraient de distinguer un « doublement » d'une « dislocation » (cf. (c) vs (d))². Nous constatons également que certaines séquences ne sont pas encore représentées, notamment celles qui impliquent à la fois l'inversion du clitique et du verbe et la présence d'un pronom tonique (cf. les séquences non attestées ci-dessous).

¹ Cf. Kayne (1972), Obenauer (1981), Bouchard & Hirschbühler (1986), Sportiche (1999), Hulk & Pollock (2001), Munaro & Pollock (2005), etc. sur les diverses problématiques soulevées par les structures interrogatives en français et Crisma (1992), Hulk (1996), Hamann (2000), Strick (2007), etc. sur l'acquisition de ces structures.

² La place du syntagme nominal par rapport à l'élément qu- est en effet considéré comme l'un des critères qui permettrait de distinguer les deux structures. Cf. Chapitre I, Section III.1.2 pour le détail.

(85) Interrogation et latéralisation à droite (4 structures, 7 réalisations) :

(a) Clitique_i ^ V ^ double_i :

LUS: *il était cassé ton biberon ?* (III.13g, 183)

EVA: *tu veux faire ça toi tout seul ?* (I.3c, 825)

(b) Clitique_i ^ V ^ qu- ^ double_i :

CEL: *elle est où ta photo ?* (II.9d, 345)

EKT: *il est où ça ?* (VIII.33c, 604)

(c) Qu- ^ clitique_i ^ V ^ double_i :

NOE: *qu+est+c(e)+(qu') elle fait Zoé ?* (V.23e, 994)

ROM: *qu+est+c(e)+que c' est ça ?* (I.3b, 523)

(d) Qu- ^ V ^ clitique_i ^ double_i :

CLA: *où es tu Maman_Ours?* (VII.30b, 1027)

XXX: non attesté avec un pronom tonique

(86) Interrogation et latéralisation à gauche (5 structures, 9 réalisations) :

(a) Double_i ^ clitique_i ^ V :

TOM: *eh Noémie tu veux ça ?* (IX.36, 946)

EVA: *et moi je peux ?* (IV.19d, 533)

(b) Double_i ^ clitique_i ^ V ^ qu- :

ANT: *ma(is) la vache elle est où ?* (III.13a, 1043)

NOE: *ça c' est à qui ?* (IV.18b, 594)

(c) Double_i ^ qu- ^ clitique_i ^ V :

NOE: *et ben le monsieur qu+est+c(e)+qu' i(l) fait ?* (II.7, 950)

NOE: *et lui comment i(l) s' appelle ?* (IV.18a, 1187)

(d) Qu- ^ double_i ^ clitique_i ^ V :

EVA: *et pourquoi Lola elle veut pas jouer ?* (III.10a, 801)

ALA: *pourquoi lui i(l) s' appelle un [/] un castor ?* (XII.45, 1226)

(e) Double_i ^ qu- ^ V ^ clitique_i :

ALA: *poisson où es tu ?* (VI.24a, 147)

XXX: non attesté avec un pronom tonique

(87) Interrogation et triplement (2 structures, 3 réalisations) :

(a) Triple_i + clitique_i + V + Triple_i :

EMA: *ça c' est <non maillot+se+bain> [*] [: son maillot+de+bain] ça ?*
(VI.24b, 864)

(b) Triple_i + clitique_i + V + *qu-* + Triple_i :

LEA: *la vache elle est où la vache ?* (III.13a, 1085)

NOE: *et c(el)ui+là ça [*] va où lui ?* (V.23a, 418)

Ces nouvelles données confirment donc que les enfants utilisent à la fois les questions à *qu-* *in situ* ainsi que l'antéposition de l'élément *qu-* dès deux ans et demi. Par ailleurs, les enfants de cet âge ne produisent aucune agrammaticalité concernant le placement relatif du double et de l'élément *qu-* (cf. (88) ci-dessous). Par contre, il est intéressant de remarquer que quelques énoncés comprennent à la fois soit un élément interrogatif et sa réponse (cf. (89)a-c)), soit deux éléments *qu-* (cf. (89d)).

(88) Séquences non attestées : *[Clitique_i ^ V ^ double_i ^ *qu-*] :

(a) XXX: **elle est ta photo où ?*

(b) XXX: **il est ça où ?*

(89) Doublement de l'adjectif :

(a) MAX: *où il est à+côté+de [*] toi le [*] oiseau ?* (III.13e, 631)

(b) QUE: *elles sont où les cartes là ?* (IV.19d, 792)

(c) MAX: *où il est le soleil là ?* (III.13e, 1072)

(d) TOM: *où tu vas mettre où [*] les livres ?* (VIII.33c, 1876)

2. Le triplement

Que ce soit lors de l'étude des pronoms toniques ou celle des syntagmes nominaux, nous avons vu que l'appellation « doublement » se révèle parfois inappropriée dans la mesure où, dans un nombre limité de cas, l'énoncé ne présente pas deux éléments co-indicés mais trois : un clitique nominatif et deux pronoms toniques, deux syntagmes nominaux ou un élément de chaque catégorie.

2.1 Un clitique, deux pronoms toniques

Tous les cas de figure impliquant deux pronoms toniques présents dans le corpus sont illustrés en (90) ci-dessous¹. Les données comprennent également une occurrence de quadruplement du clitique nominatif, reprise en (90i).

(90) Triplements avec deux pronoms toniques :

- (a) ALA: *moi j' ai pas encore fini moi* [*] . (I.3a, 28)
- (b) ROM: *moi c' est le chat moi* [*] . (V.23a, 1656)
- (c) EMA: <toi t(u) es pas> [/] *toi t(u) es pas un cochon toi* [*] dit maman . (XII.44e, 32)
- (d) ALA: après *lui i(l) va à l' école lui* [*] . (XI.42, 147)
- (e) MAX: ah mais *ça i(l) va où ça* [*] ? (XIII.47b, 781)
- (f) ZOE: *ça c' est bleu ça* [*] ! (I.3a, 801)
- (g) JUL: *ça ça existe pas ça* [*] . (XIII.48a, 823)
- (h) NOE: et *c(el)ui+là ça* [*] va où *lui* ? (V.23a, 418)
- (i) ZOE: *moi je veux un aut(r)e livre moi* [*] *moi* [*] . (II.9f, 757)

Comme l'indique l'astérisque dans les transcriptions ci-dessus, le triplement du clitique peut également être repéré dans la base de données en tant qu'agrammaticalité. En effet, cette configuration est codifiée sur la ligne %err en tant que « répétition d'un pronom » (\$SYN \$RED \$PRO)². Ces deux requêtes livrent un total de 79 triplements (dont un quadruplement) produits par quinze enfants différents. Ces premiers résultats sont ensuite filtrés en enlevant les occurrences produites une seule fois (*moi+je+moi+moi*, *celui-là+ça+lui*, *toi+tu+toi*, *ça+il+ça*) ainsi que les occurrences produites par un seul enfant (ROM : *moi+c'+moi*) afin d'obtenir la vue d'ensemble représentative de cette structure présentée dans le Tableau 50 ci-dessous :

¹ Requête : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"pro*:dg:nom|^*^pro*:dd:nom|^" +u @ >TonDG&DD.

² La requête correspondante est alors : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"\$SYN^\$RED^\$PRO" +u @ > TonDG&DD_Err.

	<i>moi+je+moi</i>	<i>lui+il+lui</i>	<i>ça+c'+ça</i>	<i>ça+ça+ça</i>	Totaux
QUE	1		2		3
EVA			2		2
EKT	1				1
TEO	2			1	3
TOM	2				2
JUL	2		1	1	4
CEL	2				2
LUS	3	1	2		6
NOE	1				1
ANT	2				2
ROM	1				1
ALA	2	1			3
EMA	3		4	1	8
MAX	17		13		30
ZOE	1		1		2
Totaux	40	2	25	3	70
%	57,14	2,86	35,71	4,29	100,00
Enfants ≠	14	2	7	3	15

Tableau 50 : répartition des triplements par type et par enfant

Le Tableau 50 ci-dessus montre que la structure la plus utilisée et la plus largement manipulée est celle qui implique P1, *moi je moi* (57,14 % ; 14 enfants différents). Les enfants utilisent donc les clitiques nominatifs de trois façons : seuls, doublés et triplés. Le triplement n'est cependant pas une exclusivité enfantine. Le corpus ne comporte qu'une seule occurrence de triplement adulte (*cf.* (91a) ci-dessous) mais il est observé par ailleurs que d'autres formes de triplement peuvent survenir chez l'adulte (*cf.* (91b)) :

(91) Le triplement chez l'adulte :

- (a) KAT: ah oui *ça c'* est les cadeaux *ça* [*] . (II.7, 1683)
- (b) XXX: *Moi perso je* pense que...

Marquer plusieurs fois le même trait semble représenter une forme de redondance qui n'est pas sans rappeler la surdétermination à l'œuvre dans le lexique¹. Il est donc intéressant de chercher la cause de cette redondance. Deux hypothèses sont possibles : (i) le doublement, ou *a fortiori* le triplement, sert à insister sur le trait [Personne] et/ou [Locuteur] ; la redondance a alors pour fonction l'emphase ou la focalisation ; (ii) le marquage multiple d'un même trait est le signe qu'au moins un

¹ *Cf.* Dalbera (2006).

des items n'est plus perçu par le locuteur comme support suffisant, d'où un besoin de remotivation ; le doublement n'est alors pas une structure emphatique mais simplement la structure par défaut qui permet de véhiculer l'information voulue. Cette partition nous renvoie à la distinction entre « dislocation » et « doublement » mentionnée au Chapitre I. En effet, la première hypothèse équivaut à une analyse en termes de « dislocation » où le clitique est considéré comme un argument et où le tonique vient renforcer cet argument alors que la deuxième hypothèse fait référence à une analyse où le clitique n'est pas perçu comme argument du verbe mais comme préfixe verbal et où le tonique fait alors office d'argument. L'un des critères qui permet de trancher entre ces deux possibilités est celui de la fréquence de cette structure par rapport à celle où le clitique n'est pas doublé. En effet, comme mentionné précédemment, si son taux est élevé, le doublement ne peut être considéré comme une mise en valeur puisqu'il acquiert alors davantage la valeur de la structure « par défaut ». En d'autres termes, si la structure majoritaire est le doublement, cette structure perd de sa valeur distinctive par rapport à la structure non doublée. Il faut donc comparer les taux de chaque structure, ce que nous faisons dans la Section VII en fin de chapitre.

2.2 Un clitique, deux syntagmes

On trouve des triplements également parmi les syntagmes, comme illustré en (92) ci-dessous, sous forme de syntagmes nominaux (avec un nom propre comme en (a) ou un nom commun comme en (b)) ou de syntagmes prépositionnels (*cf.* (c)) :

(92) Triplements avec deux syntagmes :

- (a) EMA: *Lola elle m' a pris la place Lola* [*]. (X.38c, 189)
LUS: *Coco_Lapin il est dans mon lit Coco_Lapin* [*]. (III.13e, 1429)
- (b) ANT: *mon sac il est super mon sac* . (X.39, 1063)
CEL: *tu sais 0 [*] les cloches i(ls) [*] sont venus <à m(a)> [/] à ma maison les cloches* [*]. (IX.37c, 19)
- (c) CEL: *à+côté+d(e) Lou_Anne c' est mon lit à+côté+d(e) Lou_Anne* [*]. (IX.37d, 727)

De nouveau, l'utilisation des codes %err, en l'occurrence \$\$SYN \$RED \$N, permet une extraction directe de ce type d'énoncés¹. Les données illustrent alors le fait que ce type de triplement est rare puisqu'il ne survient que treize fois dans la totalité du corpus. Néanmoins, nous notons qu'il ne s'agit pas d'une structure particulière à un enfant dans la mesure où les occurrences se répartissent entre sept enfants différents² et que, par ailleurs, les enfants qui produisent ce type de triplements utilisent également le triplement avec les pronoms toniques. Nous remarquons également que le triplement peut être représenté par trois éléments différents, comme illustré en (93) ci-dessous :

(93) Triplements avec syntagme et pronom tonique :

- (a) EMA: *lui il est ici le petit poussin* [*]. (X.38b, 547)
- (b) EKT: *Quentin toi tu casses pas mon château d+accord ?* (VIII.33a, 1668)

Toutefois, comme l'indique l'astérisque dans l'énoncé (a) uniquement, ce type de structures n'est pas signalé systématiquement dans l'ensemble de la base de données par le biais d'un codage %err³ ; il n'est donc pas possible d'extraire ce type d'énoncés par une requête simple et, à l'heure actuelle, nous ne disposons donc pas d'un compte exact de ces phrases.

¹ Requêtes : `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"$SYN^$RED^$N" +u @ > Triple_DP` et `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"$SYN^$RED^!($N+$PRO)" +u @ > Triple_DP&Pro`.

² Il s'agit d'Antoine (4 occurrences), Lucie (2), Célia (2), Maxime (2), Noémie (1), Emma (1) et Julien (1).

³ En l'occurrence, il s'agit du code \$\$SYN \$RED. Requête : `combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err +s"$SYN^$RED^!($N+$PRO)" +u @ > Triple_DP&Pro`.

2.3 Le triplement et la théorie des copies

Le dernier type de triplement mis à part, le corpus comprend ainsi une centaine d'énoncés qui comporte deux constituants identiques, réalisés de part et d'autre de l'ensemble <clitique+verbe fléchi> et co-indicés avec le clitique nominatif. Ce type de structures nous paraît alors intéressant à étudier dans le cadre de la théorie des copies que nous avons illustrée au Chapitre II avec un exemple de Chomsky (1995a) que nous reprenons en (94) ci-dessous :

(94) Illustration de la théorie des copies¹ :

John wondered [which picture of himself] Bill saw [which picture of himself]
« Jean se demandait [quelle photo de lui-même] Bill a vu [quelle photo de lui-même] »

Cette théorie stipule que, lorsqu'un élément se déplace dans la structure syntaxique, ce mouvement équivaut à « recopier » cet élément dans sa nouvelle position. Ainsi, deux exemplaires du même élément sont présents dans la structure². Néanmoins, dans la forme phonique de l'énoncé, seule une copie de cet élément est effectivement réalisée. C'est donc lors de l'interface entre la syntaxe et le système articulatoire et perceptuel que toutes les copies sauf une sont effacées. Cet (ou ces) effacement(s) est(sont) expliqué(s) par le fait qu'un énoncé qui comprend plusieurs copies d'un même élément ne satisfait pas au Principe de Linéarisation requis par la composante phonologique. En effet, pour que cette dernière prenne en charge un énoncé avec succès, tous les constituants de cet énoncé doivent être ordonnés linéairement ; cette information quant à l'ordre des items est donc déjà contenue dans la structure syntaxique et est transmise à la composante phonologique par le biais d'un algorithme tel que celui proposé par Kayne (1994) sous le nom d'Axiome de Correspondance Linéaire (*Linear Correspondance Axiom* ou *LCA*) que nous détaillons en (95) ci-dessous :

¹ Chomsky (1995a:206).

² Dans le cadre du Programme Minimaliste, cette théorie remplace la théorie des traces qui ne remplit pas la Condition d'Inclusivité (cf. Chapitre II, Section III.1). Il peut y avoir plus de deux copies, si l'élément se déplace plusieurs fois.

(95) L'Axiome de Correspondance Linéaire (*LCA*)¹ :

Un item lexical α précède un item lexical β si et seulement si :

- (a) α c-commande asymétriquement β ou
- (b) un XP dominant α c-commande asymétriquement β .

Dans ce cadre, un même élément α ne peut être prononcé deux fois dans la mesure où α c-commanderait et serait concomitamment c-commandé par le ou les élément(s) intermédiaire(s) β . Il est alors généralement postulé que seule la copie la plus élevée dans la structure hiérarchique est conservée par le système articulatoire et perceptuel car c'est la copie pour laquelle tous les traits ininterprétables sont effacés ; elle est donc « plus digeste aux interfaces »². Néanmoins, comme le soulignent Hornstein *et al.* (2005), cette théorie rend également compte : (i) de la possibilité que ce soit une copie située plus bas dans la structure qui est réalisée, « si la réalisation phonétique d'une copie supérieure pose des problèmes à la composante phonologique »³ et (ii) de la possibilité que plusieurs copies soient réalisées dans le même énoncé « si certaines de ces copies subissent une fusion morphologique » avant la prise en charge par la composante phonologique, perdant ainsi leur statut de « copie » aux yeux de cette composante et devenant alors invisibles pour cette dernière⁴.

Comme mentionné précédemment pour d'autres thématiques, cette théorie des copies constitue à elle seule un domaine d'investigations. Dans le cadre de nos travaux actuels, nous ne pouvons donc qu'initier des pistes en vue de recherches ultérieures ; notamment, nous suggérons que l'analyse des triplements enfantins, effectuée à travers le prisme de cette théorie des copies, puisse apporter des informations nouvelles sur la difficile distinction entre « dislocation » et « doublement ». À cet effet, reprenons en (96) ci-dessous les deux types principaux

¹ Cette version est la reformulation du *LCA* de Kayne (1994) telle que proposée par Hornstein *et al.* (2005:227) dans le cadre du Programme Minimaliste.

² Hornstein *et al.* (2005:242).

³ Cf. Hornstein *et al.* (2005:242) et l'exemple du roumain qui autorise la réalisation d'éléments *qu-* dans leur position initiale lorsqu'une réalisation dans la position d'arrivée enfreint certains principes phonologiques (en l'occurrence, l'interdiction de réaliser de façon adjacente plusieurs items identiques).

⁴ Cf. Hornstein *et al.* (2005:243-245), d'après une formalisation de Nunes (2004) et d'après des données de Koopman (2004) sur le *vata* (langue africaine de la famille nigéro-congolaise) qui montrent qu'un verbe peut être réalisé deux fois dans un énoncé : dans une position haute où le verbe est focalisé et une position basse, qui est sa position d'origine.

de triplements dans lesquels un même élément est réalisé deux fois : cet élément est soit un pronom tonique (*cf.* (a)), soit un nom (*cf.* (b)) :

(96) Les triplements enfantins :

(a) ALA: *moi j' ai pas encore fini moi* [*] . (I.3a, 28)

(b) EMA: *Lola elle m' a pris la place Lola* [*]. (X.38c, 189)

Si ces deux réalisations identiques sont bien deux copies d'un même élément, cela implique que cet élément est généré là où se trouve la copie la plus basse dans la structure (la plus à droite dans l'ordre linéaire) et que cet élément s'est déplacé vers autant de positions qu'il y a de copies. Le présupposé est donc qu'il s'agit d'un constituant qui doit se déplacer au cours de la dérivation. Or, dans le cadre du débat autour de l'existence et/ou de la différenciation entre une « dislocation » et un « doublement », nous suggérons que ce présupposé puisse éventuellement apporter un éclairage nouveau à la discussion. En effet, *a priori*, depuis au moins les travaux fondateurs de Koopman & Sportiche (1991) sur la position des sujets, il est admis que le sujet est un argument qui se déplace dans la structure syntaxique, de sa position initiale en spécifieur de *vP* (lorsqu'il est agent) vers sa position en spécifieur de *TP*, là où il vérifie le trait *EPP*¹. Dans ce cadre, l'élément copié (et indiqué en italique dans les énoncés (a) et (b) ci-dessus) représenterait le *sujet syntaxique*, ce qui relèguerait le clitique nominatif au rang de préfixe morphologique. Une telle analyse des copies nous placerait alors dans le cadre d'une explication en termes de « doublement » avec un argument sujet et un clitique nominatif préfixé au verbe.

Contrairement à l'analyse du sujet (qui est donc consensuelle quant à ses déplacements), les études sur la « dislocation » ne sont pas unanimes. En effet, les deux hypothèses existent quant au mouvement d'un tel constituant dans la mesure où certains postulent que la dislocation est le résultat d'un déplacement alors que d'autres favorisent l'hypothèse selon laquelle l'élément disloqué est adjoint à la structure syntaxique dans sa position effective et qu'il ne se déplace donc pas². Pour le moment, face à ce débat, nous proposons quelques remarques. Si les éléments identiques dans les énoncés enfantins présentés en (96) ci-dessus sont bien deux copies d'un même constituant (représentant ainsi les étapes d'un déplacement), ceux-

¹ *Cf.* Chapitre II, Section III.3 pour le détail.

² *Cf.* par exemple Alexiadou (2006) et De Cat (2007) qui proposent tous les deux une synthèse des principaux points qui font débat. Ces deux auteurs optent pour l'absence de déplacement. Le point de vue opposé est soutenu par exemple par Kayne (1994) ou Sportiche (1996).

ci ne peuvent pas être analysés en termes de « dislocations » (ni à gauche, ni à droite) par les linguistes qui proposent par ailleurs que la dislocation n'implique pas de déplacement. En d'autres termes, si la dislocation est une adjonction sans déplacement et si les deux occurrences de *moi* et de *Lola* sont des copies, ces deux items ne sont pas des dislocations et ils représentent alors des doublements, à moins bien sûr de postuler deux dislocations identiques et concomitantes. *A priori*, il nous semble néanmoins que l'existence de ces triplements favorise légèrement l'hypothèse du doublement dans la mesure où celle-ci n'est pas falsifiée par de telles structures mais au contraire légèrement renforcée. *A contrario*, ces triplements n'œuvrent pas en faveur d'une analyse en termes de dislocations, tout au moins pour ceux qui considèrent que la dislocation est une adjonction sur place. Nous ouvrons donc le débat quant à une éventuelle corrélation entre l'existence de ces triplements enfantins et l'hypothèse du doublement. Dans tous les cas, il faudra également rendre compte de la réalisation phonétique de ces éléments identiques puisque, pour le vata mentionné précédemment, la « fusion morphologique » du verbe à l'origine de la réalisation des deux copies ne semble pas être une explication que l'on puisse étendre aux énoncés enfantins.

V. LES SYNTAGMES NOMINAUX SEULS

Jusqu'à présent, nous avons étudié les sujets enfantins qui comprennent un clitique nominatif seul, doublé voire triplé. Puis nous avons continué notre tour d'horizon par l'analyse des éléments *qu*- interrogatifs et relatifs. Afin de compléter l'analyse des sujets grammaticaux phonétiquement réalisés par les enfants, il nous reste alors à examiner les syntagmes nominaux utilisés seuls. Cet ordre de progression peut paraître surprenant si l'on considère que nous n'abordons une structure qui est considérée comme la structure canonique de la phrase française qu'à un stade avancé de nos recherches. Cependant, le Chapitre I a montré la différence importante qui existe entre le français « standard » et le français que nous appelons « spontané ». De plus, notre premier corpus enfantin met en avant une absence totale de cette structure chez les jeunes enfants. Voyons ce que révèlent nos nouvelles données.

1. Définition

Par « syntagmes nominaux seuls », nous entendons les syntagmes non doublés par un clitique, qui comprennent un nom commun, un nom propre ou un élément autre qu'un clitique nominatif, c'est-à-dire notamment un pronom indéfini tel que *personne, chacun, tout* (codés simplement « pro » dans cette base de données) ou un pronom démonstratif tel que *celui+là* (codé « pro:dem:nom »). Deux difficultés surgissent alors dans le cadre de cette étude. Celles-ci ont déjà été mentionnées concernant l'analyse des doubles. La première ressortit à la question du syntagme. En effet, dans le cadre du codage %mor, seuls les mots graphiques reçoivent un code, pas les syntagmes. La seconde difficulté relève de la notion de fonction grammaticale qui n'est indiquée que pour les pronoms par le biais du terme « nominatif ». Ainsi, notre grille de codage morphosyntaxique ne présente pas de code direct fléchant les syntagmes nominaux sujets. La requête utilisée pour extraire ce type de constituants est donc une programmation indirecte dont le détail se trouve en (97) ci-dessous. Cette demande permet d'extraire les énoncés qui présentent la combinaison suivante : (i) l'absence des éléments étudiés jusqu'à présent (clitiques nominatifs et éléments *qu-*) et (ii) la présence d'un verbe fléchi en excluant l'impératif.

(97) Requête utilisée pour l'extraction des syntagmes nominaux sujets :

```
combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor
+s"!pro:cli*:nom*|^!pro:int|qui^!pro:lo:int:nom|^v*|^v^!v*|^imp&|^!pro:rel|
*" +x +u @ > DP_Children
```

Malgré ces restrictions, cette requête extrait tout de même deux types de sujets : les sujets recherchés et les énoncés verbaux qui ne présentent aucun sujet (sur lesquels nous revenons dans la Section VI.2)¹. L'extraction finale se fait donc manuellement afin de relever et de comptabiliser les deux types d'énoncés séparément.

¹ Plus nous ajoutons des restrictions, plus nous risquons de perdre des énoncés qui auraient été à étudier. C'est pourquoi nous n'ajoutons pas la restriction 0 [*] qui marque l'absence d'un élément car un syntagme nominal sujet peut survenir dans le même énoncé que l'absence d'un autre constituant (une préposition par exemple).

2. L'extraction des données

Le corpus présente alors un total de 39 énoncés enfantins qui comprennent un sujet sous forme de syntagme nominal non doublé ; deux de ces énoncés sont interrogatifs. Les exemples en (98) ci-dessous illustrent cette structure¹ :

(98) Exemples d'énoncés présentant un syntagme nominal seul :

- (a) EVA: alors *Max* prend la poule en chocolat . (XIII.46a, 1063)
- (b) EKT: *le facteur* n' est pas passé . (IX.36, 2126)
- (c) LUS: *tout+le+monde* a gagné . (XI.43c, 1492)
- (d) ALA: *sa maman et son papa* vont faire le lait . (VII.29a, 287)
- (e) EMA: *quelqu+un* veut choisir un p(e)tit bonbon ? (XIII.46e, 420)

Cependant, au sein de cette liste de 39 énoncés, il est visible que deux types d'occurrences se côtoient. En effet, dix de ces énoncés n'appartiennent pas au registre spontané des enfants dans la mesure où deux d'entre eux reprennent des paroles de chansons (*cf.* (99a) et (99b) ci-dessous) et huit autres, tous produits par Noémie les uns à la suite des autres, sont des répétitions d'énoncés produits par la maîtresse qui, le matin même, avait raconté cette histoire aux enfants au cours de leur séance de « lecture » en groupe (*cf.* (99c)).

(99) Les énoncés non spontanés :

- (a) EKT: le facteur n' est pas passé . (IX.36, 2126)
- (b) LUS: vole mon petit chagrin xxx . (XIII.46b, 765)
- (c) NOE: <pas moi > [""] dit le chat . (VIII.32a, 885)

Sur nos treize séances d'enregistrements, les enfants ont donc produit un total de 29 énoncés spontanés du type DP[^]Verbe. Si nous prenons comme nombre de référence le total des verbes fléchis enfantins (soit 11027) et si nous comptabilisons les deux types d'énoncés décrits ci-dessus (soit un total de 39), le pourcentage que représente cette structure dans les productions enfantines est de 0,35 %. Celle-ci est donc, de toute évidence, loin d'être productive à cet âge. Il convient alors de s'interroger sur les raisons de la rareté dans le système enfantin d'une structure qui est, *a contrario*, la structure canonique du français « standard ».

¹ La liste exhaustive de ces énoncés se trouve en Annexes, page 382.

3. Les caractéristiques de cette structure

Ces énoncés combinant un syntagme nominal seul avec un verbe fléchi, bien que peu nombreux, ne sont pas aléatoirement répartis dans le corpus. Trois phénomènes les caractérisent en effet : leur distribution entre les différents enfants, leur progression au fil des séances et leur composition.

3.1 Une augmentation des occurrences au fil des séances

Concernant les locuteurs, onze enfants sur les vingt qui composent le groupe ne produisent jamais cette structure. Sur les neuf qui l'utilisent, la répartition n'est pas homogène dans la mesure où Lucie produit à elle seule 46,15 % de ces énoncés, suivie par Noémie (avec 23,08 % mais dont huit énoncés sur neuf ne sont pas spontanés) et Eva (avec 10,26 %). Les six enfants restants utilisent ce type d'occurrences une ou deux fois sur l'ensemble des séances. En ce qui concerne la ventilation de ces énoncés sur les treize séances, lorsque nous considérons les 29 énoncés spontanés, les pourcentages du Tableau 51 ci-après montrent une tendance à une utilisation globalement croissante de cette structure, avec un taux nettement plus élevé pour la dernière séance. L'étude des enregistrements subséquents s'annonce donc prometteuse dans ce domaine également.

Séances	1	2	3	4	5	6	7	8	9	10	11	12	13	Totaux
Totaux	1	1	2	1	0	1	2	2	1	3	3	3	9	29
%	3,4	3,4	6,9	3,4	0,0	3,4	6,9	6,9	3,4	10,3	10,3	10,3	31,0	100,0

Tableau 51 : l'évolution des syntagmes nominaux utilisés seuls

3.2 Les éléments indéfinis et quantifiés

La troisième caractéristique de ces syntagmes concerne la nature des éléments qui les composent. En effet, nous retrouvons dans ce type d'énoncés l'élément indéfini *quelqu'un* (cf. (a) ci-dessous) et les éléments quantifiés *personne* et *tout+le+monde* (cf. (b) et (c)) qui sont tous trois emblématiques du débat sur le statut des clitiques nominatifs¹ :

¹ Cf., entre autres, Chapitre I, Section III.1.2.3.

(100) Les « fameux » éléments indéfinis et quantifiés :

(a) EVA: et *quelqu+un* regarde la main . (XIII.46a, 1565)

(b) LUS: alors je le retourne et *personne* me le retourne comme ça . (XII.44c, 500)

(c) LUS: *tout+l(e)+monde* a déjà fait . (XI.43a, 513)

Quelqu+un apparaît un total de quinze fois dans le corpus, dont deux fois en tant que sujet (Eva et Emma), sans la présence d'un clitique nominatif concomitant. *Personne* est produit sept fois en tout dont deux fois dans un énoncé non verbal (Zoé). Sur les cinq énoncés verbaux restants (Eva, Hector et Lucie), *personne* est sujet une seule fois (Lucie). Ainsi, la seule occurrence de ce pronom utilisé en tant que sujet dans l'ensemble de ces données survient dans le cadre de cette structure où le DP est utilisé sans le clitique nominatif. *Tout+le+monde* est produit six fois, dont cinq occurrences sont de Lucie et trois sont des sujets sans clitics concomitants. Nous avons donc un total de six éléments indéfinis ou quantifiés qui, lorsqu'ils sont sujets, ne sont pas accompagnés d'un clitique nominatif. Nous avons vu que les partisans du statut syntaxique du clitique expliquent cette non cooccurrence par le fait que, lorsqu'un clitique nominatif et un syntagme nominal surviennent dans le même énoncé, le clitique occupe la position canonique du sujet (en spécifieur de TP) et le syntagme occupe la position disloquée du Topique ; or un Topique ne peut pas être indéfini ou quantifié dans la mesure où il est postulé qu'il s'agit d'une information déjà connue de l'interlocuteur. Dans ce cadre, dès qu'un élément indéfini ou quantifié est utilisé, il apparaît seul, dans la position du sujet, ce qui est le cas pour ces énoncés enfantins. Les occurrences de ce corpus ne soulèvent donc aucune polémique puisqu'elles n'incluent pas de clitics nominatifs. Elles indiquent également que, lorsque les enfants utilisent *quelqu+un*, *personne* et *tout+l(e)+monde*, ils ne les manipulent pas concomitamment avec un clitique nominatif. Chez Lucie (dès la deuxième séance) et chez Eva et Emma (à partir de la treizième séance), il semble donc qu'une certaine conception de la notion de quantification empêche l'utilisation concomitante d'un clitique. En revanche, avant cette période chez ces trois enfants et dans tout le corpus pour les autres, un quantifieur tel que l'article indéfini *un* peut éventuellement être associé à un clitique nominatif, comme le montrent les cinq occurrences du corpus reprises en (101) ci-

dessous¹. Mais la rareté de ces occurrences² indique qu'une certaine incompatibilité est présente dans le système linguistique des enfants dès cette période³.

(101) Indéfinis et clitiques nominatifs :

- (a) ALA: *un nuage il* arrive et xx la pluie ! (I.1b, 53)
- (b) EMA: *un garçon i(l) [/]* il [/] il attrape un zoeu [*] [: œuf] . (IX.34a, 1136)
- (c) MAX: où [=? ouh] *il* [*] est où *un* [*] *autre page* ? (VIII.31b, 1211)
- (d) CEL: <un canard> [/] *un canard i(l) [/]* i(l) crie avec sa maman et son papa et son frère . (XI.41b, 403)
- (e) EMA: *il* était tout seul *un* [*] *petit lapin* . (XI.41d, 1389)

Néanmoins, notamment par le biais des données enregistrées au cours des deux années qui suivent cette première collecte, nous prévoyons des discussions intéressantes dans la mesure où les éléments *quelqu+un*, *personne* et *tout+l(e)+monde* émergent progressivement dans le système enfantin et, de fait, apparaissent ainsi plus fréquemment au fur et à mesure que les enfants grandissent. Nous avons ainsi d'ores et déjà noté que Tom produit plusieurs cooccurrences du pronom indéfini *personne* et du clitique *il* à 4;5.12 et nous avons relevé une cooccurrence de *tout le monde* et de *il* chez une petite fille de huit ans au cours d'une émission télévisée (cf. les exemples en (102) ci-dessous). Par ailleurs, ce type de cooccurrence n'est pas non plus exclu du discours adulte, comme l'illustrent les exemples repris en (103) ci-dessous.

¹ Requête : combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"det|un&MASC&SING^pro:cli:d:nom|il&3S" +u +x @ > Un&Il.

² Le corpus comprend un total de 1092 articles indéfinis *un*.

³ Dans ce cadre, il faudrait alors également étudier les autres articles indéfinis.

(102) Cooccurrences enfantines attestées :

(a) TOM: *personne i(l) m(e) l' a dit* . (Corpus 3, IX.35b/c)

(b) SEF: *tout le monde i(l) s'assoit dessus* .¹

(103) Cooccurrences adultes attestées :

(a) *Personne i(l) veut ça*.²

(b) *Le monde entier i(l) te remercie pour ce que tu as fait*.³

Au regard de l'incompatibilité attestée entre l'aspect indéfini/quantifié d'un syntagme et de sa présence en position disloquée de Topique, l'existence de tels énoncés implique que ces syntagmes doivent alors être considérés comme présents dans une position plus basse (donc celle du spécifieur de TP). Cette première constatation implique à son tour que le clitique nominatif ne peut pas occuper cette même position dans la hiérarchie syntaxique. L'existence de toutes ces occurrences représente donc un argument en faveur du statut morphologique du clitique dans ces réalisations.

4. Une structure rare : pourquoi ?

Concernant la structure DP[^]Verbe, les données montrent donc que les jeunes enfants ne l'utilisent que de façon parcimonieuse. D'après ce que nous avons vu jusqu'à présent, ils sont donc nettement plus prompts à insérer un clitique nominatif (doublé ou non) devant un verbe fléchi que de l'omettre et ainsi produire une structure sans clitique. Une tâche de production induite conduite par l'enseignante de la classe des tout-petits vient corroborer cette observation. En effet, l'enseignante raconte une petite comptine mimée à l'ensemble des enfants. Il s'agit de l'histoire d'un escargot qui attend la pluie pour aller se promener (*cf.* (104) ci-dessous). Le ton et la structure des énoncés produits par l'adulte sont extrêmement surveillés dans ce contexte. En effet, plusieurs des caractéristiques du français « standard » sont présentes (*cf.* (104a)) : pas de sujets doublés et deux élisions du sujet due à la coordination. Ce qui est très intéressant avec la « répétition » d'Alan (*cf.* (104b)), c'est que l'enfant introduit systématiquement un clitique préverbal devant tous les verbes fléchis, y compris là où l'enseignante n'en avait pas prononcé.

¹ En parlant d'un lit. Occurrence produite par Séphora, 8 ans (émission D&Co sur M6 le 08/06/08).

² Adulte, émission sur M6 le 20/07/08.

³ Adulte, journal télévisé de France 2 le 07/07/09, à l'occasion de la mort de Michael Jackson.

(104) Production induite d'une comptine mimée (I.1b) :

(a) Modèle donné par l'enseignante :

SOF: *l'escargot* dort parc(e)+qu' *il* fait beau dehors .

SOF: *un nuage* arrive et fait tomber la pluie !

SOF: *l'escargot* sort une corne deux cornes et s' en va se promener .

(b) Production subséquente d'Alan :

ALA: *l'escargot i(l)* dort parc(e)+qu' <i(l) fait> [/] *i(l)* fait beau dehors .

ALA: un nuage *il* arrive et xx la pluie !

ALA: deux cornes .

ALA: *i(l)* s' en va se promener .

Les principes d'économie paraissent donc malmenés dans ce cadre car l'enfant préfère une structure qui, en apparence, semble plus lourde dans la mesure où, lorsqu'il utilise un syntagme nominal sujet, il préfère l'utilisation conjointe d'un clitique. Mais les apparences sont peut-être trompeuses et il faudrait alors observer les faits en supposant le contraire. Reconsidérons donc les deux structures (a) et (b) ci-dessous et postulons que (b) est plus économique que (a). Quelle analyse faudrait-il alors faire de (b) pour que ce postulat soit correct ?

(105) Les deux structures :

(a) SOF: *l'escargot* dort...

(b) ALA: *l'escargot i(l)* dort...

5. Vers une redistribution de la vérification des traits

Ce qui est coûteux en syntaxe, ce sont les déplacements ; d'ailleurs, ils n'ont lieu que s'ils sont indispensables à la vérification de traits ininterprétables. Pour que (b) ci-dessus soit plus économique que (a), il faut donc postuler : (i) non seulement que les déplacements effectués par Alan ne sont pas plus nombreux que ceux effectués par l'adulte ; (ii) mais également que la dérivation de l'enfant possède une caractéristique attractive supplémentaire qui fait que l'enfant l'applique spontanément.

Commençons par analyser l'énoncé de l'enseignante. Il s'agit d'une structure canonique du français « standard » dans laquelle le sujet est généré dans le VP

(puisqu'il s'agit d'un thème et non d'un agent) puis dans laquelle cet élément est copié en spécifieur de TP afin de vérifier les traits [*u*Nominatif] et [*u*D*] portés par la tête T. Il y a donc un déplacement argumental. En revanche, pour un énoncé du type (b) ci-dessus, deux analyses sont possibles : (i) soit le clitique est analysé comme l'argument syntaxique, auquel cas le syntagme est considéré comme topique disloqué, (ii) soit le syntagme nominal est analysé comme l'argument du verbe, auquel cas le clitique est morphologique. Le cas de figure (bi) implique le même type de dérivation qu'en (a) avec un pronom à la place d'un syntagme nominal mais à laquelle il faut ajouter l'adjonction (ou le déplacement, selon la théorie choisie) de l'élément disloqué. *A priori*, il n'y a donc aucune raison pour que l'enfant, spontanément, choisisse ce type de solution dans la mesure où elle est computationnellement plus lourde que celle proposée en (a). Le cas de figure (bii) requiert le même type de dérivation syntaxique qu'en (a) en ce qui concerne le déplacement du sujet. Par contre, nous pouvons supposer que la présence de la préfixation verbale, par le biais du clitique *i(l)*, implique que les traits phi du verbe possèdent déjà leurs propres valeurs en termes de [Personne], de [Genre] et de [Nombre] lorsque le verbe arrive en T (ce qui n'est pas le cas dans le cadre de (a) puisque ces traits prennent une valeur par le biais de l'accord qui s'opère entre le verbe en T et le sujet). Ce type d'attribution des valeurs par le préfixe au verbe et non par le sujet expliquerait alors les occurrences du type *nous on dort* dans lesquelles c'est bien le clitique qui attribue ses valeurs au verbe et non le pronom tonique. Nous serions ainsi dans une configuration telle que celle préconisée par Pollock (1997) avec un « accord verbal fini [+ pronominal] »¹. Il reste alors à déterminer si cet accord verbal possède également le trait [+N], c'est-à-dire s'il possède la capacité à vérifier le trait casuel et le trait *EPP*.

Si nous postulons que le clitique est affixal, cette hypothèse implique que, dans les énoncés qui ne possèdent qu'un clitique préverbal comme *il dort*, il n'y a pas d'argument sujet phonétiquement réalisé. Il convient alors de décider quel élément se voit attribuer le thème-rôle du verbe. Nous avons vu au Chapitre II que plusieurs pistes sont envisageables : soit nous avons recours à *pro*, qui en tant qu'élément pronominal nul peut porter un thème-rôle, un Cas et le trait [D] nécessaire à la vérification de *EPP* (puisque les pronoms sont des têtes D) ; soit nous adoptons une hypothèse comme celle d'Alexiadou & Anagnostopoulou (1999) qui n'utilise pas la

¹ Cf. Chapitre II, Section III.6 pour le détail.

catégorie vide *pro* mais qui fait porter le trait [D] et le cas par la flexion verbale. Néanmoins, nous avons également mentionné le fait que cette hypothèse implique que, lorsqu'un syntagme nominal est également présent dans l'énoncé (comme dans *l'escargot il dort*), celui-ci doit obligatoirement occuper une position non-argumentale (du fait que, dans la proposition d'Alexiadou & Anagnostopoulou (1999), le spécifieur de TP n'est pas projeté). Or, cette hypothèse nous semble trop « forte » pour le français dans la mesure notamment où nous avons également vu au Chapitre II que le français ne fait pas usage de toutes les projections syntaxiques d'une langue comme l'italien (notamment la projection syntaxique FocusP dans la périphérie gauche¹) et que le français a donc besoin du spécifieur de TP pour véhiculer cette notion de Focus². Nous choisissons alors une voie intermédiaire que nous détaillons en (106) ci-dessous et dont nous présentons un exemple de dérivation en (107).

(106) Notre proposition en termes de vérification de traits :

- (a) Les valeurs des traits phi sont attribuées au verbe V grâce à un préfixe morphologique (*cf.* l'étape (a) de la dérivation (107)) ; l'attribution des valeurs aux traits phi de *v* peut alors se faire dès la fusion de *v* et de VP (*cf.* l'étape (b)), rendant ces traits interprétables plus tôt dans le cours de la dérivation et soulageant ainsi la computation de traits phi ininterprétables.
- (b) Lorsque la numération ne comprend pas d'élément phonétiquement spécifié, susceptible de porter le thème-rôle, le Cas nominatif et le trait [D] nécessaires à la convergence de la dérivation, ce thème-rôle et ces traits sont portés par un équivalent phonétiquement non-réalisé qui suit le même parcours qu'un sujet syntaxique « canonique ». Cet élément relève ainsi davantage d'une appellation du type « sujet silencieux » que du type *pro*, comme proposé par Kayne (2009)³.
- (c) Lorsque la numération comprend un élément phonétiquement spécifié, susceptible de porter le thème-rôle, le Cas nominatif et le trait [D], c'est cet élément qui suit le parcours du sujet syntaxique jusqu'en spécifieur de TP

¹ *Cf.* Chapitre II, Section III.5.

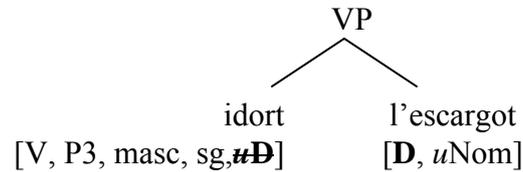
² D'après De Cat (2007).

³ L'auteur limite néanmoins l'apparition de ces éléments silencieux à une réflexion sur le clitique *on* et les éléments concomitants *nous* et *tous* (exemple : *nous on a tous ri*). Kayne (2009) postule ainsi que l'énoncé *on a tous ri* est possible parce qu'il comprend un *nous* silencieux en spécifieur de IP : "it may be that *nous* is actually in spec of IP" (« il est possible que *nous* se trouve en fait dans le spécifieur de IP »).

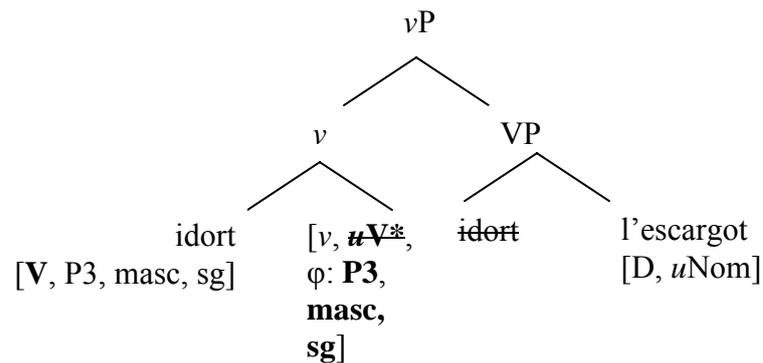
(la dérivation (107) ci-dessous propose un exemple avec un sujet qui porte le thème-rôle « thème » et non « agent »).

(107) Dérivation de la phrase « l'escargot i(l) dort » :

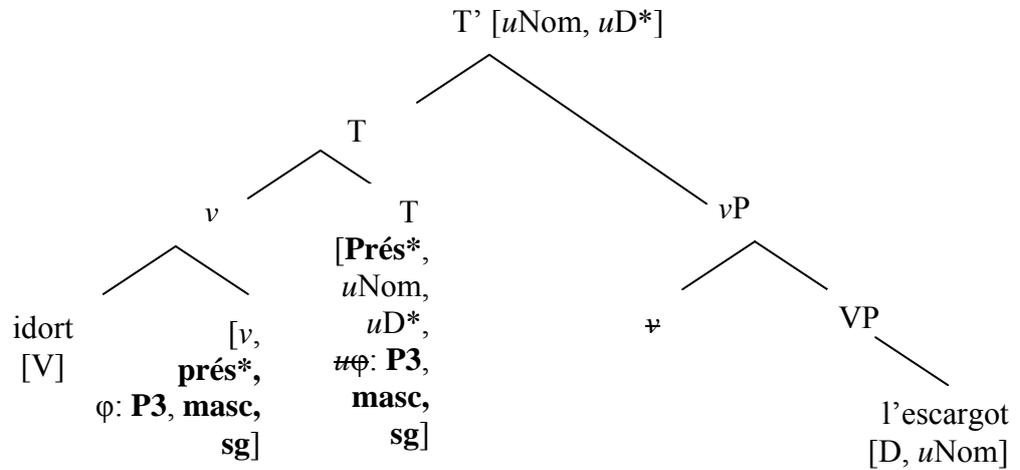
- (a) Fusion du verbe fléchi et de son thème, effacement du trait μD sur V, V porte des traits phi interprétables :



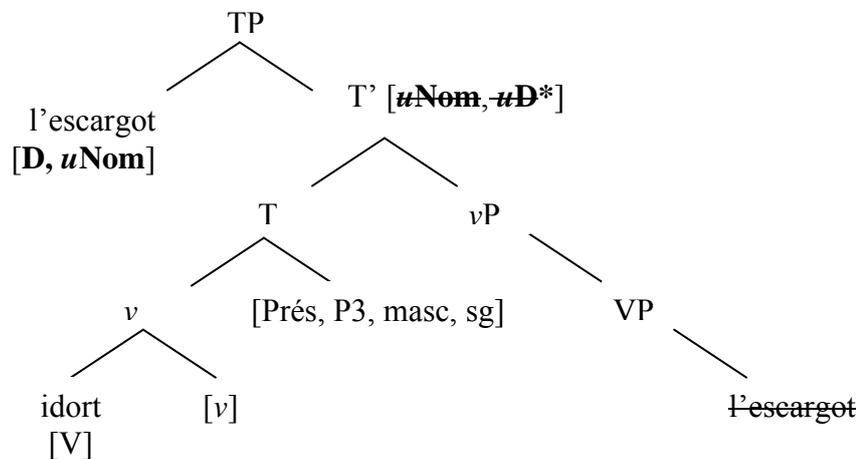
- (b) Fusion de VP et de v , adjonction de V à v ; V peut d'ores et déjà transférer les valeurs de ses traits phi à v qui cessent donc d'être ininterprétables dès ce stade de la dérivation ; v dans le cas d'un verbe inaccusatif ne projette pas de spécifieur et ne vérifie pas le Cas accusatif :



- (c) Fusion de vP et de T ; le trait T [présent*] est fort provoquant l'adjonction de v à T afin d'attribuer la valeur « présent » à v ; [$uNom$] sur T n'est pas encore vérifié du fait de l'absence d'un agent en vP ; les traits phi de T s'accordent directement avec ceux de v^1 :



- (d) Il reste un trait fort ininterprétable sur T : [uD^*] (trait *EPP*) qui se projette donc sur T' et force le déplacement du sujet en spécifieur de TP afin de l'effacer ; le trait casuel est également effacé dans cette configuration lorsque le trait [$uNom$] est porté par un thème :



Cette analyse des faits permet alors de supposer qu'Alan préfère *l'escargot i(l) dort* à *l'escargot dort* parce que son système linguistique contient le matériau morphologique préfixal qui lui permet d'attribuer des valeurs aux traits phi de v plus tôt dans le cours de la dérivation, ce qui soulage alors le système computationnel de cette tâche qui, autrement, est « en attente ». Cette hypothèse simplifie également la

¹ Mais nous nous interrogeons sur la nécessité de postuler des traits phi sur T dans le mesure où le Temps grammatical est distinct de l'accord et que le verbe est déjà marqué pour l'accord.

formalisation du processus d'accord qui, d'après la description d'Adger (2003), consiste en une transmission des traits phi du sujet en spécifieur de vP aux traits phi de T qui, ensuite, permet l'attribution de ces valeurs aux traits de v lorsque v s'adjoit à T.

Dans le cadre de notre proposition, il reste alors à rendre compte de la présence de ce matériau morphologique dans le système enfantin. Pour cela, il faut dans un premier temps terminer l'observation du système dans sa totalité et notamment confronter à nos nouvelles données l'hypothèse initiée à partir de Palasis (2005) quant à l'éventuel statut morphologique des clitiques nominatifs. Nous terminons ainsi notre tour d'horizon des sujets enfantins par l'analyse des sujets que l'enfant ne réalise pas phonétiquement : les sujets des verbes à l'impératif et les fameux « sujets nuls » enfantins.

VI. L'ABSENCE DE TOUT ÉLÉMENT RÉALISÉ

1. Les impératifs

D'après Platzack & Rosengren (1998), l'intérêt que portent les linguistes aux formes de l'impératif est relativement récent (depuis les années 1980 environ). Cela semble être lié à leur apparente simplicité. Or, une incursion dans la littérature publiée dans ce domaine au cours de ces vingt dernières années montre que ces paradigmes à l'allure dépouillée cachent en réalité des variations importantes d'une langue à une autre et un nombre non négligeable de phénomènes complexes tels qu'une certaine incompatibilité avec la négation¹. Voyons comment est analysé le français dans ce paysage extrêmement divers.

1.1 La théorie

Tout d'abord, quelle que soit la langue, il est généralement postulé que la forme verbale de l'impératif (ou de l'autre mode utilisé) se trouve, en fin de dérivation, dans la périphérie gauche de la phrase. Dans le modèle de Rizzi (1997a), il s'agit par exemple de la projection ForceP ; d'une façon plus générale et dans la littérature plus ancienne, le mouvement est décrit comme étant un déplacement de V vers C. En termes minimalistes, ce mouvement doit être motivé par une vérification d'un ou de

¹ Un nombre important de langues, dont le grec moderne par exemple, ne peuvent pas combiner le mode impératif et la négation. C'est alors souvent le mode subjonctif qui est utilisé.

plusieurs traits. Les propositions à ce sujet sont très variables. Notons simplement à titre d'exemple la proposition de Han (1998) dont la projection C contient deux traits : [directif] et [irréalis]. Cette position en C est par ailleurs motivée par l'ordre des constituants : lorsqu'une langue qui peut réaliser le sujet d'un impératif le fait, celui-ci est postverbal. Pour le français, la position postverbale des clitiques objets sert également à postuler cette place haute du verbe dans la hiérarchie. Leur position dans la périphérie gauche est donc une première caractéristique de ces formes verbales. La seconde concerne le sujet. La différence entre l'impératif et les autres types d'énoncés est particulièrement saillante en français dans la mesure où ce sont les seuls paradigmes du français standard qui valident l'absence d'un sujet phonétiquement réalisé. Han (1998) propose alors que ce sujet non réalisé soit PRO, c'est-à-dire dans le cadre *GB* une catégorie vide anaphorique, pronominale et non gouvernée, utilisée par ailleurs comme sujet dans les propositions infinitives ; cette proposition est illustrée en (108) ci-dessous.

(108) La structure d'une proposition impérative en français¹ :

[CP V_i[directif, irréalis] [IP PRO t_i]]

Ce rapprochement avec les infinitives est courant dans la mesure où, dans de nombreuses langues, le paradigme de l'impératif se réduit à une personne unique, celle de l'interlocuteur. Il n'est donc pas rare de concevoir la proposition impérative comme une structure dépouillée dans laquelle les projections de temps sont absentes². C'est ce que propose également Han (1998) pour le français :

“The imperatives in French have the syntactic structure of infinitivals. They have infinitive INFL, and so only PRO is licensed, whose reference is determined by the interpretational property of the imperative operator.”³

La question de la personne grammaticale est ainsi prise en charge par un opérateur impératif qui a « un trait interlocuteur implicite qui contrôle le sujet PRO, lui assignant la Personne 2 »⁴. Une proposition impérative en français présente ainsi un

¹ D'après Han (1998:140).

² Cf. Zanuttini (1997) ou Platzack & Rosengren (1998) par exemple.

³ Han (1998:141) : « Les impératifs en français ont la structure syntaxique des propositions infinitives. Ils ont un INFL infinitif, licenciant PRO dont la référence est déterminée par la propriété interprétationnelle de l'opérateur impératif ».

⁴ Han (1998:136).

nombre non négligeable de particularités : montée du verbe dans la périphérie gauche, absence de sujet réalisé concomitant avec une flexion verbale dont le paradigme est défectif et distinctif. Face à ce dernier faisceau de propriétés, nous nous interrogeons alors quant au rapprochement souvent effectuée entre l'impératif et l'infinitif et nous nous demandons s'il est bien tenable pour le cas du français. Voyons ce que nous disent les données enfantines.

1.2 Les données

La requête adéquate¹ montre que presque 14,81 % des verbes fléchis produits par les enfants dans un énoncé déclaratif ou exclamatif sont des verbes à l'impératif. Ce résultat est donc très proche du taux du corpus précédent qui est de 15,9 % et il établit également que cette tournure est utilisée relativement couramment par les jeunes enfants. La diversité lexicale de ces 1275 formes à l'impératif n'est pas très élevée puisque deux verbes représentent à eux seuls plus de la moitié de ces productions : *regarder* et *tenir*. Par ailleurs, la très grande majorité de ces impératifs se concentre sur une seule personne du paradigme qui en compte trois en français. En effet, les impératifs sont produits avec la Personne 2 dans 90 % des cas, le reste relevant de la Personne 5 (notamment *regardez*, *arrêtez*, *sortez*, *écoutez*, *attendez*). Ces caractéristiques sont illustrées par les exemples en (109) ci-dessous et le classement détaillé des verbes les plus fréquents est indiqué dans le Tableau 52.

(109) Les impératifs du corpus :

- (a) TOM: *regarde* y+a des renards ici . (I.2, 1251)
- (b) LEA: *tiens* ça . (II.9c, 315)
- (c) EMA: *attends* je vais chercher un aut(r)e . (I.3d, 1518)
- (d) LUS: *regardez* ma p(e)tite [/] p(e)tite cabane . (II.9b, 132)
- (e) EVA: *arrêtez arrêtez arrêtez arrêtez arrêtez* j' ai dit ! (XIII.46e, 1552)

¹ freq -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%mor +s"v*|*IMP&*"+u +o @ > Imperatives_Children_Freq.

Verbes	Personne 2	Personne 5	Totaux
<i>regarder</i>	465	80	545
<i>tenir</i>	201	1	202
<i>attendre</i>	156	7	163
<i>arrêtez</i>	39	14	53
<i>manger</i>	29	0	29
<i>donner</i>	20	0	20
<i>prendre</i>	20	0	20
<i>fermer</i>	16	1	17
<i>venir</i>	15	1	16
Totaux	961	104	1065

Tableau 52 : classement des verbes impératifs les plus fréquents

La structure de ces énoncés, illustrée en (110) ci-dessous, présente souvent le verbe à l'impératif utilisé de façon intransitive, soit parce qu'il est seul dans l'énoncé (*cf.* (a) ci-dessous), soit parce qu'il précède une proposition syntaxiquement indépendante (*cf.* (b) et (c)). Néanmoins, les enfants utilisent également les verbes à l'impératif avec des objets mais ceux-ci sont majoritairement non clitiques (*cf.* (d) à (g)). En effet, seuls quinze énoncés impératifs comprennent un enclitique (*cf.* (h) et (i)).

(110) La structure des énoncés impératifs :

- (a) TOM: regarde regarde ! (I.2, 488)
- (b) LEA: regarde *elle m' a pris ça maman* . (II.9c, 23)
- (c) NOE: oh regarde *c' est des lapins* ! (IV.18a, 563)
- (d) ANT: eh eh eh Lucie touche touche touche *ça* ! (III.13b, 412)
- (e) ROM: prends *c(el)ui+là* . (IV.19e, 931)
- (f) ZOE: (é)coute *moi* Noémie . (II.8, 53)
- (g) EVA: frappe *lui* . (IV.19e, 1481)
- (h) NOE: donne *le* à Lisa [/] euh elle . (IV.18b, 583)
- (i) ALA: rends *la moi* ! (III.13e, 1931)

Ainsi, la présence de ces pronoms postverbaux indique que les enfants qui les utilisent effectuent bien le déplacement du verbe vers la périphérie gauche de la structure. Par contre, en ce qui concerne l'analyse de la forme verbale, il nous semble difficile de postuler que l'enfant puisse établir un parallèle entre une structure infinitive et une structure impérative pour deux raisons : (i) ces deux formes verbales divergent en français (*regarde vs regarder* par exemple) et (ii) la forme de l'impératif P2 est phonétiquement identique à celle du présent de l'indicatif qui est

un mode fléchi. En toute logique, si l'enfant doit établir un parallèle entre l'impératif et un autre temps ou mode qu'il connaît, il nous semble donc que le présent de l'indicatif vient plus rapidement à l'esprit que l'infinitif, d'autant plus que le mécanisme de cliticisation des objets est également différent entre le mode impératif (enclise : *fais-le*) et le mode infinitif (proclise : *tu dois le faire*). Ainsi, si nous suivons Han (1998), il faut postuler que les deux formes verbales phonétiquement identiques de l'impératif et de l'indicatif présent sont analysées différemment par les enfants sur la base de la présence *vs* l'absence du clitique nominatif (*cf.* (111a) ci-dessous *vs* (111b)) :

(111) Comparaison entre le présent de l'indicatif et l'impératif :

- (a) Tu chantes : [CP [IP *tu chantes*]]
- (b) Chante ! : [CP *chante*_{i[directif, irréalis]} [IP PRO t_i]]

Or, cette approche pose un problème dans la mesure où les jeunes enfants entendent des énoncés non impératifs sans clitique¹ et qu'ils produisent par ailleurs des structures sans sujet réalisé qui sont considérées comme agrammaticales. Il s'agit des fameux « sujets nuls » enfantins, dont quelques exemples sont d'ailleurs co-occurrents de l'impératif, comme illustré en (112) ci-dessous :

(112) Les « sujets nuls » enfantins :

- (a) ZOE: *regarde moi aussi* 0 [*] *sais le faire* . (I.3a, 57)
- (b) EMA: *attends attends attends* 0 [*] (*v*)a [?] *te prêter Winnie* . (III.10a, 1125)

Nous suggérons donc qu'un rapprochement de l'impératif avec l'indicatif, plutôt qu'avec l'infinitif, semble plus adéquat dans le cadre du système enfantin sur la base des constatations suivantes : (i) les paradigmes des présents de l'indicatif et de l'impératif comportent des formes homophones en français, comme illustré en (111) ci-dessus ; (ii) deux différences existent néanmoins entre ces deux modes : le nombre de personnes dans les paradigmes et la réalisation du clitique nominatif. Or, une corrélation peut être proposée entre ces deux différences : le paradigme de l'indicatif comporte six personnes non distinctives alors que celui de l'impératif n'en comporte que trois (voire deux, dans le cas des enfants) et ces formes verbales sont distinctives (*regarde vs regardez* par exemple). Le paradigme non-distinctif requiert la présence

¹ Par exemple : *faut pas faire ça*.

de clitiques nominatifs alors que le paradigme distinctif ne la requiert pas. La question suivante peut alors être posée : si le système de l'enfant comporte d'autres paradigmes défectifs que ces paradigmes de l'impératif, comment son système les gère-t-il en termes de présence/absence du clitique nominatif ?

2. Les « sujets nuls »

Hormis les cas cités ci-dessus, l'absence totale d'un constituant sujet réalisé est généralement considérée comme une agrammaticalité en français. Nous avons vu que ce type d'absences marque néanmoins une étape incontournable dans le développement linguistique infantin puisqu'il est largement admis que tous les enfants, quelle que soit leur langue première (à « sujet non-nul »), traversent une période pendant laquelle ils produisent des énoncés sans aucun élément sujet apparent.

2.1 Définition et nouvelle appellation : le phénomène du clitique nul

En dehors des cas particuliers déjà cités, un énoncé est considéré comme grammatical lorsqu'il comporte soit : (i) un syntagme nominal préverbal seul, (ii) un pronom tonique seul (uniquement de Personne 3 et dans certaines conditions), (iii) un clitique nominatif seul, (iv) la combinaison d'un clitique avec un syntagme nominal ou un pronom tonique. Ce qui est donc considéré comme agrammatical est : l'absence de tout élément (*cf.* (113a) ci-dessous), la présence d'un syntagme nominal postverbal seul (*cf.* (113b)) et la présence d'un pronom tonique seul lorsqu'il n'est pas de Personne 3 (*cf.* (113c)).

(113) Les cas d'agrammaticalités :

- (a) MAX: euh 0 [*] sais pas . (X.40f, 154)
- (b) TEO: oh 0 [*] est pas là la grenouille . (III.10a, 650)
- (c) EMA: moi 0 [*] veux partir +... (X.38c, 350)

Dans Palasis (2005), nous discutons la question de « l'inversion libre » du sujet en rapport avec le paramètre *pro-drop* et celle des sujets « postverbaux » enfantins¹. En effet, dans les deux cas, nous estimons que l'agrammaticalité de ces deux structures peut être attribuée non pas à la position postverbale du syntagme nominal sujet mais à l'omission d'un clitique nominatif préverbal :

¹ *Cf.* Palasis (2005:45-46 et 102-103) et les exemples en (114) ci-après.

(114) Deux analyses divergentes :

- (a) « L'inversion libre » du sujet¹ :
 - i. Italien : *_ è stato dato un premio al presidente.*
 - ii. Français : **_ a été donné un prix au président.*
 - iii. Cible généralement proposée : *Un prix a été donné au président.*
 - iv. Notre proposition : *Il a été donné un prix au président.*
- (b) Le « sujet postverbal » enfantin² :
 - i. Français enfantin : **est mouillée voiture.* (Grégoire, 1;11)
 - ii. Cible généralement proposée : *La voiture est mouillée.*
 - iii. Notre proposition : *Elle est mouillée la voiture.*

Nous privilégions cette analyse en termes d'absence d'un clitique nominatif dans la mesure où les données enfantines montrent que la structure Clitique[^]V est prégnante dans le système enfantin alors que la structure DP[^]V est quasi-absente³. Ainsi, lorsque nous examinons les trois types d'« agrammaticalités » présentés en (113) ci-dessus, nous maintenons qu'ils peuvent être subsumés sous un seul critère : l'absence d'un clitique nominatif.

2.2 L'extraction des énoncés

Afin d'extraire et d'analyser le plus finement possible l'omission de ces clitiques, nous avons codé l'absence de ces éléments de manière détaillée sur la ligne dépendante spécifique %err ; le codage générique est \$SYN \$LOS \$PRO \$SUBJ. Ainsi, la requête qui permet l'extraction de ce type d'énoncés dans CLAN est directe, comme le montre son détail repris en (115) ci-dessous. Cette requête, ciblée sur les vingt enfants du corpus, livre un total de 348 énoncés produits par dix-neuf enfants sur vingt⁴.

¹ Cf. Rizzi (1997b:271).

² Cf. Friedemann (1993:224).

³ Cf. Section V de ce chapitre.

⁴ L'enfant qui ne produit pas ce type d'énoncés est aussi celui pour lequel nous avons le moins de données. Il s'agit de Clara qui produit un total de 43 énoncés dans ce corpus.

(115) Détail de la requête pour l'extraction des sujets « agrammaticaux » :

```
combo -t*ELO -t*LOL -t*UNI -t*KAT -t*SOF -t*TAA +t%err
+s"$SYN^$LOS^$PRO^$SUBJ*" +u @> NoCli_Children
```

Ayant été conçus au fur et à mesure de la transcription des données, les codes %err constituent une liste exhaustive des agrammaticalités présentes dans cette base de données, c'est-à-dire que chaque code renvoie à au moins un item effectivement produit et qu'il ne s'agit donc pas d'une liste préétablie d'agrammaticalités potentielles. Dans un premier temps, il est donc intéressant de consulter cette liste afin de connaître le détail des possibilités que recouvre l'omission du clitique nominatif dans le corpus. En effet, le codage \$SYN \$LOS \$PRO \$SUBJ est notre code par défaut : il n'est utilisé que lorsque la personne grammaticale n'est pas clairement identifiée dans l'énoncé. Notre corpus comprend deux énoncés de ce type repris en (116) ci-dessous¹ :

(116) Codage par défaut :

(a) TOM: 0 [*] veux [=? faut] remett(r)e ça . (VI.26b, 198)

```
%err: 0 = ? $SYN $LOS $PRO $SUBJ
```

(b) MAX: i(l) va dire 0 [*] veut [=? veux] monter de [*] l' éléphant .
(VIII.31b, 1260)

```
%err: 0 = il/je $SYN $LOS $PRO $SUBJ ; de = sur $LEX $SUB
$PREP
```

Ailleurs, dès lors que la personne grammaticale est identifiée clairement (notamment grâce au contexte), ce codage par défaut est complété par une étiquette supplémentaire qui ajoute cette information. Un total de sept codes, dont le détail figure dans le Tableau 53 ci-dessous, est ainsi créé pour cette base, chaque code renvoyant à un type de clitique différent. Chacune de ces possibilités est ensuite illustrée en (117).

¹ La requête utilisée pour extraire ces énoncés de façon isolée est : `combo -t*ELO -t*LOL -t*UNI +t%err +s"$SYN^$LOS^$PRO^$SUBJ^*^!($1S+$2S+$3S+$IMPRS+$3P+$DEM)" +u @>NoCli_All_NonID.`

Codages %err	Personnes	Éléments manquants
\$\$SYN \$LOS \$PRO \$\$SUBJ	non identifiée	?
\$\$SYN \$LOS \$PRO \$\$SUBJ \$1S	1	je
\$\$SYN \$LOS \$PRO \$\$SUBJ \$2S	2	tu
\$\$SYN \$LOS \$PRO \$\$SUBJ \$3S	3 référentiel	il, elle
\$\$SYN \$LOS \$PRO \$\$SUBJ \$IMPRS	3 explétif	il
\$\$SYN \$LOS \$PRO \$\$SUBJ \$DEM	3 démonstratif	ce
\$\$SYN \$LOS \$PRO \$\$SUBJ \$3P	6	ils, elles

Tableau 53 : liste exhaustive des codes en rapport avec l'omission du clitique nominatif

(117) Codages spécifiques :

- (a) MAX: euh 0 [*] sais pas . (III.13d, 270)
%err: 0 = je \$\$SYN \$LOS \$PRO \$\$SUBJ \$1S
- (b) ALA: 0 [*] as vu j' ai rangé . (I.3b, 304)
%err: 0 = tu \$\$SYN \$LOS \$PRO \$\$SUBJ \$2S
- (c) TOM: 0 [*] veut pas manger . (I.2, 135) (0 = *il*)
%err: 0 = il \$\$SYN \$LOS \$PRO \$\$SUBJ \$3S
- (d) CEL: 0 [*] faut pas la casser . (II.9b, 913)
%err: 0 = il \$\$SYN \$LOS \$PRO \$\$SUBJ \$IMPRS
- (e) EMA: 0 [*] est des feuilles . (I.3c, 668)
%err: 0 = c' \$\$SYN \$LOS \$PRO \$\$SUBJ \$DEM
- (f) EKT: 0 [*] [?] mangent [?] des herbes ! (II.9a, 582)
%err: 0 = ils \$\$SYN \$LOS \$PRO \$\$SUBJ \$3P

Le nombre total de clitics nuls est relativement bas puisque, par rapport à l'ensemble des 11027 verbes fléchis du corpus, le taux global moyen du corpus est de 3,16 %¹. Cela s'explique notamment par la tranche d'âge des enfants enregistrés. En effet, alors que dans Palasis (2005) les enfants les plus âgés ont 3;1, dans le cas présent huit enfants ont déjà atteint leurs trois ans au moment de la première séance et la tranche d'âge lors de la dernière session se situe entre 3;0.13 et 4;0.1. Ce phénomène du clitique nul s'estompe donc bien progressivement aux alentours de l'âge de trois ans. Les taux globaux des treize séances illustrent d'ailleurs cette décroissance :

¹ Ce qui est donc très éloigné du taux de Rasetti (1996) avec 26,2 % ou Palasis (2005) avec 22,4 %.

	1	2	3	4	5	6	7	8	9	10	11	12	13	Tot.
V	438	704	874	785	838	807	825	870	1022	1084	993	504	1037	10781 ¹
-Cl	23	25	54	35	34	26	31	29	24	22	29	5	11	348
%	5,25	3,55	6,18	4,46	4,06	3,22	3,76	3,33	2,35	2,03	2,92	0,99	1,06	3,23

Tableau 54 : évolution de la proportion des clitiques nuls par rapport aux verbes fléchis

2.3 Quelles personnes verbales ?

Une première série d'observations se rapporte à l'éventail des occurrences. La ventilation présentée dans le Tableau 55 ci-après illustre tout d'abord que, à une exception près, les omissions enfantines interviennent dans le cadre du système du [Singulier]. Nous avons alors testé la piste d'une éventuelle corrélation entre l'émergence de la distinction de [Nombre] chez les enfants² et la disparition de l'omission des clitiques et nous n'en avons trouvé aucune (les données de Lucie par exemple montrent une émergence de [\pm Nombre] à 2;10.5 et une production de clitiques nuls jusqu'à 3;4.16).

Types	Tous	1S	2S	3S	Imprs	Dem	Non Id	3P
Totaux	348	219	11	35	71	9	2	1
%	100,00	62,93	3,16	10,06	20,40	2,59	0,57	0,29
%	100,00	62,93	3,16	33,05			0,57	0,29
%	100,00	99,71						0,29

Tableau 55 : répartition des omissions

La seconde caractéristique de ces énoncés est de référer majoritairement au locuteur (Personne 1 : 62,93 % des omissions). Une interprétation qui vient alors à l'esprit est de supposer que l'omission du clitique est liée à la période égocentrique dans laquelle se trouve le jeune enfant. Mais cette analyse ne rend pas compte de l'ensemble de ses omissions dans la mesure où une partie non négligeable d'entre elles relève de la Personne 3 (33,05 %). L'un des objectifs de cette recherche est donc de vérifier l'hypothèse proposée à partir de Palasis (2005) sur le lien que nous établissons entre la richesse morphologique du système verbal enfantin à ce stade et l'omission du clitique nominatif³. Nous procédons ainsi selon le même schéma que précédemment et, dans un premier temps, nous détaillons les verbes impliqués dans ces énoncés à clitique nul.

¹ Les formes du type *y+a* et *ça+y+est* ne sont pas incluses.

² Cf. Section II.3 de ce chapitre sur l'émergence des traits.

³ Les détails de cette analyse se trouvent Chapitre II, Section I.3.4.

2.4 Quels verbes ?

Les recherches sont effectuées par enfant, ce qui nous permet de filtrer les données. Ainsi, le Tableau 56 détaille la liste des verbes présents dans les énoncés à clitique nul. Un filtre est appliqué à la liste des 348 verbes initiaux : ne sont conservés à des fins de représentativité que les verbes produits par plus de trois enfants différents au fil des treize séances. Ce critère permet ainsi d'éviter d'exploiter un nombre d'occurrences total pour un verbe qui paraît significatif sur l'ensemble des séances mais qui, en fin de compte, fausserait l'analyse¹.

	QU	LA	LE	EV	EK	TE	IS	TO	JU	CE	LU	NO	AN	RO	AL	EM	MA	ZO	T
<i>veux</i>	9			1	1	17	1	42	3	1	3	2	2	23	20	6	10	2	143
<i>faut</i>		2	1	1	1		1	7	1	2	8	2	10		5	15		1	57
<i>peux</i>					1	1		12			1	2			3	1	5		26
<i>est</i>						7	1	1			2			2	1	5	1	1	21
<i>a</i>					1	2		1				2	1		6	1		2	16
<i>sais</i>											4		1				2	1	8
T	9	2	1	2	4	27	3	63	4	3	18	8	14	25	35	28	18	7	271

Tableau 56 : les verbes dans les énoncés à clitique nul

Le verbe de modalité *vouloir* à la Personne 1 arrive ainsi nettement en tête devant le verbe défectif *falloir* et un autre verbe de modalité, également à la Personne 1; puis viennent les auxiliaires à la Personne 3 et le verbe *savoir* à la Personne 1. Ce classement est proche de celui obtenu dans Palasis (2005) en termes de verbes (*être*, *vouloir*, *pouvoir*, *falloir*), la différence essentielle étant que le verbe *être* n'est plus le premier du classement, ce qui peut être interprété comme un signe de diversification lexicale. En termes de morphologie verbale, si l'on considère les paradigmes des personnes du singulier de ces verbes, celle-ci est distinctive pour *falloir* puisqu'il s'agit d'un verbe défectif et pour les deux auxiliaires (soit pour 34,7 % des 271 verbes ci-dessus). En ce qui concerne les autres verbes (*vouloir*, *pouvoir* et *savoir*), la morphologie verbale n'est pas distinctive pour ces trois personnes. Cependant, le Tableau 56 ne présente qu'une vue partielle du système verbal des enfants dans la mesure où seuls les verbes utilisés avec un clitique nul y apparaissent. Complétons donc cette première liste par l'ensemble des verbes utilisés par les enfants afin notamment de reconstituer les paradigmes effectivement manipulés dans la totalité de ce système.

¹ Cf. Chapitre III, Section I.4.2.1 sur la question de la variation individuelle. La liste complète des verbes se trouve en Annexes, page 384.

2.5 Les verbes dans l'ensemble du corpus

L'ensemble des verbes du corpus utilisé par les dix-neuf enfants qui produisent des clitiques nuls est examiné après l'application du même filtre que précédemment. Un total de 8902 verbes fléchis est ainsi extrait ; les verbes les plus courants et notamment ceux qui sont utilisés par l'ensemble de ces dix-neuf enfants se trouvent dans le Tableau 57 ci-après¹.

	Totaux	Enfants
v:*\ être&pres&3sv	3110	19
v:*\ avoir&pres&1sv	951	19
v:*\ avoir&pres&3sv	803	19
v:*\ vouloir&pres&1sv	590	19
v:*\ aller&pres&3sv	250	19
v:*\ faire&pres&3sv	215	16
v:*\ falloir&pres&3sv	163	16
v:*\ avoir&pres&2sv	151	19
v:*\ vouloir&pres&3sv	123	19
v:*\ être&impf&3sv	109	15
v \ voir&pres&1sv	102	13
v \ appeler&pres&3sv	100	17
v \ mettre&pres&3sv	98	14
v:*\ être&pres&1sv	79	19

Tableau 57 : liste des verbes les plus courants dans le corpus

Globalement, cette liste correspond déjà au classement des verbes les plus fréquents du français adulte². Confrontons ces totaux avec ceux obtenus dans la section précédente. Le Tableau 58 présente ainsi les taux de clitiques nuls par verbe et par rapport à l'ensemble des formes pour ce verbe et cette personne ; ils sont classés par ordre décroissant de fréquence.

¹ Les astérisques dans le tableau indiquent que les distinctions entre verbes lexicaux et verbes auxiliaires ou modaux n'est pas conservée dans ce cadre (par exemple, les 3110 formes de *être* regroupent à la fois le verbe utilisé en tant que copule et le verbe utilisé dans les temps composés).

² Cf. par exemple les classements donnés dans Muller (1979).

Formes verbales	Personnes	Totaux	Clitiques nuls	% clitiques nuls
<i>peux</i>	1	72	26	36,11
<i>faut</i>	3	163	57	34,97
<i>veux</i>	1	590	143	24,24
<i>sais</i>	1	76	8	10,53
<i>a</i>	3	803	16	1,99
<i>est</i>	3	3110	21	0,68
Totaux		4814	271	5,63

Tableau 58 : pourcentages de clitiques nuls par forme verbale (première partie)

Ces pourcentages sont les plus élevés du corpus et il est flagrant qu'ils se situent bien au-dessus de la moyenne de 3,23 % établie dans le Tableau 54. Le clitique de Personne 1 est donc très fortement omis avec *peux*, *veux* et *sais*. L'explication à ce phénomène apparaît lorsque l'on considère les paradigmes de ces verbes, tels qu'ils sont manipulés par les jeunes enfants.

2.6 Notre hypothèse sur le phénomène du clitique nul

Complétons tout d'abord nos informations avec les personnes qui n'apparaissent pas dans le Tableau 58 ci-dessus :

Formes verbales	Personnes	Totaux	Clitiques nuls	% clitiques nuls
<i>peux</i>	2	57	6	10,53
<i>peut</i>	3	48	4	8,33
<i>veux</i>	2	67	0	0
<i>veut</i>	3	123	8	6,50
<i>sais</i>	2	59	0	0
<i>sait</i>	3	8	0	0

Tableau 59 : pourcentages de clitiques nuls par forme verbale (seconde partie)

Le Tableau 60 ci-dessous regroupe les informations contenues dans les deux tableaux précédents ; les paradigmes du singulier sont alors complets pour les trois verbes¹. Le paradigme de *savoir* est particulièrement révélateur :

¹ Les parenthèses indiquent que l'élément n'est pas toujours réalisé.

pouvoir	% d'omission du clitique	vouloir	% d'omission du clitique	savoir	% d'omission du clitique
<i>(je) peux</i>	36,11	<i>(je) veux</i>	24,24	<i>(je) sais</i>	10,53
<i>(tu) peux</i>	10,53	<i>tu veux</i>	0	<i>tu sais</i>	0
<i>(i) peut</i>	8,33	<i>(i) veut</i>	6,50	<i>i sait</i>	0

Tableau 60 : les paradigmes enfantins de *pouvoir*, *vouloir* et *savoir* (synthèse)

L'idée qui se dégage de l'observation de ces systèmes est que, si un paradigme verbal comprend trois formes homophones (qui ne permettent donc pas de distinguer la personne grammaticale), il faut ajouter une marque distinctive sur deux de ces trois formes afin de rendre le paradigme distinctif. Bien sûr, une autre stratégie pourrait consister à marquer les trois formes concomitamment mais cette procédure serait moins économique. Dans ce cadre, lorsqu'il est décidé de marquer deux formes sur trois, il reste à choisir quelles formes marquer (ou quelle forme ne pas marquer). Toujours dans un souci d'économie, c'est la forme la plus couramment utilisée du paradigme qui n'est pas marquée, c'est-à-dire, pour les trois verbes ci-dessus, la Personne 1. Cette stratégie est particulièrement bien illustrée par le verbe *savoir* et la distribution complémentaire parfaite entre l'omission possible du clitique de Personne 1 et la réalisation systématique des clitics de Personnes 2 et 3. Les faits, dans leur ensemble, sont un peu moins tranchés pour le verbe *vouloir* et nettement moins tranchés pour *pouvoir*. Cependant, souvenons-vous que nous avons affaire à des moyennes.

Ainsi, l'étude de cette stratégie au niveau individuel permet de maintenir l'hypothèse proposée. Pour le verbe *pouvoir*, les clitics nuls de P2 proviennent tous de Maxime et ceux de P3 sont tous dus à Lucie ; les clitics nuls de P1 proviennent de huit enfants différents mais c'est Tom qui recourt à cette stratégie le plus souvent (12 clitics nuls sur un total de 26). Pour ces trois enfants, les paradigmes de *pouvoir* sont donc les suivants :

Personnes	TOM		MAX		LUS	
	Totaux	Clitics nuls	Totaux	Clitics nuls	Totaux	Clitics nuls
1	20	12	8	5	5	1
2	3	0	9	6	7	0
3	5	0	2	0	11	4
Totaux	28	12	19	11	23	5

Tableau 61 : les stratégies individuelles pour le verbe *pouvoir*

Tom applique à la lettre la stratégie de l'omission du clitique de P1 dans la mesure où 100 % de ses omissions relèvent de cette personne. Lucie quant à elle choisit également le même type de procédure puisque 4 omissions sur 5 relèvent d'une seule personne mais elle applique cette stratégie non pas à P1 mais à P3. Cette option reste néanmoins compatible avec la stratégie proposée dans la mesure où P3 est la personne la plus employée par Lucie dans ce paradigme. Seules les données de Maxime n'illustrent pas complètement ces préceptes. D'une part, la personne qu'il omet le plus correspond bien à la forme qu'il produit le plus (P2) et le clitique n'est pas omis sur la forme qu'il produit le moins (P3). Ces données illustrent donc la même stratégie que Tom et Lucie. Cependant, le problème avec Maxime est qu'il applique la même stratégie à deux personnes (non seulement P2 mais aussi P1), ce qui, en fin de compte, annule les effets distinctifs souhaités, d'autant plus que les omissions sur ces deux personnes surviennent au cours de la même séance. Après vérification, il ne nous semble pas y avoir d'erreur d'interprétation quant à la personne concernée lors de ces omissions (*cf.* l'énoncé (118) ci-dessous où Maxime nous demande bien la boîte). Cependant, cette transcription attire notre attention dans la mesure où, contextuellement, il s'agit bien d'un *tu* mais l'enfant, lors de la reprise (indiquée par le signe [/]), utilise un *je*. Il se pourrait alors que ce que nous pensons être une omission de P2 relève en fait d'une omission de P1 et que Maxime utilise le verbe *donner* en inversant les rôles thématiques, comme le montre le deuxième code %err en (118) ci-dessous.

(118) *MAX: 0 [*] peux [/] je [*] peux 0 [*] donner la boîte ? (IV.19b, 1319)
 %err: 0 = tu \$SYN \$LOS \$PRO . je = tu \$MOR \$SUB \$PRO \$SUBJ

Ainsi, malgré un taux de clitiques nuls moins élevé que dans Palasis (2005), l'hypothèse d'un lien étroit entre la morphologie verbale suffixale et l'adjonction ou non d'une information préverbale supplémentaire peut être maintenue. Antoine nous offre une illustration exemplaire de ce phénomène. En effet, la très grande majorité des verbes fléchis enfantins (84,93 %) est au présent de l'indicatif (qu'il s'agisse d'un verbe lexical au présent ou d'un auxiliaire au présent accompagné d'un participe, construisant ainsi un passé composé). En effet, les enfants dans un premier temps, s'expriment sur le « ici et maintenant » utilisant ainsi également l'impératif (11,96 %). Puis, dans un second temps, ils relatent le passé et l'imparfait entre alors dans leur système aux côtés du passé composé. À ce stade, l'imparfait reste

néanmoins peu employé (2,84 % des verbes fléchis). Quelques rares occurrences du futur, du conditionnel présent et du subjonctif présent viennent compléter ce tableau (0,27 %). Antoine introduit ainsi le verbe *vouloir* à l'imparfait à partir de la troisième séance (à 2;10.18) et produit un total de 35 occurrences de cette forme sur l'ensemble du corpus. Il utilise un total de quatre personnes grammaticales différentes (P1, P2, P3 et P6) mais ces quatre personnes présentent la particularité de constituer un paradigme de formes totalement homophones ([vule]). Le Tableau 62 ci-dessous nous livre la stratégie d'Antoine :

Personnes	Totaux	Clitiques nuls
1	30	11
2	1	0
3	3	0
6	1	0
Totaux	35	11

Tableau 62 : Antoine et le verbe *vouloir* à l'imparfait

Ainsi, au sein d'un paradigme verbal qui comprend quatre formes homophones, Antoine utilise toujours un clitique pour les trois personnes les moins fréquentes et n'omet le clitique qu'avec la forme qu'il utilise le plus et de façon massive puisque ces clitiques nuls représentent 36,67 % des formes de P1. La stratégie du clitique nul reste donc productive chez Antoine dans la mesure où il l'applique à un paradigme récemment introduit dans son système linguistique alors qu'il a entre 2;10.18 et 3;4.21. Cependant, une telle application n'est pas surprenante au regard des principes d'économie récurrents dans le langage. Les clitiques nuls sont beaucoup plus rares chez les adultes mais le fait que les seules omissions adultes dans ce corpus portent sur le verbe défectif *falloir* est une autre illustration de ce principe :

(119) KAT: 0 [*] *faut* fermer les yeux pour pas tricher . (V.23a, 947)

L'hypothèse du lien fort établi dans Palasis (2005) entre la richesse morphologique d'un paradigme verbal et l'omission des clitiques préverbaux est ainsi confirmée malgré la baisse substantielle du taux d'omission entre les deux corpus. Il reste alors à expliquer pourquoi les clitiques nuls disparaissent quasi-entièrement du fonctionnement linguistique des locuteurs francophones.

2.7 La disparition progressive des clitiques nuls

Rappelons la configuration du système verbal du corpus exploité dans Palasis (2005) : (i) une opposition P1 vs P3 dans 95,9 % des énoncés et (ii) des paradigmes verbaux distinctifs dans 72,9 % des cas. Ces faits sont visibles dans le Tableau 63 ci-dessous.

		<i>être</i>	<i>vouloir</i>	<i>faire</i>	<i>avoir</i>	<i>savoir</i>	<i>voir</i>	1 ^{er} groupe
Pers.	%	48,7	17,2	16,2	7,0	3,6	2,2	5,1
1sg	35,0	sɥi	vø	fe	e	se	vwa	mãʒ
3sg	60,9	e	-	fe	a	se	vwa	mãʒ

Tableau 63 : les formes verbales enfantines (corpus 1)

Comparons cette première configuration à celle du corpus actuel dont les caractéristiques sont les suivantes¹ : (i) introduction de P2, (ii) des paradigmes verbaux moins distinctifs du fait de l'introduction de P2 et d'une plus grande diversité lexicale (plus de verbes différents au sein du troisième groupe et une augmentation du pourcentage des verbes du premier groupe). Cette deuxième série de faits est récapitulée dans le Tableau 64 ci-après.

		<i>être</i>	<i>avoir</i>	3 ^{ème} groupe	1 ^{er} groupe
Pers.	%	38,6	22,7	27,6	11,1
1sg	27,3	sɥi	e	vø	mãʒ
2sg	5,7	e	a	vø	mãʒ
3sg	66,1	e	a	vø	mãʒ

Tableau 64 : les formes verbales enfantines (corpus 2)

Dans Palasis (2005), les faits résumés dans le Tableau 63 conduisent à expliquer le phénomène du clitique nul par une ressemblance de ce système verbal avec le système morphologiquement riche des langues *pro-drop* : les enfants omettent le clitique dans 22,4 % des énoncés parce que le clitique n'apporte alors aucune information supplémentaire concernant la personne grammaticale. Ils insèrent néanmoins le clitique dans la majorité des cas car ils savent en amont que le français n'est pas une langue *pro-drop*². Un phénomène en particulier nous permet de formuler une telle affirmation. Dans la Section II.3.4 de ce chapitre consacrée à la référence des clitiques, nous avons mentionné la corrélation forte envisagée par

¹ Ces taux sont calculés sur la base des verbes produits au présent de l'indicatif.

² Cette hypothèse fait notamment référence aux travaux de Wexler (1998) sur la fixation précoce et correcte des paramètres et aux données que l'enfant entend au quotidien qui contiennent majoritairement des clitiques.

certaines entre l'absence de sujets explétifs dans une langue et son classement en tant que système *pro-drop*¹. Ce corpus enfantin montre que les enfants omettent bien un certain nombre de clitiques explétifs (puisque le taux d'omission de ce type de clitique s'élève à 28,51 %²). Néanmoins, ce taux est loin d'être dominant et un examen attentif de ces omissions indique que, même chez les jeunes enfants, des formes telles que **pleut* ne sont pas attestées. En effet, leurs omissions ne concernent que quatre formes verbales différentes, comme illustré en (120) ci-dessous et 85,92 % d'entre elles impliquent le verbe *falloir* (cf. l'exemple (a)). Ainsi, si nous reprenons la distinction établie entre les différents types de *pro* par Rizzi (1986b)³, il est visible que les éléments nuls dans les énoncés enfantins relèvent uniquement de la catégorie des *pro* « purement explétifs » (du type *il y a*) et qu'il n'y a donc pas de *pro* « quasi-argumentaux » (du type *il pleut*). Le mécanisme qui s'applique dans le système enfantin en français n'est donc pas identique en tous points à celui d'une langue *pro-drop* comme l'italien.

(120) Les types d'explétifs omis par les enfants (71 occurrences) :

- (a) LUS: 0 [*] faut ranger ! (IV.18b, 186) (61 occurrences ; 85,92 %)
- (b) ALA: 0 [*] a plus de place ! (III.13h, 371) (8 occurrences ; 11,26 %)
- (c) EKT: oui non mais que s(e) passe t 0 [*] ah ! (IV.19d, 43) (1 ; 1,41 %)
- (d) NOE: oh 0 [*] me manque ça et ça et ça . (V.23a, 1437) (1 ; 1,41 %)

Dans le cadre de notre hypothèse, l'information que véhicule le clitique préverbal, lorsque le jeune enfant l'insère dans son énoncé, est donc purement grammaticale dans la mesure où elle permet d'identifier le trait [Personne]. Les jeunes enfants considèrent donc cet élément préverbal de la même façon qu'ils considèrent les suffixes verbaux ; ce sont des outils grammaticaux. Rappelons que les enfants de cet âge ne sont pas encore en contact ni avec un discours métalinguistique quelconque ni avec la version écrite de la langue. Par rapport au continuum phonique qu'ils entendent, ils n'ont donc aucune raison d'établir une distinction entre l'information postverbale (dont le statut morphologique ne fait aucun doute) et l'information préverbale (dont le statut fait débat chez les linguistes).

¹ Cf. Haider (2001:285) cité par Kaiser (2009:150).

² Sur 249 formes verbales impersonnelles, les enfants omettent 71 fois le clitique explétif.

³ Cf. Chapitre II, Section II.2.1.2.

Dans Palasis (2009), nous suggérons alors que ces clitiques préverbaux démarrent leur carrière dans le système linguistique des jeunes enfants sous forme de préfixes verbaux. En tant que tels, ils entrent dans la dérivation syntaxique de la même façon que les suffixes et ils ne représentent donc pas des éléments que la syntaxe peut manipuler indépendamment du verbe (au même titre que les suffixes)¹. Ces clitiques ne peuvent donc pas avoir un statut argumental et ne peuvent pas recevoir de thème-rôle. Les données étudiées ici corroborent cette hypothèse. Bien que les clitiques soient omises dans une proportion moindre, leur omission relève bien du même mécanisme : ils ne sont pas réalisés lorsqu'ils n'apportent aucune information supplémentaire quant au trait [Personne] dans un paradigme verbal. L'exemple d'Antoine et de son paradigme de *vouloir* à l'imparfait est particulièrement révélateur. Avec ce type d'hypothèse, nous touchons alors à la dichotomie initiale induite par le paramètre *pro-drop* mais nous avons vu précédemment que : (i) ce type de macro-paramétrage ne rend pas compte de la réalité de la micro-variation linguistique et (ii) il serait envisageable de poser une partition entre les mécanismes de suffixation verbale et ceux de préfixation. Dans le chapitre suivant, nous clôturons ainsi ce travail de recherches en proposant de nouvelles pistes de réflexions non pas sur un paramètre unique mais sur deux paramètres qui nous semblent plus à même de rendre compte de cette micro-variation : le paramètre du suffixe distinctif et le paramètre du préfixe. Mais terminons tout d'abord ce chapitre avec une synthèse chiffrée des informations recueillies au cours de cette analyse.

VII. SYNTHÈSE

1. Les principales catégories de sujets potentiels

Reprenons notre parcours en rappelant les structures et les chiffres attenants à chaque étape. Cinq catégories principales de sujets potentiels ont été étudiées au cours de cette analyse : les clitiques nominatifs utilisés seuls (*cf.* Tableau 65 ci-après), les clitiques nominatifs produits concomitamment à un constituant co-référent (*cf.* Tableau 66), les éléments *qu-* (*cf.* Tableau 67), les syntagmes nominaux seuls (*cf.* Tableau 68) et les absences de tout élément réalisé (*cf.* Tableau 69). Néanmoins, nous avons également vu que chacune de ces catégories se scinde en plusieurs sous-

¹ *Cf.* Section V.5 de ce chapitre pour une illustration.

classes. Les tableaux ci-dessous indiquent les types d'occurrences ainsi que leurs nombres et pourcentages à l'intérieur de chacune de ces catégories principales¹.

Le Tableau 65 illustre que les clitiques nominatifs sont quasiment tous préverbaux dans ce corpus enfantin. Nous avons souligné le fait que les rares clitiques postverbaux se rencontrent dans des occurrences qui ne semblent pas relever du langage enfantin spontané².

Types	N	%	Exemples
+PrC	5827	99,9	ah là <i>j'</i> ai fait un rêve (LUS, I.2)
+PoC	4	0,1	que se passe t <i>il</i> ? (EKT, IV.19d)
Totaux	5831	100,0	

Tableau 65 : les clitiques nominatifs utilisés seuls

Le Tableau 66 synthétise les requêtes effectuées afin d'établir la liste des constituants co-occurents d'un clitique nominatif. Nous avons noté que la partition pré- vs postverbal semble être liée aux différents types d'énoncés dans la mesure où la majorité des constituants préverbaux se trouve dans des énoncés non-interrogatifs³.

Constituant	N	%	Types	N	%	Exemples
Préverbal	2019	72,0	+PrN +C	412	14,7	et le chat <i>i(l)</i> joue (ANT, VII.29a)
			+PrP +C	1513	54,0	<i>moi je</i> prends le rouge (ALA, I.3)
			+PrAdv +C	94	3,4	là <i>c'</i> est très beau
Postverbal	689	24,6	+C +PoN	414	14,8	<i>i(l)</i> veut venir <i>Téo</i> (ANT, VII.29a)
			+PoC +PoN	2	0,1	où es tu <i>Maman_Ours</i> ? (CLA, VII.30b)
			+C +PoP	262	9,3	<i>c'</i> est une chenille <i>ça</i> (ROM, X.38c)
			+C +PoAdv	11	0,4	<i>c'</i> est Maxime là (MAX, IV.19b)
Pré & postverbal	96	3,4	+PrN +C +PoN	14	0,5	le feu <i>il</i> a brûlé le feu (QUE, VIII.33a)
			+PrP +C +PoP	82	2,9	<i>ça ça</i> se mange <i>ça</i> (TEO, XI.43c)
Totaux	2804	100,0		2804	100,0	

Tableau 66 : les clitiques nominatifs utilisés avec un autre constituant

¹ Dans les tableaux : +=présent, -=absent, Pr=préverbal, Po=postverbal, C=clitique, N=nom, P=pronom tonique, Adv=adverbe.

² Cf. Section II.1.2 de ce chapitre.

³ Cf. Section IV de ce chapitre.

Le total du Tableau 67 nous rappelle la parcimonie avec laquelle les jeunes enfants utilisent les structures interrogatives et relatives dont l'élément *qu-* a pour fonction le sujet de la proposition.

Types	N	%	Sous-types	N	%	Exemples
Interrogatif	99	26,3	<i>qui</i>	17	17,2	<i>qui</i> veut jouer ? (EKT, IV.19b)
			<i>c'est qui qu(i)</i>	36	36,4	<i>c'est qui qu'</i> a l(e) chat ? (LUS, V.23a)
			<i>qui+c+est+qui</i>	32	32,3	<i>qui+c+est+qui</i> veut les étoiles ? (EVA, I.3a)
			<i>qu+est+c(e)+qui</i>	14	14,1	<i>qu+est+c(e)+qui</i> se cache ici ? (MAX, XIII.47b)
			Sous-totaux	99	100,0	
Relatif	278	73,7	<i>c'est X qu(i)</i>	171	61,6	<i>c'est moi qui</i> pioche ! (LUS, X.38b)
			<i>y+a X qu(i)</i>	15	5,4	<i>y+a Tom qui</i> me l' a pris ! (NOE, XIII.48d)
			<i>N qu(i) V</i>	67	24,0	le chien <i>qui</i> se baigne (ISI, VII.30a)
			« vraies relatives »	25	9,0	j' avais le cochon <i>qui</i> va là (NOE, V.23a)
			Sous-total	278	100,0	
Totaux	377	100,0				

Tableau 67 : les éléments *qu-*

Le Tableau 68 confirme la disparité qui existe réellement entre la structure canonique du français « standard » (syntagme nominal + verbe) et la structure du français « spontané » puisque ces 39 énoncés représentent seulement 0,4 % des propositions qui contiennent un verbe fléchi dans le corpus. Nous avons suggéré une hypothèse afin d'expliquer pourquoi les jeunes enfants n'utilisent pas une telle structure et privilégient au contraire une dérivation syntaxique qui comprend un clitique préverbal morphologique dans la mesure où celui-ci permet de régler les questions d'accord plus tôt dans le cours de la dérivation¹.

¹ Cf. Sections V.5 et VI.2.6 de ce chapitre.

Types	N	%	Exemples
+PrN	30	76,9	alors <i>Max</i> prend la poule en chocolat (EVA, XIII.46a)
+PoN	9	23,1	<pas moi > [""] dit <i>le chat</i> (NOE, VIII.32a)
Totaux	39	100,0	

Tableau 68 : les syntagmes nominaux utilisés seuls

Le Tableau 69 indique que les enfants manipulent également deux types d'énoncés qui ne comprennent aucun élément réalisé. Nous avons suggéré un rapprochement entre ces deux catégories en termes de richesse de la morphologie verbale et nous avons constaté que la catégorie des clitiques nuls diminue au fur et à mesure que le système enfantin s'étoffe¹.

Types	N	%	Exemples
Impératif	1275	78,6	<i>regardez</i> ma p(e)tite [/] p(e)tite cabane (LUS, II.9b)
Clitique nul	348	21,4	euh 0 [*] sais pas (MAX, III.13d, 270)
Totaux	1623	100,0	

Tableau 69 : les absences de tout élément réalisé

2. La répartition de ces catégories dans le système enfantin

Pour terminer, regroupons dans le Tableau 70 ci-dessous les totaux de la section précédente afin de présenter une vue d'ensemble du système enfantin.

Types	N	%	Exemples
+C	5831	54,6	ah là <i>j'</i> ai fait un rêve (LUS, I.2)
+C +P/N	2804	26,3	<i>moi je</i> prends le rouge (ALA, I.3)
Impératif	1275	11,9	<i>regardez</i> ma p(e)tite [/] p(e)tite cabane (LUS, II.9b)
-C - P/N	348	3,3	euh 0 [*] sais pas (MAX, III.13d, 270)
+ <i>qu</i> rel	278	2,6	<i>c'</i> est moi <i>qui</i> pioche ! (LUS, X.38b)
+ <i>qu</i> int	99	0,9	<i>c' est qui qu'</i> a l(e) chat ? (LUS, V.23a)
+N	39	0,4	alors <i>Max</i> prend la poule en chocolat (EVA, XIII.46a)
Totaux	10674	100,0	

Tableau 70 : les types de sujets potentiels dans le système enfantin

Le Tableau 70 ci-dessus permet ainsi de formuler trois caractéristiques principales du système enfantin entre trois et quatre ans, caractéristiques que nous synthétisons en (121) ci-dessous :

¹ Cf. Section VI de ce chapitre.

- (121) Les principales caractéristiques du système enfantin entre trois et quatre ans :
- (a) Les enfants francophones continuent très majoritairement à utiliser un clitique nominatif associé à un verbe fléchi dans leurs énoncés (dans 80,9 % des cas).
 - (b) Le phénomène du clitique nul décroît fortement pendant cette période (3,3 % des énoncés en moyenne sur tout le corpus).
 - (c) La structure <syntagme nominal+verbe> est quasiment inexistante dans ce système (0,4 % des énoncés).

Il convient à présent de rendre compte de ces caractéristiques dans le cadre des paramètres de la grammaire générative et de définir lesquels de ces éléments sont effectivement des sujets syntaxiques dans ce système.

CHAPITRE V : CONCLUSIONS

Le débat sur le statut du clitique nominatif a été notre fil rouge tout au long de ce travail de recherches. Cela est dû au fait que cette problématique présente des tenants et des aboutissants dans plusieurs domaines concomitamment. Nous tirons parti de ce fait en explicitant nos conclusions au regard des trois domaines concernés : la question des « variétés » du français, celle du statut de ce clitique dans les différentes « variétés » en question et la façon d'en rendre compte dans le cadre des paramètres de la grammaire générative.

I. LES FRANÇAIS SPONTANÉ ET NORMÉ

Dans un premier temps, nous avons vu que ce débat se limite à une variété particulière de français : il s'agit de celle qui relève de la langue généralement qualifiée de « non-standard » ou de « parlée »¹. Ces systèmes sont alors généralement définis par rapport à un système « standard », considéré comme ensemble de référence ; ainsi, le système oral est celui dans lequel on « élimine » la particule négative *ne* par exemple². Ce type de description présuppose alors que les locuteurs de ce système oral maîtrisent par ailleurs le système pris comme référence : ils connaissent ainsi l'existence du *ne* à l'écrit mais l'omettent lorsqu'ils parlent. Or, aussi bien historiquement et qu'individuellement, la langue est toujours d'abord orale et seulement dans un deuxième temps écrite. Le point de vue didactique de Zribi-Hertz (2006) et notre point de vue acquisitionnel, différents des approches traditionnelles, permettent d'envisager les types de français autrement. En effet, dans les deux cas, il n'est pas considéré que l'écrit et l'oral représentent deux « variétés » d'un seul et unique français. *A contrario*, il est postulé que les « variations »

¹ Cf. Chapitre I. Ces termes sont ceux de Lambrecht (1981) et de Blanche-Benveniste (2000), respectivement.

² Terminologie utilisée par Lambrecht (1981).

répertoriées sont bien trop profondes et diverses et que ce qui est généralement considéré comme deux variétés ou registres sont en fait deux systèmes linguistiques, certes apparentés, mais distincts. Ainsi, Zribi-Hertz (2006) qualifie ces deux grammaires internes de « standard » et de « dialectale » ; nous proposons les termes de systèmes « normé » et « spontané ».

Par rapport aux dichotomies traditionnellement posées, nos points de vue impliquent une seconde différence fondamentale. En effet, s'agissant de deux systèmes distincts, le lien qui existe entre les deux ne peut pas être de même nature que celui qui est formulé par Lambrecht (1981) par exemple ; ainsi, le français spontané n'est pas un système dans lequel le locuteur « élimine » ou « remplace » des constituants du français normé. Le point de vue acquisitionnel offre alors une perspective chronologique radicalement différente dans la mesure où l'enfant naît en manipulant un premier système linguistique que nous qualifions de « spontané » dans la mesure où les paramètres de cette grammaire sont fixés de façon inductive par l'enfant, grâce à son environnement linguistique mais sans que celui-ci ne reçoive jamais la moindre instruction explicite (ce que nous qualifions donc d'*acquisition*). Puis, dans un second temps, notamment au fur et à mesure de sa scolarisation, l'enfant reçoit de plus en plus d'informations explicites, par le biais d'un discours métalinguistique puis d'enseignements écrits sur ce qu'il *faut* dire et écrire afin d'utiliser un français « correct » et afin de procéder aux analyses grammaticales requises lors de son parcours scolaire. Cet autre système n'est pas un français spontané auquel on aurait apporté quelques « corrections » ; il s'agit d'un *second* système linguistique, le français « normé », qui est ainsi davantage *enseigné* au jeune enfant. Celui-ci doit par exemple faire des exercices de grammaire dans lesquels il lui est demandé de transformer des phrases affirmatives en phrases négatives afin de manipuler le morphème discontinu de la négation correctement. Néanmoins, il est par ailleurs évident que ces deux systèmes présentent tout de même un certain nombre de caractéristiques communes et que le français normé n'entre pas dans la vie de l'enfant uniquement par le biais de l'école (et donc d'un enseignement explicite). Un jeune enfant auquel on lit beaucoup d'histoires aura ainsi un contact précoce, fréquent et inductif avec ce français normé. Il est à signaler que l'intonation joue alors un rôle majeur dans la distinction de ces deux systèmes¹. Ainsi, si tous les

¹ Cf. par exemple l'intonation qu'utilisent Hector et Quentin lorsqu'ils utilisent les clitiques postverbaux (cf. Chapitre IV, Section II.1.1).

enfants démarrent leur acquisition avec un système spontané, le moment auquel ils établissent les valeurs de ce deuxième système normé, la proportion et l'aptitude avec lesquelles ils manipulent l'un et l'autre sont extrêmement variables d'un enfant à un autre.

II. LE CLITIQUE EST D'ABORD MORPHOLOGIQUE

Nous avons vu que le statut du clitique nominatif en français normé ne provoque aucun débat linguistique dans la mesure où il est consensuellement qualifié de « pronom » au sens étymologique du terme, c'est-à-dire qu'il remplace un nom, qu'il est un argument syntaxique à part entière et qu'il représente ainsi le sujet syntaxique du verbe fléchi, comme illustré en (1) ci-dessous avec les exemples (a) et (b)¹. Le français normé admet néanmoins un second type de cas qui comporte également un clitique nominatif : il s'agit de la dislocation (*cf.* l'exemple (c) ci-dessous), c'est-à-dire une structure où le clitique nominatif reste le sujet syntaxique du verbe fléchi mais dans laquelle un élément co-indiqué au clitique est produit de manière disloquée, en dehors du noyau de la phrase et hors de la valence du verbe. Cet élément n'est donc pas le sujet syntaxique du verbe mais le topique de l'énoncé, ce dont on parle. De ce fait, un topique est généralement considéré comme relevant du domaine du défini. Ainsi, l'énoncé (d) ci-dessous qui comporte un article indéfini est perçu comme agrammatical en français normé.

(1) Le statut syntaxique du clitique nominatif en français normé :

- (a) *La thèse* traite d'acquisition.
- (b) *Elle* traite d'acquisition.
- (c) La thèse, *elle* traite d'acquisition.
- (d) *Une thèse, *elle* traite d'acquisition.

Or, de telles cooccurrences entre un syntagme indéfini et un clitique nominatif existent : la littérature sur le français « oral » l'atteste² et notre étude du système infantin le confirme³. Puisque ces énoncés sont réels, il faut alors que la théorie linguistique en rende compte. C'est notamment dans ce cadre qu'émerge l'analyse concurrente à la dislocation : celle qui propose une structure en termes de

¹ *Cf.* également Chapitre I, Section II.2.

² *Cf.*, entre autres, Chapitre I, Section III.1.2.2.

³ *Cf.* Chapitre IV, Section V.3.2.

doublement dans laquelle le clitique n'est plus syntaxique mais morphologique¹. Le sujet syntaxique n'est donc plus le clitique mais le syntagme co-indicé. Cette structure est illustrée en (2) ci-dessous pour le trentin (*cf.* (a)), langue pour laquelle une telle analyse morphologique ne fait pas débat et pour le français « non-standard » (*cf.* (b)), là où une telle analyse n'est pas consensuelle².

(2) Le statut morphologique du clitique préverbal :

(a) [_{IP} *El Gianni* [_I *el* [_{VP} *magna*]]]

« *le Jean* il mange »

(b) *Jean* il mange.

Ainsi, depuis les années 1980 et l'analyse de Rizzi (1986a) notamment, le débat sur le statut de ce clitique préverbal en français « oral » perdure avec des arguments de valeur de part et d'autre. L'un des objectifs de cette thèse a été d'apporter un regard différent sur cette question grâce à des données acquisitionnelles spontanées³. Nous synthétisons à présent les observations développées dans le courant du Chapitre IV, qui nous amènent à conclure que, en français spontané, les clitiques nominatifs sont manipulés comme des préfixes verbaux et qu'ils ressortissent ainsi à la morphologie. À cette fin, nous avons souhaité confronter nos données aux arguments développés par De Cat (2005) notamment qui, *a contrario*, propose une argumentation en faveur du statut syntaxique de ces éléments en français « parlé »⁴. Nous rappelons que les questions abordées par l'auteur le sont sous forme de problèmes que soulèverait une analyse morphologique des clitiques nominatifs et qu'ils concernent : la redondance des marques d'accord, l'optionnalité des clitiques préverbaux, la rareté du mécanisme de préfixation, l'inversion du clitique, le statut des éléments interférents (les clitiques non-nominatifs et la particule négative *ne*) et le lien entre la structure syntaxique et l'interprétation de l'énoncé. Nous avons déjà traité de la question de l'inversion (quasiment inexistante dans le système enfantin) et de la question des clitiques interférents (pour lesquels seul *ne* a été considéré comme bloquant la réalisation morphologique du clitique nominatif)⁵; nous développons à présent notre interprétation des autres thématiques proposées par De Cat (2005).

¹ *Cf.* Chapitre I, Section III.2.2.

² *Cf.* Chapitre II, Section IV.2.

³ *Cf.* Chapitre III.

⁴ *Cf.* Chapitre II, Section IV.2 pour le détail.

⁵ *Cf.* Chapitre IV, Sections II.1.2. et II.2.2 (sur l'inversion) et Section II.4 (sur les clitiques interférents).

1. Le préfixe ne marque pas l'accord de façon redondante

De Cat (2005) mentionne le fait que la présence d'un préfixe verbal « génère un certain niveau de redondance aléatoire dans le paradigme morphologique de l'accord verbal en français »¹. D'une manière générale, il nous semble tout d'abord que la redondance fait partie de tout système linguistique et que son existence ne suscite par ailleurs pas de polémique particulière ; les marques de genre par exemple sont extrêmement redondantes, comme l'illustre l'exemple (3) ci-dessous dans lequel le genre féminin est visible et audible un total de cinq fois.

(3) La redondance est un phénomène fréquent :

La petite fille intelligente est attentive.

Néanmoins, si la redondance devait être considérée comme un phénomène indésirable, nous pouvons argumenter qu'elle est loin d'être majoritaire dans le système verbal du français dans la mesure où les verbes *avoir*, *être* et *aller* mis à part, ce système comprend un nombre conséquent de formes verbales homophones, comme le montre le Tableau 71 ci-dessous. Il est également consensuel de constater que ce phénomène d'homophonie est renforcé par l'absence de la forme verbale de *Personne 4* dans les paradigmes oraux².

Groupes	Exemples	Formes homophones	Réalisations
1	<i>parler</i>	1=2=3=6	[paʁl]
2	<i>finir</i>	1=2=3	[fini]
3	<i>lire</i>	1=2=3	[li]
3	<i>rire</i>	1=2=3=6	[ri]

Tableau 71 : les formes verbales homophones en français

Par ailleurs, les deux systèmes verbaux enfantins dont nous reprenons les descriptions dans les tableaux ci-dessous³, permettent d'établir une corrélation entre la qualité de l'information véhiculée par la suffixation verbale et la réalisation du clitique préverbal. En effet, le premier corpus présente des formes verbales dont la suffixation est distinctive dans 72,9 % des cas (*cf.* Tableau 72) et un taux de clitiques nuls de 22,4 % alors que le deuxième corpus ne présente que deux formes verbales distinctives et un taux de clitiques nuls inversement proportionnel de 3,3 %.

¹ De Cat (2005:1197).

² Nos données vont également dans ce sens : ni les données enfantines ni les données adultes de ce corpus ne comportent de réalisations du clitique *nous*.

³ Ces deux descriptions sont issues respectivement de Palasis (2005) et de notre présente étude.

		<i>être</i>	<i>vouloir</i>	<i>faire</i>	<i>avoir</i>	<i>savoir</i>	<i>voir</i>	1 ^{er} groupe
Pers.	%	48,7	17,2	16,2	7,0	3,6	2,2	5,1
1sg	35,0	sqi	vø	fe	e	se	vwa	mãʒ
3sg	60,9	e	-	fe	a	se	vwa	mãʒ

Tableau 72 : les formes verbales enfantines (corpus 1)

		<i>être</i>	<i>avoir</i>	3 ^{ème} groupe	1 ^{er} groupe
Pers.	%	38,6	22,7	27,6	11,1
1sg	27,3	sqi	e	vø	mãʒ
2sg	5,7	e	a	vø	mãʒ
3sg	66,1	e	a	vø	mãʒ

Tableau 73 : les formes verbales enfantines (corpus 2)

Dans le corpus 2, les stratégies enfantines adoptées au niveau individuel quant à l'omission de certains clitiques pour certaines personnes dans des paradigmes bien particuliers¹ illustrent également cette complémentarité entre les parties suffixale et préfixale du verbe. La présence du clitique préverbal ne peut donc pas être qualifiée de « redondante » dans ce système.

2. Le préfixe n'est pas aléatoirement optionnel

D'après De Cat (2005), la présence non obligatoire des clitiques nominatifs dans les énoncés du français représente un second argument en faveur de leur analyse syntaxique. En effet, dans ce cadre, lorsqu'ils sont absents d'un énoncé, le sujet syntaxique est alors représenté par un autre élément tel qu'un syntagme nominal ou un pronom tonique de troisième personne (*les convoyeurs attendent ; eux pourront t'aider*). Par contre, s'ils sont analysés comme des affixes morphologiques, leur présence semble alors revêtir un caractère optionnel qu'il faut effectivement justifier.

De nouveau, nous suggérons que cet argument ne s'applique pas au système enfantin. En effet, d'une part, la configuration des paradigmes verbaux enfantins mentionnée dans la section précédente rend compte de la réalisation ou non de ces clitiques : dans ce cadre, leur présence n'est donc pas aléatoire. D'autre part, le modèle utilisé par De Cat (2005) pour évoquer cette « optionalité dans certains cas » est un modèle qui n'existe quasiment pas dans le système enfantin dans la mesure où les enfants ne produisent que 0,4 % d'énoncés qui comprennent un syntagme

¹ Cf. notamment l'exemple d'Antoine, Section VI.2 du Chapitre IV.

préverbal sans clitique et 3,3 % d'énoncés à clitique nul, comme illustré de nouveau dans le Tableau 74 ci-dessous¹.

Types	N	%	Exemples
+C	5831	54,6	ah là <i>j'</i> ai fait un rêve (LUS, I.2)
+C +P/N	2804	26,3	<i>moi je</i> prends le rouge (ALA, I.3)
Impératif	1275	11,9	<i>regardez</i> ma p(e)tite [/] p(e)tite cabane (LUS, II.9b)
-C - P/N	348	3,3	euh <i>0</i> [*] sais pas (MAX, III.13d, 270)
+ <i>qu</i> rel	278	2,6	<i>c'</i> est moi <i>qui</i> pioche ! (LUS, X.38b)
+ <i>qu</i> int	99	0,9	<i>c'</i> est <i>qui qu'</i> a l(e) chat ? (LUS, V.23a)
+N	39	0,4	alors <i>Max</i> prend la poule en chocolat (EVA, XIII.46a)
Totaux	10674	100,0	

Tableau 74 : les types de sujets potentiels dans le système enfantin

Les résultats de la tâche de production induite effectuée par l'enseignante auprès d'Alan² illustrent également le fait que les qualificatifs d'« aléatoires » et « optionnels » ne s'appliquent pas aux clitiques préverbaux du français spontané dans la mesure où nous avons constaté que, lors de cette tâche, Alan place un clitique devant chaque verbe fléchi alors que le modèle donné par l'enseignante relève *a contrario* du français normé.

3. La proclise est un mécanisme productif

D'après De Cat (2005), l'analyse morphologique des clitiques préverbaux implique que ces éléments « seraient les seuls marqueurs d'accord préfixaux en français oral »³. *Stricto sensu*, les propos de l'auteur décrivent effectivement la situation actuelle du français. Toutefois, nous aimerions argumenter en faveur du statut préfixal de ces clitiques grâce à des remarques qui concernent la diachronie de la langue.

Tout d'abord et d'une manière générale dans l'histoire des langues indo-européennes, le mécanisme de l'enclise précède celui de la proclise⁴. Plus précisément, en ce qui concerne les clitiques nominatifs du français, nous avons vu que ceux-ci ont d'abord été enclitiques avant de devenir progressivement proclitiques à partir de l'ancien français⁵. Ce schéma n'est pas atypique dans la diachronie du français puisque d'autres signifiés véhiculés en latin par le biais de la suffixation

¹ Cf. Section VII du Chapitre IV.

² Cf. Chapitre IV, Section V.

³ De Cat (2005:1198).

⁴ Cf. par exemple De Dardel (2001) sur l'évolution des pronoms en roman.

⁵ Cf. Chapitre I, Section II.2.3.1.

deviennent des proclitiques : les désinences casuelles par exemple sont complétées par des particules préposées puis sont remplacées par des prépositions¹. Nous avons déjà également mentionné le fait que certains éléments proclitiques deviennent parfois des préfixes (comme dans les exemples *s'enfuir* ou *s'envoler*²). Les déterminants du français moderne sont également des constituants proclitiques qui apparaissent progressivement dans l'histoire de la langue à partir de l'ancien français. L'hypothèse qui consiste à poser qu'une information d'ordre grammatical, d'abord véhiculée par la voie suffixale, est par la suite traduite de façon proclitique est donc compatible avec d'autres phénomènes observés dans la diachronie du français. Il reste alors à préciser ce que l'on entend par « proclitique » dans la mesure où un tel constituant peut être soit une tête syntaxique (D, s'il s'agit d'un pronom) soit un préfixe morphologique.

4. Les statuts morphologique et syntaxique coexistent

Nous avons noté dans le cours de l'analyse que : (i) les proclitiques nominatifs ne se rencontrent qu'avec des verbes fléchis et (ii) que le clitique *il* obéit à des mécanismes d'ajustements phonétiques extrêmement contraints, contrairement à d'autres éléments dans le corpus tels que les prépositions, les conjonctions ou les déterminants³. Le premier comportement fait partie des caractéristiques mises en avant par Zwicky & Pullum (1983) afin de distinguer les constituants syntaxiques des affixes morphologiques. En effet, les auteurs postulent que « les clitics peuvent se montrer peu sélectifs quant à leur hôte [phonologique], alors que les affixes sont très sélectifs quant au radical auquel ils s'attachent »⁴. Ainsi, l'élément enclitique *'s* par exemple (élision de *is* « est » en anglais) utilise tout hôte adjacent dans le continuum phonique, quel que soit son statut grammatical (préposition, verbe, etc.) alors que le suffixe du pluriel *-s* par exemple ne peut s'attacher qu'à des noms. Le second comportement évoqué (au sujet du clitique *il*) appelle deux remarques. La première concerne l'ajustement systématique réalisé par les enfants sur ce clitique. En effet, si tous les éléments du corpus présentaient les mêmes mécanismes, ce comportement ne serait pas significatif mais nous avons montré que les prépositions, les conjonctions ou les déterminants ne sont pas systématiquement

¹ Par exemple : latin : *e patria fugit* « il s'enfuit hors de sa patrie », avec la particule *e* « hors de » et le Cas ablatif sur le nom « patrie ».

² Cf. Chapitre IV, Section II.4.1.2.

³ Cf. Chapitre IV, Section II.4.2.

⁴ Zwicky & Pullum (1983:503).

éolidés par les jeunes enfants. Il semble donc que le clitique nominatif soit considéré comme plus proche du verbe fléchi que le déterminant ne l'est du nom par exemple. Nous interprétons cette différence comme étant un signe en faveur du statut morphologique des clitics préverbaux, qui se distinguent ainsi des déterminants par exemple *qui*, *eux*, sont des têtes syntaxiques. Cette remarque est notamment corroborée au niveau physiologique par des recherches effectuées sur l'activité du cerveau lors de la manipulation d'enclitiques verbaux en italien (par exemple : *portalo* « amène-le ») par Finocchiaro *et al.* (2008) qui concluent que « le différentiel [constaté pendant l'expérience] lors de la production d'énoncés verbe+clitique et verbe+déterminant+nom doit être attribué à la relation particulière qu'il y a entre les verbes et les clitics ».

Néanmoins, une partition a été établie entre le comportement majoritaire de *i(l)* mentionné ci-dessus et le comportement d'un petit nombre de clitics *il* *qui*, *eux*, ne montrent pas le même mécanisme d'ajustement devant consonne¹. Un contexte particulier a alors été mis au jour dans ces cas et il a été démontré qu'il s'agit de la réalisation phonétique de la tête négative *ne*. Cette différence nette de comportement nous a alors conduite à envisager l'existence de deux objets distincts chez les quatre enfants du groupe qui manipulent la particule négative, comme illustré en (4) ci-dessous : un affixe morphologique dont l'ajustement est fortement contraint (*cf.* (a)) et un argument syntaxique qui ne s'élide pas devant la consonne de la particule négative (*cf.* (b)). Ces deux objets n'apparaissant jamais concomitamment, il est également supposé qu'ils appartiennent à deux grammaires distinctes : le *i(l)* morphologique relevant du français spontané et le *il* argumental appartenant au français normé.

¹ *Cf.* Chapitre IV, Section II.4.4.2.

(4) Deux objets distincts :

(a) *i(l)* morphologique :

*il*_C : LUS: *i* nage sur moi . (XIII.46e, 1080)

EMA: si le loup était *i* nous mangerait ! (VII.27b, 1080)

*il*_V : EMA: regarde *il* a déchiré son livre . (III.10b, 79)

(b) *il* syntaxique :

*il*_ne : EMA: *il* ne chante plus . (XI.41d, 1404)

Une corrélation a ainsi été envisagée entre la réalisation de la tête négative *ne* et l'émergence de la réalisation du clitique *il* en tant que tête syntaxique, corrélation que nous reprenons en (5) ci-dessous :

(5) La présence de la particule *ne* joue sur le statut de *il* :

(a) Les clitiques nominatifs et non nominatifs entrent dans la numération en tant qu'affixes du verbe (exemple : N = {i-m(e)-donne, *argument silencieux*_[agent], *argument silencieux*_[bénéficiaire], un, livre}.

(b) Ce processus morphologique n'est pas bloqué par la présence du marqueur négatif *pas* (N = {i-m(e)-donne, *argument silencieux*_[agent], *argument silencieux*_[bénéficiaire], pas, un, livre}.

(c) L'analyse morphologique de *i(l)* décrite en (a) est totalement bloquée lorsque la numération comprend le marqueur négatif clitique *ne* (*N = {i-m(e)-donne, *argument silencieux*_[agent], *argument silencieux*_[bénéficiaire], ne, pas, un, livre} puisqu'une telle numération produirait la structure agrammaticale **ne i-m(e)-donne pas un livre*.

Une corrélation est également envisageable entre l'utilisation de *ne* et l'instanciation de la structure canonique du français normé. En effet, deux des seize énoncés qui comprennent *ne* sont des structures qui ne comprennent pas de clitique nominatif, comme illustré en (6) ci-dessous. Par ailleurs, les quatre enfants qui manipulent la particule *ne* font partie des neuf enfants qui produisent au moins une structure DP[^]V dans le corpus.

- (6) La corrélation entre la présence de *ne* et l'absence du clitique dans le corpus :
- (a) *le facteur n' est pas passé* . (EKT, IX.36, 2126)
 - (b) *ta photo n' est pas là ?* (LUS, II.5, 145)

Un faisceau de corrélations semble alors se dégager autour des trois phénomènes repris en (7) ci-dessous :

- (7) Faisceau de corrélations entre :
- (a) L'utilisation des clitiques préverbaux syntaxiques.
 - (b) La réalisation de la particule négative *ne*.
 - (c) L'instanciation de la structure DP[^]V.

À ce propos, Zribi-Hertz (2006) remarque que :

« Les corrélations entre propriétés sont solidaires de la cohérence interne qui préside nécessairement à toute grammaire interne. Le principe de cohérence est la clef de l' « apprenabilité » des grammaires. »¹

L'existence d'une telle corrélation constitue ainsi un indice qui œuvre en faveur des hypothèses formulées dans ce travail de recherches. Elle nous renvoie également à l'avantage d'utiliser des données spontanées dans la mesure où celles-ci apportent des informations sur la globalité d'un système linguistique, à la différence de certaines tâches de productions induites qui ciblent les investigations sur une structure en particulier². Par ailleurs, la possibilité de traduire une telle corrélation en termes de paramètres constituera également une indication positive de l' « apprenabilité » de ces grammaires du français spontané et du français normé.

5. Focalisation et préfixation sont compatibles

Par opposition au topique qui véhicule une information déjà connue, le focus est défini comme présentant une information nouvelle³. Pour le français, il est proposé que ce type d'information se trouve en spécifieur de TP, la position canonique du

¹ Zribi-Hertz (2006:2) d'après Barra-Jover (2004, 2007).

² Cf. Chapitre III, Section I.2.

³ Cf. Chapitre II, Section III.5.

sujet syntaxique¹. Dans le cadre de notre corpus enfantin, deux remarques peuvent alors être formulées.

Tout d'abord, dans la mesure où les enfants ne produisent pas (corpus 1) ou quasiment pas (corpus 2) la structure canonique DP[^]V, il faut envisager qu'en français spontané les informations nouvelles soient véhiculées par un autre type de configuration. Les exemples en (8) ci-dessous illustrent ainsi les deux stratégies majoritairement employées dans ce système : il s'agit essentiellement d'énoncés non-verbaux (*cf.* (a)) ou d'énoncés introduits par un présentatif (*cf.* (b)).

(8) Les structures qui apportent une information nouvelle :

(a) Les énoncés non-verbaux :

i. KAT: qui+est+c(e)+qui mange à la cantine ? (VII.27a, 2058)

LUS: *pas moi* .

ii. KAT: +, et qui+c+est+qui habite là+dedans ? (X.40b, 777)

NOE: *une souris !*

(b) Les présentatifs :

i. KAT: qui+c+est+qui fait ça ? (IX.34a, 485)

EMA: <c' est les> [/] <c' est> [/] c' est le [*] oiseau qui fait ça .

ii. KAT: qui+c+est+qui mange qui ? (VII.27b, 1677)

EMA: <c' est le> [/] ben [/] ben hm c' est le petit cochon qui le mange !

Toutefois, dans la mesure où nous postulons que les clitiques préverbaux sont des affixes morphologiques en français spontané et vu que les enfants produisent 26,3 % d'énoncés qui comprennent un syntagme et un clitique concomitants (et que ces énoncés ne sont pas considérés comme des topicalisations), parmi ces énoncés, il doit en exister qui apportent une information nouvelle dans la conversation. Les exemples en (9) ci-dessous illustrent cette possibilité².

¹ *Cf.* De Cat (2007).

² La nouveauté d'une information n'étant pas codée dans notre base, nous ne disposons pas à l'heure actuelle de taux précis sur cette thématique. Il s'agit donc d'un domaine dont l'étude pourra également être approfondie ultérieurement.

(9) Information nouvelle et préfixes verbaux :

(a) Le tout début d'une conversation avec Alan (I.2, 471) :

KAT: ça va Alan ?

ALA: *mon papa i(l)* vient me chercher +...

ALA: +, *ma maman i(l) [*]* vient me chercher .

(b) Antoine veut attraper un jeton sous un lit mais le lit est trop bas (XIII.49c, 254) :

ANT: nan je peux plus l' attraper .

KAT: tu peux pas l' attraper ?

ANT: oui [*] *ma tête elle* est trop grosse !

ANT: *i(l)* faut qu(e) je pousse mon lit .

(c) Une séance de « lecture » (IX.34a, 1127) :

KAT: qui c' est ?

LUS: un ours !

EMA: *un garçon i(l) [/]* il [/] il attrape un zoeu [*] [: œuf] .

Le français spontané autorise donc l'apport d'une information nouvelle par le biais d'un syntagme et d'un clitique dans la mesure où ce clitique est affixal, ce qui implique que le syntagme se trouve dans la position attendue afin de véhiculer cette information nouvelle, c'est-à-dire en spécifieur de TP. Dans le cadre de cette analyse, les structures qui présentent un syntagme indéfini (*cf.* (c) ci-dessus) sont donc possibles puisque le syntagme n'est pas un topique.

6. Les sujets silencieux

Afin de clore cette discussion sur le statut du clitique nominatif, il convient de décrire précisément quels sont les constituants *syntaxiques* dans les différents cas de figure évoqués et notamment quels sont les constituants *sujets* potentiels. Le Tableau 75 rend ainsi compte des différentes réalisations possibles et impossibles dans chacune des projections syntaxiques, étant entendu que TopicP accueille la dislocation, TP le sujet et vP le verbe fléchi.

		Dislocation	Sujet	Verbe fléchi
		TopicP	TP	vP
Français spontané	a			idort
	b		lui	idort
	c		l'escargot	idort
	d		un escargot	idort
	e	l'escargot		idort
	f	*un escargot		idort
	g	l'escargot	lui	idort
Français normé	a'		il	dort
	b'		lui	dort
	c'		l'escargot	dort
	d'		un escargot	dort
	e'	l'escargot	il	dort
	f'	*un escargot	il	dort
	g'	l'escargot	lui	dort

Tableau 75 : les constituants syntaxiques en français spontané et français normé

Dans ce cadre, seuls deux phénomènes distinguent alors le français spontané du français normé : toutes choses égales par ailleurs (comme l'illustrent les paires minimales x-x' dans le tableau ci-dessus), le français spontané utilise une préfixation verbale et la possibilité d'avoir recours à un sujet silencieux. Ainsi, les traits vérifiés par le sujet en français normé sont répartis en français spontané : les traits phi sont portés par le préfixe, ce qui permet d'attribuer une valeur aux traits phi de ν (en termes de [Personne], de [Genre] et de [Nombre]) dès que V se déplace vers ν . Les autres traits ainsi que le thêta-rôle sont portés par un sujet silencieux, ce qui permet la vérification du trait [Nominatif] en spécifieur de ν P et du trait *EPP* [μ D*] en spécifieur de TP. Lorsqu'une dislocation fait partie de l'énoncé, quel que soit le système de français, celle-ci est générée en dehors de la valence verbale et n'intervient donc pas dans le domaine du sujet syntaxique. Dans le cadre du français spontané, l'appellation de clitique « nominatif » se révèle alors partiellement inexacte dans la mesure où le clitique possède bien une forme morphologique qui évoque le Cas nominatif mais, pour autant, il ne vérifie pas effectivement le trait casuel [Nominatif] dans le cours de la dérivation. Par ailleurs, cet élément est bien un « clitique » mais ce terme fait uniquement référence à son statut phonologique, ce qui ne le différencie donc pas de l'argument pronominal qui est également un clitique mais dont la cliticisation au verbe fléchi n'intervient que de façon post-syntaxique¹.

¹ Dans le cadre du français normé, c'est cette cliticisation post-syntaxique qui permet notamment d'expliquer la différence en termes de liaison entre *les garçons ont* (sans liaison entre l'argument et le verbe puisque, dans ce cas, il n'y a pas de cliticisation) et *ils [z]ont* (où une proclise post-syntaxique a lieu entre l'argument *ils* et le verbe).

D'un point de vue morphologique, ce « clitique nominatif » pourrait alors porter l'appellation de « préfixe verbal », ce qui permettrait de le distinguer de l'argument pronominal mais une telle appellation subsumerait à la fois les éléments « nominatifs » et les clitiques qui sont morphologiquement marqués pour d'autres cas, tels que l'accusatif ou le datif. D'un point de vue syntaxique, le « clitique/préfixe nominatif » véhicule des valeurs de [Personne], de [Genre] et de [Nombre], ce qui permet de les attribuer à *v* dans le cours de la dérivation. Ces traits possèdent à leur tour la caractéristique d'être « interprétables », c'est-à-dire d'être manipulables par la composante sémantique. Ces « clitiques/préfixes nominatifs » sont donc des éléments qui sont visibles à la fois en morphologie, en syntaxe, en phonologie et en sémantique, ce qui rend chacune des appellations utilisées partielle et ce qui leur a notamment valu le surnom de « petits monstres » de la part de Lambert (2001). Cette question terminologique n'est cependant pas nouvelle et les termes existants ne semblent pas apporter de solution. L'appellation « préfixe nominatif » nous semble néanmoins raisonnablement rendre compte du statut de cet élément (dans le cadre du français dans la mesure où les sujets sont nominatifs) car il permet une opposition d'une part avec l'appellation « argument nominatif » et d'autre part avec les termes « préfixes accusatifs », etc.

Par ailleurs, les taux de latéralisation observés dans notre corpus lorsque le clitique est doublé par un syntagme nominal ou un pronom tonique sont également compatibles avec l'analyse morphologique de ces éléments. En effet, la majorité des énoncés non-interrogatifs comporte un double latéralisé à gauche du verbe préfixé¹ : il s'agit de la position canonique du sujet en spécifieur de TP. La tendance inverse pour les énoncés interrogatifs (la latéralisation à droite) s'explique par la montée du verbe fléchi vers la périphérie gauche, laissant ainsi le sujet en spécifieur de TP, donc à sa droite. Il nous reste à présent à rendre compte des occurrences proposées dans le Tableau 75 ci-dessus en termes de paramètres.

III. LA FISSION DU PARAMÈTRE *PRO-DROP*

En français spontané, l'importance accordée à la préfixation verbale ainsi que la répartition des traits entre un préfixe (morphologique) et un sujet (syntaxique)

¹ Cf. Chapitre IV, Section IV.1.4 : la non-interrogation montre un taux de latéralisation à gauche de 78,53 % et l'interrogation affiche un taux de latéralisation à droite de 69,05 %.

obligent à considérer le paramètre syntaxique *pro-drop* sous un angle nouveau. Tout d’abord, il faut noter que les différences qui sont observées entre les deux français ne relèvent pas du domaine strict de la syntaxe mais que la différenciation s’opère à deux niveaux : la morphologie *et* la syntaxe. Dans un second temps, nous devons intégrer à notre réflexion des données d’une langue *pro-drop* afin de rendre compte des différences entre les différents systèmes. Le Tableau 76 ci-après intègre ainsi les données de l’italien normé et la projection FocusP proposée par Rizzi (1997a)¹ :

	Dislocation	Focus	Sujet	Verbe fléchi		
	TopicP	FocusP	TP	prfx	V	suffix
Français spontané				i	dort	
			l’escargot	i	dort	
	l’escargot		lui	i	dort	
Français normé			il		dort	
			l’escargot		dort	
	l’escargot		il		dort	
Italien normé					dorm	e
		la chiocciola			dorm	e
		ella			dorm	e

Tableau 76 : la représentation du sujet dans les différents systèmes linguistiques

1. Le paramètre de la suffixation verbale distinctive

Si nous reprenons le principe du miroir de Baker (1985) mentionné précédemment, nous devons considérer ces informations de droite à gauche. Dans cette perspective, l’italien se distingue des deux français dans la mesure où il s’agit du seul système linguistique sur les trois dont la suffixation verbale apporte une information distinctive quant au trait [Personne]². Une première dichotomie apparaît ainsi entre deux types de langues : celles qui possèdent des suffixes verbaux distinctifs et celles pour lesquelles l’information suffixale n’est pas aussi riche. Nous proposons alors qu’une première partition puisse être établie au sein des langues quant à cette suffixation verbale et que l’enfant, lors de son acquisition, ait à fixer la valeur d’un premier paramètre *morphologique* : le paramètre de la suffixation verbale distinctive. Comme l’indique le Tableau 77 ci-dessous, la valeur de ce paramètre est négative pour les français spontané et normé et positive pour l’italien.

¹ Cf. Chapitre II, Section III.5 sur la fission de la périphérie gauche.

² Le paradigme de *dormire* (*dormo, dormi, dorme, dormiamo, dormite, dormono*) est en effet distinctif.

Systèmes	Suffixation verbale distinctive
Français spontané	–
Français normé	–
Italien normé	+

Tableau 77 : le paramètre de la suffixation verbale distinctive

2. Le paramètre de la préfixation verbale

Puis, procédant toujours de droite à gauche, nous constatons que la morphologie peut disposer d'un second moyen pour véhiculer une information grammaticale : il s'agit de la préfixation verbale. Le Tableau 76 ci-dessus montre alors que, des trois systèmes représentés, seul le français spontané dispose de tels préfixes. Nous suggérons ainsi qu'une seconde dichotomie puisse être établie entre les différents systèmes linguistiques : ceux dont la morphologie contient des préfixes verbaux et ceux dont la morphologie ne dispose pas de tels morphèmes. Un second paramètre entre ainsi en jeu : le paramètre de la préfixation verbale. Le Tableau 78 ci-dessous illustre alors cette seconde partition : le français spontané dispose de cet outil préfixal ; la valeur du paramètre de préfixation verbale est donc positive. *A contrario*, le français normé ne dispose pas d'un tel outil ; la valeur est ainsi négative. Dans Palasis (2009), nous avons proposé que ce second paramètre ne s'applique pas dans le cas des langues qui disposent déjà d'une suffixation verbale distinctive et nous avons posé la valeur 0 au paramètre de préfixation verbale pour l'italien normé par exemple. Or, les paradigmes de certains dialectes de l'Italie du nord, tels que le trentin, illustrent qu'un paradigme verbal peut disposer d'une suffixation distinctive et d'une préfixation de façon concomitante¹. Nous insérons alors le trentin dans le Tableau 78 ci-dessous et illustrons les deux valeurs du paramètre de préfixation verbale pour deux langues dont la valeur du paramètre de suffixation verbale est positive.

¹ Ce que les dialectologues italiens qualifient de « réduplication de l'accord ». Cf. Chapitre II, Section IV.2.1.

Systèmes	Préfixation verbale	Suffixation verbale distinctive
Français spontané	+	–
Français normé	–	–
Italien normé	–	+
Trentin	+	+

Tableau 78 : les paramètres de suffixation verbale distinctive et de préfixation verbale

A priori, le paramètre de préfixation verbale semble également ressortir à la morphologie. Toutefois, dans le cadre du Tableau 78 ci-dessus, il est visible qu'à ce stade (qui est effectivement purement morphologique), les paradigmes du français normé ne disposent pas d'informations grammaticales suffisantes quant à la [Personne] dans la mesure où les deux paramètres sont négatifs. La conséquence immédiate de cette lacune morphologique est que le système doit apporter cette information autrement : c'est alors l'argument syntaxique sujet qui remplit ce rôle (par le biais de la transmission de ses traits phi à T). L'absence d'information morphologique quant aux traits phi lorsque V se déplace en *v* se répercute ainsi directement dans la syntaxe qui requiert la réalisation d'un sujet. Il semble alors que ce second paramètre puisse être qualifié de morphosyntaxique dans la mesure où il a des implications dans les deux domaines (à moins que ce soit la morphosyntaxe qui soit un domaine unique).

Le français spontané est ainsi proche du français normé en termes de suffixation verbale mais est également proche du trentin en termes de préfixation, les deux systèmes les plus éloignés étant alors le français normé et le trentin. Dans ce cadre, ce qui est généralement qualifié de français « oral » adulte ne représente donc pas un système linguistique à proprement parler. Il s'agit d'un ensemble de réalisations produit par des locuteurs qui manipulent deux systèmes (le français spontané et le français normé), ce que Zribi-Hertz (2006) qualifie de « diglossie ». Les variations qui sont observées au sein de ces réalisations proviennent alors de la fréquence avec laquelle chacune de ces deux grammaires est activée. Cette répartition provient à son tour d'un ensemble complexe de facteurs environnementaux et pragmatiques. Il nous semble néanmoins qu'une constante peut être dégagée en ce qui concerne l'ordre d'implémentation de ces deux grammaires : d'abord le français spontané puis le français normé. En effet, l'émergence vers l'âge de quatre ans de phénomènes corrélés relevant du français normé, tels que la réalisation de la particule négative, appuient cette hypothèse.

IV. POUR TERMINER

1. De nombreuses pistes à développer

Tout au long de ces investigations, notre base de données spontanées a apporté un nombre considérable d'informations par le biais des données elles-mêmes mais également grâce aux différents niveaux de codages appliqués à ces données. Ainsi, de nombreux domaines qui ont été abordés dans la perspective de l'étude du sujet syntaxique pourront être étudiés ultérieurement sous des angles différents et complémentaires. Nous avons notamment survolé les questions concernant l'acquisition du genre, du nombre, des structures interrogatives et impératives ainsi que la question de l'interprétation du triplement dans le cadre de la théorie des copies. Nous avons également notamment seulement abordé la question du placement des éléments *qu-* par rapport aux doubles, comme illustré de nouveau en (10) ci-dessous¹. Néanmoins, une étude approfondie de ces positions dans le système de chaque enfant pourrait apporter de plus amples informations sur la fréquence de l'utilisation de structures telles que (a) et (b)². Les questions d'ordres prosodique et sémantique représentent également des domaines d'étude extrêmement riches mais qui restent encore à exploiter.

(10) Les positions relatives de l'élément *qu-* et du double :

(a) Double_i ^ *qu-* ^ clitique_i ^ V :

NOE: et ben *le monsieur qu+est+c(e)+qu' i(l)* fait ? (II.7, 950)

NOE: et *lui comment i(l)* s' appelle ? (IV.18a, 1187)

(b) *Qu-* ^ double_i ^ clitique_i ^ V :

EVA: et *pourquoi Lola* elle veut pas jouer ? (III.10a, 801)

ALA: *pourquoi lui i(l)* s' appelle un [/] un castor ? (XII.45, 1226)

Une autre piste que nous aimerions également suivre est celle de la quantification. En effet, depuis le Chapitre I, nous avons croisé le chemin de ce trait plusieurs fois. Il a bien sûr été évoqué à plusieurs reprises dans le cadre de la discussion sur la topicalisation³ mais nous notons que ce trait émerge également dans l'analyse que propose Schapansky (2009) de la négation dans la mesure où l'auteur postule que la particule négative *ne* possède non seulement un trait de négation positif [+N] mais

¹ Cf. Chapitre IV, Section IV.1 sur le doublement.

² Cf., entre autres, Chapitre I, Section III.1.2 sur cette distinction.

³ Cf., entre autres, Chapitre I, Section III.1.2.3.

également un trait de quantification négatif et ininterprétable [-uQ] qui doit donc être vérifié par l'homologue interprétable présent sur *pas* [+N, +Q]¹. Lorsque la particule *ne* est réalisée dans un énoncé, le trait [Q] doit alors être manipulé pour que la dérivation converge. Or, nous avons vu que la topicalisation d'un constituant ne peut se faire que si l'élément topicalisé est [+Q] (**un escargot, il dort*). Les deux structures (le morphème discontinu de la négation et la topicalisation) requièrent certes des valeurs différentes pour [Q] mais il nous semble intéressant que, dans les deux cas, elles requièrent que ce trait ait une valeur, quelle qu'elle soit. Vu le très faible nombre de pronoms indéfinis produits par les enfants au cours des enregistrements², nous nous interrogeons notamment sur le moment et les conditions d'émergence de ce trait dans le système enfantin.

2. Une base de données destinée à s'agrandir

Les énoncés en (10) ci-dessus illustrent que certaines investigations portent sur des questions de positions dans la hiérarchie syntaxique. Les lignes dépendantes implémentées jusqu'à présent dans cette base (%mor et %err), bien que permettant des requêtes en termes d'ordre et/ou d'adjacence des constituants, semblent alors trouver leurs limites. En effet, nous avons notamment déjà évoqué les problèmes soulevés par : (i) l'étiquetage du mot graphique et non du syntagme et (ii) l'absence d'un étiquetage global en termes de fonctions syntaxiques³. Dans la mesure où ces deux types de codes relèvent d'une ligne dépendante supplémentaire (%syn), il convient alors de s'interroger sur l'ajout éventuel de ce codage à la base de données déjà réalisée et/ou aux bases à venir. En effet, ce premier corpus ne restitue que la première année des enregistrements effectués au sein de cette école maternelle. Dix-neuf de ces vingt-deux enfants⁴ ont ainsi été suivis pendant leur seconde année de maternelle en 2007-2008 puis le groupe se réduit à dix-huit enfants⁵ pour la troisième années d'enregistrements. De nombreuses données primaires additionnelles attendent ainsi d'être traitées et il est fort probable qu'elles apporteront à leur tour des indications extrêmement intéressantes sur l'évolution du sujet syntaxique chez les enfants francophones entre quatre et six ans ainsi que sur de nombreuses autres thématiques relevant du développement du langage.

¹ Cf. Chapitre IV, Section II.4.2.

² Cf. Chapitre IV, Section V.3.2.

³ Cf. Chapitre IV, Sections IV et V.

⁴ Éva, Claire et Zoé déménagent entre les deux premières années d'enregistrements.

⁵ Alan déménage entre la deuxième et la troisième année.

Ainsi, à l'heure où nous clôturons cette étude, nous repensons avec émotion et gratitude à nos petits informateurs qui entrent à présent à l'école élémentaire, ainsi qu'à leurs enseignants. Les moments partagés avec toutes ces personnes au cours de ces trois années font que cette collecte de données, avant de devenir une étape dans un travail de recherches, représente une somme de souvenirs ainsi qu'un ensemble d'échanges forts et enrichissants.

ANNEXES

I. LA CONSTITUTION DE LA BASE DE DONNÉES

1. Le dossier d'information destiné aux parents

Université de Nice-Sophia Antipolis
Faculté des Lettres, Sciences Humaines et Sociales
98 bd Edouard Herriot, 06204 Nice Cedex 3
Contact : Katerina.PALASIS@unice.fr

CNRS UMR 6039 « Bases, Corpus, Langage »

IMPORTANT

Note d'information à l'attention des parents des élèves
de la petite section de l'école maternelle de XXX
Année scolaire 2006/2007

Pourquoi ce document ?

La classe de petite section de l'école maternelle de XXX a été choisie pour faire l'objet d'une **étude scientifique sur l'acquisition du langage**. Ce document a donc pour but de **vous informer** des modalités de cette étude et de **solliciter votre autorisation** en tant que responsable légal de l'un des enfants concernés.

Pourquoi la classe de votre enfant ?

L'étude scientifique envisagée porte sur **l'acquisition de la syntaxe par des enfants francophones à partir de deux ans**. Elle sera menée par Katerina Palasis, doctorante de l'Université de Nice-Sophia Antipolis et du laboratoire CNRS (Centre National de la Recherche Scientifique) « Bases, Corpus, Langage » (UMR 6039), sous la direction de Jean-Philippe Dalbera, Professeur des Universités, et de Michèle Oliiviéri, Maître de Conférences en Sciences du Langage. Cette doctorante est également mère de trois enfants et a donc l'habitude d'être en contact avec des jeunes enfants.

La classe de votre enfant a été choisie car elle correspond à la tranche d'âge recherchée et se trouve dans la zone géographique de l'Université de Nice-Sophia Antipolis.

Que va-t-il se passer exactement dans la classe ?

Le but de cette étude est d'effectuer des **enregistrements audio et vidéo** des enfants de la classe au cours de leurs activités scolaires, pendant les heures habituelles de classe, dans les locaux de l'école. Le chercheur fera donc tout pour rester le plus discret possible afin de ne pas perturber le déroulement de la classe. Les séances d'enregistrements s'effectueront à un rythme le plus régulier possible (probablement une demi-journée toutes les deux ou trois semaines) et s'échelonneront sur l'année scolaire, en accord avec la directrice de l'école et l'enseignante.

Que vont devenir les enregistrements ?

Dans un premier temps, ces enregistrements sont destinés à être **transcrits et analysés** par la doctorante présente dans la classe dans le cadre de sa thèse de doctorat dont le titre est *Syntaxe générative et acquisition : le sujet dans le développement du système linguistique du jeune enfant*. L'analyse consiste donc, entre autres, en une **étude statistique** de l'utilisation d'éléments tels que les pronoms et les groupes nominaux sujets.

L'étude porte uniquement sur la structure des phrases produites par les enfants et aucun jugement de valeur ne sera effectué, à aucun moment, sur le contenu des productions des enfants ni sur l'état du développement langagier des enfants. Les transcriptions des enregistrements apparaîtront en annexe de la thèse. **Le cadre d'utilisation sera strictement scientifique** (aucun but commercial ou lucratif). Les données recueillies serviront par exemple à confirmer ou infirmer des hypothèses existantes sur l'acquisition de la syntaxe et, peut-être, à proposer des hypothèses nouvelles.

Ces données audio et vidéo pourront également être reprises dans d'autres communications scientifiques (revues, conférences, etc.). Elles seront aussi destinées à être partagées avec la **communauté scientifique internationale**, notamment par le biais d'une base de données informatique internationale du nom de CHILDES (Child Language Data Exchange System) dont vous pouvez visiter le site à l'adresse <http://childes.psy.cmu.edu> (sachant que les données ne sont accessibles qu'aux chercheurs abonnés ayant adhéré à une **charte déontologique** que vous pouvez également consulter sur ce site).

Notre vie privée sera-t-elle préservée ?

Le nom et la localisation exacte de l'école maternelle où se déroule l'enquête ne seront jamais mentionnés. Le corpus constitué sera donc identifié comme « provenant de la petite section d'une école maternelle des Alpes-Maritimes, France, pendant l'année scolaire 2006/2007 ».

Les noms de famille des enfants ne seront jamais mentionnés.

Un **devoir d'anonymisation** oblige le chercheur à ne pas utiliser les vrais prénoms des enfants mais à employer des prénoms pseudonymes.

La seule donnée divulguée par le chercheur concernera l'âge des enfants car cette information représente une donnée scientifique importante lorsqu'il s'agit d'étudier le développement du langage. Cependant, il va de soi que, vu la nature des enregistrements (audio et vidéo), un anonymat total est impossible, la voix et le visage des enfants ne pouvant que difficilement être modifiés ou masqués.

Le chercheur s'engage à **respecter la vie privée des enfants et de leurs familles**. Ainsi, toutes les informations personnelles telles que les prénoms des membres de la famille et les noms de lieux seront également modifiées. Le chercheur s'engage également à ne pas discuter du contenu des enregistrements en dehors du cadre strictement scientifique.

Quels sont mes avantages à autoriser ces enregistrements ?

Il doit être clair qu'il n'y a **aucun avantage financier** à participer à cette étude. Personne n'est et ne sera rémunéré. Par contre, le fait de pouvoir enregistrer votre enfant représente un **atout scientifique majeur**. En effet, la France ne dispose pas d'un grand nombre d'enregistrements enfantins (contrairement aux pays anglophones comme les Etats-Unis) et elle a maintenant pris du retard dans ce domaine. Votre autorisation représenterait donc vraiment une **contribution importante** au développement d'une base de données francophone. Cependant, vous seriez également tout à fait en droit de ne pas participer à cette étude.

Si vous souhaitez en savoir plus sur la constitution de corpus oraux, le cadre juridique, le devoir d'anonymisation, le droit à l'image, etc., vous pouvez par

exemple consulter le site de la Délégation Générale à la Langue Française et aux Langues de France (www.dglff.culture.gouv.fr) qui vient de publier un guide intitulé *Corpus oraux. Guide des bonnes pratiques 2006*.

Aurai-je accès aux enregistrements et aux résultats de l'étude ?

A tout moment, vous pouvez contacter le chercheur en charge de l'enquête à l'adresse suivante : Katerina.PALASIS@unice.fr. Elle se fera un plaisir de vous faire accéder aux données de votre enfant et de répondre à toutes vos questions, interrogations ou doutes éventuels.

Un ou plusieurs entretiens pourront également être organisés si certains parents souhaitent en savoir plus sur l'avancement des travaux.

Que dois-je faire pour participer ?

Il suffit de **lire, remplir et signer l'autorisation parentale** ci-jointe et de la retourner à l'école **le plus rapidement possible**. Cette autorisation est le dernier document administratif dont nous avons besoin pour procéder à cette étude. Nous avons en effet déjà sollicité et obtenu les autorisations de Monsieur l'Inspecteur d'Académie, de Madame Grattarola, Inspectrice de l'Education Nationale chargée de la circonscription de Nice 1, de Madame Claire XXX, Directrice de l'Ecole Maternelle de XXX et de Madame Sophie XXX, votre enseignante.

**Nous vous remercions vivement
de votre attention et de votre coopération.**

2. Le formulaire d'autorisation parentale

Université de Nice-Sophia Antipolis
Faculté des Lettres, Sciences Humaines et Sociales
98 bd Edouard Herriot, 06204 Nice Cedex 3
Contact : Katerina.PALASIS@unice.fr

CNRS UMR 6039 « Bases, Corpus, Langage »

AUTORISATION PARENTALE

(à retourner le plus rapidement possible à l'école, merci)

Je, soussigné(e), (NOM, Prénom)

représentant légal de l'enfant (NOM, Prénom, Date de naissance)

- déclare avoir pris connaissance de la note d'information en pages 1 à 3 ci-jointes, concernant les modalités de l'étude menée par Katérina PALASIS, doctorante au sein de l'Université de Nice-Sophia Antipolis et du laboratoire CNRS UMR 6039 « Bases, Corpus, Langage ».
- En conséquence de quoi, j'autorise K. Palasis à enregistrer en audio et vidéo mon enfant pendant les heures habituelles de classe, dans les locaux de l'école maternelle de XXX, selon les modalités mentionnées dans la note d'information ci-jointe.
- J'ai bien noté que ces enregistrements en audio et vidéo sont destinés à être utilisés à des fins scientifiques uniquement et que ceci implique une diffusion et un partage de ces données avec d'autres membres de la communauté scientifique (revues, conférences, bases de données informatiques, etc.).
- Il est entendu que ni les enfants ni leurs familles ne bénéficieront d'une rémunération, ni pour avoir autorisé les enregistrements, ni pour avoir autorisé leur diffusion.

Fait à _____, Le _____,

(Signature du responsable légal de l'enfant)

Merci de renseigner les rubriques ci-dessous :

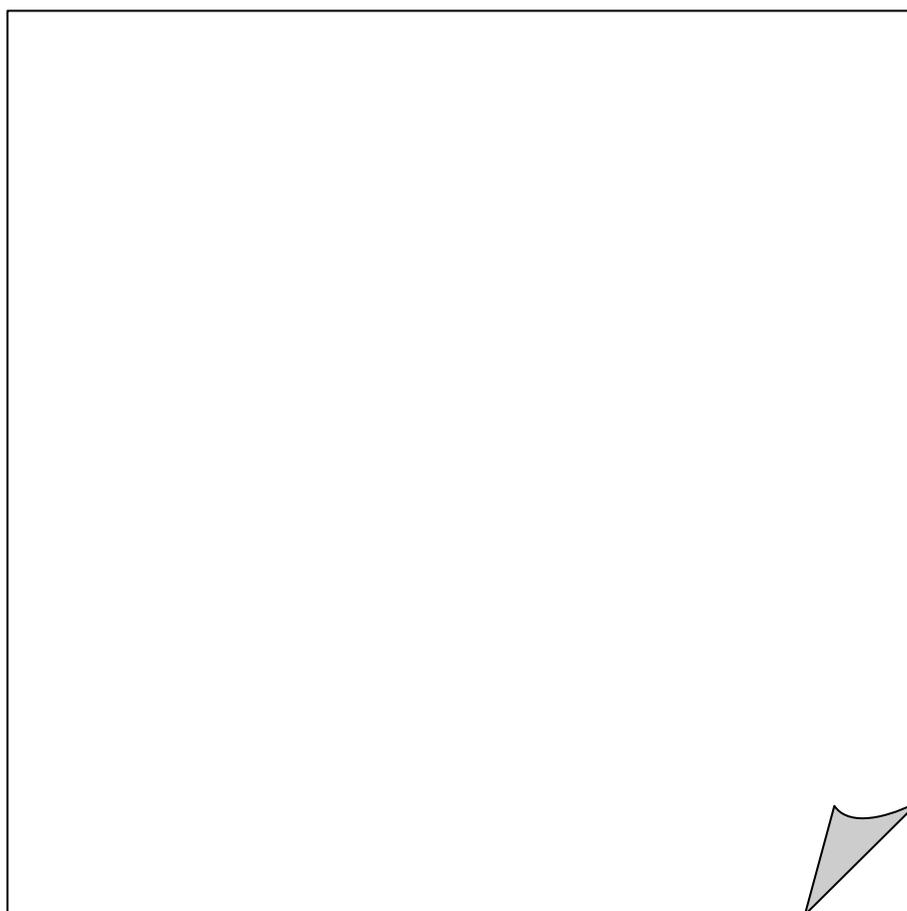
- Langue(s) parlée(s) par la mère à l'enfant (s'il y a plusieurs langues, merci de les indiquer par ordre décroissant de fréquence) : _____
- Langue(s) parlée(s) par le père à l'enfant (s'il y a plusieurs langues, merci de les indiquer par ordre décroissant de fréquence) : _____
- L'enfant a-t-il des frères et/ou sœurs ? Si oui, merci d'indiquer leurs prénoms et dates de naissance : _____

Et merci de cocher la case ci-dessous si vous souhaitez être informé de l'avancement des travaux :

Je souhaite être informé de l'avancement des travaux à l'adresse e-mail suivante : _____

3. Les fichiers informatiques de la base de données

Sur ce CD se trouvent les 121 fichiers informatiques qui composent notre base de données. Chaque fichier comporte : des entêtes @, des lignes de transcription *XXX ainsi que les lignes de codages %mor et %err adéquates.



4. Liste des codes %err

4.1 Tournures incorrectes : \$INC ou \$POS

Descriptions	Codes	Exemples
Général	\$INC	et i(l) vient parc(e)+que avec sa maman avec le panier (VI.26a)
Négation incorrecte	\$\$SYN \$INC \$NEG	on a pas jamais joué beaucoup aux arbres ! (X.39)
Adverbe incorrect	\$\$SYN \$INC \$ADV	très beaucoup (XIII.46d)
Mauvaise position	\$\$SYN \$POS	un <pas bon> médicament
Inversion	\$\$SYN \$POS \$EX	<i(l) lui> (=lui i(l)) va au panier (VIII.32a)
Adjectif	\$\$SYN \$POS \$ADJ	je fais un château gros (VIII.33a)
Adverbe	\$\$SYN \$POS \$ADV	j' ai pas mis bien ça (VII.29c)
Conjonction	\$\$SYN \$POS \$CONJ	parc(e)+que i(l) pleure 0 i(l) peut pas descendre (VIII.31b)
Clitique objet	\$\$SYN \$POS \$PRO \$OBJ	i(l) m' a le volé (V.23e)
Mauvaise position négation	\$\$SYN \$POS \$NEG	c'était le fantôme pas (I.4)

4.2 Omissions d'éléments : \$LOS

Descriptions	Codes	Exemples
Élément non spécifié	\$\$SYN \$LOS	nan j(e) suis en 0 (=train de) li(re) (IV.19a)
Adverbe interrogatif	\$\$SYN \$LOS \$ADV \$INT	et ça 0 (=comment) i(l) s' appelle hein? (IV.19d)
Conjonction	\$\$SYN \$LOS \$CONJ	va dire à Isia 0 (=que) i(l) tombe le livre (I.1a)
Déterminant (pourrait être soit déf soit indéf)	\$\$SYN \$LOS \$DET	et lui i(l) a 0 (=une/la) cass(e)role (VI.24b)
Déterminant défini	\$\$SYN \$LOS \$DET \$DEF	i(l) mange 0 (=la) souris le chat (I.4)
Déterminant indéfini	\$\$SYN \$LOS \$DET \$INDEF	oh y+a 0 (= un) oiseau dedans ! (IV.18a)
Déterminant possessif	\$\$SYN \$LOS \$DET \$POSS	(j)e veux 0 (=ma) maison (II.7)
Déterminant partitif	\$\$SYN \$LOS \$DET \$PARTIT	j' ai entendu 0 (= du) bruit (VII.27b)
Préposition	\$\$SYN \$LOS \$PREP	0 (= à) qui c' est le cochon ? (IV.19b)
Pronom sujet non identifié	\$\$SYN \$LOS \$PRO \$SUBJ	0 (je/il) veux [=? faut] remett(r)e ça (VI.26b)

Pronom sujet 1S	\$\$SYN \$LOS \$PRO \$\$SUBJ \$1S	moi 0 (=je) veux le jaune (I.1a)
Pronom sujet 2S	\$\$SYN \$LOS \$PRO \$\$SUBJ \$2S	0 (=tu) as vu j' ai rangé (3b)
Pronom sujet 3S	\$\$SYN \$LOS \$PRO \$\$SUBJ \$3S	0 (=il) veut pas manger (I.2)
Pronom sujet explétif	\$\$SYN \$LOS \$PRO \$\$SUBJ \$IMPRS	0 (=il) faut ranger ! 0 (=il y/y') a un hamster (VII.29b)
Pronom c' ou ce	\$\$SYN \$LOS \$PRO \$\$SUBJ \$DEM	0 (=c') est le panda (IX.36) est+(ce)+que 0 [*] t(u) as le chien ? (X.40f)
Pronom sujet 3P	\$\$SYN \$LOS \$PRO \$\$SUBJ \$3P	0 (=ils?) mangent [?] des herbes (II.9a)
Complément	\$\$SYN \$LOS \$OBJ	moi je vois 0 (=?: me, y, DP, ?, etc.) (IV.19e)
Pronom accusatif 1S	\$\$SYN \$LOS \$PRO \$OBJ \$1S	on 0 (=m') écoute moi aussi (I.2)
Pronom accusatif 3S	\$\$SYN \$LOS \$PRO \$OBJ \$3S	elle va 0 (=le) donner à maman (I.1a)
Pronom accusatif 3P	\$\$SYN \$LOS \$PRO \$OBJ \$3P	j' 0 (=les) ai déchiré (II.7)
Pronom réfléchi 1S	\$\$SYN \$LOS \$PRO \$REFL \$1S	je 0 (=m') en vais (VI.26c)
Pronom réfléchi 3S	\$\$SYN \$LOS \$PRO \$REFL \$3S	i(l) 0 (=se) réveille (VI.26a)
Pronom datif 1S	\$\$SYN \$LOS \$PRO \$DAT \$1S	tu peux 0 (=me) donner la boîte ? (IV.19b)
Pronom « y » locatif	\$\$SYN \$LOS \$PRO \$LOC	i(l) veut pas 0 (=y) aller. (VI.24b)
Pronom « en » ablatif	\$\$SYN \$LOS \$PRO \$ABL	il 0 (=en) a besoin (VI.24b)
Pronom « en » quantitatif	\$\$SYN \$LOS \$PRO \$QUANT	et moi j' 0 (=en) ai plus (VI.26a)
Pronom interrogatif	\$\$SYN \$LOS \$PRO \$INT	euh 0 (=que) s(e) passe+t+il ? (IV.19d)
Pronom relatif	\$\$SYN \$LOS \$PRO \$REL	qui c' est 0 (=qui) se cache ? (IX.37a)
Verbe de modalité	\$\$SYN \$LOS \$V \$MDL	ben moi je 0 (=veux/vais) prendre le li(vre) (IV.19e)
Auxiliaire	\$\$SYN \$LOS \$AUX	quelqu+un qui 0 (= est) coincé là (VII.27b)
« Root infinitive »	\$\$SYN \$LOS \$INFL	0 (=il faut) fermer les yeux (IV.19e)
S+V	\$\$SYN \$LOS \$MAIN	0 (=c'est) Katarina qui m(e) la donné (VIII.31b)
Participe passé	\$\$SYN \$LOS \$PART	elle est pas encore 0 (=finie) (VII.27b)

4.3 Ajouts d'éléments : \$ADD ou \$RED

Descriptions	Codes	Exemples
Ajout d'une préposition	\$\$SYN \$ADD \$PREP	je mets le casque avec (II.5)
Ajout d'un pronom	\$\$SYN \$ADD \$PRO	qu+est+c(e)+qui 0 [*] passe+t+il [*] ? (19d)
Ajout d'un pronom dans une relative	\$\$SYN \$ADD \$PRO \$REL	c' est le xx qui i(l) va tout les manger (II.7)
Répétition d'un adverbe	\$\$SYN \$RED \$ADV	moi j' ai presque gagné presque (II.9a)
Répétition d'un circonstanciel	\$\$SYN \$RED \$CIRC	tout+à+l+heure t(u) étais pas là tout+à+l+heure (VIII.31a)
Répétition d'un pronom	\$\$SYN \$RED \$PRO	moi je fais la fourmi moi (I.1a)
Répétition d'un syntagme nominal	\$\$SYN \$RED \$N	<mon petit chien> il est mon amou(r) <mon petit chien> (I.4)
Triplement	\$\$SYN \$RED	lui il est ici le petit poussin (X.38b)
Répétition de la possession	\$\$SYN \$ADD \$POSS	ma maman à moi (II.5)
Réduplication possession	\$MOR \$RED \$POSS	ça c' est <mon mien> (= le mien) (I.2)

4.4 Substitutions d'éléments : \$SBL ou \$SUB

Descriptions	Codes	Exemples
Amalgame non réalisé	\$MOR \$LOS \$SBL	c' est la couronne <de le> (=du) monsieur (VII.27a)
Amalgame en trop	\$MOR \$ADD \$SBL	i(ls) veulent aller au (=à) l' eau (VI.24b)
Substitution de catégorie	\$\$SYN \$SUB \$CAT	alors c' est l' histoire de quoi ? de Grande (=adj) ! (VIII.31b)
Erreur d'accord sur le Nom / déterminant	\$MOR \$SUB \$AGC \$N	c'est un animaux (=animal) (IV.19c)
Erreur d'accord sur le Déterminant / nom	\$MOR \$SUB \$DET \$NUM	mes (= mon) coffre à bijoux (VII.27a)
Erreur d'accord sur le Déterminant	\$MOR \$SUB \$ACC \$DET	mais où il est ce (=cet) oiseau ! (IV.19c) ça c' est un (= une) pâte à crêpes (VI.26b)
Pronom non élidé	\$MOR \$SUB \$ACC \$PRO	faut le (=l') habiller ! (IV.18a)
Conjonction non élidée	\$MOR \$SUB \$ACC \$CONJ	pa(r)c(e)+que (=qu') i(l) veut sa maman encore alors (VI.26a)
Préposition non élidée	\$MOR \$SUB \$ACC \$PREP	de (=d') une carotte et de (=d') une tomate. (IX.34a)

Mauvaise forme déterminant défini/indéfini	\$LEX \$\$SUB \$DET \$DEF	elle est gentille une (=l') zabelle (I.4) un (= le) chapeau de Lucie (V.23e)
Mauvais genre sur le pro	\$MOR \$\$SUB \$AGA \$PRO	il (=elle) a pris ça à moi Noémie (II.8)
Mauvais genre sur l'adj	\$MOR \$\$SUB \$AGA \$ADJ	moi j' étais <tout seul> (=toute seule) dans l' eau (VI.26c)
Mauvais genre sur participe	\$MOR \$\$SUB \$AGA \$PART	y+a Tom qui me l' a pris (=prise) XIII.48d
Mauvais genre sur nom	\$MOR \$\$SUB \$AGA \$N	copain = copine (XIII.48d)
Mauvais nombre sur SN	\$MOR \$\$SUB \$NUM \$N	c' est ta meilleure couleur le rouge et le bleu ? (IX.34b)
Mauvaise personne sur V	\$MOR \$\$SUB \$AGC \$V	je m'en va (= vais)
Mauvaise personne sur Clitique+V	\$MOR \$\$SUB \$AGC \$PRO \$V	les monsieurs elle a (= ils ont) (V.23b)
Mauvais nombre sur le Cl+V	\$MOR \$\$SUB \$NUM \$PRO \$V	il a (=ils ont) perdu leurs mitaines (VII.29c)
Mauvais temps / mauvaise forme du temps sur V	\$MOR \$\$SUB \$TNS	ma maman elle veut pas que j(e) <vais [=? veux] aller> (=aille) au Carnaval (VI.26c) il s'assit (=assoit/assied) (XI.41a)
Mauvaise forme verbale : flexion /non-flexion	\$MOR \$\$SUB \$NFL	je veux le range (= ranger) (V.20) il ouvre sa bouche pour <il mange> (= manger) (VI.26a)
Substitution forme verbale non-fléchie	\$MOR \$\$SUB \$NONFIN	il a finir (= fini) (IV.19b)
Mauvais nombre sur adj	\$MOR \$\$SUB \$NUM \$ADJ	trois petit (=petits) ours (VII.30a)
Surgénération participe passé	\$MOR \$\$SUB \$REG \$PART	moi je l' ai ouvri (= ouvert) (IV.19a)
Surgénération infinitif	\$MOR \$\$SUB \$REG \$INF	prendre (=prendre) les cartes (VIII.32b)
Surgénération pluriel	\$MOR \$\$SUB \$REG \$PL	ça c' est des chevaux (= chevaux) (II.7)
Mauvais pronom sujet	\$MOR \$\$SUB \$PRO \$SUBJ	là i(l) (=je) par(le) moi (II.9b)
Mauvais cas sur le pronom	\$MOR \$\$SUB \$PRO \$K	elle le (=lui) fait quoi là 0 [*] le garçon (XII.44e)
Mauvais pronom objet / mauvaise forme du pronom (fort au lieu de faible)	\$MOR \$\$SUB \$PRO \$OBJ	je le (=me) vois (IV.19e) on écoute moi (=m') ! (II.9d)

Mauvais pronom réfléchi	\$MOR \$SUB \$PRO \$REFL	je s' (=m') appelle pas Maxime (II.9a)
Pronom au lieu de adj int	\$SYN \$SUB \$ADJ \$INT	c' est <le train qui> (= quel train) ? (IV.19d)
Pronom au lieu de adj poss	\$SYN \$SUB \$ADJ \$POS	c'est le perroquet <de moi> (= mon) (VII.29b)
Quelque chose au lieu de pronom interrogatif	\$SYN \$SUB \$PRO \$INT	et toi quelle (=laquelle) tu as ? (IV.19b)
Quelque chose au lieu d'un adverbe interrogatif	\$SYN \$SUB \$ADV \$INT	i(l) s'appelle qui (=comment) ? (IV.18a)
Quelque chose à la place d'un pronom relatif	\$SYN \$SUB \$PRO \$REL	c' est moi j' (=qui) ai fini (3a)
Quoi/que au lieu de qu+est+ce+que	\$SYN \$SUB \$WH	que (=qu+est+ce+que) tu as donné à Lucie ? (VI.24b)
Même code pour est+ce+que au lieu de ce+que		je sais qu+est+c(e)+que (=ce que) c' est (X.40c)

4.5 Liaisons : \$ALL

Descriptions	Codes	Exemples
Absence de liaison obligatoire	\$ALL \$LOS	[oewazo] = [oenwazo] (VII.29b)
Présence de liaison interdite	\$ALL \$ADD	les mamans zarrivent (V23e) Les chat qui zont (VIII.33b)
Mauvaise forme du nom	\$ALL \$SUB \$N	j'ai le noiseau (= oiseau) (IV.19d)
Mauvaise forme du verbe	\$ALL \$SUB \$V	j'ai louvert = ouvert (V23e)

4.6 Phonologie : \$PHO

Descriptions	Codes	Exemples
Ajout de consonne	\$PHO \$ADD \$CON	on blit (= lit) (V.22)
Ajout de consonne par persévération	\$PHO \$ADD \$PER \$CON	un crocodile (I.3) Nonël (II.8)
Ajout de consonne par anticipation	\$PHO \$ADD \$ANT \$CON	manrionette (III.10a) trigre (VII.29a)
Inversion de consonnes	\$PHO \$EX \$CON	bédut (= début) (I.3)
Inversion de voyelles	\$PHO \$EX \$VOW	camaré (= caméra) (IX.34b)
Substitution de voyelle	\$PHO \$SUB \$VOW	un harisson (= hérisson) ! (XII.44c)
Substitution de consonne ou glide / voyelle	\$PHO \$SUB \$CON	poude (= poule) (III.10a) ploché = pioché (VII.30b)

Substitution par persévération	\$PHO \$SUB \$CON \$PER	la caméma (= caméra) (IV.19a)
Substitution de consonne impliquant un seul trait	\$PHO \$SUB \$CON \$FEA	sais (= fais) voir (V23d)
Déplacement d'une consonne	\$PHO \$SH \$CON	toture = tortue (VI.26d) rous = ours (X.38a)
Blend	\$PHO \$BLE	l(e m)édicament (V.23a)
Perte d'une syllabe	\$PHO \$LOS \$SYL	hi(ppo)potame (III.10a) (hé)risson (IX.34b)
Ajout d'une syllabe	\$PHO \$ADD \$SYL	la caraméra (=caméra) (XII.45a)
Réduplication d'une syllabe	\$PHO \$RED \$SYL	nonocéros (=rhinocéros) (III.13h)
Simplification d'un cluster	\$PHO \$LOS \$CC	ézargot (= escargot) (IV.19e)

4.7 Lexique : \$LEX

Descriptions	Codes	Exemples
Blend	\$LEX \$BLE	par(ce)+quoi (= parce+que/pourquoi) (VI.24a)
Substitution	\$LEX \$SUB	pourquoi (= parce+que) (III.3)
Substitution oui/non	\$LEX \$SUB \$YN	ça ça vous intéresse pas ? oui (=non). (VIII.31a)
Mauvais adjectif	\$LEX \$SUB \$ADJ	elle est blanc (= blonde) (II.5)
Mauvais verbe	\$LEX \$SUB \$V	et moi j' ai (=suis) avec ma mamie (II.5)
Mauvais auxiliaire	\$LEX \$SUB \$AUX	et moi je suis (=ai) fait un escargot. (I.1a)
Mauvais adverbe	\$LEX \$SUB \$ADV	je vois pas arrière (= derrière) là (VI.26a)
Mauvaise préposition	\$LEX \$SUB \$PREP	j' étais à [/] à (=dans) ma chambre (VI.26a)
Mauvaise conjonction	\$LEX \$SUB \$CONJ	j'ai un pansement donc (= parce+que) j'ai un bouton (VII.27a)
Même code pour que que = ce que		oh moi je sais pas que (=ce) que c' est ! (X.40d)
Mauvais nom mais existant	\$LEX \$SUB \$N	i(l) pleut des cornes (= cordes)
Mauvais nom (approximation)	\$LEX \$SUB \$N \$NW	c' est un [nystɛʁ] (=hamster) (VII.30b)

II. LES ÉNONCÉS

1. Base de données complète

	1	2	3	4	5	6	7	8	9	10	11	12	13	T/loc
QUE			51	38	14			137						240
LAU				47			22			24				93
LEA		63	50						56					169
EVA	32	21	14	106			89						257	519
EKT		171	84	252	40	174		198	125	75				1119
CLA	3	22					18							43
TEO		50	216				47	66		155	132		44	710
ISI	1			61	10		100					31		203
ELO	7	38		59					61	173		68	74	480
TOM	111	232	3	35	16	174	203	156	123	73			123	1249
JUL	13						100			265	79	24	107	588
CEL		38		6					148		115	11		318
LUS	89	329	261	273	250	159	226	115	201	345	260	346	282	3136
NOE		108		180	222	193	30	141	84	194	105	52	181	1490
ANT	132		303		168	89	181	223	356	76	375		268	2171
KLE					40	36		29			7			112
ROM	18			136	135	58	31		101	66	42			587
ALA	57	43	183		139	105	36	109	89	40	50	59	96	1006
EMA	181		106		64	88	232	70	160	175	275	234	144	1729
LOL	75	66	8				56		100	128			85	518
MAX	30		77	130				47		249		32	60	625
ZOE	94	123	1		126	5					6			355
UNI	29	11												40
KAT	803	1131	822	1094	1031	965	1154	732	792	1432	1118	705	1202	12981
SOF	26	2		4	1									33
TAA			5	3	9		7	4						28
T/sce	1701	2448	2184	2424	2265	2046	2532	2027	2396	3470	2564	1562	2923	30542

Tableau 79 : nombres d'énoncés par locuteur et par séance¹

¹ Requêtes sur le modèle : mlu +u @ > SousTotX (avec X = n° de la séance et @ = les fichiers correspondants à cette séance).

2. Phrases verbales et non verbales

	+V	%	-V	%	Tot
QUE	159	66,25	81	33,75	240
LAU	66	70,97	27	29,03	93
LEA	87	51,48	82	48,52	169
EVA	386	74,37	133	25,63	519
EKT	573	51,21	546	48,79	1119
CLA	25	58,14	18	41,86	43
TEO	386	54,37	324	45,63	710
ISI	121	59,61	82	40,39	203
TOM	771	61,73	478	38,27	1249
JUL	289	49,15	299	50,85	588
CEL	225	70,75	93	29,25	318
LUS	1682	53,64	1454	46,36	3136
NOE	911	61,14	579	38,86	1490
ANT	1367	62,97	804	37,03	2171
KLE	52	46,43	60	53,57	112
ROM	376	64,05	211	35,95	587
ALA	608	60,44	398	39,56	1006
EMA	979	56,62	750	43,38	1729
MAX	422	67,52	203	32,48	625
ZOE	263	74,08	92	25,92	355
KAT	9782	75,36	3199	24,64	12981
SOF	24	72,73	9	27,27	33
TAA	14	50	14	50	28
Totaux	19568	66,32	9936	33,68	29504

Tableau 80 : Phrases verbales et non verbales par locuteur

III. LES SYNTAGMES NOMINAUX SEULS

1. Statistiques

1.1 Tous locuteurs confondus

	1	2	3	4	5	6	7	8	9	10	11	12	13	Totaux	%	T/Gr	%/Gr
QUE				1										1	1,45	39	56,52
LAU														0	0,00		
LEA														0	0,00		
EVA													4	4	5,80		
EKT									1					1	1,45		
CLA														0	0,00		
TEO											1			1	1,45		
ISI														0	0,00		
TOM														0	0,00		
JUL														0	0,00		
CEL														0	0,00		
LUS		1				1	1	2	1	3	2	3	4	18	26,09		
NOE								8					1	9	13,04		
ANT	1													1	1,45		
KLE														0	0,00		
ROM														0	0,00		
ALA			1				1							2	2,90		
EMA			1										1	2	2,90		
MAX														0	0,00		
ZOE														0	0,00		
KAT		3		3	4		1	3		3	3	3	2	25	36,23	30	43,48
SOF	2													2	2,90		
TAA			1		1		1							3	4,35		
Totaux	3	4	3	4	5	1	4	13	2	6	6	6	12	69	100,00	69	100,00

1.2 Enfants

Tous	1	2	3	4	5	6	7	8	9	10	11	12	13	T	%
QUE				1										1	2,56
LAU														0	0,00
LEA														0	0,00
EVA													4	4	10,26
EKT									1					1	2,56
CLA														0	0,00
TEO											1			1	2,56
ISI														0	0,00
TOM														0	0,00
JUL														0	0,00
CEL														0	0,00
LUS		1				1	1	2	1	3	2	3	4	18	46,15
NOE								8					1	9	23,08
ANT	1													1	2,56
KLE														0	0,00
ROM														0	0,00
ALA			1				1							2	5,13
EMA			1										1	2	5,13
MAX														0	0,00
ZOE														0	0,00
T	1	1	2	1	0	1	2	10	2	3	3	3	10	39	100,00
%	2,56	2,56	5,13	2,56	0,00	2,56	5,13	25,64	5,13	7,69	7,69	7,69	25,64	100	

2. Liste exhaustive

Loc.	Références	Enoncés
QUE	IV.19d, 166	oh <i>le train</i> est là .
EVA	XIII.46a, 1063	alors <i>Max</i> prend la poule en chocolat .
	XIII.46a, 1112	sûr(e)ment que [/] que <i>sa soeur</i> a pris tous les oeufs .
	XIII.46a, 1469	et <i>sa soeur</i> cherche <i>Max</i> .
	XIII.46a, 1565	et <i>quelqu'un</i> regarde la main .
EKT	IX.36, 2126	<i>le facteur</i> n' est pas passé .
TEO	XI.43a, 604.	<i>le beau vert</i> ici [/] est pas ici .
LUS	II.5, 145	<i>ta photo</i> n' est pas là ?
	IX.34b, 923	et pour écrire <i>les enfants</i> vient [*] avec toi .
	VI.24a, 343	là <i>tous les oiseaux</i> s' en vont .
	VII.27a, 2584	alors <i>un autre garçon</i> va à sa place !
	VIII.31a, 526	et après+que <i>Caramel</i> voit par la fenêtre +...
	VIII.31a, 534	+, si <i>maman</i> est là .
	X.38a, 899	et <i>la souris</i> veut casser ses pics .
	X.38a, 1740	alors <les taupes> [/] <les taupes> [/] <les taupes> [/] <i>les taupes</i> va [*] dans l(e) nid .
	X.38c, 1436	+, <i>les vers+de+terre</i> sont tout gluants .
	XI.43a, 513	<i>tout+l(e)+monde</i> a déjà fait .
	XI.43c, 1492	<i>tout+le+monde</i> a gagné .
	XII.44a, 1119	<i>tout+le+monde</i> a répondu .
	XII.44c, 500	alors je le retourne et <i>personne</i> me le retourne comme ça .
	XII.44d, 201	alors <i>la maman</i> prépare quelque+chose dans le four .
	XIII.46a, 70	<et mon> [/] et <i>ma petite fourchette</i> marche marche .
	XIII.46a, 76	et <i>ma petite cuillère</i> marche marche .
	XIII.46e, 1065	euh <i>le poisson</i> nage sur moi .
	XIII.46b, 765	vole <i>mon petit chagrin</i> xxx .
	NOE	VIII.32a, 885
VIII.32a, 890		<pas moi > ["] dit <i>le canard</i> .
VIII.32a, 912		<pas moi > ["] dit <i>le cochon</i> .
VIII.32a, 917		<pas moi > ["] dit <i>le chat</i> .
VIII.32a, 922		<pas moi > ["] dit <i>le chat</i> .
VIII.32a, 927		<pas moi > ["] dit <i>le canard</i> .
VIII.32a, 932		<pas moi > ["] dit <i>le cochon</i> .
VIII.32a, 970		<pas moi > ["] dit <i>le cochon</i> .
XIII.48b, 847		et comme <i>les chats</i> ont besoin de xx .
ANT	I.4, 347	mon [/] <i>mon petit chat</i> 0 [*] appelle <i>Mouna</i> .
ALA	III.13e, 585	<et ses> [/] et <i>ses pieds</i> sont pieds nus .
	VII.29a, 287	<i>sa maman</i> et <i>son papa</i> vont faire le lait .
EMA	III.10b, 393	si quand [/] quand euh le [/] <i>le papa</i> est [/] est fâché .
	XIII.46e, 420	<i>quelqu'un</i> veut choisir un p(e)tit bonbon ?

KAT	II.9a, 388	alors <i>tout+le+monde</i> va me raconter une histoire .
	II.9a, 1540	peut-être que <i>Tom</i> le sait .
	II.9d, 188	<i>Lucie</i> a des dents magnifiques .
	IV.19c, 249	<i>personne</i> te le prend .
	IV.19f, 237	comme ça <i>chacun</i> a un carton .
	V.20, 1616	<i>mon sac</i> est vide .
	V.20, 1902	ben quand <i>Antoine</i> a fini d'accord ?
	V.23e, 559	<i>quelqu+un</i> t' a volé ta casquette ?
	V.23g, 42	<i>tout+le+monde</i> a envie de jouer .
	VII.29b, 1281	<i>personne</i> a le papillon ?
	VIII.31a, 930	et voilà <i>maman</i> arrive .
	VIII.31b, 1131	est+c(e)+que <i>quelqu+un</i> sait ?
	VIII.32a, 431	voilà comme ça <i>tout+le+monde</i> voit bien .
	VIII.32a, 1274	ah ben <i>chacun</i> peut choisir un jeu différent .
	X.40b, 449	comme ça <i>chacun</i> pioche ?
	X.40b, 790	alors est+c(e)+que <i>quelqu+un</i> a la souris ?
	X.40b, 874	est+c(e)+que <i>quelqu+un</i> a la maison de l' écureuil ?
	XI.42, 493	<i>tout</i> est là .
	XI.43b, 1938	et est+c(e)+que <i>quelqu+un</i> a un chat ?
	XI.43b, 2446	est+c(e)+que <i>quelqu+un</i> d' autre veut aller dans la classe ?
	XII.44b, 469	<i>tout+le+monde</i> se souvient ?
	XII.44d, 55	est+c(e)+que <i>quelqu+un</i> veut me raconter ?
	XII.44d, 87	<i>Lucie</i> l' a dit d'abord .
XIII.46a, 619	est+c(e)+que <i>quelqu+un</i> veut me raconter autre chose sur son week+end ?	
XIII.46e, 1815	est+c(e)+que <i>quelqu+un</i> d' autre veut partir ?	
SOF	I.1b, 20	<i>un nuage</i> arrive et fait tomber la pluie !
	I.1b, 28	<i>l' escargot</i> sort une corne deux cornes et s' en va se promener .
TAA	III.13h, 140	est+c(e)+que <i>Sophie</i> pourrait avoir <i>Lucie</i> ?
	V.23a, 1835	ben quand <i>Katarina</i> t' appellera .
	VII.30b, 1106	aujourd'hui <i>maman</i> vient te chercher trésor d'accord ?

IV. LES VERBES

1. Les verbes dans les énoncés à clitique nul

Verbes	QU	LA	LE	EV	EK	TE	IS	TO	JU	CE	LU	NO	AN	RO	AL	EM	MA	ZO	Tot
<i>veux</i>	9			1	1	17	1	42	3	1	3	2	2	23	20	6	10	2	143
<i>faut</i>		2	1	1	1		1	7	1	2	8	2	10		5	15		1	57
<i>peux</i>					1	1		12			1	2			3	1	5		26
<i>est</i>						7	1	1			2			2	1	5	1	1	21
<i>a</i>					1	2		1				2	1		6	1		2	16
<i>voulais</i>													11						11
<i>veut</i>								5					2				1		8
<i>sais</i>											4		1				2	1	8
<i>vais</i>								4										2	6
<i>peux 2</i>																	6		6
<i>peut</i>											4								4
<i>ferme 3</i>									1			2	1						4
<i>ai</i>						1								2	1				4
<i>prête 1</i>					2									1					3
<i>ferme 1</i>				1										1		1			3
<i>fais 1</i>	1				1							1							3
<i>va</i>												1				1			2
<i>ouvre 3</i>																		2	2
<i>enlève 3</i>					2														2
<i>dis 1</i>										2									2
<i>appelles</i>			2																2
<i>relève 1</i>													1						1
<i>ramasse 3</i>													1						1
<i>prends 1</i>													1						1
<i>pose 1</i>						1													1
<i>passe se</i>					1														1
<i>nage 3</i>													1						1
<i>mets 1</i>													1						1
<i>manque 3</i>												1							1
<i>mangent</i>					1														1
<i>mange 1</i>																1			1
<i>jette 1</i>											1								1
<i>habille se</i>					1														1
<i>fait</i>												1							1
<i>cache 3</i>					1														1
<i>as</i>															1				1
Totaux	10	2	3	3	13	29	3	72	5	5	23	14	33	29	37	31	25	11	348

2. Les verbes enfantins dans l'ensemble du corpus

Codes %mor	T	Enf ≠
v:*\ être&pres&3sv	3110	19
v:*\ avoir&pres&1sv	951	19
v:*\ avoir&pres&3sv	803	19
v:*\ vouloir&pres&1sv	590	19
v:*\ aller&pres&3sv	250	19
v:*\ faire&pres&3sv	215	16
v:*\ falloir&pres&3sv	163	16
v:*\ avoir&pres&2sv	151	19
v:*\ vouloir&pres&3sv	123	19
v:*\ être&impf&3sv	109	15
v\ voir&pres&1sv	102	13
v\ appeler&pres&_3sv	100	17
v\ mettre&pres&3sv	98	14
v:*\ être&pres&1sv	79	19
v:*\ savoir&pres&1sv	76	13
v\ manger&pres&_3sv	72	11
v:*\ pouvoir&pres&1sv	72	12
v:*\ vouloir&pres&2sv	67	15
v:*\ aller&pres&1sv	61	17
v:*\ savoir&pres&2sv	59	12
v:*\ pouvoir&pres&2sv	57	10
v\ ranger&pres&_3sv	54	10
v\ cacher&pres&_3sv	53	11
v:*\ être&impf&1sv	53	14
v:*\ faire&pres&1sv	48	14
v:*\ pouvoir&pres&3sv	48	9
v\ mettre&pres&1sv	44	11
v:*\ être&pres&3pv	42	11
v\ prendre&pres&1sv	42	9
v\ prendre&pres&3sv	41	12
v:*\ avoir&impf&3sv	36	7
v:*\ être&pres&2sv	33	11
v\ voir&pres&3sv	33	9
v\ pleurer&pres&_3sv	31	6
v\ dire&pres&3sv	29	11
v\ fermer&pres&_1sv	28	11
v\ fermer&pres&_3sv	28	9
v\ jouer&pres&_3sv	28	8
v\ dormir&pres&3sv	25	8
v\ voir&pres&2sv	25	12
v\ tomber&pres&_3sv	25	11
v\ pleuvoir&pres&3sv	24	8
v:*\ aller&pres&2sv	23	8

(suite)	T	Enf ≠
v\ arriver&pres&_1sv	23	9
v\ couper&pres&_3sv	23	5
v:*\ avoir&impf&1sv	23	8
v\ monter&pres&_3sv	21	7
v:*\ vouloir&impf&1sv	20	5
v\ manquer&pres&_3sv	20	10
v\ raconter&pres&_1sv	20	6
v\ marcher&pres&_3sv	19	7
v\ rester&pres&_3sv	19	7
v\ sauter&pres&_3sv	19	5
v\ croire&pres&1sv	18	4
v\ enlever&pres&_3sv	17	8
v:*\ faire&pres&2sv	16	7
v\ ouvrir&pres&3sv	16	6
v\ écouter&pres&_3sv	16	7
v:*\ avoir&pres&3pv	16	7
v\ habiller&pres&_3sv	15	7
v\ piquer&pres&_3sv	15	6
v\ courir&pres&3sv	14	6
v\ donner&pres&_1sv	14	6
v\ boire&pres&3sv	13	5
v\ jouer&pres&_1sv	13	7
v\ tirer&pres&_3sv	13	7
v\ dire&cond&3sv	12	6
v\ adorer&pres&_1sv	12	4
v\ regarder&pres&_3sv	12	8
v\ rentrer&pres&_3sv	12	6
v\ connaître&pres&1sv	11	8
v\ lire&pres&1sv	11	6
v\ manger&pres&_1sv	11	6
v\ manger&pres&_3pv	11	5
v\ piocher&pres&_1sv	11	4
v\ tourner&pres&_3sv	11	5
v\ trouver&pres&_1sv	11	7
v\ vas+y	10	5
v\ lire&pres&3sv	10	5
v\ mettre&pres&2sv	10	5
v\ remettre&pres&3sv	10	5
v\ donner&pres&_2sv	10	5
v\ parler&pres&_3sv	10	4
v\ passer&pres&_3sv	10	5
v\ casser&pres&_3sv	9	5
v\ nager&pres&_3sv	9	4

(fin)	T	Enf ≠
v connaître&pres&2sv	8	5
v dire&pres&1sv	8	4
v:* faire&pres&3pv	8	6
v:* savoir&pres&3sv	8	5
v aimer&pres&_2sv	8	4
v arriver&pres&_3sv	8	6
v griffer&pres&_3sv	8	4
v regarder&pres&_1sv	8	6
v voler&pres&_3sv	8	4
v attendre&imp&2pv	7	4
v:* venir&pres&2sv	7	4
v arracher&pres&_3sv	7	5
v donner&pres&_3sv	7	6
v ranger&pres&_1sv	7	6
v:* être&impf&2sv	6	5
v dire&pres&2sv	6	4
v chercher&pres&_1sv	6	4
v:* devoir&pres&3sv	5	4
v entendre&pres&1sv	5	4
v:mdllex venir&pres&3sv	5	4
v crier&pres&_3sv	5	4
v trouver&pres&_3sv	5	4
v appeler&pres&_1sv	4	4
v jeter&pres&_3sv	4	4
v jouer&pres&_2sv	4	4
Total	8902	

BIBLIOGRAPHIE

- Abney, Steven P., 1987, *The English Noun Phrase in its Sentential Aspect*, Doctoral Dissertation, Massachusetts Institute of Technology.
- Adams, Marianne, 1987, "From Old French to the Theory of Pro-Drop", *Natural Language & Linguistic Theory*, Vol. 5, N° 1, p. 1-32.
- Adger, David, 2003, *Core Syntax. A Minimalist Approach*, New York: Oxford University Press.
- Alexiadou, Artemis, 2006, "Left Dislocation (including CLLD)", in *The Blackwell Companion to Syntax*, Vol. 2, Everaert, Martin & Henk van Riemsdijk (eds.), Oxford: Blackwell, p. 668-700.
- Alexiadou, Artemis & Elena Anagnostopoulou, 1998, "Parametrizing Agr: Word Order, V-movement and EPP-checking", *Natural Language & Linguistic Theory*, Vol. 16, p. 491-539.
- Alexiadou, Artemis & Elena Anagnostopoulou, 1999, "EPP Without Spec, IP", in *Specifiers. Minimalist Approaches*, Adger, David, Susan Pintzuk, Bernadette Plunkett & George Tsoulas (eds.), New York: Oxford University Press, p. 93-109.
- Anagnostopoulou, Elena, 2006, "Clitic Doubling", in *The Blackwell Companion to Syntax*, Vol. 1, Everaert, Martin & Henk van Riemsdijk (eds.), Oxford: Blackwell, p. 519-581.
- Aristote, [1994], *De anima*, Paris: Gallimard.
- Ashby, William J., 1982, "The Drift of French Syntax", *Lingua*, Vol. 57, p. 29-46.
- Auger, Julie, 1993, "More Evidence for Verbal Agreement-marking in Colloquial French", in *Linguistic Perspectives on the Romance Languages: Selected Papers from the 21st Linguistic Symposium on Romance Languages*, Ashby, J., M. Mithun, G. Perissinotto & E. Raposo (eds.), Amsterdam: John Benjamins, p. 177-198.
- Auger, Julie, 1994, *Pronominal Clitics in Québec Colloquial French: A Morphological Analysis*, PhD Dissertation, University of Pennsylvania.
- Auroux, Sylvain, 1996, *La philosophie du langage*, Paris: Presses Universitaires de France.
- Ayoun, Dalila, 2003, *Parameter-setting in Language Acquisition*, London & New York: Continuum Press.

- Baker, Mark C., 1985, "The Mirror Principle and Morphosyntactic Explanation", *Linguistic Inquiry*, Vol. 16, p. 373-415.
- Baker, Mark C., 1997, "Thematic Roles and Syntactic Structure", in *Elements of Grammar. Handbook of Generative Grammar*, Haegeman, Liliane (ed.), Dordrecht: Kluwer Academic Publishers, p. 73-137.
- Barra-Jover, Mario, 2004, "Interrogatives, négatives et évolution des traits formels du verbe en français parlé", *Langue française*, Vol. 141, p. 110-125.
- Barra-Jover, Mario, 2007, "Comment évolue un trait grammatical: le pluriel en français dans une perspective romane", *Romance Philology*, Vol. 60, p. 1-46.
- Bauche, Henri, 1928, *Le langage populaire*, Paris: Éditions Payot.
- Baude, Olivier (ed.), 2006, *Corpus Oraux. Guide des bonnes pratiques 2006*, Orléans & Paris: Presses Universitaires d'Orléans & CNRS Éditions.
- Belletti, Adriana, 1990, *Generalized Verb Movement*, Turin: Rosenberg & Sellier.
- Bentolila, Alain, 2007, *La maternelle: au front des inégalités linguistiques et sociales*, Rapport commandé par Xavier Darcos, Ministre de l'Éducation nationale, <http://www.cafepedagogique.net/lexpresso/Documents/docsjoint/Bentomater.pdf>.
- Benveniste, Émile, 1946, "Structure des relations de personne dans le verbe", in *Problèmes de linguistique générale, 1*, Paris: Éditions Gallimard, 1966, p. 225-236.
- Benveniste, Émile, 1956, "La nature des pronoms", in *Problèmes de linguistique générale, 1*, Paris: Éditions Gallimard, 1966, p. 251-257.
- Benveniste, Émile, 1958, "De la subjectivité dans le langage", in *Problèmes de linguistique générale, 1*, Paris: Éditions Gallimard, 1966, p. 258-266.
- Berko, Jean, 1958, "The Child's Learning of English Morphology", *Word*, Vol. 14, p. 150-177; reprinted in Lust, Barbara C. & Claire Foley (eds.), 2004, *First Language Acquisition. The Essential Readings*, Oxford: Blackwell, p. 253-273.
- Bideaud, Jacqueline, Olivier Houdé & Jean-Louis Pardinielli, 1993, *L'homme en développement*, Paris: Presses Universitaires de France.
- Blanche-Benveniste, Claire, 1994, "Quelques caractéristiques grammaticales des "sujets" employés dans le français parlé des conversations", in *Subjecthood and Subjectivity. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory"*, London, 19-20 March 1993, Yaguello, Marina (ed.) p. 77-107.
- Blanche-Benveniste, Claire, 2000, *Approches de la langue parlée en français*, Paris: Éditions Ophrys.
- Blanche-Benveniste, Claire, 2003, "Les formes grammaticales de réalisation des sujets et leur inégale représentation en français contemporain", in *Le sujet*, Merle, Jean-Marie (ed.) Paris: Éditions Ophrys, p. 73-90.
- Bloomfield, Leonard, 1933, *Language*, New York: Holt, Rinehart & Winston.
- Boeckx, Cedric, 2008, *Bare Syntax*, Oxford: Oxford University Press.

- Bonnard, Henri, 1997, *Grammaire française à l'usage de tous*, Paris: Éditions Magnard.
- Borer, Hagit, 1984, *Parametric Syntax: Case Studies in Semitic and Romance Languages*, Dordrecht: Foris.
- Borer, Hagit, 1986, "Introduction", in *The Syntax of Pronominal Clitics*, Borer, Hagit (ed.), New York: Academic Press, p. 1-11.
- Bouchard, Denis, 2001, "La pauvreté des stimuli: quels sont les faits? Trente ans de syntaxe", *Revue québécoise de linguistique*, Vol. 30, N° 1, p. 43-62.
- Bouchard, Denis & Paul Hirschbühler, 1986, "French *Quoi* and its Clitic Allomorph *Que*", in *Studies in Romance Languages*, Neidle, Carol & Rafael Nunez Cedeno (eds.), Dordrecht: Foris, p. 39-60.
- Brandi, Luciana & Patrizia Cordin, 1989, "Two Italian Dialects and the Null Subject Parameter", in *The Null Subject Parameter*, Jaeggli, Osvaldo & Kenneth Safir (eds.), Dordrecht: Kluwer Academic Publishers, p. 111-142.
- Brown, Roger, 1973, *A First Language: The Early Stages*, Cambridge, MA: Harvard University Press.
- Cabredo Hofherr, Patricia, 2003, "Inflected Complementizers and the Licensing of Non-referential Pro-drop", in *The Role of Agreement in Natural Language: TLS 5 Proceedings*, Texas Linguistics Forum, 53, Griffin, William E. (ed.), p. 47-58.
- Cabredo Hofherr, Patricia, 2004, "Les clitiques sujets du français et le paramètre du sujet nul", *Langue française*, Vol. 141, p. 99-109.
- Chevrot, Jean-Pierre, Céline Dugua & Michel Fayol, 2009, "Liaison Acquisition, Word Segmentation and Construction in French: A Usage Based Account", *Journal of Child Language*, Vol. 36, N° 3, p. 557-596.
- Chomsky, Noam, 1957, *Syntactic Structures*, La Haye: Mouton; traduction française *Structures syntaxiques*, 1969, Paris: Éditions du Seuil.
- Chomsky, Noam, 1959, "Review of B.F. Skinner's *Verbal Behavior*", *Language*, Vol. 35, N° 1, p. 26-58; reprinted in Lust, Barbara C. & Claire Foley (eds.), 2004, *First Language Acquisition. The Essential Readings*, Oxford: Blackwell, p. 25-55.
- Chomsky, Noam, 1965, *Aspects of the Theory of Syntax*, Cambridge, MA: The MIT Press; traduction française *Aspects de la théorie syntaxique*, 1971, Paris: Éditions du Seuil.
- Chomsky, Noam, 1966, *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, New York: Harper & Row; traduction française *La Linguistique cartésienne: un chapitre de l'histoire de la pensée rationaliste*, 1969, Paris: Éditions du Seuil.
- Chomsky, Noam, 1968, *Language and Mind*, New York: Harcourt, Brace & World, Inc.
- Chomsky, Noam, 1972, *Studies on Semantics in Generative Grammar*, La Haye: Mouton; traduction française *Questions de sémantique*, 1975, Paris: Éditions du Seuil.

- Chomsky, Noam, 1975, *Reflections on Language*, New York: Pantheon Books; traduction française *Réflexions sur le langage*, 1981, Paris: Flammarion.
- Chomsky, Noam, 1977a, *Essays on Form and Interpretation*, New York: Elsevier North-Holland; traduction française *Essais sur la forme et le sens*, 1980, Paris: Éditions du Seuil.
- Chomsky, Noam, 1977b, "On WH-movement", in *Formal Syntax*, Culicover, Peter W., Thomas Wasow & Adrian Akmajian (eds.), New York: Academic Press, p. 71-132.
- Chomsky, Noam, 1981, *Lectures on Government and Binding: The Pisa Lectures*, Dordrecht: Foris (seventh edition, 1993, The Hague: Mouton de Gruyter); traduction française *Théorie du gouvernement et du liage: les conférences de Pise*, 1991, Paris: Éditions du Seuil.
- Chomsky, Noam, 1982, *Some Concepts and Consequences of the Theory of Government and Binding*, Cambridge, MA: The MIT Press; traduction française *La Nouvelle Syntaxe*, 1987, Paris: Éditions du Seuil.
- Chomsky, Noam, 1986, *Barriers*, Cambridge, MA: The MIT Press.
- Chomsky, Noam, 1995a, "Categories and Transformations", in *The Minimalist Program*, Cambridge, MA: The MIT Press, p. 219-394.
- Chomsky, Noam, 1995b, *The Minimalist Program*, Cambridge, MA: The MIT Press.
- Chomsky, Noam, 1999, "Derivation by Phase", *MIT Working Papers in Linguistics*, Vol. 18; reprinted in *Ken Hale: A Life in Language*, Kenstowicz, Michael (ed.), 2001, Cambridge, MA: The MIT Press, p. 1-52.
- Chomsky, Noam, 2000, "Minimalist Inquiries: The Framework", in *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Martin, Roger, David Michaels & Juan Uriagereka (eds.), Cambridge, MA: The MIT Press, p. 89-155.
- Chomsky, Noam, 2001, "Beyond Explanatory Adequacy", *MIT Occasional Papers in Linguistics*, Vol. 20; reprinted in *Structures and Beyond. The Cartography of Syntactic Structures* (volume 3), Belletti, Adriana (ed.), 2004, Oxford: Oxford University Press, p. 104-131.
- Chomsky, Noam, 2004, "On Phases", in *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud*, Freidin, Robert, Carlos P. Otero & Maria Luisa Zubizarreta (eds.), Cambridge, MA: The MIT Press (2008).
- Chomsky, Noam, 2005, "Three Factors in Language", *Linguistic Inquiry*, Vol. 36, N° 1, p. 1-22.
- Chomsky, Noam, 2007, "Approaching UG from Below", in *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*, Sauerland, Uli & Hans-Martin Gartner (eds.), Berlin & New York: Mouton de Gruyter, p. 1-29.
- Chomsky, Noam, Adriana Belletti & Luigi Rizzi, 2002, "An Interview on Minimalism", in *On Nature and Language*, Chomsky, Noam, Adriana Belletti & Luigi Rizzi (eds.), Cambridge: Cambridge University Press, p. 92-161.
- Cinque, Guglielmo, 1990, *Types of A'-Dependencies*, Cambridge, MA: The MIT Press.

- Cinque, Guglielmo, 1999, *Adverbs and Functional Heads: A Cross-Linguistic Perspective*, New York & Oxford: Oxford University Press.
- Cohen, Marcel, 1924, "Sur les langages successifs de l'enfant", *L'acquisition du langage oral et écrit*, Vol. 31 (1993), p. 33-43.
- Cordemoy, Géraud de, 1666, *Discours physique de la parole*, Paris: Éditions du Seuil (1968).
- Côté, Marie-Hélène, 1999, *Issues in the Analysis and Acquisition of Clitics in (Spoken) French*, Ms., MIT.
- Côté, Marie-Hélène, 2001, "On the Status of Subject Clitics in Child French", in *Proceedings of the 8th Conference of the IASCL*, Almgren, Margareta, Andoni Barrena, Marijose Ezeizabarrena, Itziar Idiazabal & Brian MacWhinney (eds.), Somerville, MA: Cascadilla Press, p. 1314-1330.
- Crain, Stephen & Cecile McKee, 1985, "The Acquisition of Structural Restrictions on Anaphora", in *Proceedings of NELS 16*, Berman, Stephen, Jae-Woong Choe & Joyce McDonough (eds.), Amherst: GLSA, p. 94-110.
- Crain, Stephen & Mineharu Nakayama, 1987, "Structure Dependence in Grammar Formation", *Language*, Vol. 63, N° 3, p. 522-543.
- Creissels, Denis, 1995, *Éléments de syntaxe générale*, Paris: Presses Universitaires de France.
- Crisma, Paola, 1992, "On the Acquisition of Wh-Questions in French", *Geneva Generative Papers*, Vol. 0, N° 1-2, p. 115-122.
- Cunha, Celso & Luis Cintra, 1984, *Nova Gramática do Português Contemporâneo*, Lisboa: Edições João Sá da Costa.
- Dalbera, Jean-Philippe, 1994, *Les Parlers des Alpes-Maritimes. Étude comparative. Essai de reconstruction*, Egham: Association Internationale d'Études Occitanes.
- Dalbera, Jean-Philippe, 2002, "Le corpus entre données, analyse et théorie", *Corpus*, Vol. 1, p. 89-104.
- Dalbera, Jean-Philippe, 2006, *Des dialectes au langage: une archéologie du sens*, Paris: Champion.
- Dalbera-Stefanaggi, Marie-José, 2001, *Essais de linguistique corse*, Ajaccio: Éditions Alain Piazzola.
- De Boysson-Bardies, Bénédicte, 1996, *Comment la parole vient aux enfants. De la naissance jusqu'à deux ans*, Paris: Éditions Odile Jacob.
- De Cat, Cécile, 2002, *French Dislocation*, PhD Thesis, University of York.
- De Cat, Cécile, 2005, "French Subject Clitics are not Agreement Markers", *Lingua*, Vol. 115, N° 9, p. 1195-1219.
- De Cat, Cécile, 2007, *French Dislocation. Interpretation, Syntax, Acquisition*, Oxford & New York: Oxford University Press.
- De Dardel, Robert, 2001, "Quelques réflexions sur l'évolution du protoroman (à propos des pronoms régimes atones)", in *Clitiques et cliticisation. Actes du Colloque de Bordeaux, octobre 1998*, Muller, Claude (ed.), Paris: Honoré Champion, p. 317-326.

- Deal, Amy Rose, 2005, "Pro-drop and Subject (non-)recoverability: The Case of Nez Perce", *Generative Grammar in Geneva*, Vol. 1, p. 93-111.
- Demuth, Katherine, 1996, "Collecting Spontaneous Production Data", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 3-22.
- Denis, Delphine & Anne Sancier-Chateau, 1994, *Grammaire du français*, Paris: Le Livre de Poche.
- Déprez, Viviane & Amy Pierce, 1993, "Negation and Functional Projections in Early Grammar", *Linguistic Inquiry*, Vol. 24, N° 1, p. 25-67.
- Descartes, René, 1637, *Discours de la Méthode pour bien conduire sa raison et chercher la vérité dans les sciences*, Leyde.
- Deshaies, Denise, Christian Guilbault & Claude Paradis, 1993, "Prosodie et dislocation à gauche par anaphore en français québécois spontané", in *Actes du XV^e Congrès International des Linguistes*, Vol. 2, Crochetière, André, Jean-Clause Boulanger & Conrad Ouellon (eds.), Sainte-Foy: Presses de l'Université de Laval, p. 31-34.
- Diderot, Denis, 1751, "Lettre sur les sourds et les muets", in *Oeuvres complètes de Diderot*, vol. 1, Paris: Garnier Frères (1875).
- Dubois, Jean, Mathée Giacomo, Louis Guespin, Christiane Marcellesi, Jean-Baptiste Marcellesi & Jean-Pierre Mével, 1994, *Dictionnaire de linguistique et des sciences du langage*, Paris: Larousse.
- Dubois, Jean, Henri Mitterand & Albert Dauzat, 1964, *Dictionnaire étymologique et historique du français*, Paris: Larousse, 1997.
- Egger, Émile, 1879, *Observations et réflexions sur le développement de l'intelligence et du langage chez les enfants*, Paris: Picard.
- Ezeizabarrena, Maria-José, 2001, "Non-finite Root Sentences: A Language Specific or a Child Language Specific Option?" *Working Papers in Multilingualism*, Vol. 26, N° 1, p. 54-72.
- Ezeizabarrena, Maria-José, 2003, "Null Subjects and Optional Infinitives in Basque", in *(In)vulnerable Domains in Multilingualism*, Müller, Natascha (ed.), Amsterdam: John Benjamins, p. 83-106.
- Ferdinand, Renata Astrid, 1996, *The Development of Functional Categories: The Acquisition of the Subject in French*, La Hague: Holland Academic Graphics.
- Finocchiaro, Chiara, Brigida Fierro, Filippo Brighina, Giuseppe Giglia, Margherita Francolini & Alfonso Caramazza, 2008, "When Nominal Features are Marked on Verbs: A Transcranial Magnetic Stimulation Study", *Brain and Language*, Vol. 104, N° 2, p. 113-121.
- Fodor, Janet Dean, 1998, "Unambiguous Triggers", *Linguistic Inquiry*, Vol. 29, N° 1, p. 1-36.
- Foisil, Madeleine (ed.), 1989, *Journal de Jean Héroard*, Paris: Fayard.
- Fontana, Josep M., 1993, *Phrase Structure and the Syntax of Clitics in the History of Spanish*, PhD Dissertation, University of Pennsylvania.
- Foulet, Lucien, 1930, *Petite syntaxe de l'ancien français*, Paris: Honoré Champion.

- François-Geiger, Denise, 1974, *Français parlé: Analyse des unités phoniques et significatives d'un corpus recueilli dans la région parisienne*, Paris: Société d'études linguistiques et anthropologiques de France.
- Frei, Henri, 1929, *La Grammaire des fautes*, Paris: Geuthner.
- Friedemann, Marc-Ariel, 1993, "The Underlying Position of External Arguments in French: A Study in Adult and Child Grammar", *Language Acquisition*, Vol. 3, N° 3, p. 209-255.
- Friedemann, Marc-Ariel & Luigi Rizzi, 2000, "The Acquisition of Syntax: Introduction", in *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics*, Friedemann, Marc-Ariel & Luigi Rizzi (eds.), London: Longman, p. 1-25.
- Gadet, Françoise, 1990, "Les outils grammaticaux au risque de l'oral non standard", *Travaux de Linguistique*, Vol. 21, p. 13-24.
- Gadet, Françoise, 1998, "Le "français avancé" à l'épreuve de ses données", in *Analyse linguistique et approches de l'oral: recueil d'études offert en hommage à Claire Blanche-Benveniste*, Bilger, Mireille, Karel van den Eynde & Françoise Gadet (eds.), Paris: Peeters, p. 59-68.
- Galilée, 1632, *Dialogue sur les deux plus grands systèmes du monde*, Florence.
- Gary-Prieur, Marie-Noëlle, 1985, *De la grammaire à la linguistique. L'étude de la phrase*, Paris: Armand Colin Éditeur.
- Gerken, LouAnn & Michele E. Shady, 1996, "The Picture Selection Task", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 125-145.
- Gide, André, 1951, *Journal*, Paris: Gallimard.
- Golinkoff, Roberta Michnick, Kathy Hirsh-Pasek, Kathleen M. Cauley & Laura Gordon, 1987, "The Eyes Have It: Lexical and Syntactic Comprehension in A New Paradigm", *Journal of Child Language*, Vol. 14, p. 23-45.
- Godard, Danièle, 1988, *La syntaxe des relatives en français*, Paris: Éditions du CNRS.
- Gordon, Peter, 1996, "The Truth-Value Judgment Task", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 211-232.
- Grégoire, Antoine, 1937, *L'apprentissage du langage: les deux premières années*, Paris & Liège: E. Droz & Faculté de Philosophie et Lettres.
- Grégoire, Antoine, 1947, *L'apprentissage du langage: la troisième année et les années suivantes*, Paris & Liège: E. Droz & Faculté de Philosophie et Lettres.
- Grevisse, Maurice & André Goosse, 2008, *Le bon usage*, Bruxelles: De Boeck & Duculot (14ème édition).
- Gross, Maurice, 1975, *Méthodes en syntaxe*, Paris: Hermann.
- Guasti, Maria Teresa & Ur Shlonsky, 1995, "The Acquisition of French Relative Clauses Reconsidered", *Language Acquisition*, Vol. 4, N° 4, p. 257-276.

- Guillaume, Gustave, 1964, *Langage et science du langage*, Paris & Laval: A.G. Nizet & Presses de l'Université de Laval.
- Haegeman, Liliane, 1997, "Adult Null Subjects in Non Pro-drop Languages", in *The Acquisition of Syntax. Studies in Comparative Developmental Linguistics*, Friedemann, Marc-Ariel & Luigi Rizzi (eds.), London: Longman, 2000, p. 129-169.
- Hagège, Claude, 1996, *L'enfant aux deux langues*, Paris: Éditions Odile Jacob.
- Haider, Hubert, 2001, "Parametrisierung in der Generativen Grammatik", in *Language Typology and Language Universals. An International Handbook*, Vol. 1, Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), Berlin: de Gruyter, p. 283-293.
- Hamann, Cornelia, 2000, "The Acquisition of Constituent Questions and the Requirements of Interpretation", in *The Acquisition of Syntax. Studies in Comparative Developmental Linguistics*, Friedemann, Marc-Ariel & Luigi Rizzi (eds.), London: Longman, p. 170-201.
- Hamann, Cornelia, Luigi Rizzi & Uli Frauenfelder, 1994, "On the Acquisition of the Pronominal System in French", *Geneva Generative Papers*, Vol. 2, N° 2, p. 91-103.
- Hamann, Cornelia, Luigi Rizzi & Uli H. Frauenfelder, 1996, "On the Acquisition of Subject and Object Clitics in French", in *Generative Perspectives on Language Acquisition.*, Vol. 14, Clahsen, Harald (ed.), Amsterdam & Philadelphia: John Benjamins, p. 309-334.
- Han, Chung-hye, 1998, *The Structure and Interpretation of Imperatives: Mood and Force in Universal Grammar*, Dissertation, University of Pennsylvania.
- Harris, James, 1995, "The Morphology of Spanish Clitics", in *Evolution and Revolution in Linguistic Theory*, Campos, Hector & Paula Kempchinsky (eds.), Washington, DC: Georgetown University Press, p. 168-197.
- Heap, David, 2002, "Split Subject Pronoun Paradigms: Feature Geometry and Underspecification", in *Current Issues in Romance Languages: Selected Papers from the 29th Linguistic Symposium on Romance Languages*, Satterfield, Teresa, Cristina Tortora & Diana Cresti (eds.), Amsterdam: John Benjamins, p. 129-144.
- Heggie, Lorie & Francisco Ordóñez, 2005, "Clitic Ordering Phenomena. The Path to Generalizations", in *Clitic and Affix Combinations. Theoretical Perspectives*, Heggie, Lorie & Francisco Ordóñez (eds.), Amsterdam & Philadelphia: John Benjamins, p. 1-29.
- Hérodote, [1930], *Histoires, Livre II, Euterpe*, Paris: Les Belles Lettres; texte établi et traduit par Philippe-Ernest Legrand, 4ème édition, 1963.
- Hirsh-Pasek, Kathy, Roberta Michnick Golinkoff, Gabriella Hermon & Diana Kaufman, 1995, "Evidence from Comprehension for the Early Knowledge of Pronouns", in *The Proceedings of the Twenty-Sixth Annual Child Language Research Forum*, Clark, Eve V. (ed.), Stanford, CA: CSLI Publications, p. 53-59.
- Hirsh-Pasek, Kathy & Roberta Michnick Golinkoff, 1996a, *The Origins of Grammar: Evidence from Early Language Comprehension*, Cambridge, MA: The MIT Press.

- Hirsh-Pasek, Kathy & Roberta Michnick Golinkoff, 1996b, "The Intermodal Preferential Looking Paradigm: A Window onto Emerging Language Comprehension", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 105-124.
- Hornstein, Norbert, Jairo Nunes & Kleanthes K. Grohmann, 2005, *Understanding Minimalism*, New York: Cambridge University Press.
- Huang, C.T. James, 1984, "On the Distribution and Reference of Empty Pronouns", *Linguistic Inquiry*, Vol. 15, p. 531-574.
- Hudelot, Christian, 1980, "Qu'est-ce que la complexité syntaxique? L'exemple de la relative", *La linguistique*, Vol. 16, N° 2, p. 5-41.
- Hugo, Victor, 1856, *Les contemplations*, Paris: Hachette (1922).
- Hulk, Aafke, 1991, "Les pronoms clitiques sujets et la théorie linguistique", in *Actes du XVIIIe Congrès International de Linguistique et de Philologie Romanes*, Tübingen: Niemeyer, p. 504-514.
- Hulk, Aafke, 1996, "The Syntax of Wh-Questions in Child French", *Amsterdam Series in Child Language Development*, Vol. 5, p. 129-172.
- Hulk, Aafke & Jean-Yves Pollock, 2001, "Subject Positions in Romance and the Theory of Universal Grammar", in *Subject Inversion in Romance and the Theory of Universal Grammar*, Hulk, Aafke & Jean-Yves Pollock (eds.), New York: Oxford University Press, p. 3-19.
- Humboldt, Wilhelm von, 1836, *Über die Verschiedenheit des Menschlichen Sprachbaues*, Bonn: F. Dümmlers Verlag, 1960.
- Hyams, Nina M., 1986, *Language Acquisition and the Theory of Parameters*, Dordrecht: D. Reidel Publishing Company.
- Hyams, Nina M., 1992, "A Reanalysis of Null Subjects in Child Language", in *Theoretical Issues in Language Acquisition: Continuity and Change in Development*, Weissenborn, Jürgen, Helen Goodluck & Thomas Roeper (eds.), Hillsdale, New Jersey: Lawrence Erlbaum Associates, p. 249-268.
- Jackendoff, Ray S., 1977, *X' Syntax: A Study of Phrase Structure*, Cambridge, MA: The MIT Press.
- Jacquesson, François, 2008, *Les personnes. Morphosyntaxe et sémantique*, Paris: CNRS Éditions.
- Jaeggli, Osvaldo, 1982, *Topics in Romance Syntax*, Dordrecht: Foris.
- Jaeggli, Osvaldo, 1986, "Three Issues in the Theory of Clitics: Case, Doubled NPs, and Extraction", in *The Syntax of Pronominal Clitics*, Borer, Hagit (ed.) New York: Academic Press, p. 15-42.
- Jakobson, Roman, 1941, "Langage enfantin, aphasie et lois générales de la structure phonique", in *Langage enfantin et aphasie*, Paris: Les Éditions de Minuit, 1969, p. 15-101.
- Jakobson, Roman, 1956, "Deux aspects du langage et deux types d'aphasies", in *Essais de linguistique générale*, Paris: Les Éditions de Minuit, 1963, p. 43-67.

- Jakobson, Roman, 1960, "Pourquoi "papa" et "maman"?" in *Langage enfantin et aphasie*, Paris: Les Éditions de Minuit, 1969, p. 121-130.
- Jakubowicz, Célia, 1996, "Crosslinguistic Investigation", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 257-285.
- Jakubowicz, Célia & Catherine Rigaut, 1997, "L'acquisition des clitiques nominatifs en français", in *Les pronoms: morphologie, syntaxe et typologie*, Zribi-Hertz, Anne (ed.) Saint-Denis: Presses Universitaires de Vincennes, p. 57-99.
- Jakubowicz, Célia & Catherine Rigaut, 2000, "L'Acquisition des clitiques nominatifs et des clitiques objets en français", *Canadian Journal of Linguistics*, Vol. 45, N° 1/2, p. 119-157.
- Jeanjean, Colette, 1980, "L'organisation des formes sujet en français de conversation: Étude quantitative et grammaticale de deux corpus", *Recherches sur le français parlé*, Vol. 3, p. 99-134.
- Jisa, Harriet, 2004, "Growing into Academic French", in *Language Development across Childhood and Adolescence*, Berman, Ruth A. (ed.) Amsterdam & Philadelphia: John Benjamins, p. 135-161.
- Kaiser, Georg A., 1994, "More about INFL-lection and Agreement: The Acquisition of Clitic Pronouns in French", in *Bilingual First Language Acquisition: French and German Grammatical Development*, Vol. 7, Meisel, Jürgen M. (ed.), Amsterdam & Philadelphia: John Benjamins, p. 131-159.
- Kaiser, Georg A., 2002, *Verbstellung und Verbstellungswandel in den romanischen Sprachen*, Vol. 465, Tübingen: Niemeyer.
- Kaiser, Georg A., 2009, "Losing the Null Subject. A Contrastive Study of (Brazilian) Portuguese and (Medieval) French", in *Null-subjects, Expletives, and Locatives in Romance*, Kaiser, Georg A. & Eva-Maria Remberger (eds.), Konstanz: Konstanz Working Papers of the Department of Linguistics, p. 131-156.
- Kaiser, Georg A. & Marc-Olivier Hinzelin, à paraître, "Le paramètre du sujet nul dans les variétés dialectales de l'occitan et du francoprovençal", in *GalRom07*, Scheer, Tobias (ed.), Paris: Presses Universitaires de Vincennes.
- Kayne, Richard S., 1972, "Subject Inversion in French Interrogatives", in *Generative Studies in Romance Languages*, Casagrande, J. & B Saciuk (eds.), Rowley, MA: Newbury House, p. 70-126.
- Kayne, Richard S., 1975, *French Syntax. The Transformational Cycle*, Cambridge, MA: The MIT Press; traduction française *Syntaxe du français. Le cycle transformationnel*, 1977, Paris: Éditions du Seuil.
- Kayne, Richard S., 1976, "French Relative *que*", in *Current Studies in Romance Linguistics*, Lujan, Marta & Fritz Hensey (eds.), Washington, DC: Georgetown University Press, p. 255-299.
- Kayne, Richard S., 1989, "Null Subjects and Clitic Climbing", in *The Null Subject Parameter*, Jaeggli, Osvaldo & Kenneth Safir (eds.), Dordrecht: Kluwer Academic Publishers, p. 239-261.
- Kayne, Richard S., 1994, *The Antisymmetry of Syntax*, Cambridge, MA: The MIT Press.

- Kayne, Richard S., 2000, "A Note on Clitic Doubling in French", in *Parameters and Universals*, Kayne, Richard S. (ed.) Oxford & New York: Oxford University Press, p. 163-186.
- Kayne, Richard S., 2009, "Some Silent First Person Plurals", in *Merging Features. Computation, Interpretation, and Acquisition*, Brucart, José M., Anna Gavarró & Jaume Solà (eds.), Oxford: Oxford University Press, p. 276-293.
- Koopman, Hilda, 2004, *The Syntax of Verbs: from Verb Movement in the Kru Languages to Universal Grammar*, Dordrecht: Foris.
- Koopman, Hilda & Dominique Sportiche, 1991, "The Position of Subjects", *Lingua*, Vol. 85, N° 2-3, p. 211-258.
- La Fontaine, Jean de, 1678, "Les Deux Coqs", in *Fables*, Vol. VII-13, Collinet, Jean-Pierre (ed.), Paris: Gallimard, 1991.
- La Forge, Louis de, 1666, *Traité de l'esprit de l'homme*, Amsterdam.
- Labelle, Marie, 1988, *Prédication et mouvement: le développement de la construction relative chez les enfants francophones*, Thèse de Doctorat, Université d'Ottawa.
- Labelle, Marie, 1990, "Predication, Wh-movement and the Development of Relative Clauses", *Language Acquisition*, Vol. 1, N° 1, p. 95-119.
- Labelle, Marie, 1996, "The Acquisition of Relative Clauses: Movement or No Movement?" *Language Acquisition*, Vol. 5, N° 2, p. 65-82.
- Labelle, Marie, 2000, "Les Infinitifs racines en langage enfantin", *Canadian Journal of Linguistics*, Vol. 45, N° 1/2, p. 159-192.
- Labiche, Eugène, 1844, *Le major Cravachon*, Paris: Flammarion (1990).
- Lagane, René, 1969, "Problèmes de définition. Le sujet", *Langue française*, Vol. 1, N° 1, p. 58-62.
- Lambert, Frédéric, 2001, "Aux sources de l'enclise: réflexions sur les théories d'Apollonios Dyscole", in *Clitiques et cliticisation. Actes du Colloque de Bordeaux, octobre 1998*, Muller, Claude (ed.), Paris: Honoré Champion, p. 17-29.
- Lambrecht, Knud, 1981, *Topic, Antitopic and Verb Agreement in Non-standard French*, Amsterdam: John Benjamins.
- Lambrecht, Knud, 1984, "A Pragmatic Constraint on Lexical Subjects in Spoken French", *Chicago Linguistic Society*, Vol. 20, p. 239-256.
- Lambrecht, Knud, 1987, "On the Status of SVO Sentences in French Discourse", in *Coherence and Grounding in Discourse*, Tomlin, R.S. (ed.) (ed.), Amsterdam: John Benjamins, p. 217-261.
- Lenneberg, Eric H., 1967, *Biological Foundations of Language*, New York: Wiley; partly reprinted in Lust, Barbara C. & Claire Foley (eds.), 2004, *First Language Acquisition. The Essential Readings*, Oxford: Blackwell, p. 103-108.
- Leonard, Laurence B., 1996, "Assessing Morphosyntax in Clinical Settings", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 287-302.

- Lidz, Jeffrey, Sandra Waxman & Jennifer Freedman, 2003, "What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months", *Cognition*, Vol. 89, p. 65-73.
- Lightbown, Patsy, 1977, *Consistency and Variation in the Acquisition of French*, PhD Dissertation, Columbia University.
- Locke, John, 1690, *Essai sur l'entendement humain, Livres I et II*, Paris: Vrin; traduit par Jean-Michel Vienne, 2001.
- Loti, Pierre, 1897, *Ramuntcho*, Paris: Calmann-Lévy.
- Lyons, John, 1970, *Linguistique générale. Introduction à la linguistique théorique*, Paris: Larousse.
- MacWhinney, Brian, 2000a, *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume I: Transcription Format and Programs*, Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, Brian, 2000b, *The CHILDES Project. Tools for Analyzing Talk, Third Edition. Volume II: The Database*, Vol. 2, Mahwah, NJ: Lawrence Erlbaum Associates.
- Manning, Christopher D., 2003, "Probabilistic Syntax", in *Probabilistic Linguistics*, Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.), Cambridge, MA: The MIT Press, p. 289-341.
- Manzini, M. Rita & Leonardo M. Savoia, 1997, "Null Subjects without pro", *UCL Working Papers in Linguistics*, Vol. 9, p. 303-313.
- Manzini, M. Rita & Leonardo M. Savoia, 2002, "Parameters of Subject Inflection in Italian Dialects", in *Subjects, Expletives and the EPP*, Svenonius, Peter (ed.) New York: Oxford University Press, p. 157-199.
- Manzini, M. Rita & Leonardo M. Savoia, 2005, *I dialetti italiani e romanci: morfosintassi generativa*, Torino: Edizioni dell'Orso.
- Martin, Roger, 1999, "Case, the Extended Projection Principle, and Minimalism", in *Working Minimalism*, Epstein, Samuel David & Norbert Hornstein (eds.), Cambridge, MA: The MIT Press, p. 2-25.
- McCloskey, James, 1997, "Subjecthood and subject positions", in *Elements of Grammar: Handbook in Generative Syntax*, Haegeman, Liliane (ed.) Dordrecht: Kluwer, p. 197-235.
- McDaniel, Dana & Helen Smith Cairns, 1996, "Eliciting Judgements of Grammaticality and Reference", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 233-254.
- McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), 1996, *Methods for Assessing Children's Syntax*, Cambridge, MA: The MIT Press.
- Merle, Jean-Marie (ed.), 2003, *Le sujet. Actes du colloque "Le sujet" organisé à l'Université de Provence, les 27 et 28 septembre 2001, avec le concours du CELA (EA851) de l'UFR LAG-LEA et de l'Université de Provence*, Gap & Paris: Éditions Ophrys.

- Miller, Philip H. & Paola Monachesi, 2003, "Les pronoms clitiques dans les langues romanes", in *Langues romanes. Problèmes de la phrase simple*, Godard, Danièle (ed.), Paris: CNRS Éditions, p. 67-123.
- Moignet, Gérard, 1965, *Le pronom personnel français. Essai de psycho-systématique historique*, Paris: Klincksieck.
- Morgenstern, Aliyah, 2006, *Un "je" en construction: genèse de l'auto-désignation chez le jeune enfant*, Paris: Éditions Ophrys.
- Morgenstern, Aliyah & Emmanuelle Prak-Derrington, 2008, "Regards croisés sur la répétition chez Beckett et dans le langage de l'enfant", *Bulletin de la société de stylistique anglaise*, Vol. 30, p. 249-272.
- Muller, Charles, 1979, "Les verbes les plus fréquents du français", in *Langue française et linguistique quantitative. Recueil d'articles*, Muller, Charles (ed.), Genève: Éditions Slatkine, p. 363-370.
- Munaro, Nicola & Jean-Yves Pollock, 2005, "Qu'est-ce-que (*qu*)-est-ce que? A Case Study in Comparative Romance Interrogative Syntax", in *The Oxford Handbook of Comparative Syntax*, Cinque, Guglielmo & Richard S. Kayne (eds.), New York: Oxford University Press, p. 542-606.
- Nash, Léa, 1997, "La partition personnelle dans les langues ergatives", in *Les pronoms. Morphologie, syntaxe et typologie*, Zribi-Hertz, Anne (ed.), Saint-Denis: Presses Universitaires de Vincennes, p. 129-149.
- Neeleman, Ad & Kriszta Szendrői, 2007, "Radical Pro Drop and the Morphology of Pronouns", *Linguistic Inquiry*, Vol. 38, N° 4, p. 671-714.
- Nunes, Jairo, 2004, *Linearization of Chains and Sideward Movement*, Cambridge, MA: the MIT Press.
- Obenauer, Hans-Georg, 1981, "Le principe des catégories vides et la syntaxe des interrogatives complexes", *Langue française*, Vol. 52, N° 1, p. 100-118.
- Obenauer, Hans-Georg, 1990, "De quelques malentendus récurrents concernant la grammaire générative", *Recherches linguistiques de Vincennes*, Vol. 19, p. 73-106.
- Ochs, Elinor, 1979, "Transcription as Theory", in *Developmental Pragmatics*, Ochs, Elinor & Bambi B. Schieffelin (eds.), New York: Academic Press, p. 43-72.
- Oliviéri, Michèle, 1983, *Le parler de Sospel*, Mémoire de Maîtrise, Université de Nice-Sophia Antipolis.
- Oliviéri, Michèle, 1994, *Phénomènes et mécanismes de pronominalisation en français*, Thèse de Doctorat, Université de Nice-Sophia Antipolis.
- Oliviéri, Michèle, 2006, "La micro-variation en syntaxe dialectale", in *La dialectologie hier et aujourd'hui (1906-2006)*, Lyon, 7-9 décembre 2006.
- Oliviéri, Michèle, 2009, "Syntactic Parameters and Reconstruction", in *Null-subjects, Expletives and Locatives in Romance*, Kaiser, Georg A. & Eva-Maria Remberger (eds.), Konstanz: Konstanz Working Papers of the Department of Linguistics.

- Oliviéri, Michèle, à paraître, "From Dialectology to Diachrony", in *Proceedings of the Thirteenth International Conference on Methods in Dialectology, Leeds, 4-8 August 2008*, Lang, Peter (ed.), Bamberg: University of Bamberg.
- Ouhalla, Jamal, 1999, *Introducing Transformational Grammar. From Principles and Parameters to Minimalism*, London: Arnold.
- Palasis, Katérina, 2005, *Problèmes d'acquisition et le Paramètre du Sujet Nul*, Mémoire de Master, Université de Nice-Sophia Antipolis.
- Palasis, Katérina, 2008, "Syntaxe, acquisition et dialectologie: la pluridisciplinarité appliquée à l'étude du sujet nul enfantin", in *Autour des langues et du langage. Perspective pluridisciplinaire*, Lidilem (ed.) Grenoble: Presses Universitaires de Grenoble, p. 241-248.
- Palasis, Katérina, 2009, "On the Variable Morpho-Syntactic Status of the French Subject Clitics", in *Null-subjects, Expletives and Locatives in Romance*, Vol. Arbeitspapier 123, Kaiser, Georg A. & Eva-Maria Remberger (eds.), Konstanz: Konstanz Working Papers of the Department of Linguistics, p. 47-62, <http://ling.uni-konstanz.de/pages/publ/arbeitspapiere.html>.
- Palasis, Katérina & Michèle Oliviéri, 2007, "Beyond Deficiency. Reconsidering the Null Subject Phenomenon", in *The Acquisition of Romance Languages. Selected Papers from The Romance Turn II 2006*, Vol. LOT Occasional Series 8, Baauw, Sergio, Jacqueline Van Kampen & Manuela Pinto (eds.), Utrecht: LOT Publications, p. 127-143.
- Passy, Paul-Edouard, 1890, *Étude sur les changements phonétiques et leurs caractères généraux*, Paris: Didot.
- Pavlov, Ivan P., 1927, *Conditioned Reflexes*, London: Routledge.
- Pei, Mario A., 1932, *The Language of the VIIIth Century Texts in Northern France*, New York: Couranza.
- Perez-Leroux, Ana Teresa, 1995, "Resumptives in the Acquisition of Relative Clauses", *Language Acquisition*, Vol. 4, N° 1, p. 105-138.
- Perlmutter, David M., 1971, *Deep and Surface Structure Constraints in Syntax*, New York: Holt, Rinehart & Winston.
- Philippaki-Warburton, Irene, 1977, "Modern Greek Clitic Pronouns and the 'Surface Structure Constraints' Hypothesis", *Journal of Linguistics*, Vol. 13, p. 259-281.
- Philippaki-Warburton, Irene, 1987, "The Theory of Empty Categories and the Pro-Drop Parameter in Modern Greek", *Journal of Linguistics*, Vol. 23, N° 2, p. 289-318.
- Piaget, Jean, 1923, *Le langage et la pensée chez l'enfant. Études sur la logique de l'enfant*, Neuchâtel: Delachaux & Niestlé.
- Piaget, Jean & Bärbel Inhelder, 1969, "Les images mentales", in *Traité de psychologie expérimentale*, Fraisse, Paul & Jean Piaget (eds.), Paris: Presses Universitaires de France.
- Pierce, Amy, 1992, *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*, Vol. 14, Dordrecht: Kluwer Academic Publishers.

- Pinker, Steven, 1994, *L'Instinct du langage*, Paris: Éditions Odile Jacob, 1999.
- Platon, [1920-1964], "Le Sophiste", in *Oeuvres complètes*, Vol. 8, Paris: Les Belles Lettres; traduction de Diès.
- Platon, [1999], *Menon*, Paris: Le Livre de Poche; traduction de Jacques Cazeaux.
- Platzack, Christer & Inger Rosengren, 1998, "On the Subject of Imperatives: A Minimalist Account of the Imperative Clause", *The Journal of Comparative Germanic Linguistics*, Vol. 1, N° 3, p. 177-224.
- Plunkett, Kim, 1995, "Connectionist Approaches to Language Acquisition", in *The Handbook of Child Language*, Fletcher, Paul J. & Brian MacWhinney (eds.) (eds.), Oxford: Blackwell, p. 36-72.
- Poletto, Cecilia, 2000, *The Higher Functional Field: Evidence from Northern Italian Dialects*, New York & Oxford: Oxford University Press.
- Pollock, Jean-Yves, 1989, "Verb Movement, Universal Grammar and the Structure of IP", *Linguistic Inquiry*, Vol. 20, N° 3, p. 365-424.
- Pollock, Jean-Yves, 1997, *Langage et cognition. Introduction au programme minimaliste de la grammaire générative*, Paris: Presses Universitaires de France.
- Popper, Karl R., 1935, *Logik der Forschung*, Vienne: Springer; traduction française *La logique de la découverte scientifique*, 1973, Paris: Éditions Payot.
- Radford, Andrew, 1997, *Syntax: A Minimalist Introduction*, Cambridge: Cambridge University Press.
- Rasetti, Lucienne, 1996, "Null Subjects and Root Infinitives in the Child Grammar of French", *Geneva Generative Papers*, Vol. 4, N° 2, p. 120-132.
- Rasetti, Lucienne, 2003, *Optional Categories in Early French Syntax: A Developmental Study of Root Infinitives and Null Arguments*, Thèse de Doctorat, Université de Genève.
- Renzi, Lorenzo & Laura Vanelli, 1982, "I pronomi soggetto in alcune varietà romanze", in *Studi linguistici in onore di Giovan Battista Pellegrini*, Pisa: Pacini, p. 121-145.
- Rivas, Alberto M., 1977, *A theory of Clitics*, PhD Dissertation, Massachusetts Institute of Technology.
- Rizzi, Luigi, 1982, *Issues in Italian Syntax*, Dordrecht: Foris.
- Rizzi, Luigi, 1986a, "On the Status of Subject Clitics in Romance", in *Studies in Romance Linguistics*, Vol. 24, Jaeggli, Osvaldo & Carmen Silva-Corvalan (eds.), Dordrecht: Foris, p. 391-419.
- Rizzi, Luigi, 1986b, "Null Objects in Italian and the Theory of *pro*", *Linguistic Inquiry*, Vol. 17, p. 501-557.
- Rizzi, Luigi, 1990, *Relativized Minimality*, Cambridge, MA: The MIT Press.
- Rizzi, Luigi, 1992a, "Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives", *Geneva Generative Papers*, Vol. 0, N° 2, p. 102-114.

- Rizzi, Luigi, 1992b, "Early Null Subjects and Root Null Subjects", in *Comparative Syntax and Language Acquisition*, Rizzi, Luigi (ed.), London & New York: Routledge, 2000, p. 299-339.
- Rizzi, Luigi, 1997a, "The Fine Structure of the Left Periphery", in *Comparative Syntax and Language Acquisition*, Rizzi, Luigi (ed.) London & New York: Routledge, 2000, p. 241-295.
- Rizzi, Luigi, 1997b, "A Parametric Approach to Comparative Syntax: Properties of the Pronominal System", in *The New Comparative Syntax*, Haegeman, Liliane (ed.) Harlow: Addison Wesley Longman, p. 268-285.
- Rizzi, Luigi, 2001, "Relativized Minimality Effects", in *The Handbook of Syntactic Theory*, Baltin, Mark & Chris Collins (eds.), Oxford: Basil Blackwell, p. 89-100.
- Rizzi, Luigi, 2002, *On the Grammatical Basis of Language Development: A Case Study*, Ms., University of Siena.
- Rizzi, Luigi & Ur Schlonsky, 2005, *Strategies for Subject Extraction*, Manuscript, http://www.ciscl.unisi.it/doc/doc_pub/rizzi_shlonsky2005.pdf.
- Robert, Paul, Josette Rey-Debove & Alain Rey, 2002, *Le Nouveau Petit Robert: dictionnaire alphabétique et analogique de la langue française*, Paris: Dictionnaires Le Robert.
- Roberts, Ian, 2007, "Introduction", in *Comparative Grammar. Critical Concepts in Linguistics. Volume II: the Null-Subject Parameter*, Roberts, Ian (ed.) London & New York: Routledge, p. 1-44.
- Roché, Michel, 1992, "Le masculin est-il plus productif que le féminin?" *Langue française*, Vol. 96, p. 113-124.
- Rossi, Mario, 1999, *L'intonation, le système du français: description et modélisation*, Paris: Éditions Ophrys.
- Rowlett, Paul, 1998, *Sentential Negation in French*, Oxford: Oxford University Press.
- Rumelhart, David E. & James L. McClelland, 1986, "On Learning the Past Tense of English Verbs", in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Psychological and Biological Models*, Vol. 2, McClelland, James L., David E. Rumelhart & the PDP Research Group (eds.), Cambridge, MA: Bradford Books, the MIT Press, p. 216-271.
- Sartre, Jean-Paul, 1964, *Les mots*, Paris: Gallimard.
- Saussure, Ferdinand de, 1916, *Cours de linguistique générale*, Paris: Éditions Payot & Rivages, 1995.
- Schapansky, Nathalie, 2009, "Further Aspects of Negation in French", *Lingua*, Vol. in press, 10.1016/j.lingua.2009.05.002.
- Scheer, Tobias, à paraître, *A Lateral Theory of Phonology. Volume 2: How Morphosyntax Talks to Phonology. A Survey of Extra-phonological Information in Phonology since Trubertzkoy's Grenzsignale*, Berlin: Mouton de Gruyter.
- Schlegel, August W., 1801, *Die Kunstlehre*, Stuttgart: W. Kohlhammer Verlag, 1963.

- Shannon, Claude E., 1949, *The Mathematical Theory of Information*, Urbana: University of Illinois Press.
- Skinner, Burrhus F., 1938, *The Behavior of Organisms*, New York: Appleton-Century-Crofts, Inc.
- Skinner, Burrhus F., 1953, *Science and Human Behavior*, London: Macmillan Company.
- Skinner, Burrhus F., 1957, *Verbal Behavior*, New York: Appleton-Century-Crofts, Inc.
- Smith, Neil & Ianthi Maria Tsimpli, 1991, "Linguistic Modularity? A Case Study of a 'Savant' Linguist", *Lingua*, Vol. 84, p. 315-351.
- Snyder, William, 2007, *Child Language: The Parametric Approach*, New York: Oxford University Press.
- Spelke, Elizabeth, 1979, "Perceiving bimodally specified events in infancy", *Developmental Psychology*, Vol. 15, p. 626-636.
- Spitz, René A., 1962, *Le non et le oui: la genèse de la communication humaine*, Paris: Presses Universitaires de France, 1994.
- Sportiche, Dominique, 1996, "Clitic Constructions", in *Phrase Structure and the Lexicon*, Rooryck, Johan & Laurie Zaring (eds.), Dordrecht: Kluwer Academic Publishers, p. 213-276.
- Sportiche, Dominique, 1999, "Subject Clitics in French and Romance, Complex Inversion and Clitic Doubling", in *Beyond Principles and Parameters: Essays in Memory of Osvaldo Jaeggli*, Johnson, Kyle B. & Ian Roberts (eds.), Dordrecht: Kluwer, p. 189-222.
- Strick, Nelleke, 2007, "L'acquisition des phrases interrogatives chez les enfants francophones", *Psychologie française*, Vol. 52, N° 1, p. 27-39.
- Stromswold, Karin, 1996, "Analyzing Children's Spontaneous Speech", in *Methods for Assessing Children's Syntax*, McDaniel, Dana, Cecile McKee & Helen Smith Cairns (eds.), Cambridge, MA: The MIT Press, p. 23-53.
- Strozer, Judith Reina, 1976, *Clitics in Spanish*, PhD Dissertation, UCLA.
- Suñer, Margarita, 1998, "Resumptive Restrictive Relatives: A Crosslinguistic Perspective", *Language*, Vol. 74, N° 2, p. 335-364.
- Suppes, Patrick, Robert Smith & Madeleine Léveillé, 1973, "The French Syntax of a Child's Noun Phrases", *Archives de Psychologie*, Vol. 42, p. 207-269.
- Svenonius, Peter, 2002, "Introduction", in *Subjects, Expletives, and the EPP*, Svenonius, Peter (ed.) New York: Oxford University Press, p. 3-27.
- Taraldsen, Knut Tarald, 2001, "Subject Extraction, the Distribution of Expletives, and Stylistic Inversion", in *Subject Inversion in Romance and the Theory of Universal Grammar*, Hulk, Aafke & Jean-Yves Pollock (eds.), New York: Oxford University Press, p. 163-182.
- Tinbergen, Nikolaas, 1951, *The Study of Instinct*, Oxford: Clarendon Press.
- Tomasello, Michael, 2000, "Do Young Children Have Adult Syntactic Competence?" *Cognition*, Vol. 74, p. 209-253.

- Tomasello, Michael, 2003, *Constructing a Language. A Usage-Based Theory of Language Acquisition*, Cambridge, MA & London: Harvard University Press.
- Tomasello, Michael, 2006, "Acquiring Linguistic Constructions", in *Handbook of Child Psychology: Cognition, Perception and Language (Volume 2)*, Kuhn, Deanna & Robert Siegler (eds.), Hoboken, NJ: John Wiley & Sons, p. 255-298.
- Touratier, Christian, 2003, "Sujet et syntaxe?" in *Le sujet*, Merle, Jean-Marie (ed.), Paris: Éditions Ophrys, p. 41-50.
- Veneziano, Edy, 2000, "Interaction, conversation et acquisition du langage dans les trois premières années", in *L'acquisition du langage. Le langage en émergence. De la naissance à trois ans.*, Vol. 1, Kail, Michèle & Michel Fayol (eds.), Paris: Presses Universitaires de France, p. 231-265.
- Vendryes, Joseph, 1920, *Le langage. Introduction linguistique à l'histoire*, Paris: Éditions Albin Michel.
- Vergnaud, Jean-Roger, 1973, *French Relative Clauses*, Thesis, MIT.
- Vigny, Alfred de, 1835, *Servitude et grandeur militaires*, Paris: Garnier (1965).
- Vygotski, Lev S., 1934, *Pensée et langage*, Paris: Éditions sociales, 1984.
- Wagner, Robert & Jacqueline Pinchon, 1962, *Grammaire du français*, Paris: Hachette.
- Watson, John B., 1924, *Behaviorism*, London: Kegan, Paul, Trench, Trubner & Co.
- Watson, John B., 1928, *Psychological Care of the Infant and Child*, New York: Allen & Unwin.
- Wexler, Ken, 1998, "Very Early Parameter Setting and the Unique Checking Constraint: A New Explanation of the Optional Infinitive Stage", *Lingua*, Vol. 106, N° 1-4, p. 23-79.
- Williams, Edwin S., 1980, "Predication", *Linguistic Inquiry*, Vol. 11, N° 1, p. 203-238.
- Yaguello, Marina (ed.), 1994, *Subjecthood and Subjectivity. The Status of the Subject in Linguistic Theory. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory", London, 19-20 March 1993*, Gap, Paris & Londres: Éditions Ophrys & Institut français du Royaume-Uni.
- Zanuttini, Raffaella, 1997, *Negation and Clausal Structure. A Comparative Study of Romance Languages*, New York: Oxford University Press.
- Zazzo, René, 1993, *Reflets de miroir et autres doubles*, Paris: Presses Universitaires de France.
- Zimmermann, Michael, 2009, "On the Evolution of Expletive Subject Pronouns in Old French", in *Null-subjects, Expletives and Locatives in Romance*, Vol. Arbeitspapier 123, Kaiser, Georg A. & Eva-Maria Remberger (eds.), Konstanz: Konstanz Working Papers of the Department of Linguistics, p. 63-92.
- Zribi-Hertz, Anne, 1994, "La syntaxe des clitiques nominatifs en français standard et en français avancé", *Travaux de Linguistique et de Philologie*, Vol. 32, p. 131-148.

Zribi-Hertz, Anne, 2006, *Français standard et français dialectal: pour une reconnaissance générale du phénomène diglossique pour la description et l'enseignement du français*, Manuscrit.

Zwicky, Arnold M. & Geoffrey K. Pullum, 1983, "Cliticization vs. Inflection: English N'T", *Language*, Vol. 59, N° 3, p. 502-513.