UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA MODÉLISATION DES MARQUEURS D'ARGUMENTS DE L'ARABE STANDARD DANS LE CADRE DES GRAMMAIRES À BASE DE CONTRAINTES

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN LINGUISTIQUE

PAR
ADEL JEBALI

JANVIER 2009

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

AVANT-PROPOS

À mes parents, qui ont cru en moi et qui m'ont poussé à devenir ce que je suis. Ils sont restés loin, dans notre Tunisie natale, mais ils étaient toujours présents.

À mes filles, Leïla et Kamila. Le bonheur de vous avoir efface toute peine.

À ma femme, Andréanne. Merci pour tout.

REMERCIEMENTS

J'aimerais avant tout remercier ma directrice de recherche, Louisette Emirkanian, sans laquelle cette aventure n'aurait pu aboutir. Louisette a su m'accorder l'autonomie dont j'avais besoin. Durant ces années de travail, j'ai apprécié en elle une personne avec beaucoup d'écoute, d'entraide et de sympathie. Sans ces qualités et sans sa présence continue pour me diriger tant sur le plan scientifique que professionnel, cette thèse n'aurait pas vu le jour. Je ne saurais assez la remercier.

Je remercie également ma compagne de route, la femme qui a su me soutenir et supporter le fardeau familial quand j'avais besoin de temps et de concentration pour travailler sur ma thèse. Ce n'était pas facile par moments, mais elle était toujours là prête à tout prendre en charge avec beaucoup d'amour et de compréhension. Merci Andréanne.

D'autres personnes ont grandement influencé mon parcours universitaire tant en Tunisie qu'au Québec. Ici, au Québec, je nomme Mohamed Guerssel, avec lequel j'ai eu des discussions linguistiques et non linguistiques fort enrichissantes, John Lumsden, qui m'a poussé, grâce à ces remarques, à poser des questions auxquelles je n'avais pas pensé et Isamail Biskri, qui a fait des remarques fort pertinentes sur des versions antérieures de cette thèse. Merci d'avoir accepté de faire partie de mon jury de thèse.

En Tunisie, mon directeur de DEA, Mohamed Slaheddine Chérif, m'a initié à la linguistique alors que je n'étais qu'un étudiant de langue arabe et depuis cette première piqûre je n'ai pas cessé d'être fasciné par cette science.

Je remercie finalement mes amis et collègues les étudiants de doctorat en linguistique à l'UQAM. Nos discussions, nos sorties, nos blagues, nos soucis partagés, et parfois même nos déceptions, tout cela me manquera cruellement.

TABLE DES MATIÈRES

Liste des figures ix

Liste des tableaux xi

liste des abréviations, sigles et acronymes xiii

Résumé xv

Introduction 1

CHAPITRE I
CADRE THÉORIQUE 6

1.1 Le choix du cadre 7

1.2 Les grammaires à base de contraintes 7

1.3 La grammaire syntagmatique endocentrique HPSG 8

1.3.1 Les traits distinctifs de la grammaire 9

1.3.1.1 Une théorie à base de contraintes 9

1.3.1.2 Une orientation surfaciste 10

1.3.1.3 Une théorie éclectique 10

1.3.1.4 Une théorie lexicaliste au sens fort 10

1.3.2 Les traits saillants du formalisme 11

1.3.3 Types, contraintes et hiérarchies 13

1.3.4 Les structures mathématiques : les structures de traits 15

1.3.4.1 Définitions 16

1.3.4.2 Les contraintes sur les structures de traits 17

1.3.4.3 Les opérations sur les structures de traits 19

1.3.5 Les principes de bonne formation 21

1.3.5.1 Le principe de la réalisation des arguments 21

1.3.5.2 Le principe des traits de tête 22

1.3.5.3 Le principe de la valence 23

1.3.5.4 Le principe généralisé des traits de tête 25

1.3.6 Les schémas de dominance immédiate 26

1.3.6.1 Schéma DI1 27

1.3.6.2 Schéma DI2 27

1.3.6.3 Schéma DI3 29
1.3.6.4 Schéma DI4 30
1.3.6.5 Schéma DI5 31
1.3.6.6 Schéma DI6 32
1.3.7 Les règles lexicales 34
1.4 La modélisation des clitiques 37
1.4.1 Une analyse lexicaliste au sens fort 38
1.4.1.1 Une analyse basée sur les règles lexicales 39
1.4.1.2 Une analyse basée sur les contraintes 40
1.4.2 Une mise en question de l'hypothèse affixale 51
1.5 HPSG et l'arabe 52
1.5.1 Une analyse des pronoms résomptifs de l'arabe 53
1.5.2 Une modélisation de la morphologie de l'arabe 54
1.6 Conclusion 55

CHAPITRE II
LE SYSTÈME PRONOMINAL DE L'ARABE 57
2.1 Sources des données 58
2.2 Une brève présentation 59
2.2.1 Les pronoms indépendants 59
2.2.2 Les marqueurs d'arguments 60
2.2.2.1 Les marqueurs d'objet 61
2.2.2.2 Les marqueurs de sujet 62
2.2.2.3 La morphologie de l'imperfectif 70
2.2.3 Les formes retenues 75
2.3 Les pronoms indépendants 76
2.3.1 Les pronoms indépendants nominatifs 76
2.3.1.1 Dans la phrase verbale 76
2.3.1.2 Dans la phrase équative 77
2.3.2 Les pronoms indépendants accusatifs 79
2.4 Les marqueurs d'arguments 80
2.4.1 Les symétries 80
2.4.2 Les asymétries 83
2.4.2.1 L'asymétrie des hôtes 83

2.4.2.2 La morphologie des marqueurs d'arguments 84
2.4.2.3 Marqueurs d'arguments et marqueurs d'accord 86
2.4.2.4 La corroboration par des réfléchis 87
2.5 Relations entre pronoms indépendants et marqueurs d'arguments 89
2.5.1 La morphophonologie des pronoms 89
2.5.2 La syntaxe 91
2.5.2.1 Marqueurs d'objet et pronoms indépendants 91
2.5.2.2 Marqueurs d'objet et SN 92
2.6 Conclusion 95

CHAPITRE III
LE STATUT MORPHOSYNTAXIQUE DES MARQUEURS D'ARGUMENTS 97
3.1 Définitions 100
3.1.1 Arguments et marqueurs d'accord 100
3.1.2 Redoublement, dislocation et topicalisation 100
3.2 Les marqueurs d'objet 103
3.2.1 Les compléments du verbe 104
3.2.1.1 Le redoublement 104
3.2.1.2 La dislocation 105
3.2.2 Les compléments de nominaux et de prépositions 105
3.2.2.1 Le redoublement 106
3.2.2.2 La dislocation 106
3.2.3 Les compléments de têtes fonctionnelles 107
3.3 Les marqueurs de sujet 109
3.3.1 L'hypothèse flexionnelle 110
3.3.2 L'hypothèse argumentale 113
3.3.3 La mixité des marqueurs de sujet 115
3.3.3.1 Relations d'accord et ambiguïté fonctionnelle 116
3.3.3.2 Le statut du référent 120
3.4 Les implications 124
3.4.1 Les marqueurs de sujet et l'aspect 125
3.4.2 L'accord partiel 127
3.4.3 Les asymétries de l'accord 132
3.4.3.1 Les données de l'arabe standard 133

3.4.3.2 Les analyses syntaxiques 138
3.4.3.3 Les analyses lexicalistes 144
3.4.3.4 Éléments pour une solution 149
3.5 Conclusion 153

CHAPITRE IV
LE STATUT MORPHOLOGIQUE DES MARQUEURS D'ARGUMENTS . . 155
4.1 Les tests du statut morphologique 157
4.1.1 Les affixes, les clitiques et les mots 157
4.1.1.1 Les affixes et les mots 157
4.1.1.2 Les affixes et les clitiques 161
4.1.1.3 Les mots et les clitiques 164
4.1.2 Clitiques et affixes syntagmatiques 169
4.1.3 Les tests de Miller 171
4.2 Le statut morphologique des marqueurs d'arguments 174
4.2.1 Prémisses 174
4.2.2 Les marqueurs d'arguments et les mots 175
4.2.2.1 L'ordre rigide 176
4.2.2.2 L'attachement 177
4.2.2.3 La prosodie 177
4.2.3 Le statut des marqueurs d'arguments 179
4.2.3.1 Les idiosyncrasies et les trous arbitraires 179
4.2.3.2 La coordination 184
4.2.3.3 L'appartenance catégorielle 186
4.3 Conclusion 187

CHAPITRE V
LA MODÉLISATION DES MARQUEURS D'ARGUMENTS 189
5.1 Modélisation 190
5.1.1 Les types syntagmatiques 190
5.1.2 L'ordre des mots en arabe 191
5.1.2.1 L'ordre VSO 192
5.1.2.2 L'ordre dit SVO 193
5.1.3 Les marqueurs d'objet 197
5.1.3.1 Rappel des données 197

5.1.3.2 La hiérarchie des types 198
5.1.3.3 Les marqueurs d'objet argumentaux 199
5.1.3.4 Les marqueurs d'objet non argumentaux 200
5.1.3.5 Les suites de clitiques 203
5.1.4 Les marqueurs de sujet 207
5.1.4.1 Rappel des données 207
5.1.4.2 La morphophonologie des verbes de l'arabe 207
5.1.4.3 La morphosyntaxe et l'insertion des marqueurs de sujet . . 216
5.1.4.4 L'accord entre le sujet et le verbe 224
5.2 Implémentation 229
5.2.1 Pourquoi LKB ? 230
5.2.2 Introduction à LKB 231
5.2.3 Grammaire 232
5.2.3.1 Les modules de base 232
5.2.3.2 Le fonctionnement de l'environnement 235
5.2.3.3 Notre grammaire 247
5.2.3.4 Les résultats 284
5.3 Conclusion 297

Conclusion **299**

Annexes **308**

Exemples cités **308**

Phrases soumises à l'analyseur **333**

RÉFÉRENCES **337**

LISTE DES FIGURES

1.1 Une hiérarchie partielle des types 14

1.2 Une représentation arborescente de *Michel donne une pomme à Marie* 25

1.3 Le schéma de DI 1 28

1.4 Le schéma DI 2 29

1.5 Le schéma DI 3 30

1.6 Le schéma DI 4 31

1.7 Le schéma DI 5 32

1.8 Le schéma DI 6 33

1.9 Une hiérarchie partielle des types proposée par Miller et Sag (1997) 42

1.10 La hiérarchie (partielle) des types lexicaux du polonais proposée par Kupsc (2000) 51

5.1 Une hiérarchie des types syntagmatiques 191

5.2 Une hiérarchie partielle des types pronominaux 198

5.3 La hiérarchie des types pronominaux (version révisée 1) 202

5.4 La hiérarchie des marqueurs d'objet 203

5.5 Une hiérarchie de types pour les verbes de l'arabe 219

5.6 Une hiérarchie partielle des types proposée par Miller et Sag (1997) 221

5.7 L'interface *trollet* 236

5.8 La hiérarchie des types telle que représentée par LKB 237

5.9 La définition du type *syntagme* 238

5.10 La définition étendue du type *syntagme* 238

5.11 Le visionnement de l'entrée lexicale du verbe *naam* 240

5.12 Le résultat de l'application des règles flexionnelles au radical *naam* 241

5.13 La représentation de la *regle-verbe-initial-2-complements* 242

5.14 Résultat de l'analyse de la phrase en (5.55) 243

5.15 L'échec de l'analyse de la phrase en (5.56) 244

5.16 Le *chart parser* lors d'un échec d'analyse 245

5.17 Le *chart parser* lors d'une analyse réussie 246

5.18 Un exemple de génération de phrases 247

5.19 L'analyse réussie d'une structure dite SVO 286

5.20 Échec de l'analyse d'une structure dite SVO mal formée (pas d'accord en nombre) 286

5.21 Échec de l'analyse d'une structure dite SVO mal formée (pas d'accord en genre) . 287

5.22 L'analyse d'une structure VS avec un sujet pluriel (type 5.58a) 288

5.23 Génération à partir d'une structure VS (type 5.58a) 288

5.24 Échec de l'analyse d'une structure VS (type 5.58b) 289

5.25 Analyse des structures sans sujet lexical 290

5.26 Génération de structures sans sujet lexical (une partie seulement est montrée) . . . 291

5.27 L'analyse assignée à la structure en (5.60a) 292

5.28 L'analyse assignée à la structure en (5.60b) 293

5.29 L'analyse assignée à la structure en (5.60c) 293

5.30 L'échec de l'analyse de la structure en (5.60d) 294

5.31 L'analyse d'une structure contenant un marqueur d'objet complément d'un nominal 294

5.32 L'analyse d'une structure contenant un marqueur d'objet complément de préposition 295

5.33 L'analyse assignée à la structure en (5.63a) 296

5.34 L'analyse assignée à la structure en (5.63b) 296

5.35 L'analyse assignée à la structure en (5.63c) 296

LISTE DES TABLEAUX

1.1 Les déclarations des traits et les contraintes sur les types (version simplifiée) 15

2.1 Les pronoms indépendants (nominatifs et accusatifs) 60

2.2 Les marqueurs d'objet 62

2.3 La conjugaison perfective du verbe *katab* « écrire » 65

2.4 L'imperfectif indicatif du verbe *katab* « écrire » 67

2.5 L'imperfectif subjonctif du verbe *katab* « écrire » 67

2.6 L'imperfectif jussif du verbe *katab* « écrire » 68

2.7 L'impératif du verbe *katab* « écrire » 69

2.8 L'injonctif du verbe *katab* « écrire » 69

2.9 Le prohibitif du verbe *katab* « écrire » 69

2.10 Les marqueurs de sujet à l'imperfectif 70

2.11 Les pronoms indépendants (nominatifs et accusatifs) 75

2.12 Les marqueurs de sujet 75

2.13 Les marqueurs d'objet 75

2.14 L'alternance M/F dans quelques marqueurs de sujet 87

2.15 Les pronoms indépendants ACC et les marqueurs d'objet 89

2.16 Les pronoms indépendants NOM et les marqueurs de sujet 90

2.17 Les formes protosémitiques (Gray, 1934) 90

3.1 La divergence entre la morphologie et la morphosyntaxe 99

3.2 Les pronoms conjoints nominatifs dans la tradition grammaticale 114

3.3 Les marqueurs de sujet pronominaux 115

3.4 Les marqueurs d'accord verbaux 116

3.5 Les marqueurs de sujet 125

3.6 L'accord des nombres entre 3 et 9 131

3.7 L'accord des nombres entre 13 et 19 131

3.8 Les asymétries et la variation dialectale 140

3.9 La morphosyntaxe des marqueurs d'arguments 153

4.1 Les clitiques selon Zwicky (1977) 160

4.2 Les types de clitiques selon Klavans 170

4.3 Le statut des marqueurs d'arguments 188

5.1 Les schèmes morphologiques des verbes trilitères et quadrilatères 209

1 Le résultat de l'analyse des marqueurs d'arguments 303

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

NOM nominatif

NN non nominatif

ACC accusatif

GÉN génitif

1 1re personne

2 2e personne

3 3e personne

M masculin

F féminin

S singulier

D duel

P pluriel

PER perfectif

IMP imperfectif

IMPÉ impératif

JUSS jussif

SUBJ subjonctif

Indé indéfini

TRANSCRIPTION DE L'ALPHABET ARABE

Lettre arabe	Transcription	Lettre arabe	Transcription
أ	ʔ	م	m
ب	b	ن	n
ت / ة	t	ه	h
ث	θ	و	w
ج	ʤ	ي	j
ح	ħ	اَ	a
خ	x	اِ	i
د	d	اُ	u
ذ	ð	ا / ى	a:
ر	r	اِي	i:
ز	z	اُو	u:
س	s		
ش	ʃ		
ص	s'		
ض	ḍ		
ط	t'		
ظ	ḍ		
ع	ʕ		
غ	ɣ		
ف	f		
ق	q		
ك	k		
گ	g		
ل	l		

RÉSUMÉ

Les marqueurs d'arguments de l'arabe standard possèdent plusieurs propriétés qui nous permettent de les traiter comme des pronoms. Malgré cette étiquette qui explique les similarités tant distributionnelles que morphologiques entre ces unités, les marqueurs de sujet se distinguent nettement des marqueurs d'objet sur plusieurs plans. Ces différences, qui se reflètent dans les différents niveaux d'analyse, sont reliées aux différences dans le statut morphologique. En fait, si les marqueurs de sujet sont des affixes, les marqueurs d'objet sont des clitiques. Sur un plan purement morphosyntaxique, par contre, les deux sortes d'unités peuvent être des arguments de la tête qui est leur hôte phonologique, mais peuvent également être des non-arguments. Les marqueurs de sujet, plus précisément ceux de la troisième personne du singulier, jouent le rôle de marqueurs d'accord dans certains contextes. Les marqueurs d'objet de troisième personne du singulier peuvent, pour leur part, jouer le rôle d'explétifs ou de thèmes lorsqu'ils sont attachés à certaines têtes fonctionnelles.

L'analyse de ces unités dans le cadre de la grammaire syntagmatique endocentrique HPSG nous a permis de consolider cette théorie basée sur les contraintes en démontrant que son application sur une langue sémitique ne nécessite pas l'adoption de solutions ad hoc et n'exige pas de modifications notables au formalisme et à la théorie. Cela prouve la flexibilité des grammaires à base de contraintes et leur puissance formelle, qui en font des théories linguistiques capables d'exprimer des analyses linguistiques provenant de plusieurs familles de langues.

Nous avons implémenté cette analyse linguistique dans une grammaire LKB pour en tester la validité et nous avons ainsi pu démonter que l'analyse lexicaliste basée sur les structures de traits typées et la hiérarchie de types à héritage multiple sont non seulement possibles mais également souhaitables.

MOTS CLÉS : marqueurs d'arguments, arabe, clitiques, affixes, HPSG, LKB.

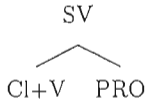
INTRODUCTION

Les langues naturelles présentent certains morphèmes dont le statut grammatical est ambigu. Ces morphèmes se présentent sous la forme d'éléments attachés morphophonologiquement à des hôtes forts dont ils dépendent. Ces formes font office, par ce fait même, d'éléments d'interface entre les différents modules de l'analyse grammaticale en faisant appel à la phonologie, à la morphologie et à la syntaxe. Souvent regroupées sous l'étiquette de *clitiques (pronominaux)*, ces unités ont occupé les débats linguistiques depuis plusieurs années comme le signalent Heap et Roberge (2001).

Le débat portant sur l'origine de ces formes dites clitiques, dans le cadre génératif chomskyen, a opposé deux courants de pensée : la génération à la base du système (dans le lexique), une hypothèse adoptée, entre autres, par Cummins et Roberge (1994), Jaeggli (1982) et Rivas (1977), et l'analyse syntaxique transformationnelle par déplacement, une hypothèse défendue, entre autres, par Cardinaletti et Starke (1994), Chomsky (1995) et Kayne (1975). La première analyse, dite affixale ou morphologique [1], prend l'attachement morphologique de ces unités pour la propriété dominante alors que la deuxième prend les propriétés argumentales de ces unités pour dominantes, comme le souligne Bouchard (2002).

Pour illustrer l'analyse qui prône une génération à la base, nous prenons pour exemple l'approche de Jaeggli (1982). Cet auteur propose en effet que les clitiques soient directement générés sur le verbe et qu'ils soient reliés à une position occupée par l'élément vide PRO. Cette position correspond à celle du sujet dans le cas des clitiques sujets et à celle de l'objet dans le cas des clitiques objets. Nous pouvons illustrer cette analyse dans la structure arborescente suivante :

(1)



Les analyses transformationnelles proposent, par contre, un déplacement qui affecte le clitique ou le verbe. Kayne (2000) par exemple, et en se basant sur l'analyse antisymétrique proposée dans

[1] Jaeggli (1982) affirme que l'approche affixale fournit une explication des phénomènes du redoublement (du clitique par un argument lexical). Ce lien entre le redoublement et le statut du clitique (comme argument ou marqueur d'accord) constitue l'un des arguments les plus utilisés pour défendre l'approche affixale. Voir, par exemple, les travaux de Auger (1994) et Miller (1992). Il faut noter, en outre, que l'approche affixale n'est pas une invention de la linguistique moderne et se trouve énoncée dans des textes aussi vieux que la grammaire de Port-Royal. Voir Heap et Roberge (2001).

son ouvrage Kayne (1994), défend une idée selon laquelle le clitique doit toujours s'adjoindre à gauche d'une catégorie fonctionnelle telle que AGR.

Il faut toutefois noter que la majorité des analyses sont en réalité hybrides et font la distinction entre le statut grammatical des clitiques sujets (réalisés sous le nœud I, ou INFL) et celui des clitiques objets (attachés au verbe) : Auger (1994), Borer (1984), Everett (1985), Roberge (1986) et Sportiche (1983).

L'apport de la théorie grammaticale syntagmatique endocentrique HPSG à ce débat est considérable. L'hypothèse qui a été la plus exploitée dans ce cadre est affixale. Les clitiques ne sont, selon cette hypothèse, que des affixes attachés à leurs hôtes dans le lexique. C'est l'idée défendue, entre autres, par Auger et Miller (1994), Miller (1992), Miller et Monachesi (2003), Miller et Sag (1997, 1995) et Monachesi (1998a) pour les langues romanes. Toutefois, l'idée préconçue selon laquelle tous les chercheurs qui travaillent dans ce cadre s'accordent à réserver un traitement affixal/lexical aux clitiques ne se vérifie pas quand on prend en considération certains travaux, dont ceux de Crysmann (2000) et Kupsc (2000, 1999). Ces deux auteurs ont en fait travaillé sur le polonais et le portugais européen respectivement et sont parvenus à une hypothèse non affixale concernant certains des clitiques étudiés. Il semble donc que la langue étudiée influence les analyses adoptées, ce qui témoigne d'un souci empirique louable[2] : les données orientent la théorie et non le contraire.

Les marqueurs d'arguments[3] de l'arabe standard[4] sont longtemps restés en périphérie par rapport à ces débats. Bien que leur statut grammatical soit problématique, peu de travaux y sont consacrés. Pour les désigner, certains auteurs utilisent des termes comme *clitique*, *affixe*, *pronom* et *marqueur d'accord*. Cette terminologie n'est pas toujours soutenue par une analyse linguistique systématique du statut et du rôle de ces unités dans la grammaire de l'arabe. En outre, les analyses sont des fois implicites et sous-tendent des travaux sur d'autres unités de la langue, telles que les types de phrases, les complémenteurs ou encore le système de l'accord.

Nous pensons que la meilleure façon d'aborder la question de ces marqueurs est de poser des

[2] Nous mettons l'accent sur ce souci empirique dans le premier chapitre, section (1.3.2).

[3] Nous définissons les marqueurs d'arguments comme étant les morphèmes qui encodent le genre, le nombre et la personne et qui réfèrent à des arguments, à savoir au sujet ou à l'objet. Cette appellation englobe donc les marqueurs de sujet et les marqueurs d'objet. Par ailleurs, l'emploi de ce terme se veut neutre par rapport au statut grammatical de ces éléments, un problème qui sera soulevé dans les chapitres 3 et 4.

[4] Nous référons à cette variété par le mot *arabe* et nous employons les qualificatifs *standard* et *dialectal* chaque fois qu'une distinction entre ces variétés est nécessaire. L'arabe standard réfère aux deux variétés historiques que sont l'arabe classique et l'arabe moderne. Ce choix est motivé par le fait qu'il n'existe, à notre connaissance, aucune différence entre ces deux variétés en ce qui a trait aux marqueurs d'arguments.

questions concernant leur statut grammatical, d'évaluer l'impact des réponses à ces questions sur le système grammatical de la langue étudiée et d'intégrer ces réponses dans une théorie linguistique capable d'exprimer des universaux.

En choisissant ces marqueurs comme sujet de recherche, nous visons à atteindre plusieurs objectifs. Nous avons pour but, tout d'abord, de fournir une description exhaustive des marqueurs de sujet et des marqueurs d'objet de l'arabe en nous appuyons sur un corpus effectif et sur notre intuition. La raison qui motive ce choix réside dans le fait qu'un grand nombre des analyses qui traitent de ces marqueurs ne prennent en compte qu'un nombre restreint de données à la fois. Bien que ce choix méthodologique permette dans certains cas de se concentrer sur l'essentiel, il a l'inconvénient de ne pas prendre en compte des données pouvant être importantes, voire cruciales. La couverture empirique de telles analyses reste donc un objectif à atteindre. À l'inverse, nous voulons que notre description (et également notre analyse) englobe le plus de données possibles pour assurer une couverture empirique adéquate.

Ensuite, ce volet descriptif de notre travail nous permettra d'atteindre notre deuxième objectif, à savoir contribuer à la compréhension de certains phénomènes morphologiques et syntaxiques impliquant les marqueurs d'arguments de l'arabe. Nous visons surtout à contribuer au débat sur les asymétries de l'accord[5]. Nous sommes d'avis qu'une analyse adéquate des marqueurs d'arguments nous permettra de mieux cerner cette problématique.

Enfin, nous visons à contribuer non seulement à l'étude des marqueurs d'arguments de l'arabe, mais également à l'étude des clitiques en général. Cet objectif peut être atteint en questionnant la validité des notions théoriques développées pour statuer sur le rôle grammatical des clitiques, comparés aux pronoms dits forts et aux syntagmes nominaux, le lien entre les trois étant très instructif en arabe, et nous y consacrons, par conséquent, une grande part de notre description.

La description mène à une analyse et l'analyse que nous visons à développer s'inscrit dans le cadre de la grammaire HPSG. Étant une théorie développée à l'origine pour les langues germaniques, à savoir essentiellement l'anglais, son utilisation pour modéliser les phénomènes linguistiques dans d'autres familles de langues, surtout la famille slave et la famille romane, a permis d'enrichir les outils théoriques et formels dont dispose cette grammaire. Nous pensons que la famille sémitique peut enrichir davantage cette théorie surtout en prenant en compte les particularités de sa morphologie non concaténative (appelée morphologie à schèmes) et l'ordre

[5] Voir le deuxième chapitre pour une introduction à ce phénomène.

relativement libre dont jouissent les constituants dans cette famille de langues. Pourtant, très peu d'études ont été consacrées à la modélisation des phénomènes linguistiques rencontrés dans les langues sémitiques dans le cadre de cette grammaire. De ce fait, nous poursuivons un double objectif : démontrer que ce cadre théorique est applicable à cette famille de langues et contribuer au développement de ce modèle et des recherches linguistiques sémitiques.

Le travail que nous entreprenons dans notre thèse est basé sur un certain nombre d'hypothèses :

1. Le statut morphologique des marqueurs d'arguments doit être séparé de leur statut morphosyntaxique, les affixes n'étant pas forcément des marqueurs d'accord. Notre analyse apportera des arguments supplémentaires en faveur de cette hypothèse, défendue également par Auger (1994) dans sa thèse sur les marqueurs d'arguments du français parlé informel.
2. On ne peut traiter les marqueurs d'arguments de manière uniforme. Les marqueurs de sujet diffèrent des marqueurs d'objet sur plusieurs plans et doivent être analysés en prenant en compte ces différences. Les marqueurs de sujet sont des affixes flexionnels alors que les marqueurs d'objet sont des clitiques. Cette différence de statut morphologique est à notre avis la source des asymétries observées quand nous comparons le comportement de ces unités.
3. On ne peut traiter les marqueurs d'objet comme des affixes. À l'encontre de la majorité des analyses des clitiques dans le cadre de HPSG, qui tendent à dissoudre ces unités sous l'étiquette affixale, nous pensons qu'il y a lieu de postuler l'existence d'une catégorie autonome, celle des clitiques, différente de celle des affixes. L'impact de cette distinction se voit au niveau de l'analyse syntaxique, où seuls les clitiques constituent des nœuds dans la structure arborescente.
4. Les marqueurs de sujet sont des affixes flexionnels faisant partie des entrées lexicales des verbes qui en sont les hôtes. Ils sont donc insérés en lexique et ne constituent pas de nœuds syntaxiques, même quand ils sont des arguments du verbe.

Notre thèse est organisée comme suit. Nous présentons notre cadre théorique, la grammaire syntagmatique endocentrique HPSG, dans le premier chapitre. Cette présentation se concentre sur les aspects théoriques et formels qui nous servent dans l'analyse des phénomènes à l'étude. Nous présentons ainsi les principes de base, les notions fondamentales et quelques exemples d'analyse en relation à notre sujet : l'analyse des clitiques en général et le traitement de quelques phénomènes impliquant la langue arabe dans ce cadre.

Le deuxième chapitre est de nature descriptive. Nous y présentons le système pronominal

de l'arabe en mettant l'accent sur les deux sortes d'éléments qui composent ce système : les pronoms indépendants et les marqueurs d'arguments. Cette présentation est basée sur des données effectives tirées de notre corpus de vérification et comporte également des exemples que nous avons construits.

La défense de nos hypothèses de recherche est entreprise dans les chapitres III et IV. Dans le chapitre III, nous posons la question que nous qualifions de morphosyntaxique : les marqueurs d'arguments sont-ils mieux traités comme arguments ou comme non-arguments ? Le statut des marqueurs d'objet est ainsi comparé à celui des marqueurs de sujet et plusieurs analyses précédentes sont exposées et critiquées. Nous testons également la validité de nos hypothèses quant à la morphosyntaxe des marqueurs d'arguments en en inspectant les effets sur le système de l'accord de l'arabe et sur certains aspects de la variation dialectale.

Nous consacrons le quatrième chapitre à la question que nous qualifions de morphologique : les marqueurs d'arguments sont-ils mieux traités comme affixes ou comme non-affixes ? Nous répondons à cette question en faisant appel aux critères et aux tests de deux principaux auteurs : Arnold Zwicky et Philip Miller.

Le cinquième chapitre est une formalisation des hypothèses défendues dans les deux chapitres précédents en employant les outils fournis par HPSG. Cette formalisation à base lexicaliste prouve que la théorie adoptée est assez flexible pour rendre compte des phénomènes étudiés même si elle a été conçue à l'origine avec, en tête, des langues très différentes de celles de la famille sémitique. Pour tester cette formalisation, nous consacrons une seconde partie de ce chapitre à l'implémentation de l'analyse dans une grammaire informatisée. Cette implémentation se fait dans le système LKB et démontre que notre analyse se justifie dans un système de performance.

CHAPITRE I

CADRE THÉORIQUE

> More surprising, however, is the fundamental observation that from a broad range of research directions -from varied work within linguistics, artificial intelligence, and computer science- researchers are converging upon a single approach to grammar of great flexibility and power. (Shieber 1986, p. 67)

Introduction

Notre travail de recherche porte sur la modélisation des marqueurs d'arguments de l'arabe standard dans le cadre des grammaires basées sur les contraintes, et notamment la grammaire syntagmatique endocentrique (connue également sous le nom de *la grammaire syntagmatique guidée par les têtes*) HPSG (pour Head-Driven Phrase Structure Grammar). Nous consacrons ce premier chapitre à introduire ce cadre théorique ainsi que le traitement des clitiques et l'étude de la langue arabe dans ce cadre. Cette introduction ne peut être que partielle et se concentre sur les aspects qui nous serviront dans l'analyse [1]

Ce chapitre est divisé comme suit : après avoir motivé notre choix de ce cadre dans la section (1.1) et présenté le paradigme des grammaires à base de contraintes dans la section (1.2), nous présentons les propriétés formelles et les notions fondamentales de HPSG dans la section (1.3). Suit la section (1.4), où nous présentons le traitement des clitiques dans le cadre de HPSG. La section subséquente, (1.5), est consacrée à la rencontre entre le cadre théorique et la langue étudiée. Nous y présentons les quelques travaux qui ont porté sur la langue arabe dans le cadre de la grammaire HPSG.

[1] Le lecteur est invité à consulter les ouvrages cités pour plus de détails.

1.1 Le choix du cadre

En plus de sa puissance formelle, nous avons choisi de travailler dans le cadre de HPSG pour plusieurs raisons. Tout d'abord, les auteurs qui travaillent dans ce cadre ont consacré un grand effort à l'étude des clitiques et des morphèmes semblables, et ce, dans des langues appartenant à des familles différentes (romanes et slaves essentiellement). En outre, et grâce à l'adoption des structures de traits [2], la modélisation des données dans ce cadre prend en compte les divers aspects du signe linguistique. Ainsi, sont représentés dans une même structure les traits phonologiques, morphologiques, syntaxiques, sémantiques et pragmatiques. Nous verrons que les marqueurs d'arguments de l'arabe impliquent plusieurs de ces aspects, notamment la phonologie, la morphologie, la syntaxe et les interfaces entre ces systèmes (à savoir la morphophonologie et la morphosyntaxe).

De plus, si nous savons que cette théorie a évolué en étant confrontée à des langues non SVO (essentiellement le gallois, une langue VSO), nous pouvons mesurer l'importance de travailler sur l'arabe, une langue à verbe initial, mais qui, comme l'anglais et le français, autorise également l'ordre dit SVO. Nous verrons que le traitement des marqueurs d'arguments implique également le traitement de la question de l'ordre de surface.

Finalement, HPSG est une théorie linguistique appropriée à l'implémentation informatique. Elle est même la théorie linguistique la plus utilisée dans ce domaine. L'implémentation informatique n'est pas un but en soi, mais elle nous permettra de tester l'analyse adoptée dans un système de performance.

1.2 Les grammaires à base de contraintes

HPSG est l'une des théories linguistiques regroupées sous l'appellation générale de *grammaires basées sur les contraintes* [3]. Elle représente même l'exemple typique de ces grammaires pour certains auteurs. De ce fait, il va sans dire que l'introduction au style d'analyse adopté par HPSG passe *ipso facto* par une introduction à ces grammaires.

Les grammaires basées sur les contraintes sont nées de la confluence de l'informatique et de la linguistique formelle comme le souligne Shieber (1992, 1991, 1986). La convergence de ces sources a

[2] Nous présentons ce formalisme dans la section (1.3.4).

[3] Ce terme désigne désormais ce qui était auparavant connu sous d'autres appellations qui se sont avérées moins adéquates, à savoir *grammaires basées sur l'information* et *grammaires basées sur l'unification*, voir Pollard (1996) et Sag *et al.* (2003).

donné naissance à plusieurs approches qui ont exploité la richesse formelle offerte par ce paradigme pour développer plusieurs outils et théories linguistiques [4]. Parmi les outils linguistiques éminents, les plus connus sont PATR II, présenté par Shieber *et al.* (1983), la Grammaire Fonctionnelle de l'Unification (ou FUG, pour *Functional Unification Grammar*) présentée par Kay (1979) et la Grammaire à Clauses Définies (ou DCG, pour *Definite Clause Grammar*) utilisée surtout avec le langage de programmation logique PROLOG et présentée, entre autres, par Abramson et Dahl (1989). Parmi les théories, la grammaire lexicale fonctionnelle (ou LFG, pour *Lexical Functional Grammar*) (Bresnan, 1982), la grammaire syntagmatique généralisée (ou GPSG, pour *Generalized Phrase Structure Grammar*) (Gazdar *et al.*, 1985), HPSG et la grammaire d'arbres adjoints (ou TAG, pour *Tree-Adjoining Grammars*), présentée, entre autres, par Joshi et Schabes (1997), sont les plus représentatives.

1.3 La grammaire syntagmatique endocentrique HPSG

La grammaire HPSG est l'œuvre des deux linguistes Carl Pollard et Ivan A. Sag. Elle a connu trois versions ponctuées par la publication des deux ouvrages « fondateurs » de Pollard et Sag (1994, 1987). La première version (appelée également HPSG I) a vu le jour avec la publication de l'ouvrage de Pollard et Sag (1987), dans lequel les auteurs ont posé les fondements théoriques et formels de cette grammaire. À cette étape, les auteurs ont défini le formalisme choisi (les structures de traits typées) et présenté des analyses linguistiques de phénomènes tels que la sous-catégorisation. Ils ont également dissocié les règles de préséance linéaire (PL) des règles de dominance immédiate (DI) et proposé une organisation hiérarchique du lexique.

La deuxième version (appelée HPSG II) est représentée par les huit premiers chapitres de l'ouvrage de Pollard et Sag (1994). Dans cette version, la grammaire comporte une meilleure organisation des structures de traits avec, notamment, la fusion des deux attributs SYN(taxe) et SEM(antique) en un seul attribut SYNSEM. Plusieurs changements ont également touché les principes de la grammaire, dont l'adoption de nouveaux principes (le principe de marquage et le principe des traces) et l'affinement d'autres principes déjà présents dans la 1re version (dont le principe des traits non locaux).

[4] Selon Shieber (1986), les outils linguistiques se distinguent par des mécanismes généraux destinés à accroître leur pouvoir expressif alors que les théories linguistiques ont tendance à incorporer des instruments définis pour des besoins spécifiques, liés aux hypothèses linguistiques auxquelles ces théories adhèrent. La compréhension des choix conceptuels faits dans les formalismes du 1er type est relativement indépendante des analyses linguistiques qu'ils servent à formuler. Au contraire, ces deux aspects sont intimement liés dans les formalismes du 2e type, où les choix formels sont censés exprimer des principes linguistiques universels.

La troisième version de la théorie (appelée HPSG III) est présentée dans le chapitre 9 de l'ouvrage de Pollard et Sag (1994) *Reflections and Revisions*. Les deux changements majeurs dans cette version consistent dans l'abandon de certaines catégories vides (comme les traces) dans l'analyse de certaines dépendances non bornées et une modification de la description de la sous-catégorisation. Cette modification adoptée sous l'influence des travaux de Robert Borsley, surtout Borsley (1987), consiste à partitionner l'attribut SUBCAT (pour la sous-catégorisation) en trois attributs : SPR (pour les spécifieurs), SUBJ (pour les sujets) et COMPS (pour les compléments).

Ces changements n'ont toutefois pas touché les idées principales qui étaient à l'origine de cette théorie et que nous présentons dans les paragraphes suivants.

1.3.1 Les traits distinctifs de la grammaire

HPSG se présente comme une théorie linguistique dont les principales propriétés sont le fait d'être basée sur les contraintes plutôt que sur les règles, l'orientation surfaciste, l'éclectisme et le lexicalisme au sens fort.

1.3.1.1 Une théorie à base de contraintes

HPSG est une grammaire générative basée sur les contraintes. Cette spécificité signifie que la grammaire ne fournit pas les règles de production, mais l'ensemble des contraintes qu'une séquence doit satisfaire pour être bien formée, comme le souligne Abeillé (1993). Les contraintes n'ont pas à être ordonnées étant donné qu'elles doivent toutes être « satisfaites » simultanément, à la différence des règles transformationnelles. La satisfaction des contraintes n'est par ailleurs pas destructive, en ce sens qu'elle n'altère pas les structures (par effacement ou par déplacement).

HPSG emploie les contraintes de manière systématique et la tendance à fonder une théorie complètement basée sur ces dernières est explicitée dans le chapitre XVI de l'ouvrage de Sag et Wasow (1999) *From Rule to Sign*. Dans ce chapitre, le modèle a évolué vers une version où aucune règle lexicale n'est employée pour relier les structures. Tout est désormais modélisé en termes de contraintes sur la bonne formation des signes, décrites par des structures de traits typées et hiérarchisées.

1.3.1.2 Une orientation surfaciste

La réhabilitation des structures de surface constitue l'une des propriétés essentielles des grammaires basées sur les contraintes. HPSG partage cette propriété avec les autres grammaires du même paradigme et n'a recours aux catégories vides que le moins possible. L'évolution du modèle révèle clairement cette tendance surfaciste. En effet, alors que Pollard et Sag (1987) légitiment la catégorie vide Trace (t) pour rendre compte des dépendances non bornées, on aboutit à un modèle sans traces dans HPSG III.

HPSG partage son orientation surfaciste et l'importance accordée aux contraintes avec les autres grammaires du même paradigme, mais elle possède des propriétés qui lui sont propres. Parmi celles-là, le caractère éclectique et le lexicalisme au sens fort sont les traits les plus proéminents.

1.3.1.3 Une théorie éclectique

HPSG est née de la rencontre de plusieurs théories linguistiques et formelles, dont la plus influente est GPSG telle que présentée par Gazdar *et al.* (1985). On y trouve également des influences de certaines théories syntaxiques, dont GB telle que conçue par Chomsky (1982), LFG, décrite par Bresnan (1982) par exemple, et les grammaires catégorielles CG, dont une version est défendue par Steedman (2000). Certaines théories sémantiques ont également participé à façonner cette grammaire, dont la *Situation Semantics* de Barwise et Perry (1983) et la *Discourse Representation Theory* DRT de Kamp et Reyle (1993). Pollard et Sag (1987) sont d'avis que cet éclectisme permet d'atteindre les deux objectifs suivants : profiter des avantages de plusieurs traditions de recherches, et démontrer que toutes ces traditions, aussi disparates puissent-elles paraître, font en fin de compte partie du même projet scientifique. C'est donc dans cet esprit de continuité que Pollard et Sag ont bâti les fondements de cette grammaire qui a su intégrer un grand nombre de notions qui ont prouvé leur force dans d'autres théories, telles que le schéma X-barre, le principe des traits de tête et celui de la valence.

1.3.1.4 Une théorie lexicaliste au sens fort

Le deuxième trait particulier à HPSG est son lexicalisme. En fait, cette grammaire localise la plupart de l'information grammaticale et sémantique dans le lexique et plus précisément dans les

entrées lexicales. Ces entrées lexicales correspondent directement aux mots réalisés *en surface* et les informations contenues dans ces entrées sont très riches et déterminent, de ce fait, la structure sémantique et syntaxique en conformité avec un nombre limité de principes (tels que le *principe de traits de tête*[5] ou le *principe de valence*[6]). C'est pour cette raison que plusieurs des phénomènes traités syntaxiquement par d'autres théories trouvent leur place dans le lexique en HPSG. C'est le traitement réservé, par exemple, aux phénomènes du liage dans les relations à longue distance ainsi que dans le traitement du passif et des explétifs. Le passif en anglais, par exemple, est traité par Sag *et al.* (2003) en faisant appel à une entrée lexicale, celle de la copule *be*, et à une règle lexicale[7] qui relie les deux formes active et passive du verbe.

> Our treatment of the active/passive alternation in English is based on a relationship between verb forms. We formalize this with a derivational lexical rule that modifies the lexeme type, the morphology, the argument structure, and some details of the HEAD values. (*Ibid.*, p. 327)

Dans les sections suivantes, nous présentons le formalisme et la grammaire développés dans le cadre de HPSG. Pour ce faire, nous introduisons les traits saillants du formalisme ainsi que les notions fondamentales de structures de traits et de principes de bonne formation. Ces définitions proviennent essentiellement des ouvrages de Pollard (1997, 1996), Pollard et Sag (1994) et Sag *et al.* (2003).

1.3.2 Les traits saillants du formalisme

Selon Shieber (1986), tout formalisme grammatical qui se réclame du paradigme basé sur les contraintes s'engage à respecter les choix généraux suivants : ce formalisme doit être surfaciste[8], informatif[9], inductif[10] et déclaratif[11]. Ces quatre choix distinguent les théories génératives basées sur les contraintes des autres courants qui s'inscrivent dans la lignée du programme génératif.

Par ailleurs, le formalisme adopté dans HPSG (et dans toute grammaire basée sur les contraintes),

[5] Nous présentons ce principe dans la section (1.3.5.2).

[6] Nous présentons ce principe dans la section (1.3.5.3).

[7] Nous définissons les règles lexicales et nous en donnons quelques exemples dans la section (1.3.7).

[8] « Providing a direct characterization of the actual surface order of string elements in a sentence. » (Shieber, 1986, p. 6)

[9] Ce caractère revient à associer aux éléments manipulés par la théorie une information provenant d'un domaine informationnel donné (Shieber, 1986, p. 7).

[10] Les grammaires basées sur les contraintes procèdent de manière inductive (ou *bottom-up*) en ceci qu'elles obtiennent des généralisations à partir des phénomènes observés au lieu de commencer par énoncer de grandes généralités (ou des principes) et d'essayer par la suite de trouver des phénomènes qui les confirment.

[11] Le caractère déclaratif (vs procédural) revient à exiger que la grammaire fournisse un modèle des entités linguistiques possibles et non un modèle de la manière dont ces entités sont traitées : « defining the association between strings and informational elements in terms of *what* associations are permissible, not *how* they are computed. » (Shieber, 1986, p. 7)

selon Pollard (1996) est caractérisé par les huit traits suivants :

1. La générativité : HPSG est une grammaire générative en ceci qu'elle se conforme aux trois critères qui définissent cette sorte de grammaires. Ces critères sont les suivants :

   (a) La structure mathématique : la grammaire générative détermine avec précision quelles sont les structures candidates qui seront considérées comme bien formées ou mal formées par la grammaire. Ce sont les structures de traits typées dans HPSG, telles que définies par Carpenter (1992). Ces structures de traits représentent, selon Shieber (1986), le domaine informationnel sur lequel tout le système grammatical est basé.

   (b) La théorie formelle : la grammaire générative se base sur une théorie formelle bien définie, comme la logique des traits en ce qui concerne HPSG.

   (c) La décidabilité : la grammaire générative nous permet de décider devant une structure donnée si cette dernière satisfait la grammaire ou non.

2. L'expressivité . la langue dans laquelle sont exprimées les grammaires à base de contraintes est expressivement riche (et peut être une langue formelle, comme PROLOG ou LISP, ou une langue naturelle quelconque). Toutefois, la condition suivante est respectée : la langue choisie ne doit pas imposer de contraintes sur la théorie, puisque c'est cette dernière qui est tenue de les imposer.

3. L'adéquation empirique : la priorité de la théorie basée sur les contraintes est d'obtenir des généralisations empiriques qui englobent plusieurs faits réels et attestés. Faire découler ces généralisations d'un principe plus « profond » reste un objectif secondaire :

   > Make a precise, falsifiable hypothesis that accounts for a wide range of facts (i.e. get the details right for a subpart of the language), then revise the hypothesis to expand coverage. (Pollard, 1997)

   > HPSG's methodology insists upon accepting linguistic analyses as explanatory only when they can be shown to be compatible with a broad and representative database of facts from a given language. (Webelhuth *et al.*, 1999, p. 3)

4. La responsabilité psychologique : les grammaires à base de contraintes ont pour but l'investigation de la compétence linguistique de l'être humain et non la manière dont le langage humain est traité (par le cerveau). Autrement dit, l'objectif est le suivant :

   > to construct a scientific theory of the system of knowledge that is embodied in the human mind/brain which makes language possible. This is distinct from constructing a psycholinguistic theory (a theory of how the mind/brain knowledge to produce and interpret utterances). (Pollard, 1997)

Toutefois, même si la grammaire n'est qu'un modèle de compétence, elle est tenue de ne pas être basée sur des calculs que le locuteur ne pourrait accomplir.

5. L'absence de l'effet destructeur : les opérations auxquelles font référence ces grammaires (telles que l'unification) n'engendrent pas la destruction de structures linguistiques déjà existantes.
6. La localité : pour déterminer si une structure est bien formée ou non, les grammaires à base de contraintes ne font pas référence à d'autres structures en compétition avec cette structure, comme dans la théorie de l'optimalité par exemple.
7. Le parallélisme : les grammaires basées sur les contraintes font référence à différents niveaux de l'analyse linguistique, mais elles admettent que ces niveaux existent en parallèle et non dérivés les uns des autres.
8. La non-autonomie radicale (vs l'autonomie de la syntaxe) : ces théories traitent tous les facteurs linguistiques pertinents que ces derniers soient phonétiques, morphologiques, syntaxiques, sémantiques ou pragmatiques. La syntaxe n'a donc pas l'autonomie dont elle jouit dans d'autres théories : « HPSG is not syntactocentric. » (*Ibid.*)

1.3.3 Types, contraintes et hiérarchies

Les objets modélisés par la grammaire HPSG sont organisés en types regroupés dans une hiérarchie à héritage multiple. En recourant au typage, toute grammaire HPSG doit inclure ce qui suit selon Ginzburg et Sag (2000) :

1. Une ontologie linguistique : une énumération des types reconnus par la grammaire.
2. Des conditions d'appropriation : définir quels sont les attributs appropriés à chaque type et quelles sont les valeurs appropriées à chaque attribut.
3. Des contraintes : définir les contraintes imposées à chaque objet d'un type donné. Ces contraintes peuvent être de deux sortes : des contraintes non violables et des contraintes par défaut (violables) selon Sag *et al.* (2003). Les premières sont des contraintes que les types supérieurs dans la hiérarchie transmettent aux types inférieurs de façon monotone, alors que les dernières sont évaluées avant chaque héritage et ne se transmettent que lorsqu'elles sont compatibles avec les contraintes imposées aux types inférieurs.
4. Une hiérarchie : les types sont organisés dans une hiérarchie à héritage multiple. Les types les plus spécifiques (appelés les sous-types ou les descendants) héritent les contraintes imposées à un ou à plusieurs types plus généraux (appelés les super-types ou les parents).

Pour comprendre comment se construit une hiérarchie des types, nous prenons pour exemple celle qui a été proposée par Sag *et al.* (2003). Dans cette hiérarchie, le type le plus général reconnu par la théorie est *str-traits* (pour « structure de traits » [12]). Ce type, qui constitue la racine de l'arbre représentant la hiérarchie des types, possède plusieurs sous-types, dont *expression*, *p-disc* (pour « parties du discours ») et *synsem* (pour « syntaxe-sémantique »). Le sous-type *p-disc* hérite les contraintes non violables et les contraintes par défaut compatibles de son super-type immédiat *str-traits*. Il possède, à son tour, plusieurs sous-types, dont *pdisc-agr* (pour parties du discours qui participent à des relations d'accord). Ce dernier type, qui est plus bas dans la hiérarchie, hérite à la fois les contraintes imposées à son super-type immédiat *p-disc* et celles imposées au super-type supérieur *str-traits*. La figure (1.1) représente une partie de cette hiérarchie.

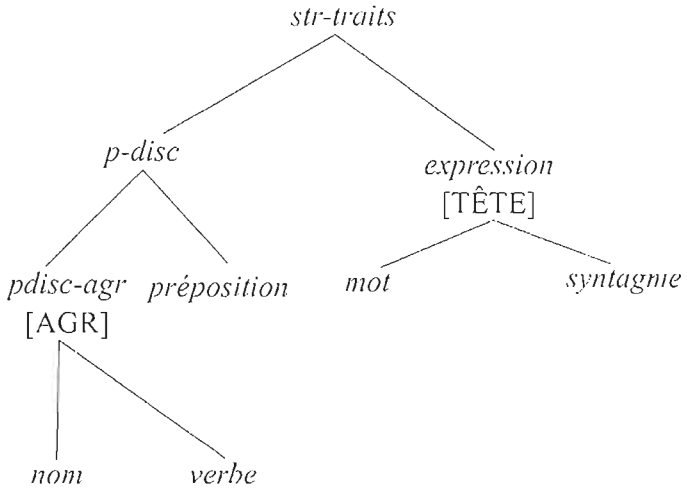


FIGURE 1.1: Une hiérarchie partielle des types

Cette hiérarchie est à héritage simple, puisque chaque type hérite d'un seul parent à la fois. Dans un héritage multiple, chaque type peut hériter de plus d'un seul super-type.

Ce que fournit une telle hiérarchie c'est surtout une déclaration des types admis par la grammaire. Certaines contraintes sont également déclarées. Pour le type *pdisc-agr*, par exemple, la contrainte notée [AGR] exige que les objets de ce type soient porteurs de l'attribut AGR (cet attribut modélise les traits d'accord). De même, le type *expression* est contraint à avoir l'attribut TÊTE parmi ses attributs appropriés.

[12] Nous définissons les structures de traits dans la section subséquente de ce travail.

Il existe toutefois une autre représentation qui déclare à la fois les types et les contraintes imposées aux types, comme le tableau en (1.1), où est indiqué pour chaque type son super-type immédiat (STI) ainsi que les contraintes directement imposées à ce type (et qui s'ajoutent aux contraintes imposées par le super-type).

Types	Contraintes	STI
str-traits		
p-disc		*str-traits*
expression	[TÊTE]	*str-traits*
pdisc-agr	[AGR]	*pdisc*
nom		*pdisc-agr*
verbe		*pdisc-agr*
mot		*expression*
syntagme		*expression*

TABLEAU 1.1: Les déclarations des traits et les contraintes sur les types (version simplifiée)

Cette hiérarchie n'est qu'un exemple simplifié, une hiérarchie plus complète sera présentée dans les sections subséquentes.

Les types ont un statut particulier dans la théorie étant donné qu'ils représentent une augmentation formelle aux structures mathématiques du formalisme, dans le but d'exprimer des généralisations que les seules structures en questions ne peuvent exprimer.

Après avoir présenté la manière dont la théorie organise le lexique, nous consacrons la sous-section subséquente aux structures mathématiques du formalisme.

1.3.4 Les structures mathématiques : les structures de traits

Loin d'être une invention de la théorie HPSG, les structures de traits représentent le noyau formel du modèle et revêtent une grande importance puisqu'elles servent à représenter non seulement les objets linguistiques étudiés, mais également les principes de la théorie et même ses différentes règles.

1.3.4.1 Définitions

Les structures de traits sont une représentation formelle, et très souvent partielle, des propriétés des objets linguistiques modélisés par la théorie. Nous pouvons lire ce qui suit dans (Sag *et al.*, 2003, p. 561) :

> A standard way of representing linguistic information is in terms of complexes of features and values. A feature can be thought of as a dimension along which different linguistic entities (such as words, phrases, or sentences) may differ, and values identify locations on those dimensions. A feature-value pair models a property of a linguistic entity that distinguishes it in a linguistically interesting way from some other entities.

Ces objets peuvent être des lexèmes, des mots, des syntagmes, des phrases et même des contraintes et des principes. Pour les linguistes travaillant dans ce cadre, ces objets sont des signes. Ce terme à saveur saussurienne laisse entrevoir une rencontre entre la forme et le sens.

> We conceive of signs as structured complexes of phonological, syntactic, semantic, discourse, and phrase-structural information. (Pollard et Sag, 1994, p. 15)

Si les objets linguistiques manipulés par la théorie sont de nature complexe et englobent des informations de différents types, il est normal que les représentations formelles de ces objets soient elles aussi complexes. C'est d'ailleurs ce qui caractérise les structures de traits. Ces dernières, à la différence des arbres syntaxiques, par exemple, ne représentent pas seulement l'information syntaxique. On y trouve également des informations de nature phonologique, morphologique, sémantique et pragmatique.

Formellement, une structure de traits est un ensemble d'attributs et de leurs valeurs (chaque paire attribut-valeur constitue un trait). Pour représenter le type *expression*, par exemple, deux attributs sont appropriés : PHON et SYNSEM. Le premier est une représentation de la *forme* de l'expression, c'est-à-dire des propriétés qui servent de base à son interprétation phonétique et phonologique. La valeur de cet attribut est une liste constituée de chaînes de phonèmes. Le deuxième attribut a pour valeur un complexe d'informations syntaxiques et sémantiques représentées par plusieurs traits.

D'autres attributs sont également reconnus dans la théorie. Parmi ces derniers, citons l'attribut STR-ARG, qui modélise la structure argumentale (la liste des arguments sélectionnés) d'un objet donné et n'est approprié qu'à certains types lexicaux (*mot* ou *lexème*). L'attribut LOC modélise les traits réalisés localement (à la différence des traits non locaux) et n'est approprié qu'aux objets de type *synsem*. L'attribut CAT modélise la catégorie de l'objet linguistique étudié, ses traits de sous-catégorisation (modélisés par l'attribut VAL) et ses traits de tête (modélisés par l'attribut

TÊTE). L'attribut CONT modélise les propriétés sémantiques de l'objet et l'attribut CTXT ses propriétés pragmatiques. Notons également les trois attributs qui constituent avec leurs valeurs la valeur de l'attribut VAL : SUJ (qui modélise le sujet), COMPS (qui modélise les compléments) et SPEC (qui modélise le spécifieur).

Les structures de traits sont représentées de manière standard sous forme de matrices d'attributs-valeurs (ou AVM pour *Attribute-Value Matrices*) [13]. Ainsi, la structure de traits qui décrit le type *expression* peut être représentée dans l'AVM (1.1).

(1.1) $\begin{bmatrix} \textit{expression} & \\ \text{PHON} & \langle\rangle \\ \text{SYNSEM} & \textit{synsem} \end{bmatrix}$

1.3.4.2 Les contraintes sur les structures de traits

Les structures de traits utilisées dans le cadre de HPSG doivent satisfaire trois contraintes formelles selon Pollard et Sag (1994) :

1. Ces structures doivent être typées :

   > This means simply that each node is labelled with a *sort symbol* that tells what type of object the structure is modelling ; that is, there is one sort symbol for each basic type (ontological category) of construct. (Pollard et Sag, 1994, p. 17)

   Dans la structure de traits représentée en (1.1), par exemple, deux types sont explicités : *expression* et *synsem*. Dans l'AVM représenté en (1.2), l'attribut SYNSEM, approprié aux objets de type *synsem*, a pour valeur des attributs qui lui sont appropriés, dont l'attribut LOC(al). Ce dernier a, à son tour, les attributs CAT(égorie), CONT(enu) et CTXT (contexte) comme valeurs. Chacun de ces attributs modélise un objet d'un type donné (l'attribut LOC modélise un objet de type *loc*, par exemple, et cet objet est une structure de traits comportant les attributs CAT, CONT et CTXT et leurs valeurs).

[13] Les conventions typographiques suivantes sont respectées dans ce formalisme : les attributs sont écrits en petites majuscules, les types en italiques, les indices atomiques en chiffres arabes encadrés et les indices de listes en lettres majuscules encadrées.

$$(1.2)\quad \begin{bmatrix} \textit{expression} & \\ \text{PHON} & \langle\rangle \\ \text{SYNSEM} & \begin{bmatrix} \textit{synsem} & \\ \text{LOC} & \begin{bmatrix} \textit{loc} & \\ \text{CAT} & \begin{bmatrix} \textit{cat} & \\ \text{TÊTE} & \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle\rangle \\ \text{COMPS} & \langle\rangle \\ \text{SPEC} & \langle\rangle \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \textit{cont} \\ \text{CTXT} & \textit{ctxt} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

2. Les structures de traits doivent être bien typées. Cela revient à exiger que les traits utilisés dans une structure donnée soient déterminés par le type de l'objet linguistique modélisé par cette structure. Ainsi, par exemple, l'attribut CAT ne peut se retrouver que dans la structure de traits qui modélise un objet de type *loc*.

3. Les structures de traits doivent être complètes : « What this means is that they are total (not merely partial) models of the objects they represent. » (Pollard et Sag, 1994, p.18). Il est cependant possible de représenter une structure de traits partielle pour modéliser un objet en mettant l'accent sur certains traits seulement. C'est le cas de la structure en (1.1) par exemple, où plusieurs traits non pertinents pour cette présentation sont omis.

En plus de ces contraintes, les structures de traits utilisés dans la grammaire HPSG possèdent deux propriétés formelles importantes : la récursivité et le partage de valeurs. En effet, la valeur d'un attribut donné peut être atomique ou complexe. Un exemple d'une valeur atomique est la valeur de l'attribut CAS, approprié aux objets de type *nominal*. Les valeurs atomiques possibles de cet attribut en arabe, par exemple, sont *nom(inatif)*, *acc(usatif)* et *gén(itif)*.

Quand la valeur d'un attribut est complexe, elle consiste en une nouvelle structure de traits [14] (qui comporte donc d'autres couples attributs-valeurs), ce qui illustre la récursivité de ces structures. Dans la structure en (1.2), par exemple, la valeur de l'attribut LOC est une structure de traits.

[14] Lorsque plusieurs structures de traits constituent la valeur d'un attribut donné, elles peuvent être regroupées dans une liste notée <>, auquel cas l'ordre des éléments est pertinent, ou dans un ensemble noté {}, où l'ordre n'est pas significatif.

Le partage de valeurs (ou la ré-entrance) se réalise lorsque deux chemins [15] ou plus mènent à la même valeur. Ce partage se note par des chiffres arabes encadrés ou des lettres majuscules encadrées dans les AVM. Dans l'exemple suivant, qui représente l'entrée lexicale [16] simplifiée du verbe *émettre*, les deux chemins SS [17]|LOC|CAT|VAL|SUJ et SS|LOC|CONT|ÉMETTEUR aboutissent à des valeurs identiques. C'est également le cas des chemins SS|LOC|CAT|VAL|COMPS et SS|LOC|CONT|MESSAGE ainsi que des chemins SS|LOC|CAT|VAL|COMPS et SS|LOC|CONT|RÉCEPTEUR. Ce partage de valeurs (ou de structures) est indiqué par la lettre majuscule encadrée A et les deux chiffres encadrés 1 et 2 [18].

(1.3)
$$\begin{bmatrix} \text{PHON} & \langle \textit{émettre} \rangle \\ \text{SS} \mid \text{LOC} & \begin{bmatrix} \text{CAT} \mid \text{VAL} & \begin{bmatrix} \text{SUJ} & \boxed{A} \\ \text{COMPS} & \langle \boxed{1}, \boxed{2} \rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \text{ÉMETTEUR} & \boxed{A} \\ \text{MESSAGE} & \boxed{1} \\ \text{RÉCEPTEUR} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

1.3.4.3 Les opérations sur les structures de traits

Les structures de traits peuvent entrer dans plusieurs relations les unes avec les autres, dont la subsomption et l'unification.

> Viewed intuitively, a feature-structure D subsumes a feature-structure D' (notated D $\sqsubseteq$ D') if D contains a subset of the information in D'. (Shieber, 1986, p. 15)

Ainsi, si nous prenons les deux structures de traits complexes (1.4) et (1.5), nous pouvons affirmer que la première subsume la deuxième).

(1.4) $\begin{bmatrix} \text{CAT} & \textit{SN} \end{bmatrix}$

[15] Le chemin est la liste des attributs qui mènent vers une valeur donnée.

[16] Dans cette perspective, une entrée lexicale est une description d'une structure de traits de type *mot* (le sous-type immédiat de *signe*).

[17] SS : abréviation de SYNSEM.

[18] Cet exemple illustre les liens entre la syntaxe et la sémantique dans ce cadre.

(1.5) $\begin{bmatrix} \text{CAT} & \textit{SN} \\ \text{ACCORD} & \begin{bmatrix} \text{NOMBRE} & \textit{singulier} \end{bmatrix} \end{bmatrix}$

L'unification [19] consiste à enrichir la description d'un objet en combinant les informations de deux structures de traits qui le décrivent partiellement. Formellement,

> We define the unification of two feature structures D' and D" as the most general feature structure D, such that D' $\sqsubseteq$ D and D" $\sqsubseteq$ D. We notate this D = D' $\sqcup$ D". (Shieber, 1986, p. 17-18)

Ainsi, nous pouvons unifier la structure de traits en (1.5) et la structure de traits en (1.6) pour obtenir la structure en (1.7). L'unification des deux structures (1.5) et (1.8) échoue à cause de l'incompatibilité des valeurs de l'attribut NOMBRE.

(1.6) $\begin{bmatrix} \text{CAT} & \textit{SN} \\ \text{ACCORD} & \begin{bmatrix} \text{GENRE} & \textit{masculin} \end{bmatrix} \end{bmatrix}$

(1.7) $\begin{bmatrix} \text{CAT} & \textit{SN} \\ \text{ACCORD} & \begin{bmatrix} \text{NOMBRE} & \textit{singulier} \\ \text{GENRE} & \textit{masculin} \end{bmatrix} \end{bmatrix}$

(1.8) $\begin{bmatrix} \text{CAT} & \textit{SN} \\ \text{ACCORD} & \begin{bmatrix} \text{NOMBRE} & \textit{pluriel} \end{bmatrix} \end{bmatrix}$

L'unification est l'opération formelle qui a donné son nom aux grammaires à base de contraintes dans les années 80 ; on parlait alors de grammaires basées sur l'unification. Pourtant, comme le notent Sag *et al.* (2003) et Shieber (1986), ce n'est pas la seule opération utilisée dans ces formalismes, puisque ces derniers ont recours à d'autres opérations, comme la généralisation, la substitution, la disjonction et la réécriture par surcharge ; l'unification est toutefois l'opération centrale.

[19] Cette notion ne doit pas être confondue avec celle d'union. En réalité, l'unification s'opère entre des structures de traits compatibles et peut aboutir à un échec si les valeurs des attributs des structures sont différentes. L'union est une opération formelle qui ne peut vérifier la valeur des attributs et consiste en une combinaison de structures qui ne peut échouer parce qu'elle est « aveugle » au contenu de ces dernières.

1.3.5 Les principes de bonne formation

Mot et *syntagme* sont les deux sous-types immédiats du type *expression* et constituent la base de la grammaire en HPSG.

> Hence, the lexical entries and construction rules work together : the lexical entries define a set of words ; the construction rules define a set of phrases built from words or phrases. (Ginzburg et Sag, 2000, p. 18)

La bonne formation des signes du premier sous-type dépend de la satisfaction de plusieurs contraintes lexicales, dont le principe de la réalisation des arguments (appelé ARP, pour *Argument Realization Principle*), que nous présentons dans la section (1.3.5.1).

La bonne formation des signes syntagmatiques dépend également de la satisfaction d'un certain nombre de contraintes présentées sous forme de principes. Ainsi, une structure syntagmatique ne peut être reconnue comme bien formée que si elle satisfait tous les principes de la grammaire, ce qui inclut les principes dits universels et les principes spécifiques à la langue étudiée. Les principes universels reconnus dans ce cadre ne sont pas nombreux, mais ils ont une large couverture. Parmi ces principes, le *principe des traits de tête* (désormais PTT) et le *principe de la valence* (désormais PV) et le *principe généralisé des traits de tête* (PGTT) sont les plus représentatifs.

1.3.5.1 Le principe de la réalisation des arguments

Pour traiter les variations observées entre les langues en ce qui a trait à la structure argumentale [20], cette information est encodée dans un principe qui constitue une contrainte sur le type *mot*. Ce type est en fait contraint à avoir un attribut particulier STR-ARG. Ce dernier modélise la grille de sous-catégorisation (la structure argumentale) d'un mot donné dans une liste qui concatène les arguments qu'il sélectionne en commençant par l'argument le moins oblique (le sujet/le spécifieur, le complément direct, le complément indirect, dans cet ordre). Ce principe peut être représenté dans la structure suivante :

[20] Les arguments ne sont pas toujours des nœuds syntaxiques dans les différentes langues. Ils peuvent en fait être des affixes comme le démontrent Auger (1994), Auger et Miller (1994), Miller (1992) et Miller et Sag (1997), entre autres auteurs, ou être nuls selon Sag *et al.* (2003).

(1.9)

$$mot \Rightarrow \begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} & \begin{bmatrix} \text{SUJ} & \boxed{A} \\ \text{SPR} & \boxed{B} \\ \text{COMPS} & \boxed{C} \end{bmatrix} \\ \text{STR-ARG} & \boxed{A} \oplus \boxed{B} \oplus \boxed{C} \end{bmatrix}$$

> This formulation of the ARP simply ensures that all arguments are realized on the appropriate valence list - and hence are selected by a given word in a headed construction. (Ginzburg et Sag, 2000, p. 23)

Manning et Sag (1999) ont plaidé pour la dissociation entre la valeur de l'attribut STR-ARG et la valeur de l'attribut VAL. Autrement dit, la valence d'un mot peut ne pas contenir tous les éléments sélectionnés dans sa liste argumentale. Cela est possible quand les arguments en question sont réalisés comme des affixes ou sont nuls.

Par ailleurs, permettre à la liste STR-ARG d'exprimer des informations qui ne sont pas exprimées dans la liste des traits de valence résout les problèmes posés par la modélisation de l'ordre des mots puisque l'attribut STR-ARG encode l'ordre de réalisation des arguments.

1.3.5.2 Le principe des traits de tête

Ce principe fait appel au caractère endocentrique de la grammaire (une grammaire guidée par les têtes). La tête syntaxique constitue un postulat de la théorie. Elle se définit intuitivement comme étant le constituant le plus important du syntagme, celui qui détermine sa catégorie et plusieurs de ses propriétés syntaxiques et sémantiques (Sag *et al.*, 2003, p. 562).

Pour comprendre le rôle de ce principe dans la grammaire HPSG, il faut noter que les syntagmes sont dotés d'un attribut non approprié aux mots : FILLES. Cet attribut prend pour valeur une structure de traits de type *str-cons* (structure de constituants), qui modélise les constituants immédiats du syntagme. Ce type (*str-cons*) domine, pour sa part, plusieurs sous-types qui dépendent des relations structurales qu'entretiennent les constituants au sein du syntagme. Le sous-type qui nous importe ici est celui qui modélise les constructions endocentriques, à savoir le type *str-tête* (structures à tête). Les attributs appropriés à ce dernier type sont FILLE-TÊTE et FILLES-COMPS. Le premier prend pour valeur un seul signe (la tête lexicale) puisque chaque syntagme possède une seule tête alors que le deuxième prend une liste de signes pour valeur[21].

[21] Dans cette liste, les compléments sont ordonnés en commençant par le moins oblique.

Le principe des traits de tête intervient pour garantir la bonne formation des syntagmes du type STR-TÊTE en imposant un partage des valeurs de l'attribut TÊTE du syntagme en question (ou la mère) et de celles de l'attribut TÊTE de sa FILLE-TÊTE. Ce principe peut être énoncé sous la forme d'une structure de traits, comme celle en (1.10).

(1.10)
$$\begin{bmatrix} \textit{str-tête} & \\ \text{PHON} & \langle\rangle \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{TÊTE} & \boxed{1} \\ \text{FILLES} \mid \text{FILLE-TÊTE} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{TÊTE} & \boxed{1} \end{bmatrix}$$

Ce principe garantit que les valeurs de l'attribut TÊTE de la phrase soient partagées par l'attribut TÊTE du syntagme verbal (SV). Dans l'exemple suivant, qui présente une structure de traits partielle de la phrase *Michel éternue*, les traits de tête de la fille-tête (le SV *éternue*) sont partagés par l'attribut TÊTE de la mère (la phrase).

(1.11)
$$\begin{bmatrix} \textit{str-tête} & & \\ \text{PHON} & & \textit{<Michel, éternue>} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{TÊTE} & & \boxed{1} \\ \text{FILLES} \mid \text{FILLE-TÊTE} & \begin{bmatrix} \text{PHON} & \textit{<éternue>} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{TÊTE} & \boxed{1}\textit{verbe} \end{bmatrix} & \end{bmatrix}$$

Les traits de tête [22] se propagent [23] donc de la fille à la mère de manière inductive *bottom-up*. Le principe de la valence illustre un autre cas où la théorie se montre inductive.

1.3.5.3 Le principe de la valence

Ce principe impose la contrainte suivante sur les objets de type *syntagme* : la valeur de l'attribut VAL de la FILLE-TÊTE d'un syntagme de type *str-tête* est la somme qui résulte de la concaténation de la valeur de l'attribut VAL de la mère et des valeurs des attributs SYNSEM des filles compléments FILLES-COMPS.

[22] Parmi les traits de tête qui se propagent de la fille à la mère, citons le trait [FIN] approprié aux têtes verbales. Ainsi, dans notre exemple, le verbe *éternue*, le SV *éternue* et la phrase *Michel éternue* sont marqués [FIN].

[23] Il faut prendre ce terme dans un sens non dérivationnel, puisqu'il s'agit essentiellement d'une opération d'unification lors de laquelle les valeurs des traits de tête de la mère et de la fille sont unifiées. L'échec d'une telle unification signifie une violation de ce principe et le syntagme qui en résulte ne peut être considéré comme un syntagme à tête.

> The effect of this principle is to 'check off' the subcategorization requirements of the lexical head as they become satisfied by the complement daughters of its phrasal projections. At the same time, the SUBCAT elements themselves are token-identical to the SYNSEM values of the corresponding complements. (Pollard et Sag, 1994, p. 34)

Ce principe peut être représenté sous la forme d'une structure de traits, comme celle en (1.12).

(1.12)
$$\begin{bmatrix} \textit{syntagme} & \\ \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} & \langle \boxed{1} \rangle \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{VAL} & \langle \boxed{1} \oplus \boxed{2} \rangle \\ \text{FILLES-COMPS} & \langle \boxed{2} \rangle \end{bmatrix} \end{bmatrix}$$

Dans l'exemple *Michel donne une pomme à Marie*, les compléments sélectionnés par la tête verbale *donne* sont le sujet, le premier complément et le deuxième complément. Ces compléments sont encodés dans la liste des arguments du verbe sous l'attribut STR-ARG. Nous constatons également leur présence comme valeurs de l'attribut FILLES-COMPS. En *montant* dans la structure pour aboutir à la mère, le SV *donne une pomme à Marie*, les deux compléments sont retirés de la liste de valence, le seul élément qui reste étant le sujet *Michel*. Finalement, au niveau de la phrase, nous aboutissons à un syntagme saturé sans valence et la liste-valeur de l'attribut VAL est vide $< >$. Le principe de la valence permet de retirer de la liste de valence du syntagme tous les arguments réalisés par les filles, de telle manière qu'au niveau de la mère, cette liste soit saturée (notée par la liste vide $< >$). Ces effets sont représentés dans la figure (1.2)[24].

[24] Dans cet arbre, l'on représente la tête par la branche qui porte le symbole T et les compléments par les branches portant le symbole C. Ces symboles illustrent les attributs FILLE-TÊTE et FILLES-COMPS. Les nœuds terminaux, par exemple *Michel*, sont une expression spatiale de l'attribut PHON et de ses valeurs.

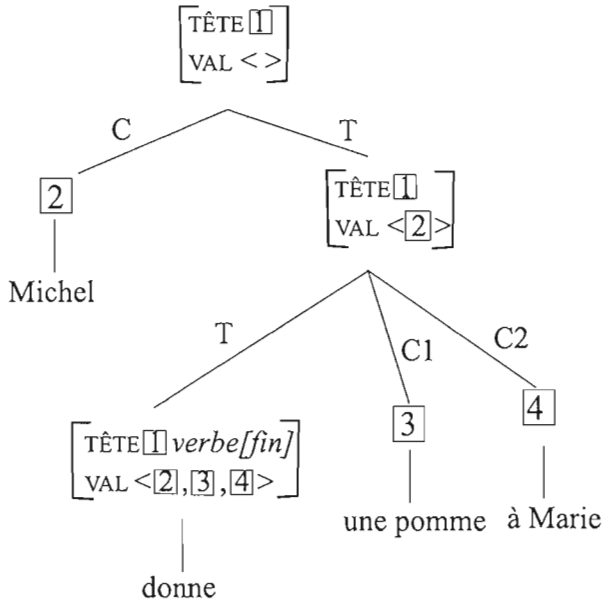


FIGURE 1.2: Une représentation arborescente de *Michel donne une pomme à Marie*

Cet arbre illustre en même temps les effets du principe des traits de tête. La tête *donne* porte la valeur [FIN] qui se propage au SV (sous forme de l'indice 1) et à la phrase (en utilisant le même indice que celui sur le SV).

1.3.5.4 Le principe généralisé des traits de tête

Ce principe a fait son apparition dans les derniers développements de la théorie et surtout dans l'ouvrage de Ginzburg et Sag (2000) et a été employé dans des travaux comme celui de Alexopoulou et Kolliakou (2002). Formulé pour simplifier la grammaire et permettre aux contraintes d'être redéfinies en annulant l'effet des principes généraux (tels que le principe de traits de tête), ce principe introduit une contrainte violable sur les structures à tête. Le fait que cette contrainte soit violable permet à d'autres principes de défaire ses effets à un niveau plus bas de la hiérarchie des types syntagmatiques.

Ce principe requiert que la valeur de l'attribut SYNSEM d'un syntagme à tête et la valeur du

même attribut de sa fille tête soient identiques par défaut, comme l'illustre cette structure de traits[25] :

(1.13) $\begin{bmatrix} \textit{str-tête} & \\ \text{SYNSEM} & /\boxed{1} \\ \text{FILLE-TÊTE} & \begin{bmatrix} \text{SYNSEM} & \boxed{1} \end{bmatrix} \end{bmatrix}$

Cette contrainte peut ainsi être défaite par les contraintes imposées aux sous-types de *str-tête*. Parmi ces derniers la théorie reconnaît, par exemple, les types *str-tête-comp*, *str-tête-suj*, *str-tête-spr*, etc.

1.3.6 Les schémas de dominance immédiate

Un quatrième principe universel joue un rôle important dans la théorie. Il s'agit du principe de dominance immédiate (ID) énoncé comme suit par Pollard et Sag (1994, p. 399)[26] : « Chaque syntagme de type *str-tête* doit satisfaire un [seul] schéma de dominance immédiate. »

En effet, si les entrées lexicales décrivent des structures de traits de type *mot*, les schémas de dominance immédiate (désormais schémas DI) sont une description des structures de traits de type *syntagme*. Nous pouvons également les considérer comme des contraintes disjonctives sur la constitution immédiate des syntagmes (*Ibid.*, p. 38). Elles sont disjonctives en ceci qu'un seul schéma DI à la fois peut décrire la constitution d'un syntagme donné.

Le nombre de ces schémas est limité, ne dépassant pas six, mais leur couverture empirique est très large. Ils peuvent ainsi décrire les relations structurales entre la tête, les différents compléments, les marqueurs, les adjoints et les éléments qui entretiennent des relations à longue distance. Nous présentons ici les six schémas proposés par Pollard et Sag (1994) : le schéma tête-sujet (schéma DI1), le schéma tête-complément (schéma DI2), le schéma tête-sujet-complément (schéma DI3), le schéma tête-marqueur (DI4), le schéma tête-adjoint (DI5) et le schéma tête-*filler* (DI6).

[25] La notation « / » indique une contrainte par défaut que d'autres contraintes peuvent annuler.
[26] Traduction libre.

1.3.6.1 Schéma DI1

Ce schéma légitime un signe syntagmatique saturé dont la valeur de l'attribut FILLE-TÊTE est un syntagme et la valeur de l'attribut FILLES-COMPS est un singleton (une liste constituée d'un seul élément). Il est l'équivalent HPSG du schéma X-barre en ceci qu'il légitime un syntagme avec deux filles : la première est une tête syntagmatique et la deuxième un complément. Ce schéma[27] décrit en fait des syntagmes saturés dont les constituants immédiats sont une tête et un sujet ou une tête et un spécifieur Dans le premier cas, le schéma décrit la constitution d'une phrase qui comporte un sujet préverbal (dans l'ordre SV). Dans le deuxième cas, le schéma décrit la constitution d'un SN (où la tête sous-catégorise un déterminant).

(1.14)
$$\begin{bmatrix} \textit{str-tête} & \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{VAL} & \langle\rangle \end{bmatrix} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} & \begin{bmatrix} \textit{syntagme} & \\ \text{SS} & \begin{bmatrix} \text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{VAL}\,| & (\text{SUJ} \vee \text{SPR}) & \boxed{2} \\ \text{TÊTE} & & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{FILLES-COMPS} & \left\langle \boxed{2} \right\rangle \end{bmatrix} \end{bmatrix}$$

Ce schéma est également présentable sous la forme d'un arbre où l'on précise les branches tête (T) et complément (C), par exemple pour la phrase *Michel ramasse des fleurs*. Voir la figure (1.3).

1.3.6.2 Schéma DI2

Ce schéma légitime un signe syntagmatique non saturé (ayant satisfait toutes les exigences de sous-catégorisation à l'exception d'une seule, celle qui concerne le sujet ou le spécifieur), et dont la valeur de l'attribut FILLE-TÊTE est un signe lexical. Ce schéma peut en fait décrire la

[27] La disjonction est marquée par le signe V.

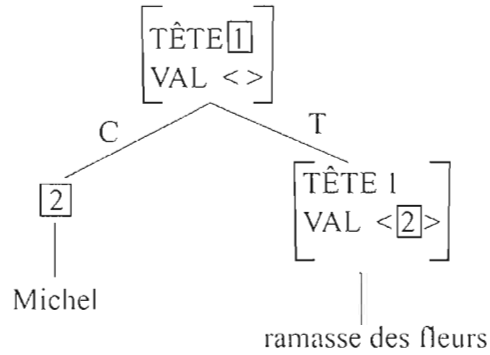


FIGURE 1.3: Le schéma de DI 1

constitution du syntagme verbal en français, par exemple. Ce dernier est un syntagme non saturé ayant une tête lexicale non syntagmatique (le verbe) et sous-catégorisant un SN sujet.

(1.15)

$$\begin{bmatrix} \textit{str-tête} & \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{VAL}\,|\,\text{(SUJ V SPR)} & \boxed{2} \end{bmatrix} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} & \begin{bmatrix} \textit{mot} & \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{(SUJ V SPR)} & \boxed{2} \\ \text{COMPS} & \left\langle \boxed{3} \ldots \boxed{n} \right\rangle \end{bmatrix} \\ \text{TÊTE} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{FILLES-COMPS} & \left\langle \boxed{3} \ldots \boxed{n} \right\rangle \end{bmatrix} \end{bmatrix}$$

Suivant ce schéma, nous pouvons présenter le syntagme verbal *ramasse des fleurs* dans l'arbre en (1.4).

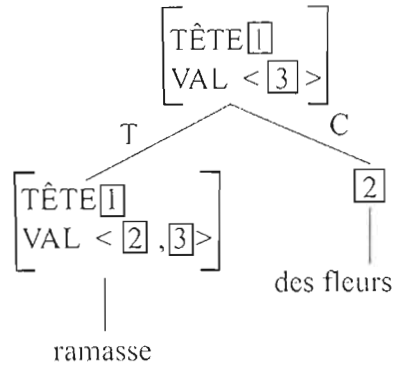


FIGURE 1.4: Le schéma DI 2

1.3.6.3 Schéma DI3

Ce schéma décrit un signe syntagmatique saturé [VAL <>] dont la tête est un signe lexical (de type *mot*). Dans les premières versions de la théorie, ce troisième schéma décrivait à la fois l'inversion du sujet et de l'auxiliaire en anglais et les structures à verbe initial dans certaines langues, comme le gallois. Cependant, les révisions apportées à la théorie dans sa troisième version, et sous l'influence des travaux de Robert Borsley et surtout Borsley (1987), ont mis en doute la validité de ce schéma pour décrire les structures appelées VSO. L'arabe étant une langue à verbe initial, comme le gallois, nous pensons qu'il faudrait vérifier la validité de ce schéma pour décrire ces structures.

(1.16)
$$\begin{bmatrix} \textit{str-tête} & \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{VAL} & \langle\rangle \end{bmatrix} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} & \begin{bmatrix} \textit{mot} & \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{VAL}\,|\,\text{COMPS} & \left\langle \boxed{2} \ldots \boxed{n} \right\rangle \\ \text{TÊTE} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{FILLES-COMPS} & \left\langle \boxed{2} \ldots \boxed{n} \right\rangle \end{bmatrix} \end{bmatrix}$$

La représentation arborescente de la phrase anglaise avec inversion *can Kim go ?* se trouve dans la figure (1.5).

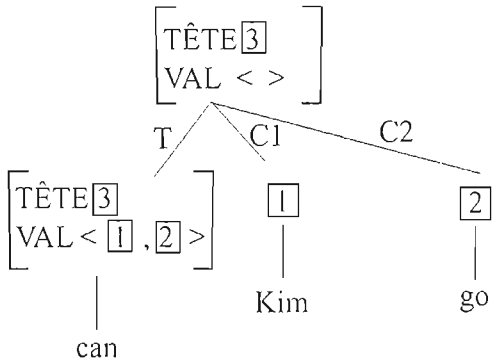


FIGURE 1.5: Le schéma DI 3

1.3.6.4 Schéma DI4

Pour comprendre ce schéma, il faut commencer par introduire le traitement réservé aux complémenteurs dans cette grammaire. Ces derniers sont modélisés par un attribut particulier MARQUEUR. Dans cette conception, les complémenteurs sont considérés comme des signes saturés [VAL < >] avec, comme valeur de l'attribut TÊTE, une structure de traits de type *marqueur* sélectionnant comme spécifieur un objet de type *phrase* à condition que cette phrase ne soit pas marquée (n'étant pas elle-même précédée d'un autre complémenteur). Les complémenteurs ont également un attribut particulier MARQUAGE dont la valeur est la forme phonologique du complémenteur. Ces contraintes sur le type *marqueur* sont exemplifiées en (1.17) pour le complémenteur anglais *that* (Pollard et Sag, 1994, p. 46).

(1.17)
$$\begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{marqueur} & \\ \text{SPEC} & \textit{phrase[non-marquée]} \end{bmatrix} \\ \text{VAL} & \langle\rangle \\ \text{MARQUAGE} & \textit{that} \end{bmatrix}$$

Le quatrième schéma de dominance immédiate légitime les syntagmes de type *str-tête-marqueur* et décrit justement les phrases introduites par un complémenteur. Le syntagme en question doit

comporter une fille marqueur (notée M dans le schéma) dont la valeur de l'attribut SPR est partagée avec la valeur de l'attribut SYNSEM de la fille-tête et dont la valeur de l'attribut MARQUAGE est partagée avec celle de la mère. L'arbre en (1.6) est une représentation de la complétive française *que Michel éternue* selon le schéma DI4.

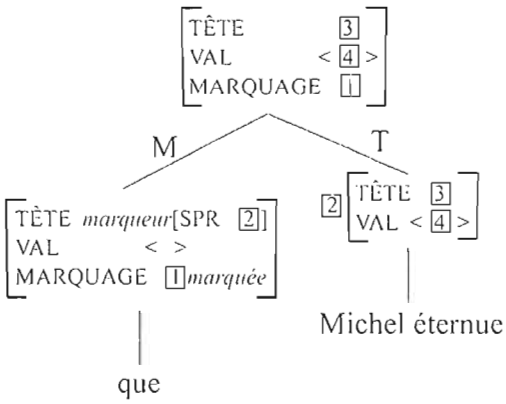


FIGURE 1.6: Le schéma DI 4

1.3.6.5 Schéma DI5

Ce schéma introduit le traitement réservé par la théorie aux adjoints. Pollard et Sag (1994, p. 55) considèrent, comme c'est le cas dans la grammaire catégorielle, que les adjoints sélectionnent la tête avec laquelle ils forment une structure particulière appelée structure tête adjoint, dont le type est *str-tête-adjoint*, un sous-type de *str-tête* dont la valeur de l'attribut FILLES-COMPS est spécifiée < > et qui a un attribut supplémentaire FILLE-ADJOINT. Pour que l'adjoint puisse sélectionner la tête, les auteurs introduisent un nouvel attribut MOD(ifié), un attribut réalisé sous l'attribut TÊTE. La valeur de cet attribut est partagée avec la valeur de l'attribut SYNSEM de la tête pour refléter cette relation de sélection.

Dans cette perspective, le schéma DI5 légitime les syntagmes de type *str-tête-adjoint* et impose la contrainte suivante sur ces syntagmes : la valeur de l'attribut MOD de la fille adjoint doit être

identique à la valeur de l'attribut SYNSEM de la fille tête. La représentation arborescente dans la figure (1.7) du SN français *petite fille* est un exemple de l'application de ce schéma[28].

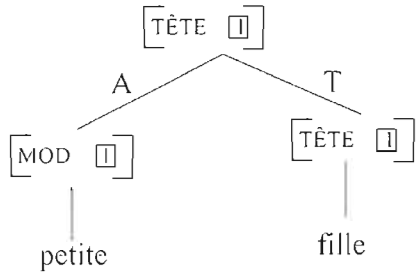


FIGURE 1.7: Le schéma DI 5

1.3.6.6 Schéma DI6

Ce schéma introduit le traitement réservé par la grammaire aux phénomènes appelés d'extraction ou encore les constructions à dépendance non bornée selon Pollard et Sag (1994). Ces constructions englobent, entre autres, la topicalisation, les questions wh et les relatives. Les exemples suivants illustrent quelques-uns de ces phénomènes[29] :

(1.18) a. Kim, Sandy loves.
b. I wonder who Sandy loves.
c. It's Kim who Sandy loves.

Ces constructions sont dites « filler-gap » en référence à la présence d'un élément extrait en quelque sorte en laissant un trou syntagmatique (le « gap ») dont le contenu est identifié grâce à la présence d'un autre élément appelé ici « filler[30] ». La relation entre ces deux éléments est

[28] La branche A est une représentation de la fille adjoint.
[29] (Pollard et Sag, 1994, p. 157)
[30] Les constructions appelées « filler-gap » ne représentent en réalité qu'un seul groupe des dépendances non bornées. Le deuxième groupe englobe des structures où le « filler » est un argument alors que ce dernier est non argumental dans les constructions du premier type. Nous ne donnons ici que des exemples des constructions « filler-gap » par souci de concision.

non locale, d'où l'emploi de l'attribut NON-LOCAL, qui reflète cette réalité. Parmi les attributs appropriés à la structure de traits de type *non-local*, l'attribut SLASH encode les traits de l'élément absent de la structure. La valeur de cet attribut est un ensemble (noté par le symbole { }) de traits partagés avec la valeur de l'attribut LOCAL du « filler ». La construction contenant un élément slashé est dite *str-tête-filler*, décrite par le schéma de dominance immédiate 6, dont une forme simplifiée est présentée dans la structure de traits suivante et une représentation arborescente est faite dans la figure (1.8) :

$$(1.19)\quad \begin{bmatrix} \textit{str-tête-filler} & \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-FILLER}\,|\,\text{SS}\,|\,\text{LOCAL} & \boxed{1} \\ \text{FILLE-COMPS} & \langle\rangle \\ \text{TÊTE-FILLE} & \begin{bmatrix} \text{SYNSEM} & \begin{bmatrix} \text{LOCAL} & \boxed{1} \\ \text{NON-LOCAL}\,|\,\text{SLASH} & \left\{\boxed{1}\right\} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

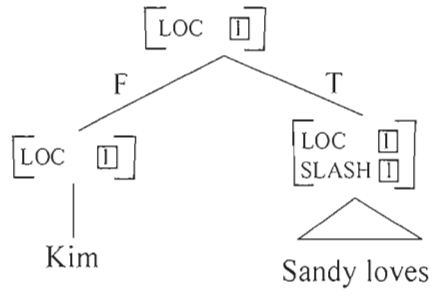


FIGURE 1.8: Le schéma DI 6

En plus de l'ontologie linguistique exprimée dans la hiérarchie des types et des contraintes de bonne formation exprimées dans les principes, les schémas de dominance immédiate définissent les limites de l'universel dans la grammaire HPSG. Cette grammaire laisse néanmoins de la place

à l'expression des spécificités de chaque langue. Ainsi, chaque langue effectue un choix parmi les choix universellement disponibles. La spécificité s'exprime également dans le lexique, où les entrées lexicales peuvent être reliées, produites ou transformées par des règles de redondance lexicale. Nous présentons ce dispositif formel dans le paragraphe suivant.

1.3.7 Les règles lexicales

Le lexique en HPSG est organisé dans une hiérarchie de types. Cette hiérarchie n'est cependant pas capable de rendre compte des opérations morphologiques de base, telles que celles qui impliquent la dérivation d'un mot à partir d'un autre. Pour que le lexique hiérarchisé soit capable de rendre compte des relations morphologiques entre les items lexicaux, la hiérarchie devrait contenir toutes les formes possibles d'un même mot (pensons, par exemple, aux formes conjuguées d'un même verbe). En optant pour une telle solution, on assisterait à l'éclatement du lexique et la hiérarchie ne pourrait exprimer les généralisations et les liens évidents entre les items. Pour remédier à cette situation, la théorie a recours aux règles lexicales[31] pour dériver de nouveaux types à partir des types de base présents dans la hiérarchie. Ces règles modélisent ainsi la productivité lexicale et constituent le locus des relations morphologiques (telles que la dérivation et la flexion). Ainsi,

> Families of related words - such as the different inflectional forms of a verb - can be derived from a single lexical entry by means of lexical rules. (Sag *et al 2003, p. 564)*

Deux remarques s'imposent concernant ces règles à saveur transformationnelle. Tout d'abord, elles n'ont pas cessé d'alimenter le débat entre les linguistes travaillant dans le cadre de HPSG. Si certains auteurs, dont Pollard et Sag (1994, 1987), Miller et Sag (1997) et Monachesi (1998a), les emploient non seulement pour modéliser les relations morphologiques, mais également pour relier des structures syntaxiques, d'autres, dont Krieger et Nerbonne (1993), Koenig (1999) et Smets (2000), trouvent que leur utilisation est une modélisation non optimale des relations lexicales.

Ensuite, leur conception a grandement évolué. Pour Pollard et Sag (1987), elles prenaient la forme d'opérations reliant deux structures de traits. La règle lexicale suivante, par exemple, modélise la dérivation d'une forme verbale à la 3e personne du singulier (3S) en anglais :

(1.20) La règle lexicale de la 3S

[31] Ces règles sont également dites *règles de redondance lexicale*, comme on peut lire dans Pollard et Sag (1987, p. 209). Ce terme garde vivant l'origine du concept, redevable à Jackendoff (1975).

$$3S : \begin{bmatrix} base & \\ \text{PHON} & \boxed{1} \\ \text{3S} & \boxed{2} \\ \text{SS} & \begin{bmatrix} \text{LOC} \mid \text{CAT} \mid \text{VAL} & \boxed{3} \\ \text{CONT} & \boxed{4} \end{bmatrix} \end{bmatrix} \rightarrow \begin{bmatrix} 3S & \\ \text{PHON} & f_{3S}\ (\boxed{1}, \boxed{2}) \\ \text{SS} & \begin{bmatrix} \text{LOC} \mid \text{CAT} \mid \text{VAL} & \boxed{3}\ cont\ \boxed{4} \end{bmatrix} \end{bmatrix}$$

Cette règle repose sur une fonction phonologique f_{3S} qui prend deux arguments : la forme phonologique du verbe et la désinence de la 3S (ici *-s*). Cette fonction produit le mot *gives* à partir du lexème[32] *give*. Il va de soi que cette fonction sera redéfinie différemment pour les verbes réguliers et les verbes irréguliers.

Les règles lexicales, telles que conçues dans HPSG I et HPSG II ne sont en fin de compte que des opérations formelles affectant des structures de traits. Cette conception a été critiquée par plusieurs auteurs, dont Krieger et Nerbonne (1993), Koenig (1999) et Smets (2000). Krieger et Nerbonne (1993), par exemple, défendent l'idée selon laquelle les règles lexicales doivent être décrites par des structures de traits et non en tant qu'opérations sur ces dernières. Les structures de traits restent de ce fait l'outil de description et de représentation des connaissances par excellence dans la théorie : « Where Pollard and Sag viewed lexical rules as operators on feature structures, we propose defining lexical rules purely in terms of feature structures. » (*Ibid.*, p. 92)

Koenig (1999) propose d'éliminer les règles lexicales parce qu'elles posent deux problèmes : un problème de parcimonie de la théorie et un problème de principe. En ce qui a trait à la parcimonie de la théorie, l'auteur pense que ces règles nécessitent souvent des ajouts *ad hoc* pour modéliser la productivité lexicale et leur coût théorique devient donc trop élevé. Le deuxième problème posé par ces règles est celui de la dépendance à l'ordre, une dépendance qui mène dans certains cas à des paradoxes quand deux règles doivent être appliquées l'une à la suite de l'autre (en relation d'alimentation).

Prenant en compte les problèmes soulevés par ces auteurs, Sag *et al.* (2003) ont grandement transformé la conception des règles lexicales en les intégrant entièrement dans le système grammatical. Ces règles s'expriment désormais de manière déclarative comme des contraintes sur un type donné. Étant des contraintes, elles peuvent prendre la forme d'une structure de traits dont les attributs principaux sont INTRANT et EXTRANT. Les règles lexicales sont ainsi conçues comme

[32] Le lexème est conçu comme étant l'unité abstraite qui donne naissance à plusieurs mots reliés entre eux. Ainsi, le lexème *marcher* donne lieu aux mots *marche, marches, marchions, marchaient, eussent marché*, etc. reliés au premier grâce aux règles lexicales de flexion.

des objets linguistiques d'un type donné *r-lexicale* et ce type a 3 sous-types : *r-deriv* (règles dérivationnelles, reliant un lexème à un lexème), *r-flex* (règles flexionnelles, reliant un lexème à un mot) et *r-pflex* (règles post-flexionnelles, reliant un mot à un mot). Ainsi, la règle lexicale en (1.20) appartient au 2e sous-type *r-flex* et prend la forme suivante dans le modèle défendu par Sag *et al.* (2003)[33] :

(1.21) La règle lexicale de la 3S

$$\begin{bmatrix} \textit{r-flex} & \\ \text{INTRANT} & \left\langle \boxed{1}, \begin{bmatrix} \textit{lxm-verbal} \end{bmatrix} \right\rangle \\ \text{EXTRANT} & \left\langle \text{F}_{3S}(\boxed{1}), \begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{TÊTE} & \begin{bmatrix} \text{FORM} & \textit{fin} \\ \text{AGR} & \textit{3S} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Notons, dans cet AVM, l'introduction des deux traits de tête [FORM] et [AGR]. Le premier modélise les traits aspectuels et temporels du verbe alors que le deuxième modélise ses traits d'accord. Un nouveau type est également introduit : *lxm-verbal* pour lexème verbal, qui modélise la forme lexémique du verbe avant toute flexion.

Nous employons, dans notre travail, des règles lexicales déclaratives dont la forme est semblable à celle présentée en (1.21). Ces règles sont ainsi conçues non comme des opérations affectant des structures de traits, mais comme des structures de traits à part entière obéissant à des contraintes de bonne formation rencontrées ailleurs dans la grammaire.

La grammaire syntagmatique HPSG fournit non seulement un cadre formel solide pour la modélisation des phénomènes linguistiques en vue d'une implémentation dans les systèmes de performance, mais également et surtout une théorie linguistique qui accorde une grande importance à la validation empirique des analyses adoptées. Ces caractéristiques s'observent dans les analyses de différents phénomènes syntaxiques, comme les dépendances à longue distance ou encore l'emplacement des clitiques. Ce dernier problème a fait l'objet d'un grand nombre de travaux dans ce cadre. Nous y consacrons la section suivante étant donné ses liens avec notre sujet de

[33] Certains détails sont omis de cette présentation, par exemple les traits sémantiques préservés entre l'intrant et l'extrant. Voir pour plus de détails (Sag *et al.*, 2003, p. 256-)

recherche.

1.4 La modélisation des clitiques

Les clitiques pronominaux les plus étudiés dans le cadre de HPSG sont les clitiques des langues romanes. Ces derniers sont caractérisés par un certain nombre de propriétés singulières que nous présentons ici.

1. Le statut morphologique : les clitiques pronominaux dans les langues romanes sont une catégorie à part à cause de leur comportement morphologique spécial. En effet, ils ont à la fois les propriétés des mots indépendants et celles des affixes.
2. Le statut syntaxique : les clitiques pronominaux en français, par exemple, peuvent apparaître dans des positions différentes des SN auxquels ils correspondent. Cette propriété fait d'eux des « clitiques spéciaux » au sens de Zwicky (1977).
3. Le statut phonologique : les clitiques pronominaux dépendent prosodiquement d'autres éléments ; ce sont des éléments non accentués qui ont besoin d'un hôte fort. Qui plus est, l'ordre de réalisation des clitiques dans une suite est extrêmement idiosyncrasique et semble ne pas être explicable par un principe indépendant de la grammaire.

Ces propriétés singulières ont fait de ces clitiques un objet d'étude très controversé et ont donné lieu à des questions qui ne cessent d'alimenter les débats théoriques : dans quel composant de la grammaire sont formés ces clitiques ? Est-ce la syntaxe, la morphologie (ou le lexique) ou la phonologie ? Et quel impact aura ce choix sur la théorie en général ? Et si ces clitiques n'étaient en fin de compte que des éléments d'interface qui impliquent toutes les composantes de la grammaire ? Quelle est donc la relation entre ces modules (dans une conception modulaire) ? Doit-on restreindre la portée des opérations syntaxiques et doit-on, par contre, leur permettre de « voir » à l'intérieur des mots ? Etc.

Bien que l'étude de ces clitiques soit depuis longtemps un sujet de choix dans la grammaire générative chomskyenne, elle n'a commencé à intéresser les linguistes de HPSG qu'après la publication de la thèse de Philip Miller [34], qui se situait encore dans le cadre de la grammaire GPSG, mais présentait des perspectives intéressantes pour les grammaires à base de contraintes et pour les théories qui se veulent lexicalistes. En fait, tous les travaux que nous allons présenter sont une adaptation des idées de Miller au cadre de la grammaire HPSG, ce qui est fait, de plus, par Miller

[34] Voir Miller (1992).

lui-même dans Abeillé *et al.* (1998) et Miller et Sag (1997, 1995). Des versions antérieures de ces travaux circulaient dans la communauté scientifique depuis 1993 et ont inspiré Paola Monachesi, qui a proposé une analyse lexicaliste des clitiques italiens qui s'apparente à celle de Philip Miller dans Monachesi (1999a, 1998a,b). Paola Monachesi a également essayé d'appliquer ce genre d'analyse au roumain et a obtenu des résultats semblables dans Monachesi (1999b). La réussite qu'ont connue ces travaux sur les clitiques des langues romanes était à l'origine des tentatives d'application sur d'autres familles de langues. C'est le cas du polonais (de la famille slave) dans Kupsc (2000, 1999) et du serbo-croate (également de la famille slave) dans Penn (1999a,b).

Nous commencerons par présenter les grandes lignes de l'analyse lexicaliste des pronoms clitiques défendue dans le cadre de HPSG, essentiellement pour les langues romanes dans la section (1.4.1). Cette analyse repose sur l'hypothèse suivante : les clitiques sont des affixes flexionnels, par conséquent ils trouvent leur place dans la morphologie (ou le lexique) plutôt que dans la syntaxe. Nous présentons la particularité des analyses proposées pour les clitiques d'autres familles que la famille romane dans la section (1.4.2).

1.4.1 Une analyse lexicaliste au sens fort

Le lexicalisme au sens fort repose sur une approche selon laquelle les mots sont les atomes de la syntaxe et qu'ils sont, par conséquent, invisibles aux opérations syntaxiques et aux contraintes qui opèrent dans la syntaxe. Les mots sont en fait traités dans le lexique et comportent toutes les informations nécessaires pour former les syntagmes et les phrases selon des principes universels et des schémas de dominance qui définissent les constructions bien formées. Cette orientation lexicaliste est expliquée par Miller et Monachesi (2003) comme suit :

> Nous adopterons l'hypothèse lexicaliste, selon laquelle les mots morphologiquement complexes sont construits dans le lexique par la morphologie et insérés tels quels dans les constructions syntaxiques.

Les clitiques sont, dans cette perspective, construits dans le lexique par la morphologie et insérés tels quels sur les têtes verbales avant que ces dernières ne soient *insérées* dans des structures syntagmatiques. Le statut qui leur confère cette possibilité est leur affixalité. Les clitiques des langues romanes sont considérés comme des affixes flexionnels créés dans le lexique et insérés dans la matrice du verbe comme information morphologique (ou morphosyntaxique). Miller et Sag (1995) affirment :

> Les soi-disant « pronoms clitiques » du français sont en fait des affixes flexionnels lexicalement attachés et non des clitiques post-lexicaux (c'est-à-dire des mots

indépendants dont les propriétés prosodiques sont telles qu'ils sont rattachés postsyntaxiquement à un mot environnant).

Les arguments pour soutenir une telle thèse proviennent en grande partie de l'application des critères de Zwicky [35] et des modifications proposées à ces critères par Miller (1992). Ainsi, le degré de sélection de l'hôte, les trous arbitraires dans l'ensemble des combinaisons possibles, les idiosyncrasies morphophonologiques, l'ordre rigide dans les suites, l'applicabilité des règles lexicales phonologiques et la coordination semblent démontrer que les clitiques des langues romanes sont mieux traités comme affixes.

La majorité des travaux qui ont traité les clitiques des langues romanes dans ce cadre théorique s'accordent sur l'analyse linguistique de ces unités et adoptent l'hypothèse affixale. Ils diffèrent cependant sur la formalisation de cette hypothèse et par conséquent sur les outils employés pour implémenter cette analyse dans le cadre de HPSG. Nous pouvons ainsi grouper ces travaux en deux classes : dans la première, nous citons les travaux qui proposent un traitement en termes de règles lexicales. Ces travaux constituent les premières tentatives dans le cadre de HPSG de formaliser les clitiques pronominaux des langues romanes. Cette étape était surtout marquée par les travaux suivants : Abeillé *et al.* (1998), Miller et Sag (1995) et Monachesi (1998a,b). Dans le deuxième groupe, nous trouvons un seul travail, celui de Miller et Sag (1997), dans lequel les auteurs proposent une implémentation exempte de règles lexicales et entièrement formulée en termes de contraintes sur des structures de traits typées.

1.4.1.1 Une analyse basée sur les règles lexicales

Nous prenons un exemple représentatif pour montrer les propriétés d'une formalisation qui repose sur les règles lexicales : l'article de Miller et Sag (1995). La modélisation proposée dans cet article essaie de capter la relation empirique entre l'extraction des compléments et la cliticisation en français [36]. Pour ce faire, les auteurs proposent deux règles lexicales en relation d'alimentation : l'extrant de la première, la règle lexicale de l'extraction des compléments en (1.22), sert d'intrant pour la deuxième, la règle lexicale de l'affixation des compléments en (1.23).

(1.22) La règle lexicale de l'extraction des compléments

[35] Nous présentons ces tests dans le quatrième chapitre.

[36] Cette relation se voit dans le cas des pronoms clitiques objets de cette langue, qui subissent la montée syntaxique.

$$\begin{bmatrix} \text{COMPS} & \left\langle \ldots, \boxed{2}\left[\text{LOC } \boxed{3}, \text{ SLASH } \{\}\right], \ldots \right\rangle \\ \text{STR-ARG} & \left\langle \ldots, \boxed{2}, \ldots \right\rangle \\ \text{SLASH} & \boxed{4} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{COMPS} & \left\langle \ldots, \ldots \right\rangle \\ \text{STR-ARG} & \left\langle \ldots, \boxed{2}\left[\text{LOC } \boxed{3}, \text{ SLASH } \left\{\boxed{3}\right\}\right], \ldots \right\rangle \\ \text{SLASH} & \left\{\boxed{3}\right\} \cup \boxed{4} \end{bmatrix}$$

(1.23) La règle lexicale de l'affixation des compléments

$$\begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{SLASH} & \textit{S2} \\ \text{PRAFS} & \textit{S1} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{SLASH} & \textit{S2}/\left[\boxed{1}\right] \\ \text{PRAFS} & \textit{S1} \cup \left\{\boxed{1}\right\} \end{bmatrix} \text{Où } \boxed{1} \in \text{S2}$$

La règle lexicale de l'extraction des compléments intervient pour dériver les entrées verbales dont les compléments sont extraits. Elle enlève un élément de la valeur de l'attribut COMPS et le place comme valeur de l'attribut SLASH. Dans cet environnement intervient la règle lexicale de l'affixation des compléments, qui enlève un élément de la liste SLASH et le place comme valeur dans la liste PRAFS [37].

Deux remarques s'imposent ici. Premièrement, cette analyse reprend une idée qui a vu le jour dans Miller (1992), selon laquelle il existe un rapport entre la cliticisation et les phénomènes de l'extraction en français. Cette hypothèse, qui fait de la cliticisation un exemple des dépendances non locales (appelées également dépendances non bornées), est contestée par Abeillé *et al.* (1998), qui ont démontré que les clitiques impliquent des phénomènes de dépendance locale bornée, même dans les cas de montée, où le clitique semble ne pas être localement lié à son hôte.

Deuxièmement, les deux règles lexicales proposées ne concernent que les clitiques objets; les auteurs ne fournissent aucune analyse des clitiques sujets. Par contre, Miller et Sag (1997) fournissent un traitement en termes de contraintes qui prend en compte toutes les sortes de clitiques pronominaux du français, dont les clitiques sujets.

1.4.1.2 Une analyse basée sur les contraintes

Les hypothèses présentées par Miller et Sag (1997) sont basées sur une hiérarchie des types lexicaux conjuguée avec des contraintes sur ces types et avec des principes universels de bonne

[37] Cet attribut modélise les objets de type affixal, à savoir ici les pronoms affixes.

formation.

1.4.1.2.1 La hiérarchie des types lexicaux

Cette hiérarchie est adaptée au français et diffère quelque peu de celle proposée par Pollard et Sag (1994, 1987) pour l'anglais. Elle repose, en outre, sur trois classifications des mots fléchis :

- Une classification selon la réalisation du clitique : le type *mot* possède deux sous-types qui sont *pl-wd* (les mots simples) et *cl-wd* (les mots cliticisés). Le type *cl-wd*, pour sa part, possède deux sous-types : *su-cl-wd* (les mots dotés de clitiques sujets) et *ns-cl-wd* (les mots dotés de clitiques non sujets).
- Une classification selon la flexion : ce typage spécifie la forme flexionnelle d'un lexème donné, par exemple *3sg-pres-indic-vb* (verbe fléchi pour la 3e personne du singulier au présent de l'indicatif).
- Une classification du lexème qui définit la base morphologique, la partie du discours, la structure argumentale et le sens commun à une famille de formes fléchies.

À la lumière de cette hiérarchie, et pour rendre compte de la différence entre les verbes hôtes de clitiques et les verbes qui ne le sont pas, les auteurs classent les verbes français en deux types différents :

1. Le type *pl-wd* : les contraintes associées à ce type exigent que chaque élément de la structure argumentale du verbe corresponde à un syntagme qui se combine syntaxiquement avec la tête verbale, il doit par conséquent être présent dans la liste SUJ ou COMPS de ce verbe, en même temps que dans sa liste de ses arguments ARG-ST.
2. Le type *cl-wd* : ce type correspond à des verbes qui ont, au moins, un argument réalisé morphologiquement (en forme d'affixes) plutôt que syntaxiquement. Ces arguments apparaissent alors dans la liste ARG-ST, mais pas comme valeurs des attributs SUJ ou COMPS.

Rien ne garantit, toutefois, la présence des affixes appropriés dans la forme phonologique des mots cliticisés (*cl-wd*). Pour résoudre ce problème, les auteurs proposent une classification des objets SYNSEM en sous-types selon deux critères : la canonicité et l'affixalité. Ces sous-types sont présentés dans la figure (1.9).

Les *synsem* canoniques sont le type correspondant à tous les signes linguistiques « canoniques ». Les *synsem* non canoniques correspondent à des membres de la structure argumentale qui ne sont pas réalisés comme des dépendants syntaxiques locaux de la tête. Les sous-types

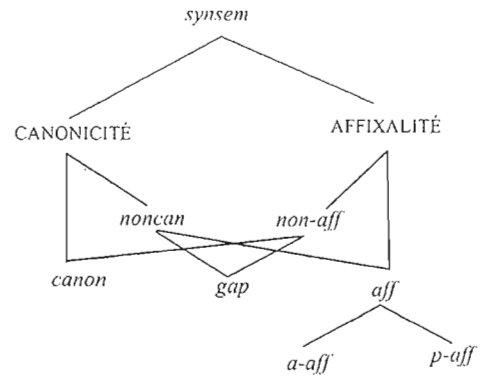


FIGURE 1.9: Une hiérarchie partielle des types proposée par Miller et Sag (1997)

regroupés sous ce type sont *aff* (affixes) et *gap* (dans les phénomènes d'extraction, par exemple). Ce typage modélise l'insertion des affixes (c'est-à-dire les clitiques) dans les formes verbales. Il suffit, en fait, qu'il existe des éléments de type *aff*[38] dans la liste argumentale (STR-ARG) du verbe pour déclencher la réalisation des affixes pronominaux qui correspondent à l'information morphologique contenue dans l'entrée lexicale du verbe. Dans cette conception, le passage de mots à des signes syntagmatiques bien formés est assuré par le principe des traits de tête et le principe de la valence, présentés dans les sections (1.3.5.2) et (1.3.5.3).

Ces deux principes, en concordance avec la hiérarchie des types lexicaux, modélisent la réalisation des clitiques pronominaux sur les têtes verbales. D'autres contraintes s'ajoutent à ces deux principes et au typage lexical, des contraintes plus spécifiques sur les types *cl-wd* et ses deux sous-types (*su-cl-wd* et *ns-cl-wd*).

1.4.1.2.2 Les contraintes spécifiques

Miller et Sag (1997) proposent les contraintes suivantes sur le type *cl-wd* (1.24) :

[38] Les deux types *p-aff* (pronoms personnels affixes) et *a-aff* (affixes anaphoriques) servent à rendre compte des phénomènes du liage. En effet, quand la morphologie réflexive est présente sur le verbe, l'un des membres de la structure argumentale de ce verbe doit être réalisé comme un élément *a-aff*, d'où la possibilité d'appliquer le principe A de la théorie du liage. Par contre, quand le verbe contient, dans sa structure argumentale, un élément de type *p-aff*, c'est le principe B qui est en vigueur. La conception de ces deux principes dans le cadre de la théorie HPSG est présentée dans l'ouvrage de Pollard et Sag (1994).

(1.24) Contraintes sur le type *cl-wd*

$$cl\text{-}wd \Rightarrow \begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & F_{PRAF}\ (\boxed{0}, \ldots) \\ \text{I-FORM} & \boxed{0} \end{bmatrix} \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \boxed{2} \\ \text{COMPS} & \boxed{3}\text{liste(non-aff)} \end{bmatrix} \\ \text{STR-ARG} & (\boxed{2} \oplus \boxed{3})\ \boxed{0}\text{nelist(aff)} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Ces contraintes garantissent ce qui suit :

- Que tous les membres de la liste COMPS ne sont pas des affixes puisque la valeur de cet attribut est une liste (potentiellement vide) d'éléments du type *non-aff*. Cela évite aux éléments affixaux d'apparaître dans la liste de valence ; ils ne sont encodés que dans la liste-valeur de la structure argumentale.
- Que la valeur de l'attribut STR-ARG comporte une liste non vide (*nelist*) d'éléments affixaux (un ou plusieurs clitiques).
- Que la phonologie des *cl-wd* est déterminée par la fonction F_{praf}, qui exige que la valeur de l'attribut FORM soit reliée à la valeur de l'attribut I-FORM (la forme flexionnelle) via l'affixation pronominale appropriée.

Les auteurs proposent également les deux contraintes suivantes sur les sous-types *su-cl-wd* (1.25) et *ns-cl-wd* (1.26) :

(1.25) Contraintes sur le type *su-cl-wd*

$$su\text{-}cl\text{-}wd \Rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{VAL} \mid \text{SUJ} & \langle\rangle \\ \text{STR-ARG} & \left\langle [\text{aff,nom}], \ldots \right\rangle \end{bmatrix} \end{bmatrix}$$

(1.26) Contraintes sur le type *ns-cl-wd*

$$su\text{-}cl\text{-}wd \Rightarrow \begin{bmatrix} \text{SYNSEM} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{VAL} \mid \text{SUJ} & \boxed{1} \\ \text{STR-ARG} & \left\langle \boxed{1}, \ldots \right\rangle \end{bmatrix} \end{bmatrix}$$

Ces contraintes garantissent :

- Que le type *su-cl-wd* ait une liste SUJ vide et un premier argument de type *aff* qui porte le cas nominatif, sans exclure la possibilité que d'autres arguments soient de ce type affixal.

- Que le premier argument de type *ns-cl-wd* apparaisse sur la liste SUJ. Cet élément peut être soit de type *canon* soit de type *gap* (correspondant dans ce cas au sujet extrait). De plus, et parce que le type *ns-cl-wd* hérite les contraintes du type supérieur *cl-wd*, il doit contenir au moins un élément de type *aff* dans sa structure argumentale.

Ces contraintes, en harmonie avec les contraintes universelles sur la formation des syntagmes à partir des informations contenues dans les entrées lexicales des mots, modélisent les relations qu'entretient le clitique avec son hôte. Nous exposons la manière dont interagissent ces composantes formelles à partir d'un exemple concret dans la section suivante.

1.4.1.2.3 Un exemple d'application

En combinant les contraintes sur le type *cl-wd* et l'entrée lexicale du verbe *donner* en (1.27), l'entrée lexicale obtenue est celle en (1.28) pour *lui-donnera* :

(1.27) L'entrée lexicale de *donner*

$$\begin{bmatrix} \textit{pl-wd \& donner} & \\ \text{MORPH} & \begin{bmatrix} \text{FORM} & \textit{donner} \\ \text{I-FORM} & \textit{donner} \end{bmatrix} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{verbe} & \\ \text{VFORM} & \textit{infin} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2}, \boxed{3} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \boxed{1}\text{SN}[\text{NOM}], \boxed{2}\text{SN}[\text{ACC}], \text{SN}[\text{À}1] \rangle \end{bmatrix} \end{bmatrix}$$

(1.28) L'entrée lexicale de *lui-donnera*

$$\begin{bmatrix} \textit{ns-cl-wd \& donner \& 3sg-fut-indic-vb} & \\ \text{MORPH} & \begin{bmatrix} \text{FORM} & \textit{lui-donnera} \\ \text{I-FORM} & \textit{donnera} \end{bmatrix} \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT} & \begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{verbe} & \\ \text{VFORM} & \textit{indic} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \boxed{1}\textit{SN[3sg]}, \boxed{2}\textit{SN[acc]}, \textit{SN[p-aff]}, \textit{à1}, \textit{3sg} \rangle \end{bmatrix} \end{bmatrix}$$

Dans cette structure, le complément du verbe est obligatoirement réalisé comme clitique attaché à la tête verbale. Il apparaît seulement dans la liste ARG-ST et non comme valeur de l'attribut COMPS. Les contraintes en (1.28) exigent que le premier élément de la liste ARG-ST doive apparaître dans la liste SUJ (l'élément indicé 1). Cette entrée respecte également une contrainte héritée du type supérieur *cl-wd*, à savoir l'appartenance d'au moins un argument au type *aff*.

Ces contraintes modélisent, d'une manière élégante, la réalisation des clitiques sur les têtes verbales, mais ne nous informent pas sur la réalisation morphologique du verbe. Qu'est-ce qui garantit que la forme morphologique obtenue pour l'entrée lexicale en (1.28), par exemple, est *lui-donnera* et non pas **donnera-lui* ? La théorie a besoin d'autres outils pour implémenter ces faits.

1.4.1.2.4 La réalisation morphologique

La fonction F_{PRAF} contraint la formation des mots de type *cl-wd*. C'est une fonction à trois arguments : la valeur de l'attribut I-FORM fournie par le type flexionnel, la valeur de l'attribut TÊTE du verbe et la valeur de la liste ARG-ST du mot. Miller et Sag (1997) présentent cette structure en (1.29) :

(1.29) Contraintes sur le type *cl-wd* 2

$$cl\text{-}wd \Rightarrow \begin{bmatrix} \text{MORPH} & \begin{bmatrix} \text{FORM} & F_{PRAF}(\boxed{0},\boxed{1},\boxed{2}) \\ \text{I-FORM} & \boxed{0} \end{bmatrix} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{STR-ARG} & \boxed{2}\ \textit{aff-liste} \end{bmatrix} \end{bmatrix}$$

Les valeurs de l'attribut FORM définies par la fonction F_{PRAF} sont des objets de type *cl-fm* (forme clitique). Ce type possède deux sous-types : *procl-fm* (pour les proclitiques) et *encl-fm* (pour les enclitiques). Ces structures de traits spécifient l'information en termes de traits de l'attribut BASE, dont la valeur est une forme fléchie (valeur d'I-FORM) et sept attributs de « position » (*slot*) dont les valeurs sont les clitiques (ou bien l'ensemble vide). La structure suivante résume ces valeurs :

(1.30)
$$\begin{bmatrix} \textit{cl-fm} & \\ \text{BASE} & \text{infl-form} \\ \text{SL-1} & \left\{ \textit{je,tu,il,elle,on,nous,vous,ils,elles,}[\ ] \right\} \\ \text{SL-2} & \left\{ \textit{me,te,nous,vous,se,}[\ ] \right\} \\ \text{SL-3} & \left\{ \textit{le,la,les,}[\ ] \right\} \\ \text{SL-4} & \left\{ \textit{lui,leur,}[\ ] \right\} \\ \text{SL-5} & \left\{ \textit{moi,toi,nous,vous,}[\ ] \right\} \\ \text{SL-6} & \left\{ \textit{y,zy,}[\ ] \right\} \\ \text{SL-7} & \left\{ \textit{en,zen,}[\ ] \right\} \end{bmatrix}$$

Pour modéliser l'ordre des clitiques, Miller et Sag (1997) proposent des contraintes sur les types *procl-fm* et *encl-fm* qui garantissent cette position en termes morphologiques. Pour ce faire, ils proposent une définition de la fonction F_{PRAF} :

(1.31) La définition de la fonction F_{PRAF} (où X est une forme fléchie, Y est la valeur de l'attribut TÊTE et Z la valeur de la liste ARG-ST) :

FPRAF(X,Y,Z) = W où W

(1) = X si Y = [VFORM *past-p*],

$$(2) = \begin{bmatrix} \textit{encl-fm} & \\ \text{BASE} & X \end{bmatrix} \text{si Y} = \begin{bmatrix} \text{VFORM} & \textit{imp} \\ \text{NEG} & - \end{bmatrix},$$

$$(3) = \begin{bmatrix} \textit{procl-fm} & \\ \text{BASE} & \text{X} \end{bmatrix} \text{autrement.}$$

Cette définition est conjuguée avec des contraintes supplémentaires sur W. Parmi ces contraintes, nous prenons comme exemple celles qui déterminent les formes des clitiques sujets en SL-1 [39] :

[39] C : abréviation du mot *condition*.

(1.32) La valeur W de l'attribut F est v, seulement si Z contient α et Y satisfait H, où :

	F	v	α	H
C1a :	SL-1	je	*[p-aff, 1sg, nom]*	*tensé*
C1b :	SL-1	tu	*[p-aff, 2sg, nom]*	*tensé*
C1c :	SL-1	il	*[p-aff, 3sgm, nom]*	*tensé*
C1d :	SL-1	elle	*[p-aff, 3sgf, nom]*	*tensé*
C1e :	SL-1	on	*[p-aff, 3sgm, nom]*	*tensé*
C1f :	SL-1	nous	*[p-aff, 1pl, nom]*	*tensé*
C1g :	SL-1	vous	*[p-aff, 2pl, nom]*	*tensé*
C1h :	SL-1	ils	*[p-aff, 3plm, nom]*	*tensé*
C1i :	SL-1	elles	*[p-aff, 3plf, nom]*	*tensé*

Ce tableau se lit comme suit pour la première ligne par exemple : la condition *C1a* doit être interprétée comme garantissant que la valeur de SL-1 de l'extrant de $\mathbf{F}_{PRAF}$ W soit *je* seulement quand :

1. Z contient un élément de type *p-aff* qui est spécifié également comme [PER *1*], [NUM *sg*], et [CAS *nom*], et
2. Le *cl-wd* en question est spécifié [VFORM *tensé*].

Si nous conjuguons ces contraintes avec celles du type *su-cl-wd* en (1.25), on garantit que les pronoms clitiques sujets apparaissent seulement quand le premier argument du verbe est un sujet à la 1re personne du singulier au nominatif. Ces contraintes garantissent du même coup le fait que les clitiques sujets doivent apparaître avant tous les autres types de clitiques. Pour modéliser le comportement des clitiques non-sujets, Miller et Sag proposent les contraintes suivantes[40] :

[40] Dans ce tableau, le super-type de *obj* est *cas* et ses sous-types sont *acc* et *dat*.

(1.33) La valeur W de l'attribut F est v, seulement si Z contient α et Y satisfait ω où :

	F	v	α	ω
C2a :	SL-12	me	*[aff,1sg,obj]*	*[procl-fm]*
C2b :	SL-2	te	*[aff,2sg,obj]*	*[procl-fm]*
C2c :	SL-12	nous	*[aff,1pl,obj]*	*[procl-fm]*
C2d :	SL-2	vous	*[aff,2pl,obj]*	*[procl-fm]*
C2e :	SL-2	se	*[a-aff,3,obj]*	*[procl-fm]*
C3a :	SL-3	la	*[p-aff,3sgf,acc]*	
C3b :	SL-3	le	*[p-aff,3sgm,acc]* $\bigvee$ *[p-aff,* +PRED*]*	
C3c :	SL-3	les	*[p-aff,3pl,acc]*	
C4a :	SL-4	lui	*[p-aff,3sg,à1]*	
C4b :	SL-4	leur	*[p-aff,3pl,à1]*	
C5a :	SL-5	moi	*[aff,1sg,obj]*	*[encl-fm]*
C5b :	SL-5	toi	*[aff,2sg,obj]*	*[encl-fm]*
C5c :	SL-5	nous	*[aff,1pl,obj]*	*[encl-fm]*
C5d :	SL-5	vous	*[aff,2pl,obj]*	*[encl-fm]*
C6a :	SL-6	y	*[p-aff,à2]*	*[procl-fm]*
C6b :	SL-6	zy	*[p-aff,à2]*	*[encl-fm]*
C7a :	SL-7	en	*[p-aff,de]*	*[procl-fm]*
C7b :	SL-7	zen	*[p-aff,de]*	*[encl-fm]*

De plus, les auteurs proposent une autre contrainte sur l'ordre de l'apparition des clitiques dans une suite :

(1.34) Contrainte sur l'ordre des clitiques non-sujets

C8 : si SL-2 ou SL-6 est non vide, alors SL-4 est vide.

La violation de cette contrainte engendre l'agrammaticalité des exemples suivants :

(1.35) a. *Il me lui présente.

b. *Présentez-moi-lui !

c. *Présentez-lui-moi !

d. *Ils se lui présentent.

La fonction F_{PRAF} se charge donc, non seulement de la forme morphologique du mot et de la réalisation de certains de ses arguments comme éléments de type *aff*, mais également, et grâce à l'interaction avec d'autres contraintes d'ordre, de contraindre l'ordre dans lequel apparaissent les clitiques en combinaison avec leur hôte.

L'approche de Miller et Sag (1997) est la plus complète concernant les clitiques du français dans le cadre de la grammaire HPSG et elle est souvent citée comme étant la base de l'étude de l'interface morphologie/syntaxe dans ce cadre. Exempte de règles lexicales, cette approche est entièrement basée sur les contraintes. Les auteurs proposent ainsi des contraintes morphologiques, morphosyntaxiques et syntaxiques qui gèrent l'emplacement des clitiques dans cette langue et proposent une modélisation non seulement des clitiques objets, mais également des clitiques sujets.

Cependant, cette modélisation n'est pas exempte des problèmes suivants. En effet, et comme le souligne Smets (2000), la combinaison d'un système de types et d'une fonction phonologique (F_{PRAF}) ne garantit pas que seules les suites de clitiques grammaticales seront permises. D'ailleurs, d'autres contraintes fonctionnent comme des filtres pour rejeter les suites mal formées. Ces contraintes sont indépendantes de la définition des types et constituent des stipulations indésirables. De plus, le statut formel de certaines de ces contraintes et conditions est problématique. Celles en (1.33), par exemple, ne sont en fin de compte qu'une liste de conditions auxquelles la fonction phonologique centrale doit avoir accès d'une manière ou d'une autre.

Finalement, la modélisation de l'ordre des clitiques dans les suites est fortement inspirée du modèle proposé par Perlmutter (1971) et repose donc sur des schèmes descriptifs[41]. Aucun principe grammatical ne semble intervenir dans cette modélisation, qui reste dépendante, après tout, des filtres[42] comme celui en (1.34).

[41] Le schème proposé par cet auteur pour les clitiques de l'espagnol repose sur un système de positions pour ordonner les clitiques dans des suites bien formées : SE II I III.

Ainsi, les clitiques sont assignés à différentes classes de positions et ils apparaissent dans l'ordre prescrit par ces classes. Ce schème impose en même temps une contrainte forte sur cet ordre puisque chaque position ne peut accueillir plus d'un seul clitique de la même classe à la fois. Évidemment, ce système de positions n'est pas suffisant pour éliminer toutes les suites mal formées et l'auteur a donc besoin de filtres qui stipulent que certaines combinaisons ne sont pas possibles.

[42] Les filtres font partie du dispositif formel mis au point par Perlmutter (1971) lui-même.

1.4.2 Une mise en question de l'hypothèse affixale

Les approches que nous avons présentées dans les sections précédentes prennent exclusivement les clitiques des langues romanes comme objet d'étude. Il existe toutefois d'autres travaux dans le cadre de HPSG qui ont porté sur d'autres familles de langues. Anna Kupsc (Kupsc, 2000, 1999) a, par exemple, consacré plusieurs travaux, dont sa thèse de doctorat, au développement d'une approche aux clitiques du polonais, fidèle aux principes de la théorie HPSG. Gerald Penn a également mené des travaux (Penn, 1999b,a) sur les clitiques du serbo-croate dans le même cadre théorique. Ces deux langues appartiennent à la famille slave et constituent un défi lorsqu'il s'agit de porter des analyses proposées à l'origine pour les langues romanes. Dans cette section, nous présenterons quelques problèmes qu'a révélés l'étude du polonais par Anna Kupsc, en guise de comparaison avec les résultats obtenus pour les langues romanes.

Les approches aux clitiques romans dans le cadre de HPSG sont basées sur l'hypothèse affixale : ces clitiques sont des affixes flexionnels [43]. Kupsc (2000, 1999) n'adhère pas à cette hypothèse. Elle considère en fait que les clitiques (les pronoms faibles) du polonais sont des items syntaxiques [44] qui correspondent à des signes (ou à des mots indépendants). Cela a un effet sur la hiérarchie des types lexicaux, comme nous pouvons le constater dans la figure (1.10).

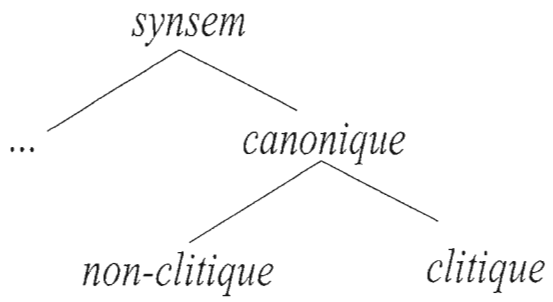


FIGURE 1.10: La hiérarchie (partielle) des types lexicaux du polonais proposée par Kupsc (2000)

Cette hiérarchie place le type *clitique* comme sous-type de *canonique*, ce qui diffère énormément

[43] Cette affirmation n'est pas absolue. En italien, par exemple, les clitiques sont quelque peu mixtes. En effet, pour Monachesi (1998a), les clitiques monosyllabiques de cette langue sont des affixes flexionnels alors que le clitique dissyllabique *loro* est un mot indépendant. Par ailleurs, Crysmann (2000), qui a étudié les clitiques du portugais européen, propose que ces derniers représentent un état de transition où ils ont déjà acquis les propriétés morphologiques des affixes alors que leurs règles de placement restent essentiellement syntaxiques.

[44] Selon cette auteure, les propriétés générales des pronoms clitiques du polonais sont les suivantes : ils sont prosodiquement déficients et apparaissent comme des enclitiques, mais ils ne forment pas une unité lexicale prosodique avec leur hôte. Ils peuvent, par ailleurs, être effacés et n'ont pas de position fixe dans la phrase.

de la hiérarchie proposée par Miller et Sag (1997) pour le français. Par contre, un seul pronom clitique polonais est lexicalement introduit dans la structure. Il s'agit du réfléchi *się* (se) des verbes pronominaux inhérents dans l'exemple (1.36)[45], qui reçoit un traitement lexical et constitue le seul clitique de cette langue qui se comporte comme les pronoms clitiques romans. Les autres pronoms clitiques polonais reçoivent donc un traitement syntaxique selon l'analyse de Kupsc, parce qu'aucun mécanisme lexical n'est nécessaire pour assurer la bonne représentation de ces clitiques. Formellement, cela a l'effet suivant : ces clitiques apparaissent dans la liste COMPS de leur prédicat, tout comme les éléments de type *non-clitique*. Les principes nécessaires pour rendre compte de leurs propriétés sont, par conséquent, les mêmes principes syntaxiques qui servent à introduire les mots indépendants dans la structure. Kupsc relègue ainsi certains clitiques polonais à la syntaxe et n'en fait pas un objet d'étude morphologique.

(1.36) *Jan chce sie przestac go hac*
Jean veut se arrêter lui avoir.peur
« Jean veut arrêter d'avoir peur de lui »

Cette comparaison entre les clitiques des langues romanes et les clitiques des langues slaves, et entre les approches des différents auteurs qui ont travaillé sur ces entités, nous révèle que l'appartenance au même cadre théorique ne signifie pas forcément l'adoption des mêmes analyses linguistiques. Les clitiques qui appartiennent à des familles de langues différentes reçoivent ainsi des analyses différentes. On s'attendrait donc à ce que ce soit le cas des marqueurs d'arguments de l'arabe.

1.5 HPSG et l'arabe

Parmi les recherches menées sur la langue arabe dans ce cadre, nous citons Moosally (1994), qui a étudié les pronoms résomptifs, Smets (2000), qui a formalisé certains aspects de la morphologie non concaténative de cette langue et Beller (2006), qui a étudié l'accord *appauvri* en arabe moderne.

Malgré ces efforts, il n'existe, à notre connaissance, aucune grammaire de l'arabe formalisée dans le cadre de HPSG. Les études que nous venons de mentionner sont généralement partielles ou sont basées sur des approches informatiques et non des analyses linguistiques. Nous nous

[45] (Kupsc, 1999, p. 77)

devons ainsi de citer les travaux de Aloulou (2003), Aloulou *et al.* (2002, 2003) et Bahou *et al.* (2006), qui ont développé un analyseur de textes arabes non voyellés *syntaxe*, basé sur HPSG, dans le cadre du système *maspar*. Citons également le travail accompli par Mutawa *et al.* (2007) pour analyser les phrases nominales simples de l'arabe dans ce cadre. La base de ces travaux est cependant informatique et grammaticale dans le sens scolaire du terme. Par ailleurs, les seules analyses qui touchent de près à notre objet d'étude, comme l'analyse de Smets (2000), ne prennent en considération qu'une seule dimension des marqueurs d'arguments, à savoir la dimension morphophonologique.

1.5.1 Une analyse des pronoms résomptifs de l'arabe

Dans son mémoire de maîtrise, Moosally (1994) a étudié les pronoms résomptifs de l'arabe dans le cadre de HPSG. Elle a également abordé la question des asymétries de l'accord et celle de l'ordre de surface. Sans contribuer au débat soulevé par ces problèmes, l'auteure adopte certaines analyses déjà proposées dans la littérature et les formalise en utilisant les outils fournis dans l'ouvrage de Pollard et Sag (1994). Elle stipule ainsi que l'arabe est une langue *pro-drop* de type VSO. Le seul argument par lequel l'auteure justifie son choix d'inclure une catégorie vide dans la liste des catégories pronominales possibles en arabe (en plus des pronoms indépendants et des pronoms clitiques) est d'ordre théorique. En effet, si elle n'adopte pas l'analyse proposée dans le cadre de LFG par Fassi Fehri (1988) (et selon laquelle les marqueurs de sujet sont des pronoms sujets) c'est que cette approche ne pourrait, selon elle, être formalisée dans le cadre de HPSG. Elle propose la structure de traits suivante pour décrire l'entrée lexicale de ce pronom nul [46].

(1.37) L'entrée lexicale du pronom nul

[46] Les abréviations employées dans cet AVM sont les suivantes : *ppro* pour pronom personnel et *nom* pour le cas nominatif.

$$\begin{bmatrix} \text{PHON} & \langle\rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{ppro} & \\ \text{CAS} & \textit{nom} \end{bmatrix} \\ \text{VAL} & \langle\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{ppro} & \\ \text{INDEX} & \begin{bmatrix} \text{PERS} \\ \text{NBRE} \\ \text{GRE} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

La contrainte casuelle (le nominatif) sur ce pronom nul (la valeur de l'attribut PHON est une liste vide) est censée l'empêcher d'apparaître autrement que postverbal (la position du sujet). Cependant, le sujet n'a pas le monopole du cas nominatif en arabe et rien dans cette représentation n'empêche donc le pronom nul d'être préverbal. Les descriptions proposées par l'auteure ont l'inconvénient de permettre la réalisation du pronom nul en position préverbale, étant un sous-type de *ppro* (pronom personnel).

1.5.2 Une modélisation de la morphologie de l'arabe

Smets (2000) a proposé une théorie morphologique pour la grammaire HPSG basée sur l'approche paradigmatique (appelée l'approche par Mot et Paradigme *Word and Paradigm*) qui soit indépendante des processus morphologiques particuliers. Pour ce faire, l'auteure a testé sa théorie non seulement sur une langue connue pour sa morphologie concaténative (le français et, plus particulièrement, les pronoms clitiques de cette langue), mais également sur une langue connue pour sa morphologie à schèmes (l'arabe). Il va sans dire que l'arabe combine les deux sortes de processus morphologiques : la concaténation et le recours aux schèmes.

L'auteure a consacré le 5e chapitre de sa thèse à deux phénomènes de l'arabe : le pluriel dit *brisé* et la morphologie verbale. Pour des raisons évidentes, nous choisissons de présenter son travail sur la morphologie verbale. L'auteure a en fait repris l'analyse proposée par McCarthy (1981) dans le cadre de la phonologie autosegmentale et l'a implémentée dans sa théorie morphologique basée sur l'unification. Le point crucial de l'analyse de McCarthy c'est que, constatant que les différentes composantes de la forme verbale en arabe expriment des informations de différents ordres (le

schème verbal est associé à l'information sémantique et aux propriétés de sous-catégorisation de la racine, la mélodie vocalique est associée aux catégories verbales comme l'aspect et la voix et la racine fournit le sens général), il faudrait être capable de représenter ces différentes informations de manière indépendante. La représentation adoptée par McCarthy repose de ce fait sur trois niveaux (*tiers*) : un pour représenter la racine consonantique, un pour la mélodie vocalique et un troisième pour représenter le schème morphologique.

Implémentée dans la théorie morphologique de Martine Smets, l'analyse de McCarthy donne lieu à une redéfinition de certains attributs. L'attribut PHON(ologie) a pour valeurs possibles trois autres attributs : RACINE, VOCAL(isme) et PATRON (qui correspond au schème) [47].

Cette analyse constitue un bon point de départ pour modéliser la morphologie verbale de l'arabe. Sa couverture empirique reste néanmoins limitée. Elle ne prend en compte que les 10 premiers schèmes de la dérivation verbale (les plus productifs) appliqués à trois verbes à racine trilitère et ne se préoccupe que de la formation morphologique de ces verbes. Elle délaisse ainsi toute information morphosyntaxique ou syntaxique associée aux morphèmes qui entrent dans la composition de ces verbes : « The account of Arabic verbal morphology is thus limited to the definition of helping and main paradigms. » (Smets, 2000, p. 114) Ce type d'information nous importe ici puisque nous entreprenons de modéliser non seulement la morphophonologie mais également la morphosyntaxe des marqueurs d'arguments.

1.6 Conclusion

Nous avons consacré ce premier chapitre à l'introduction du cadre théorique et formel dans lequel nous proposerons notre analyse, la grammaire HPSG. Cette théorie est dotée de plusieurs propriétés qui lui procurent une grande flexibilité dans le traitement des phénomènes langagiers en préservant sa puissance formelle. Cette flexibilité et cette puissance émanent essentiellement de l'adoption d'une seule structure de données, les structures de traits typées, et d'une opération formelle centrale, l'unification. De plus, les dispositifs formels fournis dans ce cadre, à savoir essentiellement la hiérarchie de types à héritage multiple et les règles lexicales, permettent de modéliser une grand nombre de phénomènes, de rendre compte des généralisations et de faire place à l'expression de la variation linguistique. Un lexique riche, un formalisme bien défini et un

[47] Ces attributs ne sont cependant valables que pour les entrées verbales. Les attributs valeurs de PHON sont issus des notions employées dans la morphologie prosodique (syllabes et pieds) en ce qui concerne le pluriel brisé (et donc pour les entrées nominales).

apport provenant de plusieurs théories et de plusieurs disciplines colorent HPSG d'une couleur unique parmi les grammaires génératives, mais également parmi le paradigme des grammaires à base de contraintes.

Cette introduction de notre cadre théorique ne saurait être complète sans illustrer ses apports à la recherche linguistique. Nous avons effectué un choix et ce dernier n'était pas arbitraire. Nous avons en fait choisi le problème des clitiques, puisqu'il a des liens évidents à notre sujet d'étude, les marqueurs d'arguments de l'arabe standard. Nous avons mis l'accent sur le traitement de la cliticisation dans la famille romane et dans la famille slave et nous avons identifié quelques problèmes que connaissent ces analyses. Nous pensons que ce qu'il faudrait retenir de ces analyses c'est surtout le respect du domaine empirique, à savoir les données pertinentes de la langue étudiée, lorsque l'on propose une analyse ou une autre. Ainsi, même si l'hypothèse affixale semble privilégiée en ce qui a trait aux langues romanes, les langues slaves suggèrent une approche différente et cette approche trouve sa place dans la théorie et dans le formalisme, puisque l'hypothèse affixale ne fait partie ni de la théorie ni du formalisme.

Nous avons également présenté quelques analyses de phénomènes propres à l'arabe dans le cadre de cette grammaire et nous avons conclu que beaucoup de travail reste à faire pour écrire une grammaire HPSG de cette langue. Le peu de travail déjà accompli souffre de lacunes théoriques redevables surtout à l'absence de la dimension linguistique et à la prépondérance de l'analyse grammaticale traditionnelle ou de l'aspect purement informatique. Les quelques analyses pouvant être qualifiées de linguistiques ont également leur lot de problèmes, c'est le cas notamment de l'approche de Moosally (1994) dans son étude des pronoms résomptifs de l'arabe.

Avant d'entreprendre l'analyse et la modélisation des marqueurs d'arguments de l'arabe, nous les présentons dans le deuxième chapitre en établissant les faits les plus saillants les impliquant. Ces faits s'établissent à deux niveaux : au niveau de la morphophonologie et au niveau de la morphosyntaxe.

CHAPITRE II

LE SYSTÈME PRONOMINAL DE L'ARABE

> Make a precise, falsifiable hypothesis that accounts for a wide range of facts (i.e. get the details right for a subpart of the language), then revise the hypothesis to expand coverage. (Pollard, 1997)

Introduction

L'arabe (dans ses variétés classique et contemporaine) présente des formes que nous qualifions de pronoms personnels en suivant la définition proposée par Bresnan (2001) :

> What universally characterizes a pronoun are its referential role and functions, not its phrase-structure category. Some elements which resemble clitic pronouns, such as the indirect object clitic copies in Spanish, are not pronominal in content, but simply markers of grammatical agreement. Likewise, some elements which resemble non-pronouns, such as deictics used anaphorically in many languages or bound agreement morphology on verbs, may actually function as pronominals.

En effet, selon cette auteure, trois types de propriétés définissent les pronoms : sémantiques, informationnelles et morphosyntaxiques. Sémantiquement, un pronom est caractérisé par sa capacité de changer de référent et par son anaphoricité. Ses propriétés informationnelles se résument dans son rôle dans la structure de l'information, en ceci qu'il peut référer à un topique [1]. Les propriétés morphosyntaxiques, quant à elles, se résument dans le fait que les pronoms sont morphologiquement distinguables (et classifiables) selon leurs traits morphosyntaxiques (de personne, de genre et de nombre).

En arabe, les formes que nous considérons comme pronominales [2] suivant cette définition se présentent sous deux formes : des formes indépendantes et des formes conjointes. Les formes

[1] Par opposition au commentaire, ce terme désigne la partie dont on parle dans un énoncé donné. Voir Givon (1983), Lambrecht (1994) et Li et Thompson (1976).

[2] Nous ne traiterons que les pronoms personnels. D'autres pronoms, démonstratifs et relatifs, sont attestés en arabe, mais ne font pas partie de notre sujet d'étude.

indépendantes, que nous désignons par l'appellation *pronoms indépendants*, sont des proformes phonologiquement et morphologiquement indépendants qui remplacent des syntagmes nominaux. Les formes conjointes, que nous désignons par le terme *marqueurs d'arguments*, emprunté à Auger (1994), sont des morphèmes prosodiquement déficients qui encodent les traits des arguments auxquels ils sont associés ou qu'ils remplacent. Les deux formes peuvent changer de référents, référer à des topiques et être classées selon leurs traits morphosyntaxiques, se conformant ainsi à la définition des pronoms proposée par Bresnan (2001).

Nous consacrons ce chapitre à la description de ces pronoms et nous le subdivisons comme suit. Dans la section (2.1), nous introduisons notre corpus et les sources de nos données. Suit une section, la (2.2), dans laquelle nous faisons une présentation brève des deux sortes de pronoms personnels de l'arabe. La section (2.3) est consacrée à une présentation plus détaillée des propriétés morphologiques et syntaxiques des pronoms indépendants, alors que le section (2.4) est consacrée à une présentation semblable des marqueurs d'arguments. Finalement, nous consacrons la section (2.5) à une comparaison entre ces deux sortes de pronoms.

2.1 Sources des données

Bien que la majorité des exemples présentés dans ce travail soient construits (conformes à notre intuition) ou proviennent des travaux cités [3], nous nous appuyons également sur trois corpus écrits :

1. Une anthologie classique : *ʔal ʕiqdu l-fari:d* de Ibn abd rabbih, auteur et poète des IXe et Xe siècles.
2. Un roman contemporain : *mawsim lhiǧra ila ʃʃama:l* de tajjib s'a:lih.
3. Un corpus écrit à l'origine, mais lu à haute voix. Ce dernier sert à vérifier la pertinence de l'analyse des structures appelées SVO en arabe. Il se compose de textes classiques et modernes lus par différents narrateurs. Les textes classiques proviennent du site Web http ://www.alwaraq.net, qui constitue une bibliothèque virtuelle du patrimoine arabe écrit. Ce site comporte également des enregistrements audio de 12 récits classiques populaires d'à peu près 5 minutes chacun. Nous avons également un corpus audio de textes contemporains provenant des médias. Il s'agit d'un documentaire scientifique de 25 minutes diffusé sur la

[3] Nous n'avons exclu aucune source possible de données. Ainsi, les livres de grammaire écrits par les premiers grammairiens arabes et ceux écrits dans l'époque contemporaine sont consultés. Nous nous sommes également basés sur les travaux des linguistes contemporains et de certains Orientalistes.

chaîne qatariote *Al Jazeera* et d'une émission de nouvelles de 7 minutes diffusée sur la même chaîne le 11 avril 2007.

2.2 Une brève présentation

En nous basant sur nos sources de données et sur notre intuition, nous classons les pronoms personnels de l'arabe en deux sortes, à la présentation desquelles nous consacrons cette section. Nous commençons par les formes pronominales autonomes, que nous qualifions de pronoms indépendants.

2.2.1 Les pronoms indépendants

Ces pronoms peuvent porter un accent principal indépendant. Ils peuvent, en outre, avoir deux cas possibles et prennent donc des formes différentes suivant ces deux cas. Le pronom de la 3MS, par exemple, est *huwa* au nominatif et *ʔijja:hu* à l'accusatif.

Les pronoms indépendants prennent par ailleurs des formes différentes selon la spécification de leurs traits morphosyntaxiques. Ainsi, le pronom indépendant nominatif de la troisième personne du singulier masculin prend la forme *huwa*, alors que le pronom indépendant nominatif de la troisième personne du singulier féminin prend la forme *hija*. Cependant, certains de ces marqueurs ne sont pas discriminés en fonction du genre et prennent la même forme au masculin et au féminin : les pronoms de la 1re personne (au nominatif *ʔana:* : moi ; *naħnu* : nous) et les pronoms du duel[4] (au nominatif *naħnu* : nous deux M/F ; *ʔantuma:* : vous deux M/F ; *huma:* eux deux/elles deux).

Introduits dans la structure, les pronoms indépendants peuvent être essentiels, comme dans la phrase équative[5] en (2.1a), et peuvent être optionnels dans les autres cas, comme en (2.1b), d'où la possibilité de (2.1c). À noter que nous adoptons la méthode suivante pour les gloses : les pronoms indépendants sont glosés par les pronoms forts du français. Nous utilisons, selon le contexte, les clitiques et les pronoms forts du français pour gloser les marqueurs d'objet. Les

[4] Le duel est propre à l'arabe standard ; il a complètement disparu de l'arabe dialectal.

[5] Les phrases dites « équatives » ne comportent pas de verbe ni de copule verbale. Elles sont construites par la juxtaposition d'un sujet et d'un prédicat (pouvant être un nom prédicatif, un adjectif prédicatif ou un syntagme prépositionnel prédicatif), tous deux au nominatif, sans copule verbale (Shulte-Nafeh, 2004). Le terme « phrase équative » est différent du terme traditionnel « phrase nominale » en ceci qu'il n'englobe pas les phrases introduites par des verbes « déficients ». Ce terme ne désigne pas non plus les phrases construites selon l'ordre de mots dit SVO, traditionnellement considérées comme des phrases nominales.

marqueurs de sujet sont toujours glosés par les traits morphosyntaxiques qu'ils représentent dans leur contexte.

(2.1) a. ***huwa*** *ʕalijj* *-un* *ʔibn* *-u* *muħammad* *-in*
lui Ali -NOM fils -NOM Mohammed -GÉN
« Il est Ali fils de Mohammed »

b. *ʔanta* *ʔakal* *-ta* *ʃaba:b* *-a* *-ka*
toi manger.PER -2MS jeunesse -ACC -ta
« Toi, tu as dilapidé ta jeunesse »

c. *ʔakal* *-ta* *ʃaba:b* *-a* *-ka*
manger.PER -2MS jeunesse -ACC -ta
« Tu as dilapidé ta jeunesse »

Une liste complète de ces proformes est présentée dans le tableau (2.1). Nous remarquons, dans ce tableau, qu'à la 1re personne, l'arabe ne fait pas de distinction de genre alors que cette distinction est opérée à la 2epersonne. Cette propriété semble être une particularité de la famille sémitique et elle se rencontre en hébreu également, comme le mentionne (Shlonsky, 1997).

	Singulier		Duel		Pluriel	
	NOM	ACC	NOM	ACC	NOM	ACC
1	ʔana:	ʔijja:ja	naħnu	ʔijja:na:	naħnu	ʔijja:na:
2M	ʔanta	ʔijja:ka	ʔantuma:	ʔijja:kuma:	ʔantum	ʔijja:kum
2F	ʔanti	ʔijja:ki	ʔantuma:	ʔijja:kuma:	ʔantunna	ʔijja:kunna
3M	huwa	ʔijja:hu	huma:	ʔijja:huma:	hum	ʔijja:hum
3F	hija	ʔijja:ha:	huma:	ʔijja:huma:	hunna	ʔijja:hunna

TABLEAU 2.1: Les pronoms indépendants (nominatifs et accusatifs)

Nous présentons les marqueurs d'arguments dans la section suivante.

2.2.2 Les marqueurs d'arguments

Le deuxième groupe de pronoms personnels est constitué de marqueurs phonologiquement dépendants d'autres catégories et ne se rencontrant dans la langue que dans un état d'attachement. Ces marqueurs sont de deux sortes : des marqueurs d'objet et des marqueurs de sujet. Nous commençons notre présentation par les marqueurs d'objet.

2.2.2.1 Les marqueurs d'objet

Les marqueurs d'objet peuvent être accusatifs ou génitifs, mais ils ont toujours la même forme phonologique dans les deux cas, à une exception près, à savoir avec le marqueur d'objet de 1S, qui prend la forme *-niː* quand il est attaché aux verbes (donc à l'accusatif) et *-iː* (*-ja*) quand il est attaché aux autres parties du discours (au génitif). Notons ici que les termes *nominatif*, *accusatif* et *génitif* correspondent *grosso modo* à ce qu'ils signifient dans les langues indo-européennes. Cette correspondance n'est cependant pas parfaite ; le génitif, par exemple, désigne aussi bien le génitif que d'autres cas comme l'oblique et le datif. Les trois cas retenus ici correspondent aux trois flexions casuelles possibles pour le nom (et les autres catégories reliées au nom) : *-u* pour le nominatif, *-a* pour l'accusatif et *-i* pour le génitif. Concernant les pronoms personnels, les désinences casuelles ne sont pas visibles ; ces pronoms prennent toutefois des formes phonologiques différentes suivant leurs fonctions dans la phrase et s'ils sont remplacés par des SN, ces derniers prennent les désinences associées aux trois cas mentionnés.

Les marqueurs d'objet sont des pronoms toujours postposés à la tête de laquelle ils dépendent (se comportant ainsi comme des enclitiques). Comme pour les pronoms indépendants, nous remarquons ici également la présence de formes communes : les distinctions de genre sont neutralisées à la 1re personne et au duel. La forme de la 1S (valable pour le M et le F), par exemple, est *-iː*, également réalisée *-niː* et *-ja* dans certains contextes.

Les exemples suivants illustrent quelques emplois possibles de ces marqueurs. En (2.2a), le marqueur d'objet de 3FS *-haː* est attaché à une tête verbale (à laquelle est déjà attaché le marqueur de sujet *-ta*), son cas est l'accusatif et sa fonction est celle de complément du verbe. En (2.2b), le marqueur d'objet *-haː* est attaché à la préposition *min* « de », son cas est le génitif et sa fonction est celle de complément de préposition.

(2.2) a. *hal* *raʔaj* *-ta* ***-haː*** *ʔ*
est-ce-que voir.PER -2MS -la
« L'as-tu vue ? »

b. *xif* *-tu* *min* ***-haː***
avoir.peur.PER -1S de -elle
« J'ai eu peur d'elle »

Une liste complète des marqueurs d'objet est présentée dans le tableau (2.2).

	Singulier	Duel	Pluriel
1	-i: (-ni:, -ja)	-na:	-na:
2M	-ka	kuma:	kum
2F	-ki	kuma:	kunna
3M	-hu	huma:	hum
3F	-ha:	huma:	hunna

TABLEAU 2.2: Les marqueurs d'objet

Nous serons amené à comparer ces marqueurs d'objet aux marqueurs de sujet sur les plans morphologique et syntaxique. Pour ce faire, nous présentons ces derniers dans la section suivante.

2.2.2.2 Les marqueurs de sujet

Avant de présenter les marqueurs de sujet, nous pensons qu'une introduction à la formation morphologique des mots en arabe, et surtout à celle des verbes, apporte un éclairage indispensable.

En arabe, et dans les langues sémitiques en général, la formation des mots repose sur deux ingrédients : la racine et le schème morphologique. La racine est une suite de phonèmes qui définissent la base lexicale des mots qui en sont dérivés. Cette suite, très souvent trilitère, se compose de consonnes dans un ordre précis. Ces consonnes forment un morphème discontinu qui exprime plus ou moins un concept de base reconnaissable dans les mots dérivés. L'exemple que nous prenons pour illustrer cette propriété est la racine */ktb/*. Cette suite consonantique exprime le concept de « l'écriture » et sert à dériver un grand nombre de noms et de verbes, comme les suivants :

(2.3) a. Des verbes

- kataba : « il a écrit »
- kutiba : « il a été écrit »
- jaktubu : « il écrit »
- jataka:tabu:na : « ils s'écrivent »

b. Des noms

- kita:b : « livre »
- kutub : « livres »

- kutajjib : « petit livre »
- kutta:b : « école coranique »
- ka:tib : « écrivain »
- maktab · « bureau »
- ʔiktita:b : « enregistrement »

Pour former ces mots, la suite consonantique, qui compose la racine, est combinée à des voyelles, à d'autres consonnes et à une information morphophonologique qui précise les modifications que doivent subir les consonnes de la racine (dans *kutta:b*, par exemple, la deuxième consonne de la racine est géminée) et leur position par rapport aux éléments insérés. Cette information, ces voyelles et ces consonnes constituent le schème morphologique. Ce dernier précise en fait quoi ajouter à la racine, où l'ajouter et sous quelles conditions. Le schème est par conséquent une sorte de morphème discontinu[6] ; sa notation se fait traditionnellement en remplaçant les consonnes de la racine par les trois consonnes abstraites /fʕl/ (dont le sens général est celui de « faire ») et en consignant les voyelles et les consonnes ajoutées ainsi que tout autre changement que peut subir la racine. Ainsi, *kutta:b*, par exemple, possède le schème morphologique *fuʕʕa:l* et *kataba* le schème *faʕala*. L'arabe comporte environ 150 schèmes et tous les mots de cette langue (à l'exception des morphèmes grammaticaux, dont les pronoms) peuvent être analysés comme issus de l'entrelacement d'une racine et d'un schème.

Les verbes sont formés selon le même modèle, c'est-à-dire en combinant des racines et des schèmes. Cependant, l'analyse morphologique des paradigmes verbaux révèle l'existence de formes « intermédiaires » qui servent de base à la conjugaison. Ces formes invariables dans chaque paradigme sont les radicaux. Le radical se compose des consonnes de la racine ainsi que des voyelles et des consonnes ajoutées par le schème. Les seuls morphèmes qui ne font pas partie du radical sont des morphèmes variables qui véhiculent des informations concernant les trois traits morpho-syntaxiques (la personne, le genre et le nombre) et le mode. De la racine */ktb/*, par exemple, est dérivé le verbe perfectif *kataba* « il a écrit ». Cette forme conjuguée peut être analysée en isolant le radical *katab* et le morphème -*a* de la 3MS. Le même radical est employé dans la forme *katabna:* « nous avons écrit » combiné au morphème -*na:* de la 1MP.

En prenant l'aspect comme critère typologique, nous classons les formes verbales possibles en

[6] SelonHoberman (1988), Marusic (2002), McCarthy (1981) et Prunet *et al.* (2000), les langues sémitiques en général connaissent le phénomène des morphèmes discontinus. Ces derniers peuvent être de deux sortes : des morphèmes discontinus fonctionnels (des circumfixes) et des morphèmes discontinus lexicaux (la racine consonantique). Les premiers ne sont pas possibles dans une langue donnée sans les deuxièmes selon Marusic (2002).

arabe en trois groupes : le perfectif[7], l'imperfectif (avec ses trois modes : l'indicatif, le subjonctif et le jussif) et l'impératif (que nous pouvons considérer comme une forme de l'imperfectif). Les paragraphes qui suivent sont consacrés aux marqueurs de sujet attachés à ces différentes formes verbales.

2.2.2.2.1 Les marqueurs de sujet au perfectif

À la forme perfective, les marqueurs de sujet sont suffixés au radical verbal et précèdent ainsi tout autre morphème (comme les marqueurs d'objet) qui pourrait être attaché au verbe, comme dans l'exemple (2.2a), répété ici en (2.4).

(2.4) *hal* *raʔaj* ***-ta*** *-ha:ʔ*
est-ce-que voir.PER -2MS -la
« L'as-tu vue ? »

Les marqueurs de sujet de la forme perfective sont classifiables selon leurs traits morphosyntaxiques en marqueurs de 1S (*-tu*), de 2MS (*-ta*), de 2FS (*-ti*), etc. Il existe cependant, et comme nous l'avons déjà mentionné concernant les pronoms indépendants et les marqueurs d'objet, des formes communes notamment à la 1re personne et à la 2D.[8]

Ces marqueurs peuvent par ailleurs être référentiels dans des exemples comme le suivant, où le marqueur de sujet du perfectif *-u:* de 3MP réfère au pronom indépendant topique préverbal *hum* « eux » :

(2.5) hum_i *daxal* $\textbf{-u:}_i$
eux entrer.PER -3MP
« Eux, ils sont entrés »

Une liste complète des marqueurs de sujet à la forme perfective est présentée dans le tableau (2.3) sous forme de la conjugaison perfective du verbe *katab* « écrire ». Dans ce tableau, les marqueurs de sujet sont en caractères gras.

[7] La distinction entre le perfectif et l'imperfectif (ou entre l'accompli et l'inaccompli selon certains auteurs) est une distinction aspectuelle et non temporelle. Le perfectif exprime une action achevée alors que l'imperfectif exprime le fait que « l'action est en train de se réaliser » selon Blachère et Gaudefroy-Demombines (2004).

[8] Cependant, les formes sont discriminées à la 3D, avec *-a:* pour la 3MD et *-ata:* pour la 3FD.

	Singulier	Duel	Pluriel
1	katab-**tu**	katab-**na:**	katab-**na:**
2M	katab-**ta**	katab-**tuma:**	katab-**tum**
2F	katab-**ti**	katab-**tuma:**	katab-**tunna**
3M	katab-**a**	katab-**a:**	katab-**u:**
3F	katab-**at**	katab-**ata:**	katab-**na**

TABLEAU 2.3: La conjugaison perfective du verbe *katab* « écrire »

2.2.2.2.2 Les marqueurs de sujet à l'imperfectif

Nous partons d'un exemple qui illustre la forme la plus simple d'un verbe à l'imperfectif. De la racine */ktb/* « écrire », par exemple, est issu le verbe imperfectif indicatif *jaktubu*. Le radical étant *aktub*, nous constatons la présence d'un préfixe *j-* et d'un suffixe *-u*. Le préfixe est le marqueur de sujet et il exprime les traits de la 3MS. Le suffixe est probablement le morphème de mode. Ce qui appuie cette idée c'est que ce suffixe subit un changement au subjonctif (*j-aktub-**a***) et est tronqué au jussif (*j-aktub∅*). Pour sa part, le préfixe varie selon les traits morphosyntaxiques exprimés : *ʔ-aktub-u* (1S), ***n**-aktb-u* (1P), ***t**-aktub-u* (3FS ou 2MS), etc.

Avec certaines formes, la composition de l'imperfectif est plus complexe. À la 2FS, par exemple, l'imperfectif indicatif issu de la racine */ktb/* est *taktubi:na*. Dans cette forme, nous reconnaissons la présence du radical commun *aktub*. Nous pouvons également isoler le morphème *-na* et postuler qu'il est le morphème de mode pour la simple raison qu'il est tronqué au subjonctif et au jussif dans la forme commune *taktubi:*. Ce morphème est également présent à la 2MP (*taktubu:na*) et à la 3MP (*jaktubu:na*) de l'indicatif et subit les mêmes changements au subjonctif et au jussif (*taktubu:* et *jaktubu:* respectivement).

Cependant, à la 2FP et à la 3FP, le morphème *-na* résiste à la troncation. L'imperfectif indicatif issu de la racine */ktb/* à la 2FP est *taktubna*, mais le subjonctif et le jussif sont également *taktubna*. Cette persistance de la forme prouve que *-na* n'est pas un morphème de mode à la 2FP et à la 3FP Pour préserver la régularité du paradigme, nous pouvons supposer qu'un morphème de mode *-na* (caractéristique du pluriel) a subi une haplologie : *taktub-na-na ⇒taktub-∅-na. Ce qui pourrait motiver une telle prise de position est l'haplologie du morphème de mode *-na* constatée avec la forme énergétique de l'imperfectif :

taktubu:na (IMP.IND.3MP) + nna (le suffixe de l'énergétique) ⇒taktub-u:nna ⇒taktubu-nna (abrègement de la voyelle dans une syllabe lourde fermée).

Au duel, le morphème que nous pouvons postuler qu'il est le morphème de mode est *-ni*, visible à l'indicatif de la 2D *taktuba:-ni*, par exemple, et tronqué au subjonctif et au jussif dans la forme commune *taktuba:*.

La question qui se pose à présent concerne les autres suffixes amalgamés aux suffixes de mode. À l'imperfectif indicatif issu de la racine */ktb/* à la 2FS *taktubi:na*, par exemple, et après avoir isolé le radical *aktub* et le morphème de mode *-na*, nous constatons la présence d'un préfixe *t-* et d'un suffixe *-i:*. Ces affixes se retrouvent aux trois modes et ne peuvent donc faire partie des affixes de mode. Ils varient par ailleurs en fonction des traits morphosyntaxiques : *t-* est remplacé par *j-* à la 3MS *jaktubu*, et *-i:* est remplacé par *-a:* à la 2D *taktuba:ni*. S'il est traditionnellement admis que les préfixes de l'imperfectif expriment le trait de la personne et les suffixes ceux du genre et du nombre, cette affirmation ne peut s'avérer justifiée dans ce cas précis. Le préfixe *t-* à la 2FS porte également une indication de genre (le féminin par opposition au masculin exprimé par le préfixe *j-*).

Cette complexité morphologique de l'imperfectif pose plusieurs problèmes à toute tentative de définir les frontières des marqueurs de sujet et leur composition interne. Les grammairiens de la tradition ont, par exemple, proposé un traitement des affixes présentés ici en termes de *pronoms conjoints*. Cependant, cette étiquette n'était valable, selon eux, que pour les suffixes suivants : *-i:* (2FS), *-a:* (D[-1]), *-u:* (MP[-1]) et *-na* (2FP et 3FP), qui apparaissent adjacents au radical verbal et précédant les suffixes de mode. Les préfixes de l'imperfectif ne sont, toujours selon ces grammairiens, que des marqueurs d'aspect. Cette analyse de la morphologie de l'imperfectif est toutefois confrontée au problème suivant : si les préfixes de l'imperfectif sont des marqueurs d'aspect, pourquoi prennent-ils des formes différentes selon les traits de genre, de nombre et de personne ? Leur variation suivant ces traits morphosyntaxiques prouve qu'ils représentent beaucoup plus qu'une simple information aspectuelle.

Une autre analyse de ces affixes est proposée par Lumsden et Haleform (2003). Ces auteurs affirment que les préfixes de l'imperfectif sont des *pronoms sujets* (qu'ils considèrent comme clitiques) alors que les suffixes sont une sorte de marqueurs d'accord liés aux premiers.

Finalement, une troisième hypothèse est défendue. Selon cette dernière hypothèse défendue, entre autres, par Fassi Fehri (1993), Habash (2007) et Noyer (1997, 1992), les préfixes et les suffixes

appartiennent à une même entité morphologique et constituent des morphèmes discontinus que nous pouvons qualifier de circumfixes en adoptant la définition de (Anderson, 1992, p. 327), qui définit les circumfixes comme étant : « simultaneous prefixation and suffixation corresponding to a single unit of morphological form. ». Ce qui pourrait appuyer ce point de vue c'est que d'autres morphèmes discontinus sont attestés en arabe, à l'instar du morphème du pluriel non concaténatif (appelé pluriel *brisé*), du morphème qui représente la racine, celui qui représente le schème morphologique, mais également le morphème de la négation *ma:*...-ʃ retrouvé dans certaines variétés dialectales.

Nous signalons que les marqueurs de sujet de l'imperfectif, et quelle que soit l'approche que l'on adopte pour traiter leur morphophonologie, sont classifiables selon les trois traits morpho-syntaxiques de genre, de nombre et de personne. Les trois tableaux suivants nous présentent la liste complète de ces formes. Dans le tableau (2.4), nous présentons la conjugaison imperfective à l'indicatif du verbe *katab* « écrire », le tableau (2.5) présente le subjonctif et le tableau (2.6) le jussif. Dans ces tableaux, les morphèmes que nous pouvons considérer comme marqueurs de sujet sont en caractères gras.

	Singulier	Duel	Pluriel
1	?-aktub-u	**n**-aktub-u	**n**-aktub-u
2M	**t**-aktub-u	**t**-aktub-**a:**-ni	t-aktub-**u:**-na
2F	**t**-aktub-**i:**-na	**t**-aktub-**a:**-ni	**t**-aktub-**na**
3M	**j**-aktub-u	**j**-aktub-**a:**-ni	**j**-aktub-**u:**-na
3F	**t**-aktub-u	t-aktub-**a:**-ni	**j**-aktub-**na**

TABLEAU 2.4: L'imperfectif indicatif du verbe *katab* « écrire »

	Singulier	Duel	Pluriel
1	(lan) ?-aktub-a	(lan) **n**-aktub-a	(lan) **n**-aktub-a
2M	(lan) t-aktub-a	(lan) **t**-aktub-**a:**	(lan) **t**-aktub-**u:**
2F	(lan) t-aktub-**i:**	(lan) **t**-aktub-**a:**	(lan) **t**-aktub-**na**
3M	(lan) j-aktub-a	(lan) **j**-aktub-**a:**	(lan) **j**-aktub-**u:**
3F	(lan) t-aktub-a	(lan) t-aktub-**a:**	(lan) **j**-aktub-**na**

TABLEAU 2.5: L'imperfectif subjonctif du verbe *katab* « écrire »

Nous terminons cette présentation brève des marqueurs de sujet à l'imperfectif en signalant que ces derniers sont, comme on s'y attendrait, semblables aux marqueurs de sujet au perfectif, aux marqueurs d'objet et aux pronoms indépendants en ceci qu'ils peuvent référer à des topiques,

	Singulier	Duel	Pluriel
1	(lam) ?-aktub	(lam) **n**-aktub	(lam) **n**-aktub
2M	(lam) **t**-aktub	(lam) **t**-aktub-**a:**	(lam) **t**-aktub-**u:**
2F	(lam) **t**-aktub-**i:**	(lam) **t**-aktub-**a:**	(lam) **t**-aktub-**na**
3M	(lam) **j**-aktub	(lam) **j**-aktub-**a:**	(lam) **j**-aktub-**u:**
3F	(lam) **t**-aktub	(lam) **t**-aktub-**a:**	(lam) **j**-aktub-**na**

TABLEAU 2.6: L'imperfectif jussif du verbe *katab* « écrire »

comme dans les exemples suivants, où les marqueurs de sujet *t-* et *j-* réfèrent aux pronoms indépendants topiques *huwa* et *hija* respectivement :

(2.6) a. *huwa*$_i$ *j*$_i$*-* *uhibb* *-u* *ha:*dh*a:*
lui 3MS- aimer.IMP -ind cei
« Lui, il aime ceci »

b. *hija*$_j$ *t*$_j$*-* *uhibb* *-u* *ha:*dh*a:*
elle 3FS- aimer.IMP -ind ceci
« Elle, elle aime ceci »

2.2.2.2.3 Les marqueurs de sujet à l'impératif

Pour les grammairiens de la tradition, il n'y avait nul doute que l'impératif est dérivé de l'imperfectif jussif à la 2e personne moyennant la troncation du préfixe, l'assimilation de la première voyelle du radical avec la dernière voyelle et l'insertion d'un coup de glotte (noté par le ?-). Voici un exemple qui illustre la formation de l'impératif : de la forme *taktub*, qui est la conjugaison imperfective (au jussif) issue de la racine */ktb/* à la 2MS, l'impératif est formé en tronquant le préfixe *t-*, en assimilant les deux voyelles du radical (a → u) et en insérant un coup de glotte en début de mot. La forme obtenue est *?uktub* « écris ! ».

Nous pouvons considérer que le radical de l'impératif est la forme réalisée à la 2MS. Les quatre autres formes sont dérivées par l'insertion des suffixes de l'imperfectif que les grammairiens considéraient comme pronoms sujets. Ces suffixes de la 2e personne ne varient qu'en nombre et en genre : *-i:* (FS), *-na* (FP) et *-u:* (MP). Cependant, le suffixe *-a:* ne marque que le nombre (le duel) et garde la même forme au masculin et au féminin. Le tableau en (2.7) présente les formes rencontrées à l'impératif. Dans ce tableau, les marqueurs de sujet sont écrits en caractères gras.

	Singulier	Duel	Pluriel
2M	ʔuktub	ʔuktub-a:	ʔuktub-u:
2F	ʔuktub-i:	ʔuktub-a:	ʔuktub-na

TABLEAU 2.7: L'impératif du verbe *katab* « écrire »

L'impératif, en tant que conjugaison, ne se réalise qu'à la 2e personne[9]. Pour la 1re et la 3e personnes, le rôle discursif de l'impératif peut être assumé par l'injonctif[10], à savoir l'emploi de l'imperfectif jussif précédé de la particule *li-* (que). Le tableau (2.8) présente l'injonctif.

	Singulier	Duel	Pluriel
1	li-ʔ-aktub	li-n-aktub	li-n-aktub
2M	li-t-aktub	li-t-aktub-a:	li-t-aktub-u:
2F	li-t-aktub-i:	li-t-aktub-a:	li-t-aktub-na
3M	li-j-aktub	li-j-aktub-a:	li-j-aktub-u:
3F	li-t-aktub	li-t-aktub-a:	li-j-aktub-na

TABLEAU 2.8: L'injonctif du verbe *katab* « écrire »

Cet emploi justifie un traitement unifié de l'impératif et de l'imperfectif, d'autant plus que ce lien est indéniable quand nous prenons en compte la forme de l'impératif négatif (ou le prohibitif). Cette dernière s'obtient par l'emploi de l'imperfectif jussif (à la 2e personne) précédé de la particule de prohibition/négation *la:* « ne pas », comme le montre le tableau en (2.9).

	Singulier	Duel	Pluriel
2M	la: t-aktub	la: t-aktub-a:	la: t-aktub-u:
2F	la: t-aktub-i:	la: t-aktub-a:	la: t-aktub-na

TABLEAU 2.9: Le prohibitif du verbe *katab* « écrire »

Cette relation entre l'impératif et l'imperfectif jussif est visible syntaxiquement dans les systèmes hypothétiques, où les deux formes sont corrélées « comme dans le célèbre : *ʔaslim taslam* (convertis-toi à l'islam, tu seras sauvé) » (Larcher, 2003) : le premier verbe *ʔaslim* est à l'impératif alors que le deuxième *taslam* est à l'imperfectif jussif.

De cette présentation brève des marqueurs de sujet attachés au perfectif, à l'imperfectif et à l'impératif, nous pouvons conclure que ces formes s'insèrent dans deux paradigmes différents : le paradigme du perfectif et celui de l'imperfectif (l'impératif pouvant être déduit de l'imperfectif

[9] Cela semble être une caractéristique universelle dans les langues naturelles.
[10] L'injonctif n'est cependant pas restreint à ces deux personnes, comme le montre le tableau (2.8).

jussif).

Une autre possibilité est de traiter l'impératif et le perfectif comme faisant partie du même paradigme, étant donné que les deux sont formés selon le modèle [radical+suffixe], alors que l'imperfectif est formé selon le modèle [préfixe+radical+(suffixe)].

Nous avons brièvement présenté les pronoms indépendants et les marqueurs d'arguments. Les pronoms indépendants sont répartis en deux sous-groupes selon leur cas . des pronoms indépendants nominatifs et des pronoms indépendants accusatifs. Les marqueurs d'arguments sont répartis en deux sous-groupes également : des marqueurs de sujet et des marqueurs d'objet. Cette présentation nous a permis de soulever certains problèmes qui touchent la composition morphologique et la catégorisation de ces marqueurs. L'un de ces problèmes concerne les marqueurs de sujet à l'imperfectif. Nous avons signalé l'existence de trois approches concernant leur morphologie : selon l'une d'elles, les suffixes sont les seuls marqueurs de sujet dans ce contexte ; selon une autre approche, seuls les préfixes le sont et selon les tenants de la troisième, les deux sortes d'affixes font partie d'un même morphème discontinu. Ce problème mérite plus d'attention et sera par conséquent exposé dans la sous-section suivante.

2.2.2.3 La morphologie de l'imperfectif

Nous avons des raisons de soutenir l'hypothèse selon laquelle les marqueurs de sujet à l'imperfectif sont les morphèmes que nous isolons dans le tableau (2.10).

	Singulier	Duel	Pluriel
1	?-	n-	n-
2M	t-	t-...-a:	t-...-u:
2F	t-...-i:	t-...-a:	t-...-na
3M	j-	j-...-a:	j-...-u:
3F	t-	t-...-a:	j-...-na

TABLEAU 2.10: Les marqueurs de sujet à l'imperfectif

Selon cette hypothèse, les marqueurs de sujet à l'imperfectif sont constitués à la fois de préfixes et de circumfixes[11]. Il n'est pas surprenant d'avoir les deux sortes d'affixes dans le même paradigme morphologique étant donné que la morphologie de la langue arabe n'est pas strictement concaténative. Le morphème du pluriel, par exemple, peut être un suffixe concaténé au nom singulier (ce qui forme le pluriel concaténatif, appelé également le pluriel *sain*). Ce morphème se réalise également sous forme de morphème discontinu (c'est le pluriel non concaténatif, appelé également le pluriel *brisé*). Le pluriel du nom singulier *kitaʔb* « livre », par exemple, est formé de manière non concaténative en changeant le schème morphologique du singulier pour obtenir *kutub*. Le pluriel du nom déverbal *qaːriʔ* « lecteur » est, par contre, concaténatif et s'obtient par l'ajout du suffixe *-uːna* au singulier. Il en résulte la forme *qaːriʔuːna* « lecteurs ».

En outre, la morphophonologie de ces marqueurs de sujet nous rappelle les formes de la conjugaison perfective, du paradigme des marqueurs d'objet et même de celui des pronoms indépendants. Pour ne prendre qu'un seul exemple représentatif, comparons les formes suivantes toutes de la 2MD : la première est le pronom indépendant nominatif *ʔantumaː*, la deuxième est le marqueur d'objet *-kumaː*, la troisième est le marqueur de sujet au perfectif *-tumaː* et la dernière est le marqueur de sujet à l'imperfectif (un morphème discontinu) *t-...-aː*.

Si nous maintenons que le tableau (2.10) représente bel et bien les marqueurs de sujet à l'imperfectif, nous constatons néanmoins que la distribution des traits de nombre, de genre et de personne se fait différemment selon que l'on est en présence de morphèmes discontinus ou en présence de préfixes :

1. Dans les morphèmes discontinus, le préfixe exprime généralement la personne, à l'exception des deux marqueurs de sujet de 3MD et de 3FD, où les deux préfixes *j-* et *t-* expriment le genre également (le masculin et le féminin respectivement). Le suffixe exprime, quant à lui, les deux autres traits, à l'exception des suffixes du duel, qui n'expriment que le nombre.
2. Quand le marqueur de sujet est un préfixe, il exprime plus d'un seul trait : la personne et le genre (*j-* et *t-*) ou la personne et le nombre (*ʔ-* et *n-*).

Cette distribution singulière des traits morphosyntaxiques semble contredire les conclusions de Fassi Fehri (2000), qui affirme que le trait de nombre n'est jamais réalisé par un préfixe et que le trait de personne n'est jamais réalisé par un suffixe alors que le genre est libre d'apparaître dans l'une ou l'autre de ces deux positions selon lui.

[11] D'autres langues comportent également des marqueurs de sujet morphologiquement mixtes. En wolof, par exemple, les clitiques sujets sont composés de proclitiques et de morphèmes discontinus : ma...-a (1S), nga (2S), mu (3S), nu (1P), ngeen...-een (2P) et ɔu (3P). Voir (N'Diaye-Correard, 2003).

Cette distribution donne lieu à deux phénomènes morphosyntaxiques : l'impératif, dans lequel le suffixe qui caractérise l'imperfectif est absent, et la possibilité de l'insertion d'un SN nominatif postverbal avec les préfixes de 3S. Les deux paragraphes suivants constituent un examen de ces deux points.

2.2.2.3.1 L'impératif et les marqueurs de sujet

Nous avons avancé l'idée selon laquelle l'impératif et l'imperfectif font partie du même paradigme, étant donné que le premier peut être déduit du deuxième au mode jussif. Cependant, une question se pose ici : si c'est le cas, comment alors expliquer cette différence entre une conjugaison traditionnellement désignée comme étant une conjugaison à *préfixes* (l'imperfectif) et une conjugaison à *suffixes* (le perfectif) ?

L'absence de préfixes dans la conjugaison impérative pourrait nous amener à conclure, comme le font Lumsden et Haleform (2003), que ces morphèmes sont les marqueurs de sujet, puisqu'ils disparaissent à l'impératif, un comportement que nous pouvons reconnaître dans plusieurs langues, dont le français comme dans l'exemple suivant :

(2.7) vous sortez → sortez !

En (2.7), le pronom clitique sujet *vous* est omis à l'impératif alors que le suffixe *-ez* est réalisé à l'indicatif et à l'impératif. Le même phénomène toucherait l'impératif en arabe dans l'exemple suivant :

(2.8)	*lam*	***t-***	*axruɣ*	*-**u:***	→	*ʔuxruɣ*	*-**u:** !*
	ne.pas	2	sortir.IMP	2MP		sortir.IMPÉ	-2MP

« Vous n'êtes pas sortis » « Sortez ! »

Cependant, nous pensons que cet argument est empiriquement inadéquat. En fait, si nous restons dans le cadre d'une comparaison entre l'arabe et le français, nous pouvons constater que le français emploie une stratégie différente de celle employée par la langue arabe quand il s'agit du prohibitif (une forme d'impératif). Le pronom clitique sujet est absent de l'impératif négatif français : *ne sortez pas !* alors que le préfixe de l'imperfectif ne subit pas le même sort en arabe : *la:* ***t-****axruɣ-**u:*** !

Nous avons donc des raisons de penser que les préfixes ne peuvent constituer à eux seuls les marqueurs de sujet quand un suffixe est également réalisé. En ce qui concerne l'impératif (dans sa forme la plus simple), l'absence de préfixes va de soi : puisque cette conjugaison n'est possible qu'à la 2e personne, le locuteur n'a pas besoin d'exprimer la personne par un morphème distinct, une information déjà présente dans la situation du discours et déduite de la conjugaison même du verbe.

2.2.2.3.2 Les SN et les marqueurs de sujet

Malgré la différence morphologique entre l'impératif et l'imperfectif, ces deux conjugaisons ont le point commun suivant : un marqueur de sujet spécifié pour le nombre (ou pour un trait de nombre différent du singulier) ne peut se réaliser en même temps que le SN nominatif postverbal qui constitue son référent, et ce, dans les contextes non marqués. Ce comportement, qui pourrait être motivé pragmatiquement [12] en ce qui concerne l'impératif, est très révélateur en ce qui concerne l'imperfectif. Ainsi, seuls les deux préfixes *j-* et *t-* de l'imperfectif peuvent coexister - dans les contextes non marqués - avec des SN nominatifs postverbaux. Ces derniers peuvent être au singulier (2.9a), au duel (2.9b) ou au pluriel (2.9c et d) :

(2.9) a. *j-* *aktub* *-u* *sa:mi:* *risa:lat* *-a* *-n*
3M- écrire.IMP -Ind Sami lettre -ACC -Indé
« Sami écrit une lettre »

b. *t-* *aktub* *-u* *l-binta:ni* *risa:lat* *-a* *-n*
3F- écrire.IMP -Ind les-deux.filles lettre -ACC -Indé
« Les deux filles écrivent une lettre »

c. *t-* *aktub* *-u* *n-nis:ʔ-u* *risa:lat* *-a* *-n*
3F- écrire.IMP -Ind les-femmes-NOM lettre -ACC -Indé
« Les femmes écrivent une lettre »

d. *j-* *aktub* *-u* *l-ʔawla:d-u* *risa:lat* *-a* *-n*
3M- écrire.IMP -Ind les-garçons-NOM lettre -ACC -Indé
« Les garçons écrivent une lettre »

Aucun autre marqueur de sujet à l'imperfectif ne partage cette distribution avec les deux préfixes *j-* et *t-*. Les mêmes remarques sont valables pour les marqueurs de sujet au perfectif.

[12] Dans le chapitre intitulé *la nature des pronoms*, Benveniste (1966) distingue les pronoms de 1re et de 2e personne, qui réfèrent à des entités présentes dans le discours, et les pronoms de 3e personne, qui réfèrent à des « non-personnes » : « La troisième personne représente en fait le membre non marqué de la corrélation de personne. » (Benveniste, 1966, p. 254) De par leur fonction discursive, les pronoms ont donc un comportement syntaxique différent : les pronoms de 3e personne peuvent coexister avec des SN alors que les pronoms de 1re et de 2e personne ne le peuvent pas.

En fait, les deux marqueurs de sujet *-a* et *-at*, qui sont des marqueurs du perfectif, exhibent ces mêmes propriétés distributionnelles, comme l'illustrent les exemples suivants :

(2.10) a. *katab* *-a* *sa:mi:* *risa:lat* *-a* *-n*
écrire.PER -3M Sami lettre -ACC -Indé
« Sami a écrit une lettre »

b. *katab* *-at* *l-binta:ni* *risa:lat* *-a* *-n*
écrire.PER -3F les-deux.filles lettre -ACC -Indé
« Les deux filles ont écrit une lettre »

c. *katab* *-at* *n-nis:ʔ-u* *risa:lat* *-a* *-n*
écrire.PER -3F les-femmes-NOM lettre -ACC -Indé
« Les femmes ont écrit une lettre »

d. *katab* *-a* *l-ʔawla:d-u* *risa:lat* *-a* *-n*
écrire.PER -3M les-garçons-NOM lettre -ACC -Indé
« Les garçons ont écrit une lettre »

Que tous ces marqueurs soient à la 3e personne peut s'expliquer par l'hypothèse de Benveniste (1966). Ce qui pose problème c'est qu'ils s'accordent en genre et en personne seulement avec le SN postverbal. Cela n'est qu'une facette d'un problème plus large appelé les *asymétries de l'accord*, que nous traiterons dans le troisième chapitre. Notons cependant que le degré de spécification des traits pourrait fournir une piste de réponse : les seuls marqueurs de sujet du singulier qui peuvent coexister avec des SN postverbaux nominatifs dans les contextes non marqués encodent le genre d'une manière marquée : *-a* et *j-* pour le masculin et *-at* et *t-* pour le féminin.

Ce comportement singulier des marqueurs de sujet de 3S pourrait expliquer, en partie du moins, la diversité des approches par rapport à ces affixes : Wright (1898), par exemple, considère que seul *-at* est un pronom, *-a* n'étant, selon lui, qu'une marque morphologique aspectuelle distinctive du perfectif. Fassi Fehri (1993) propose deux analyses possibles des deux marqueurs de sujet *-a* et *-at* : il les considère comme des pronoms quand ils n'apparaissent pas avec des SN nominatifs postverbaux non marqués, auquel cas ils sont spécifiés pour tous leurs traits (3MS pour *-a* et 3FS pour *-at*). Quand ils apparaissent avec de tels SN, il les considère comme des marqueurs d'accord, auquel cas ils ne sont spécifiés que pour le trait de genre (M pour *-a* et F pour *-at*). Lumsden et Haleform (2003) considèrent que les préfixes de l'imperfectif sont des pronoms clitiques, alors que la majorité des grammairiens les considèrent comme étant des marqueurs morphologiques qui désignent l'imperfectif (les appelant ainsi *ħuru:f lmuḍa:riʕ* « les particules de l'imperfectif »).

2.2.3 Les formes retenues

Les tableaux suivants présentent les formes qui constituent notre objet d'étude dans ce chapitre : le tableau (2.11) est une reproduction du tableau (2.1) (les pronoms indépendants), le tableau (2.12) est une reproduction du tableau (2.3) (les marqueurs de sujet au perfectif) et du tableau (2.10) (les marqueurs de sujet à l'imperfectif) et le tableau (2.13) est une reproduction du tableau (2.2) (les marqueurs d'objet).

	Singulier		Duel		Pluriel	
	NOM	ACC	NOM	ACC	NOM	ACC
1	ʔana:	ʔijja:ja	naħnu	ʔijja:na:	naħnu	ʔijja:na:
2M	ʔanta	ʔijja:ka	ʔantuma:	ʔijja:kuma:	ʔantum	ʔijja:kum
2F	ʔanti	ʔijja:ki	ʔantuma:	ʔijja:kuma:	ʔantunna	ʔijja:kunna
3M	huwa	ʔijja:hu	huma:	ʔijja:huma:	hum	ʔijja:hum
3F	hija	ʔijja:ha:	huma:	ʔijja:huma:	hunna	ʔijja:hunna

TABLEAU 2.11: Les pronoms indépendants (nominatifs et accusatifs)

	Singulier		Duel		Pluriel	
	PER	IMP	PER	IMP	PER	IMP
1	-tu	ʔ-	-na:	n-	-na:	n-
2M	-ta	t-	-tuma:	t-...-a:	-tum	t-...-u:
2F	-ti	t-...-i:	-tuma:	t-...-a:	-tunna	t-...-na
3M	-a	j-	-a:	j-...-a:	-u:	j-...-u:
3F	-at	t-	-ata:	t-...-a:	-na	j-...-na

TABLEAU 2.12: Les marqueurs de sujet

	Singulier	Duel	Pluriel
1	-i: (-ni:, -ja)	-na:	-na:
2M	-ka	-kuma:	-kum
2F	-ki	-kuma:	-kunna
3M	-hu	-huma:	-hum
3F	-ha:	-huma:	-hunna

TABLEAU 2.13: Les marqueurs d'objet

Dans les sections qui suivent, nous abordons la morphologie et la syntaxe de ces différentes formes. Nous montrons que les marqueurs d'arguments et les pronoms indépendants sont étroitement reliés ; nous ferons également la part des différences et des ressemblances entre les marqueurs d'objet et les marqueurs de sujet. La section (2.3) est consacrée aux pronoms indépendants et à leur distribution, la section (2.4) est consacrée aux marqueurs d'arguments et la section (2.5) aux relations morphophonologiques et morphosyntaxiques entre les marqueurs d'arguments et les pronoms indépendants.

2.3 Les pronoms indépendants

Dans cette section, nous présentons, avec plus de détails, les deux sortes de pronoms indépendants : les nominatifs et les accusatifs. Nous mettons l'accent, essentiellement, sur leur distribution.

2.3.1 Les pronoms indépendants nominatifs

Les pronoms indépendants nominatifs (désormais NOM) n'ont pas la même distribution selon le type de phrase (équative [13] ou verbale [14]). Ils sont toujours non essentiels dans la phrase verbale, mais ils peuvent constituer un élément essentiel de la phrase équative.

2.3.1.1 Dans la phrase verbale

Les pronoms indépendants NOM ne peuvent faire partie des constituants immédiats de la phrase verbale. Autrement dit, ils ne peuvent occuper la fonction du sujet et leur présence dans ce genre de phrases est conditionnel à l'emploi des marqueurs de sujet (2.11a) ou d'objet (2.11b). Étant associé à des effets d'emphase, de topicalisation ou de focalisation [15], leur emploi dans

[13] La phrase équative, dans notre travail, désigne toute phrase construite sans verbe et dont la tête est nominale.

[14] Nous entendons par « phrase verbale » toute phrase dont la tête est verbale, que son élément initial soit ce genre de tête ou un élément dont la tête est nominale.

[15] La traduction des exemples en français reflète ces différences pragmatiques. La traduction par la tournure « c'est ... qui/que » exprime une focalisation, alors que les traductions par un SN ou un pronom fort détaché expriment une topicalisation.

ce contexte est pragmatiquement marqué. Cela s'observe quand ils sont préposés, comme en (2.11a,b), ou postposés, comme en (2.11c,d) :

(2.11) a. *huwa*$_i$ *katab* *-a*$_i$ *risa:lat* *-a* *-n*
lui écrire.PER -3MS lettre -ACC -Indé
« Lui, il a écrit une lettre »

b. *huwa*$_i$ *raʔaj* *-tu* *-hu*$_i$ *ʔamsi*
lui voir.PER -1S -le hier
« Lui, je l'ai vu hier »

c. *katab* *-a*$_i$ *huwa*$_i$ *risa:lat* *-a* *-n*
écrire.PER -3MS lui lettre -ACC -Indé
« Il a écrit une lettre, lui »

d. *raʔaj* *-tu* *-hu*$_i$ *huwa*$_i$ *ʔamsi*
voir.PER -1S -le lui hier
« Je l'ai vu hier, lui »

Nous pouvons de ce fait postuler que ces pronoms sont *périphériques* [16] dans la phrase verbale ; ils sont ainsi prédisposés à être associés non seulement à des marqueurs de sujet, comme en (2.11a,c), mais également à des marqueurs d'objet, comme en (2.11b,d). Leur caractère périphérique les prive d'être des arguments de la tête verbale et constitue un argument pour considérer les marqueurs d'arguments en (2.11) comme étant les arguments du verbe. Cette hypothèse sera exposée et défendue dans le troisième chapitre.

2.3.1.2 Dans la phrase équative

Dans la phrase équative, les pronoms indépendants NOM peuvent jouer plusieurs rôles qui affectent la spécification de leurs traits morphosyntaxiques et leur degré de référence. Ils peuvent en fait occuper l'une ou l'autre des deux fonctions principales (le sujet et le prédicat), auquel cas ils s'accordent en tous traits avec le syntagme qui occupe l'autre fonction. Un exemple est donné en (2.12).

(2.12) ***hunna*** *ʔaxawa:t* *-u* *l-* *munðir* *-i*
elles sœurs -NOM le- Mundhir -GÉN
« Elles sont les sœurs d'El Mundhir »

[16] Ce terme est employé dans un sens proche de ce que Adger *et al.* (2004) appellent les *périphéries*.

Par ailleurs, les pronoms indépendants NOM de 3e personne peuvent jouer un rôle dans la phrase équative qui nous rappelle celui joué par les marqueurs de sujet de la 3e personne [17]. Ils marquent un accord en nombre et en genre seulement entre le sujet et le prédicat et peuvent être employés avec un sujet à la 1re personne, comme en (2.13a), ou à la 2e personne, comme en (2.13b). Ils sont ainsi sous-spécifiés pour la personne, tout comme les marqueurs de sujet de 3e personne peuvent être sous-spécifiés pour le nombre, et n'assument pas de rôle référentiel.

(2.13) a. *ʔana:* ***huwa*** *l-* *fa:ʕil* *-u* (1re personne)
moi lui le- faiseur -NOM
« C'est moi le malfaiteur »

b. *ʔanta* ***huwa*** *l-* *fa:ʕil* *-u* (2e personne)
toi lui le- faiseur -NOM
« C'est toi le malfaiteur »

c. *zajd* *-un* ***huwa*** *l-* *fa:ʕil* *-u* (3e personne)
Zayd -NOM lui le- faiseur -NOM
« C'est Zayd le malfaiteur »

Dans un troisième contexte, les pronoms indépendants NOM de la 3e personne peuvent être des explétifs dénués de toute référentialité, comme dans les exemples suivants [18] :

(2.14) a. ***huwa*** *l-* *kawn* *-u* *ħajj* *-u* *-n*
lui l'- univers -NOM vivant -NOM -Indé
« Le fait est que l'univers est un vivant »

b. ***hija*** *l-* *ʔumm* *-u* *madrasat* *-u* *-n*
elle la- mère -NOM école -NOM -Indé
« Le fait est que la mère est une [véritable] école »

Ils s'accordent ainsi avec le premier élément de la phrase équative qu'ils introduisent, en genre et en personne seulement, puisque seuls les pronoms de 3S peuvent être employés dans ce contexte.

Les pronoms indépendants NOM ont donc une distribution très élargie ; ils peuvent être des arguments ou des éléments périphériques dans la structure et ils apparaissent aussi bien dans les phrases verbales que dans les phrases équatives. Nous constatons, dans la section suivante, que la distribution des pronoms indépendants accusatifs est plus restreinte.

[17] Voir Fassi Fehri (1992).
[18] L'exemple en (2.14b) provient de (Hassan, 1987, tome 1, p. 253).

2.3.2 Les pronoms indépendants accusatifs

L'ordre non marqué en arabe est VSO. Dans cet ordre, l'objet est postverbal dans les contextes non marqués, comme en (2.15a). Toutefois, quand l'objet est pronominalisé sous forme de pronom indépendant accusatif (désormais ACC), ce dernier doit être topicalisé (2.15c) [19] et ne peut rester postverbal (2.15d). Notons que le SN également peut être préposé à la tête verbale, comme en (2.15b). Cette topicalisation de l'objet est associée à une valeur pragmatique marquée, que l'élément topicalisé soit le SN ou le pronom indépendant.

(2.15) a. *qara?* *-tu* *r-* *risa?lat* *-a*
lire.PER -1S la- lettre -ACC
« J'ai lu la lettre »

b. *?ar-* *risa?lat* *-a* *qara?* *-tu*
la- lettre -ACC lire.PER -1S
« C'est la lettre que j'ai lue (et non le roman) »

c. *?ijja:ha:* *qara?* *-tu*
elle(ACC) lire.PER -1S
« C'est elle que j'ai lue (et non une autre) »

d. **qara?* *-tu* *?ijja:ha:*
lire.PER -1S elle(ACC)

La pronominalisation du SN objet se fait par le recours à un marqueur d'objet dans les cas non marqués. Ce marqueur d'objet est alors postposé à la tête verbale comme illustré par l'exemple suivant :

(2.16) *qara?* *-tu* *-ha:*
lire.PER -1S -la
« Je l'ai lue »

L'argument interne du verbe (l'objet) est donc postverbal quand il est un marqueur d'objet et préverbal quand il est un pronom indépendant ACC. L'argumentalité des pronoms indépendants ACC ne se limite cependant pas à ce contexte. Dans les structures à double objet, lorsque deux marqueurs d'objet forment une suite, le deuxième complément peut se réaliser comme un pronom indépendant ACC, comme en (2.17b) [20] :

[19] La topicalisation est entendue ici comme étant le procédé par lequel un constituant est préposé.

[20] La différence entre l'emploi du marqueur d'objet et l'emploi du pronom indépendant est surtout une différence de registre, comme expliqué dans la section (2.5.2.1).

(2.17) a. *wahab* *-tu* *-ka* *-ha:*
donner.PER -1S -te -la
« Je te l'ai donnée »

b. *wahab* *-tu* *-ka* *ʔijja:ha:*
donner.PER -1S -te elle(ACC)
« Je te l'ai donnée »

Nous avons présenté la distribution des pronoms indépendants NOM et ACC dans le but de les comparer aux marqueurs d'arguments. La question de l'argumentalité des uns et des autres sera au coeur de cette comparaison. Mais avant d'examiner cette question, nous avons besoin d'avoir une meilleure compréhension du comportement des marqueurs d'arguments. La section (2.4) est dédiée à une étude descriptive de la distribution et de la morphosyntaxe de ces marqueurs.

2.4 Les marqueurs d'arguments

Les marqueurs d'arguments constituent un groupe homogène quand ils sont contrastés aux pronoms indépendants. Toutefois, ce groupe comporte plusieurs asymétries qui le scindent en deux sous-groupes distincts. La section (2.4.1) portera sur les ressemblances distributionnelles et morphologiques entre ces deux sortes de marqueurs alors que la section (2.4.2) mettra l'accent sur les différences.

2.4.1 Les symétries

Les marqueurs d'arguments partagent certaines propriétés distributionnelles dont la plus importante est celle mentionnée par Eloussfourri (1998). L'auteure souligne que les deux sous-classes de marqueurs se comportent de la même manière vis-à-vis de certaines structures. Dans les contextes non marqués, quand un marqueur d'argument est spécifié pour le nombre, il ne peut coexister avec un SN nominatif postposé dans les cas non marqués. En (2.18a), le marqueur d'objet de 3MS *-hu* et le SN MS *s'adï:qï:* « mon ami » sont coréférentiels mais incompatibles. La phrase qui en résulte est agrammaticale. Ce comportement caractérise également le marqueur de

sujet de 3MP *-u:* en (2.18b), qui ne peut co-apparaître avec le SN MP *ʔas'diqa:ʔi:* « mes amis » dans les contextes non marqués :

(2.18) a. **raʔj -tu -hu$_i$ s'adi:q -i:$_i$* (marqueur d'objet)
voir.PER -1S -le ami -mon

a'. *raʔj -tu -hu*
voir.PER -1S -le
« Je l'ai vu »

a". *raʔj -tu s'adi:q -i:*
voir.PER -1S ami -mon
« J'ai vu mon ami »

b. **nadʒaħ -u:$_i$ ʔas'diqa:ʔ -i:$_i$* (marqueur de sujet)
réussir.PER -3MP amis -mes

b'. *nadʒaħ -u:*
réussir.PER -3MP
« Ils ont réussi »

b". *nadʒaħ -a$_i$ ʔas'diqa:ʔ -i:$_i$*
réussir.PER -3M amis -mes
« Mes amis ont réussi »

En plus de cette distribution particulière, les deux sortes de marqueurs, quand ils sont spécifiés pour le nombre, peuvent coexister avec de tels SN si ces derniers ne sont pas des arguments de la tête dont dépend le marqueur. Cela est également possible avec des pronoms indépendants. Ces derniers ont, en outre, la particularité d'être au cas nominatif même si le marqueur coréférentiel est à l'accusatif :

(2.19) a. *raʔaj -tu -**hu**$_i$ huwa$_i$* (marqueur d'objet)
voir.PER -1S -le lui(NOM)
« Je l'ai vu, lui »

b. *radʒaʕ -**a**$_i$ huwa$_i$* (marqueur de sujet)
revenir.PER -3MS lui(NOM)
« Il est revenu, lui »

Ce qui est commun aux marqueurs d'arguments dans ces contextes c'est que leur présence dans la structure rend la présence des SN et des pronoms indépendants superflue et les confine à la périphérie de la phrase, où ils ne peuvent constituer des arguments de la tête.

Une autre similitude distributionnelle s'observe dans la coordination. Les deux classes de marqueurs d'arguments ne peuvent être coordonnés les uns aux autres et il faut plutôt les répéter sur leurs hôtes, comme le montrent les exemples suivants :

(2.20) a. **raʔaj -tu -**hu** wa -**ha:*** (marqueur d'objet)
voir.PER -1S -le et -la

b. *raʔaj -tu -**hu** wa raʔaj -tu -**ha:***
voir.PER -1S -le et voir.PER -1S -la
« Je l'ai vu et je l'ai vue »

c. **ʤiʔ -**tu** wa -**ta*** (marqueur de sujet)
venir.PER -1S et -2MS

d. *ʤiʔ -**tu** wa ʤiʔ -**ta***
venir.PER -1S et venir.PER -2MS
« Je suis venu et toi également »

Finalement, les marqueurs de sujet, tout comme les marqueurs d'objet, ne peuvent être accentués, focalisés ou mis en valeur, comme le soulignent Bloch (1986) et Eloussfourri (1998). Ils sont toujours pragmatiquement neutres et seuls leurs équivalents indépendants peuvent être accentués :

(2.21) a. **raʔaj -tu -**HU*** (marqueur d'objet focalisé)
voir.PER -1S -LE

b. *ʔijja:**hu** raʔaj -tu* (Pro.Ind.ACC focalisé)
lui(ACC) voir.PER -1S
« C'est lui que j'ai vu »

c. **nim -**TU*** (marqueur de sujet focalisé)
dormir.PER -1S

d. *ʔan:**a** nim -tu* (Pro.Ind.NOM focalisé)
moi dormir.PER -1S
« C'est moi qui me suis endormi »

Les similitudes distributionnelles mentionnées dans cette section nous donnent des raisons de penser que les marqueurs de sujet et les marqueurs d'objet, même s'ils sont contrastés sur plus d'un plan, appartiennent à une certaine classe commune de formes, et c'est cette classe que nous appelons ici marqueurs d'arguments.

2.4.2 Les asymétries

Les marqueurs d'arguments sont hétérogènes[21] sur plus d'un plan : sur le plan phonologique, et en prenant en considération les différents hôtes possibles de ces formes prosodiquement déficientes, les marqueurs de sujet sélectionnent un seul type d'hôte, à savoir les verbes fléchis, alors que les marqueurs d'objet sélectionnent des hôtes de plusieurs catégories (section 2.4.2.1). Sur le plan morphologique, seuls les marqueurs de sujet sont analysables dans ce module de la grammaire. Cela s'explique par le fait qu'ils s'attachent à un radical verbal pour former un mot (le verbe fléchi), alors que les marqueurs d'objet s'attachent à un mot déjà formé (section 2.4.2.2). Sur le plan morphosyntaxique, seuls les marqueurs de sujet ont un comportement et des formes semblables à ceux de certains marqueurs morphosyntaxiques, à savoir les marqueurs d'accord nominaux (section 2.4.2.3). Et finalement, sur le plan syntaxique, le comportement des marqueurs de sujet et des marqueurs d'objet est dissemblable dans certaines structures, à savoir notamment dans la corroboration (section 2.4.2.4).

2.4.2.1 L'asymétrie des hôtes

Les marqueurs d'objet peuvent s'attacher à plusieurs catégories : les verbes (2.22a), les nominaux (2.22b), les prépositions (2.22c) et les complémenteurs (2.22d) :

(2.22) a. *katab* *-ta* ***-ha:***
écrire.PER -2MS -la
« Tu l'as écrite »

b. *kita:b* *-u* ***-ha:***
livre -NOM -son
« son livre »

c. *xif* *-tu* *min* ***-ha:***
craindre.PER -1S de -elle
« J'ai eu peur d'elle »

d. *ʔinna* ***-ha:*** *qari:bat* *-u* *-n*
que -elle proche -NOM -Indé
« Elle est proche, certes»

[21] Mullen (1986) a signalé la présence d'asymétries semblables en amharique, la deuxième langue sémitique la plus parlée au monde, après l'arabe. Cette langue est parlée en Éthiopie par une majorité de la population, mais également par des minorités dispersées dans plusieurs pays, dont l'Égypte et l'Érythrée.

Les marqueurs de sujet n'ont, quant à eux, qu'un seul hôte possible : le verbe fléchi perfectif (2.23a), imperfectif (2.23b) et impératif (2.23c). Ils ne peuvent s'attacher ni à des noms, ni à des prépositions, ni à des complémenteurs :

(2.23) a. *rakʒaʕ* *-u:*
revenir.PER -3MP
« Ils sont revenus »

b. *j-* *arkʒiʕ* *-u:* *-na*
3- revenir.IMP -MP -Ind
« Ils reviennent »

c. *ʔirkʒiʕ* *-u:*
revenir.IMPÉ -2MP
« Revenez ! »

À cet égard, les marqueurs de sujet se comportent comme les clitiques verbaux des langues romanes. Leur domaine syntaxique d'attachement est le verbe. Le domaine d'attachement des marqueurs d'objet n'est cependant pas limité aux verbes. Ils peuvent, à cet égard, être rapprochés du clitique du possessif anglais *'s*[22].

2.4.2.2 La morphologie des marqueurs d'arguments

Prenant en compte ce que nous avons avancé par rapport à la formation des mots en arabe, nous pouvons affirmer que seuls les marqueurs de sujet peuvent être considérés comme des atomes de la morphologie : ils font partie intégrante de la forme morphologique des verbes fléchis. En fait, les marqueurs de sujet s'attachent au radical verbal pour former le verbe fléchi, alors que les marqueurs d'objet sont attachés une fois que les premiers le sont et ne font donc pas partie du verbe en tant qu'unité morphologique. Cette *atomicité* morphologique des marqueurs de sujet se traduit dans le fait que le verbe sélectionne un SN, un pronom indépendant ou un marqueur d'objet seulement quand il comporte déjà un marqueur de sujet. Cette propriété est captée par le concept de l'incorporation de Fassi Fehri (1993)[23], mais également par l'hypothèse postulant que les marqueurs de sujet sont des affixes (faisant partie de la tête verbale).

Les exemples suivants mettent l'accent sur la formation morphologique des verbes avec des marqueurs d'arguments des deux sortes discutées ici. En (2.24a), seul un marqueur de sujet est

[22] Ce clitique s'attache à un SN en position finale comme enclitique, mais son hôte phonologique (le dernier élément du SN) peut être de plusieurs types catégoriels. Voir Zwicky (1987).
[23] L'incorporation dont il s'agit ici est un déplacement de tête à tête à la Baker (1988).

attaché au radical verbal *katab* et compose avec ce radical la forme morphologique du verbe fléchi. En (2.24b), un marqueur d'objet est attaché immédiatement à droite du marqueur de sujet. Nous pouvons avancer que cet attachement est de deux ordres : il est tout d'abord d'ordre phonologique et sert à pallier la déficience accentuelle du marqueur d'objet ; il est également d'ordre syntaxique et sert de ce fait à fournir le complément du verbe. Cet attachement n'a cependant aucun caractère morphologique. Ce qui le prouve c'est que l'attachement du seul marqueur d'objet produit une forme agrammaticale (2.24c), et que la violation de cet ordre (radical + marqueur de sujet + marqueur d'objet) est tout aussi agrammaticale (2.24d)[24]. En (2.24e), nous avons un exemple plus complexe où le radical verbal *aktub* est muni du marqueur de sujet discontinu *t-...-a:*, du morphème de l'indicatif *-ni* et d'un marqueur d'objet *-ha:*. Ici encore, le marqueur d'objet est attaché à l'extérieur des frontières morphologiques du verbe.

(2.24) a. *katab* *-tu*
écrire.PER -1S
« J'ai écrit »

b. *katab* *-tu* *-ha:*
écrire.PER -1S -la
« Je l'ai écrite »

c. **katab* *-ha:*
écrire.PER -la

d. **katab* *-ha:* *-tu*
écrire.PER -la -1S

e. *t-* *aktub* *-a:* *-ni* *-ha:*
2- écrire.IMP -MD -Ind -la
« Vous (deux) l'écrivez »

L'atomicité morphologique des marqueurs de sujet ainsi que leur attachement à une seule catégorie (les verbes) en font des candidats éligibles à l'étiquette d'affixes. Cette question sera discutée dans le quatrième chapitre.

Les marqueurs d'objet sont différents des marqueurs de sujet sur les deux plans phonologique et morphologique. Morphosyntaxiquement, une troisième différence vient approfondir cette hétérogénéité ; elle se résume dans les liens entre les différents marqueurs d'arguments et les marqueurs d'accord.

[24] Une comparaison avec les clitiques sujets du français s'impose ici. En fait, ces clitiques sont toujours attachés au verbe et ne peuvent en être séparés que par des clitiques objets. Ce comportement donne lieu à des phrases comme *je te le donne*, où le clitique sujet est attaché à l'extérieur des deux clitiques objets.

2.4.2.3 Marqueurs d'arguments et marqueurs d'accord

Les marqueurs de sujet, à l'opposé des marqueurs d'objet, ressemblent très visiblement aux marqueurs d'accord que l'on trouve sur les noms et les adjectifs. Cette ressemblance est mise en relief dans les exemples en (2.25). En (2.25a), le marqueur d'accord *-u:na* attaché au nom *l-muʕallim-u:na* (les enseignants) se décompose en deux morphèmes : *-u:* et *-na*. Le premier est un morphème portemanteau qui exprime à la fois trois traits : le cas (NOM), le genre (le masculin) et le nombre (le pluriel). Le deuxième est un indicateur de l'état libre du nom (le fait qu'il n'entre pas dans un état construit avec un autre nom). Un morphème composé semblable *-u:na* fait partie du verbe imperfectif *j-uɣa:dir-u:-na* « ils quittent ». Il se laisse décomposer en deux morphèmes : *-u:* et *-na*. Le premier de ces deux morphèmes exprime le genre (le masculin) et le nombre (le pluriel), alors que le deuxième est le morphème de l'indicatif. Le suffixe *-u:* exprime donc les mêmes traits morphosyntaxiques (de genre et de nombre) que son hôte soit un verbe ou un nom. Le fait qu'il désigne le cas nominatif du nom peut servir d'indicateur important si nous considérons que le marqueur de sujet à l'imperfectif est un pronom sujet (et donc au nominatif), hypothèse dont nous discuterons dans le troisième chapitre. En (2.25b), le même constat est valable en ce qui a trait au duel masculin marqué sur le verbe et sur le nom par le même suffixe *-a:*. Ce dernier est également le morphème qui exprime le nominatif sur le nom. Le suffixe *-ni* dans ces formes est comparable au suffixe *-na* en (2.25a).

(2.25) a. *ʔal- muʕallim**u:na** j- uɣa:dir **-u: -na** l- qa:ʕat -a*
les- enseignants 3- quitter.IMP -MP-Ind la- salle -ACC
« Les enseignants, ils quittent la salle »

b. *ʔal- muʕallim**a:ni** j- uɣa:dir **-a:-ni** l- qa:ʕat -a*
les- deux.enseignants 3- quitter.IMP -MD-Ind la- salle -ACC
« Les deux enseignants, ils quittent la salle »

La ressemblance est toutefois limitée à l'imperfectif indicatif. Cette conjugaison est en effet appelée *ʔalmuʃa:riʕ* (littéralement « la forme qui ressemble à celle des noms ») par les grammairiens pour mettre l'accent sur cette ressemblance. Notons également que les marqueurs de l'accord (qui s'attachent aux noms et aux adjectifs) *-u:na* et *-a:ni* sont propres au cas nominatif et qu'ils se réalisent *-i:na*[25] et *-ajni* respectivement aux autres cas. L'alternance est autre pour les marqueurs de sujet formés en partie de *-u:na* et de *-a:ni*, qui subissent une troncation de leur 2e syllabe au subjonctif et au jussif[26].

[25] La forme *-i:na* se retrouve dans la conjugaison imperfective des verbes à la 2FS.

[26] La troncation de cette même syllabe se produit quand le nom qui porte les marqueurs d'accord *-u:na*, *-i:na*, *-a:ni* et *-ajni* est en état construit, comme dans l'exemple suivant

Nous pouvons, par ailleurs, signaler d'autres ressemblances entre les marqueurs de sujet et les marqueurs d'accord en isolant les morphèmes qui encodent le genre, par exemple. Le féminin est généralement marqué par le morphème *-t* sur les noms et les adjectifs. Ce même morphème semble entrer dans la composition de plusieurs marqueurs de sujet de la 3e personne pour distinguer le féminin du masculin, comme illustré dans le tableau (2.14).

	Singulier	Duel
Perfectif 3M	katab-a	katab-a:
Perfectif 3F	katab-at	katab-ata:
Imperfectif 3M	j-aktub-u	j-aktub-a:-ni
Imperfectif 3F	t-aktub-u	t-aktub-a:-ni

TABLEAU 2.14: L'alternance M/F dans quelques marqueurs de sujet

Toutefois, au pluriel de la 3e personne et aux autres personnes, cette relation est moins transparente. Le paradigme contient en fait un trou morphologique : à la 1re personne, aucune distinction n'est faite entre le masculin et le féminin, alors qu'aux autres personnes les moyens adoptés sont variables. À la 2e personne, par exemple, la distinction des deux genres se fait par l'alternance vocalique *a/i* (entre *-ta* 2MS et *-ti* 2FS) ou le marquage du féminin par un morphème spécifique (*-i:* à l'imperfectif). À la 3e personne du pluriel, la stratégie supplétive est adoptée : *-u:* pour la 3MP et *-na* pour la 3FP, etc.

2.4.2.4 La corroboration par des réfléchis

Le comportement des marqueurs de sujet et des marqueurs d'objet dans les structures à apposition corroborative[27] fournit un appui important aux asymétries signalées dans les paragraphes précédents. Seuls les marqueurs d'objet peuvent en fait être corroborés par les réfléchis[28]. En (2.26a), le marqueur d'objet accusatif *-hu* est corroboré par le réfléchi *nafs-a-hu* « lui-même ». En (2.26b), le marqueur d'objet génitif *-hu* est corroboré par le réfléchi *nafs-i-hi*. Cette corroboration est impossible pour les marqueurs de sujet, d'où l'agrammaticalité de (2.26c), où le marqueur de sujet *-u:* du perfectif est corroboré par le réfléchi *ʔanfus-u-hum*. Ce comportement caractérise les marqueurs de sujet attachés à l'imperfectif et à l'impératif également :

(i) qa:riʔa:ni → qa:riʔa: r-risa:lat-i
lecteurs.NOM.D → lecteurs.NOM.D la-lettre-GÉN
« Les deux lecteurs de la lettre »

[27] La corroboration se fait par des réfléchis ou par des quantifieurs en arabe. Seule la corroboration par des réfléchis donne lieu à l'asymétrie observée dans ce paragraphe. La corroboration par un quantifieur, comme *kull* « tout/chaque », est possible pour les deux sortes de marqueurs d'arguments.

[28] Les réfléchis sont formés de certains noms, comme *nafs*, *ðaːt* et *ʕajn*, qui signifient « âme » et « personne », annexés à un marqueur d'objet.

(2.26) a. *qa:bal* *-tu* ***-hu*** *nafs -a -hu* (marqueur d'objet ACC)
rencontrer.PER -1S -le âme-ACC-son
« Je l'ai rencontré, lui-même »

b. *farar* *-tu* *min* ***-hu*** *nafs -i -hi* (marqueur d'objet GÉN)
fuir.PER -1S de -lui âme-GÉN-son
« Je l'ai fui, lui-même »

c. **ʤa:ʔ* *-u:* *ʔanfus -u -hum* (marqueur de sujet)
venir.PER -3MP âmes-NOM-leurs

Pour accomplir la corroboration dans la phrase (2.26c), un pronom indépendant NOM, coréférentiel avec le marqueur de sujet, doit obligatoirement être inséré, comme l'illustre l'exemple suivant :

(2.27) *ʤa:ʔ* *-u:* ***hum***$_i$ *ʔanfus -u -hum*$_i$
venir.PER -3MP eux âmes-NOM-leurs
« Ils sont venus, eux-mêmes »

Ces exemples montrent une hétérogénéité syntaxique entre les deux sortes de marqueurs d'arguments qui vient s'ajouter aux autres asymétries morphologiques et morphosyntaxiques déjà présentées.

De ce qui précède, les marqueurs d'arguments semblent constituer une classe homogène. Ils présentent en fait certaines propriétés qui les distinguent nettement des pronoms indépendants sur le plan morphophonologique, sur le plan (morpho)syntaxique et sur le plan sémantique. Sur ce dernier point, notons que les marqueurs d'arguments ne sont jamais associés à des valeurs pragmatiques marquées, alors que les pronoms indépendants NOM et ACC le sont dans la majorité de leurs emplois. Les grammairiens de la tradition affirmaient, par ailleurs, cette valeur ajoutée par l'emploi des pronoms indépendants en déclarant : « ne pas recourir aux pronoms indépendants si l'on peut employer les pronoms conjoints » [29]. La section suivante apporte plus d'éléments à cette comparaison.

[29] (Sibawayhi, 1977, auteur du XIIIe siècle), traduction libre.

2.5 Relations entre pronoms indépendants et marqueurs d'arguments

Les marqueurs d'arguments de l'arabe entretiennent plusieurs rapports avec les pronoms indépendants. Sur un plan purement morphophonologique, ces derniers sont majoritairement composés de marqueurs d'arguments agglutinés à d'autres éléments, ce qui témoigne d'un processus diachronique probable de dérivation. Sur le plan morphosyntaxique, les deux sortes de formes encodent des traits semblables, à savoir les traits de genre, de nombre et de personne, en plus de la pertinence de la discrimination sur la base du cas morphologique. Sur le plan syntaxique, finalement, si certains contextes admettent l'emploi de l'une ou l'autre de ces deux formes, il n'est généralement pas possible de les substituer les unes aux autres.

2.5.1 La morphophonologie des pronoms

Nous remarquons une relation morphophonologique incontestable entre plusieurs marqueurs d'arguments et les pronoms indépendants équivalents. Dans le tableau (2.15), nous mettons l'accent sur la composition des marqueurs d'objet et des pronoms indépendants ACC. Il s'avère que ces derniers sont invariablement formés selon le modèle suivant : le support *?ijja:-*[30] suivi d'un marqueur d'objet.

Pers.	Gre	Nbre	Marqueur d'objet	Pronom indépendant ACC
1	M/F	S	-(n)i: (-ja)	?ijja:-ja
		D/P	-na:	?ijja:-na:
2	M	S	-ka	?ijja:-ka
		D	-kuma:	?ijja:-kuma:
		P	-kum	?ijja:-kum
	F	S	-ki	?ijja:-ki
		D	-kuma:	?ijja:-kuma:
		P	-kunna	?ijja:-kunna
3	M	S	-hu	?ijja:-hu
		D	-huma:	?ijja:-huma:
		P	-hum	?ijja:-hum
	F	S	-ha:	?ijja:-ha:
		D	-huma:	?ijja:-huma:
		P	-hunna	?ijja:-hunna

TABLEAU 2.15: Les pronoms indépendants ACC et les marqueurs d'objet

Cette relation est également transparente dans le cas des pronoms indépendants NOM. Ces

[30] Ce support serait à l'origine un démonstratif, ce qui appuie l'hypothèse des historiens des langues sémitiques selon laquelle les pronoms sont issus de démonstratifs en arabe (Fleisch, 1979).

pronoms sont en fait formés des marqueurs de sujet du perfectif précédés du support *ʔan-*[31] pour la 2e personne et d'autres supports pour la 1re et la 3e personnes.

Pers.	Gre	Nbre	Marqueur de suj. PER	Pronom indépendant NOM
1	M/F	S	-tu	ʔana:
		D/P	-na:	naħnu
2	M	S	-ta	ʔan-ta
		D	-tuma:	ʔan-tuma:
		P	-tum	ʔan-tum
	F	S	-ti	ʔan-ti
		D	-tuma:	ʔan-tuma:
		P	-tunna	ʔan-tunna
3	M	S	-a	huwa
		D	-a:	huma:
		P	-u:	hum
	F	S	-at	hija
		D	-ata:	huma:
		P	-na	hunna

TABLEAU 2.16: Les pronoms indépendants NOM et les marqueurs de sujet

Bien que certains marqueurs de sujet (par exemple, à la 1S et à la 3S) ne semblent pas être directement reliés aux pronoms indépendants équivalents, nous ne pouvons nous empêcher de faire un rapprochement entre les deux formes sur la base de leur constitution morphophonologique. Cette ressemblance est plus frappante si nous prenons les formes protosémitiques qui sont censées être à l'origine des formes actuelles, telles que reconstituées par Gray (1934), par exemple, dans le tableau (2.17)[32].

Pers.	Gre	Nbre	MSu IMP	MSu PER	Pro. ind. NOM
1	M/F	S	ʔ-	-ku	ʔan-a:(ku)
	M/F	D/P	n-	-na:	naħ-nu
2	M	S	t-	-t-a	ʔan-t-a
	F	S	t-...-i:	-ti:	ʔan-t-i
	M	P	t-	-t-um-u	ʔan-t-um-u
	F	P	t-	-tinn-a	ʔan-t-inn-a

TABLEAU 2.17: Les formes protosémitiques (Gray, 1934)

Ces observations sont à l'origine de l'hypothèse suivante qu'on retrouve dans plusieurs ouvrages, dont Fleisch (1979), Gray (1934) et Moscati (1964) : le système des marqueurs de sujet de l'arabe a vraisemblablement résulté de l'incorporation des pronoms indépendants nominatifs au verbe. Fleisch (1979), par exemple, affirme que les préfixes de l'imperfectif et les suffixes du

[31] Un homophone de ce support *ʔan* est employé comme complémenteur.
[32] Cité dans (Benmamoun, 2000b, p. 60).

perfectif sont des pronoms réduits agglutinés au verbe.

À la lumière de cette relation morphophonologique avérée entre les marqueurs d'arguments et les pronoms indépendants, nous nous posons la question suivante : existe-t-il un pendant syntaxique à cette relation ? Nous consacrons la section suivante à cette question.

2.5.2 La syntaxe

2.5.2.1 Marqueurs d'objet et pronoms indépendants

Les marqueurs d'objets et les pronoms indépendants ACC n'occupent généralement pas les mêmes fonctions. Seuls les marqueurs d'objet peuvent être des compléments de nom, de préposition et de complémenteur et des compléments directs postposés au verbe, comme le montrent les contrastes suivants :

(2.28) a. *bajt* *-u* *-ha:*
maison -NOM -sa
« sa maison »

a'. **bajt* *-u* *ʔijja:ha:*
maison -NOM elle

b. *min* *-ha:*
de -elle
« d'elle »

b'. **min* *ʔijja:ha:*
de elle

c. *ʔanna* *-ha:*
que -elle
« qu'elle »

c'. **ʔanna* *ʔijja:ha:*
que elle

d. *laqi:* *-tu* *-ha:*
trouver.PER -1S -la
« Je l'ai trouvée »

d'. **laqi:* *-tu* *ʔijja:ha:*
trouver.PER -1S elle

Nous pourrions penser à une distribution complémentaire, mais une structure semble contredire cette conclusion. Il s'agit du deuxième complément des verbes dits *de don*. D'autres verbes sélectionnent deux compléments en arabe, mais le deuxième complément ne peut se réaliser sous la forme d'un marqueur d'objet. Il s'agit notamment des verbes dits de *croyance* comme *xa:la* « croire » dans les exemples suivants :

(2.29) a. *xil* *-tu* *r-* *raʒul* *-a* *lis's'* *-a* *-n*
croire.PER -1S l'- homme -ACC voleur -ACC -Indé
« J'ai confondu l'homme avec un voleur »

b. *xil* *-tu* *-hu* *lis's'* *-a* *-n*
croire.PER -1S -le voleur -ACC -Indé
« Je l'ai confondu avec un voleur »

b. **xil* *-tu* *-hu* *-hu*
croire.PER -1S -le -le

Quand le deuxième complément des verbes de don est pronominalisé, les deux structures avec un marqueur d'objet et avec un pronom indépendant sont grammaticales :

(2.30) a. *wahab* *-tu* *-ka* ***l- kita:b -a***
donner.PER -1S -te le-livre-ACC
« je t'ai donné le livre »

b. *wahab* *-tu* *-ka* ***-hu***
donner.PER -1S -te -le
« Je te l'ai donné »

c. *wahab* *-tu* *-ka* ***ʔijja:hu***
donner.PER -1S -te lui(ACC)
« Je te l'ai donné »

Les deux exemples en (2.30b) et en (2.30c), bien qu'ils expriment le même sens, représentent deux registres différents de la langue. Le premier est archaïsant alors que le deuxième est plus souvent rencontré dans le style contemporain. Attendu cette différence de registre et la différence entre les verbes de *don* et les verbes de *croyance*, nous pouvons maintenir l'hypothèse selon laquelle les marqueurs d'objet et les pronoms indépendants ACC sont en distribution complémentaire.

2.5.2.2 Marqueurs d'objet et SN

Si la distribution des marqueurs d'objet et des pronoms indépendants ACC est complémentaire, la comparaison des contextes dans lesquels sont insérés des SN (non nominatifs) et des

marqueurs d'objet donne lieu à une conclusion différente. En fait, et comme le montrent les exemples en (2.31), les marqueurs d'objet occupent les mêmes fonctions qu'occupent les SN, à condition qu'ils soient réalisés à droite de la tête. Quand le SN est à gauche de la tête, et ce n'est possible qu'avec une tête verbale, seul un pronom indépendant ACC peut remplacer le SN :

(2.31) a. *bajt* *-u* ***zajd -in***
maison -NOM Zayd-GÉN
« la maison de Zayd »

a'. *bajt* *-u* ***-hu***
maison -NOM -sa
« sa maison »

b. *ʕinda* ***zajd -in***
chez Zayd-GÉN
« chez Zayd »

b'. *ʕinda* ***-hu***
chez -lui
« chez lui »

c. *ʔanna* ***zajd -an***
que Zayd-ACC
« que Zayd »

c'. *ʔanna* ***-hu***
que -il
« qu'il »

d. *laqi:* *-tu* ***zajd -an***
trouver.PER -1S Zayd-ACC
« J'ai trouvé Zayd »

d'. *laqi:* *-tu* ***-hu***
trouver.PER -1S -le
« Je l'ai trouvé »

e. ***zajd -an*** *laqi:* *-tu*
Zayd-ACC trouver.PER -1S
« C'est Zayd que j'ai trouvé (et non Michel) »

e'. ***ʔijja:hu*** *laqi:* *-tu*
lui(ACC) trouver.PER -1S
« C'est lui que j'ai trouvé (et non elle) »

Nous pouvons donc conclure que, pour remplacer un SN (dont le cas est autre que le nominatif) par un pronom, le choix se fait entre les pronoms indépendants ACC et les marqueurs d'objet. Quand le SN est à droite de la tête, seul un marqueur d'objet peut le remplacer en préservant

cette position. À gauche de la tête, par contre, seul un pronom indépendant peut remplacer le SN.

Nous pouvons ainsi comprendre pourquoi le deuxième complément pronominal en (2.30c) est indépendant. N'étant pas adjacent à la tête verbale, il n'est pas forcé de s'y attacher. Cette indépendance est même obligatoire dans certains contextes où l'ordre des marqueurs d'objet constitue une violation de la contrainte thématique[33] ou de la contrainte de la personne[34], comme dans les exemples suivants :

(2.32) a. **wahab* *-tu* *-hu* *-ka*
donner.PER -1S -lui -te

b. *wahab* *-tu* *-hu* *ʔijja:ka*
donner.PER -1S -lui toi
« Je t'ai donné à lui »

Cette indépendance est également obligatoire avec les verbes de croyance, comme le montre le contraste entre les exemples suivants :

(2.33) a. *xil* *-tu* *-ka* *lis's'-a-n*
croire.PER -1S -te voleur-ACC-Indé
« J'ai cru que tu étais un voleur "

b. **xil* *-tu* *-ka* *-hu*
croire.PER -1S -te -le

c. *xil* *-tu* *-ka* *ʔijja:hu*
croire.PER -1S -te lui
« J'ai cru que tu l'étais »

Par ailleurs, nous pouvons également rendre compte de la relation morphophonologique entre les marqueurs d'objet et les pronoms indépendants ACC. Ainsi, nous pouvons postuler qu'en arabe, les pronoms objets sont toujours attachés à la tête qui constitue un hôte potentiel (c'est-à-dire une tête adjacente à gauche du pronom) quand un tel hôte est disponible ; autrement, ils apparaissent attachés au support morphologique *ʔijja:-*, et forment avec ce dernier une unité pronominale autonome (le pronom indépendant).

[33] Dans une suite de marqueurs d'objet, le bénéficiaire doit précéder le thème.
[34] Dans une suite de marqueurs d'objet, les personnes doivent être ordonnées comme suit : $1 < 2 < 3$.

Nous n'avons discuté dans cette section que de la syntaxe des marqueurs d'objet et surtout des relations distributionnelles entre ces pronoms, les pronoms indépendants et les syntagmes nominaux que ces deux sortes de pronoms sont censées remplacer. Aussi ces remarques ne concernent-elles en réalité que les marqueurs d'objet. Les marqueurs de sujet ont une tout autre histoire et nous ne pouvons envisager d'étudier leur syntaxe sans soulever plusieurs problèmes qui concernent leur statut morphosyntaxique.

2.6 Conclusion

Nous avons présenté les deux classes qui forment ce que nous pouvons considérer comme étant le système des pronoms personnels de l'arabe, à savoir les marqueurs d'arguments et les pronoms indépendants. De cette présentation, nous pouvons tirer plusieurs conclusions.

Ce système comporte deux sortes de pronoms personnels morphophonologiquement et morphosyntaxiquement reliés : les marqueurs d'arguments et les pronoms indépendants. L'étiquette de *pronoms* attribuée aux marqueurs d'arguments pourrait être motivée par leurs propriétés référentielles et par leur expression des traits de nombre, de genre et de personne. Elle pourrait également être contestée en ce qui concerne certains marqueurs de sujet, étant donné que quatre de ces derniers (à savoir les marqueurs de sujet de la 3e personne du singulier) peuvent coexister avec des SN postposés dans les contextes non marqués. Cette question est reliée à celle de la morphosyntaxe de ces marqueurs et sera traitée dans le troisième chapitre.

Cette typologie des pronoms personnels de l'arabe met l'accent sur les similarités et les différences morphophonologiques et syntaxiques. Elle nous permet non seulement de bien comprendre les rôles endossés par ces deux classes, mais également de bien situer les marqueurs d'arguments, notre sujet d'étude, dans un système plus global.

Les marqueurs d'arguments partagent certaines propriétés du fait de leur appartenance à cette sous-classe, mais ils ne sont pas homogènes et se divisent en deux sous-groupes : les marqueurs de sujet et les marqueurs d'objet. Plusieurs asymétries morphologiques et syntaxiques appuient cette typologie. Parmi ces asymétries, nous pensons que le comportement vis-à-vis des structures à corroboration par des réfléchis revêt une grande importance. Les marqueurs de sujet ne peuvent en fait être corroborés par des réfléchis que lorsqu'on insère un pronom indépendant NOM équivalent entre le marqueur de sujet et le syntagme qui le corrobore. Cette exigence ne concerne

cependant que les marqueurs de sujet ; les marqueurs d'objet peuvent être corroborés librement sans l'insertion d'un pronom indépendant.

Cette hétérogénéité qui caractérise les marqueurs d'arguments sera exploitée pour répondre à deux questions : quel est le statut morphosyntaxique de ces marqueurs ? Et quel est leur statut morphologique ? Autrement dit : doit-on considérer que ces marqueurs sont des arguments du verbe ou des marqueurs d'accord ? Et doit-on les traiter comme des affixes ou comme des non-affixes (clitiques ou mots indépendants) ? Ces questions seront soulevées dans les deux chapitres subséquents.

CHAPITRE III

LE STATUT MORPHOSYNTAXIQUE DES MARQUEURS D'ARGUMENTS

> There is a large body of lore concerning the notion [of subject], and any proposed definition must at least largely agree with the traditional, and to some extent, pretheoretical usage of the term. (Keenan, 1976, p. 306)

Introduction

Pour modéliser le comportement des marqueurs d'arguments dans le cadre de la théorie HPSG, il nous faut prendre certaines décisions concernant la relation entre ces éléments et leurs hôtes. Nous avons essentiellement besoin de savoir s'ils sont des arguments des têtes desquelles ils dépendent ou de simples dépendants phonologiques. Nous avons également besoin de savoir s'ils font partie de ces têtes-mots (en étant des atomes morphologiques invisibles à la syntaxe [1], par exemple des affixes) ou constituent des nœuds syntaxiques autonomes (des atomes syntaxiques, par exemple des mots ou des clitiques).

Dans le cadre de la grammaire HPSG, les informations concernant les arguments de la tête sont consignées dans deux attributs : STR-ARG, approprié aux objets de type *lexème* et VAL, approprié aux objets de type *mot*. Dans le premier attribut sont listés les arguments sélectionnés par la tête dans un ordre précis. Quand la tête est réalisée comme mot, ces arguments sont projetés *(mapped)* comme valeurs de l'attribut VAL, à condition qu'ils ne soient pas nuls ou affixes, comme le soutiennent Miller et Sag (1997).

Deux ordres de questions se posent donc concernant les marqueurs d'arguments de l'arabe : ces marqueurs sont-ils des arguments ou des non-arguments ? Et sont-ils des atomes de la morphologie (des affixes) ou des atomes de la syntaxe (des non-affixes) ? La première question est d'ordre morphosyntaxique alors que la deuxième est d'ordre morphologique.

[1] L'hypothèse en vigueur est lexicaliste au sens fort.

Ces deux questions établissent une distinction que nous empruntons à Auger (1994) dans son étude des clitiques du français parlé informel du Québec. En effet, l'auteure distingue deux dimensions dans l'étude de ces clitiques : une dimension morphosyntaxique et une dimension morphologique. La distinction morphologique se fait entre des éléments qui dépendent morphologiquement d'autres éléments (typiquement, des affixes) et des éléments qui jouissent d'une indépendance morphologique (des clitiques ou des mots indépendants). La distinction morphosyntaxique se fait, quant à elle, entre des arguments et des non-arguments. En ce qui a trait aux clitiques étudiés par Julie Auger, cette dernière étiquette désigne surtout des marqueurs d'accord.

L'auteure affirme que les affixes ne sont pas forcément des marqueurs d'accord et que les mots indépendants ne sont pas forcément des arguments. Elle donne comme exemple les suffixes verbaux des verbes synthétiques dans les langues celtes. Ces suffixes ne sont jamais employés avec un sujet lexical, laissant ainsi croire qu'ils sont des arguments de ces verbes : « We thus have a case of an affixal element which functions like a syntactic argument. » (Auger, 1994, p. 27). Par ailleurs, certains éléments morphologiquement indépendants peuvent fonctionner comme des marqueurs d'accord. Les exemples donnés par l'auteure proviennent surtout des constructions à redoublement du clitique en polonais, en pirahã et en nganhcara. Ces divergences entre le statut morphologique et le statut morphosyntaxique sont représentées dans le tableau (3.1) [2].

Ce tableau montre également que certains affixes fonctionnent comme des marqueurs d'accord et que certains non-affixes fonctionnent comme des arguments, ce qui constitue la corrélation attendue.

Séparer la dimension morphologique de la dimension morphosyntaxique est important pour plusieurs raisons. Tout d'abord, en ce qui concerne les marqueurs d'objet, nous avons des raisons de nous interroger sur leur statut morphosyntaxique et leur statut morphologique. Devrions-nous toujours les traiter comme arguments, ce qui va à l'encontre de certaines données empiriques comme nous le verrons dans ce chapitre ? Et devrions-nous les considérer comme des affixes ou comme des clitiques, les deux termes ayant déjà servi dans la littérature pour les désigner ?

En ce qui a trait aux marqueurs de sujet, séparer ces deux dimensions est une nécessité. Certaines étiquettes employées pour les classer cachent, en fait, un amalgame entre la morphologie et la morphosyntaxe. Celle de *pronoms conjoints*, par exemple, employée surtout par les grammairiens, laisse croire qu'ils sont des arguments d'un certain type, à savoir des arguments

[2] (Auger, 1994, p. 48).

	Marqueurs d'accord	Arguments
Affixes	· Marqueurs de sujet en français parlé informel (FPI) · Suffixes d'accord avec le sujet et marqueurs d'objets indirects (OI) en espagnol · Marqueurs de sujet dans les langues bantoues	· Marqueurs d'objet en FPI · Pronoms clitiques en français standard · Marqueurs d'objet direct (OD) en espagnol · Marqueurs d'objet en chichewa · Marqueurs de sujet dans les langues celtiques · Suffixes possessifs en finnois
Non-affixes (clitiques ou mots indépendants)	· Clitiques OI en nganhcara · Marqueurs de 2e et 3e personnes en pirahã · Clitiques du 2e type en yagua · Marqueurs de sujet en polonais archaïque	· Pronoms en anglais · Pronoms en allemand

TABLEAU 3.1: La divergence entre la morphologie et la morphosyntaxe

conjoints, subordonnés à d'autres éléments, sans toutefois préciser la nature de cette subordination. L'étiquette de *marqueurs d'accord*, employée par certains linguistes laisse entendre une subordination morphologique, alors que la corrélation entre le rôle morphosyntaxique et la nature morphologique n'est pas biunivoque, comme le démontre Julie Auger dans sa thèse de doctorat.

Ce chapitre est consacré à la question de la morphosyntaxe des marqueurs d'arguments de l'arabe[3]. Nous démontrons que ces derniers ne sont pas morphosyntaxiquement homogènes. La majorité des marqueurs de sujet sont des arguments, mais quatre d'entre eux appartiennent à deux types d'objets différents : des arguments pronominaux et des marqueurs d'accord. Les marqueurs d'objet assument également des rôles morphosyntaxiques différents. Ils peuvent ainsi fonctionner comme arguments, comme explétifs ou comme topiques.

Ce chapitre est organisé comme suit : nous commençons par introduire certaines notions clés dans l'étude de la morphosyntaxe des marqueurs d'arguments. Nous abordons, par la suite, la question du statut des marqueurs d'objet. Suit une section dans laquelle nous abordons la question

[3] La question du statut morphologique sera abordée dans le quatrième chapitre.

tant débattue du statut des marqueurs de sujet. Nous testons la validité de l'hypothèse que nous y proposons dans une section consacrée aux implications de cette approche.

3.1 Définitions

Nous introduisons dans cette section quatre notions clés qui nous permettront d'étudier la morphosyntaxe des marqueurs d'arguments : arguments, marqueurs d'accord, redoublement, dislocation et topicalisation.

3.1.1 Arguments et marqueurs d'accord

Nous définissons le marqueur d'accord comme étant un élément dépourvu de fonction syntaxique propre. Cet élément réfère à un autre, avec lequel il partage certains traits [4] grammaticaux dans une configuration donnée (Auger, 1994, p. 27). Le marqueur d'accord fait donc typiquement partie d'un mot (ou d'un syntagme) qui, lui, occupe une fonction syntaxique. C'est pour cette raison que certains auteurs, comme Kaiser (1994) [5], considèrent que les marqueurs d'accord doivent être des affixes, puisqu'ils sont nécessairement attachés à un autre élément. D'autres, au contraire, Auger (1994) et Stump (1980) par exemple, insistent sur la nécessité de séparer les deux dimensions puisque certains marqueurs d'accord ne sont pas des affixes et que certains affixes fonctionnent comme des arguments.

Nous définissons l'argument syntaxique ici comme étant un objet linguistique sélectionné par une tête et qui est donc le complément de cette tête, en notant que le terme « complément » a une acception large dans la théorie HPSG et désigne tout dépendant de la tête (complément au sens classique, sujet, spécifieur, etc.).

3.1.2 Redoublement, dislocation et topicalisation

La question de la morphosyntaxe des marqueurs d'arguments est reliée à celle de la dislocation, du redoublement et de la dislocation des SN et des pronoms indépendants qui constituent les

[4] Ces traits varient d'une langue à l'autre et peuvent varier d'une variété d'une langue à une autre variété. Le cas, par exemple, constitue un trait d'accord nominal en arabe standard, mais il ne fait pas partie des traits d'accord nominal dans les dialectes arabes contemporains. Nous constaterons que l'accord entre le sujet et le verbe en arabe standard exclut le nombre, alors que ce trait est activé dans les dialectes.

[5] C'est également l'approche qui sous-tend l'analyse de Shlonsky (1997) dans son étude des marqueurs de sujet.

antécédents de ces marqueurs. En effet, le fait d'affirmer qu'un marqueur d'argument est un argument ou un marqueur d'accord présuppose une analyse du rôle syntaxique joué par le SN ou le pronom indépendant antécédents si ces derniers sont également présents dans la structure. Et puisque c'est souvent le cas en arabe, nous entreprenons une telle analyse chaque fois que nous serons amené à étudier la morphosyntaxe des marqueurs d'arguments.

Le redoublement est entendu ici comme étant la coexistence du marqueur d'argument et d'un constituant (SN ou pronom indépendant) associé (= co-indicé / coréférentiel) à ce marqueur, à condition que ce constituant soit à droite du marqueur d'argument et qu'il n'y ait pas de coupure intonatoire qui placerait le constituant en question à la périphérie de la phrase. Cette coupure est souvent notée par une virgule à l'écrit et marquée par une pause à l'oral, comme le notent Philippaki-Warburton *et al.* (2002). Un exemple de redoublement est donné en (3.1) de l'arabe tunisien, où le SN *l-wla:d* « les garçons » redouble le marqueur d'objet *-hum* « eux » :

(3.1) a. *wi:n* ***-hum*** *l-wla:ad*
où -eux les-garçons
« Où sont les garçons ? »

b. *wi:n* ***-hum***
où -eux
« Où sont-ils ? »

c. *??wi:n* *l-wla:ad*
où les-garçons
« Où sont les garçons ? »

Dans cet exemple, le redoublement se fait sans recourir à l'insertion d'une préposition dite explétive devant le SN redoublé. Guerssel (1995) signale un comportement semblable en berbère tamazight. Toutefois, dans d'autres langues, le redoublement n'est possible que lorsqu'une préposition est insérée devant le SN. C'est le cas de l'espagnol et des dialectes arabes du Levant. En arabe libanais, comme mentionné par Aoun (1996), et en arabe palestinien, comme mentionné par Shlonsky (1997), par exemple, le SN qui redouble un marqueur d'objet est obligatoirement précédé d'une préposition. L'exemple en (3.2) est de l'arabe palestinien. Dans cet exemple, le marqueur d'objet est *-ha:* « la » et le SN qui le redouble est *l-mʕalme* « l'enseignante ». Ce SN est obligatoirement précédé de la préposition *la*[6], comme le montre le contraste entre les exemples suivants :

[6] (Shlonsky, 1997, p. 194).

(3.2) a. *fhim* *-t* *-ha:ᵢ* *la-* *l-* *mʕalmeᵢ*
comprendre.PER -1S -la à- l'- enseignante
« J'ai compris l'enseignante »

b. **fhim* *-t* *-ha:ᵢ* *l-* *mʕalmeᵢ*
comprendre.PER -1S -la l'- enseignante

En (3.1) et en (3.2), les structures présentées sont conformes à la définition du redoublement adoptée ici. Le SN associé au marqueur d'objet apparaît à droite de ce dernier et il n'est accompagné d'aucune coupure intonatoire perceptible à l'écrit ou à l'oral.

La dislocation s'apparente au redoublement, mais s'en distingue par une coupure intonatoire qui détache le SN du reste de la phrase, ce dernier pouvant alors se retrouver à la périphérie gauche ou à la périphérie droite (Philippaki-Warburton *et al.*, 2002, p. 58). Cette structure est associée à des effets « pragmatiques », comme le contraste, l'emphase ou le changement de topique, comme le démontre Auger (1994). La dislocation est connue en français standard dans les phrases suivantes, par exemple :

(3.3) a. Je **lui** ai raconté mes secrets, le traître.
b. Je **lui** ai raconté mes secrets, au traître.
c. Martine **le** lit, le livre.

Finalement, la topicalisation consiste à antéposer un argument sans qu'il soit associé à un marqueur d'argument. Dans l'exemple suivant, le SN objet *ʔar-risa:lat-a* « la lettre » est topicalisé. Ce SN est en fait antéposé au verbe tout en gardant sa marque de l'accusatif et sans être co-indicé avec un marqueur d'argument :

(3.4) *ʔar-* *risa:lat* *-a* *qaraʔ* *-tu*
la- lettre -ACC lire.PER -1S
? « La lettre, j'ai lu »

Prenant ces définitions en considération, nous traitons la question du statut morphosyntaxique des marqueurs d'arguments dans deux sections. La première est consacrée aux marqueurs d'objet. Nous montrons qu'ils peuvent être des arguments d'une tête lexicale, des topiques ou des explétifs attachés à une tête fonctionnelle. La deuxième section sera consacrée aux marqueurs de sujet. Le débat concernant l'argumentalité de ces derniers touche plusieurs problèmes syntaxiques, tels que la légitimation des pronoms vides, la syntaxe des SN pré- et postverbaux, mais également

les structures de redoublement et de dislocation. Nous passons en revue l'approche des auteurs qui considèrent que tous les marqueurs de sujet sont des marqueurs d'accord. Nous présentons également l'approche de ceux qui soutiennent que ce sont plutôt des arguments, du moins dans certains contextes. Nous présentons finalement les arguments qui nous incitent à soutenir une approche selon laquelle, si la majorité des marqueurs de sujet sont des arguments pronominaux, quatre d'entre eux sont des homophones des marqueurs d'accord verbaux. Cette approche a plusieurs répercussions dont nous discutons dans la section (3.4).

3.2 Les marqueurs d'objet

Comme nous l'avons montré dans le deuxième chapitre, les marqueurs d'objet peuvent occuper plusieurs fonctions, qui sont toutes des fonctions de compléments : d'un verbe (3.5a), d'un nominal (3.5b), d'une préposition (3.5c) ou d'un complémenteur (3.5d,e) :

(3.5) a. *katab* *-tu* ***-ha:***
écrire.PER -1S -la
« Je l'ai écrite »

b. *bajt* *-u* ***-ha:***
maison -NOM -sa
« sa maison »

c. *xif* *-tu* *min* ***-ha:***
craindre.PER -1S de -elle
« J'ai eu peur d'elle »

d. *ʔinna* ***-hu*** *ʤa:ʔ* *-a* *-na:* *raʤul* *-u* *-n*
que -il venir.PER -3M -nous homme -NOM -Indé
« Le fait est qu'un homme est venu vers nous »

e. *ʔinna* ***-ka*** *t-* *uri:d* *-u* *l-* *ʕira:k* *-a*
que -toi 2MS- vouloir.IMP -Ind la- zizanie -ACC
« Toi, tu cherches la zizanie, certes »

Toutefois, en fonction de la nature de l'hôte, ces marqueurs exhibent certaines différences distributionnelles qui nous incitent à les grouper en trois sous-classes : les compléments du verbe, les compléments de nominaux ou de prépositions et les compléments de têtes fonctionnelles. Nous consacrons les sections suivantes à la morphosyntaxe de ces trois sous-classes en mettant l'accent sur les structures à redoublement et celles où il y a dislocation.

3.2.1 Les compléments du verbe

3.2.1.1 Le redoublement

Les marqueurs d'objet attachés au verbe ne peuvent être redoublés par des SN ou des pronoms indépendants en arabe. Pour que ce redoublement soit possible, il faudrait que les conditions citées dans la définition du redoublement soient réunies, à savoir la coréférence, la position du double à droite du marqueur d'objet et l'absence de toute coupure intonatoire réalisable à l'oral ou à l'écrit. Ces conditions ne peuvent être réunies dans le cas des compléments du verbe. Ici les traits du SN ou du pronom indépendant en question, à savoir le fait que ces composants soient animés[7] ou non, ne changent rien à l'impossibilité du redoublement, comme le montrent les exemples suivants :

(3.6) a. **katab* *-tu* ***-ha:*** *r-risa:lat-a* / *ʔijja:aha:*
écrire.PER -1S -la la-lettre-ACC / elle(ACC)

b. **raʔaj* *-tu* ***-ha:*** *raba:b -a* / *ʔijja:aha:*
voir.PER -1S -la Rabab-ACC / elle(ACC)

De même, la définitude et la spécificité n'ont aucun impact :

(3.7) a. **katab* *-tu* ***-ha:*** *risa:lat*$_i$ *-a* *-n*
écrire.PER -1S -la lettre -ACC -Indé
(Pour signifier : « J'ai écrit une lettre »)

b. **katab* *-tu* ***-ha:***$_i$ *risa:lat*$_i$ *-a* *ʃukr* *-i* *-n*
écrire.PER 1S -la lettre -ACC remerciement -GÉN -Indé
(Pour signifier : « J'ai écrit une lettre de remerciement »)

Cette distribution des marqueurs d'objet compléments de verbes nous permet d'affirmer que ces pronoms sont des arguments et non des marqueurs d'accord puisqu'ils ne peuvent être redoublés. Ils peuvent toutefois être associés à des SN ou à des pronoms indépendants disloqués.

[7] Ce trait a, par contre, un effet sur la possibilité du redoublement dans certaines langues. En espagnol, par exemple, seuls les pronoms objets directs animés sont obligatoirement redoublés (Suner, 1992).

3.2.1.2 La dislocation

Les marqueurs d'objet compléments du verbe peuvent apparaître dans des structures où des SN ou des pronoms indépendants sont disloqués et co-indicés avec ces marqueurs. Ces structures sont [8] pragmatiquement marquées et expriment l'emphase ou le contraste.

L'élément disloqué (le SN ou le pronom indépendant) apparaît au nominatif, comme dans les exemples en (3.8) :

(3.8) a. *ʔar* *-risa:lat*$_i$ *-u* *katab* *-tu* ***-ha:***$_i$
la- lettre -NOM écrire.PER -1S -la
« La lettre, je l'ai écrite »

b. *hija*$_i$ *katab* *-tu* ***-ha:***$_i$
elle(NOM) écrire.PER -1S -la
« Elle, je l'ai écrite »

c. *katab* *-tu* ***-ha:***$_i$ *r-* *risa:lat*$_i$ *-u*
écrire.PER -1S -la la- lettre -NOM
« Je l'ai écrite, la lettre »

d. *katab* *-tu* ***-ha:***$_i$ *hija*$_i$
écrire.PER -1S -la elle(NOM)
« Je l'ai écrite, elle»

Le cas nominatif de l'élément disloqué constitue un critère déterminant pour ne pas le considérer comme le complément du verbe. Dans les deux structures, le marqueur d'objet est un argument de la tête verbale, alors que le SN et le pronom indépendant ne le sont pas. Nous verrons dans la section suivante que ces remarques sont valables pour les marqueurs d'objet compléments d'autres têtes lexicales malgré certaines différences de distribution.

3.2.2 Les compléments de nominaux et de prépositions

Les marqueurs d'objet compléments de nominaux et de prépositions ont un comportement semblable à celui des marqueurs d'objet compléments de verbes en ce qui a trait à l'impossibilité du redoublement et sous les mêmes conditions. Cependant, certaines différences émergent dans les structures où il y a dislocation de l'antécédent.

[8] Pour une étude détaillée de ces structures en arabe et dans d'autres langues, voir (Aoun et Benmamoun, 1998), ainsi que Alexopoulou *et al.* (2004) et Adger *et al.* (2004).

3.2.2.1 Le redoublement

Les marqueurs d'objet compléments de nominaux et de prépositions ne peuvent être redoublés par des SN, comme en témoignent les exemples suivants :

(3.9) a. **ha:ðaː bajt -u -**ha:**$_i$ raba:b -a$_i$*
celle-ci maison -NOM -sa Rabab-GÉN

b. *haːðaː bajt -u -**haː***
celle-ci maison -NOM -sa
« Celle-ci est sa maison »

c. *haːðaː bajt -u raba:b -a*
celle-ci maison Rabab -GÉN
« Celle-ci est la maison de Rabab »

Cette impossibilité du redoublement démontre que ces marqueurs d'objet sont des arguments de la tête qui constitue leur hôte.

3.2.2.2 La dislocation

Les possibilités de dislocation des SN ou des pronoms indépendants associés à ce type de marqueurs d'objet sont limitées. La dislocation à gauche n'est possible que lorsque le composant disloqué est au nominatif, auquel cas ce composant est un topique au sens que confèrent Li et Thompson (1976) et Lambrecht (1994) à ce terme :

(3.10) a. *raba:b -u$_i$ ʔaxawa:t -u -**ha:**$_i$ ʕaʃr -u -n*
Rabab-NOM sœurs -NOM -ses dix -NOM -Indé
« Rabab, ses sœurs sont au nombre de dix »

b. *hija$_i$ ʔaxawa:t -u -**ha:**$_i$ ʕaʃr -u -n*
elle sœurs -NOM -ses dix -NOM -Indé
« Elle, ses sœurs sont au nombre de dix »

À droite, l'élément disloqué ne peut être qu'un pronom indépendant nominatif et l'emploi d'un SN quel que soit son cas morphologique est proscrit :

(3.11) a. *ʔaxawa:t -u -**ha:**$_i$ hija$_i$ ʕaʃr -u -n*
sœurs -NOM -sa elle(NOM) dix -NOM -Indé
« Ses sœurs à elle sont au nombre de dix »

b. *ʔaxawa:t -u -ha:$_i$ rabaːb -a$_i$ ʕaʃr -u -n
sœurs -NOM -ses Rabab-GÉN dix -NOM -Indé

Les marqueurs d'objet en (3.10) et en (3.11a) demeurent des arguments de la tête à laquelle ils s'attachent. Les SN et les pronoms indépendants co-indicés avec ces marqueurs d'objet sont disloqués, d'où le cas nominatif et les valeurs pragmatiques marquées qu'ils expriment. En (3.10), la dislocation peut exprimer un changement de topique ou un contraste selon le contexte et en (3.11a) la dislocation exprime l'emphase.

3.2.3 Les compléments de têtes fonctionnelles

L'attachement des marqueurs d'objet à un complémenteur se fait dans trois sortes de structures. Ils peuvent en fait s'attacher à un complémenteur qui introduit une phrase équative, comme en (3.12a), à un complémenteur qui introduit un topique, comme en (3.12b), ou à un complémenteur qui introduit un explétif, comme en (3.12c).

(3.12) a. ʔinna -hu s'adiiq -i:
que -lui ami -mon
« Il est mon ami, certes »

b. ʔinna -hu j- uħibb -u l- muzaʔħ -a
que -lui 3MS- aimer.IMP -Ind la- plaisanterie -ACC
« Lui, il aime plaisanter, certes »

c. ʔinna -ha: ʕa:d -at l- fata:ta:ni
que -elle revenir.PER -3F les- deux.filles.NOM
« Le fait est que les deux filles sont revenues »

Les marqueurs d'objet dans ces exemples sont à l'accusatif, un cas qui leur est assigné par le complémenteur. Ils ne sont toutefois pas des arguments du verbe. Dans la phrase en (3.12a), aucun verbe ou particule verbale ne sont réalisés. Le marqueur d'objet est le sujet[9] de la phrase équative et il remplace un SN accusatif qui pourrait être lexicalisé, comme dans l'exemple suivant :

(3.13) ʔinna zajd -an s'adi:q -i:
que Zayd -ACC ami -mon
« Zayd est mon ami, certes »

[9] Le problème qui se pose est évidemment celui du conflit apparent entre le cas morphologique (l'accusatif) et la fonction du marqueur d'objet (le sujet de la phrase équative). Ce problème peut être résolu si nous éliminons toute correspondance biunivoque entre la fonction et le cas morphologique dans les phrases équatives. Cette correspondance est d'autant superflue dans le cas de ces phrases que les deux éléments qui les composent (c'est-à-dire le sujet et le prédicat) peuvent tour à tour être au nominatif ou à l'accusatif.

Dans la phrase (3.12b), malgré la présence d'un verbe, le marqueur d'objet n'en est pas l'argument [10]. Il est plutôt un topique disloqué et il remplace un SN accusatif que nous pouvons lexicaliser, comme dans l'exemple suivant :

(3.14) *ʔinna zajd -an j- uħibb -u l- muzaːħ -a*
que Zayd -ACC 3MS- aimer.IMP -Ind la- plaisanterie -ACC
« Zayd, il aime plaisanter, certes »

En (3.12c), le marqueur d'objet est un explétif. Il ne peut s'agir d'un topique disloqué puisqu'il n'est pas lié à un pronom résomptif équivalent en termes de traits. En outre, cet explétif demeure à la 3S et s'accorde en genre avec un constituant principal (comme le sujet) de la phrase ainsi introduite, comme le montrent les exemples suivants :

(3.15) a. *ʔinna* ***-hu*** *ʕaːd -a l- fitjat -u*
que -il revenir.PER -3M les- jeunes.garçons -NOM
« Le fait est que les jeunes garçons sont revenus »

b. *ʔinna* ***-haː*** *ʕaːd -at n- nisaːʔ -u*
que -elle revenir.PER -3F les- femmes -NOM
« Le fait est que les femmes sont revenues »

Les marqueurs d'objet attachés aux complémenteurs représentent les seuls cas où ces marqueurs ne sont pas des arguments, notamment quand ils remplacent des SN topiques disloqués et quand ils sont des explétifs. Dans la phrase équative, leur statut est toutefois comparable à celui des pronoms indépendants nominatifs en ceci qu'ils représentent un élément essentiel de la phrase.

D'un autre côté, les marqueurs d'objet attachés aux complémenteurs ne sont pas différents des marqueurs d'objet attachés aux autres catégories en ceci qu'ils écartent toute possibilité de redoublement tout en ayant la possibilité d'avoir pour antécédents des SN ou des pronoms indépendants disloqués :

(3.16) a. *zajd -un ʔinna* ***-hu*** *sʿadiːq -iː*
Zayd-NOM que -lui ami -mon
« Zayd, il est mon ami, certes »

[10] Les auteurs qui considèrent le marqueur d'objet comme étant le sujet du verbe *juħibbu* se heurtent au problème de justifier le cas accusatif assigné à ce marqueur. Le même problème se pose avec le SN accusatif *zajdan* dans la phrase en (3.14). Voir Letourneau (1993).

b. *ʔax-i:* *ʔinna* *-hu* *j-* *uħibb* *-u* *l-* *muza:ħ* *-a*
frère-mon que -lui 3MS- aimer.IMP -Ind la- plaisanterie -ACC
« Mon frère, lui, il aime plaisanter, certes »

Nous avons présenté plusieurs structures dans lesquelles peuvent être réalisés des marqueurs d'objet. Parmi ces structures, nous avons mis l'accent sur le redoublement et la dislocation. Nous avons montré que le redoublement de ces marqueurs est impossible en arabe standard et que, lorsqu'ils apparaissent coréférentiels avec des SN ou des pronoms indépendants, ces derniers sont disloqués (à gauche ou à droite). Nous pouvons ainsi affirmer que les marqueurs d'objet ne peuvent fonctionner comme des marqueurs d'accord [11]. Ils peuvent cependant fonctionner comme des arguments, comme des explétifs ou comme des topiques disloqués. Cette richesse des possibilités provient de leur capacité à remplacer des SN dans un grand nombre de contextes et de la diversité de leurs hôtes possibles.

Même si leur distribution est en réalité plus restreinte, le statut morphosyntaxique des marqueurs de sujet représente un plus grand défi. En effet, ils n'apparaissent attachés qu'au verbe, d'où les deux approches qui les considèrent comme des arguments incorporés au verbe et comme des marqueurs d'accord avec les arguments de ce dernier.

3.3 Les marqueurs de sujet

Rappelons, en quelques lignes, la distribution des marqueurs de sujet. Ces marqueurs peuvent être *redoublés* par des SN (3.17a) et peuvent également être associés à des éléments disloqués (3.17b) :

(3.17) a. *katab* *-a* *l-* *walad* *-u* *risa:lat* *-a* *-n* (redoublement)
écrire.PER -3M le- garçon -NOM lettre -ACC -Indé
« Le garçon a écrit une lettre »

b. *katab* *-a* *huwa* *risa:lat* *-a* *-n* (dislocation)
écrire.PER -3MS lui lettre -ACC -Indé

[11] Selon Shlonsky (1997), les marqueurs d'objet sont des têtes AGR° (une forme d'accord). Toutefois, cette hypothèse n'est pas corroborée par une étude exhaustive des structures dans lesquelles apparaissent ces marqueurs. L'auteur consacre son effort à l'analyse syntaxique qui en découle et affirme : « Semitic clitics are Agr° elements to which a lower head adjoins as it raises out of the complement of Agr° » (*Ibid.*, p. 187)

« Il a écrit une lettre, lui »

Leur statut morphosyntaxique dans un cas comme dans l'autre pose problème, d'où les deux approches suivantes. La première, que nous appelons ici *l'approche argumentale*, est adoptée par les grammairiens arabes, comme Astarabaadii (1978), Hassan (1987) et Sibawayhi (1977), et par Akkal (1996), Bulk (2004), Fassi Fehri (1993), Jamari (1993) et Lumsden et Haleform (2003). Malgré certaines différences quant au nombre et à la nature des marqueurs de sujet reconnus comme tels par ces auteurs, ils ont en commun cette hypothèse qui considère qu'au moins certains marqueurs de sujet sont des arguments du verbe.

La deuxième approche, que nous appelons ici *l'approche flexionnelle*, est adoptée dans les travaux de Shlonsky (1997) et de Eloussfourri (1998) sur la nature flexionnelle des marqueurs de sujet et de Benmamoun (1995), Eid (1993), Jelinek (2002) et Kenstowicz (1989) sur la légitimation du pronom nul *pro* en arabe. Elle constitue également l'une des prémisses des analyses consacrées aux asymétries de l'accord en arabe, comme Ackema et Neeleman (2003), Aoun *et al.* (1994, 1999), Bahloul et Harbert (1992) et Harbert et Bahloul (2002), et se retrouve également dans les ouvrages de grammaire arabe écrits par des orientalistes, tels que Fleisch (1979). Ces divers auteurs sont d'avis que les marqueurs de sujet de l'arabe sont des marqueurs d'accord avec un sujet lexical ou nul, et ce, dans tous les contextes.

Nous présentons ces deux approches dans les sections subséquentes. Nous commençons par l'approche flexionnelle, que nous présentons et dont nous discutons dans la section (3.3.1) et nous discuterons des arguments avancés pour soutenir cette hypothèse. Nous présentons, par la suite, l'approche argumentale dans la section (3.3.2) et nous y discutons des fondements de cette hypothèse. Finalement, nous présentons nos hypothèses sur la question dans la section (3.3.3).

3.3.1 L'hypothèse flexionnelle

Nous ne donnons ici qu'un seul exemple que nous trouvons représentatif de l'approche flexionnelle, à savoir celle de Shlonsky (1997) [12]. Cet auteur stipule que l'arabe, et les langues sémitiques en général, ne connaissent pas le phénomène typiquement roman des clitiques sujets puisque, selon lui, les marqueurs de sujet ne sont pas attachés au verbe dans la composante syntaxique, comme les clitiques postlexicaux, mais dans la composante lexicale avant toute projection syn-

[12] Implicitement, l'auteur est d'avis que les éléments introduits lexicalement (les affixes) ne peuvent être des arguments et croit donc à une relation biunivoque entre le statut morphologique et le statut morphosyntaxique.

taxique. Pour appuyer cette approche, l'auteur avance plusieurs arguments. Il discute de certains phénomènes en hébreu et en berbère[13] et généralise cette discussion pour inclure les phénomènes observés en arabe.

En hébreu, l'auteur souligne que la particule de négation *ʔejn* peut être suivie d'un pronom indépendant sujet ou d'un marqueur de sujet, comme dans les exemples suivants[14] :

(3.18) a. *ʔejn* ***ʔani*** *ʔohev* *xacilim*
NÉG moi aimer(BENONI).MS aubergines
« Je n'aime pas les aubergines »

b. *ʔejn* *(ə**n**)**i*** *ʔohev* *xacilim*
NÉG 1S aimer(BENONI).MS aubergines
« Je n'aime pas les aubergines »

En se basant sur ces exemples, il soutient que, si l'on considère que *(ə**n**)**i*** en (3.18b) est un pronom sujet, on ne pourrait expliquer pourquoi, à la 3e personne, la phrase en (3.19b) est agrammaticale, alors que (3.19c) est tout à fait grammaticale, et ce, grâce à la lexicalisation du sujet :

(3.19) a. *ʔejn* ***hu*** *ʔohev* *xacilim* (pronom indépendant)
NÉG lui aimer(BENONI).MS aubergines
« Il n'aime pas les aubergines »

b. **ʔejn* ***-o*** *ʔohev* *xacilim* (marqueur de sujet)
NÉG -3MS aimer(BENONI).MS aubergines

c. ***hu*** *ʔejn* ***-o*** *ʔohev* *xacilim*
lui NÉG -3MS aimer(BENONI) aubergines
« Il n'aime pas les aubergines »

Il affirme :

> The paradigm strongly suggests that the suffix on *ʔejn* should not be analysed as an incorporated pronoun but, rather, as an inflectional marker, a manifestation of subject agreement. (*Ibid.*, p. 184).

L'auteur en conclut que l'obligation d'insérer un sujet lexical avec les marqueurs de sujet de la 3e personne de l'hébreu suggère un traitement unifiant dans lequel tous les marqueurs de sujet

[13] L'auteur adresse par ailleurs la critique suivante à l'approche argumentale, et plus précisément à Fassi Fehri (1993) : si les marqueurs de sujet sont des pronoms incorporés, la question se pose de savoir pourquoi ces marqueurs sont influencés par l'aspect du verbe. Nous traitons cette question dans la section (3.4.1).

[14] (Shlonsky, 1997, p. 184).

(même ceux qui ne sont pas à la 3e personne) doivent être considérés comme des marqueurs d'accord et non comme des arguments.

Cependant, nous pensons que généraliser cette analyse en essayant de l'appliquer à l'arabe pose un problème, et non des moindres. En fait, l'arabe se distingue de l'hébreu en ceci que l'emploi des sujets lexicaux avec les marqueurs de sujet de 3e personne n'est pas requis. Comme le reconnaît l'auteur lui-même, l'hébreu est une langue qu'on peut qualifier de *semi pro-drop*, où le sujet doit être lexicalisé à la 3e personne : « Null subjects are admitted in conjunction with first and second person, but not with third person inflection. » (*Ibid.*, p. 111). L'arabe, au contraire, n'impose pas cette contrainte. La possibilité d'une phrase, comme celle en (3.20), où seul un marqueur de sujet de 3e personne *-u:* est réalisé sans avoir besoin d'un sujet lexical, illustre cette particularité de la langue arabe, comparée à l'hébreu :

(3.20)	*saʔal*	*-u:*	*-ni:*	*s'-*	*s'afħ*	*-a*	*fa-*	*s'afaħ*	*-tu*
	demander.PER	-3MP	-me	le-	pardon	-ACC	et-	pardonner.PER	-1S

« Ils m'ont demandé pardon et j'ai pardonné »

Cette différence entre l'hébreu et l'arabe mine le premier argument avancé par Shlonsky. Son deuxième argument, basé sur des observations du berbère [15], connaît le même sort. En effet, l'auteur souligne qu'en berbère tamazight d'Ait-Seghrouchen [16], tout comme en arabe standard, les pronoms objets directs sont toujours attachés à l'extérieur des marqueurs de sujet, comme l'illustrent les exemples suivants [17] :

(3.21) a. *ttci* *-x* *aysum*
manger.PER -1S viande
« J'ai mangé de la viande »

b. *ttci* *-x* *-t*
manger.PER -1S -la
« J'en ai mangé »

Puisque les marqueurs d'objet du berbère, à la différence des marqueurs de sujet, doivent occuper la 2e position dans la proposition (c'est-à-dire à droite du verbe quand ce dernier est initial, ou à gauche du verbe quand ce dernier est précédé d'un membre d'une certaine catégorie), Shlonsky en conclut que le marqueur de sujet *-x* en (3.21) n'est pas un pronom clitique parce qu'il

[15] Langue non sémitique de la famille afro-asiatique.
[16] Voir Guerssel (1985).
[17] (Shlonsky, 1997, p. 186)

n'a pas cette position variable. Il affirme : « Subject agreement in Berber is a conjugation, while object clitics are second position clitics. » (*Ibid.*, p. 187). En généralisant cette analyse, l'auteur affirme que les marqueurs de sujet de l'arabe se comportent comme leurs semblables du berbère et doivent donc être analysés de la même manière.

La conclusion à laquelle aboutit l'auteur repose sur une prémisse que nous ne pouvons partager avec lui. Il suppose en fait que les différents pronoms clitiques d'une langue donnée doivent avoir une distribution similaire. Transposée aux langues romanes, cette affirmation s'affaiblit. En français, par exemple, les clitiques sujets n'ont pas la même distribution que les clitiques objets [18], sans que cela ne soit interprété de manière que les clitiques objets soient les seuls pronoms clitiques de cette langue.

Nous avons donc des raisons de penser que les arguments avancés par Shlonsky (1997) ne peuvent justifier ses prises de position en faveur de l'approche flexionnelle. Le seul argument à prendre en considération est alors le suivant : tout affixe est passible d'être un marqueur d'accord, jamais un argument. Nous pensons, tout comme le font Auger (1994) et Stump (1980), que le statut morphologique ne doit pas être corrélé au statut morphosyntaxique. Les données de plusieurs langues, dont l'arabe, appuient cette hypothèse. Faudrait-il donc pencher vers une approche radicalement opposée à celle de Shlonsky (1997) et prétendre que les marqueurs de sujet sont des arguments ? Nous examinons cette question dans la section subséquente.

3.3.2 L'hypothèse argumentale

La deuxième approche à la question du statut morphosyntaxique des marqueurs de sujet considère que ces entités sont des arguments et non des marqueurs d'accord, ne serait-ce que dans certains contextes. Cette hypothèse est défendue dans la tradition grammaticale arabe, mais également dans plusieurs recherches linguistiques contemporaines. Pour exposer les idées principales qui fondent cette approche, nous prenons comme exemple la vision grammaticale traditionnelle de la morphosyntaxe des marqueurs de sujet.

La pronominalité de certains marqueurs de sujet ne faisait aucun doute pour les grammairiens depuis l'apparition du premier livre de grammaire arabe au VIIIe siècle, le *kita:b* de Sibawayhi, publié dans Sibawayhi (1977). Cet auteur a commencé une longue tradition dans laquelle l'on

[18] Nous pensons, entre autres, à deux différences importantes : les clitiques sujets sont toujours proclitiques, alors que les clitiques objets peuvent être enclitiques également. De plus, seuls les clitiques objets peuvent subir une « montée » syntaxique. Les clitiques sujets sont toujours attachés à leur hôte verbal.

reconnaissait deux sortes de pronoms : des pronoms indépendants (*munfas'ila*) et des pronoms conjoints (*muttas'ila*). Les pronoms conjoints sont à leur tour analysés comme appartenant à deux types : des pronoms conjoints visibles et des pronoms conjoints invisibles (*mustatira*). Ils sont également classés selon le cas en pronoms conjoints nominatifs et pronoms conjoints non nominatifs. Dans cette conception, les pronoms conjoints invisibles sont une sous-classe des pronoms conjoints nominatifs. Le tableau (3.2) présente la liste traditionnelle des pronoms conjoints nominatifs visibles.

	Singulier		Duel		Pluriel	
	PER	IMP	PER	IMP	PER	IMP
1	-tu	?-	-na:		-na:	
2M	-ta		-tuma:	-a:	-tum	-u:
2F	-ti	-i:	-tuma:	-a:	-tunna	-na
3M			-a:	-a:	-u:	-u:
3F			-ata:	-a:	-na	-na

TABLEAU 3.2: Les pronoms conjoints nominatifs dans la tradition grammaticale

Les pronoms conjoints invisibles sont censés remplir les cases vides dans ce tableau et compléter le paradigme de la conjugaison impérative à la 2MS (la seule forme à laquelle le verbe est complètement nu de tout autre morphème que le radical). La tradition grammaticale a donc proposé une solution au problème posé par les 4 marqueurs de sujet de 3S du perfectif et de l'imperfectif en les excluant de la classe des pronoms ; ils ne sont, pour les grammairiens, que des marques morphologiques aspectuelles (c'est le cas des marqueurs de sujet de 3S *-a*, *j-* et *t-*) et qui marquent le féminin par opposition au masculin (c'est le cas du marqueur de sujet de 3FS *-at*).

Cette conception des pronoms conjoints comporte également une proposition de solution au problème de la conjugaison imperfective ; selon les grammairiens, les préfixes ne sont pas des pronoms, mais des marqueurs d'aspect. Nous avons critiqué cette conception dans le deuxième chapitre étant donné qu'elle n'explique pas la variabilité de ces *marqueurs d'aspect* en genre et en personne.

Cette approche a cependant plusieurs avantages. Tout d'abord, les pronoms conjoints qui y sont reconnus comme tels sont des suffixes pour la majorité d'entre eux. Ensuite, une contrainte forte est mise sur les pronoms invisibles : ces pronoms sont une sous-classe des pronoms conjoints nominatifs et ils ne sont pas légitimés par la présence d'une morphologie riche sur le verbe, mais par l'absence d'une telle morphologie. Cette conception est donc très différente de l'approche contemporaine concernant les pronoms nuls. Elle garde en fait un aspect empirique important :

on ne stipule pas l'existence de pronoms nuls si des candidats lexicalisés sont présents dans la structure.

Malgré ces avantages, l'approche grammaticale est confrontée à un certain nombre de problèmes. Elle reste, en fait, très influencée par sa propre conception de la morphologie du verbe en arabe. En classant les verbes en figés (*mabnijja*) et fléchis (*muʕraba*), elle ne pouvait se détacher de l'idée de trouver les marques de figement, qui sont généralement la voyelle ou la consonne finale. Dans le cas du perfectif, le marqueur de sujet *-a* a été pris pour une marque de figement caractéristique de l'aspect perfectif. Cet élément ne pouvait donc être un pronom conjoint pour les grammairiens, qui stipulaient que le pronom conjoint sujet est invisible dans ce contexte.

3.3.3 La mixité des marqueurs de sujet

Nous avons des raisons de penser que le terme *marqueurs de sujet* couvre deux réalités différentes. Il sert en fait d'étiquette à des arguments pronominaux et à des marqueurs d'accord verbaux. Quatre parmi les éléments du premier type sont des homophones d'éléments du deuxième type, ce qui leur confère un statut morphosyntaxique apparemment mixte. Ces marqueurs homophones sont ceux de la 3S [19] perfectifs (*-a* et *-at*) et imperfectifs (*j-* et *t-*). Les deux tableaux suivants illustrent cette hypothèse.

	Singulier		Duel		Pluriel	
	PER	IMP	PER	IMP	PER	IMP
1	-tu	ʔ-	-na:	n-	-na:	n-
2M	-ta	t-	-tuma:	t-...-a:	-tum	t-...-u:
2F	-ti	t-...-i:	-tuma:	t-...-a:	-tunna	t-...-na
3M	-a	j-	-a:	j-...-a:	-u:	j-...-u:
3F	-at	t-	-ata:	t-...-a:	-na	j-...-na

TABLEAU 3.3: Les marqueurs de sujet pronominaux

[19] Le fait que ces marqueurs sont à la 3e personne pourrait s'expliquer par la relation d'accord qui se fait entre le sujet et le verbe. Le sujet encode la 3e personne puisque c'est un SN et le verbe encode le même trait pour réaliser cette relation.

	Perfectif	Imperfectif
3M	-a	j-
3F	-at	t-

TABLEAU 3.4: Les marqueurs d'accord verbaux

Cette approche se distingue des deux approches précédentes en ceci qu'elle envisage le système pronominal de l'arabe dans sa totalité, qu'elle intègre la question des asymétries de l'accord et qu'elle traite le problème du statut du SN/pronom indépendant préverbal/postverbal en employant des critères indépendamment motivés. Ainsi, nous ne contentons pas d'exposer notre vision de la morphosyntaxe des marqueurs de sujet et d'y fournir des arguments, nous examinons également les répercussions de cette approche sur le système pronominal, le système de l'accord et l'analyse des données dialectales.

Nous commençons par présenter les notions d'accord grammatical, d'accord anaphorique et d'ambiguïté fonctionnelle proposées par Bresnan et Mchombo (1987) et nous utilisons les tests cités dans leur travail pour appuyer l'hypothèse suivante : les marqueurs de sujet entrent dans deux relations d'accord différentes ; une relation d'accord anaphorique, dans laquelle ils jouent le rôle d'arguments, et une relation d'accord grammatical, dans laquelle ils jouent le rôle de marqueurs d'accord. Cette mixité illustre un cas de ce que ces auteurs appellent « ambiguïté fonctionnelle ».

3.3.3.1 Relations d'accord et ambiguïté fonctionnelle

Bresnan et Mchombo (1987) distinguent deux sortes d'accord : l'accord *grammatical* et l'accord *anaphorique*. Dans l'accord grammatical entre un SN et un marqueur de sujet, le SN est l'argument du verbe alors que le marqueur sur ce dernier exprime d'une manière redondante la personne, le nombre et le genre de ce SN. Dans l'accord anaphorique, par contre, le marqueur est un argument pronominal incorporé et le SN coréférentiel n'est donc pas argumental ; il fonctionne ainsi comme un adjoint du pronom incorporé ou comme un focus ou un topique de la proposition ou de la structure discursive. Cette approche de l'accord s'inscrit dans le cadre de la grammaire lexicale fonctionnelle (LFG) et sert à fournir une analyse des marqueurs de sujet/d'objet du chichewa (une langue bantoue). Les auteurs nous démontrent que ces marqueurs n'exhibent pas la même sorte d'accord . si les marqueurs d'objet sont toujours employés pour marquer un accord

anaphorique et qu'ils sont toujours des pronoms incorporés, les marqueurs de sujet sont employés d'une manière ambiguë pour marquer les deux sortes d'accord ; ils sont des pronoms incorporés dans certains contextes et des marqueurs d'accord dans d'autres. Le critère décisif est ici la présence ou l'absence d'un SN sujet, ou dans les termes des auteurs : « the 3rd person pronominal interpretation of S[ubject] M[arker] will arise when and only when there is no subject NP in the phrase structure » (*Ibid.*, p. 745). De plus, plusieurs tests permettent de distinguer l'accord anaphorique de l'accord grammatical, comme la localité, le statut (périphérique vs central) du SN co-indicé avec le marqueur et l'appartenance de la langue étudiée à un type donné. Nous présentons ici le test de la localité et nous abordons la question de la typologie de la langue arabe [20].

Selon les auteurs, les relations d'accord grammatical sont toujours locales alors que les relations d'accord anaphorique peuvent ne pas l'être. La localité est entendue ici en termes de proximité des éléments en relation d'accord dans la proposition :

> A local agreement relation is one which holds between elements of the same simple clause, while a non-local agreement relation is one which may hold between elements of different clauses. (*Ibid.*, p. 752).

Ce critère implique que si le SN ou le pronom indépendant co-indicés avec le marqueur de sujet sont placés en dehors de la proposition qui contient ce dernier, la relation d'accord entre les deux est anaphorique et non grammaticale.

Pour que ce critère soit valable en arabe, il faut que ses prédictions soient correctes en ce qui a trait à ces deux types d'accord dans cette langue. Il est donc primordial de vérifier si l'accord anaphorique (par exemple, entre un marqueur d'objet et un SN ou un pronom indépendant coréférentiels) peut ne pas être local et si l'accord grammatical (par exemple, entre un SN et un adjectif) est toujours local.

En ce qui a trait à l'accord anaphorique, nous avons déjà signalé que le redoublement d'un marqueur d'objet par un SN ou un pronom indépendant est impossible, mais que la dislocation est possible. Les exemples suivants illustrent des cas de dislocation :

(3.22) a. *raba:b -u* *ʔaxawa:t* *-u* *-ha:* *ʕaʃr* *-u* *-n*
Raba-NOM sœurs -NOM -ses dix -NOM -Indé
« Rabab, ses sœurs sont au nombre de dix »

[20] Les autres tests sont également concluants.

b. *hija* *ʔaxawa:t* *-u* ***-ha:*** *ʕaʃr* *-u* *-n*
elle sœurs -NOM -ses dix -NOM -Indé
« Elle, ses sœurs sont au nombre de dix »

Le SN en (3.22a) et le pronom indépendant en (3.22b) sont co-indicés avec le marqueur d'objet *-ha:*. L'accord ici est anaphorique et non grammatical. En outre, le SN et le pronom indépendant ne se trouvent pas dans le même domaine de localité que le marqueur d'objet puisqu'ils sont disloqués.

L'accord grammatical entre un SN et un adjectif exige une stricte localité : les deux doivent se trouver au sein du même SN, comme le montre le contraste suivant, où (3.23b) est agrammaticale malgré le fait que l'arabe est une langue à ordre (relativement) libre :

(3.23) a. *ʔinna* *-ha:* *lajlat* *-u* *-n* *t'awi:lat* *-u* *-n*
que -elle nuit -NOM -Indé longue -NOM -Indé
« C'est une longue nuit, certes »

b. **t'awi:lat* *-u* *-n* *ʔinna* *-ha:* *lajlat* *-u* *-n*
longue -NOM -Indé que -elle nuit -NOM -Indé

Nous pouvons donc affirmer que le test de la localité reflète des propriétés réelles dans le système de l'accord de l'arabe. En l'appliquant aux marqueurs de sujet, nous découvrons que la relation d'accord entre les marqueurs de sujet et le SN/pronom indépendant préverbal/postverbal n'est pas toujours locale. Quand le composant co-indicé avec le marqueur de sujet est disloqué, il se trouve dans une relation d'accord non local avec le marqueur de sujet et nous pouvons ainsi parler d'accord anaphorique. Toutefois, ce composant se trouve dans certains contextes en proximité du marqueur de sujet dans une relation que nous pouvons qualifier de locale. L'accord, dans un tel cas, est donc grammatical. Ces contextes sont cependant très limités. Le composant en question doit être un SN postverbal et le marqueur de sujet à la 3S, comme dans les exemples suivants :

(3.24) a. *za:r* ***-a*** *-na:* *l-wa:li:*
visiter.PER -3M -nous le-gouverneur
« Le gouverneur nous a rendu visite »

b. *za:r* ***-a*** *-na:* *l-wa:lija:ni*
visiter.PER -3M -nous les-deux.gouverneurs
« Les deux gouverneurs nous ont rendu visite »

c. *za:r* *-a* *-na:* *l-wula:la:t -u*
visiter.PER -3M -nous les-gouverneurs-NOM
« Les gouverneurs nous ont rendu visite »

Les exemples en (3.24) pourraient être traités comme des cas de redoublement puisqu'ils sont conformes à la définition du redoublement qu'on a adoptée dans la section (3.1.2). Ce n'est cependant pas à strictement parler un redoublement étant donné que ces exemples échouent au critère suivant, appelé critère des dépendances à longue distance : un SN qui redouble un marqueur de sujet ne peut entretenir une relation à longue distance avec son hôte selon Aoun et Benmamoun (1998). Toutefois, les SN en (3.24) peuvent entretenir de telles relations (quand ils sont remplacés par des éléments Qu « *Wh* »), comme le montrent les exemples suivants :

(3.25) a. *man*$_i$ *za:r* *-a*$_i$ *-kum*
qui visiter.PER -3MS -vous
« Qui vous a rendu visite ? »

b. *man*$_i$ *za:r* *-a:*$_i$ *-kum*
qui visiter.PER -3MD -vous
« Qui sont les deux personnes qui vous ont rendu visite ? »

c. *man*$_i$ *za:r* *-u:*$_i$ *-kum*
qui visiter.PER -3MP -vous
« Qui sont les personnes qui vous ont rendu visite ? »

L'accord n'est pas local en (3.25). Il s'agit donc d'un accord anaphorique. Le marqueur de sujet est, dans ces structures, un argument incorporé au verbe et associé à un élément Qu (lui-même non argumental). D'autres exemples de cet accord anaphorique sont donnés en (3.26) et en (3.27) et impliquent des SN et des pronoms indépendants préposés et postposés. Il s'agit dans tous ces cas de dislocations à gauche ou à droite :

- Dislocation à gauche :

(3.26) a. *ʔal- wa:lija:ni* *za:r* *-a:* *-na:*
les- deux.gouverneurs.NOM visiter.PER -3MD -nous
« Les deux gouverneurs, ils nous ont rendu visite »

b. *huma:* *za:r* *-a:* *-na:*
eux.deux visiter.PER -3MD -nous
« Eux deux, ils nous ont rendu visite »

- Dislocation à droite :

(3.27) a. *za:r* *-a:* *-na:* *l- wa:lija:ni*
visiter.PER -3MD -nous les- deux.gouvernenurs.NOM
« Ils nous ont rendu visite, les deux gouvernenurs »

b. *za:r* *-a:* *-na:* *huma:*
visiter.PER -3MD -nous eux.deux
« Ils nous ont rendu visite, eux deux »

Nous pouvons donc avancer que les marqueurs de sujet exhibent deux sortes d'accord : un accord grammatical, auquel cas ils sont des marqueurs d'accord avec un SN postposé (l'argument nominatif) et un accord anaphorique, auquel cas ils sont des pronoms incorporés à la tête verbale et associés anaphoriquement à des SN ou à des pronoms indépendants disloqués (non argumentaux). Ce double rôle joué par certains marqueurs de sujet (à savoir ceux de la 3S) illustre ce que Bresnan et Mchombo (1987) appellent *l'ambiguïté fonctionnelle.*

Il nous est, par voie de conséquence, possible d'affirmer que l'arabe est une langue qui incorpore les pronoms à la tête verbale, comme le fait le chichewa. L'arabe présente, d'ailleurs, les trois traits qui distinguent de telles langues :

1. La possibilité de l'omission du SN sujet, et ce, grâce à la présence d'un marqueur de sujet pronominal muni de tous les traits, dont le nombre.
2. Le verbe n'assigne pas de cas au SN/pronom indépendant avec lequel le marqueur s'accorde anaphoriquement étant donné que ces composants sont périphériques. Ils apparaissent ainsi toujours au nominatif en arabe, un cas assigné par défaut.
3. Dans les langues qui connaissent deux sortes de formes pronominales anaphoriques, les formes au contenu phonétique le plus complexe sont employées pour exprimer le contraste, l'emphase ou la focalisation. En arabe, ce rôle est joué par les pronoms indépendants.

3.3.3.2 Le statut du référent

En chichewa, les marqueurs de sujet qui entrent dans une relation d'accord grammatical (avec des SN) sont coréférentiels avec des sujets, alors que les marqueurs de sujet qui entrent dans une relation d'accord anaphorique sont coréférentiels avec des topiques :

> When the S[ubject] M[arker] is used as a grammatical agreement, it agrees with a nominal that has the SUBJ[ECT] function ; when the SM is used for anaphoric binding, its antecedent within the sentence has the TOP[IC] function. (Bresnan et Mchombo, 1987, p. 757)

Par conséquent, on s'attendrait à ce que les référents des marqueurs de sujet anaphoriques de l'arabe soient des topiques et à ce que les référents des marqueurs de sujet grammaticaux soient des sujets. Pour vérifier la validité de cette hypothèse, nous proposons l'emploi de tests

indépendamment motivés pour opérer la distinction entre le sujet et le topique [21]. Li et Thompson (1976) proposent de tels tests. L'application de ces derniers nous permet de conclure que le SN postverbal référent d'un marqueur de sujet à la 3S est le sujet alors que le SN préverbal et le pronom indépendant NOM (qu'ils soient préverbaux ou postverbaux) sont des topiques [22].

Sept tests sont proposés par Li et Thompson (1976) : la définitude, la sélection, la détermination, l'accord avec le verbe, la position initiale et les processus grammaticaux. Nous en présentons ici trois : la définitude, la position initiale et les processus grammaticaux. [23]

3.3.3.2.1 La définitude

Selon Li et Thompson (1976), le topique est obligatoirement défini alors que le sujet n'a pas à se plier à cette condition. En arabe, le SN postverbal référent (que nous pouvons considérer comme étant le sujet) peut être indéfini comme en (3.28a) et peut également être défini comme en (3.28b) :

(3.28) a. ʤa:ʔ -a *raʤul* -u -n (sujet indéfini)
arriver.PER -3M homme -NOM -Indé
« Un homme est arrivé »

b. ʤa:ʔ -a *r-* *raʤul* -u (sujet défini)
arriver.PER -3M l'- homme -NOM
« L'homme est arrivé »

Le SN préverbal référent doit toutefois être défini [24] (3.29a). Quand le SN préverbal est indéfini, il faut que le contexte soit marqué, sinon la phrase est agrammaticale, comme c'est le cas en (3.29b). Par ailleurs, le SN/pronom indépendant topique doit être défini qu'il soit associé à un marqueur de sujet comme en (3.29a) ou à un marqueur d'objet comme en (3.29c) :

(3.29) a. *ʔar-* *rajul*$_i$ -u ʤa:ʔ -a$_i$ (topique défini)
l'- homme -NOM venir.PER -3MS
« L'homme, il est arrivé »

[21] Lambrecht (1994, p. 118) affirme ce qui suit concernant la différence entre ces deux notions : « Topics are not necessarily grammatical subjects, and grammatical subjects are not necessarily topics. » Le terme « sujet » que nous employons ici est une abréviation du terme « sujet grammatical » tel qu'employé par Lambrecht.

[22] Toutefois, dans plusieurs travaux, dont ceux de Ackema et Neeleman (2003), Aoun *et al.* (1994) et Harbert et Bahloul (2002), l'on défend l'idée selon laquelle le SN ou le pronom indépendant NOM associés aux marqueurs de sujet sont des arguments nominatifs (sujets) quelle que soit leur emplacement.

[23] Les autres tests sont également concluants, mais nous nous contenterons ici de ces exemples.

[24] De même, le pronom indépendant est défini.

b. **rakʒul -u -n ʤa:ʔ -a*[25] (topique indéfini)
homme -NOM -indé venir.PER -3MS

c. *ʔar- rakʒul$_i$ -u raʔj -tu -hu$_i$* (topique objet défini)
l'- homme -NOM voir.PER -1S -le
« L'homme, je l'ai vu »

Selon ce premier critère, nous avons des raisons de soutenir l'idée selon laquelle le SN/pronom préverbal sont des topiques alors que le SN postverbal est le sujet. Les deux tests suivants mènent à la même conclusion.

3.3.3.2.2 La position initiale

Selon les auteurs, la position initiale dans la phrase est la position *sui generis* des topiques. Le sujet peut lui aussi apparaître dans cette position (en français, par exemple), mais il peut également apparaître dans d'autres positions (le verbe peut être initial en arabe et, en malgache, le verbe apparaît en position finale). Nous pouvons donc en déduire que la tendance générale du topique est d'être en position initiale[26]. En arabe, le constituant préverbal référent a toutes les chances d'être un topique par le simple fait qu'il est en position initiale et le SN référent postposé a de fortes chances d'être le sujet. Ce test est bien évidemment insuffisant et doit être jumelé à d'autres tests, comme celui notamment des processus grammaticaux.

3.3.3.2.3 Les processus grammaticaux

Selon Li et Thompson (1976), le sujet, et non le topique, peut être impliqué dans certains processus grammaticaux spécifiques au domaine interne d'une phrase. Parmi ces processus, les auteurs citent la formation des pronoms réfléchis, du passif et de l'impératif. L'exemple du passif est très représentatif de ces processus. En fait, la formation de ce dernier dans plusieurs langues implique le sujet et non le topique. C'est le cas en français, mais également en arabe. Dans cette langue, le passif s'obtient par le changement du patron (ou schème) morphologique du

[25](Ackema et Neeleman, 2003) citent un exemple semblable qu'ils considèrent grammatical : baqarat-u-n takallama-at (vache-NOM-Indé parler.PER-3FS) « Une vache a parlé ». Cette construction n'est possible en arabe que lorsque le contexte est marqué. Dans ce cas-ci, le locuteur relate un événement hors du commun et doit, à l'oral, opérer une coupure intonatoire nette.

[26] Il s'agit bien sûr d'une tendance générale. Lambrecht (1994) conteste ce critère pourtant accepté par plusieurs auteurs.

verbe, doublé d'une élision du sujet et de la promotion du complément du verbe en sujet[27] (une promotion accompagnée de l'assignation du cas nominatif). En (3.30a), nous avons un exemple d'une phrase à la voix active avec à la fois un topique (le SN préverbal) et un sujet (le marqueur de sujet). Seul ce dernier contribue à la formation du passif en (3.30b) alors que le premier demeure inchangé :

(3.30) a. *ʔal- falla:ħ -u:na ħalab -u: baqara:t -i -him*
les- agriculteurs -NOM traire.PER -3MP vaches -ACC -leurs
« Les agriculteurs, ils ont trait leurs vaches »

b. *ʔal- falla:ħ -u:na ħulib -at baqara:t -u -hum*
les- agriculteurs -NOM traire.PASSIF.PER -3F vaches -NOM -leurs
« Les agriculteurs, leurs vaches ont été traites »

La phrase passive reliée à une phrase verbale sans SN topique disloqué à gauche est différente de la phrase obtenue en (3.30b), comme en témoignent les deux exemples suivants :

(3.31) a. *ħalab -a l- falla:ħ -u:na l- baqara:t -i*
traire.PER -3M les- agriculteurs -NOM les- vaches -ACC
« Les agriculteurs ont trait les vaches »

b. *ħulib -at l- baqara:t -u min qibal -i l- falla:ħ -i:na*
traire.PASSIF.PER -3F les- vaches -NOM de la.part -GÉN les- agriculteurs -GÉN
« Les vaches ont été traites par les argriculteurs »

Dans ces exemples, le SN postverbal de la phrase passive, que nous considérons comme étant le sujet, donne lieu à une sorte de complément d'agent dans la phrase passive, démontrant ainsi son rôle syntaxique de sujet.

En appliquant les tests de Li et Thompson (1976), nous sommes en mesure de soutenir l'hypothèse selon laquelle le SN/pronom indépendant nominatif préverbal associé à un marqueur de sujet est un topique.

[27] Les grammairiens de la tradition parlent, dans ce cas, d'un « délégué du sujet ». Ce constituant se comporte comme le sujet en ce qui a trait à l'accord.

Nous avons également utilisé les tests de Bresnan et Mchombo (1987) pour statuer sur la nature de l'accord observé entre le composant nominal ou pronominal nominatif et les marqueurs de sujet. Il s'est avéré que ces marqueurs de sujet sont mieux traités comme des pronoms incorporés à la tête verbale et que l'accord observé est anaphorique et non grammatical, sauf dans un contexte : celui où le SN *redouble* le marqueur de sujet. Cette hypothèse s'inscrit dans la lignée des idées avancées par plusieurs auteurs qui ont étudié la grammaticalisation des pronoms dans les langues naturelles. Moravcsik (1974) et Givon (1976), par exemple, ont défendu l'hypothèse selon laquelle l'accord entre le verbe et ses arguments résulte de la grammaticalisation de pronoms qui étaient redoublés par ces arguments[28]. Le statut des marqueurs de sujet de l'arabe est compatible avec cette hypothèse. Cette dernière nous permet par ailleurs de résoudre le problème soulevé par la question de savoir pourquoi les marqueurs d'objet de cette langue ne donnent pas lieu à des marqueurs d'accord : les marqueurs d'objet ne peuvent être redoublés par des arguments lexicaux, et cela ne constitue pas un contexte propice au développement de marqueurs d'accord.

Ces hypothèses ont des répercussions sur le traitement de plusieurs phénomènes morphosyntaxiques en arabe. En fait, si nous considérons que la seule relation d'accord possible entre le sujet et le verbe est celle qui implique un accord *partiel* (avec un marqueur de sujet toujours au singulier), il faudrait justifier cette hypothèse. La question des asymétries de l'accord reste par ailleurs entière . pourquoi le nombre n'est encodé que lorsque le verbe n'est pas initial ? Nous abordons ces questions dans la section subséquente.

3.4 Les implications

Quatre marqueurs de sujet sont ambigus, à savoir *-a*, *-at*, *j-* et *t-*, et peuvent entrer dans deux sortes de relations d'accord : un accord grammatical et un accord anaphorique. L'accord entre le sujet et le verbe est de nature grammaticale et ne se réalise que dans la configuration V+SN. Lorsque le pronom indépendant nominatif est postverbal, dans les deux configurations où le verbe n'est pas initial et lorsque seul le marqueur de sujet est présent, ce marqueur est un pronom incorporé argumental et le SN ou le pronom indépendant, si présents, sont disloqués. Cette vision de l'accord entre le sujet et le verbe en arabe pourrait poser certains problèmes. L'un de ces problèmes est le suivant[29] : si les marqueurs de sujet sont des arguments (autrement dit, sujets) dans certains contextes du moins, pourquoi prennent-ils des formes différentes selon

[28] Voir également Miller et Monachesi (2003).
[29] Voir (Shlonsky, 1997, p. 185).

l'aspect ? N'est-ce pas là la preuve qu'ils sont plutôt des marqueurs d'accord qui s'amalgament aux autres marqueurs du verbe (les marqueurs de mode, de temps, d'aspect, etc.) ? Nous dédions la section (3.4.1) à cette question.

Un autre problème concernerait l'accord en lui-même ; on s'attendrait en fait à ce que l'accord authentique soit riche et non partiel. Comment peut-on justifier le fait que l'accord entre le sujet et le verbe en arabe est partiel, alors que cette langue est traditionnellement considérée comme étant une langue à morphologie riche ? Nous consacrons la section (3.4.2) à cette question.

Le dernier problème que nous soulevons est le suivant : compte tenu de cette analyse, comment rendre compte des asymétries de l'accord ? Nous tentons de répondre à cette question dans la section (3.4.3).

3.4.1 Les marqueurs de sujet et l'aspect

Le premier problème se pose surtout quand nous comparons les marqueurs de sujet de l'arabe aux clitiques sujets de certaines langues romanes. En effet, les clitiques de ces langues semblent prendre les mêmes formes quels que soient le mode et le temps de conjugaison et semblent donc indépendants de la morphologie du verbe. Les marqueurs de sujet de l'arabe, d'un autre côté, sont influencés par les propriétés flexionnelles du verbe et apparaissent tantôt comme des suffixes (avec le perfectif et l'impératif), tantôt comme des préfixes et des circumfixes dans la conjugaison imperfective, comme illustré dans le tableau (3.5).

	Singulier		Duel		Pluriel	
	PER	IMP	PER	IMP	PER	IMP
1	-tu	?-	-na:	n-	-na:	n-
2M	-ta	t-	-tuma:	t-...-a:	-tum	t-...-u:
2F	-ti	t-...-i:	-tuma:	t-...-a:	-tunna	t-...-na
3M	-a	j-	-a:	j-...-a:	-u:	j-...-u:
3F	-at	t-	-ata:	t-...-a:	-na	j-...-na

TABLEAU 3.5: Les marqueurs de sujet

Ce comportement pourrait paraître singulier, mais les marqueurs de sujet de l'arabe ne sont pas uniques en leur genre. Étant des morphèmes qui interagissent avec le verbe, il n'est pas surprenant qu'ils soient influencés par les propriétés flexionnelles de ce dernier. Les clitiques des langues romanes, qui ont également le verbe pour hôte, sont influencés dans une large mesure par les propriétés flexionnelles de cet hôte. Ainsi, le choix entre la proclise et l'enclise, par exemple,

dans des langues telles que l'italien, le catalan, l'espagnol et le roumain, comme le soulignent Miller et Monachesi (2003), dépend de la flexion du verbe : l'enclise est obligatoire avec l'infinitif et la proclise est exigée avec les verbes finis. Les exemples suivants de l'italien illustrent cette influence :

(3.32) a. *Martina vuole leggerlo*
Martine veut lire-le
« Martine veut le lire »

b. *Martina lo legge*
Martine le lit
« Martine le lit »

En français standard, par contre, la proclise est de mise avec l'infinitif, le participe présent et l'impératif négatif, alors que l'enclise est de mise avec l'impératif affirmatif. C'est dire que le site d'attachement de ces marqueurs d'arguments est influencé par la morphologie de cet hôte particulier qu'est le verbe et que cette influence peut être différente d'une langue à l'autre, même au sein d'une même famille.

En outre, si nous restons dans le cadre restreint des marqueurs de sujet, nous remarquons d'autres ressemblances avec la famille romane. Les dialectes du Nord italien sont probablement les dialectes les plus étudiés quand il s'agit des pronoms clitiques sujets. Dans certains de ces dialectes, les pronoms clitiques sujets peuvent avoir des formes différentes selon qu'ils sont proclitiques ou enclitiques. Dans le dialecte de Padua, par exemple, selon Cardinaletti et Repetti (2006), le clitique sujet de 3MS a la forme *el* quand il est proclitique et *lo* quand il est enclitique :

(3.33) a. *el magna*
il mange
« Il mange »

b. *magne-lo*
mange-il
« Mange-t-il ? »

La ressemblance avec l'arabe est d'autant plus frappante que le pronom clitique sujet dans le dialecte parlé à Milan, par exemple, se compose de deux parties : un proclitique et un enclitique, comme l'illustre l'exemple suivant [30] :

[30] Cité par Heap (1997).

(3.34) *Ti* *te* *parle* *-t*
toi tu parles -2S
« Toi, tu parles »

Dans cette construction, l'enclitique *-t* fait partie de la morphologie verbale, le proclitique *te* est obligatoire et le pronom indépendant *ti* exprime l'emphase.

En poussant la comparaison un peu plus loin, nous découvrons que l'analyse de certains clitiques sujets dans ces dialectes du Nord italien par certains auteurs se fait suivant les mêmes lignes que celles que nous avons adoptées pour l'analyse des marqueurs de sujet de l'arabe. Ainsi, Brandi et Cordin (1982), cités par Ramiro (1986), soutiennent une idée selon laquelle, dans les deux dialectes trentin et florentin, les clitiques sujets sont des arguments en absence de tout sujet lexical et des flexions verbales (des marqueurs d'accord) quand un sujet préverbal ou postverbal est réalisé. Dans les exemples suivants, la présence d'un sujet lexical (le pronom indépendant *lei*, qui ne déclenche pas une lecture emphatique) en (3.35a) force une analyse du clitique sujet *la* comme étant un marqueur d'accord. Ce même clitique est analysé comme un argument du verbe en l'absence d'un sujet lexical dans l'exemple[31] (3.35b) :

(3.35) a. *Lei* *la* *parla*
elle elle parle
« Elle parle »

b. *La* *parla*
elle parle
« Elle parle »

Nous sommes de ce fait en mesure d'affirmer que les marqueurs de sujet de l'arabe, dans leur ambiguïté fonctionnelle et dans leur sensibilité aux propriétés flexionnelles de leur hôte verbal, ne sont en aucun cas uniques. Un comportement semblable est observable dans la famille romane.

3.4.2 L'accord partiel

L'accord entre le sujet et le verbe ne se réalisant que lorsqu'un SN postverbal NOM associé au marqueur de sujet est réalisé, il s'ensuit que cet accord se fait seulement en genre et en personne (excluant ainsi le nombre). Nous avons des raisons de penser que cela est dû non à un statut

[31] (Ramiro, 1986, p. 207).

particulier de ce trait en arabe, mais à une propriété singulière de l'accord en général dans cette langue.

Plusieurs auteurs, dont Cantarino (1975) et Mohammad (2000), signalent que, dans les structures où le sujet est séparé du verbe par un autre constituant (l'objet par exemple), l'accord en genre peut être neutralisé si bien qu'on obtient un accord au masculin avec un sujet au féminin, comme dans l'exemple suivant :

(3.36) *j-* *askun* *-u* *l-* *ʕira:q* *-a* *ʔumam* *-u* *-n* *muxtalifat* *-u*
3M- habiter.IMP -Ind l'- Irqa -ACC peuples -NOM -Indé différents -NOM
-n
-Indé
« Différents peuples habitent l'Iraq »

En (3.36), le sujet est *ʔumamun muxtalifatun*. Ce pluriel collectif est traité comme un féminin singulier en arabe, comme le montre l'accord avec l'adjectif *muxtalifatun* (« différente », donc FS). Cependant, le verbe imperfectif *jaskunu* porte le trait du masculin et non du féminin (rappelons qu'à la 3FS le préfixe de l'imperfectif est *t-* et qu'à la 3MS le préfixe est *j-*).

D'une manière plus générale, comme le montrent Blachère et Gaudefroy-Demombines (2004), plusieurs contextes permettent d'accorder un sujet féminin comme s'il était masculin en arabe. Ces contextes sont les suivants : le sujet est un singulier ou un duel d'objet asexué ayant une terminaison typique du féminin (3.37a), un pluriel interne (3.37b) ou un collectif naturel (3.37c)[32] :

(3.37) a. *ðahab* *-a* *sa:ʕat -u -n* *mina* *l-* *lajl* *-i*
partir.PER -3M heure -NOM -Indé de la- nuit -GÉN
« Une fraction de la nuit s'écoula »

b. *qut'iʕ* *-a* *ʔajdi: -him*
trancher.PASSIF -3M mains -leurs
« Leurs mains furent tranchées »

c. *baʕaθ* *-a* *qurajʃ -un* ...
envoyer.PER -3M Quraych -NOM
« Quraych envoya ...»

Blachère et Gaudefroy-Demombines (2004) emploient le terme « accord flottant » pour désigner ce comportement, puisque le genre féminin est possible dans ces mêmes contextes :

[32] (Blachère et Gaudefroy-Demombines, 2004, p. 301).

(3.38) a. *ðahab* *-at* *sa:ʕat -u -n* *mina* *l-* *lajl* *-i*
partir.PER -3F heure -NOM -Indé de la- nuit -GÉN
« Une fraction de la nuit s'écoula »

b. *qut'iʕ* *-at* *ʔajdi: -him*
trancher.PASSIF -3F mains -leurs
« Leurs mains furent tranchées »

c. *baʕaθ* *-at* *qurajʃ -un* ...
envoyer.PER -3F Quraych -NOM
« Quraych envoya ... »

Cet accord *flottant* n'est pas propre à l'accord entre le sujet et le verbe (ou l'accord verbal). L'accord entre le nom et l'adjectif (ou l'accord nominal) présente les mêmes propriétés. En (3.39a), nous présentons un exemple de l'accord nominal flottant en genre et en nombre :

(3.39) a. *ʔaʃʤa:r* *-u* *-n* *ʕiɮa:m* *-u* *-n*
arbres -NOM -Indé grands -NOM -Indé
« de grands arbres »

b. *ʔaʃʤa:r* *-u* *-n* *ʕaɮi:mat* *-u* *-n*
arbres -NOM -Indé grande -NOM -Indé
« de grands arbres »

En arabe, le pluriel [33] des noms non animés est considéré comme étant au féminin singulier aux fins de l'accord. Cependant, en (3.39a), l'adjectif qui devrait porter les traits FS pour s'accorder avec le nom non animé *ʔaʃʤa:r-u-n* « arbres » porte les traits MP. L'exemple en (3.39b) présente un accord nominal régulier.

Cette similitude entre l'accord verbal et l'accord nominal est d'autant plus notable que nous retrouvons les mêmes asymétries dans les deux systèmes. Dans les exemples suivants, l'accord entre le déverbal et le SN ne se fait pas en nombre :

(3.40) a. *raʤul* *-u* *-n* *qa:dim -u -n* *ʔaxu: -hu*
homme -NOM -Indé arrivant -NOM -Indé frère -son
« un homme dont le père est en train d'arriver »

b. *raʤul* *-u* *-n* *qa:dim -u -n* *ʔaxawa: -hu*
homme -NOM -Indé arrivant -NOM.D -Indé frères.NOM.D -ses
« Un homme dont les deux frères sont en train d'arriver»

[33] Au singulier, un nom non animé peut être au masculin (comme *qalam* « crayon ») ou au féminin (comme *ma:ʔidat* « table »).

c. **raʤul* *-u* *-n* *qa:dim -a:ni* *ʔaxawa: -hu*
homme -NOM -Indé arrivants -NOM.D frères.NOM.D -ses

d. *raʤul* *-u* *-n* *qa:dim -u -n* *ʔixwat -u -hu*
homme -NOM -Indé arrivant.S -NOM -Indé frères.P -NOM -ses
« un homme dont les frères sont en train d'arriver »

e. **raʤul* *-u* *-n* *qa:dim -u:na* *ʔixwat -u -hu*
homme -NOM -Indé arrivant.P -NOM-Indé frères.P -NOM -ses

Le déverbal (un participe actif) dans ces exemples est placé à gauche du nom qu'il qualifie. La relation entre les deux est morphosyntaxiquement une relation d'accord entre un SN et un adjectif, mais sur le plan sémantique c'est une relation entre un prédicat et son argument. Le participe actif ne s'accorde pas en nombre avec le SN, d'où le singulier qui nous rappelle l'accord partiel entre le sujet et le verbe quand le verbe est initial.

Dans les exemples suivants, le participe actif est placé à droite du SN. L'accord en nombre obtenu dans ce contexte est semblable à l'accord obtenu dans l'ordre dit SVO (un accord riche) :

(3.41) a. *raʤul* *-u* *-n* *ʔaxu: -hu* *qa:dim -u -n*
homme -NOM -Indé frère -son arrivant -NOM -Indé
« un homme dont le frère est en train d'arriver »

b. **raʤul* *-u* *-n* *ʔaxawa: -hu* *qa:dim -u -n*
homme -NOM -Indé frères.NOM.D -ses arrivant -NOM.D -Indé

c. *raʤul* *-u* *-n* *ʔaxawa: -hu* *qa:dim -a:ni*
homme -NOM -Indé frères.NOM.D -ses arrivants -NOM.D
« un homme dont les deux frères sont en train d'arriver»

d. **raʤul* *-u* *-n* *ʔixwat -u -hu* *qa:dim -u -n*
homme -NOM -Indé frères.P -NOM -ses arrivant.S -NOM -Indé

e. *raʤul* *-u* *-n* *ʔixwat -u -hu* *qa:dim -u:na*
homme -NOM -Indé frères.P -NOM -ses arrivant.P -NOM-Indé
« un homme dont les frères sont en train d'arriver »

Ce comportement n'est pas l'apanage du participe actif. Les adjectifs qualificatifs, comme *kari:m* « généreux » dans les exemples suivants, ont la même distribution que le participe actif concernant l'accord (riche/partiel selon l'ordre) :

(3.42) a. *raʤul* *-u* *-n* *kari:mat -u -n* *ʔuxta: -hu*
homme -NOM -Indé généreuse.S -NOM -Indé sœurs.NOM.D -ses
« un homme dont les deux sœurs sont généreuses»

b. **radʒul -u -n kari:m -ata:ni ʔuxta: -hu*
homme -NOM -Indé généreuses -NOM.D sœurs.NOM.D -ses

c. *radʒul -u -n ʔuxta: -hu kari:m -aat:ni*
homme -NOM -Indé sœurs.NOM.D -ses généreuses -NOM.D
« un homme dont les deux sœurs sont généreuses »

L'accord en arabe, qu'il soit verbal ou nominal, donc partiel dans bien des cas. Cette caractéristique se rencontre dans d'autres systèmes également. Si nous prenons le système des nombres, par exemple, nous constatons que l'accord entre le nombre et le SN quantifié tient plus du désaccord que de l'accord. Les nombres 1, 2 et 10 s'accordent en genre avec le SN quantifié, mais les nombres de 3 à 9 se désaccordent avec le SN quantifié en genre, comme le montre le tableau (3.6).

Nombre	SN quantifié	Exemples
Féminin	Masculin	θala:θat-u ridʒa:l -in trois.F -NOM hommes-GÉN « trois hommes »
Masculin	Féminin	θala:θ -u nisa:ʔ -in trois.M -NOM femmes -GÉN « trois femmes »

TABLEAU 3.6: L'accord des nombres entre 3 et 9

Les nombres entre 11 et 19 sont composés de deux parties. Pour les nombres 11 et 12, les deux parties s'accordent en genre avec le SN quantifié, mais les nombres entre 13 et 19 ont un comportement différent : seule la 2e partie du nombre composé s'accorde avec le SN, la 1re partie se désaccorde comme avec les nombres entre 3 et 9, comme le montre le tableau (3.7).

1re partie	2e partie	SN quantifié	Exemples
Féminin	Masculin	Masculin	θala:θata ʔaʃara radʒulan trois.F dix.M homme « treize hommes »
Masculin	Féminin	Féminin	θala:θa aʃrata mraʔatan trois.M dix.F femme « treize femmes »

TABLEAU 3.7: L'accord des nombres entre 13 et 19

Ces exemples démontrent que le système de l'accord en arabe est basé sur plusieurs asymétries. Il n'est donc pas surprenant que l'accord entre le sujet et le verbe soit partiel, voire dans certains cas absent (quand ni le nombre ni le genre ne sont spécifiés).

En somme, l'accord en arabe est plus partiel que ne le laissent présager certains auteurs. L'accord entre le verbe et le sujet, qui ne s'obtient que lorsque le verbe est initial, est un accord partiel

qui n'implique qu'une spécification morphosyntaxique minimale des traits d'accord classiques[34]. Nous avons démontré que cet état de fait n'est pas propre à l'accord entre le sujet et le verbe et qu'il se trouve dans d'autres systèmes en arabe : l'accord nominal et l'accord entre les nombres et les SN quantifiés. Une question reste en suspens cependant : si l'accord entre le sujet et le verbe ne se réalise que partiellement et dans l'ordre dit VSO, peut-on encore parler d'asymétries de l'accord entre le sujet et le verbe ? Nous consacrons la section (3.4.3) à cette question.

3.4.3 Les asymétries de l'accord

Les asymétries de l'accord constituent le problème le plus débattu de la linguistique arabe. Ce problème, connu depuis l'aube de la grammaire arabe, a alimenté plusieurs débats dans le cadre de la théorie du gouvernement et du liage et dans le cadre du programme minimaliste, sans que les différentes approches proposées ne soient capables de le résoudre, et ce, en partie parce que certaines données ne sont pas prises en considération par certains auteurs. Nous essaierons dans cette section de donner une image complète de la question et des différentes approches proposées pour la résoudre. Nous contribuons également au débat en proposant quelques éléments de réflexion en rapport avec notre étude des marqueurs de sujet. Dans la section (3.4.3.1), nous présentons les données qu'il faudrait prendre en considération quand on aborde la question des asymétries de l'accord. Les sections (3.4.3.2) et (3.4.3.3) seront consacrées à donner quelques exemples représentatifs des analyses proposées dans le cadre du programme de recherches chomskyen et des quelques propositions dans le cadre des grammaires à base de contraintes. La section (3.4.3.4) comportera ce que nous pensons être une voie qui mène vers une solution possible de ce problème.

[34] Blachère et Gaudefroy-Demombines (2004) ont essayé d'expliquer ces singularités du système de l'accord en arabe par l'évolution diachronique de cette langue :

> On peut penser (...) qu'à une époque ancienne l'arabe connaissait seulement la construction verbe+sujet, où le verbe restait toujours à la 3e pers. du masc. sing., verbe et sujet formaient alors un tout dans lequel le premier exprimait le procès et le second l'agent réalisant le procès ou le subissant (...) À côté de cette construction, s'en rencontre toutefois une autre sujet+verbe, dans laquelle le second joue le rôle d'un véritable attribut et s'accorde comme tel. Il semble que lorsque l'arabe est libre entièrement de son choix, il préfère la première construction (verbe+sujet) à la seconde (sujet+verbe). (*Ibid.*, p. 299)

3.4.3.1 Les données de l'arabe standard

3.4.3.1.1 L'accord avec des SN

Depuis Sibawayhi au VIIIe siècle et son maître ʔal-xaliːl, les linguistes intéressés par la langue arabe et surtout par la variété standard, se sont posé plus d'une question face à des phrases comme celles en (3.43) à (3.48)[35].

(3.43) a. *ʔal- ʔawlaːd -u daxal -uː* (MP)
les- garçons -NOM entrer.PER -3MP
« Les garçons, ils sont entrés »

b. *daxal -a l- ʔawlaːd -u*
entrer.PER -3M les- garçons -NOM
« Les garçons sont entrés »

c. **ʔal- ʔawlaːd -u daxal -a*
les- garçons -NOM entrer.PER -3MS

d. **daxal -uː l- ʔawlaːd -u*
entrer.PER -3MP les- garçons -NOM

(3.44) a. *ʔal- banaːt -u daxal -na* (FP)
les- filles -NOM entrer.PER -3FP
« Les filles, elles sont entrées »

b. *daxal -at l- ʔbanaːt -u*
entrer.PER -3F les- filles -NOM
« Les filles sont entrées »

c. **ʔal- ʔbanaːt -u daxal -at*
les- filles -NOM entrer.PER -3FS

d. **daxal -na l- ʔbanaːt -u*
entrer.PER -3FP les- filles -NOM

(3.45) a. *ʔal- walad -aːni daxal -aː* (MD)
les- garçons -NOM.D entrer.PER -3MD
« Les deux garçons, ils sont entrés »

b. *daxal -a l- ʔwalad -aːni*
entrer.PER -3M les- garçons -NOM.D
« Les garçons sont entrés »

[35] L'agrammaticalité des phrases avec un SN postverbal s'obtient dans un contexte non marqué, sans recourir à la dislocation.

c. *ʔal- ʔwalad -a:ni daxal -a
les- garçons -NOM.D entrer.PER -3MS

d. *daxal -a: l- ʔwalad -a:ni
entrer.PER -3MP les- garçons -NOM.D

(3.46) a. ʔal- bint -a:ni daxal -ata: (FD)
les- filles -NOM.D entrer.PER -3FD
« Les deux filles, elles sont entrées »

b. daxal -at l- ʔbint -a:ni
entrer.PER -3F les- filles -NOM.D
« Les deux filles sont entrées »

c. *ʔal- ʔbint -a:ni daxal -at
les- filles -NOM.D entrer.PER -3FS

d. *daxal -ata: l- ʔbint -a:ni entrer.PER -3FP
les- filles -NOM.D

(3.47) a. ʔal- walad -u daxal -a (MS)
le- garçon -NOM entrer.PER -3MS
« Le garçon, il est entré »

b. daxal -a l- walad -u
entrer.PER -3M le- garçon -NOM
« Le garçon est entré »

(3.48) a. ʔal- bint -u daxal -at (FS)
la- fille -NOM entrer.PER -3FS
« La fille, elle est entrée »

b. daxal -at l- bint -u
entrer.PER -3F la- fille -NOM
« La fille est entrée »

Nous avons varié ces exemples selon deux paramètres :

1. L'ordre des mots : un SN nominatif est tantôt préverbal, tantôt postverbal.
2. Les traits de genre et de nombre de ce SN . les combinaisons possibles sont MP, FP, MD, FD, MS et FS.

Cette variation entraîne des changements dans les marqueurs de sujet, qui varient en fonction des traits du SN et en fonction de sa position. En effet, dans l'ordre SN+V, et quelles que soient les valeurs des traits nombre et genre du SN, le marqueur de sujet porte les mêmes traits que

ce SN (nous parlons ici d'accord riche). C'est le cas dans les exemples (3.43a), (3.44a), (3.45a) et (3.46a). Toutefois, dans l'ordre V+SN, le marqueur de sujet ne semble pas spécifié pour le nombre (nous parlons ici d'accord partiel). Les traits de genre et de personne ont par contre les mêmes valeurs que celles sur le SN en question : (3.43b), (3.44b), (3.45b) et (3.46b). Les phrases en (3.43c), (3.44c), (3.45c) et (3.46c), dans lesquelles le SN préverbal ne dicte pas tous les traits d'accord sur le marqueur de sujet, sont agrammaticales. Les phrases en (3.43d), (3.44d), (3.45d) et (3.46d) sont agrammaticales parce que le SN postverbal dicte ces traits. Dans les cas où le SN est singulier, dans les exemples en (3.47) et (3.48), nous pouvons généraliser en considérant que le marqueur de sujet n'exprime pas le nombre dans l'ordre V+SN et qu'il l'exprime dans l'ordre SN+V.

3.4.3.1.2 L'accord avec des pronoms indépendants

Cette asymétrie entre l'accord partiel et l'accord riche ne concerne que la distribution des SN. Le remplacement de ces SN par des pronoms indépendants nominatifs ne donne pas lieu aux mêmes phénomènes, comme le montrent les exemples suivants :

(3.49) a. *hum* *daxal* *-u:* (MP)
eux entrer.PER -3MP
« Eux, ils sont entrés »

b. **daxal* *-a* *hum*
entrer.PER -3M eux

c. **hum* *daxal* *-a*
eux entrer.PER -3MS

d. *daxal* *-u:* *hum*
entrer.PER -3MP eux
« Ils sont entrés, eux »

(3.50) a. *hunna* *daxal* *-na* (FP)
elles entrer.PER -3FP
« Elles, elles sont entrées »

b. **daxal* *-at* *hunna*
entrer.PER -3F elles

c. **hunna* *daxal* *-at*
elles entrer.PER -3FS

d. *daxal* ***-na*** *hunna*
entrer.PER -3FP elles
« Elles sont entrées, elles »

(3.51) a. *huma:* *daxal* ***-a:*** (MD)
eux.D entrer.PER -3MD
« Eux deux, ils sont entrés »

b. **daxal* ***-a*** *huma:*
entrer.PER -3M eux.D

c. **huma:* *daxal* ***-a***
eux.D entrer.PER -3MS

d. *daxal* ***-a:*** *huma:*
entrer.PER -3MP eux.D
« Ils sont entrés, eux deux »

(3.52) a. *huma:* *daxal* ***-ata:*** (FD)
elles.D entrer.PER -3FD
« Elles deux, elles sont entrées »

b. **daxal* ***-at*** *huma:*
entrer.PER -3F elles.D

c. **huma:* *daxal* ***-at***
elles.D entrer.PER -3FS

d. *daxal* ***-ata:*** *huma:*
entrer.PER -3FP elles.D
« Elles sont entrées, elles deux»

(3.53) a. *huwa* *daxal* ***-a***
lui entrer.PER -3MS (MS)
« Lui, il est entré »

b. *daxal* ***-a*** *huwa*
entrer.PER -3M lui
« Il est entré, lui »

(3.54) a. *hija* *daxal* ***-at*** (FS)
elle entrer.PER -3FS
« Elle, elle est entrée »

b. *daxal* ***-at*** *hija*
entrer.PER -3F elle
« Elle est entrée, elle »

Dans l'ordre pronom indépendant + V, le marqueur de sujet montre un accord riche avec le pronom indépendant préverbal, tout comme si c'était un SN dans les phrases (3.49a), (3.50a), (3.51a), (3.52a), (3.53a) et (3.54a). Cependant, il exhibe la même sorte d'accord dans l'ordre V+pronom indépendant, c'est-à-dire un accord riche dans les exemples en (3.49d), (3.50d), (3.51d), (3.52d), (3.53b) et (3.54b). Ce comportement instaure une asymétrie entre les SN et les pronoms indépendants.

3.4.3.1.3 Questions de recherche

L'accord entre le SN/pronom indépendant NOM préverbal/postverbal et le verbe est caractérisé par plusieurs asymétries en arabe standard. Cet accord peut être partiel ou riche selon deux paramètres principaux : l'ordre des mots (verbe initial vs verbe non initial) et la nature du constituant nominatif associé au marqueur de sujet. Selon le premier paramètre, un SN préverbal déclenche un accord riche alors qu'un SN postverbal déclenche un accord partiel dans les contextes non marqués. Selon le deuxième paramètre, si le constituant en question est un SN, son ordre par rapport au verbe déclenche l'asymétrie accord riche / accord partiel. Mais si ce constituant est un pronom indépendant, aucune asymétrie de ce genre n'est déclenchée et l'accord est riche dans les deux positions. Ce comportement pose deux problèmes : pourquoi l'ordre des mots a-t-il une incidence sur l'accord ? Et pourquoi le pronom indépendant ne se comporte pas de la même manière qu'un SN dans ces structures ?

Pour répondre à ces questions, plusieurs approches étaient proposées. Nous en présenterons ici quelques exemples représentatifs que nous groupons en deux catégories : des approches dans le cadre génératif chomskyen et des approches dans le cadre des grammaires à base de contraintes. La première catégorie regroupe des approches essentiellement syntaxiques préminimalistes dans lesquelles les auteurs tentent de résoudre les problèmes posés par les asymétries de l'accord en se basant sur les représentations syntaxiques, les niveaux d'analyse et les relations configurationnelles (le gouvernement vs la relation entre le spécifieur et la tête). Les tenants des approches minimalistes proposent des analyses qui font intervenir les interfaces, notamment l'interface FP (Forme Phonologique). Les auteurs qui ont travaillé dans le cadre de la grammaire HPSG ont proposé des analyses essentiellement lexicalistes. Lors de notre présentation de ces approches, nous serons amené à introduire de nouvelles données qui proviennent des dialectes, ce qui nous permettra dans un deuxième temps de discuter d'autres données et de proposer des éléments de réponse.

3.4.3.2 Les analyses syntaxiques

Les premières analyses purement syntaxiques des phénomènes de l'accord en arabe sont proposées dans le cadre de la théorie du gouvernement et du liage (ou GB). Dans le cadre de ces analyses, l'on se posait la question de savoir dans quelle position syntaxique apparaissent les SN préverbal et postverbal et quelle est la nature de la relation qu'entretiennent ces SN avec le verbe. Ainsi, pour Aoun *et al.* (1994, 1999) et Mohammad (1990), l'accord se réalise dans une relation entre le spécifieur et la tête (selon le schéma $\bar{X}$), alors qu'il se réalise dans une relation de gouvernement pour Bahloul et Harbert (1992) et Harbert et Bahloul (2002). Cela a donné lieu à plusieurs hypothèses dont nous n'exposerons que les plus représentatives, à savoir l'hypothèse explétive de Mohammad (1990) dans le paragraphe (3.4.3.2.1), l'hypothèse de la perte d'accord de Aoun *et al.* (1994, 1999) dans le paragraphe (3.4.3.2.2) et l'hypothèse du gouvernement de Bahloul et Harbert (1992) et Harbert et Bahloul (2002) dans le paragraphe (3.4.3.2.3). Nous exposons également une analyse minimaliste, celle de Benmamoun et Lorimor (2006), dans le paragraphe (3.4.3.2.4).

3.4.3.2.1 L'hypothèse explétive

L'hypothèse défendue par Mohammad (1990) peut être énoncée de la manière suivante : l'ordre de base est SVO[36] et dans cet ordre le SN préverbal en (3.55a)[37] est réalisé dans la position du spécifieur de IP et entretient donc une relation Spec-tête avec la tête I (sous laquelle le verbe est réalisé), d'où l'accord riche observé dans ce contexte. Quand le SN est postverbal, en (3.55b), la position [Spec, IP] est remplie par un pronom explétif nul. L'accord se fait donc entre le verbe et cet explétif nul, ce qui donne lieu, selon l'auteur, à un accord partiel et à une phrase avec deux sujets : un sujet réel et un sujet explétif.

(3.55) a. SV : $[_{IP}$ Subj$_j$ V$_i$+I [VP t$_j$ t$_i$...]]
b. VS : $[_{IP}$ *pro*EXPL V$_i$+I $[_{VP}$ Subj t$_i$...]]

Pour justifier cette analyse, l'auteur souligne que cet explétif nul est lexicalisé dans certains contextes et apparaît donc en même temps que le SN postverbal considéré comme sujet, comme en (3.56)[38] où l'explétif *-hu* attaché au complémenteur *?anna* apparaît en position préverbale et

[36] L'auteur reconnaît que l'ordre VSO est l'ordre pragmatiquement neutre (non marqué), mais pense que l'ordre SVO est l'ordre de base.
[37] Ces représentations proviennent de Soltan (2004).
[38] (Mohammad, 1990, p. 110).

le SN sujet *r-raʤul-u* apparaît en position postverbale :

(3.56) *ʔiddaʕa:* *ʔaħmad* *-u* *ʔanna* *-hu* *ʤa:ʔ* *-a* *r-* *raʤul*
prétendre.PER.3M Ahmed -NOM que -il venir.PER -3M l'- homme
-u
-NOM
« Ahmed a prétendu que l'homme est arrivé »

Pour que deux sujets puissent être présents dans la même phrase, l'un d'eux ne doit pas avoir de rôle thématique, en conformité au *critère thêta*, et c'est justement le cas de l'explétif, selon l'auteur.

Toutefois, pour que cette analyse soit acceptable, il faudrait que les explétifs soient toujours sous-spécifiés pour le nombre en arabe, ce qui n'est pas le cas, comme le souligne Fassi Fehri (1993). En fait, les explétifs peuvent avoir plusieurs valeurs pour les traits genre et nombre selon le contexte, comme le montrent les exemples suivants :

(3.57) a. *hum(u)* *l-* *ʤun:d* *-u* (MP)
eux les- soldats -NOM
« Ce sont les soldats »

b. *hunna* *n-* *nisa:ʔ* *-u* (FP)
elles les- femmes -NOM
« Ce sont les femmes »

Par ailleurs, cette analyse ne rend pas compte de l'asymétrie entre les SN et les pronoms indépendants postverbaux. Les deux constituants sont censés être réalisés dans la position [Spec, VP] dans l'ordre VSO, selon l'auteur. Or, l'accord dans cet ordre se ferait avec l'explétif préverbal, qui est le même que le constituant postverbal soit un SN ou un pronom indépendant. On s'attendrait donc à un accord partiel avec les deux constituants, ce qui n'est pas le cas.

L'hypothèse explétive, en plus d'être une solution peu élégante (l'existence de deux sujets dans une même phrase), ne prédit pas toutes les données déjà mentionnées, mais également des données supplémentaires qui proviennent des dialectes arabes contemporains. C'est pour rendre compte de cette variation dialectale que Aoun *et al.* (1994) ont proposé une analyse syntaxique différente de celle que défend Mohammad (1990).

3.4.3.2.2 L'accord perdu

Aoun *et al.* (1994) soutiennent, tout comme Mohammad (1990), que l'accord est une relation structurale entre la tête et le spécifieur. Cependant, l'approche adoptée par ces auteurs se veut plus globale, et ce, en embrassant un plus grand nombre de données, et notamment des données dialectales. Les structures qui intéressent les auteurs sont les deux suivantes[39] :

(3.58) a. SN V
b. V SN

Les variétés étudiées sont l'arabe libanais (désormais AL), l'arabe marocain (désormais AM) et l'arabe standard (désormais AS). Le tableau (3.8) récapitule ces données.

	SN V	V SN
AS	Accord riche	Accord partiel
AM et AL	Accord riche	Accord riche

TABLEAU 3.8: Les asymétries et la variation dialectale

Nous avons vu des exemples qui illustrent ces structures pour AS dans la section (3.4.3.1). En ce qui a trait aux dialectes, les exemples suivants illustrent les phénomènes de l'accord avec un SN préverbal (3.59a) et un SN postverbal (3.59b). Ces données sont de l'arabe tunisien (désormais AT), une variété dialectale qui exhibe le même comportement que les deux autres variétés discutées par les auteurs[40].

(3.59) a. *l-* *wla:d* *xurʁ* *-u:* (SN V)
les- garçons sortir.PER -3MP
« Les garçons sont sortis »

b. *xurʁ* *-u:* *l-* *wla:d* (V SN)
sortir.PER -3MP les- garçons
« Les garçons sont sortis »

Aoun *et al.* (1994) fournissent les réponses suivantes aux questions qui concernent les asymétries de l'accord : les traits de l'accord sont légitimés dans une relation entre la tête et le spécifieur. Cette légitimation s'obtient aux différents stades de la dérivation. Ainsi, en AS, l'accord en genre dans l'ordre V+SN est légitimé parce que le sujet et le verbe étaient en relation de spécifieur à

[39] (Aoun *et al.*, 1994, p. 195). Les auteurs traitent également la question de l'accord du premier conjoint dans les structures [SN1 et SN2 V] et [V SN1 et SN2]. Nous ne traitons pas cette question ici.
[40] Voir Mahfoudhi (2002).

tête à un certain stade de la dérivation avant le déplacement du verbe vers une position de tête par-dessus le sujet[41]. Dans ce même ordre des mots, l'accord en nombre n'est pas légitimé parce que cette information est perdue en cours de dérivation, suite au déplacement de la tête verbale en dehors d'une relation de spécifieur à tête (la relation qui légitime les traits d'accord selon les auteurs). Cette explication est censée être valide pour les dialectes également : en AL et en AM, le verbe ne subit pas le déplacement supplémentaire qu'il subit en AS, d'où la préservation de tous les traits d'accord, y compris le trait nombre, et l'accord riche observé en (3.59b). L'accord riche dans l'ordre SN+V découle du même principe puisque le verbe et le sujet se trouvent en relation de spécifieur à tête dans cette structure.

Le problème de cette approche, comme le soulignent Munn (1999) et Soltan (2004), c'est que la perte de l'accord en nombre suite au déplacement du verbe est *ad hoc* et ne découle d'aucun principe général, surtout si on prend en compte deux restrictions à cette « perte » : la première intervient avec un pronom indépendant postverbal, où aucune « perte » de l'accord en nombre ne se produit malgré le déplacement présumé du verbe, et la deuxième est dans les dialectes qui accordent le verbe et le SN sujet de la même manière que l'AS accorde le verbe et le pronom indépendant postverbal.

Cela dit, les deux analyses de Mohammad (1990) et de Aoun *et al.* (1994) ont en commun une vision syntaxique de l'accord qui le relie à une relation entre la tête verbale et son spécifieur. Il suffit que cette relation soit établie à un niveau ou à un autre de la dérivation syntaxique pour légitimer les traits d'accord observés en surface. Nous avons vu qu'une telle approche ne saurait rendre compte de toutes les données reliées aux asymétries de l'accord, notamment l'accord avec un pronom indépendant nominatif postverbal et la variation dialectale. D'autres analyses syntaxiques ont essayé d'éviter ces problèmes en proposant le gouvernement comme étant la relation qui légitime l'accord. Les auteurs qui ont mis l'accent sur cette relation sont surtout Bahloul et Harbert (1992) et Harbert et Bahloul (2002).

3.4.3.2.3 Accord et gouvernement

Bahloul et Harbert (1992) et Harbert et Bahloul (2002) défendent l'hypothèse suivante : les asymétries de l'accord attribuables à l'ordre de surface s'expliquent par le fait que l'accord se réalise sous des conditions syntaxiques différentes dans les configurations dites VSO et SVO.

[41] Cette position est la tête F selon les auteurs (tête de FP). Notons que ce choix de tête indique l'embarras face à cette position vers laquelle se déplace le verbe en dehors de IP.

Dans l'ordre dit SVO, l'accord encode une relation entre la tête (le verbe) et son spécifieur (le sujet), alors que dans l'ordre dit VSO, l'accord est légitimé dans une relation de gouvernement entre ces éléments. De plus, la différence dans la morphologie de l'accord (entre un accord partiel et un accord riche) se produit parce que les ensembles de traits accessibles sous ces deux types d'accord ne sont pas les mêmes.

Pour expliquer cette différence, les auteurs ont recours à la théorie de Ritter (1997) sur la projection d'une tête fonctionnelle Num (le nombre) en NumP. En adoptant cette théorie, les deux auteurs défendent l'idée selon laquelle la différence entre les deux traits genre et nombre est une différence entre un trait inhérent aux têtes nominales (le genre) et un trait acquis lors d'un déplacement dans un nœud fonctionnel (le trait nombre). Lors de la dérivation, le trait de genre, en tant que trait inhérent aux noms, est associé à la strate inférieure de la projection, alors que le trait de nombre est associé à la strate supérieure. Seuls les traits de la strate inférieure d'un système de projection sont visibles à l'accord sous gouvernement, alors que tous les traits de genre, de nombre et de personne sont visibles à l'accord dans la configuration Spécifieur-Tête. Les auteurs affirment :

> Gender features originate on lexical Ns, hence on the Bases of Projection Systems. They are thus visible for agreement under government. Number features are features of the functional head Num. They are accordingly not features of the Base node in Projection Systems terminating in lexical Ns, and are invisible to such agreement. (Harbert et Bahloul, 2002, p. 47)

Cependant, cette théorie n'explique pas encore l'asymétrie entre les SN et les pronoms indépendants nominatifs postverbaux. Pour y parvenir, et pour expliquer cette particularité qu'a le pronom indépendant (sujet, selon les auteurs) à déclencher un accord riche, Harbert et Bahloul (2002) soutiennent que le trait nombre est inhérent aux pronoms alors qu'il est acquis en cours de dérivation en ce qui concerne les SN. Comme tout trait inhérent, ce trait est donc accessible dans une configuration de gouvernement, d'où l'accord en genre, en nombre et en personne entre le pronom indépendant et le verbe qui le gouverne.

Qu'en est-il des dialectes discutés dans Aoun *et al.* (1994) ? Rappelons qu'en AM, en AL et en AT, l'interdiction de l'accord en nombre avec le SN sujet postverbal est levée. Peut-on affirmer que le trait nombre est inhérent aux têtes nominales dans ces dialectes alors qu'il est acquis en cours de dérivation en arabe standard ? Une telle solution aurait besoin d'arguments forts pour la justifier. Aussi, les auteurs se contentent-ils de signaler cette différence entre les variétés de

l'arabe sans essayer de l'expliquer.

Les trois analyses syntaxiques présentées ici ne sont que des exemples des approches strictement syntaxiques adoptées pour résoudre les nombreux problèmes soulevés par les asymétries de l'accord en arabe. Que ces analyses choisissent de considérer l'accord comme une relation qui s'obtient dans une configuration de gouvernement ou dans une relation de la tête avec son spécifieur, le résultat est presque le même : on n'arrive pas à embrasser toutes les données. Il faut dire que ces données sont complexes, surtout quand on prend en compte les variétés dialectales. Les approches minimalistes ont tenté de remédier à ces problèmes en proposant de prendre en considération les interfaces (l'interface FP, surtout). Nous présentons un exemple de ces analyses, celle de Benmamoun et Lorimor (2006), dans la section suivante.

3.4.3.2.4 L'accord encore une fois perdu

Selon Benmamoun (2000a) et Benmamoun et Lorimor (2006), les items lexicaux entrent dans la dérivation syntaxique avec tous leurs traits (y compris le trait de nombre) et l'opération de fusion (Merger) se produit dans le composant morphologique postsyntaxique pour fusionner les membres de la séquence VS (verbe-sujet). Ainsi, la fusion du verbe et du sujet à l'interface FP (forme phonologique), cette interface où les traits morphosyntaxiques sont épelés selon la théorie de la morphologie distribuée, donne lieu à l'épellation du trait de nombre déjà contenu comme trait inhérent au SN : « the merger of the verb and the subject amounts to spelling-out the number feature, making the presence of the number affix redundant. » (Benmamoun, 2000a, p. 11). Cette hypothèse n'est pas sans nous rappeler l'approche de Aoun *et al.* (1994), selon laquelle l'accord en nombre est perdu en cours de dérivation. Cependant, si ce trait est perdu à cause d'un déplacement supplémentaire de la tête verbale en dehors du IP selon Aoun *et al.* (1994), il est perdu lors de la fusion du verbe avec le SN sujet (portant lui-même ce trait) selon Benmamoun. De plus, cette perte se produit dans une composante morphologique postsyntaxique et non dans la composante syntaxique, comme le soutiennent Aoun *et al.* (1994).

Cette approche n'explique cependant pas pourquoi il y aurait fusion entre le sujet et le verbe en arabe standard seulement quand le sujet est postverbal. Rappelons que cette fusion est censée expliquer l'accord partiel dans l'ordre VSO. Qu'est-ce qui empêcherait le sujet et le verbe de fusionner dans l'ordre dit SVO ? En fait, l'auteur ne fournit aucune théorie spécifique de la fusion à l'interface FP. Les SN préverbaux sont adjacents au verbe eux également, il n'est donc pas clair

ce qui devrait empêcher la fusion et la suppression du trait de nombre dans ce contexte, comme le signalent Ackema et Neeleman (2003).

Admettons qu'un principe syntaxique motive la fusion dans l'ordre VSO seulement, alors si le trait de nombre est « perdu » dans l'ordre VSO parce qu'il est devenu redondant à l'interface FP, comment explique-t-on le fait que cette fusion du sujet et du verbe ne produise pas ce même effet dans les variétés dialectales ? Pourquoi le trait de nombre semble être redondant en AM, en AL et en AT et non en AS ? Benmamoun répond à ces questions en stipulant tout simplement que l'opération de fusion est propre à l'AS et absente des dialectes :

> Moroccan Arabic has one single option to spell-out number agreement on the verb, while Standard Arabic has two options, affixation and merger between the verb and the subject. (Benmamoun, 2000a, p. 133)

Nous pensons que cette affirmation manque d'éléments de conviction et d'arguments pour l'appuyer. Notons, par ailleurs, la place du trait de nombre dans cette analyse. Ce trait est en fait censé être un trait inhérent à la tête nominale, suivant Chomsky (1995) Cette idée est difficilement conciliable avec les données qui impliquent des pronoms indépendants nominatifs en position postverbale. Peut-on en fait affirmer que ce trait ne fait pas partie des traits inhérents à ces pronoms, contre toute apparence ?

Les différentes approches syntaxiques qui ont tenté de résoudre le problème des asymétries de l'accord, qu'elles soient dans le cadre de GB ou dans le cadre du programme minimaliste, apportent des éléments de réponse qu'il faudrait considérer. Elles ne parviennent cependant pas à englober toutes les données pertinentes sur la question.

3.4.3.3 Les analyses lexicalistes

Les analyses lexicalistes de la question des asymétries de l'accord optent pour des méthodes d'analyse différentes que celles employées dans le cadre syntaxique strict des approches déjà présentées. Elles apportent également une contribution notable à ce débat et méritent sans doute qu'on s'y attarde.

Les seules analyses lexicalistes dont nous connaissons l'existence sont les deux que nous pré-

sentons ici : l'analyse de Moosally (1994) et celle de Beller (2006).

3.4.3.3.1 L'ordre VSO est dérivé

Moosally (1994) est d'avis que l'ordre non marqué en arabe est l'ordre VSO. Pour concilier cette analyse avec les données empiriques (le fait que le verbe ne puisse être initial dans certains cas) et en restant fidèle au type d'approche admise dans Pollard et Sag (1994), elle propose d'employer le schéma de dominance immédiate 3, déjà présenté en (1.16) et répété ici en (3.60), pour décrire les structures VSO en arabe.

(3.60)
$$\begin{bmatrix} \textit{str-tête} & \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{VAL} & \langle\rangle \end{bmatrix} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} & \begin{bmatrix} \textit{mot} & \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{VAL} \mid \text{COMPS} & \langle \boxed{2} \ldots \boxed{n} \rangle \\ \text{TÊTE} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{FILLES-COMPS} & \langle \boxed{2} \ldots \boxed{n} \rangle \end{bmatrix} \end{bmatrix}$$

Ce schéma a été proposé pour décrire les structures en anglais où il y a inversion du sujet et de l'auxiliaire. Quand l'entrée lexicale d'un verbe auxiliaire comporte le trait [+INV], cette spécification déclenche l'inversion. De même, Moosally propose que chaque verbe[42] en arabe possède deux entrées lexicales : une entrée avec le trait [-INV] pour les structures dites SVO et l'autre avec le trait [+INV] pour les structures VSO. Quand l'entrée lexicale comporte le trait [-INV], l'auteure propose qu'elle soit sujette au schéma DI1, alors que les entrées comportant le trait [+INV] sont sujettes au schéma DI3. Pour illustrer cette analyse, nous présentons les deux entrées lexicales du verbe *darasa* « il a étudié », que l'auteure propose[43].

(3.61) L'entrée lexicale de *darasa* (1)

[42] Théoriquement, chaque verbe en arabe peut apparaître dans l'un ou l'autre des deux ordres dits VSO et SVO.

[43] SN : une abréviation de tous les sous-types de *N-form* incluant *ppro* (pronom personnel), *npro* (non pronominal) et *apro* (pronom anaphorique).

$$\begin{bmatrix} \text{PHON} & \langle \text{DARASA} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{verbe} \\ \text{INV} \quad + \end{bmatrix} \\ \text{VAL} & \left\langle \begin{matrix} \textit{ppro}\ \boxed{1}\begin{bmatrix} \text{3MS} \end{bmatrix} \\ \textit{npro}\ \boxed{1}\begin{bmatrix} \text{M} \end{bmatrix} \end{matrix} \right\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{psoa} & \\ \text{RELN} & \textit{étudier} \\ \text{ÉTUDIANT} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(3.62) L'entrée lexicale de *darasa* (2)

$$\begin{bmatrix} \text{PHON} & \langle \text{DARASA} \rangle \\ \text{SYNSEM} & \begin{bmatrix} \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{verbe} \\ \text{INV} \quad - \end{bmatrix} \\ \text{VAL} & \left\langle \text{SN}\ \boxed{1}\begin{bmatrix} \text{3MS} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{psoa} & \\ \text{RELN} & \textit{étudier} \\ \text{ÉTUDIANT} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Ces deux entrées lexicales sont censées refléter deux propriétés empiriques différentes :

- Le fait que les *sujets* préposés déclenchent un accord riche est reflété dans la liste de sous-catégorisation du verbe, dont le trait de tête est [-INV] (cette liste ne comportant pas les compléments du verbe dans cette représentation).
- Le fait que l'accord dans l'ordre VSO soit partiel est capté par les traits d'accord du sujet des verbes dont les traits de tête comportent la spécification [+INV]. Le sujet peut ainsi être un *ppro* (incluant ainsi les pronoms nuls) ou un *npro* (un SN plein). Dans le premier cas, l'accord est riche, alors qu'il est partiel dans le dernier (seul le genre est accordé).

Cette analyse repose sur des présupposés que nous ne partageons pas avec l'auteure. Parmi ces présupposés, le statut morphosyntaxique des SN préverbaux et des pronoms indépendants pose plus d'un problème. En outre, traiter l'ordre VSO comme étant un « dérivé » de l'ordre SVO

ne repose pas sur une analyse de la structure syntagmatique de l'arabe. Si l'inversion est traitée par l'ajout de l'attribut INV dans le cas des auxiliaires en anglais c'est qu'il y a justement une inversion. L'ordre VSO en arabe, au contraire, ne comporte aucune inversion.

3.4.3.3.2 Le rôle de la structure argumentale

Beller (2006) utilise les outils théoriques et formels fournis par Sag *et al.* (2003). Il propose une contrainte additionnelle sur l'accord qui relie l'accord entre le sujet et le verbe à la structure argumentale et teste les implications de cette prise de position sur des structures de l'anglais. Il propose également une nouvelle règle lexicale pour modéliser l'alternance dans l'ordre des mots.

L'auteur est d'avis que l'alternance VSO/SVO est une alternance lexicale et non syntaxique et il la modélise, par conséquent, par une règle lexicale dont l'intrant est un verbe avec des valeurs pour les attributs SUJ et SPEC et dont l'extrant est un verbe dont le sujet est placé comme premier membre de la liste COMPS. On remarquera dans cette règle que la liste COMPS est sous-spécifiée pour permettre de modéliser les verbes quelle que soit leur valence (intransitifs, transitifs, etc.). Par ailleurs, la valeur de l'attribut NBRE de la tête dans l'extrant est *sing*(ulier). Cette valeur est censée être la valeur par défaut quand aucun accord ne se réalise avec le sujet, selon l'auteur[44]

(3.63) La règle lexicale de VSO

[44] Cette règle est dérivationnelle par opposition aux règles lexicales flexionnelles. Cette distinction est faite par Sag *et al.* (2003), qui insistent sur le fait que ces règles relient en réalité deux séquences lexicales et ne créent pas une séquence à partir d'une autre. Dans la structure de traits qui décrit ces règles, deux attributs sont appropriés INTRANT et EXTRANT, dont les valeurs sont des listes de lexèmes. Les lexèmes en question dans la règle présentée ici sont verbaux (de type *lexème-verbal*). Dans cette représentation, la lettre encadrée A renvoie à une liste. Les chiffres arabes encadrés revoient à des items.

$$\begin{bmatrix} \textit{règle-dérivationnelle} & \\ \text{INTRANT} & \left\langle \boxed{0}, \begin{bmatrix} \textit{lexème-verbal} & \\ \text{SYN} & \begin{bmatrix} \text{VAL} & \begin{bmatrix} \text{SPEC} & \left\langle \boxed{1} \right\rangle \end{bmatrix} \end{bmatrix} \\ \text{STR-ARG} & \boxed{A} \end{bmatrix} \right\rangle \\ \text{EXTRANT} & \left\langle \boxed{0}, \begin{bmatrix} \textit{lexème-verbal} & \\ \text{SYN} & \begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{NBRE} & \textit{sing} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SPEC} & \langle\rangle \\ \text{COMPS} & \left\langle \boxed{1} \ldots \right\rangle \end{bmatrix} \\ \text{STR-ARG} & \boxed{A} \end{bmatrix} \right\rangle \end{bmatrix}$$

Pour modéliser l'accord qu'il appelle *appauvri* dans l'ordre VSO, il propose une contrainte (ASAP, pour Argument Structure Agreement Principle) sur le type *lexème-verbal* qui contraigne un mot à s'accorder avec le premier membre de sa structure argumentale, qui est le *sujet* dans les deux ordres VSO et SVO : « ASAP tags the AGR features PER and GEND of a lexeme with the PER and GEND features of the first member of that lexeme's ARG-ST ⟨STR-ARG⟩ list ». (Beller, 2006, p. 22)[45] :

(3.64) ASAP
$$\begin{bmatrix} \textit{lexème-verbal} & \\ \text{SYN} & \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{PG} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{STR-ARG} & \left\langle \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{PG} & \boxed{1} \end{bmatrix} \end{bmatrix}, \ldots \right\rangle \end{bmatrix}$$

L'accord en nombre n'est pas garanti par cette contrainte. Ce rôle est donc légué par l'auteur à la contrainte SHAC[46] (Specifier Head Agreement Constraint) proposée par Sag *et al.* (2003) pour modéliser l'accord entre le sujet et le verbe dans l'ordre dit SVO.

[45] Dans cette contrainte, l'attribut PG modélise un sous-type de *accord* qui englobe les deux traits personne et genre.

[46] L'auteur n'est cependant pas sûr que le *partage de pouvoirs* entre les deux contraintes soit aussi direct et déclare :

> The ASAP must constrain the PG features, person and gender, and the SHAC must minimally constrain number agreement. It is hard to discern whether the SHAC only constrains number or constrains all of the AGR features. (Beller, 2006, p. 25)

(3.65) SHAC
$$\begin{bmatrix} \text{TÊTE} \begin{bmatrix} \text{ACCORD} & \boxed{1} \end{bmatrix} \\ \text{VAL} \begin{bmatrix} \text{SPEC} & \left\langle \begin{bmatrix} \text{ACCORD} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Cette modélisation des asymétries de l'accord de l'arabe ne rend pas compte de toutes les données, puisque l'auteur ne soulève pas les problèmes posés par les pronoms indépendants. De plus, en niant toute différence pragmatique (ou de sens) entre les structures dites SVO et les structures dites VSO[47], et en considérant que l'ordre VSO est d'une manière ou d'une autre dérivé de l'ordre SVO, l'auteur s'expose aux mêmes problèmes qu'affronte l'analyse proposée par Moosally (1994).

Les analyses syntaxiques et lexicalistes que nous avons présentées ici partagent certains problèmes dont deux nous paraissent particulièrement importants : le traitement de l'ordre dit SVO et l'aspect partiel des données couvertes. En ce qui a trait au premier point, ces analyses s'accordent presque toutes à considérer que les structures à verbe initial en arabe sont dérivées d'un ordre universel où le *sujet* précède le verbe. Cette dérivation, qu'elle soit syntaxique ou lexicale, cache une différence importante entre les deux sortes de structures, qui est une différence dans l'interprétation et dans la fonction pragmatiques.

Le deuxième problème concerne la couverture empirique de ces analyses. La majorité de ces dernières utilisent en effet les mêmes exemples et ne traitent pas soit les asymétries impliquant des pronoms, soit les données dialectales, soit les deux.

La section suivante comporte quelques éléments qui contribuent à la question des asymétries de l'accord en essayant d'éliminer les défauts des analyses précédentes.

3.4.3.4 Éléments pour une solution

Nous proposons une analyse qui s'inspire, certes, de certaines des approches précédentes et surtout concernant la position du SN préverbal, le statut du marqueur de sujet et la nature de

[47] L'auteur s'appuie sur les travaux de deux auteurs qui ne signalent pas, selon lui, cette différence, à savoir Benmamoun (2000a) et Mohammad (2000).

la relation d'accord entre le sujet et le verbe en arabe. Nous intégrons les données des dialectes en plus des données de l'arabe standard, les données de l'accord nominal en plus des données de l'accord verbal et les propriétés des marqueurs d'arguments en plus des propriétés des marqueurs d'accord. Nous détaillons notre approche en plusieurs points en nous basant sur les idées déjà avancées dans les sections précédentes et dans le deuxième chapitre :

1. Le SN préverbal est un topique disloqué : dans l'ordre dit SVO, le « S » est en réalité un topique disloqué (un non-argument) et non un sujet (un argument du verbe).
2. Le marqueur de sujet dans l'ordre dit SVO est un pronom incorporé : en nous appuyant sur les critères de Bresnan et Mchombo (1987), nous avons démontré que l'accord constaté entre le SN préverbal et le marqueur de sujet est un accord anaphorique et non un accord grammatical. Étant un accord anaphorique, il implique un topique et un pronom incorporé et il ne s'agit donc pas d'un accord entre le sujet et le verbe.
3. Le marqueur de sujet est un marqueur d'accord dans l'ordre dit VSO (avec un SN postverbal NOM que nous considérons le sujet) : l'accord dans ce cas est grammatical. C'est la seule relation d'accord entre le sujet et le verbe en arabe standard. Le fait que cet accord est partiel découle des propriétés générales du système de l'accord dans cette langue.
4. Le pronom indépendant NOM dans l'ordre dit SVO et dans l'ordre dit VSO est disloqué : nous avons décrit la distribution de ces pronoms dans le deuxième chapitre et nous avons conclu que leur apparition dans la phrase verbale est toujours marquée et qu'ils sont disloqués à gauche ou à droite. L'argument dans ce contexte est le marqueur de sujet incorporé au verbe et l'accord observé ici est un accord anaphorique et non grammatical.

Il nous est donc possible d'avancer l'hypothèse suivante : ce que l'on appelle les asymétries de l'accord n'est en réalité que l'expression de deux sortes d'accord au sein du même système des marqueurs d'arguments : un accord grammatical et un accord anaphorique. L'asymétrie entre les SN et les pronoms indépendants peut être décrite de manière différente de celle pour laquelle optent plusieurs auteurs. En fait, on ne peut comparer les SN postverbaux aux pronoms indépendants postverbaux. Ces derniers sont toujours disloqués alors que les premiers peuvent être des arguments quand les marqueurs de sujet marquent l'accord verbal et disloqués quand ces derniers sont des arguments pronominaux.

Nous pensons que cette hypothèse rend compte de toutes les données déjà discutées. Elle reste cependant incapable d'expliquer les différences entre l'arabe standard et l'arabe dialectal. Rappelons que la différence principale entre ces deux variétés est la suivante : dans les dialectes, le

SN postverbal peut apparaître avec l'accord riche sans qu'il soit disloqué. C'est dire que l'accord entre le sujet et le verbe est riche, à l'encontre de ce que nous constatons en arabe standard.

Pour résoudre ce problème, il faut noter ce qui suit :

1. L'accord dans les dialectes arabes n'est pas partiel, comme il l'est en arabe standard. L'accord en genre est strict même avec les collectifs. Le système de l'accord semble donc plus riche qu'en arabe standard et implique plus de traits.
2. Le redoublement des marqueurs d'objet est possible dans certains dialectes, dont l'arabe maltais et l'arabe tunisien. En arabe tunisien, par exemple, ce redoublement peut se faire sans l'insertion d'une préposition [48] devant le SN double, comme dans les exemples suivants :

(3.66) a. *wi:n* *-hum*$_i$ *l- wla:d*$_i$
où -ils les- garçons
« Où sont les garçons ? »

b. *ma-* *ħla:* *-ha:*$_i$ *tu:nis*$_i$
que- belle -elle Tunisie
« Que la Tunisie est belle ! »

Nos informateurs, qui parlent ce dialecte, nous ont confirmé qu'aucune coupure intonatoire ne se produit dans ce contexte. C'est le cas également quand le SN redouble un marqueur de sujet exhibant un accord riche, comme dans l'exemple suivant :

(3.67) *dixl* *-u:*$_i$ *l- wla:d*$_i$
entrer.PER -3MP les- garçons
« Les garçons sont entrés »

Il est clair que l'arabe dialectal (du moins l'arabe tunisien) permet le redoublement des arguments. De manière plus générale, la différence entre la variété standard et la variété dialectale semble être la suivante : l'arabe standard n'autorise aucun redoublement des arguments pronominaux incorporés, alors que les dialectes le permettent. Nous pensons que l'explication de cette différence ne peut être différente de l'explication des différences observées entre les dialectes également. En effet, et si nous ne prenons en considération que les dialectes arabes contemporains [49], ces dialectes peuvent être classés en trois sous-groupes concernant les possibilités du

[48] Les dialectes arabes du Levant permettent une forme de redoublement avec insertion d'une préposition. L'arabe tunisien et l'arabe maltais semblent donc différents à la fois de ces dialectes et de l'arabe standard. Cela prouve que la variation existe même au sein des groupes qu'on considère homogènes (les dialectes).

[49] Au VIIIe siècle, Sibawayhi nous apprenait l'existence d'un dialecte dont le comportement était semblable à celui des dialectes arabes contemporains : il l'appelle le dialecte de *ʔakal-u:-ni: l-bara:ɣi:θ* « *ils m'ont dévoré les puces ». Ce parlé est différent de la norme grammaticale défendue par Sibawayhi et tous les grammairiens depuis cette époque sur deux plans : l'accord riche dans l'ordre VSO et l'accord du pluriel d'un nom non-animé (qui doit être au FS dans la variété standard) comme un nom animé (au MP).

redoublement :

1. Des dialectes qui permettent le redoublement des marqueurs de sujet, mais ne permettent aucun redoublement des marqueurs d'objet (que ce soit par l'insertion d'une préposition ou de manière directe) : l'arabe marocain appartient à ce groupe.
2. Des dialectes qui permettent le redoublement des marqueurs de sujet et qui permettent le redoublement des marqueurs d'objet à condition que le redoublement se fasse par l'insertion d'une préposition : l'arabe du Levant appartient à ce groupe.
3. Des dialectes qui permettent le redoublement des marqueurs de sujet et des marqueurs d'objet, sans que le redoublement ne soit conditionnel à l'insertion d'une préposition : l'arabe tunisien appartient à ce groupe.

Nous pouvons désormais mettre l'arabe standard à la tête de cette liste comme étant la variété qui ne permet aucun redoublement des arguments pronominaux incorporés. Ces différences s'expliquent-elles sur un axe diachronique ? Nous pensons que la diachronie fournit des éléments de réponse. Il s'agit du processus de grammaticalisation mentionné par plusieurs auteurs, dont Ariel (2000), Givon (1976), Hale (1973), Meillet (1912) et Zwicky (1977) :

(3.68) mot $\longrightarrow$ clitique $\longrightarrow$ affixe

> After the development from independent word to clitic, the next step is, of course, the incorporation of clitics into morphology proper : what is a clitic at one stage is reinterpreted as a derivational or inflectional affix at the next. (Zwicky, 1977, p. 6)

Les trois étapes de cette évolution sont représentées dans les variétés de l'arabe. Les pronoms indépendants sont des mots et ils ont vraisemblablement donné naissance aux différents marqueurs d'arguments. Par ailleurs, certains des marqueurs d'arguments ont changé ou sont en cours de changement pour ne devenir que des marqueurs d'accord. Les quatre marqueurs de sujet de l'arabe standard *-a*, *-at*, *j-* et *t-* ont cette particularité de pouvoir être employés comme arguments ou comme marqueurs d'accord. Tous les marqueurs de sujet des dialectes sont devenus de simples marqueurs d'accord. Et finalement, les marqueurs d'objet de certains dialectes (en l'occurrence l'arabe tunisien) sont fonctionnellement ambigus et peuvent fonctionner comme des arguments ou comme des marqueurs d'accord. Cette affirmation est confirmée par les données empiriques : l'accord dans les dialectes est plus riche, que ce soit dans le système nominal ou dans le système verbal, alors que l'accord (donc la grammaticalisation complète des pronoms) reste partiel en arabe standard.

Il est évident que ces hypothèses ont besoin de plus d'investigations et qu'une étude plus poussée de cette question pourrait révéler d'autres découvertes. Nous pensons cependant que les quelques éléments de réponse que nous fournissons ici sont sur la bonne voie.

3.5 Conclusion

Nous avons consacré ce chapitre à une étude morphosyntaxique des marqueurs d'arguments de l'arabe standard. Les résultats de cette étude peuvent être résumés dans le tableau (3.9).

Marqueurs d'accord	Arguments	Explétifs, topiques
- Marqueurs de sujet co-indicés avec un sujet lexical.	- Marqueurs de sujet co-indicés avec un élément disloqué. - Marqueurs de sujet employés seuls. - Marqueurs d'objet compléments de têtes lexicales. - Marqueurs d'objet compléments de têtes fonctionnelles introduisant des phrases équatives.	- Marqueurs d'objet compléments de têtes fonctionnelles introduisant des phrases verbales.

TABLEAU 3.9: La morphosyntaxe des marqueurs d'arguments

Les marqueurs de sujet sont fonctionnellement ambigus et quatre d'entre eux peuvent ainsi fonctionner comme arguments et comme marqueurs d'accord. Ils sont des arguments (du verbe) quand le SN ou le pronom indépendant auquel ils sont associés sont disloqués (fonctionnant comme des topiques ou des focus) et quand ils apparaissent seuls sans être co-indicés avec ces constituants. Ils sont des marqueurs d'accord quand la structure comporte un SN sujet, lequel est postverbal. L'accord entre le sujet et le verbe se réalise ainsi seulement dans l'ordre de surface VSO (où S = SN). Quand le verbe n'est pas initial, le constituant préverbal est disloqué à gauche et il ne peut donc être l'argument du verbe.

Cette analyse de la morphosyntaxe des marqueurs de sujet a plusieurs implications, dont deux nous ont paru importantes : la pauvreté généralisée de l'accord en arabe standard et ce qui est communément appelé les asymétries de l'accord. En ce qui a trait à ces dernières, nous avons signalé qu'elles sont la manifestation de l'ambiguïté fonctionnelle des marqueurs de sujet et donc de la coexistence de deux sortes d'accord : un accord grammatical et un accord anaphorique. L'accord

grammatical implique l'emploi de marqueurs d'accord alors que l'accord anaphorique implique des arguments pronominaux dont les antécédents sont des constituants non argumentaux.

Les marqueurs d'objet ne connaissent pas cette ambiguïté fonctionnelle qui caractérise les marqueurs de sujet. De par cette propriété, les SN ou les pronoms indépendants associés aux marqueurs d'objet sont des non-arguments disloqués et l'accord observé est donc toujours un accord anaphorique.

Cette analyse morphosyntaxique des marqueurs d'arguments doit logiquement être suivie d'une analyse morphologique pour expliciter la nature de l'attachement observé entre les marqueurs d'arguments et leurs hôtes. Est-ce un attachement morphologique, auquel cas les marqueurs en question sont des affixes, ou un attachement phonologique, auquel cas ils sont des clitiques ? Nous consacrons le quatrième chapitre au traitement de cette question en nous basant essentiellement sur les tests de Zwicky.

CHAPITRE IV

LE STATUT MORPHOLOGIQUE DES MARQUEURS D'ARGUMENTS

> It is easy to mistake the nature of familiar tests for membership (...). The temptation is to see these tests as necessary and sufficient conditions for the applicability of a theoretical term, i.e. as DEFINITIONS of the term. But what is normally intended, when such tests are appealed to, is more analogous to medical diagnosis than to operations using an axiomatic system. The tests point to characteristic SYMPTOMS of a linguistic state of affairs, not to invariant concomitants of it. (Zwicky et Pullum, 1983, p. 284-285)

Introduction

En étudiant la morphosyntaxe des marqueurs d'arguments de l'arabe standard, nous avons proposé de les traiter comme des unités appartenant à deux classes hétérogènes. Les marqueurs d'objet sont morphosyntaxiquement des arguments, des explétifs ou des topiques suivant la tête de laquelle ils sont compléments. Les marqueurs de sujet sont caractérisés par une ambiguïté fonctionnelle qui touche certains d'entre eux. En fait, les deux marqueurs du perfectif *-a* et *-at* et les deux marqueurs de l'imperfectif *j-* et *t-* couvrent deux réalités différentes : celle d'arguments pronominaux et celle de marqueurs d'accord. Certains marqueurs de sujet pronominaux sont ainsi des homophones de marqueurs d'arguments.

Si nous adoptons l'approche de Anderson (2005), qui affirme que les marqueurs de sujet qui excluent tout argument nominal sont dotés de propriétés pronominales, nous pouvons affirmer que les marqueurs d'objet sont des pronoms puisqu'ils ne peuvent être redoublés. Nous pouvons également affirmer que les marqueurs de sujet arguments le sont au même titre que les marqueurs d'objet. Les quatre marqueurs d'accord verbal ne sont de toute évidence pas des pronoms selon cette analyse. Cela rejoint également la définition des pronoms avancée par Bresnan (2001) [1].

[1] Voir l'introduction du deuxième chapitre de notre travail.

Suite au travail accompli dans le troisième chapitre, nous savons désormais que certains marqueurs d'arguments font partie des arguments de leurs hôtes. Cela se traduit dans la théorie HPSG par l'encodage de ces éléments dans la liste qui constitue la valeur de l'attribut STR-ARG. Il s'agit des marqueurs d'objet compléments de verbes, de noms et de prépositions, en plus des marqueurs de sujet argumentaux. Les non-arguments, à savoir les marqueurs d'accord, les explétifs et les topiques, ne peuvent, cependant, être membres de cette liste.

Dans la théorie HPSG standard, les arguments sélectionnés par la tête ne sont pas toujours « projetés » dans la structure syntagmatique, cette projection qui se traduit par l'encodage des éléments projetés comme valeurs de l'attribut VAL(ence). Pour que cette « projection » soit possible, les arguments sélectionnés ne doivent pas être des affixes, puisque les affixes font partie du mot et ne peuvent constituer des filles (ou des nœuds indépendants) dans l'arbre. Nous n'avons cependant pas encore abordé la question du statut morphologique des marqueurs d'arguments. Le problème se pose dans ces termes : les arguments de la tête sont-ils forcément tous projetés dans la liste de valence ? Certains auteurs parlent d'affixes en désignant les marqueurs d'arguments, est-ce motivé ? Peut-on, comme le font certains, les désigner par le terme « clitiques » ?

Notre objectif dans ce quatrième chapitre est d'assembler le plus grand nombre possible de tests et de critères nous permettant de traiter ces marqueurs d'arguments sous l'une ou l'autre des étiquettes suivantes : affixes ou non-affixes (essentiellement clitiques).

La distinction entre les affixes et les non-affixes n'est toutefois pas des plus aisées. Plusieurs auteurs ont essayé de proposer des critères qui permettent d'opérer de telles distinctions entre des unités qui sont en apparence très semblables et qu'on regroupe souvent sous les appellations « clitiques », « particules » et « affixes ». En incluant les modifications proposées par Anderson (2005), Klavans (1995, 1985, 1983) et Miller (1992), nous utilisons les critères de Zwicky, proposés par Zwicky (1977, 1985b,a) et Zwicky et Pullum (1983), pour soulever le problème du statut morphologique des marqueurs d'arguments de l'arabe. Nous commençons par présenter ces critères dans la section (4.1). De cette présentation, nous retenons quelques tests, dont nous prouvons la validité pour la langue arabe. Ces tests seront appliqués dans la section (4.2) pour montrer que les marqueurs d'arguments de l'arabe sont marqués par une hétérogénéité morphologique : les marqueurs de sujet sont des affixes alors que les marqueurs d'objet sont mieux traités comme des non-affixes (en l'occurrence, des clitiques).

4.1 Les tests du statut morphologique

Les travaux de Zwicky concourent à prouver que les clitiques sont à cheval entre les mots et les affixes. Ce statut intermédiaire est, depuis, reconnu comme étant la pierre angulaire de la définition de ces unités. L'auteur est parvenu à cette conclusion dès son ouvrage de 1977, où il a tenté de fournir des critères qui servent à distinguer les affixes des mots. Les idées exprimées dans cet ouvrage seront raffinées au fil du temps et l'auteur sera amené à proposer d'autres critères pour distinguer les affixes des clitiques et les clitiques des mots.

4.1.1 Les affixes, les clitiques et les mots

4.1.1.1 Les affixes et les mots

Zwicky (1977) marque le début de l'étude moderne des clitiques dans le cadre génératif. Ce manuscrit a servi de base pour les recherches ultérieures et les idées ainsi que la terminologie de l'auteur ont longtemps constitué un véritable programme de recherche pour les travaux ultérieurs selon Anderson (2005). La question qui était à l'origine de cette longue tradition concernait le morphème réduplicatif R du madurais[2] : est-ce un mot ou un affixe ? Pour répondre à cette question, Zwicky a proposé six critères pour distinguer les affixes des mots :

1. L'ordre : l'ordre des affixes dans un mot est rigide alors que l'ordre des mots dans un syntagme peut être variable d'une langue à l'autre. En outre, un changement dans l'ordre des affixes entraîne un changement de sens, alors que le changement de l'ordre des mots n'entraîne pas forcément un tel changement. En ce qui a trait au morphème R, Zwicky observe que ce dernier se comporte comme un mot : il peut être ordonné de plusieurs manières par rapport à certains autres morphèmes, comme le préfixe causatif *pa-* sans que le sens change. Dans les exemples suivants, cette variabilité de l'ordre est mise de l'avant : le radical est *kumul* « ramasser ».

   (4.1) a. *i-* *pa-* ***pul-*** *kumul*
   passif- causatif- R- radical

   b. *i-* ***pul-*** *pa-* *kumul*
   passif- R- causatif- radical

[2] Une langue austronésienne parlée essentiellement en Indonésie dans l'île de Madura.

c. ***pul-*** *i-* *pa-* *kumul*
R- passif- causatif- radical

2. Le Sandhi interne : les règles d'ajustement phonologique caractéristiques des frontières entre les mots n'ont pas d'effet à l'intérieur des mots. Le morphème R du madurais se comporte comme un mot à cet égard : les règles qui affectent les frontières des morphèmes à l'intérieur du mot (comme l'harmonie vocalique et la syllabation) n'affectent pas les frontières qui séparent R des morphèmes adjacents.

3. L'attachement : les morphèmes qui sont toujours attachés à d'autres éléments sont des affixes ; les mots sont libres. Le morphème R se comporte comme un affixe à cet égard ; il ne peut être isolé en réponse à une interrogative ou séparé des morphèmes auxquels il s'attache par une parenthèse par exemple.

4. La construction avec des affixes : si un élément est construit avec un affixe, il doit être un radical ou lui-même un affixe. Le morphème R est un affixe à cet égard, puisqu'il est construit avec des affixes.

5. L'immunité aux règles d'effacement : les règles syntaxiques, par exemple l'effacement sous identité, n'affectent pas les affixes. Les exemples que donne l'auteur proviennent de l'anglais : les affixes *-ish* et *-ing* doivent être répétés sur leurs hôtes et ne peuvent être effacés dans la coordination (4.2). Le morphème R, tout comme ces affixes de l'anglais, ne peut subir un tel effacement.

   (4.2) a. yellow**ish** or grey**ish**
   b. *yellow or grey**ish**
   c. dan**cing** and sing**ing**
   d. *dance and sing**ing**

6. L'accent : les éléments non porteurs d'accent indépendant sont des affixes et non des mots. R est un affixe à cet égard n'étant pas porteur d'accent indépendant.

Le morphème R serait donc un mot selon les deux critères 1 et 2 et un affixe selon les autres critères. Par ailleurs, l'auteur reconnaît que certains des six critères ont besoin d'être raffinés. Ce raffinement est d'autant plus nécessaire qu'il existe certains morphèmes qui ne peuvent être clairement classés selon ces critères, ni parmi les mots indépendants, ni parmi les affixes : ce sont les clitiques[3] spéciaux, les clitiques simples et les morphèmes conjoints.

[3] L'auteur emploie le terme « clitique » dans un sens préthéorique, c'est-à-dire sans émettre des hypothèses sur la nature lexicale ou syntaxique de ces unités.

Les clitiques spéciaux sont des formes conjointes non accentuées qui sont des variantes de formes libres accentuées, gardant un sens et une constitution phonologique semblables, tout en exhibant une syntaxe spéciale. Parmi les exemples que donne Zwicky, nous trouvons les marqueurs d'objet de l'arabe égyptien familier : *-(a)k* « te » et *-ha:* « la » dont les variantes libres sont *ʔinta* « toi » et *hijja* « elle ». Il explique les différences de distribution entre ces deux sortes de pronoms dans les termes suivants : « the enclitic pronouns are objects or possessives, while the independent pronouns are normally subjects, though they can be used with possessive enclitics for emphasis. » (Zwicky, 1977). En fait, la forme disjointe est typiquement employée quand on fait appel à un pronom accentué pour des raisons syntaxiques ou sémantiques, en particulier quand un pronom doit être accentué ou quand il doit être isolé (par exemple, dans les réponses à des questions). Cette syntaxe, dite spéciale, est plus claire dans le cas des clitiques objets du français standard : ces derniers apparaissent dans des positions que ne peuvent occuper leurs équivalents indépendants (la série des pronoms forts *moi*, *toi*, etc.), ni, dans certains contextes, les SN objets. Finalement, la phonologie de ces clitiques est spéciale également ; ils ne sont pas dérivés des formes libres correspondantes par des règles phonologiques d'une portée générale (= indépendantes).

Les clitiques simples sont le résultat de la réduction phonologique de morphèmes libres. Cette réduction est accompagnée d'un affaiblissement prosodique et le clitique simple finit par s'attacher à un autre morphème avoisinant. Cette réduction est généralement conditionnée par le style. Elle se rencontre, par exemple, dans le système pronominal de l'anglais : les pronoms objets *him* et *her* ont des variantes réduites (les clitiques simples *'m* et *'r*). Ces clitiques n'exhibent pas de syntaxe spéciale : ils apparaissent dans les mêmes positions que leurs variantes libres. De plus, leur réduction se fait selon des règles phonologiques d'une portée générale.

Les mots conjoints sont des morphèmes qui apparaissent toujours dans cet état, qui sont toujours non accentués et qui montrent une liberté syntaxique considérable, en ceci qu'ils peuvent être attachés à des mots de plusieurs catégories morphosyntaxiques. Ces morphèmes sont généralement localisés aux marges des mots, se situant même à l'extérieur des affixes flexionnels. Les exemples que donne l'auteur sont la particule latine *-que*, les particules du tagalog (les indicateurs de fonction, quelques particules adverbiales, les monosyllabes et quelques conjonctions) et le possessif anglais *'s*.

Ces trois sortes de morphèmes peuvent être considérées comme des clitiques. Ils constituent des cas d'exceptions aux six critères de distinction entre les mots et les affixes parce qu'ils représentent des mixtes de ces critères. Une première définition de ce qui est un clitique, dans un sens

préthéorique, se laisse entrevoir [4] :

(4.3) Les clitiques sont des morphèmes qui ne sont pas clairement identifiables comme des mots indépendants ou comme des affixes puisque leurs propriétés sont un complexe composé des propriétés des uns et de celles des autres.

Suivant cette définition, Zwicky considère que le morphème R du madurais est un clitique, puisqu'il se comporte comme un mot indépendant eu égard aux deux premiers critères (l'ordre alternatif et le manque de Sandhi) alors qu'il se comporte comme un affixe eu égard aux autres critères.

Le tableau (4.1) résume les points communs et les points de divergence entre les trois sortes de clitiques reconnues par l'auteur.

	Déficience accentuelle	Position spéciale	Allomorphe
Clitiques spéciaux	+	+	+
Clitiques simples	+	-	+
Mots conjoints	+	-	-

TABLEAU 4.1: Les clitiques selon Zwicky (1977)

La déficience accentuelle est le seul point commun entre ces trois sortes de clitiques. À noter que c'est également un critère qui distingue les affixes des mots indépendants. La position syntaxique spéciale est le critère décisif pour distinguer les clitiques spéciaux des deux autres sortes de clitiques. L'existence d'un allomorphe indépendant semble distinguer les clitiques (spéciaux et simples) des mots conjoints. Cependant, ce critère reste le moins significatif, comme le prouve Anderson (2005). C'est sans doute pour cette raison que Zwicky et Pullum (1983) ont aboli cette distinction tripartite pour ne garder que deux sous-classes de clitiques : les clitiques spéciaux et les clitiques simples.

Deux dimensions semblent donc définir un clitique : la dimension phonologique (la déficience accentuelle) et la dimension morphosyntaxique (la position spéciale ou ordinaire). Ces deux dimensions constituent la base de la définition des clitiques dans plusieurs travaux, dont Anderson

[4] C'est la définition première de ce qui est un clitique dans plusieurs ouvrages, dont Miller et Monachesi (2003) :

> Dans la grande majorité des langues, on trouve certains éléments dont le statut est problématique parce que leur comportement est en apparence intermédiaire entre celui des mots indépendants et celui des affixes habituels. S'ils semblent jouir d'une plus grande autonomie que ces derniers, ils s'appuient phonologiquement à un hôte, contrairement aux mots, et forment avec lui un seul mot prosodique.

(2005). L'auteur définit ce qu'il entend par *clitique phonologique* (4.4a)[5] et *clitique morphosyntaxique*[6] (4.4b)[7] :

(4.4) a. *Clitique phonologique*

Un élément linguistique dont la forme phonologique est déficiente et qui est privé d'une structure prosodique au niveau du mot (prosodique).

b. *Clitique morphosyntaxique*

Un élément linguistique dont la position est régie par un ensemble de principes distincts de ceux qui régissent les positions des éléments indépendants dans la langue.

La contribution de Zwicky à l'étude des clitiques n'est pas restreinte à ses critères de distinction entre les mots et les affixes ; ses travaux ultérieurs ont été voués à fournir des critères qui permettent de distinguer les clitiques des affixes et les clitiques des mots. Ces critères seront présentés dans les deux sections suivantes.

4.1.1.2 Les affixes et les clitiques

La déficience accentuelle caractérise aussi bien les clitiques que les affixes. Les deux types de morphèmes apparaissent en fait toujours attachés à d'autres constituants, d'où la nécessité d'établir des critères distinctifs explicites. De même, cette nécessité émane d'une problématique posée par les auxiliaires contractés[8] et le négatif contracté *'nt* en anglais. La question de recherche derrière l'article de Zwicky et Pullum (1983) est la suivante : ces morphèmes sont-ils mieux traités comme affixes ou comme clitiques ? Pour répondre à cette question, six critères sont proposés :

Critère A Les affixes sont très sélectifs de leurs radicaux alors que les clitiques peuvent montrer un moindre degré de sélection par rapport à leurs hôtes. Le degré de sélection n'étant pas quantifiable, il signifie tout au plus qu'un morphème qui apparaît avec n'importe quel mot est moins sélectif qu'un morphème qui apparaît avec seulement des verbes, par exemple. Ce

[5] (*Ibid.*, p. 23), traduction libre.

[6] Nous verrons cependant que ces deux dimensions ne sont pas suffisantes pour entièrement distinguer les clitiques des affixes, par exemple. La définition du clitique phonologique en (4.4a) est applicable aux affixes également. De plus, comme le signale Klavans (1983), certains clitiques reconnus comme tels ne sont pas déficients prosodiquement au niveau du mot. Nous pourrions conclure que la dimension morphosyntaxique est décisive dans la définition des clitiques, mais cette conclusion exclut les clitiques simples (dont la position n'a rien de spécial). Pour remédier à cette situation, nous pensons qu'il est possible de changer la dimension morphosyntaxique pour englober tous les clitiques postlexicaux, que ces derniers soient spéciaux ou simples.

[7] (*Ibid.*, p. 31), traduction libre.

[8] Un exemple est l'auxiliaire *have*, qui peut avoir une forme contractée *'ve* dans *they've done*. Notons que la variante contractée peut être considérée comme un clitique simple dans les termes de Zwicky.

dernier est à son tour moins sélectif qu'un morphème qui n'apparaît qu'avec une sous-classe de verbes. Et ce dernier est moins sélectif qu'un morphème qui apparaît avec quelques verbes en particulier Zwicky (1985a) utilise ce test et les cinq tests suivants pour s'interroger sur le statut morphologique des morphèmes de mode (qu'il note M) en hidatsa. Ces derniers se comportent comme des affixes étant donné qu'ils n'apparaissent attachés qu'après des verbes. Ce même critère a permis à d'autres auteurs, comme Miller (1992) et Miller et Sag (1997, 1995), d'affirmer que les clitiques verbaux du français ont certaines propriétés affixales de par leur attachement exclusif à des verbes. En anglais, les affixes du pluriel ne s'attachent qu'à des noms, les affixes du passé ne s'attachent qu'à des verbes et les affixes du superlatif ne s'attachent qu'à des adjectifs et des adverbes. Par contre, les clitiques simples s'attachent à plusieurs catégories lexicales, comme le montrent les exemples suivants[9] .

(4.5) a. The person I was talking <u>to</u>'s going to be angry with me.
b. The ball you <u>hit</u>'s just broken my dining room window.
c. Any answer not entirely <u>right</u>'s going to be marked as an error.
d. The drive home <u>tonight</u>'s been really easy.

Critère B Les trous dans la distribution sont plus caractéristiques des composés [radical+affixe] que des composés [hôte+clitique]. En anglais, comme le mentionnent les auteurs, aucune combinaison possible entre un hôte et un clitique simple ne manque à l'appel. Par contre, les paradigmes flexionnels souffrent de tels trous. Ainsi, le verbe *stride*, par exemple, n'a pas de participe passé en anglais. Certains auteurs ont employé ce critère pour démontrer le statut affixal des clitiques des langues romanes. Ces derniers sont caractérisés par plusieurs trous qui se manifestent de deux manières :

- Dans la combinaison des clitiques les uns avec les autres. L'italien, par exemple, ne permet pas la combinaison d'un clitique accusatif de la première ou de la deuxième personne avec un clitique datif, d'où l'agrammaticalité de (4.6)[10] :

  (4.6) **Martina gli mi presenta.*
  Martine le(ACC) me(DAT) présente

- Dans la combinaison des clitiques avec leurs hôtes. En français, par exemple, pour la plupart des verbes, il n'y a pas de forme acceptable pour l'inversion de la première personne du singulier *je* et d'un verbe au présent de l'indicatif, en opposition aux autres personnes, d'où le contraste en (4.7)[11] :

[9] (Zwicky et Pullum, 1983, p. 504).
[10] (Monachesi, 1998a).
[11] (Miller et Sag, 1997).

(4.7) a. Sors-tu ? Chantes-tu ?

b. *Sors-je ? * Chante-je ?

Critère C Les idiosyncrasies morphophonologiques sont plus caractéristiques des composés [radical+affixe] que des composés [hôte+clitique]. En anglais, quand un clitique simple est attaché à un hôte donné, ce dernier n'est pas affecté par cet attachement et aucune forme phonologique inattendue n'émerge de cette cliticisation. Les affixes, par contre, provoquent des changements phonologiques d'une grande ampleur qui peuvent aller jusqu'à l'apparition de formes supplétives. Parmi les exemples fournis par les auteurs, citons les formes idiosyncrasiques du pluriel dans les mots *dice*, *oxen* et *feet* et du passé dans les verbes *slept*, *thought* et *went*. Des idiosyncrasies semblables ont été observées dans les combinaisons des clitiques romans avec leurs hôtes verbaux. Citons comme exemple la réalisation idiosyncrasique de *je suis* en *chuis*, forme possible seulement avec le verbe/auxiliaire *être* et non permise avec *suivre*.

Critère D Les idiosyncrasies sémantiques sont plus caractéristiques des composés [radical+affixe] que des composés [hôte+clitique]. Ainsi, en anglais, la contribution sémantique des clitiques simples *'ve* et *'s*, par exemple, n'est pas différente de la contribution sémantique de leurs allomorphes indépendants *have* et *is/has*. Cependant, la flexion ne préserve pas toujours le sens de son hôte et la combinaison des deux n'est souvent pas compositionnelle. Un exemple est celui du superlatif anglais *last* (dérivé de l'adjectif *late*), qui a la syntaxe d'un superlatif, mais un sens idiosyncrasique. Ainsi, l'expression *last words* signifie « mots finals » et non pas « mots très en retard » ou « mots récents ».

Critère E Les règles syntaxiques peuvent affecter les composés [radical+affixe], mais ne peuvent affecter les composés [hôte+clitique]. Ainsi, aucune opération syntaxique ne traite les mots composés des clitiques simples *'ve* et *'s* et de leurs hôtes comme des unités, alors que c'est le cas des affixes du pluriel, du passé et du superlatif, qui forment avec leurs hôtes des unités manipulées par la syntaxe.

Critère F Les clitiques peuvent s'attacher à du matériel qui contient déjà des clitiques, mais les affixes ne le peuvent pas. Les clitiques simples peuvent ainsi s'attacher à des hôtes qui contiennent des clitiques, comme en (4.8) [12], ce qui n'est pas possible pour des affixes :

(4.8) I'd've done it if you'd asked me.

Les critères E et F sont le résultat de l'adoption de l'hypothèse suivante : toute cliticisation est postsyntaxique et aucune opération syntaxique n'est possible après la cliticisation. Nous pouvons

[12] (Zwicky et Pullum, 1983, p. 506).

également postuler que tous les critères, y compris E et F, découlent de l'hypothèse lexicaliste. Telle qu'elle est explicitée par Anderson (2005), cette hypothèse repose sur deux prémisses. La première est que les mots (et toute affixation) sont formés dans le lexique. Par conséquent, l'affixation a accès à la forme et au sens des radicaux et peut les affecter de manière idiosyncrasique (d'où les différentes idiosyncrasies qui caractérisent les mots affixés). La deuxième prémisse est que les mots sont combinés les uns aux autres postlexicalement. La syntaxe ne peut donc avoir accès à la composition interne des mots.

Sans cette hypothèse, les 6 critères de Zwicky et Pullum (1983) ne sont que des observations descriptives concernant les différences entre deux types de morphèmes. Adopter ces critères revient donc en quelque sorte à adopter l'hypothèse qui les sous-tend, à savoir l'hypothèse lexicaliste. Il n'est donc pas surprenant de voir plusieurs auteurs qui travaillent dans le cadre d'une théorie lexicaliste comme HPSG employer ces critères : Kupsc (2000), Miller (1992), Miller et Sag (1997, 1995) et Monachesi (1999b).

L'application de ces critères par Zwicky et Pullum les a conduits à considérer que le négatif contracté de l'anglais *'nt* (dans *isn't*, *don't*, *won't*, etc.) est un affixe flexionnel et non un clitique simple. Miller (1992), ainsi que Miller et Sag (1997, 1995), sont arrivés à la même conclusion par rapport aux clitiques du français moyennant quelques changements aux six critères. Abu Nahleh (1985) et Broselow (1983) ont utilisé ces mêmes tests dans leurs études de variétés dialectales de l'arabe (l'arabe palestinien de Ramallah et l'arabe égyptien du Caire respectivement).

Pour compléter la panoplie des tests de Zwicky, nous consacrons la section suivante aux tests qui servent à distinguer les clitiques des mots. En effet, si les clitiques sont postlexicaux par définition, il conviendrait d'avoir des critères explicites pour les distinguer des mots qui le sont également.

4.1.1.3 Les mots et les clitiques

Constatant que certains morphèmes sont à tort classés parmi les particules, comme si ces dernières constituaient une catégorie à part entière, Zwicky (1985b) s'interroge sur le statut morphologique de plusieurs de ces *particules*. Pour ce faire, il propose des tests susceptibles de distinguer les clitiques des mots et ainsi compléter la série de tests qui existent déjà pour distinguer les affixes des mots et les clitiques des affixes. Ces tests, comme d'ailleurs les précédents, doivent être pris comme un *diagnostic médical*, qui repose sur des *symptômes* observés, nous prévient

l'auteur. Ces tests sont groupés dans six catégories : des tests phonologiques, un test accentuel, des tests comparatifs aux affixes, des tests syntaxiques, un test dérivé des conditions sur les interfaces et un test en forme de métaconsidération, au cas où les autres tests seraient insuffisants et non concluants.

4.1.1.3.1 Tests phonologiques

Les tests phonologiques sont basés sur une considération générale : les combinaisons de mots [mot+mot] forment un syntagme phonologique, alors que les combinaisons [mot+clitique] forment un mot phonologique. Les opérations (et les règles) phonologiques permises au sein d'un mot phonologique ne sont pas les mêmes opérations permises au sein d'un syntagme phonologique, d'où les tests phonologiques suivants :

- Sandhi : les règles de Sandhi interne s'appliquent à l'intérieur des mots phonologiques (et affectent donc les clitiques) alors que les règles de Sandhi externe s'appliquent entre les mots phonologiques (et affectent les mots indépendants).
- Les domaines prosodiques : le domaine dans lequel un trait prosodique est attribué peut être soit le mot phonologique soit le syntagme phonologique (ou une autre unité prosodique comme la syllabe). Par conséquent, si un élément est considéré comme appartenant à un mot phonologique à des fins d'assignation de l'accent, du ton ou de la longueur, alors il devrait être un clitique. Si un élément est considéré comme appartenant à un syntagme phonologique pour ces fins, il devrait être considéré comme un mot.
- Les domaines segmentaux : certaines règles (par exemple, l'harmonie vocalique) affectent les traits segmentaux, mais elles sont prosodiques en nature, puisque leur domaine d'applicabilité est une certaine unité prosodique. Par conséquent, si un élément est considéré comme appartenant à un mot phonologique aux fins de telles règles, alors il devrait être un clitique. Si un élément est considéré comme appartenant à un syntagme phonologique à ces fins, alors il devrait être un mot indépendant.

4.1.1.3.2 Test accentuel

Ce test est de loin le plus utilisé, mais il est également le moins fiable. Il s'agit de distinguer un mot d'un clitique par leurs propriétés accentuelles : si une unité est dépourvue d'accent et apparaît dépendante d'une autre unité pour recevoir un accent, elle a toutes les chances d'être un

clitique et non un mot. Si, au contraire, une unité est accentuellement indépendante, elle a toutes les chances d'être un mot. Ainsi, les marqueurs de discours en anglais (par exemple, *well, oh, like, uh, say* etc.) sont prosodiquement indépendants et doivent par conséquent être considérés comme des mots et non comme des clitiques, selon l'auteur

Zwicky (1985b) signale deux problèmes posés par ce test : un problème mineur et un problème majeur. Le problème mineur c'est que certaines langues permettent aux clitiques d'être accentués dans certains contextes [13]. Le problème majeur consiste dans le fait que plusieurs unités considérées comme mots ne portent pas d'accent indépendant. Tel est le cas des prépositions, des déterminants et des auxiliaires en anglais. Compte tenu de ces deux problèmes, l'auteur nous invite à ne jamais prendre ce test en isolation ; il peut, tout au plus, appuyer des résultats obtenus grâce aux autres tests.

4.1.1.3.3 Tests comparatifs

Ces tests s'appuient sur les similitudes entre les affixes et les clitiques. Autrement dit, un clitique a certaines propriétés de par sa ressemblance aux affixes (surtout flexionnels) qu'un mot n'a pas. Les cinq tests suivants découlent de cette comparaison :

1. L'attachement : les éléments conjoints ont plus de chances d'être des clitiques, alors que les éléments libres ont plus de chances d'être des mots. Ce test a permis de traiter certaines « particules » de l'allemand (par exemple, *doch* « oui ») comme mots et non comme clitiques, étant donné que ces dernières peuvent être isolées en réponse à des questions, par exemple.
2. Le verrouillage : typiquement, certains affixes flexionnels *verrouillent* les mots devant toute affixation supplémentaire. Par conséquent, un élément qui constitue une clôture pour les combinaisons devant une affixation ou une cliticisation supplémentaire est probablement un clitique.
3. La construction : les affixes flexionnels se combinent avec des radicaux ou des mots entiers, alors que les mots se combinent avec d'autres mots ou avec des syntagmes.
4. L'ordre : on ne change pas l'ordre des morphèmes dans un mot sans affecter le sens. L'ordre des mots est cependant stylistique et n'affecte pas le sens. Par conséquent, si un élément est strictement ordonné par rapport à des morphèmes adjacents, il est probablement un clitique

[13] Klavans (1983) et Anderson (2005) citent des exemples où des clitiques reçoivent l'accent par des règles générales d'accentuation, par emphase ou par contraste.

(ou un affixe), alors qu'un élément dont l'ordre est libre par rapport à d'autres mots est certainement un mot indépendant.

5. La distribution : les affixes ont généralement un principe simple qui gouverne leur distribution, alors que la distribution des mots ne peut généralement pas être décrite par un tel principe. Les clitiques ont des principes de combinaison simples, comme « combiner avec la tête verbale d'une proposition », « combiner avec le premier mot d'une proposition » ou « combiner avec un SN », etc. Il s'en suit qu'un élément avec une simple distribution est probablement un clitique (ou un affixe) et qu'un élément avec une distribution complexe est presque sûrement un mot indépendant.
6. La complexité : les mots sont morphologiquement complexes en ceci qu'ils peuvent être analysés comme étant composés de deux morphèmes ou plus. Les unités affixales le sont rarement. À cet égard, les clitiques se comportent comme les affixes.

4.1.1.3.4 Tests syntaxiques

Ces tests sont basés sur l'hypothèse suivante : un mot peut servir de constituant syntaxique et donc être sujet aux processus syntaxiques. Le clitique devrait être immunisé contre de tels processus. Il s'ensuit les trois tests syntaxiques suivants :

1. L'effacement : les parties d'un mot ne sont pas sujettes à l'effacement sous identité ; les mots entiers peuvent (dans certains contextes) être effacés. Les combinaisons [clitique+mot] sont également immunisées contre l'effacement : « It follows that, in an X+Y combination, if either X or Y is deletable under identity, then X and Y are words ; neither is a clitic. » (Zwicky, 1985b, p. 288)
2. Le remplacement : les parties d'un mot ne sont pas sujettes à remplacement par une proforme sous identité (à l'encontre des mots entiers). Les parties des combinaisons [mot+clitique] sont également immunisées contre ce remplacement : « It follows that, in a X+Y combination, if either X or Y is replaceable by a pro-form, then X and Y are words ; neither is a clitic. » (*Ibid.*, p. 289)
3. Le déplacement : les parties de mots ne sont pas sujettes aux « règles de déplacement », c'est-à-dire qu'ils ne peuvent servir de *gaps* dans des relations à longue distance (ou non bornées) *filler-gaps* avec d'autres constituants. Les mots entiers (dans les circonstances appropriées) peuvent participer à de telles relations. De même, les parties des combinaisons [mot+clitique] ne peuvent pas entrer dans de telles relations : « It follows that, in an X+Y

combination, if either X or Y can be moved without the other, then X and Y are words; neither of them is a clitic. » (*Ibid.*, p. 289)

4.1.1.3.5 Test dérivé des conditions sur les niveaux d'analyse

L'auteur défend l'idée selon laquelle toute cliticisation est postsyntaxique. Il s'en suit que les règles syntaxiques ne devraient pas faire appel à des composés [mot+clitique] et que, le cas échéant, l'unité combinée au mot n'est pas un clitique.

4.1.1.3.6 Métaconsidération

Cette métaconsidération est appelée à régler les cas dans lesquels aucune évidence ne nous permet de clairement statuer qu'une unité est un mot ou un clitique. Dans de tels cas, il faudrait supposer que cet élément est un mot. Pourquoi ? L'auteur fait appel à la faculté du langage : les clitiques sont plus marqués que les affixes flexionnels et les unités syntaxiques indépendantes (les mots). Deux arguments appuient cette hypothèse : la morphologie flexionnelle est plus marquée que la syntaxe et plusieurs langues sont entièrement isolantes [14] (analytiques), mais aucune langue n'est presque totalement synthétique. En prenant cette métaconsidération en compte, le test que propose Zwicky est le suivant : « An item whose standing is unclear is most likely to be an independent word, next most likely to be an inflectional affix, and least likely to be a clitic. » (*Ibid.*, p. 289)

Nous avons présenté les trois séries de tests proposés par Zwicky pour établir une distinction claire et fondée entre trois catégories apparentées : les mots, les affixes et les clitiques. Dans la première série de tests, la distinction se fait entre les mots et les affixes, dans la deuxième le souci est de distinguer les clitiques des affixes et dans la troisième la distinction se fait entre les mots et les clitiques. Nous retrouvons, dans cette dernière, des critères qui étaient présentés dans la première, surtout quand l'auteur fait appel aux propriétés communes aux affixes et aux

[14] Une langue isolante est une langue dans laquelle les mots ont tendance à être invariables, avec une morphologie pauvre (comme c'est le cas en chinois). Une langue synthétique combine plusieurs morphèmes au sein du mot de manière agglutinante ou fusionnelle.

clitiques. Klavans (1995, 1985, 1983) a basé son étude des clitiques en grande partie sur ces critères typologiques. Elle propose en outre un nouveau terme pour caractériser les clitiques, celui d'*affixes syntagmatiques*, qui efface certaines frontières entre les clitiques (dans le sens des tests de Zwicky) et les affixes.

4.1.2 Clitiques et affixes syntagmatiques

Selon Klavans (1985), aucun critère ne peut à lui seul définir ce qui est un clitique et la déficience accentuelle est un exemple de ces critères qu'on ne peut tenir pour décisifs puisque certains clitiques peuvent être accentués par des règles phonologiques générales, alors que certains mots peuvent être dépourvus d'accent indépendant. Ce qui définit le clitique pour cet auteur est le fait qu'il est plutôt un affixe d'un certain type : un affixe syntagmatique.

Ce qui a amené l'auteure à faire cette proposition dans Klavans (1983) c'était sa conception du placement des clitiques. Les positions possibles étant limitées, les paramètres responsables de ce placement sont limités également et ils sont au nombre de trois : les deux premiers sont syntaxiques et le dernier est phonologique.

P1 La dominance (initial/final) : la possibilité qu'un clitique s'attache au constituant initial ou final dominé par un syntagme spécifique. Le possessif *'s* de l'anglais, par exemple, est toujours dominé par un nœud SN alors que les hôtes des clitiques de deuxième position sont toujours dominés par un nœud P (phrase).

P2 La préséance (avant/après) : ce paramètre spécifie si un clitique apparaît avant ou après l'hôte choisi par P1.

P3 La liaison phonologique (proclitique/enclitique) : ce paramètre précise la direction de l'attachement phonologique ; c'est la propriété du clitique en tant que tel.

Ces paramètres[15] ne permettent que 8 possibilités de cliticisation, comme illustré dans le tableau (4.2).

[15] Anderson (1992) propose une nouvelle caractérisation de ces paramètres comme suit :

1. Le clitique est localisé dans un constituant syntaxique qui représente son domaine.
2. Ce domaine peut être une P(hrase), un SN, un SV, etc.
3. Le clitique est localisé en se référant à un élément spécifique de ce domaine.
4. Cet élément peut alors être le premier, le dernier ou la tête.
5. Le clitique précède ou suit ce point de référence.

	P1	P2	P3	Exemples
Type 1	Initial (sous N')	Avant	Enclitique	Les marqueurs de SN du kwakwala
Type 2	Initial (sous N')	Avant	Proclitique	L'article en grec
Type 3	Initial (sous P)	Après	Enclitique	Les enclitiques du ngiyambaa
Type 4	Initial (sous P)	Après	Proclitique	Le clitique =an du tepecano
Type 5	Final (sous P)	Avant	Enclitique	Les clitiques du nganhcara
Type 6	Final (sous P)	Avant	Proclitique	Les préverbes du sanskrit
Type 7	Final (sous V[-T])	Après	Enclitique	Les clitiques pronominaux de l'espagnol
Type 8	Final (sous P)	Après	Proclitique	Le négatif grec ou=

TABLEAU 4.2: Les types de clitiques selon Klavans

Étant donné que le domaine de la cliticisation (défini par P1) est le plus souvent un N' ou un P, Klavans en conclut que la cliticisation est en réalité une affixation au niveau syntagmatique. Le seul problème auquel cette analyse semble ne pas trouver de solution est celui des clitiques des langues romanes. Ces derniers semblent en fait attachés à une catégorie mineure (le verbe) et non à des syntagmes (des SV, par exemple). Cette remarque a poussé l'auteure à proposer que ces clitiques soient en réalité des traits du verbe et qu'ils soient en voie de devenir des affixes au niveau du mot (par contraste aux affixes syntagmatiques).

Le terme d'affixe syntagmatique a été repris par plusieurs auteurs depuis qu'il a été proposé par Klavans. Parmi ces derniers, citons Anderson (2005, 1992), Nevis (1988) et Zwicky (1987). Anderson (1992), par exemple, emploie ce terme pour démontrer que les clitiques ne sont pas très différents des affixes. Ces derniers semblent en fait occuper certaines positions suivant des paramètres qui nous rappellent les paramètres responsables du placement des clitiques. Les paramètres de placement des affixes sont les suivants selon l'auteur :

1. L'affixe est localisé dans un domaine. Ce dernier peut être soit un constituant morphologique (la tête structurale du mot ou le mot entier), soit un constituant prosodique (le mot prosodique).
2. L'affixe est localisé en se référant à un élément {premier vs dernier vs accentué} d'un type donné dans le constituant dans lequel il apparaît.
3. L'affixe {précède vs suit} ce point de référence.

> There are very substantial similarities between the principles governing the placement of clitics and those for the placement of affixes : in fact, the two sets of parameters appear to be identical in their essence. Prefixes and suffixes obviously correspond to initial and final clitics. (Anderson, 1992, p. 210)

L'auteur pousse cette comparaison plus loin en nous rappelant que la déficience accentuelle unit les deux types d'affixes et que la typologie des clitiques (la distinction faite entre des clitiques simples et des clitiques spéciaux) ressemble beaucoup à la distinction bien connue entre les affixes dérivationnels et les affixes flexionnels.

Klavans (1983) et Anderson (1992) ne sont pas les seuls auteurs à parler des clitiques comme étant des affixes syntagmatiques. Nevis (1988) a proposé de remplacer le terme peu clair de « clitique » par deux notions distinctes : les mots conjoints et les affixes syntagmatiques. Zwicky (1987) reprend cette même distinction et démontre que le seul critère qui puisse distinguer un affixe syntagmatique d'un affixe flexionnel est leur position les uns par rapport aux autres : les affixes syntagmatiques sont toujours localisés à l'extérieur des affixes flexionnels. Ce critère n'est en réalité qu'une reformulation du test proposé par Zwicky et Pullum (1983) pour distinguer les affixes des clitiques : le critère F.

Nous pouvons conclure que la désignation des clitiques en tant qu'affixes syntagmatiques, bien qu'il nous permette de cerner certaines propriétés communes entre les affixes et les clitiques, ne peut refléter toutes les propriétés des clitiques eux-mêmes, d'autant plus qu'il exclut les clitiques des langues romanes (qui ne sont de toute évidence pas des affixes syntagmatiques). Nous pensons donc qu'il est possible de garder le terme *clitique*, puisque c'est un terme plus général qui embrasse la réalité de plusieurs langues. Toutefois, certains auteurs, dont Miller (1992), ont réexaminé les tests de Zwicky à la lumière de cette proposition pour en fin de compte les adapter aux données empiriques des langues étudiées.

4.1.3 Les tests de Miller

Dans une approche entièrement lexicaliste des clitiques sujets et objets du français, Miller (1992) discute des critères proposés par Zwicky et Pullum (1983) et propose d'autres critères plus appropriés au français. Selon l'auteur, le critère A (le degré de sélection) a perdu beaucoup de sa validité à la lumière des propositions de Nevis (1988) et de Zwicky (1987). Le fait que les clitiques sont désormais considérés comme des affixes syntagmatiques efface cette distinction entre les affixes et les clitiques sur la base de leur sélectivité. Les critères B, C et D (les trous arbitraires et les idiosyncrasies) ont, selon Miller, l'inconvénient de ne jamais rendre possible la décision en faveur du statut de clitique, puisque l'absence des trous arbitraires et des idiosyncrasies morphophonologiques et sémantiques ne rend pas le statut d'affixe possible, ni celui de clitique

nécessaire [16]. En fait, si la présence de telles idiosyncrasies peut constituer un argument pour le statut affixal, leur absence n'est pas concluante. L'auteur propose par ailleurs de se passer du critère D puisque, comme l'a démontré Labelle (1985), les expressions idiomatiques françaises en forme de SV contenant des clitiques (par exemple, *se la couler douce, l'échapper belle*, etc.) n'ont pas un statut différent des autres expressions idiomatiques en forme de SV qui contiennent des SN pleins. Il n'y a donc aucune raison de penser que les combinaisons [hôte+clitique] ont plus d'idiosyncrasies sémantiques que les combinaisons qui contiennent des SN pleins.

Le critère E est également discrédité par Miller. Selon ce dernier, le fait qu'une règle syntaxique peut s'appliquer à un groupe formé d'un hôte et d'un clitique ne peut distinguer le statut affixal du statut de clitique postlexical, à moins qu'il n'existe des raisons indépendantes de croire que le clitique et l'hôte se trouvent dans des constituants différents en syntaxe.

Finalement, le critère F est jugé trop faible en raison des contre-exemples fournis par Klavans (1985) : certains mots composés d'un radical et d'un affixe constituent des hôtes de clitiques.

En dépit de ces critiques, l'auteur garde les critères A, B, C et E et ajoute les critères suivants : l'haplologie, la coordination, la liaison, l'accent, l'ordre rigide et l'applicabilité des règles lexicales phonologiques.

1. L'haplologie [17] Ce phénomène consiste à n'exprimer qu'une seule fois des sons ou des groupes de sons identiques (ou partiellement identiques) qui se succèdent immédiatement. Selon Miller, les pronoms clitiques du français subissent cette règle, comme dans l'exemple suivant :

   (4.9) a. Il remplit un verre de ce vin.
   b. Il en (*en) remplit un.

   Ils sont mieux traités comme affixes pour cette raison. Le statut affixal s'explique par le fait que les éléments qui subissent une telle règle ne peuvent être que des affixes puisque ces derniers obéissent à un principe général : on ne peut avoir plus d'une seule réalisation affixale du même trait morphosyntaxique.

2. La coordination. Ce critère est dérivé du critère E de Zwicky et Pullum (1983), qui exclut deux configurations : une coordination de clitiques attachés à un hôte, et un clitique avec une large portée sur une coordination d'hôtes. Selon Miller, exclure la 1re configuration n'est pas empiriquement adéquat en français puisque certains affixes peuvent être coordonnés devant

[16] Voir également Labelle (1985)
[17] Voir également Zwicky (1987).

un hôte (exemples : *bi- et trihebdomadaire*, *pré- ou infra-vie*, etc.). C'est la 2e configuration qu'il retient donc, d'où les trois sous-critères suivants :

(a) Un item qui ne peut avoir une large portée sur une coordination d'hôtes ne peut être un clitique postlexical, mais un affixe. (Miller, 1992, p. 155)

(b) Un clitique postlexical peut être répété sur chaque conjoint de son hôte s'il forme avec son hôte un constituant syntaxique dans la phrase correspondante qui ne contient pas de coordination. (*Ibid.*, p. 156)

(c) Quand la répétition sur chaque conjoint est obligatoire, alors l'item en question est un affixe et ne peut être un clitique postlexical. (*Ibid.*, p. 157)

3. La liaison. La liaison obligatoire est un critère qui prouve l'attachement lexical en français. Dans cette langue, les prépositions monosyllabiques (telles que *dans* et *sans*) et les auxiliaires monosyllabiques (tels que *est*) n'exhibent pas de liaison obligatoire même s'ils sont les meilleurs candidats pour être considérés comme des clitiques postlexicaux.

4. L'accent. Si l'unité formée de l'item dont on veut connaître le statut et de son hôte fonctionne comme un mot eu égard aux règles de l'assignation de l'accent, alors cet item est un affixe et non un clitique postlexical.

5. L'ordre. L'ordre rigide est caractéristique des affixes et non des clitiques postlexicaux.

6. L'applicabilité des règles lexicales phonologiques. Si l'unité formée de l'item dont on veut connaître le statut et de son hôte subit des règles lexicales phonologiques typiques des combinaisons [radical+affixe], alors l'item en question est un affixe. Miller cite trois de ces règles : la formation des glissantes, la règle de Littré et la liaison de la consonne nasale [n].

Il faut noter que Miller (1992) établit des critères pour distinguer les affixes des clitiques postlexicaux. Il emploie le terme *clitique* dans un sens théorique précis absent des critères de Zwicky (du moins dans son ouvrage de 1977). Ce dernier employait le terme dans un sens qu'il qualifiait de préthéorique, c'est-à-dire sans émettre des hypothèses sur la nature lexicale ou syntaxique de ces unités. Miller, au contraire, l'emploie pour désigner des entités dépourvues d'accent autonome et formées dans une composante postlexicale. Les affixes sont selon lui des unités dépourvues d'accent autonome et formées dans la composante lexicale. De cette distinction découlent tous les critères qu'il retient dans son étude du statut morphologique des pronoms clitiques du français. Dans des travaux ultérieurs (par exemple, Miller et Monachesi (2003)), le terme préféré pour désigner les clitiques postlexicaux est celui de « clitiques » (sans le qualificatif).

En somme, les critères de Miller sont, dans leur majorité, une adaptation des critères de

Zwicky aux données empiriques du français. Ces critères ont constitué un cadre de travail pour d'autres auteurs qui ont étudié les clitiques du français comme Miller et Sag (1997, 1995), de l'italien comme Monachesi (1999b), du roumain comme Monachesi (1999b) et du polonais comme Kupsc (2000, 1999), pour ne citer que quelques exemples. Nous pouvons donc les utiliser dans l'étude du statut morphologique des marqueurs d'arguments de l'arabe tout en gardant à l'esprit qu'il faudrait les adapter aux particularités empiriques de cette langue. Dans la section qui suit, nous apportons des arguments basés sur ces critères pour prouver qu'il est préférable de traiter les marqueurs de sujet comme affixes et les marqueurs d'objet comme clitiques.

4.2 Le statut morphologique des marqueurs d'arguments

4.2.1 Prémisses

Avant d'exposer notre hypothèse concernant le statut morphologique des marqueurs d'argument, nous présentons des définitions de ce que nous entendons par les termes « clitique » et « affixe » en nous basant essentiellement sur les travaux de Anderson (2005), Auger (1995) et Miller (1992).

(4.10) a. Un clitique est un morphème qui n'est pas clairement identifiable comme un mot indépendant ou comme un affixe puisque ses propriétés sont un complexe composé des propriétés de l'un et de celles de l'autre.

b. Un affixe est un morphème qui fait partie d'un mot et qui n'est pas accessible aux règles syntaxiques. Tout comme le clitique, il n'est pas accentué de façon autonome.

La différence principale entre les clitiques et les affixes réside dans le fait que les clitiques constituent des nœuds syntaxiques alors que les affixes sont invisibles à la syntaxe et ne peuvent constituer de branches (ou filles) dans l'arbre. Cette différence présuppose l'adoption de l'hypothèse lexicaliste : la structure des mots et la structure syntagmatique sont gouvernées par des principes différents, ou autrement dit, la syntaxe n'a pas accès à la structure interne des mots [18].

Cette distinction entre les affixes et les clitiques donne un sens à l'affirmation selon laquelle les clitiques ont un statut intermédiaire entre celui des mots et celui des affixes [19]. Tout comme

[18] Voir Anderson (2005) et Sag *et al.* (2003).

[19] Halpern (1995, p. 1) définit les clitiques dans ces termes : « the term is used here as a rough-and-ready label for an item which is problematic for canonical approaches to defining a word or affix. »

les affixes, les clitiques ne reçoivent pas d'accent autonome et apparaissent toujours attachés à des hôtes phonologiques pour pallier cette déficience. Cependant, à la différence des affixes, mais tout comme s'ils étaient des mots, les clitiques sont sensibles aux contraintes syntaxiques. En outre, si les clitiques sont accessibles aux contraintes syntagmatiques et les affixes aux contraintes de formation des mots, alors on s'attendrait à ce que les combinaisons qui impliquent des affixes exhibent plus d'idiosyncrasies morphophonologiques et de phénomènes typiques à la phonologie lexicale que les combinaisons qui impliquent des clitiques.

Dans la section qui suit, nous nous basons sur ces définitions ainsi que sur les tests et critères de Zwicky et de Miller pour démontrer que les marqueurs d'arguments de l'arabe sont hétérogènes : les marqueurs de sujet sont mieux traités comme affixes alors que les marqueurs d'objet sont mieux traités comme clitiques. Les deux ne peuvent être considérés comme des mots pour les raisons que nous exposons dans la section suivante.

4.2.2 Les marqueurs d'arguments et les mots

Les marqueurs d'objet ont une distribution semblable à celle des mots indépendants à certains égards. Étant des pronoms objets, ils remplacent des SN et s'insèrent dans les mêmes positions que ces derniers. Dans les exemples suivants, les mêmes positions occupées par des SN en (4.11a) sont occupées par des marqueurs d'objet en (4.11b) :

(4.11) a. *ʔinna* *d- dali:l -a* *j-* *aqu:d* *-u* *n- na:s -a* *ʔila:*
que le- guide -ACC 3MS- guider.IMP -Ind les- gens -ACC vers
l- qas'r -i
le- palais -GÉN
« Le guide, il dirige les gens vers le palais, certes »

b. *ʔinna* ***-hu*** *j-* *aqu:d* *-u* ***-hum*** *ʔilaj* ***-hi***
que -lui 3MS- guider.IMP -Ind -les vers- lui
« Lui, il les y dirige, certes »

Toutefois, il n'est pas toujours possible de remplacer un SN par un marqueur d'objet, comme en témoignent les exemples suivants :

(4.12) a. *kull* *-a* *jawm -i -n*
chaque -ACC jour -ACC -Indé
« chaque jour »

a'. **kull* *-a* ***-hu***
chaque -ACC -le

b. *kita:b* *-u* *t'- t'ibb -i*
livre -NOM la- médecine -GÉN
« le livre de médecine »

b'. **kita:b* *-u* ***-hu***
livre -NOM -le

c. *raʔaj* *-tu* *ʔumm* *-a* *zajd* *-in*
voir.PER -1S mère -ACC Zayd -GÉN
« J'ai vu la mère de Zayd »

c'. **raʔaj* *-tu* *-ha:* *zayd* *-in*
voir.PER -1S -la Zayd -GÉN

Cette disparité de distribution démontre que les marqueurs d'objet, même s'ils présentent certaines propriétés caractéristiques des mots, restent différents de ces derniers. Ils présentent, en fait, certaines propriétés affixales qui nous permettent d'affirmer qu'ils ont un statut intermédiaire entre celui des mots et celui des affixes : leur ordre est très rigide, ils sont toujours attachés à des hôtes et ils ont une prosodie déficiente.

4.2.2.1 L'ordre rigide

L'ordre des mots est relativement libre en arabe grâce à la visibilité du cas morphologique. Les ordres possibles étant SVO, VSO, SOV, VOS, etc. L'ordre des marqueurs d'arguments est toutefois rigide. Cette rigidité se manifeste, en ce qui concerne les marqueurs de sujet, dans le fait qu'ils sont toujours attachés au radical verbal dans un site d'attachement immuable. Les marqueurs d'objet attachés aux verbes le sont à la combinaison [radical+marqueur de sujet] et aucun autre ordre n'est permis. Quand ils sont attachés aux autres catégories, ils apparaissent toujours postposés (hôte+marqueur d'objet). Finalement, quand plusieurs marqueurs d'objet sont attachés, leur ordre les uns par rapport aux autres est très rigide. Deux contraintes régissent cet ordre : la contrainte de la personne et la contrainte thématique.

(4.13) a. *La contrainte de la personne*

Les marqueurs d'objet sont ordonnés en commençant par la personne la plus proche du locuteur selon cette hiérarchie : 1 < 2 <3.

b. *La contrainte thématique*

Les marqueurs d'objet sont ordonnés comme suit : B < T (B : bénéficiaire, T : thème).

Ces deux contraintes s'appliquent de façon simultanée pour éliminer les suites mal formées, comme celles en (4.14), et accepter les suites bien formées, comme celles en (4.15) :

(4.14) a. **wahab* *-tu* *-hu* *-ka*
donner.PER -1S -le -te

b. **wahab* *-ta* *-hu:* *-ni:*
donner.PER -2MS -le -me

(4.15) a. *wahab* *-tu* *-ka* *-hu*
donner.PER -1S -te -le
« Je te l'ai donné »

b. *wahab* *-ta* *-ni:* *-hi*
donner.PER -2MS -me -le
« Tu me l'as donné »

Cette rigidité de l'ordre des marqueurs d'objet les rend plus susceptibles d'être des affixes ou des clitiques que des mots. Le même résultat est obtenu en appliquant le critère de l'attachement.

4.2.2.2 L'attachement

Selon Zwicky (1977), l'attachement phonologique est un critère décisif dans la distinction entre les affixes/clitiques et les mots. Les éléments morphologiques qui n'apparaissent que dans un état d'attachement à d'autres éléments ont toutes les chances d'être des affixes ou des clitiques, alors que les mots sont généralement libres.

Les marqueurs d'arguments de l'arabe sont toujours attachés à d'autres éléments (verbaux, nominaux, prépositionnels, etc.) et n'ont aucune existence dans la langue en tant qu'éléments libres. Ils sont donc plus susceptibles d'être des affixes ou des clitiques que d'être des mots.

4.2.2.3 La prosodie

Ce troisième critère a été souvent attaqué comme étant inadéquat dans la distinction entre les affixes/clitiques et les mots, puisque certains affixes/clitiques peuvent porter un accent et que

certains mots sont prosodiquement déficients, comme le mentionne à juste titre Klavans (1985). Nous prenons donc ce test seulement comme complément des deux premiers. Si les marqueurs d'arguments sont des affixes ou des clitiques selon ces derniers et qu'ils sont généralement dépourvus d'accent autonome, alors il faudrait les considérer comme des affixes ou comme des clitiques et non comme des mots. C'est justement le cas ici, puisque les marqueurs d'arguments ont toutes les propriétés qui définissent les affixes/clitiques et qu'ils sont généralement dépourvus d'accent.

Il faut toutefois noter que ces marqueurs peuvent porter un accent de par leur appartenance à un mot phonologique. En fait, suivant les règles d'assignation de l'accent en arabe, l'accent peut être assigné à un marqueur d'argument. Ces règles[20] sont les suivantes :

(4.16) Mettre l'accent sur la syllabe ultime quand cette dernière est très lourde (CVCC ou CVVC),
sinon sur la syllabe lourde (CVV ou CVC) non finale la plus à droite,
sinon sur la première syllabe.

Dans les exemples en (4.17), des marqueurs d'arguments sont accentués ou constituent le noyau de syllabes accentuées. L'accent est ainsi assigné de manière « aveugle » [21] quand la suite de syllabes formée de l'hôte et du marqueur fournit un contexte approprié à l'application de la règle de l'assignation de l'accent. En (4.18), par contre, les marqueurs d'arguments ne portent pas d'accent ou ne font pas partie de syllabes accentuées :

(4.17) a. *ki/ta:/bu* *-/**kún/na*** (marqueur d'objet)
livre-NOM -votre(FP)
« Votre livre »

b. *ka/tab* *-**tún/na*** (marqueur de sujet)
écrire.PER -3FP
« Vous(F) avez écrit »

(4.18) a. *ki/tá:/bu* *-**hu*** (marqueur d'objet)
livre-NOM -son
« son livre »

b. *ká/ta/b-**u:*** (marqueur de sujet)
écrire.PER-3MS
« Ils ont écrit »

[20] Voir McCarthy (1985, p. 127).
[21] Voir Klavans (1983).

Nous constatons que, dans les deux cas, l'hôte et le marqueur d'argument sont considérés comme faisant partie du même mot prosodique eu égard aux règles de l'assignation de l'accent. Ce comportement est caractéristique des affixes et des clitiques et non des mots indépendants.

L'ordre rigide, l'attachement et la prosodie des marqueurs d'arguments prouvent que ces derniers ne peuvent être considérés comme mots indépendants. Ils sont mieux traités comme affixes ou comme clitiques.

4.2.3 Le statut des marqueurs d'arguments

En nous basant sur l'hypothèse lexicaliste, de laquelle découlent plusieurs des critères de Zwicky et de Miller, nous démontrons que les marqueurs d'arguments sont hétérogènes : si les marqueurs de sujet sont mieux traités comme affixes, les marqueurs d'objet sont mieux traités comme clitiques. Les premiers présentent des propriétés qui ne sont compatibles qu'avec un statut affixal, alors que les derniers présentent des propriétés mixtes (d'affixes et de mots). Cette approche est appuyée par plusieurs tests : les idiosyncrasies et les trous arbitraires (section 4.2.3.1), la coordination (section 4.2.3.2) et l'appartenance catégorielle (section 4.2.3.3).

4.2.3.1 Les idiosyncrasies et les trous arbitraires

Selon Miller (1992) et Zwicky et Pullum (1983), les idiosyncrasies sont caractéristiques des combinaisons [radical+affixe]. Cette vision de la morphologie (et donc du lexique), comme étant le locus des irrégularités, est très caractéristique de la grammaire générative, dans le sens le plus large de ce terme. Les affixes, étant des éléments attachés dans le lexique, exhibent plusieurs irrégularités que les éléments introduits par des contraintes syntagmatiques n'exhibent pas. Parmi ces idiosyncrasies, Zwicky et Pullum (1983) ont signalé des idiosyncrasies morphophonologiques et sémantiques, ainsi que la présence des trous arbitraires dans l'ensemble des combinaisons possibles. Miller (1992) a, avec raison, fortement critiqué la validité des idiosyncrasies sémantiques et a également amoindri l'importance des trous arbitraires, puisque l'absence de ces derniers ne rend pas le statut de clitique [postlexical] obligatoire. Il a toutefois utilisé ce critère en conjonction à d'autres tests dans le but de démontrer le statut affixal des clitiques du français. Nous optons pour cette solution et nous exposons ici certains trous arbitraires qui caractérisent les combinaisons [hôte+marqueur d'argument], ainsi que certaines idiosyncrasies morphophonologiques. Ces trous

et ces idiosyncrasies supportent l'idée selon laquelle les marqueurs d'arguments présentent des propriétés typiques des affixes.

4.2.3.1.1 Les trous arbitraires

Parmi les trous arbitraires que nous pouvons recenser, nous en signalons deux sortes : des trous dans les relations entre les marqueurs d'arguments et leurs hôtes et des trous dans les relations entre les marqueurs d'arguments.

En ce qui a trait au premier type de trous, nous signalons que les marqueurs d'objet ne peuvent être attachés à certains hôtes, et cette interdiction ne peut être expliquée par un principe général. En fait, certaines prépositions ne peuvent être des hôtes de ces marqueurs. Ce sont les prépositions *ħatta:*, *ka*, *wa rubba*, *wa* du sermon, *ta* du sermon, *muð* et *munðu*[22] .

(4.19) a. **ħatta:* *-hu*
jusque -lui

b. **ka* *-hu*
comme -lui

Pourtant, ces prépositions peuvent être hôtes de clitiques non pronominaux, comme dans les exemples suivants :

(4.20) a. *ħatta:* *-ma*
jusque -quoi
« Jusqu'à quand ? »

b. *ka* *-ma:*
comme -cela
« Comme cela »

Elles peuvent également former des syntagmes prépositionnels avec des SN, comme dans les exemples suivants .

(4.21) a. *ħatta:* *l-* *masa:?* *-i*
jusque le- soir -GÉN
« Jusqu'au soir »

[22] Voir (Haysharii, 2003, p. 260-261).

b. *ka* *s-* *sajl* *-i*
comme le- fluide -GÉN
« comme le fluide »

Les combinaisons des radicaux avec des marqueurs de sujet connaissent certains trous arbitraires dans les formes possibles également. Tout comme en anglais, où aucun affixe du participe passé ne peut s'attacher au radical du verbe *stride*, nous constatons qu'en arabe aucun marqueur de sujet du perfectif ne peut s'attacher au radical dérivé de la racine verbale */wdʕ/* « laisser ». La conjugaison perfective n'étant pas possible pour ce verbe, un autre verbe (avec un sens proche) est employé : *taraka* « laisser ». Ce trou est d'autant plus surprenant que la conjugaison imperfective est possible (et par conséquent l'impératif également) : *j-adaʕ-u* « il laisse ». Un autre exemple est celui de la racine */wðr/* « laisser », dont on ne peut obtenir la forme perfective.

En ce qui a trait aux relations entre les marqueurs d'arguments, nous constatons que certaines suites ne sont pas possibles. Par exemple, si le premier de la suite est le marqueur d'objet *-ki* (2FS), aucun marqueur d'objet de 3e personne ne peut le suivre. Si le premier est le marqueur *-kunna* (2FP), le deuxième ne peut être *-hunna* (3FP). De plus, on ne peut avoir une suite dont le premier membre est *-hu* (3MS). Ces combinaisons proscrites ne peuvent être éliminées par des contraintes syntaxiques puisque les contraintes syntaxiques qui régissent les suites de marqueurs d'objet en arabe ne font pas référence à la forme phonologique de ces marqueurs. Cette tâche doit, par contre, être assumée par une composante lexicale de la grammaire.

4.2.3.1.2 Les idiosyncrasies morphophonologiques

Parmi les idiosyncrasies morphophonologiques qui caractérisent les composés [hôte+marqueur d'argument], nous citons les exemples apparents d'harmonie vocalique. En fait, ce qu'on pourrait croire être une harmonie vocalique authentique s'avère une idiosyncrasie dans les combinaisons de certaines prépositions avec des marqueurs d'objet. Dans les exemples suivants, les marqueurs d'objet de 3e personne dont le noyau vocalique est */u/* subissent un changement de ce noyau en */i/* en s'assimilant au noyau de la préposition *bi* « avec, par » :

(4.22) bi + -h**u** → bi-h**i** / *bi-h**u**
« avec lui »
bi + -h**u**m → bi-h**i**m / *bi-h**u**m
« avec eux »

bi + -huma: → bi-hima: / *bi-huma:

« avec eux (deux) »

bi + -hunna → bi-hinna / *bi-hunna

« avec elles »

Cependant, ce comportement est très singulier et ne consiste probablement pas en une harmonie vocalique. En fait, les marqueurs d'objet de 2e personne, qui sont phonologiquement très semblables aux marqueurs d'objet de 3e personne, ne subissent pas ce changement :

(4.23) bi + -kum → bi-kum / *bi-kim

« avec vous(M) »

bi + -kuma: → bi-kuma: / *bi-kima:

« avec vous (deux) »

bi + -kunna → bi-kunna / *bi-kinna

« avec vous(F) »

On ne peut formuler une règle d'harmonie vocalique qui tienne compte des traits morphosyntaxiques puisque les règles phonologiques sont censées être aveugles au contenu des items phonologiques. Nous pouvons cependant être amené à penser à cette règle en termes phonologiques en rendant notre règle d'harmonie sensible à la présence des phonèmes *h* (dans les marqueurs d'objet de 3e personne) et *k* (dans les marqueurs d'objet de 2e personne). Le problème que pose cette solution c'est que cette distinction n'est pas productive ailleurs dans la langue, l'arabe n'étant pas de toute façon une langue à harmonie vocalique[23]. Nous pouvons en conclure que le changement que subissent les marqueurs d'objet de 3e personne est sans doute idiosyncrasique.

Une autre idiosyncrasie morphophonologique s'observe avec le marqueur d'objet -*i*: (1S) quand ce dernier est attaché à certaines prépositions ou à des verbes. Dans ces deux contextes, une consonne *n* est insérée entre le marqueur et son hôte :

(4.24) a. Avec un verbe :

· laqij-a + -hu → laqij-a-hu

trouver.PER-3MS-le

« Il l'a trouvé »

[23] Les principales langues qui connaissent l'harmonie vocalique sont les langues agglutinantes de la famille ouralo-altaïque (le turc, le mongol, etc.), certaines langues bantoues, les langues nigéro-congolaises et le télougou.

· laqij-a + -i: → laqij-a-**ni**:
trouver.PER-3MS-me
« Il m'a trouvé »

b. Avec une préposition :

· min + -hu → min-hu
de-lui
« de lui »

· min + -i: → min-**ni**:
de-moi
« de moi »

· li + -i: → li: / *li-**n**-i:
à-moi
« à moi »

L'insertion du *n*, bien qu'elle soit productive avec des verbes, est idiosyncrasique en ce qui concerne les hôtes prépositionnels. De plus, cette insertion est optionnelle avec certaines particules comme dans les exemples suivants :

(4.25) a. kaʔanna + -i: → kaʔanna-**ni**: / kaʔann-i:
comme.si-moi
« Comme si j'étais »

b. laʕalla + -i: → laʕalla-**ni**: / laʕall-i:
peut.être-moi
« Je suis peut-être »

c. ʔinna + -i: → ʔinna-**n**-i: / ʔinn-i:
que-moi
« Je suis, certes, ... »

Ces idiosyncrasies, en plus des trous arbitraires, qui caractérisent les composés [hôte+marqueur d'argument] démontrent que ces marqueurs présentent certaines propriétés typiques des affixes. Le test de la coordination apporte un argument supplémentaire à cette affirmation et sera exposé dans le paragraphe suivant.

4.2.3.2 La coordination

La coordination est très révélatrice de la structure syntaxique ; elle dénote généralement une parité syntaxique entre les éléments coordonnés et peut donc constituer un critère crucial dans la décision concernant le statut des éléments qui entrent dans ce type de relation, ou qui ne peuvent le faire. Selon Zwicky (1977), les règles syntaxiques, l'effacement sous identité par exemple, n'affectent pas les affixes et peuvent affecter les mots. Miller (1992) récupère ce critère, en garde une interprétation en termes de portée (large vs restreinte), et le dissout en trois sous-critères :

A Un item qui ne peut avoir une portée large sur une coordination d'hôtes ne peut être un clitique postlexical, mais un affixe. (*Ibid.*, p. 155)

B Un clitique postlexical peut être répété sur chaque conjoint de son hôte s'il forme avec ce dernier un constituant syntaxique dans la phrase correspondante qui ne contient pas de coordination. (*Ibid.*, p. 156)

C Quand la répétition sur chaque conjoint est obligatoire, alors l'item en question est un affixe et ne peut être un clitique postlexical. (*Ibid.*, p. 157)

Selon le critère A, un affixe ne peut avoir une portée large sur une coordination d'hôtes, ou dans les termes de Zwicky (1977), un affixe ne peut être effacé sous identité. L'application de ce critère démontre que les marqueurs d'arguments se comportent comme les affixes à cet égard. Ce comportement caractérise aussi bien les marqueurs de sujet que les marqueurs d'objet, comme le montrent les contrastes suivants :

(4.26) a. *kita:b* *-u* ***-ha:*** *wa* *qalam* *-u* ***-ha:*** (marqueur d'objet)
livre -NOM -son et crayon -NOM -son
« son livre et son crayon »

b. **kita:b* *-u* ***-ha:*** *wa* *qalam* *-u*
livre -NOM -son et crayon -NOM

c. **kita:b* *-u* *wa* *qalam* *-u* ***-ha:***
livre -NOM et crayon -NOM -son

(4.27) a. *ʔakal* ***-tu*** *wa* *ʃarib* ***-tu*** (marqueur de sujet)
manger.PER -1S et boire.PER -1S
« J'ai mangé et j'ai bu »

b. **ʔakal* ***-tu*** *wa* *ʃarib*
manger.PER -1S et boire.PER

c. *ʔakal wa ʃarib -tu
manger.PER et boire.PER -1S

Cependant, les marqueurs d'objet et les marqueurs de sujet présentent une différence importante concernant la coordination : (4.26c) est une structure acceptable pour certains locuteurs. Malgré la résistance des grammairiens et des puristes de la langue, cette structure est très répandue dans les écrits et dans les médias et s'emploie en alternance avec des structures dans lesquelles des SN ont une portée large, comme en (4.28a) :

(4.28) a. *maqtal -u wa ʤarħ -u l- ʕaʃraːt -i*
mort -NOM et blessure -NOM les- dizaines -GÉN
« La mort et la blessure de dizaines [de gens] »

b. *maqtal -u l- ʕaʃraːt$_i$ -i wa ʤarħ -u -hum$_i$*
mort -NOM les- dizaines -GÉN et blessure -NOM -leur
« La mort et la blessure de dizaines [de gens] »

La structure acceptable d'un point de vue grammatical est (4.28b), mais (4.28a) fait partie de la grammaire de nombreux locuteurs de l'arabe moderne (surtout dans les médias). La grammaire de ces locuteurs permet donc à un marqueur d'objet d'avoir une portée large sur une structure de coordination, tout comme s'il était un mot indépendant. Il faut noter, en outre, que cette possibilité d'avoir une large portée sur une structure de coordination caractérise ce que Zwicky (1977) appelle les clitiques simples. Ces derniers peuvent porter sur deux éléments coordonnés comme dans les deux exemples[24] suivants de l'anglais :

(4.29) a. [Pat and Leslie]'ll be there.
b. [Pat and Leslie]'s father.

Les deux critères B et C de Miller reviennent à affirmer que les affixes doivent être obligatoirement répétés sur chaque conjoint alors que les clitiques peuvent l'être sous certaines conditions. Si nous considérons que la grammaire de l'arabe accepte aussi bien les structures en (4.26a) que celles en (4.26c), nous pouvons affirmer que les marqueurs d'objet se comportent à cet égard comme des clitiques simples. Les marqueurs de sujet n'ont, quant à eux, qu'une seule option : il faut les répéter sur leurs hôtes, ce qui constitue le *symptôme* d'un état affixal.

[24] Voir Miller et Sag (1997).

4.2.3.3 L'appartenance catégorielle

Selon Zwicky et Pullum (1983), la différence essentielle entre un affixe et un clitique réside dans les rapports qu'entretiennent ces deux unités avec leurs hôtes. Les affixes sont ainsi considérés comme plus sélectifs et n'apparaissent généralement qu'avec une classe limitée d'hôtes. Ce critère, même s'il perd de sa force dans les langues où les clitiques tendent à être des affixes syntagmatiques (tout aussi sélectifs de leurs hôtes syntagmatiques que les affixes ordinaires de leurs radicaux), garde son effet diagnostique dans les langues où les clitiques ne sont de toute évidence pas attachés à des syntagmes. C'est le cas en français, mais également en arabe. Dans cette dernière langue, les marqueurs de sujet sont très sélectifs de leurs radicaux et ne s'attachent qu'à des radicaux verbaux. Les marqueurs d'objet sont par contre très peu sélectifs et peuvent s'attacher à des verbes, à des nominaux, à des prépositions et à des complémenteurs ((4.30). Nous pouvons par conséquent considérer que les premiers présentent des *symptômes* compatibles avec un statut affixal alors que les deuxièmes présentent des *symptômes* compatibles avec un *diagnostic* de clitique.

(4.30) a. *raʔaj* *-tu* ***-ha:***
voir.PER -1S -la
« Je l'ai vue »

b. *kita:b* *-u* ***-ha:***
livre -NOM -son
« son livre »

c. *min* ***-ha:***
de -elle
« d'elle »

d. *ʔinna* ***-ha:***
que -elle
« Elle est certes ... »

Ce test comporte également une autre facette déjà contenue dans le critère 4 de Zwicky (1977) et que nous pouvons formuler ainsi : un clitique s'attache à un mot qui, autrement, peut apparaître indépendamment de la présence du clitique. Un clitique ne complète donc pas la forme morphologique d'un mot, alors que l'affixe en fait partie.

Prenant ce critère en considération, nous constatons que les marqueurs d'arguments de l'arabe sont de deux sortes · les marqueurs de sujet s'attachent à des radicaux qui, autrement, ne forment pas de mots complets d'un point de vue morphologique. Les marqueurs d'objet s'attachent à des mots déjà complets qui peuvent, autrement, apparaître sans ces marqueurs.

Cette différence entre les deux sortes de marqueurs d'arguments est très révélatrice : les marqueurs de sujet sont plus proches des affixes que ne le sont les marqueurs d'objet et ces derniers semblent plus susceptibles d'être des clitiques que des affixes.

4.3 Conclusion

Nous avons appliqué certains des tests de Zwicky et de Miller sur les marqueurs d'arguments de l'arabe pour identifier leur statut morphologique. Les tests que nous avons retenus sont les plus significatifs alors que d'autres sont inadéquats ou sans effet ; c'est le cas des tests de l'haplologie et de la liaison, par exemple, qui sont uniquement adéquats pour des langues qui connaissent ces phénomènes, comme le français.

Les critères ainsi retenus nous ont permis d'affirmer que les marqueurs d'arguments sont hétérogènes : certains sont des affixes (les marqueurs de sujet) alors que d'autres sont des clitiques postlexicaux (les marqueurs d'objet). Cette conclusion découle de plusieurs observations concernant le comportement de ces unités eu égard aux règles de l'assignation de l'accent, à l'ordre observable, à l'attachement et à la distribution.

Par ailleurs, nous pouvons désormais fournir un appui à l'hypothèse qui prône une distinction entre le statut morphologique et le statut morphosyntaxique. Les affixes ne sont en fait pas forcément des marqueurs d'accord (c'est le cas des marqueurs de sujet argumentaux) et les arguments ne sont pas forcément postlexicaux.

Cette hétérogénéité reflète les asymétries des marqueurs d'arguments que nous avons présentées dans le deuxième chapitre. Ces asymétries trouvent ainsi un début d'explication. En outre, nous sommes désormais en mesure de remplir le tableau (4.3), qui résume nos hypothèses quant aux statuts morphosyntaxique et morphologique des marqueurs d'arguments de l'arabe.

Nous pouvons désormais désigner les marqueurs de sujet argumentaux par le terme « pronoms affixes » et les marqueurs d'objet par le terme « pronoms clitiques ».

Comment ces éléments doivent-ils être traités dans une grammaire monostratale comme HPSG ? Quels principes régissent leur distribution et leur placement ? Nous apportons des réponses à ces questions dans le cinquième chapitre.

	Marqueurs d'accord	Arguments	Explétifs, topiques
Affixes	- Marqueurs de sujet co-indicés avec un sujet lexical.	- Marqueurs de sujet co-indicés avec un élément disloqué. - Marqueurs de sujet employés seuls.	
Clitiques		- Marqueurs d'objet compléments de têtes lexicales. - Marqueurs d'objet compléments de têtes fonctionnelles introduisant des phrases équatives.	- Marqueurs d'objet compléments de têtes fonctionnelles introduisant des phrases verbales.

TABLEAU 4.3: Le statut des marqueurs d'arguments

CHAPITRE V

LA MODÉLISATION DES MARQUEURS D'ARGUMENTS

Introduction

Les deux chapitres précédents nous ont permis d'établir les liens qui unissent les marqueurs d'arguments et leurs hôtes en morphologie et en syntaxe. Ici nous traduisons l'analyse linguistique défendue dans ces chapitres dans les termes du formalisme et de la théorie linguistique proposés dans le cadre de HPSG.

Nous avons proposé de traiter les marqueurs de sujet, ceux de la 3S, comme appartenant à deux classes d'éléments d'un point de vue morphosyntaxique : des pronoms (arguments) et des marqueurs d'accord. En termes morphologiques, ces deux classes sont des affixes et font partie intégrante de leurs hôtes verbaux.

En ce qui a trait aux marqueurs d'objet, nous avons proposé de les traiter comme arguments ou comme explétifs et topiques en morphosyntaxe et comme clitiques postlexicaux en morphologie. Les marqueurs d'objet arguments sont des compléments de verbes, de noms et de prépositions et de têtes fonctionnelles introduisant des phrases équatives. Les marqueurs d'objets compléments de têtes fonctionnelles introduisant des phrases verbales ne sont cependant pas des arguments.

Cette analyse linguistique des marqueurs d'arguments est modélisé dans ce chapitre en s'inspirant de l'approche de Miller et Sag (1997) et en s'inscrivant dans le cadre de la grammaire HPSG telle que conçue par Ginzburg et Sag (2000), Pollard et Sag (1994, chapitre IX) et Sag *et al.* (2003).

Ce chapitre est organisé comme suit : nous proposons, dans la première section, une modélisation des marqueurs d'arguments. Nous commençons, d'abord, par aborder certains problèmes communs aux deux types de marqueurs et nous explicitons notre conception de la structure syntag-

matique de l'arabe standard. Nous consacrons, ensuite, une première sous-section aux marqueurs d'objet après avoir passé en revue les données empiriques déjà discutées dans les chapitres II, III et IV. Nous consacrons, enfin, une deuxième sous-section aux marqueurs de sujet en prenant soin de préciser le domaine empirique auquel s'applique la théorie.

La deuxième grande section de ce chapitre est dédiée à l'implémentation informatique de l'analyse et du formalisme adoptés afin d'en tester la validité dans un système de performance. Nous présentons l'environnement de développement grammatical choisi, à savoir LKB, et nous montrons comment notre grammaire est capable de reconnaître les séquences bien formées contenant des marqueurs d'arguments et de rejeter celles qui sont mal formées. Cette grammaire est également capable de générer des suites grammaticales et rien d'autre que des suites grammaticales.

5.1 Modélisation

La grammaire que nous proposons est lexicaliste en ceci qu'elle réduit le nombre des règles syntagmatiques au profit d'un lexique riche et hiérarchisé. Aussi démontrons-nous qu'en adoptant une analyse à la base de laquelle se trouve le lexique avec ses entrées lexicales, ses hiérarchies de types et ses contraintes sur ces types, nous pouvons convenablement modéliser le comportement des marqueurs d'arguments de l'arabe. Nous démontrons également que, dans sa forme actuelle, la théorie HPSG dispose des outils théoriques et formels nécessaires pour aborder les questions qui touchent la famille sémitique, telle la morphologie non concaténative. Les principes universels reconnus dans ce cadre ne perdent pas leur universalité en abordant les phénomènes syntaxiques et morphosyntaxiques de l'arabe. Seul le lexique demeure le locus de la spécificité.

Cette analyse lexicaliste des marqueurs d'arguments repose, en une grande part, sur la richesse des informations contenues dans les entrées lexicales des mots, eux-mêmes organisés dans une hiérarchie de types. La structure syntagmatique est analysée selon le même modèle, en optant non pour des règles syntaxiques, mais pour une hiérarchie de types syntagmatiques conjuguée à des contraintes de bonne formation imposées à ces types.

5.1.1 Les types syntagmatiques

Nous proposons, suivant Alexopoulou et Kolliakou (2002), Ginzburg et Sag (2000), et Sag *et al.* (2003), de décrire les syntagmes non en termes de schémas de dominance immédiate, qui ne

sont en fin de compte que l'expression de règles syntagmatiques, mais comme appartenant à un type faisant partie d'une hiérarchie. Les types syntagmatiques de l'arabe peuvent être représentés dans la figure (5.1) [1] :

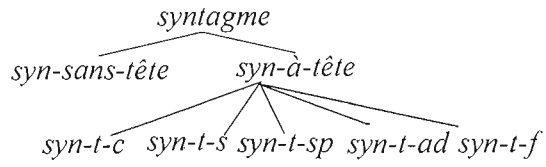


FIGURE 5.1: Une hiérarchie des types syntagmatiques

Au sommet de cette hiérarchie se trouve le type *syntagme*. Ses deux sous-types immédiats sont *syn-sans-tête*, qui couvre les structures avec coordination [2], et *syn-à-tête*, qui couvre les syntagmes dotés de têtes. Ce dernier type possède 5 sous-types selon l'élément avec lequel la tête entre en relation : un complément (*syn-t-c*), un sujet (*syn-t-s*), un spécifieur (*syn-t-sp*), un adjoint (*syn-t-ad*) ou un « filler » (*syn-t-f*). Chaque type est associé à des contraintes spécifiques que les sous-types héritent de façon monotone ou par défaut de leur super-type. Ainsi, tous les syntagmes à tête héritent la contrainte imposée au super-type immédiat, à savoir l'appropriation de l'attribut FILLES [FILLES *filles*]. Les contraintes sur les types *syn-t-c*, *syn-t-s*, *syn-t-sp*, *syn-t-ad* et *syn-t-f* seront explicitées au besoin.

Cette hiérarchie permet d'exprimer les contraintes sur les mots, les lexèmes, les règles lexicales et les syntagmes dans un système formel unifié. Ce formalisme nous permettra d'aborder plusieurs questions qui touchent l'analyse des marqueurs d'arguments de l'arabe, dont la question de l'ordre des mots. Dans la mesure où cette question touche à la fois les marqueurs de sujet et les marqueurs d'objet, nous y consacrons la sous-section suivante.

5.1.2 L'ordre des mots en arabe

En surface, plusieurs ordres de mots sont possibles en arabe. Cependant, deux sont principaux : VSO (le verbe est initial) et l'ordre dit SVO (le verbe n'est pas initial). Nous proposons, dans cette section, une modélisation de notre analyse de ces deux ordres en arabe standard. Il s'avère que l'ordre VSO est l'ordre non marqué et qu'il peut être décrit comme un syntagme ap-

[1] Cette hiérarchie est une adaptation de celle qui est proposée par Ginzburg et Sag (2000).

[2] Il va de soi que la question de savoir si les syntagmes à coordination comportent une tête ou non reste ouverte ici. Nous nous prononçons pas sur cette question puisqu'elle dépasse le cadre de notre travail.

partenant au type syntagmatique *syn-t-s* (syntagme-tête-sujet). L'ordre dit SVO n'est par contre que l'expression de phénomènes syntaxiques bien connus sous le nom de dislocation à gauche et peut par conséquent être décrit en faisant appel au type *syn-t-f* (syntagme-tête-filler), qui remplace, dans cette analyse, le recours au schéma DI6. Rappelons que ce schéma, déjà présenté dans la section (1.3.6) du premier chapitre, décrit les syntagmes qui comportent des relations à longue distance (ou des dépendances non bornées), telles que la dislocation et la topicalisation.

5.1.2.1 L'ordre VSO

Nous considérons que l'ordre non marqué des mots en arabe standard est l'ordre VSO, comme l'illustre l'exemple suivant :

(5.1) *katab* *-a* *l-* *walad* *-u* *risa:lat* *-a* *-n*
écrire.PER -3M le- garçon -NOM lettre -ACC -Indé
« Le garçon a écrit une lettre »

Cette phrase est la réponse que l'on obtient en posant la question « qu'est-il arrivé ? ». N'étant pas pragmatiquement marquée, cette phrase ne peut servir pour mettre l'emphase sur le sujet par exemple. Cette mise en relief est cependant possible en antéposant le SN nominatif dans une structure que nous qualifions de dislocation à gauche.

Nous optons pour une analyse non dérivationnelle. L'ordre VSO n'est pas dérivé de l'ordre dit SVO, comme le propose Moosally (1994). Certes, le lien entre les deux est indéniable, mais nous pouvons concevoir ce lien sans recourir à la dérivation[3]. Nous n'optons également pour aucune inversion dans l'ordre VSO, comme le laisse entendre cette auteure. Nous pensons que l'ordre de surface VSO peut convenablement être décrit par un syntagme de type *syn-t-s* et que la contrainte suivante est imposée à ce type syntagmatique en arabe :

[3] De toute façon, si dérivation il y a, l'ordre dit SVO serait dérivé de l'ordre VSO en arabe et non l'inverse puisque le premier comporte une valeur marquée absente du second.

(5.2) $\begin{bmatrix} \textit{syn-t-s} & \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & / \left\langle \boxed{2} \right\rangle \\ \text{COMPS} & \left\langle \ldots \right\rangle \end{bmatrix} \\ \text{STR-ARG} & \left\langle \boxed{2}, \ldots \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix}$

Dans cette contrainte, le sujet est lexicalisé par défaut, mais il est requis dans la structure argumentale de la tête verbale. Cela laisse la place à la possibilité de réaliser le sujet comme affixe. Le sujet affixal, à l'encontre du sujet lexical, ne constitue pas un nœud syntaxique et ne peut donc être encodé comme valeur de l'attribut SUJ. Il fait cependant partie des membres de la liste STR-ARG puisqu'il est un argument. Cette possibilité est illustrée dans l'exemple suivant, où la phrase en (5.3b) constitue une réponse possible à la question posée en (5.3a) :

(5.3) a. *ma:ða: faʕal -a l- ʔawla:d -u ʔ*
que faire.PER -3M les- garçons -NOM
« Qu'ont fait les garçons ? »

b. *xaraɢ -u:*
sortir.PER -3MP
« Ils sont sortis. »

La question de l'accord entre le verbe et le sujet reste à clarifier à la lumière de cette modélisation, ce que nous ferons dans la section (5.1.4.4) en nous basant sur la modélisation des marqueurs de sujet que nous proposerons dans la section (5.1.4).

5.1.2.2 L'ordre dit SVO

L'ordre qui paraît en surface comme étant SVO n'est en réalité qu'une dislocation à gauche. Cette dislocation est exemplifiée dans les phrases suivantes :

(5.4) a. *ʔal- walad -u katab -a risa:lat -a -n*
le- garçon -NOM écrire.PER -3MS lettre -ACC -Indé
« Le garçon, il a écrit une lettre »

b. *ʔal- walad -u raʔaj -tu -hu*
le- garçon -NOM voir.PER -1S -le
« Le garçon, je l'ai vu »

L'élément disloqué à gauche, ici le SN *ʔalwaladu* « le garçon », est associé à un pronom argumental attaché à la tête verbale. Ce pronom est un marqueur de sujet en (5.4a) et un marqueur d'objet en (5.4b). Nous proposons de traiter ces deux structures, avec marqueur de sujet et avec marqueur d'objet, comme étant des syntagmes de type *syn-t-f* et impliquant donc une forme de dépendance à longue distance (ou non bornée).

Nous basons notre analyse sur celle qui a été proposée par Alexopoulou et Kolliakou (2002) pour rendre compte des phénomènes appelés CLLD (pour Clitic-Left Dislocation) en grec. Nous n'optons cependant pas pour une analyse qui considère que le pronom dans ces structures est toujours un affixe étant donné que seuls les marqueurs de sujet le sont comme nous l'avons démontré dans le chapitre IV.

Pour expliquer les fondements de cette analyse, nous explicitons ici la théorie adoptée dans le cadre de la grammaire HPSG pour rendre compte des phénomènes de dépendance non bornée, essentiellement par Ginzburg et Sag (2000) et Miller et Sag (1997). Dans cette théorie, les arguments de la tête peuvent être réalisés de différentes manières :

- Les arguments peuvent être réalisés localement comme sœurs de la tête, auquel cas la valeur de l'attribut VAL et celle de l'attribut STR-ARG sont identiques. Ces arguments sont dits canoniques dans ce cas-ci, comme le proposent Miller et Sag (1997).
- Les arguments non réalisés ou réalisés en tant que dépendants non locaux de la tête sont dits non canoniques. Ils peuvent de ce fait être des *gaps* ou des *affixes*, auquel cas ils n'apparaissent que dans la liste STR-ARG.

Alexopoulou et Kolliakou (2002) partent de cette théorie et proposent d'encoder la relation entre le pronom clitique attaché au verbe et l'élément disloqué à gauche en grec en faisant appel à un attribut non local, équivalent à l'attribut SLASH utilisé pour rendre compte des phénomènes d'extraction dans la grammaire HPSG standard. Cet attribut approprié aux objets de type *non-local* est CL[4] (pour clitique). Ainsi, l'affixe argumental partage avec l'élément disloqué à gauche (le « filler ») la valeur de cet attribut. Cela les amène à proposer la contrainte suivante sur les objets de type *affixe* :

[4] Les auteurs insistent sur le fait que les affixes et les « gaps » ne doivent en aucun cas être traités de la même manière compte tenu des différences qui les distinguent. Parmi ces dernières, citons le fait que les affixes (argumentaux) peuvent apparaître dans des phrases qui ne contiennent pas de « filler », alors que les « gaps » ne sont légitimés qu'en présence d'un autre élément, le « filler ».

(5.5)
$$\begin{bmatrix} \textit{affixe} & \\ \text{CAT} \mid \text{TÊTE} & \boxed{1}\begin{bmatrix} \textit{nominal} & \\ \text{CAS} & \textit{cas} \\ \text{INDEX} & i \\ \text{ACCORD} & \begin{bmatrix} \text{PERS} & \textit{pers} \\ \text{NBRE} & \textit{nbre} \\ \text{GRE} & \textit{gre} \end{bmatrix} \end{bmatrix} \\ \text{CONT} & \begin{bmatrix} \textit{pronom} & \\ \text{INDEX} & i \end{bmatrix} \\ \text{CL} & \left\{ (\boxed{1}) \right\} \end{bmatrix}$$

La valeur de l'attribut CL d'un argument affixal contient optionnellement un seul objet partagé avec la valeur de l'attribut TÊTE de l'affixe. Ce dernier est de type *nominal* et porte de ce fait l'attribut CAS. Suivant Sag et Wasow (1999), les auteures emploient également un attribut de tête ACCORD (qui encode les valeurs de nombre, de genre et de personne). L'attribut INDEX, inclus dans les traits de tête de l'affixe, sert à rendre compte du partage du rôle thématique entre le « filler » et l'affixe pronominal. Les auteures unifient ainsi la valeur de cet attribut à celle de l'attribut INDEX (approprié aux objets de type *cont*(enu)) puisque ce dernier est l'attribut impliqué dans l'assignation des rôles thématiques en HPSG selon ce que proposent Pollard et Sag (1994).

Étant un trait non local, CL est sujet à la contrainte de propagation des traits non locaux (Non-Local Amalgamation Constraint) de Ginzburg et Sag (2000). Cette contrainte définit la valeur des traits non locaux des mots en fonction de la valeur des traits non locaux de leurs arguments (contenus dans la liste STR-ARG) de la manière suivante : la valeur de chaque trait non local F d'un mot est égale à la somme des valeurs des traits non locaux des arguments que ce mot sélectionne. Cette contrainte, imposée aux objets de type *slash* et *cl*, est une contrainte violable et peut donc être outrepassée par d'autres contraintes propres à certains mots.

(5.6) Pour chaque trait non local F
$$\begin{bmatrix} \textit{mot} & \\ / \text{ SYNSEM} & \begin{bmatrix} \text{F} & \sum_1 \cup \ldots \cup \sum_n \end{bmatrix} \\ / \text{ STR-ARG} & \left\langle [\text{F } \sum_1], \ldots, [\text{F } \sum_n] \right\rangle \end{bmatrix}$$

En outre, la propagation des traits non locaux de la fille à la mère, qui est en réalité une unification de ces traits, s'opère grâce à un deuxième principe universel, le principe généralisé des traits de tête, tel que nous l'avons défini dans la section (1.3.5.4) : la valeur de l'attribut SYNSEM de la mère d'un syntagme à tête est identique, par défaut, à la valeur de l'attribut SYNSEM de la fille.

Compte tenu des différences entre les deux types non canoniques *slash* et *cl*, ces objets apparaissent dans deux sortes de syntagmes à tête que les auteures nomment respectivement *head-filler-phrase* (*syn-t-f*) et *clitic-left-dislocation-phrase* (*syn-clld*). Ces types sont sous la jougue des contraintes suivantes, qui défont l'effet du principe généralisé des traits de tête. La contrainte sur le type *syn-t-f* impose le partage des traits locaux entre le « gap » et le « filler », alors que la contrainte sur le type *syn-clld* impose le partage des traits de tête entre l'élément disloqué et le *clitique*, ce qui inclut dans cette théorie la valeur de l'attribut INDEX (imposant ainsi la coréférentialité).

(5.7)
$$\begin{bmatrix} \textit{syn-t-f} & \\ \text{SLASH} & \sum 2 \\ \text{LOC} & \boxed{1} \\ \text{FILLES} \mid \text{FILLE-TÊTE} & \begin{bmatrix} \textit{syntagme} & \\ \text{TÊTE} & \textit{verbe} \\ \text{SLASH} & \left\{\boxed{1}\right\} \uplus \sum 2 \end{bmatrix} \end{bmatrix}$$

(5.8)
$$\begin{bmatrix} \textit{syn-clld} & \\ \text{CL} & \sum 2 \\ \text{TÊTE} & \boxed{1} \\ \text{FILLES} \mid \text{FILLE-TÊTE} & \begin{bmatrix} \textit{syntagme} & \\ \text{TÊTE} & \textit{verbe} \\ \text{CL} & \left\{\boxed{1}\right\} \uplus \sum 2 \end{bmatrix} \end{bmatrix}$$

En arabe, nous pouvons considérer que les syntagmes dans lesquels un SN préverbal disloqué est associé à un marqueur de sujet ou à un marqueur d'objet sont des instances du type *syn-clld* et sont donc régis par la contrainte en (5.8). Cette contrainte exige que le « filler » (l'élément disloqué) et le pronom attaché au verbe partagent la valeur de l'attribut TÊTE, à savoir leur cas,

l'index et les traits d'accord. Le cas est toutefois sous-spécifié puisque le SN préverbal disloqué est toujours nominatif alors que le pronom peut être nominatif (dans le cas des marqueurs de sujet) ou non nominatif (dans le cas des marqueurs d'objet). Le partage de l'index assure que les deux éléments (le pronom et le SN) réfèrent à la même entité et le partage des traits d'accord assure la relation d'accord anaphorique dans les termes de Bresnan et Mchombo (1987).

Après avoir présenté la théorie syntagmatique qui sous-tend notre approche et analysé les structures à verbe initial et celles où le verbe est précédé d'un SN disloqué[5], nous consacrons les sections subséquentes à la modélisation de phénomènes propres à chacun des deux types de marqueurs. Nous commençons par les marqueurs d'objet en adoptant une approche lexicaliste.

5.1.3 Les marqueurs d'objet

5.1.3.1 Rappel des données

Les faits suivants constituent le domaine empirique de référence en ce qui a trait à la modélisation des marqueurs d'objet de l'arabe :

1. Ces marqueurs ont une phonologie déficiente en raison de laquelle ils apparaissent toujours attachés à des hôtes forts.
2. Ils ont un statut morphologique de clitiques et s'insèrent donc dans des nœuds syntaxiques autonomes (des filles).
3. Ils ont un statut morphosyntaxique d'arguments, d'explétifs ou de topiques.
4. Ils peuvent faire partie d'une suite d'au plus 2 marqueurs d'objet avec les verbes de don, de type *donner, offrir*, etc. Dans ces suites, l'ordre est régi par la contrainte de personne et la contrainte thématique.
5. Certaines différences distributionnelles existent entre les marqueurs d'objet ACC et les marqueurs d'objet GÉN. Les premiers peuvent être coordonnés, mais pas les deuxièmes.

[5] Les structures avec un pronom indépendant nominatif disloqué sont traitées suivant la même analyse que l'on adopte pour les SN disloqués.

5.1.3.2 La hiérarchie des types

Nous adoptons une analyse lexicaliste à la base de laquelle se trouve le lexique avec une hiérarchie des types et les entrées lexicales auxquelles sont associés les marqueurs d'objet. Nous présentons la hiérarchie de types partielle comportant ces marqueurs dans la figure (5.2).

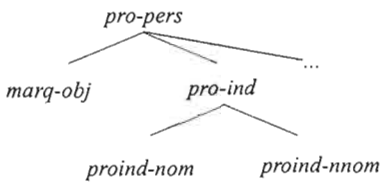


FIGURE 5.2: Une hiérarchie partielle des types pronominaux

Dans cette hiérarchie, le type racine est *pro-pers* (pour pronom personnel), lui-même sous-type de *pronom*. Deux contraintes sont imposées aux objets de type *pron-pers*, à savoir [TÊTE|ACCORD|PERS][6] et [TÊTE|CAS]. PERS est un attribut à valeurs atomiques parmi les suivantes : *1*, *2* ou *3*. CAS est un attribut qui peut prendre pour valeurs *nom*(inatif) et *n-nom* (non nominatif).

$$(5.9)\quad \begin{bmatrix} \textit{pro-pers} & \\ \text{ACCORD} & \begin{bmatrix} \text{PERS} & \textit{pers} \end{bmatrix} \\ \text{CAS} & \textit{cas} \end{bmatrix}$$

Les pronoms personnels à l'étude ici sont de deux types : *marq-obj* (pour les marqueurs d'objet) et *pro-ind* (pour les pronoms indépendants)[7]. Ces deux sous-types héritent de leur super-type les deux contraintes déjà mentionnées, mais une contrainte supplémentaire est imposée à chacun de ces deux types. Ainsi, *marq-obj* est spécifié [FORM[8] *enclitique*] alors que *pro-ind* est spécifié [FORM *libre*]. Les objets de type *enclitique* ont une prosodie déficiente et s'attachent à droite de leurs hôtes en phonologie alors que les objets de type *libre* sont phonologiquement indépendants.

[6] Cela distingue les pronoms personnels des pronoms démonstratifs, par exemple, qui ne sont pas spécifiés pour la personne tout en encodant le genre et le nombre en arabe. La contrainte casuelle se justifie également lorsque nous comparons les pronoms personnels, qui changent généralement de forme en changeant de cas, aux pronoms démonstratifs qui gardent la même forme quelle que soit leur cas morphologique.

[7] Ce dernier domine deux sous-types *proind-nom* et *proind-nnom*.

[8] Cet attribut sert à modéliser la forme morphophonologique d'un objet linguistique donné. Il va de soi que les différences entre les marqueurs d'objet et les pronoms indépendants ne sont pas que morphophonologiques. Dans une modélisation complète du système pronominal de l'arabe, il faudrait également prendre en compte les différences distributionnelles entre ces deux sortes de pronoms. Cette tache dépasse le cadre de notre thèse, mais pourrait être accomplie en se basant sur la description et l'analyse linguistique fournies dans les chapitres II et III et sur la modélisation faite dans le présent chapitre.

5.1.3.3 Les marqueurs d'objet argumentaux

Pour illustrer l'effet de ces contraintes sur le type *marq-obj*, nous prenons un exemple de syntagme prépositionnel où le complément de la préposition est un marqueur d'objet : *bi-ha:* « avec elle ». La préposition *bi* est associée à l'entrée lexicale en (5.10) et le marqueur d'objet *-ha:* à l'entrée en (5.11). L'unification de ces deux structures de traits donne lieu à la structure partielle en (5.12), qui décrit le syntagme prépositionnel *bi-ha:*, de type *syn-t-c* (syntagme-tête-complément). Dans cette structure, la disjonction est marquée par le symbole V.

(5.10) L'entrée lexicale de la préposition *bi*

$$\begin{bmatrix} \textit{préposition} & \\ \text{PHON} & < \textit{bi} > \\ \text{TÊTE} & \textit{préposition} \\ \text{STR-ARG} & <\text{SN}[\textit{n-nom}]\ \text{V}\ \textit{marq-obj}[\textit{n-nom}]> \end{bmatrix}$$

(5.11) L'entrée lexicale de *-ha:*1

$$\begin{bmatrix} \textit{marq-obj} & \\ \text{PHON} & < \textit{ha:} > \\ \text{TÊTE} & \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fém} \\ \text{NBRE} & \textit{sing} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{CAS} & \textit{n-nom} \end{bmatrix} \\ \text{FORM} & \textit{enclitique} \end{bmatrix}$$

(5.12)
$$\begin{bmatrix} \textit{syn-t-c} & \\ \text{PHON} & \langle\, \textit{bi-ha:} \,\rangle \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{TÊTE}\boxed{1} & \textit{préposition} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} & \left\langle \boxed{1} \right\rangle \\ \text{FILLES-COMPS} & \left\langle \begin{bmatrix} \textit{marq-obj} & \\ \text{TÊTE} & \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fém} \\ \text{NBRE} & \textit{sing} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{CAS} & \textit{n-nom} \end{bmatrix} \\ \text{FORM} & \textit{enclitique} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

La préposition *bi* est spécifiée dans le lexique comme devant sélectionner un seul argument non nominatif qui peut se réaliser sous la forme d'un SN [9] ou d'un marqueur d'objet. Cette spécification permet à cette préposition de sélectionner le marqueur d'objet *-ha:* comme complément, ce qui donne lieu à un syntagme prépositionnel dont la fille tête est la préposition et la fille complément est le marqueur d'objet. La relation d'attachement entre les deux est rendue par l'attribut FORM qui précise le site d'attachement du marqueur.

5.1.3.4 Les marqueurs d'objet non argumentaux

Les marqueurs d'objet peuvent, par ailleurs, être des explétifs ou des topiques. Cela est possible quand ils sont attachés à des têtes fonctionnelles et plus précisément à des complémenteurs. Ces derniers sont introduits dans le cadre de la théorie HPSG par l'attribut MARQUEUR. Dans ce cadre, le complémenteur arabe *?inna* [10], par exemple, peut être associé à l'entrée lexicale suivante :

[9] Nous employons cette abréviation pour référer à une structure de traits de type nominal dont la tête est un nom. Les contraintes imposées à ce genre de structure de traits devraient faire partie d'une théorie complète et détaillée portant sur ce genre de syntagmes en arabe.

[10] Ce complémenteur peut sélectionner deux sortes de phrases en arabe : des phrases verbales (modélisées par le type *ph-verb*) spécifiées [TÊTE *verbe*] et des phrases nominales (modélisées par le type *ph-nom*) spécifiées [TÊTE *nominal*]. Ces deux types sont des sous-types de *phrase*, lui-même sous-type de *syntagme*.

(5.13) $\begin{bmatrix} \text{TÊTE} & \begin{bmatrix} \textit{marqueur} & \\ \text{SPEC} & \textit{phrase[non-marquée]} \end{bmatrix} \\ \text{VAL} & \langle\rangle \\ \text{MARQUAGE} & \textit{ʔinna} \end{bmatrix}$

Ce complémenteur sélectionne comme spécifieur une phrase elle-même non introduite par un autre complémenteur. Le syntagme qui en résulte est de type *syn-t-sp* (syntagme-tête-spécifieur).

Pour rendre compte du fait que les deux marqueurs d'objet *-hu* (3MS) et *-ha:* (3FS) s'emploient dans deux sortes différentes de structures, à savoir comme arguments (compléments de verbe, de préposition, de nominal ou de complémenteur dans les phrases équatives) et comme non arguments (attachés à des complémenteurs, comme topiques et comme explétifs, dans les phrases verbales), nous proposons une modélisation selon laquelle chacun de ces deux marqueurs est associé à deux entrées lexicales différentes. Dans la première, celle en (5.11), ces marqueurs sont spécifiés pour tous leurs traits d'accord (genre, nombre et personne). Dans la deuxième, ces pronoms ne sont spécifiés que pour la personne et le genre, comme en témoignent les exemples suivants [11] :

(5.14) a. *ʔinna -ha: daxal -at leila*
que -elle entrer.PER -3F Leila
« Le fait est que Leila est entrée »

b. *ʔinna -hu daxal -a l- ʔawla:d -u*
que -il entrer.PER -3M les- garçons -NOM
« Le fait est que les garçons sont entrés »

Ainsi, l'entrée lexicale de l'explétif *-ha:* en (5.14a) prend la forme suivante :

(5.15) L'entrée lexicale de *-ha:*2 $\begin{bmatrix} \textit{marq-obj} & \\ \text{TÊTE} & \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fem} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{CAS} & \textit{n-nom} \end{bmatrix} \\ \text{FORM} & \textit{enclitique} \end{bmatrix}$

Nous modifions ainsi la hiérarchie des types pour inclure deux sous-types de *marq-obj*, à savoir *mo-comp* (marqueurs-d'objet-complets) et *mo-par* (marqueurs-d'objet-partiels). Le premier est

[11] Voir la section (3.2.3) dans le chapitre III pour plus de détails.

spécifié pour les traits de tête NBRE, GRE et PERS, alors que le deuxième est spécifié pour deux attributs seulement, GRE et PERS. La hiérarchie présentée dans la figure (5.3) inclut ces nouveaux types.

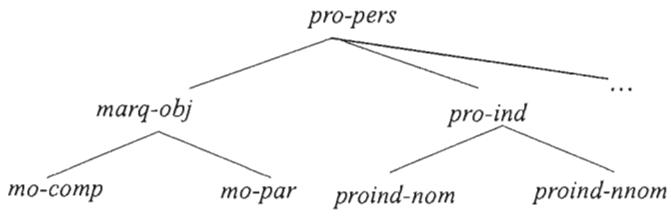


FIGURE 5.3: La hiérarchie des types pronominaux (version révisée 1)

Cependant, les marqueurs d'objet peuvent également être des topiques comme dans l'exemple suivant :

(5.16) *ʔinna* *-hum* *j-* *adxul* *-u:* *-na* *l-* *manzil* *-a*
que -eux 3- entrer.IMP -MP -Ind la- maison -ACC
« Eux, ils entrent dans la maison, certes »

Les topiques, tout comme les explétifs, ne sont pas des arguments de la tête. Cela nous permet de classer les éléments non argumentaux, de type *non-arg*, en deux sous-types *topique* et *explétif*. Le premier est spécifié pour les trois traits d'accord alors que le deuxième, comme nous l'avons déjà mentionné concernant le marqueur d'objet *-ha:*, n'est spécifié que pour deux traits. Le croisement entre cette hiérarchie et celle présentée dans la figure (5.3) nous donne une hiérarchie à héritage multiple où le type *mo-comp* [GRE, NBRE, PERS] hérite de deux super-types : *topique* et *arg*(umental). Le type *mo-par* [GRE, PERS] est le seul sous-type de *explétif*, lui-même sous-type de *non-arg*(umental). Cette hiérarchie des marqueurs d'objet est présentée dans la figure (5.4).

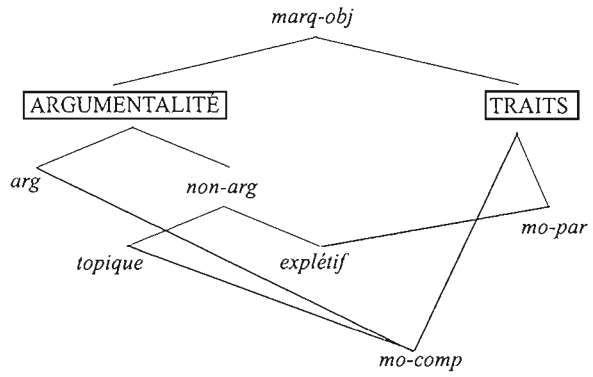


FIGURE 5.4: La hiérarchie des marqueurs d'objet

Les marqueurs d'objet de type *mo-par* sont deux : *-hu* (3M) et *-ha*: (3F). Ils sont les homophones des deux marqueurs d'objet de type *mo-comp*, à savoir *-hu* (3MS) et *-ha*: (3FS). Cette homophonie caractérise également les marqueurs de sujet et les pronoms indépendants nominatifs de 3S, comme le souligne Fassi Fehri (1992).

Pour modéliser le comportement des marqueurs d'objet de l'arabe, le traitement que nous avons proposé est entièrement lexicaliste. Grâce à la richesse des entrées lexicales et aux possibilités que fournissent les hiérarchies de types, nous n'avons eu recours à aucune règle syntagmatique ou de réécriture. Les contraintes imposées au type lexical *marq-obj* et à ces sous-types sont suffisantes pour modéliser le comportement des unités étudiées dans deux environnements différents : lorsqu'ils sont des arguments et lorsqu'ils ne le sont pas. Cette modélisation ne couvre cependant pas d'autres questions qui touchent les marqueurs d'objet, dont celle des suites de clitiques, à laquelle nous consacrons la section subséquente.

5.1.3.5 Les suites de clitiques

En arabe, les suites de clitiques sont possibles avec une seule classe de verbes à deux compléments, à savoir les verbes dits de don, dont la structure argumentale comporte un bénéficiaire et

un thème. Ces verbes ont donc des entrées lexicales qui reflètent cette propriété. Le verbe *wahaba* « donner, offrir », par exemple, est associé à l'entrée lexicale présentée en (5.17), où sont précisés ses différents arguments.

(5.17) L'entrée lexicale de *wahaba*

$$\begin{bmatrix} \textit{verbe-don} & \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{VAL} & / \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2}, \boxed{3} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \end{bmatrix} \end{bmatrix}$$

L'ordre des compléments, lorsque l'un d'eux au moins est un SN ou un pronom indépendant, est régi par la contrainte thématique : le bénéficiaire précède le thème. Cette contrainte prend la forme suivante, où nous introduisons les attributs sémantiques BÉNÉF et THÈME :

(5.18) La contrainte thématique

$$\begin{bmatrix} \textit{verbe-don} & \\ \text{SYNSEM} \mid \text{LOC} & \begin{bmatrix} \text{CAT} & \begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2}, \boxed{3} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \boxed{1}, \boxed{2}, \boxed{3} \rangle \end{bmatrix} \\ \text{CONT} \mid \text{RESTRICTIONS} & \begin{bmatrix} \text{DONNEUR} & \boxed{1} \\ \text{BÉNÉF} & \boxed{2} \\ \text{THÈME} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Une deuxième contrainte intervient lorsque les deux compléments sont des marqueurs d'objet : la contrainte de la personne. Cette dernière instaure un ordre qui prend en compte la valeur de l'attribut PERS, de manière que la première personne précède toujours la deuxième et que cette dernière précède la troisième. Cette contrainte prend la forme suivante :

(5.19) $$\begin{bmatrix} \textit{verbe-don} & \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & / \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2}, \boxed{3} \rangle \end{bmatrix} \\ \text{STR-ARG} & \left\langle \boxed{1}, \boxed{2} \begin{bmatrix} \textit{marq-obj} \\ \text{PERS} \quad \alpha \end{bmatrix}, \boxed{3} \begin{bmatrix} \textit{marq-obj} \\ \text{PERS} \quad \beta \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Où $\alpha \leq \beta$

La contrainte de la personne instaure un ordre entre les deux compléments clitiques des verbes de don de telle manière que le premier de la suite peut être de personne inférieure ou strictement égale à la personne du deuxième. Les suites suivantes sont donc acceptables puisqu'elles satisfont cette contrainte :

(5.20) a. *wahab* *-ta* *-niː* *-hi*
offrir.PER -2MS -me -le
« Tu me l'as offert »

b. *wahab* *-tu* *-ka* *-hu*
offrir.PER -1S -te -le
« Je te l'ai offert »

c. *wahab* *-tu* *-haː* *-hu*
offrir.PER -1S -la -le
« Je le lui ai offert (à elle) »

Notons que la suite en (5.20c) est possible avec deux marqueurs d'objet de 3e personne. Les suites suivantes sont agrammaticales parce qu'elles violent la contrainte de la personne :

(5.21) a. **wahab* *-ta* *-huː* *-niː*
offrir.PER -2MS -le -me

b. **wahab* *-tu* *-hu* *-ka*
offrir.PER -1S -le -te

c. **wahab* *-a* *-ka* *-niː*
offrir.PER -3MS -te -me

Les suites qui respectent la contrainte de personne respectent du même coup la contrainte thématique. Toutefois, certaines suites sont agrammaticales même quand ces deux contraintes sont respectées. Nous avons signalé dans le chapitre IV que cette agrammaticalité constitue un trou arbitraire et nous avons mentionné les cas suivants :

- Les deux marqueurs ne peuvent avoir la même forme phonologique (exemple : *-*hu-hu*).
- Si le premier de la suite est le marqueur d'objet -*ki* (2FS), aucun marqueur d'objet de 3e personne ne peut le suivre.
- Si le premier est le marqueur -*kunna* (2FP), le deuxième ne peut être -*hunna* (3FP).
- On ne peut avoir une suite dont le premier membre est -*hu* (3MS).

Étant des idiosyncrasies (probablement explicables en termes phonologiques et historiques), ces suites ne peuvent être éliminées par les deux contraintes thématique et de personne. Elles doivent donc être filtrées par un autre procédé. La grammaire HPSG utilise des contraintes de filtres, comme celles proposées par Miller et Sag (1997) dans le traitement des suites de clitiques du français. Cependant, si ces deux auteurs ont recours à ces filtres pour éliminer toutes les suites non acceptables, nous y avons recours seulement pour compléter l'effet des deux contraintes thématique et de personne précitées. Les filtres que nous proposons prennent la forme d'une contrainte sur le type *verbe-don* :

(5.22) Contraintes sur le type *verbe-don*

$$\begin{bmatrix} \textit{verbe-don} & \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \textit{verbe} \\ \text{VAL} & / \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2}, \boxed{3} \rangle \end{bmatrix} \\ \text{STR-ARG} & \left\langle \boxed{1}, \boxed{2}\begin{bmatrix} \textit{marq-obj} \\ \text{PERS} \quad \alpha \end{bmatrix}, \boxed{3}\begin{bmatrix} \textit{marq-obj} \\ \text{PERS} \quad \beta \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix}$$

Où $\boxed{2} \neq \boxed{3}$; si $\boxed{2}$ = -ki alors $\beta \neq 3$; si $\boxed{2}$ = -kunna alors $\boxed{3} \neq$ -hunna ; et $\boxed{2} \neq$ -hu.

Pour analyser le comportement et le placement des marqueurs d'objet de l'arabe dans le cadre de HPSG, les éléments essentiels à cette analyse sont de type lexical : la hiérarchie des types, les entrées lexicales et les contraintes sur les types. Seul le lexique est spécifique à la langue étudiée, alors que les principes universels de la grammaire restent valides. L'unification des différentes entrées lexicales, par exemple d'une préposition et d'un marqueur d'objet pour former un syntagme prépositionnel, se fait en respectant et ces principes et les contraintes imposées aux types contenus dans la hiérarchie. La modélisation des marqueurs de sujet se fait en respectant ces lignes directrices, comme nous le démontrons dans la section subséquente.

5.1.4 Les marqueurs de sujet

5.1.4.1 Rappel des données

Les faits suivants constituent le domaine empirique de référence en ce qui a trait à la modélisation des marqueurs de sujet de l'arabe :

1. Les marqueurs de sujet sont des affixes et doivent donc être traités comme tels par la théorie.
2. Ces marqueurs couvrent deux réalités différentes en ce qui a trait à quatre d'entre eux (les marqueurs de 3S) : des arguments et des non arguments.
3. Les arguments sont des pronoms affixes pleinement spécifiés pour leurs traits d'accord.
4. Les non-arguments, en l'occurrence les marqueurs d'accord, ne sont pas spécifiés pour le nombre et peuvent donc être unifiés avec des SN sujets dont la valeur de l'attribut NBRE est variable.
5. Ces propriétés soulèvent des questions sur l'ordre de surface et son influence sur le marqueur de sujet sélectionné par la tête verbale et sur l'accord entre le sujet et le verbe.

Nous proposons une modélisation des marqueurs de sujet qui tient compte des propriétés du domaine empirique et qui reste fidèle au modèle adopté, basant notre analyse ainsi sur la richesse du lexique. La première question que nous traitons ici, en rapport avec le statut affixal de ces marqueurs, est celle de la morphophonologie des verbes de l'arabe.

5.1.4.2 La morphophonologie des verbes de l'arabe

Étant donné que les marqueurs de sujet ont un statut morphologique d'affixes, nous pouvons postuler qu'ils sont *insérés* avec leurs hôtes dans la structure syntagmatique. Cela revient à dire

que les marqueurs de sujet ne sont pas des items lexicaux indépendants et n'ont donc ni type ni entrée lexicale. Pour rendre compte de ces propriétés et de la distribution des marqueurs de sujet par rapport aux SN et aux pronoms indépendants, nous devons tout d'abord modéliser la formation morphologique des hôtes de ces marqueurs, à savoir les verbes [12].

La théorie de Smets (2000), présentée en partie dans le premier chapitre dans la section (1.5.2), constitue notre point de départ dans la modélisation de la formation des verbes de l'arabe. Basée sur une théorie morphologique bien établie, celle de McCarthy (1981), cette analyse couvre exclusivement les aspects morphophonologiques. Les aspects morphosyntaxiques et syntaxiques constituent notre contribution à cette théorie.

5.1.4.2.1 La morphologie à patrons

Nous commençons par présenter les données sur lesquelles Smets (2000), et avant elle McCarthy (1981), ont basé leur analyse et leur formalisme. En arabe, chaque verbe possède l'un des schèmes représentés dans le tableau (5.1) [13].

Les 10 premiers schèmes trilitères sont les plus productifs. Toutefois, ce ne sont pas toutes les racines qui peuvent donner lieu à des verbes dans chacun de ces schèmes. La racine /drs/, par exemple, ne peut être combinée qu'avec les schèmes I, II, III, VI et VII selon Smets (2000, p. 111).

Ces données sont connues et discutées dans la grammaire arabe depuis ses débuts au VIIIe siècle. L'apport de McCarthy (1981) à la question réside dans le fait qu'il a défini un patron prosodique duquel tous les schèmes précités sont issus. Le patron en question est celui en (5.23) [14] :

$$(5.23)\quad \left[\left(\left\{ \begin{array}{c} C \\ CV \end{array} \right\} \right) CV\,([+seg])\,CVC \right]$$

Ce patron est tellement général qu'il sur-génère des formes non attestées et l'auteur se voit

[12] Il va de soi que cette modélisation ne touche que les marqueurs de sujet puisque ces derniers font partie de la structure morphologique des verbes. Les marqueurs d'objet peuvent avoir d'autres catégories comme hôtes et ne font pas partie de la structure morphologique de leurs hôtes verbaux.

[13] Nous avons choisi de représenter ces schèmes en nous basant sur les trois consonnes abstraites utilisées par les grammairiens de la tradition dans le même but : f, ʕ, l. Nous avons également allégé la présentation en excluant les schèmes du participe (actif et passif). La question de savoir si ce que McCarthy (1981) appelle le participe est une forme verbale ou nominale n'est pas soulevée ici et mériterait plus d'investigations.

[14] (McCarthy, 1981, p. 387). Dans ce patron, la notation [+seg] dénote un segment quelconque, à savoir un élément consonantique ou vocalique.

Verbes trilitères				
Schèmes	Perfectif		Imperfectif	
	Actif	Passif	Actif	Passif
I	faʕal	fuʕil	afʕal	ufʕal
II	faʕʕal	fuʕʕil	ufaʕʕil	ufaʕʕal
III	fa:ʕal	fu:ʕil	ufa:ʕil	ufa:ʕal
IV	ʔafʕal	ʔufʕil	ufʕil	ufʕal
V	tafaʕʕal	tufuʕʕil	atafaʕʕal	utafaʕʕal
VI	tafa:ʕal	tufu:ʕil	atafa:ʕal	utafa:ʕal
VII	nfaʕal	nfuʕil	anfaʕil	unfaʕal
VIII	ftaʕal	ftuʕil	aftaʕil	uftaʕal
IX	fʕalal	-	afʕalal	-
X	stafʕal	stufʕil	astafʕil	ustafʕal
XI	fta:ʕal	-	afta:ʕil	-
XII	ftawʕal	-	aftawʕil	-
XIII	fʕawwal	-	afʕawwil	-
XIV	fʕanlal	-	afʕanlil	-
XV	fʕanlaj	-	afʕanlij	-
Verbes quadrilatères				
QI	faʕlal	fuʕlil	ufaʕlil	ufaʕlal
QII	tafaʕlal	tufuʕlil	atafaʕlal	utafaʕlal
QIII	fʕanlal	fʕunlil	afʕanlil	ufʕanlal
QIV	fʕallal	fʕullil	afʕallal	ufʕallal

TABLEAU 5.1: Les schèmes morphologiques des verbes trilitères et quadrilatères

donc contraint d'ajouter une règle phonologique *ad hoc* qui restreint l'application du patron général. Cette règle élimine le schème non attesté CVCVCVC et prend la forme suivante [15] :

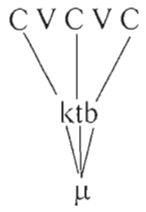
(5.24) V → ∅ / [CVC_CVC]

Les schèmes ainsi acceptables sont les suivants : CVCVC, CVCCVC, CVVCVC, CVCVCCVC, CVCVVCVC, CCVCVC et CCVCCVC.

La deuxième contribution majeure de McCarthy (1981) réside dans sa définition de conventions universelles qui associent les racines et les vocalismes aux patrons morphologiques de gauche à droite, comme dans l'exemple suivant, qui représente l'association de la racine /ktb/ au patron du schème I [16].

[15] *Ibid.*
[16] *Ibid.*, p. 388.

(5.25)



Cependant, un problème se pose en ce qui a trait au VIIIe schème : le matériel *t*, qui est préfixé ailleurs, est infixé dans ce schème. McCarthy (1981) résout ce problème en postulant l'existence d'une règle propre à cet affixe et au schème VIII. Une autre règle *ad hoc* est stipulée par l'auteur pour rendre compte des particularités des schèmes II et V. Il dispose ainsi d'un appareillage composé des ingrédients suivants pour rendre compte de la morphologie non concaténative de l'arabe :

- Un patron prosodique.
- Une règle d'effacement de la voyelle dans le contexte CVC_CVC.
- Les affixes ?, *t*, *n* et *st*.
- Les deux règles d'ajustement des schèmes VIII, II et V.

Cette théorie repose sur l'hypothèse suivante, déjà exprimée dans la tradition grammaticale arabe : la racine, le schème et la mélodie vocalique véhiculent des informations différentes. La racine exprime le sens général du verbe. Le schème exprime toute modification dans le contenu sémantique et dans les propriétés de sous-catégorisation de la racine. La mélodie vocalique est, finalement, associée à la catégorie du verbe (l'aspect et la voix).

5.1.4.2.2 La formalisation de la morphologie à patrons

Smets (2000) a formalisé cette théorie autosegmentale de la morphologie non concaténative de l'arabe. Le formalisme qu'elle propose reste fidèle à la théorie de McCarthy (1981) sur deux points :

1. Son choix des attributs reflète la distinction faite entre les trois paliers : racine, schème et vocalisme.
2. Seul l'aspect morphophonologique est pris en compte dans cette formalisation. Les informations morphosyntaxiques et syntaxiques associées aux morphèmes en étude ici ne sont pas prises en considération.

Les attributs introduits par Smets (2000) sont au nombre de trois : RACINE, VOC(alisme) et PATRON (pour le schème). Ces trois attributs constituent des valeurs de l'attribut PHON, mais pourraient également servir de valeurs de l'attribut MORPH en introduisant une composante morphologique dans le formalisme.

En plus de définir ces attributs, l'auteure définit deux sortes de paradigmes : des paradigmes sources et des paradigmes principaux. Dans les premiers, elle définit les paradigmes de patrons possibles et les paradigmes des mélodies vocaliques. Les paradigmes principaux combinent, par unification, les premiers pour obtenir les formes spécifiées de la langue.

Les paradigmes sources

L'auteure définit un paradigme de patron pour chacun des deux aspects : le perfectif et l'imperfectif. Le paradigme en (5.26) définit les patrons communs aux dix premières formes trilitères, les plus productives. La spécialisation de ces patrons en fonction de l'aspect est introduite en (5.27) pour le perfectif et en (5.28) pour l'imperfectif.

(5.26)
$$\begin{bmatrix} \textit{PS-radic-trilit} & \\ \text{PHON} & \begin{bmatrix} \text{RACINE} & \begin{bmatrix} \boxed{1}\text{C1},\boxed{2}\text{C2},\boxed{3}\text{C3} \end{bmatrix} \\ \text{VOC} & \begin{bmatrix} \boxed{4}\text{V1},\boxed{5}\text{V2} \end{bmatrix} \end{bmatrix} \& \\ \text{SCHÈME I :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\boxed{1},\boxed{4},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME II :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\boxed{1},\boxed{4},\boxed{2},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME III :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\boxed{1},\boxed{4},\boxed{4},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME IV :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [?,\boxed{4},\boxed{1},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME V :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\text{T},\boxed{4},\boxed{1},\boxed{4},\boxed{2},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME VI :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\text{T},\boxed{4},\boxed{1},\boxed{4},\boxed{4},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME VII :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\text{N},\boxed{1},\boxed{4},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME VIII :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\boxed{1},\text{T},\boxed{4},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME IX :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\boxed{1},\boxed{2},\boxed{4},\boxed{3},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \\ \text{SCHÈME X :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & [\text{S},\text{T},\boxed{4},\boxed{1},\boxed{2},\boxed{5},\boxed{3}\ ] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(5.27)
$$\begin{bmatrix} \textit{PS-perf-trilit} & \\ \text{PERF :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{PATRON} & \boxed{1} \\ \text{MOT} & \boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Cette formalisation du perfectif repose sur l'hypothèse suivante : en ce qui a trait au perfectif des verbes trilitères, la forme du radical précisée dans le paradigme (5.26) et la forme du *mot* [17] sont identiques. L'imperfectif se distingue du perfectif par la présence d'une voyelle qui précède le radical et cette propriété est donc rendue par le paradigme suivant :

[17] L'auteure emploie ce terme de manière informelle ici en voulant parler du radical verbal, qui, lui, n'est pas en réalité un mot.

(5.28) $\begin{bmatrix} \textit{PS-imper-trilit} & \\ \text{IMPER :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{RADICAL} & \boxed{1} \\ \text{MOT} & \begin{bmatrix} \text{V},\boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$

En plus de ces paradigmes, l'auteure définit des paradigmes sources pour les mélodies vocaliques possibles en optant pour la même partition entre le perfectif et l'imperfectif. Le paradigme suivant est celui de la mélodie vocalique du perfectif :

(5.29) $\begin{bmatrix} \textit{PS-mélodie-perf} & \\ \text{PHON} & \begin{bmatrix} \begin{bmatrix} \text{PATRON} & \boxed{1} \\ \text{MOT} & \boxed{1} \end{bmatrix} \& \\ \begin{bmatrix} \text{perf \& actif \& schème I :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{VOC} & [\boxed{1}\text{V1},\boxed{2}\text{V2}] \\ \text{VOCI} & [\boxed{1},\boxed{2}] \end{bmatrix} \end{bmatrix} \\ \text{perf \& actif \& non-schème I :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{VOC} & [\text{a,a}] \end{bmatrix} \end{bmatrix} \\ \text{perf \& passif \& non-schème IX :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{VOC} & [\text{u,i}] \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$

Ce paradigme reflète les observations empiriques suivantes : les schèmes de II à X partagent la mélodie [a,a] au perfectif actif et [u,i] au perfectif passif. Pour le schème I, la qualité des voyelles n'est pas prévisible et doit donc être spécifiée dans les entrées lexicales de chaque verbe.

Le paradigme suivant représente la mélodie vocalique de l'imperfectif[18] :

[18] P : patron.

(5.30)

$$\left[\begin{array}{l} \textit{PS-mélodie-imper} \\ \text{imper \& actif \& schème I :} \\ \text{imper \& actif \& } \left\{\text{schème II v schème III v schème IV}\right\} : \\ \text{imper \& actif \& } \left\{\text{schème V v schème VI}\right\} : \\ \text{imper \& actif \& } \left\{\text{schème III v schème VIII v schème IX v schème X}\right\} : \\ \text{imper \& passif \& schème I :} \\ \text{imper \& passif \& non-scème I \& non-schème IX :} \end{array}\right. \begin{array}{l} \left[\text{PHON} \left[\begin{array}{ll} \text{VOC} & [\text{nil},\boxed{1}\text{V1}] \\ \text{VOY-I} & [\boxed{1}] \\ \text{MOT} & [\text{a,P}] \end{array}\right]\right] \\ \left[\text{PHON} \left[\begin{array}{ll} \text{VOC} & [\text{a,i}] \\ \text{MOT} & [\text{u,P}] \end{array}\right]\right] \\ \left[\text{PHON} \left[\begin{array}{ll} \text{VOC} & [\text{a,a}] \\ \text{MOT} & [\text{a,P}] \end{array}\right]\right] \\ \left[\text{PHON} \left[\begin{array}{ll} \text{VOC} & [\text{a,i}] \\ \text{MOT} & [\text{a,P}] \end{array}\right]\right] \\ \left[\text{PHON} \left[\begin{array}{ll} \text{VOC} & [\text{nil,a}] \\ \text{MOT} & [\text{u,P}] \end{array}\right]\right] \\ \left[\text{PHON} \left[\begin{array}{ll} \text{VOC} & [\text{a,a}] \\ \text{MOT} & [\text{u,P}] \end{array}\right]\right] \end{array}$$

La mélodie vocalique de l'imperfectif est caractérisée par les faits suivants : la mélodie du radical de l'actif dans les schèmes II à IV est [a,i] et la voyelle préfixée est [u]. Dans les schèmes V et VI, les voyelles du radical et de l'affixe sont [a] alors que dans les schèmes VII à X, les voyelles du schème sont [a,i] et celle du préfixe est [a]. À la voix passive, la mélodie du radical est [a,a] dans les schèmes autres que I et IX et la voyelle du préfixe est [u].

Encore une fois, le schème I présente une exception parce que la voyelle du radical actif n'est pas prévisible. L'auteure propose d'encoder cette information dans l'entrée lexicale des verbes, dans la valeur de l'attribut VOY-I. La voyelle du préfixe dans ce schème est [a]. Au passif, la voyelle du radical est [a] et celle du préfixe est [u].

En somme, ces paradigmes fournissent les patrons de conjugaison possibles pour chacun des 10 premiers schèmes des verbes trilitères ainsi que les mélodies vocaliques possibles. Ces paradigmes, différents pour le perfectif et l'imperfectif, peuvent être unifiés en respectant certaines contraintes que l'auteure désigne par le terme « paradigmes principaux ».

Les paradigmes principaux

Le rôle de ces paradigmes est d'unifier les informations provenant des paradigmes sources pour définir des formes acceptables. Le paradigme suivant combine des paradigmes définissant les schèmes et les mélodies vocaliques pour donner lieu à des verbes trilitères.

(5.31) *PP-verbe-trilit :*

PS-radic-trilit & PS-perf-trilit & PS-mélodie-perf

PS-radic-trilit & PS-imper-trilit & PS-mélodie-imper

Ce paradigme définit ainsi les combinaisons possibles. La première entrée, par exemple, légitime l'unification de chaque entrée du paradigme *PS-radic-trilit* avec chacune des entrées des deux paradigmes *PS-perf-trilit* et *PS-mélodie-perf*. Pour donner un exemple de cette unification, considérons la première entrée du paradigme source des verbes trilitères *PS-radic-trilit*, celle du schème I. Cette entrée peut être unifiée avec la seule entrée du paradigme *PS-perf-trilit*, qui est en réalité une contrainte sur la réalisation du perfectif, et la première entrée du paradigme *PS-mélodie-perf* (*perf & actif & schème I*). Il en résulte la structure de traits suivante :

(5.32)
$$\begin{bmatrix} V\ perfectif\ actif & \\ \text{PHON} & \begin{bmatrix} \text{RACINE} & [\boxed{1}C1,\boxed{2}C2,\boxed{3}C3] \\ \text{VOC} & [a,a] \\ \text{PATRON} & \boxed{4}[\boxed{1},a,\boxed{2},a,\boxed{3}] \\ \text{MOT} & \boxed{4} \end{bmatrix} \\ \text{TÊTE} & \begin{bmatrix} verbe & \\ \text{ASPECT} & perfectif \\ \text{VOIX} & actif \end{bmatrix} \\ \text{FORM} & \text{V} \end{bmatrix}$$

Les entrées lexicales des verbes

Dans le cadre de cette théorie de la morphophonologie arabe, les entrées lexicales de chaque verbe doivent encoder un certain type d'informations morphophonologiques qu'on ne peut inférer à partir des paradigmes précités et de l'unification des entrées de ces paradigmes. Parmi ces

informations, citons le vocalisme des verbes du premier schème et les schèmes possibles pour chaque racine. Sur ce dernier point, il faut noter que la combinaison de certains schèmes et de certaines racines donne lieu à des formes non attestées. Le fait d'encoder cette information dans l'entrée lexicale de chaque verbe permet d'éliminer ces formes.

Smets (2000) prend l'exemple du verbe *daras* « étudier ». Dans l'entrée lexicale de ce verbe sont précisés la racine consonantique, les schèmes possibles que le verbe peut prendre et le sens associé à chaque schème ainsi que le changement de sous-catégorisation qui en résulte.

(5.33) L'entrée lexicale de *daras*

$$\begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{RACINE} & \textit{[d,r,s]} \end{bmatrix} \\ \textit{I :} & \begin{bmatrix} \text{PHON} & \begin{bmatrix} \text{VOCI} & \textit{[a,a]}, \text{ IVOW} & \textit{[u]}, \text{ COMPS} & \textit{[SN,SP]}, \text{ CONT} & \textit{[étudier quelque chose]} \end{bmatrix} \end{bmatrix} \\ \textit{II :} & \begin{bmatrix} \text{COMPS} & \textit{[SN,SP]}, \text{ CONT} & \textit{[enseigner quelque chose]} \end{bmatrix} \\ \textit{III :} & \begin{bmatrix} \text{COMPS} & \textit{[SN]}, \text{ CONT} & \textit{[étudier avec quelqu'un]} \end{bmatrix} \\ \textit{VI :} & \begin{bmatrix} \text{COMPS} & \textit{[SN]}, \text{ CONT} & \textit{[étudier attentivement avec quelqu'un]} \end{bmatrix} \\ \textit{VII :} & \begin{bmatrix} \text{COMPS} & \textit{[]}, \text{ CONT} & \textit{[disparaître]} \end{bmatrix} \end{bmatrix}$$

Cette formalisation de la morphologie verbale de l'arabe permet, tout en maintenant un système minimal basé sur les paradigmes et une seule opération, l'unification, de rendre compte de l'aspect morphophonologique de la formation des verbes. Les aspects syntaxiques et morphosyntaxiques sont cependant absents. Les entrées lexicales des verbes, telles que conçues par Smets (2000), et même si elles contiennent une information de type syntagmatique (la valeur de l'attribut COMPS), ne fournissent aucune information sur les marqueurs de sujet. Ces derniers devant faire partie de l'entrée lexicale de chaque verbe, nous avons à étendre ce formalisme pour les inclure. La section suivante propose une manière de traiter la morphosyntaxe des verbes de l'arabe, compatible avec l'approche de Smets (2000).

5.1.4.3 La morphosyntaxe et l'insertion des marqueurs de sujet

La formation des verbes en arabe peut être conçue comme impliquant deux étapes. Dans la première, le radical est formé en unifiant les différentes entrées fournies dans les paradigmes

sources selon la spécification contenue dans le paradigme principal. La deuxième étape est de nature concaténative. Lors de cette phase, les affixes qui expriment les différents traits morpho-syntaxiques ainsi que les affixes de mode, en ce qui concerne l'imperfectif, sont concaténés au radical de trois manières différentes. Dans la première, celle qui caractérise le perfectif et l'impératif, un suffixe est concaténé au radical. Dans la deuxième, un préfixe est concaténé en plus d'un affixe de mode. Dans la troisième, un préfixe et un suffixe sont concaténés en plus de l'affixe de mode. Ces deux derniers processus sont caractéristiques de l'imperfectif.

Nous traitons les différents affixes concaténés au radical verbal, une fois que ce radical est bien formé par l'unification des entrées contenues dans les paradigmes sources, comme étant des affixes flexionnels. Le traitement que nous proposons de la flexion est purement lexical et fait appel à des règles lexicales de flexion en premier lieu.

5.1.4.3.1 Les règles lexicales flexionnelles

Pour modéliser la flexion verbale, nous introduisons trois règles lexicales flexionnelles qui relient un radical (pouvant être considéré comme un lexème) à un mot bien formé (le verbe fléchi). Ces règles ont pour intrant un objet de type *radical* et pour extrant un objet de type *verbe-mot*.

La forme générale d'une règle lexicale flexionnelle est la suivante :

(5.34) $$\begin{bmatrix} \textit{r-flex-verbe} & \\ \text{INTRANT} & \begin{bmatrix} \textit{radical} \end{bmatrix} \\ \text{EXTRANT} & \begin{bmatrix} \textit{verbe-mot} \end{bmatrix} \end{bmatrix}$$

Le type auquel appartient cette règle est *r-flex-verbe*. Trois sous-types sont dominés par ce type : *r-flex-perf/impé*, *r-flex-imper1* et *r-flex-imper2*. Les règles du premier type servent à concaténer un suffixe flexionnel au radical du perfectif ou de l'impératif, les règles du deuxième type à concaténer un préfixe flexionnel et un suffixe de mode au radical de l'imperfectif et les dernières concatènent un préfixe et un suffixe flexionnel en plus d'un suffixe de mode au radical de l'imperfectif.

(5.35)
$$\begin{bmatrix} \textit{r-flex-perf/impé} & \\ \text{INTRANT} & \begin{bmatrix} \textit{radical} \\ \boxed{1} \end{bmatrix} \\ \text{EXTRANT} & \begin{bmatrix} \textit{verbe-mot} & \\ \text{PHON} & \left\langle \boxed{1} \oplus \text{SUFFIXE} \right\rangle \\ \text{ASPECT} & \textit{perfectif} \\ \text{TÊTE} & \boxed{2} \\ \text{STR-ARG} & \left\langle \ldots \right\rangle \end{bmatrix} \end{bmatrix}$$

(5.36)
$$\begin{bmatrix} \textit{r-flex-imper1} & \\ \text{INTRANT} & \begin{bmatrix} \textit{radical} \\ \boxed{1} \end{bmatrix} \\ \text{EXTRANT} & \begin{bmatrix} \textit{verbe-mot} & \\ \text{PHON} & \left\langle \text{PRÉFIXE} \oplus \boxed{1} \oplus \text{SUFFIXE-MODE} \right\rangle \\ \text{ASPECT} & \textit{imperfectif} \\ \text{TÊTE} & \boxed{2} \\ \text{STR-ARG} & \left\langle \ldots \right\rangle \end{bmatrix} \end{bmatrix}$$

(5.37)
$$\begin{bmatrix} \textit{r-flex-imper2} & \\ \text{INTRANT} & \begin{bmatrix} \textit{radical} \\ \boxed{1} \end{bmatrix} \\ \text{EXTRANT} & \begin{bmatrix} \textit{verbe-mot} & \\ \text{PHON} & \left\langle \text{PRÉFIXE} \oplus \boxed{1} \oplus \text{SUFFIXE} \oplus \text{SUFFIXE-MODE} \right\rangle \\ \text{ASPECT} & \textit{imperfectif} \\ \text{TÊTE} & \boxed{2} \\ \text{STR-ARG} & \left\langle \ldots \right\rangle \end{bmatrix} \end{bmatrix}$$

La valeur de l'attribut STR-ARG n'est pas spécifiée dans ces règles et fera l'objet d'une discus-

sion plus loin dans ce chapitre.

Dans ces règles, les objets de type *radical* sont le résultat de l'unification des entrées contenues dans les paradigmes sources selon les spécifications que fournissent les paradigmes principaux.

5.1.4.3.2 Les modes de l'imperfectif

Les marqueurs de mode sont introduits par les deux règles lexicales *r-flex-imper1* et *r-flex-imper2*. Cependant, seuls les verbes au mode indicatif, au singulier et à la 1P du mode subjonctif portent des marqueurs de mode. Au jussif et aux autres personnes du subjonctif, les marques de mode sont absentes. Ce comportement suggère une certaine sensibilité des marqueurs de mode aux traits morphosyntaxiques en ce qui a trait au subjonctif. L'indicatif et le jussif sont semblables en ceci que le premier possède de marqueurs de mode alors que le deuxième n'en a pas.

Pour modéliser ce comportement, nous proposons la hiérarchie de types présentée dans la figure (5.5).

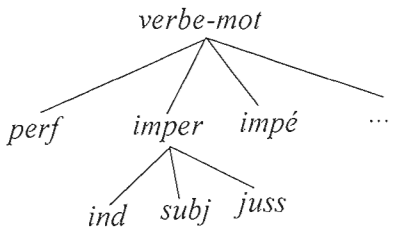


FIGURE 5.5: Une hiérarchie de types pour les verbes de l'arabe

Les contraintes suivantes sont imposées à ces types de verbes imperfectifs :

(5.38) $\begin{bmatrix} \textit{ind} & \\ \text{MORPH} & \begin{bmatrix} \text{SUFFIXE-MODE} & + \end{bmatrix} \end{bmatrix}$

(5.39) $\begin{bmatrix} \textit{juss} & \\ \text{MORPH} & \begin{bmatrix} \text{SUFFIXE-MODE} & - \end{bmatrix} \end{bmatrix}$

(5.40) $\begin{bmatrix} \textit{subj} & \\ \text{MORPH} & \begin{bmatrix} \text{SUFFIXE-MODE} & \alpha[\textit{bool}] \end{bmatrix} \end{bmatrix}$

Où α = + si [NBRE *sing*] ou [NBRE *pluriel*] et [PERS *1*], sinon α = -

Ces contraintes garantissent que chaque verbe imperfectif sera doté des affixes de mode appropriés à son type. Les verbes jussifs ne contiennent pas de tels affixes du fait de leur appartenance au type *juss*. Les verbes à l'indicatif et au subjonctif comportent les affixes de mode appropriés à leurs types respectifs *ind* et *subj*. Ces affixes interagissent avec les traits morphosyntaxiques exprimés par les marqueurs de sujet, que nous modélisons dans le paragraphe suivant.

5.1.4.3.3 Les traits morphosyntaxiques

Nous avons, jusqu'à présent, modélisé l'aspect proprement morphophonologique de l'affixation des marqueurs de mode. Les traits exprimés par les marqueurs de sujet et leur effet sur la structure syntagmatique ne sont pas pris en compte. Une manière d'y arriver est de modifier les règles lexicales flexionnelles pour introduire les traits de genre, de nombre et de personne. Ces traits sont en relation avec la structure argumentale du verbe puisque la réalisation du trait de nombre permet à l'affixe d'apparaître dans la valeur de la liste STR-ARG (en tant qu'argument). Un affixe non spécifié pour le nombre n'est pas un argument et ne constitue donc pas un membre de cette liste.

Pour modéliser cette analyse, nous faisons appel à la typologie de Miller et Sag (1997), à savoir leur distinction entre des types canoniques et des types non canoniques. Dans cette typologie, le type *aff* (pour les affixes) est un type non canonique puisqu'il peut correspondre à des éléments membres de la structure argumentale qui ne sont pas réalisés en tant que dépendants syntaxiques de la tête. Cette typologie est représentée dans la figure (1.9) et reprise ici dans la figure (5.6).

Conjointement à cette hiérarchie partielle, nous proposons d'imposer la contrainte en (5.41) au type *verbe-mot* :

(5.41) $\begin{bmatrix} \textit{verbe-mot} & \\ \text{MORPH} & \begin{bmatrix} \text{FORM} & \text{F}_{\textit{pref}}\ (\boxed{1}\ ,\ \ldots) \\ \text{I-FORM} & \boxed{1} \end{bmatrix} \end{bmatrix}$

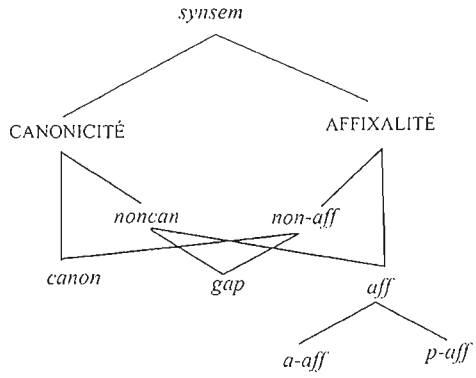


FIGURE 5.6: Une hiérarchie partielle des types proposée par Miller et Sag (1997)

Cette contrainte morphologique nous garantit que la valeur de l'attribut FORM soit le résultat de l'application d'une fonction phonologique dont l'un des arguments est la valeur de l'attribut I-FORM. Cette contrainte exige ainsi que la valeur de l'attribut FORM soit reliée à la valeur de l'attribut I-FORM par le biais de l'affixation flexionnelle appropriée.

Nous proposons que le type *verbe-mot* possède deux sous-types selon le type de l'affixe flexionnel qui y est attaché : *verbe-mot-accord* et *verbe-mot-MS*. Les deux contraintes suivantes sont imposées à ces deux types :

(5.42)
$$\begin{bmatrix} \textit{verbe-mot-accord} & \\ \text{TÊTE}\,|\,\text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{gre} \\ \text{PERS} & \textit{pers} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \ldots \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \boxed{1}, \ldots \rangle \end{bmatrix}$$

$$(5.43)\quad \begin{bmatrix} \textit{verbe-mot-MS} & \\ \text{TÊTE} \mid \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{gre} \\ \text{PERS} & \textit{pers} \\ \text{NBRE} & \textit{nbre} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle\rangle \\ \text{COMPS} & \langle \ldots \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle [\textit{aff}], \ldots \rangle \end{bmatrix}$$

Les contraintes sur le type *verbe-mot-accord* assurent la réalisation d'un sujet lexical (apparaissant à la fois dans la liste de valence et dans la structure argumentale) quand le marqueur de sujet est sous-spécifié pour le nombre.

Les contraintes sur le type *verbe-mot-MS* assurent la réalisation du sujet (le premier membre de la liste STR-ARG) en tant qu'affixe. Dans ce cas-ci, l'attribut NBRE est présent dans les traits de tête du verbe.

La différence entre ces deux sous-types implique que les verbes arabes munis de marqueurs de sujet de 3MS et de 3FS aient deux entrées lexicales différentes selon leur appartenance à l'un ou à l'autre de ces deux sous-types. Ainsi, le verbe *kataba* « il a écrit », par exemple, est compatible avec deux lectures possibles : dans l'une, le marqueur de sujet est un marqueur d'accord et, dans l'autre, ce dernier est un pronom sujet incorporé et fait partie des arguments de cette tête. Cette double lecture possible se reflète dans les deux entrées lexicales en (5.44) et (5.45). D'un autre côté, le verbe *katabna* « elles ont écrit », par exemple, n'est compatible qu'avec une seule lecture et n'a, de ce fait, qu'une seule entrée lexicale, celle en (5.46).

(5.44) $$\begin{bmatrix} \textit{verbe-mot-MS} & \\ \text{PHON} & \textit{<kataba>} \\ \text{TÊTE} \mid \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{masc} \\ \text{PERS} & \textit{3} \\ \text{NBRE} & \textit{sing} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle\rangle \\ \text{COMPS} & \langle \boxed{1} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \textit{[aff]} \oplus \boxed{1} \rangle \end{bmatrix}$$

(5.45) $$\begin{bmatrix} \textit{verbe-mot-accord} & \\ \text{PHON} & \textit{<kataba>} \\ \text{TÊTE} \mid \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{masc} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle \boxed{1} \rangle \\ \text{COMPS} & \langle \boxed{2} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \boxed{1} \oplus \boxed{2} \rangle \end{bmatrix}$$

(5.46) $$\begin{bmatrix} \textit{verbe-mot-MS} & \\ \text{PHON} & \textit{<katabna>} \\ \text{TÊTE} \mid \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fém} \\ \text{PERS} & \textit{3} \\ \text{NBRE} & \textit{plur} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \langle\rangle \\ \text{COMPS} & \langle \boxed{2} \rangle \end{bmatrix} \\ \text{STR-ARG} & \langle \textit{[aff]} \oplus \boxed{2} \rangle \end{bmatrix}$$

Par application du principe des traits de tête et du principe de la valence, les effets syntagmatiques de ces entrées lexicales sont les suivants :

- Un verbe dont l'entrée lexicale spécifie le nombre ne peut se combiner avec un sujet lexical (un SN). Si un tel composant est présent, il ne peut faire partie de la liste des arguments du verbe et sera donc disloqué. Le syntagme qui en résulte est de type *syn-t-f*.
- Un verbe dont l'entrée lexicale ne spécifie pas le nombre peut se combiner avec un sujet lexical (un SN). Ce dernier fait ainsi partie de la liste des membres de la structure argumentale du verbe et est réalisé comme membre de la liste de valence.

Ainsi, notre grammaire élimine toute possibilité d'unifier la structure de traits qui décrit un SN plein avec celle qui décrit un verbe de type *verbe-mot-MS* dans un syntagme de type *syn-t-s*. Elle permet par contre l'unification d'un SN et d'un verbe-mot-accord dans ce type de syntagme et d'un SN et d'un *verbe-mot-MS* dans un syntagme de type *syn-t-f*.

Cette analyse repose clairement sur notre conception de l'accord entre le sujet et le verbe, à savoir la position initiale du verbe et la sous-spécification du nombre sur ce dernier. Nous proposons une modélisation de cette analyse dans la section subséquente.

5.1.4.4 L'accord entre le sujet et le verbe

Nous avons avancé une hypothèse selon laquelle l'accord entre le sujet et le verbe ne s'obtient que dans l'ordre VSO en arabe. Le sujet ici est un SN et le verbe porte un marqueur d'accord partiel (non spécifié pour le nombre). Pour rendre compte de cette particularité, nous avons stipulé que les verbes portant de tels marqueurs de sujet possèdent des entrées lexicales différentes des verbes portant des marqueurs de sujet pronominaux (spécifiés pour le nombre). Ainsi, chaque verbe conjugué à la troisième personne du singulier possède deux entrées lexicales et chacune de ces entrées donne lieu à une structure syntagmatique particulière. L'entrée lexicale où le nombre est sous-spécifié donne lieu à une structure où un SN est l'argument sujet. Ce SN s'accorde avec le verbe en genre et en personne seulement. Cette modélisation ne fournit cependant aucun moyen de s'assurer que les phrases agrammaticales, telles que celle en (5.47c) et celle en (5.47d) soient exclues :

(5.47) a. *daxal* *-a* *l-* *walad* *-u*
entrer.PER -3M le- garçon -NOM
« Le garçon est entré »

b. *daxal* *-at* *l-* *bint* *-u*
entrer.PER -3F la- fille -NOM
« La fille est entrée »

c. **daxal* *-a* *l-* *bint* *-u*
entrer.PER -3M la- fille -NOM

d. **daxal* *-at* *l-* *walad* *-u*
entrer.PER -3F le- garçon -NOM

En (5.47c) et (5.47d), le sujet et le verbe portent des informations incompatibles sur le genre et aucune unification n'est possible dans ces contextes. Pour exclure ces phrases, nous avons besoin d'une contrainte qui assure que les traits d'accord, essentiellement le genre, sur le sujet et ceux sur le verbe sont compatibles et peuvent donc être unifiés. Pour y parvenir, nous devons tout d'abord rappeler la théorie de l'accord développée essentiellement par Pollard et Sag (1994), Kathol (1999), Sag et Wasow (1999) et Sag *et al.* (2003).

Le traitement de l'accord en HPSG

HPSG conçoit l'accord dans une perspective non dérivationnelle. En fait, au lieu de le concevoir comme un processus qui copie certains traits du donneur à la cible, la théorie avance l'idée selon laquelle l'accord se réalise quand les deux éléments qui entrent dans ce genre de relation spécifient une information partielle concernant le même objet linguistique : « Agreement is simply the systematic variation in form that arises from the fact that information coming from two sources about a single object must be compatible. » (Pollard et Sag, 1994, p. 60).

Dans les premières versions de la théorie, ce partage de structures concerne essentiellement la valeur de l'attribut INDEX (qui est un attribut sémantique réalisé sous l'attribut local CONT) [19]. Dans cette approche, l'accord entre le sujet et le verbe est conçu comme étant le partage des valeurs de l'attribut INDEX de ces deux constituants. Ce partage est illustré dans la structure de traits suivante pour la phrase française *les cigales chantent* :

[19] Pollard et Sag (1994) définissent trois sortes d'accord : l'accord des indices, la concordance (l'accord en cas, par exemple) et l'accord pragmatique. Cette distinction est censée rendre compte des cas de l'accord hybride comme dans l'exemple français « vous êtes belle », où l'accord entre le sujet et le verbe est un accord d'indices et celui entre le sujet et le prédicat est pragmatique. Voir Kathol (1999).

$$(5.48)\quad \begin{bmatrix} \text{PHON} & < \textit{les cigalent chantent} > \\ \text{SS}\,|\,\text{LOC}\,|\,\text{CONT}\,|\,\text{INDEX} & \boxed{1} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE}\,|\,\text{SS}\,|\,\text{LOC}\,|\,\text{CONT}\,|\,\text{INDEX} & \boxed{1}\begin{bmatrix} \text{NBRE} & \textit{plur} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{FILLE-COMPS}\,|\,\text{SS}\,|\,\text{LOC}\,|\,\text{CONT}\,|\,\text{INDEX} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

Notons que l'accord entre le sujet et le verbe en français n'implique que deux traits de l'index : le nombre et la personne.

L'évolution du modèle a amené un changement dans cette théorie de l'accord. Sag et Wasow (1999) et Sag *et al.* (2003) proposent un nouvel attribut AGR (que nous traduisons ici par ACCORD) approprié aux objets de type *verbe*, *nominal* et *déterminant* en anglais. Ces trois catégories sont ainsi regroupées sous le même type lexical *agr-cat* (une catégorie pour laquelle l'accord est approprié). Ce nouveau trait de tête [AGR *agr-cat*] joue le rôle jadis assumé par l'index.

Dans cette théorie, l'accord entre le sujet et le verbe est conçu dans les mêmes lignes que l'accord entre le déterminant et le nom. Le sujet et le déterminant étant regroupés sous le même attribut SPEC (spécifieur), cela permet aux auteurs d'imposer une contrainte lexicale aux lexèmes fléchis[20] de l'anglais, qui appartiennent au type lexical *infl-lxm*, la contrainte de l'accord entre le spécifieur et la tête ou la *SHAC* (pour Specifier-Head Agreement Constraint)[21] :

$$(5.49)\quad \begin{bmatrix} \textit{infl-lxm} & \\ \text{TÊTE}\,|\,\text{ACCORD} & \boxed{1} \\ \text{VAL}\,|\,\text{SPEC} & \left\langle \begin{bmatrix} \text{TÊTE}\,|\,\text{ACCORD} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

Les effets de cette contrainte sont tels que chaque élément du lexique qui sélectionne un spécifieur s'accorde avec ce spécifieur (le sujet ou le déterminant).

La modélisation de l'accord entre le sujet et le verbe

La SHAC est une contrainte lexicale et elle fait donc partie des contraintes propres à une langue donnée (ici l'anglais). Elle devrait de ce fait être modifiée pour rendre compte des faits

[20] Cela exclut les lexèmes non fléchis d'être affectés par cette contrainte, tels les noms propres et les pronoms, qui appartiennent à un autre type.
[21] (Sag *et al.*, 2003, p. 238)

suivants propres à l'arabe :

1. Il n'y a aucune raison de regrouper le déterminant et le sujet sous la même catégorie, étant donné que les déterminants dans cette langue ne s'accordent pas avec la tête[22] Nous pouvons ainsi stipuler que l'attribut approprié pour encoder le sujet en arabe est SUJ (et non SPEC)[23].
2. Cet accord implique deux attributs que sont le genre et la personne.
3. La contrainte ne concerne en réalité que les verbes en arabe.

Pour rendre compte de ces particularités, nous proposons une contrainte sur l'accord entre le sujet et le verbe, imposée au type lexical *verbe-mot-accord* et qui prend la forme suivante :

(5.50)
$$\begin{bmatrix} \textit{verbe-mot-accord} & \\ \text{TÊTE} \mid \text{ACCORD} & \boxed{1}\begin{bmatrix} \text{GRE} & \textit{gre} \\ \text{PERS} & \textit{pers} \end{bmatrix} \\ \text{VAL} \mid \text{SUJ} & \left\langle \begin{bmatrix} \text{TÊTE} \mid \text{ACCORD} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix}$$

Cette contrainte impose un partage de structures entre les traits d'accord de la tête verbale et ceux du sujet sélectionné par cette tête. Les traits d'accord en question sont définis comme étant le genre et la personne seulement.

Un exemple d'application

Pour illustrer les effets de cette contrainte sur la sélection opérée par un verbe de son sujet, nous donnons l'exemple suivant :

(5.51) *daxal* *-at* *l-* *bana:t* *-u*
entrer.PER -3F les- filles -NOM
« Les filles sont entrées »

Deux entrées lexicales sont associées aux verbes conjugués à la 3S, dont l'une seulement est sujette à la contrainte proposée en (5.50), étant donné que cette contrainte impose des restrictions

[22] Rappelons que le déterminant défini *?al-* en arabe est le même quels que soient les traits du nom. Le déterminant indéfini *-n* est également invariable.

[23] Il va de soi que nous avons besoin d'une théorie de ce que nous pouvons appeler « spécifieur » en arabe. Cette théorie dépasse le cadre de notre travail.

sur la réalisation du type *verbe-mot-accord*. L'entrée lexicale du verbe *daxalat*, qui nous concerne ici, est présentée dans la structure de traits partielle suivante :

(5.52)
$$\begin{bmatrix} \textit{verbe-mot-accord} & \\ \text{PHON} & \textit{<daxalat>} \\ \text{TÊTE} \mid \boxed{1}\text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fém} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{VAL} & \begin{bmatrix} \text{SUJ} & \left\langle \boxed{2}\begin{bmatrix} \text{ACCORD} & \boxed{1} \end{bmatrix} \right\rangle \end{bmatrix} \\ \text{STR-ARG} & \left\langle \boxed{2} \right\rangle \end{bmatrix}$$

Cette entrée lexicale est légitimée par la contrainte en (5.50). Le sujet compatible avec cette entrée lexicale doit contenir une information compatible dans ses traits d'accord, c'est-à-dire [GRE *fém*] et [PERS *3*]. Ce SN peut contenir une information supplémentaire concernant son nombre [NBRE *sing* V *duel* V *pluriel*] en préservant la possibilité de l'unification avec l'entrée lexicale du verbe en (5.52). Le SN présent dans notre phrase contient une information compatible puisqu'il est à la 3FP :

(5.53)
$$\begin{bmatrix} \textit{syntagme} & \\ \text{PHON} & \textit{<lbana :tu>} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} \mid \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{TÊTE} & \begin{bmatrix} \text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fém} \\ \text{NBRE} & \textit{pluriel} \\ \text{PERS} & \textit{3} \end{bmatrix} \\ \text{CAS} & \textit{nom} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

La structure de traits suivante représente la phrase qui résulte de cette unification :

(5.54)
$$\begin{bmatrix} \textit{syntagme} & \\ \text{PHON} & \textit{<daxalat lbana :atu>} \\ \text{SS} \mid \text{LOC} \mid \text{CAT} & \begin{bmatrix} \text{TÊTE} & \boxed{1} \\ \text{VAL} & \langle\rangle \end{bmatrix} \\ \text{FILLES} & \begin{bmatrix} \text{FILLE-TÊTE} & \begin{bmatrix} \boxed{1}\text{TÊTE} \mid \text{ACCORD} & \boxed{2} \\ \text{VAL} \mid \boxed{3}\text{SUJ} & \left\langle \begin{bmatrix} \boxed{2}\text{ACCORD} & \begin{bmatrix} \text{GRE} & \textit{fém} \\ \text{PERS} & \textit{3} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \\ \boxed{3}\text{FILLE-COMPS} & \begin{bmatrix} \text{TÊTE} \mid \text{ACCORD} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Cette structure de traits illustre l'effet de deux principes et d'une contrainte lexicale. Les deux principes sont le principe de traits de tête, qui impose l'unification des valeurs de l'attribut TÊTE de la mère et de la fille, une unification rendue par le chiffre encadré 1, et le principe de la valence, qui annule les éléments de la valence réalisés par les filles (ici le sujet) une fois que ces éléments sont réalisés pour aboutir à un syntagme saturé (la phrase dans notre exemple) [VAL < >]. L'effet de la contrainte sur le type *verbe-mot-accord* s'observe sur la fille tête, dont les valeurs d'accord sont partagées avec les valeurs d'accord de la seule FILLE-COMPS (le sujet), un partage exprimé par le chiffre encadré 2.

5.2 Implémentation

La modélisation des marqueurs d'arguments que nous avons proposée dans ce chapitre repose en grande partie sur le lexique. Les entrées lexicales et la hiérarchie des types, en plus des contraintes sur ces types, constituent les éléments essentiels qui nous permettent d'analyser les marqueurs d'objet et les marqueurs de sujet.

Pour comprendre comment interagissent les différents modules qui composent la grammaire proposée pour légitimer des suites bien formées contenant des marqueurs d'arguments, nous avons recours à l'implémentation de cette grammaire dans un système de performance qui permet de la tester. Cette implémentation est faite dans un système informatique conçu pour les grammaires à base de contraintes, à savoir LKB (pour *Linguistic Knowledge Building*). Ce système nous permet non seulement d'analyser des suites bien formées et de rejeter celles qui sont mal formées, mais

encore de générer des suites bien formées et rien d'autres que celles-là. La génération est un test fort efficace de la grammaire proposée, puisqu'elle permet de vérifier si les suites produites selon les informations linguistiques fournies au système sont grammaticales ou non et, par conséquent, de vérifier les forces et les faiblesses de la grammaire implémentée.

Cette implémentation a pour seul but de tester notre analyse linguistique et sa faisabilité dans un système de performance. Nous n'avons donc pas l'intention de développer une grammaire complète de l'arabe, une tâche qui dépasse le cadre restreint de notre travail. Toutefois, l'implémentation d'une grammaire des marqueurs d'arguments nous permettra de proposer un noyau qui pourrait servir à développer une grammaire à plus grande échelle, puisque ces marqueurs sont des éléments d'interfaces et font appel à plusieurs composantes de la grammaire de l'arabe.

LKB n'étant pas la seule option disponible quand il s'agit de choisir un système pour implémenter les grammaires à base de contraintes, nous justifions notre choix dans la sous-section subséquente.

5.2.1 Pourquoi LKB ?

Il existe deux environnements de développement capables d'implémenter des grammaires formelles écrites *à la main* dans le cadre de HPSG : LKB et TRALE[24]. LKB sera présenté plus bas. TRALE est une extension récente de ALE (pour *the Attribute-Logic Engine*), un système gratuit de génération et d'analyse grammaticale basé sur la programmation logique (et le langage PROLOG). Ce système a été développé dans le *Computational Linguistics Program* et *the Language Technologies Institute* à la *Carnegie Mellon University* par Bob Carpenter et Gerald Penn[25] dans les années 1990.

> [ALE] integrates phrase structure parsing, semantic-head-driven generation and constraint logic programming with typed feature structures as terms. This generalizes both the feature structures of PATR-II and the terms of Prolog II to allow type inheritance and appropriateness specifications for features and values. Arbitrary constraints may be attached to types, and types may be declared as having extensional structural identity conditions. Grammars may also interleave unification steps with logic program goal calls (as can be done in DCGs), thus allowing parsing to be interleaved with other system components. ALE was developed with an eye toward Head-Driven

[24] Pour une comparaison détaillée des deux systèmes, voir Melnik (2007).
[25] Voir essentiellement Carpenter et Penn (1999, 1995).

> Phrase Structure Grammar (HPSG), but it can also execute PATR-II grammars, definite clause grammars (DCGs), Prolog, Prolog-II, and LOGIN programs, etc. With suitable coding, it can also execute several aspects of Lexical-Functional Grammar (LFG) [26].

TRALE (tout comme ALE) présente plusieurs limitations, dont les suivantes :

- Ce système dépend d'une version payante de PROLOG (SICStus Prolog 3.8.6) pour le faire fonctionner.
- L'interface graphique GRISU ne fonctionne que sous le système d'exploitation Linux. Le système est cependant distribué sous la forme d'un disque compact amorçable qui contient un système d'exploitation complet, TRALE, l'interface graphique et des exemples de grammaires. Ce système nommé GRAMMIX [27] est conçu pour les linguistes qui connaissent et qui travaillent sur l'allemand, avec la possibilité de l'exécuter en anglais.
- Ce système permet la récursion et rend donc le déboggage des grammaires écrites dans le langage PROLOG (les clauses définies) extrêmement difficile pour les linguistes non informaticiens.

À cause de ces limitations, nous avons choisi d'implémenter notre grammaire dans le système LKB, dont nous présentons les avantages dans la section suivante.

5.2.2 Introduction à LKB [28]

LKB, dont nous trouvons une présentation complète dans l'ouvrage de Copestake (2002), a été développé aux débuts des années 1990 par Ann Copestake, John Carroll, Rob Malouf et Stephan Oepen dans le cadre des projets ACQUILEX. Conçu à l'origine comme outil lexical (d'où son nom d'origine *Lexical Knowledge Building*), essentiellement pour construire des lexiques basés sur les structures de traits typées (les TFS), il a évolué pour ressembler à un vrai GDE (*Grammar Development Environment*). On y implémente désormais des grammaires entières pour différentes langues [29] grâce surtout à l'outil Matrix disponible sur le Web à l'adresse http ://www.delph-in.net/matrix/, qui permet, à partir de paramètres bien définis, de créer une ébauche de grammaire compatible avec les données de la langue voulue. Suivant cette évolution, l'outil qu'est LKB est

[26] Voir le site Web du projet à l'adresse : http ://www.cs.toronto.edu/~gpenn/ale.html

[27] Ce système est disponible à l'adresse suivante : http ://hpsg.fu-berlin.de/Software/Grammix/

[28] Cette introduction est basée sur la présentation faite par Copestake (2002) et celle fournie avec le système en guise de manuel de l'utilisateur.

[29] Citons, en guise d'exemple, *the English Resource Grammar* (ERG), développée à l'origine par Copestake et Flickinger (2000), une grammaire du japonais appelée JACY, développée par Siegel et Bender (2002), et *the Modern Greek Resource Grammar*, une grammaire du grec basée sur les deux premières et sur l'outil Matrix en développement à l'*University of Saarland*.

devenu capable d'analyser et de générer des phrases relativement longues.

Cet environnement de développement grammatical multiplateforme et à source ouverte est conçu pour implémenter les grammaires à base de contraintes. En partant d'une seule structure de données, les structures de traits typées, et d'une seule opération formelle, l'unification, la combinaison des deux donne lieu à un outil puissant capable autant de générer que d'analyser. Cet environnement comporte par ailleurs plusieurs outils, dont les suivants :

- Un compilateur de grammaires basées sur les structures de traits typées.
- Un analyseur syntaxique dont l'extrant (en cas de réussite) est une structure arborescente.
- Un générateur d'expressions basé autant sur la syntaxe (les règles) que sur la sémantique.
- Une interface graphique pour interagir avec le système.
- Un *débogger* pour tracer les erreurs.

L'intérêt d'un tel système pour le linguiste est évident compte tenu des deux faits suivants :

1. L'utilisation de ce système ne nécessite pas des connaissances poussées en programmation informatique. Cela s'explique par le fait que les structures de traits typées et l'unification sont conçues à l'origine par des linguistes pour formaliser le comportement d'unités linguistiques.
2. L'implémentation d'une grammaire HPSG dans ce système est facilitée par le fait que le linguiste n'a qu'à reproduire ses représentations formelles des types, des règles et des entrées lexicales telles quelles (ou presque) dans le système. Cela rend évidemment service aux informaticiens, puisque ces derniers peuvent avoir accès aux techniques de la linguistique théorique.

Ce système est employé ici pour écrire une grammaire formelle capable de reconnaître de simples séquences (des phrases) contenant des marqueurs d'arguments. Cette grammaire se compose de plusieurs modules que nous pouvons qualifier de lexical, syntaxique et morphologique. Nous présentons ces modules dans la section suivante.

5.2.3 Grammaire

5.2.3.1 Les modules de base

La grammaire implémentée dans l'environnement LKB comporte les modules suivants :

1. Une hiérarchie de types : ce sous-système agit comme un cadre de définition pour le reste de la grammaire puisqu'il détermine les types admis, les conditions d'appropriation des traits

et le degré de compatibilité des types. Cette hiérarchie permet également d'exprimer des généralisations grâce à l'héritage. Ces informations sont encodées dans un langage simple appelé TDL très proche de la description basée sur les AVM. L'exemple suivant nous donne la définition du type *syntagme* :

```
syntagme := signe & [ COMPS <> ].
```

Cette définition précise le super-type immédiat (ou le parent), ici *signe*, ainsi que les contraintes associées au type en question. La seule contrainte imposée au type *syntagme* ici est le fait que sa liste de compléments soit vide.

La hiérarchie des types implémentée dans LKB doit respecter 3 conditions :

- Avoir une racine unique (ici *top*) dont tous les autres types sont des descendants. Ce type-racine est une exigence formelle et ne reçoit pas en soi de description linguistique (en terme de traits).
- Les cycles ne sont pas permis.
- Si des types sont différents dans la hiérarchie c'est qu'ils sont ou incompatibles (sans aucun descendant commun) ou compatibles et dans ce dernier cas ils ne peuvent avoir qu'un seul descendant commun supérieur (*a unique greatest lower bound*). Voir Copestake (2002, p. 39)

2. Les entrées lexicales . ces entrées sont des structures de traits partielles qui décrivent les mots de la langue. Formellement, une entrée lexicale définit la relation qui unit une suite phonologique (ici tout simplement les caractères qui composent un mot, modélisés par l'attribut PHON) et la description linguistique de ce mot (autrement dit les contraintes). La couverture de notre grammaire étant limitée, nous nous contentons ici de donner quelques entrées lexicales représentatives. L'exemple suivant fournit l'entrée lexicale du mot arabe *walad* « garçon » :

```
walad := lexeme-nom &
[ PHON.LIST.FIRST "walad",
  TETE.ACCORD.GRE masc,
  SEM.RELS.LIST.FIRST.PRED "walad_rel" ].
```

Cette entrée définit le type lexical auquel appartient le mot, ici lexème nominal (*lexeme-nom*), ainsi que plusieurs informations linguistiques : la phonologie[30], modélisée par l'attribut PHON, la sémantique, modélisée par l'attribut SEM et le trait d'accord lexical GRE. Notons ici que l'information sémantique contenue dans cette entrée, et dans toute grammaire

[30] Cette information est essentiellement de nature orthographique.

LKB en général, est basée sur la théorie sémantique connue sous le nom MRS (pour *Minimal Recursion Semantics*), une théorie sémantique computationnelle dont nous pouvons lire une introduction dans l'ouvrage de Copestake *et al.* (1999).

Nous pouvons désigner ces deux premiers modules de la grammaire (la hiérarchie et les entrées) comme définissant le module lexical.

3. Les règles de la grammaire : ces règles sont des structures de traits typées qui décrivent la manière dont les mots et les syntagmes sont combinés pour former d'autres syntagmes. Dans le système LKB, les filles sont encodées dans l'attribut ARGS. La valeur de cet attribut est une liste dont l'ordre des éléments correspond à l'ordre linéaire des filles dans le syntagme. L'exemple suivant fournit la règle correspond aux syntagmes dont la tête est un verbe intransitif :

```
regle-tete-complement-0 := tete-initiale-unaire &
[ SUJ #suj,
  THEME #theme,
  ARGS < mot & [ SUJ #suj,
                 THEME #theme,
                 COMPS <> ] > ].
```

Les règles et les entrées lexicales sont traitées de la même façon dans le système : elles sont considérées comme des entrées, c'est-à-dire que la partie qui suit le symbole « := » ne correspond pas à un nouveau type à ajouter à la hiérarchie, mais à un type déjà contenu dans la hiérarchie et duquel l'entrée en question est une instance.

4. Un symbole de départ : le symbole de départ définit le type de suites acceptées par la grammaire. Étant donné que nous voulons tester la grammaticalité des phrases, notre symbole de départ est la phrase :

```
start := syntagme & [ TETE verbe, SPEC <> ].
```

Nous pouvons qualifier ces derniers modules (les règles de la grammaire et le symbole de départ) comme délimitant le module syntaxique de la grammaire.

5. Des règles flexionnelles : ces règles créent, à la volée et seulement dans l'analyse ou à la demande de l'utilisateur, des mots bien formés à partir des lexèmes fournis dans le lexique. Ainsi, par exemple, les verbes contenus dans notre lexique sont introduits sous forme de radicaux. Les règles flexionnelles ajoutent tout affixe nécessaire à la bonne formation des mots verbaux. L'exemple suivant présente la règle lexicale de la dérivation des verbes perfectifs à la 1MS :

```
rlexicale-verbe-1ms-per :=
%suffix (* tu) (naam nimtu) (xaaf xiftu)
verbe-1ms & [TETE.ACCORD accord-masc-1-sing].
```

Cette règle précise le suffixe qui sera concaténé au radical (ici *-tu*), toute forme non régulière (ici pour les verbes *naam* « dormir » et *xaaf* « craindre »), le type lexical auquel appartient la forme obtenue (ici *verbe-1ms*) et les traits morphosyntaxiques associés à cette forme.

Bien que ce module ait été conçu à l'origine pour encoder la morphophonologie stricte des items lexicaux, nous l'avons étendu pour traiter la morphophonologie en parallèle au traitement des traits morphosyntaxiques (de genre, de nombre et de personne), du mode/aspect et des cas morphologiques. Nous pouvons donc désigner ce module comme étant le module morphologique de la grammaire.

Ces modules sont contenus dans des fichiers textes simples séparés dont l'extension est généralement **.tdl*. La hiérarchie des types est introduite dans le fichier `types.tdl`, les entrées lexicales dans le fichier `lexique.tdl`, les règles dans le fichier `regles.tdl`, le symbole de départ dans le fichier `debut.tdl` et les règles flexionnelles dans le fichier `rflex.tdl`.

D'autres éléments font partie de la grammaire également, comme celui qui définit les types globaux déjà implémentés dans le système et que l'on désire redéfinir (`globaux.lsp`[31] et `globaux-mrs.lsp`[32]), celui des fonctions personnalisées (`fonct-utilisateur.lsp`[33] et `pref-util.lsp`[34]) et les nœuds de l'analyse syntaxique (`noeuds-analyse.tdl`[35]).

5.2.3.2 Le fonctionnement de l'environnement

Nous ne présentons ici que les grandes lignes du fonctionnement du système qui sont susceptibles d'aider le lecteur à avoir une idée générale de ce fonctionnement[36].

[31] Ce fichier contient, entre autres, la définition du type racine de la hiérarchie, à savoir *top*.
[32] Ce fichier contient la définition des fonctions qui constituent la composante sémantique de la grammaire.
[33] Ce fichier fournit les fonctions essentielles pour le compilateur, l'analyseur et le générateur.
[34] Ce fichier contient les options choisies dans l'interface de LKB ainsi que les options implémentées par défaut.
[35] Ce fichier définit les contraintes sur l'assignation des étiquettes que l'on observe dans les structures arborescentes qui résultent d'une analyse réussie d'une suite.
[36] Le lecteur désireux d'avoir plus de détails pourra se référer à l'ouvrage de Copestake (2002).

5.2.3.2.1 Charger une grammaire

Pour charger les modules de la grammaire dans la mémoire du système, nous ouvrons un fichier nommé `script` dont le rôle est d'appeler les différents modules dans un ordre précis. Ce script écrit en Lisp [37] est chargé en faisant appel à la commande *Load grammar* à partir du menu *Grammar* dans la fenêtre *trollet* [38]. Quand la grammaire est chargée et que tous les modules sont vérifiés et qu'ils ne comportent pas d'erreurs, comme nous pouvons le constater dans la figure (5.7), LKB affiche une représentation graphique de la hiérarchie des traits contenue dans le fichier `types.tdl` [39]. En ce qui a trait à notre grammaire, la fenêtre affichée est représentée dans la figure (5.8).

FIGURE 5.7: L'interface *trollet*

[37] L'environnement LKB est programmé en Lisp.

[38] Trollet (pour *TROndheim LingLab Engineering Tool*) est une interface graphique écrite en Perl et GTK+ par Pavel Mihaylov pour remplacer l'interface standard livrée avec LKB, à savoir LKB Top, écrite en Lisp et Motif. Le but de cette interface est de fournir un meilleur support des polices de caractères et de la norme Unicode tout en intégrant les Widgets de l'interface LUI (pour *Linguistic User Interface*) destinée à fournir un affichage proche des AVM. Les captures d'écran réalisés dans ce travail sont faits grâce à Trollet sous Ubuntu Linux.

[39] La hiérarchie n'est affichée que si elle est exempte d'erreurs et respecte les 3 conditions (la racine unique, la non-cylicité et le descendant commun unique).

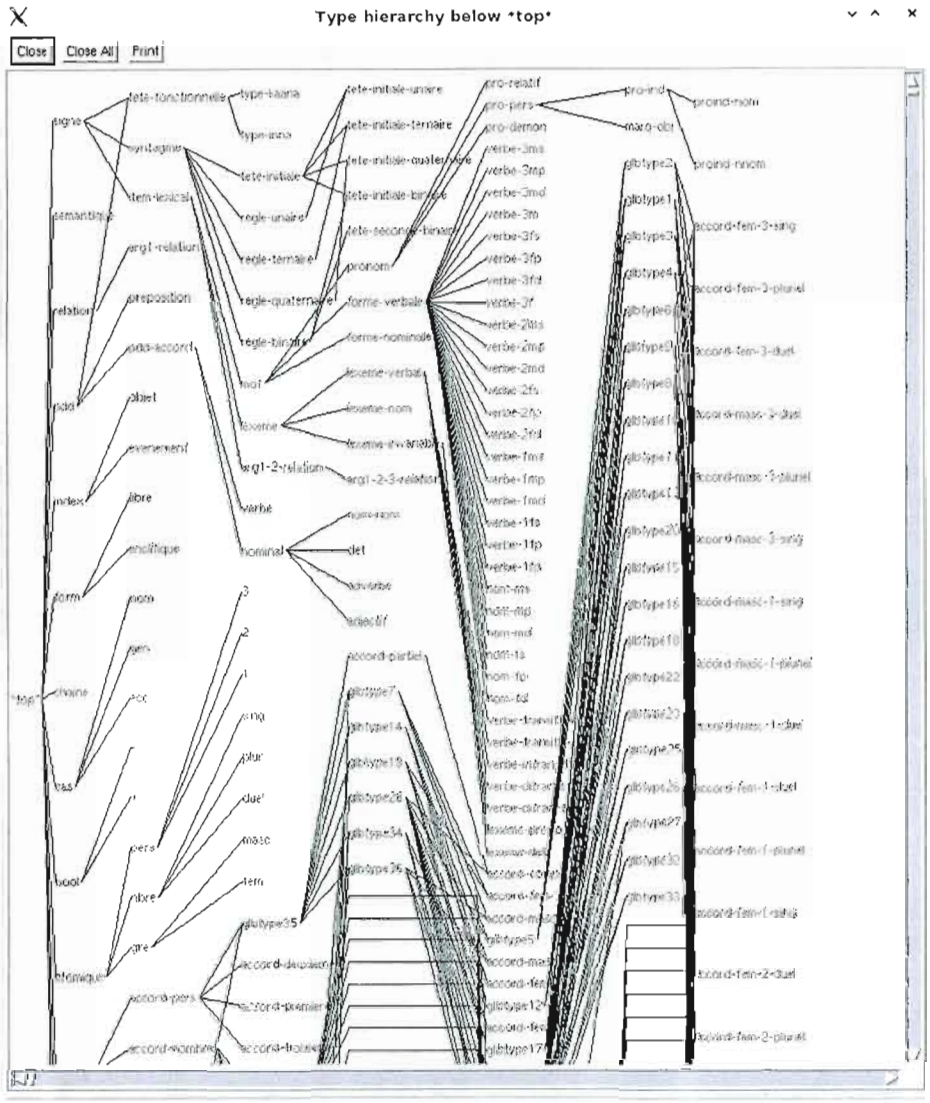


FIGURE 5.8: La hiérarchie des types telle que représentée par LKB

5.2.3.2.2 Inspecter une grammaire

LKB nous permet d'inspecter les différents modules de notre grammaire. En affichant la hiérarchie des traits, il nous permet d'inspecter un type donné et ses contraintes. Cela se fait graphiquement en cliquant sur le nom du type choisi. Un menu avec plusieurs choix se présente

alors. Pour inspecter les contraintes imposées au type voulu, nous avons deux possibilités : montrer les contraintes imposées au type (commande *type definition*) et montrer ces dernières en plus de celles que ce type hérite de ses parents (commande *extended type*). La figure (5.9) représente le résultat du premier choix pour le type *syntagme* et la figure (5.10) le résultat du deuxième choix pour le même type.

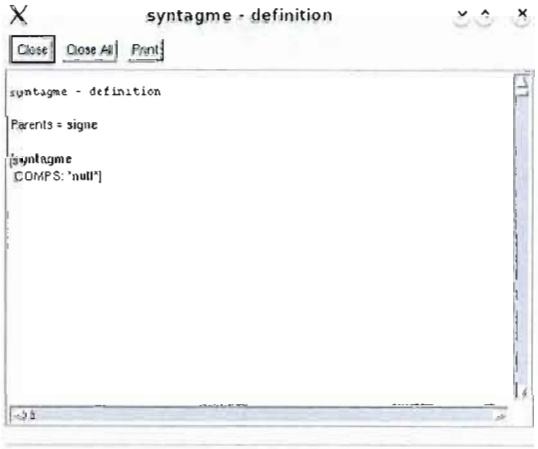


FIGURE 5.9: La définition du type *syntagme*

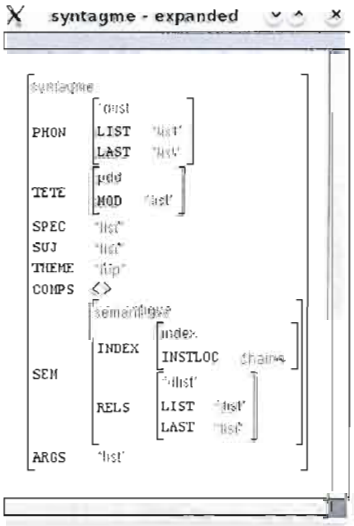


FIGURE 5.10: La définition étendue du type *syntagme*

Nous pouvons également inspecter les règles (tant syntaxiques que flexionnelles) et les entrées lexicales. Cela se fait grâce au menu *View* de la fenêtre trollet et en choisissant *Lex entry, Word entries, Other entry, Grammar rule* ou *Lexical rule*. Prenons comme exemple l'entrée lexicale du verbe *naam* « dormir ». Comme nous l'avons déjà mentionné, notre lexique ne contient que le radical et les contraintes imposées à ce verbe. Les règles flexionnelles se chargent de la création de toutes les formes conjuguées de ce verbe et leur effet peut être constaté (et inspecté) dans le système. La figure (5.11) représente l'entrée du verbe *naam*. Cette entrée est beaucoup plus riche que celle que nous avons consignée dans le lexique puisqu'elle contient non seulement toutes les informations associées à l'entrée lexicale, mais également les contraintes imposées au type auquel appartient le verbe en question et celles héritées des parents de ce type.

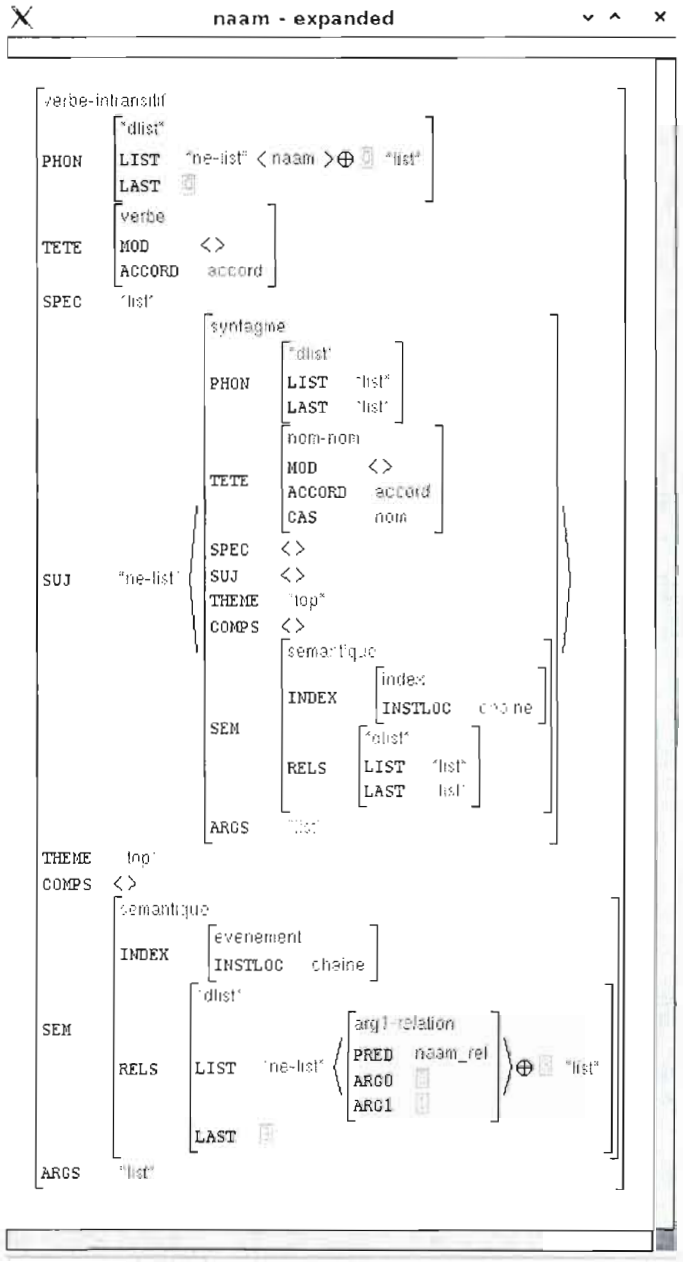


FIGURE 5.11: Le visionnement de l'entrée lexicale du verbe *naam*

En cliquant sur *naam-expended* dans la fenêtre qu'affiche le système et en choisissant l'option *apply all lex rules*, LKB nous présente une fenêtre qui contient le résultat de l'application de

toutes les règles flexionnelles appropriées à ce type. La fenêtre en question est représentée dans la figure (5.12).

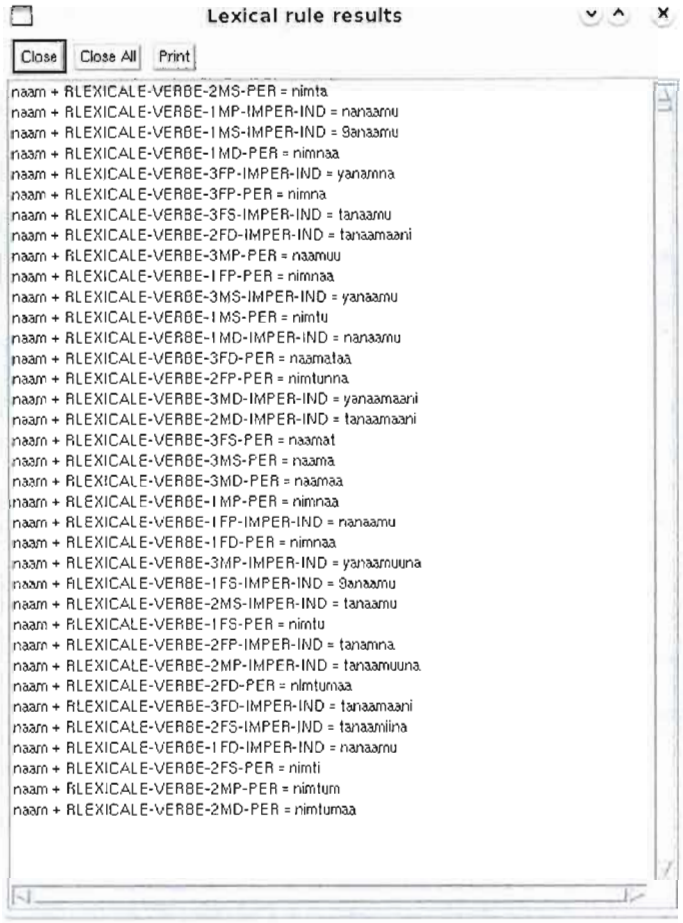


FIGURE 5.12: Le résultat de l'application des règles flexionnelles au radical *naam*

Dans notre travail d'implémentation, nous avons inclus des règles flexionnelles pour les verbes perfectifs, les verbes imperfectifs et les noms. L'ajout d'autres règles morphologiquement productives est relativement aisé, mais le système, dans sa forme actuelle, ne permet pas de traiter la morphologie non concaténative ni une double affixation (la concaténation d'un préfixe et d'un suffixe à la fois).

Nous verrons que pour analyser une phrase qui contient le verbe *naam*, par exemple, ce dernier

est saisi sous sa forme fléchie et le système se charge de faire correspondre cette forme au radical et aux contraintes imposées à son type et à ses parents.

De même, le système nous permet d'inspecter les règles de la grammaire. La figure (5.13) représente le résultat de l'inspection de la règle dite *regle-verbe-initial-2-complements*, une contrainte sur les syntagmes de type *tete-initiale-quaternaire* (avec 4 filles).

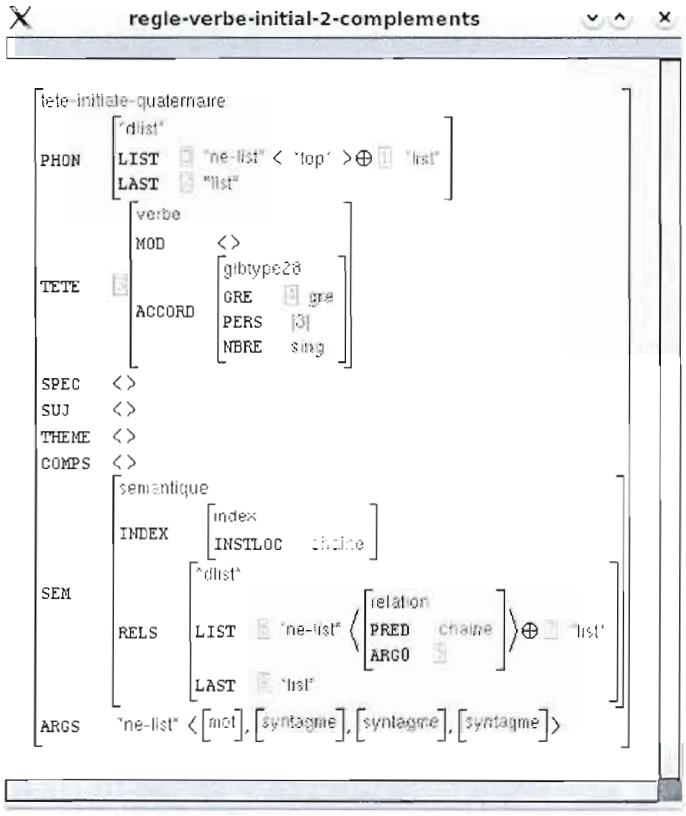


FIGURE 5.13: La représentation de la *regle-verbe-initial-2-complements*

5.2.3.2.3 Analyser une suite

Le système LKB nous permet d'analyser des syntagmes correspondant à la définition du symbole initial fournie dans le fichier `debut.tdl`. Le symbole initial dans notre grammaire est une phrase (le symbole PH) ou un syntagme verbal (le symbole SV). Cela nous permet de rendre compte des suites où le sujet lexical (le SN) est absent de la structure.

Pour analyser une suite, nous cliquons sur le menu *Parse* et nous choisissons l'option *Parse input*. La figure (5.14)[40] représente le résultat de l'analyse réussie de la phrase en (5.55). Une analyse est réussie lorsque la suite soumise à l'analyseur respecte toutes les contraintes de la grammaire, y compris celles qui ont trait à la forme phonologique.

(5.55) *ʔal walad -u katab -a ʔal- qissat -a*
le garçon -NOM écrire.PER 3MS l'- histoire -ACC
« Le garçon, il a écrit l'histoire »

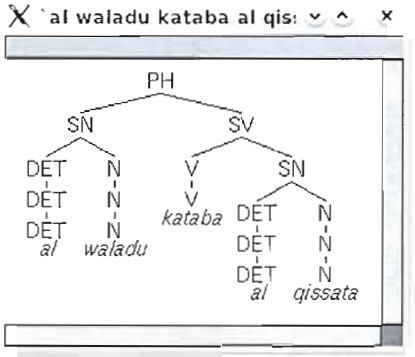


FIGURE 5.14: Résultat de l'analyse de la phrase en (5.55)

Si la suite saisie dans l'analyseur est mal formée, la fenêtre principale de Trollet affiche la mention « `No parses found` ». C'est le résultat obtenu en essayant d'analyser, par exemple, la phrase agrammaticale suivante :

(5.56) **katab -u: ʔal ʔawla:d -u ʔal- qissat -a*
écrire.PER -3MP les- garçons -NOM l'- histoire -ACC

Ce résultat est illustré dans la figure (5.15).

[40] Les représentations arborescentes de LKB contiennent les étiquettes des noeuds (PH, SN, SV, etc.) ainsi qu'une cascade d'étiquettes qui représentent le résultat de l'application des différentes règles (syntaxiques et lexicales) de la grammaire. Ces règles peuvent être inspectées en cliquant sur les étiquettes.

FIGURE 5.15: L'échec de l'analyse de la phrase en (5.56)

Suite à un échec de l'analyse, le système nous permet d'en inspecter la cause grâce au *Chart Parser*. Ce dernier est affiché à partir du menu *Debug*, option *Print parser chart*. Les étapes par lesquelles passe l'analyseur sont ainsi explicitées. Nous pouvons constater, dans la figure (5.16), que l'analyseur ne peut unifier le verbe et le SN en employant les règles syntaxiques fournies dans la grammaire et cet échec de l'unification entraîne le rejet de la phrase.

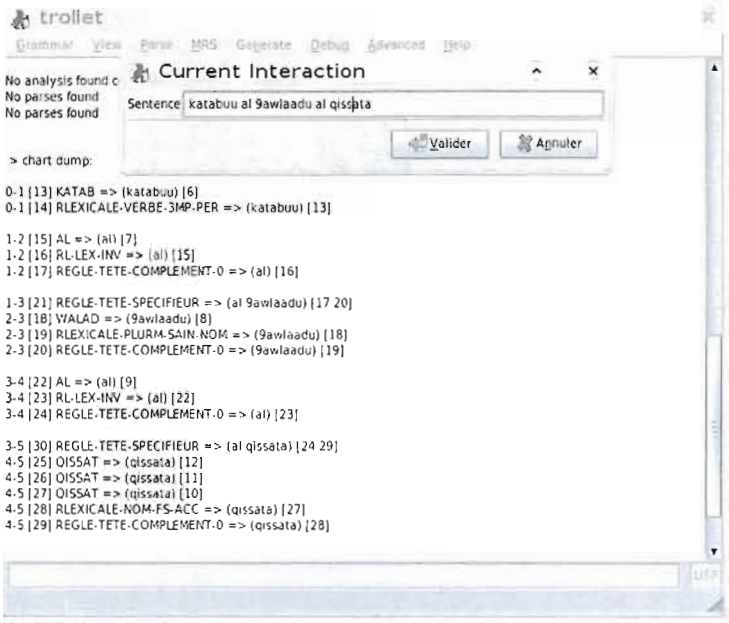


FIGURE 5.16: Le *chart parser* lors d'un échec d'analyse

Nous pouvons ainsi comparer cette figure à celle en (5.17), qui représente le *chart parser* lors d'une analyse réussie de la phrase grammaticale (avec un accord partiel) équivalente à celle en (5.56).

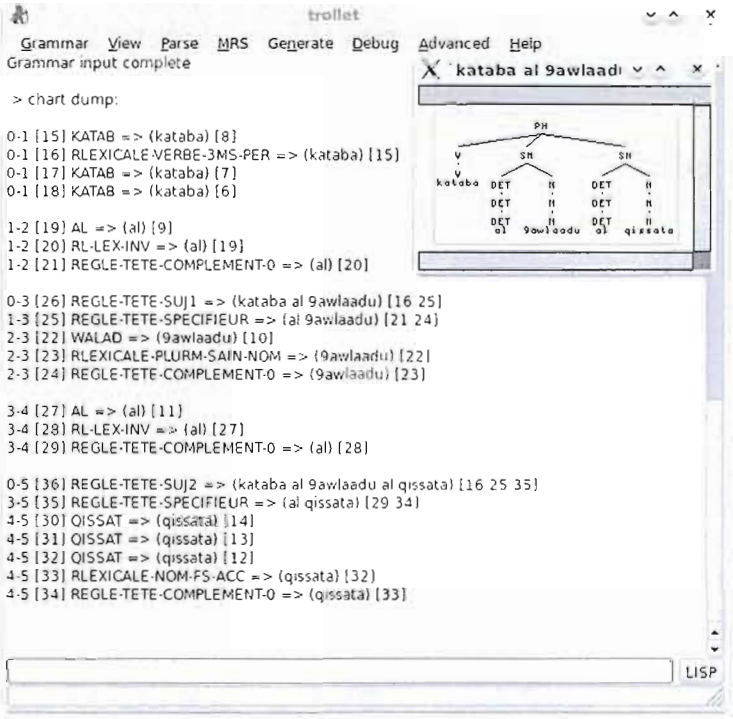


FIGURE 5.17: Le *chart parser* lors d'une analyse réussie

5.2.3.2.4 Générer des suites

Le système nous permet également de générer des phrases en partant des mêmes lexèmes qui sous-tendent les formes réalisées dans la phrase en (5.55), à savoir les 3 lexèmes *walad*, *katab* et *qissat*, en appliquant toutes les règles lexicales appropriées et les règles syntaxiques qui permettent la réalisation des différents nœuds. Le résultat de cette génération est représenté dans la figure (5.18), où nous pouvons constater que toutes les phrases générées sont grammaticales[41].

[41] Notre grammaire n'est pas contrainte sur le plan sémantique et peut donc générer des suites où les rôles thématiques ne sont pas respectés, du genre « l'histoire a écrit le garçon », où le premier SN est nominatif. La cause en est que cette grammaire implémente un nombre restreint d'informations linguistiques pertinentes. Une extension sémantique serait souhaitable dans une grammaire à plus grande échelle.

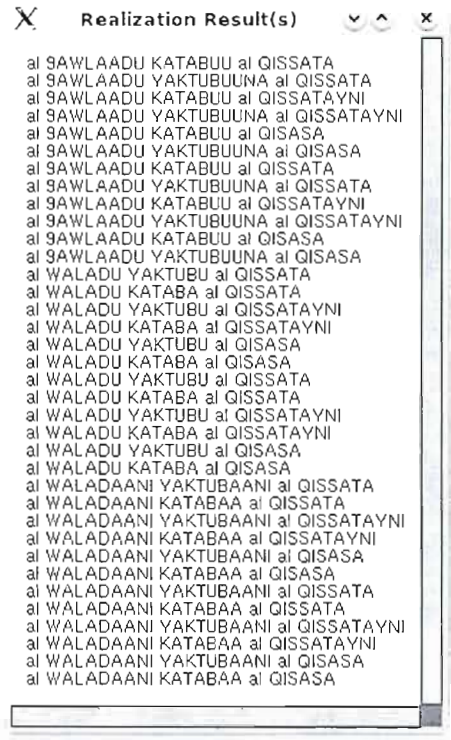


FIGURE 5.18: Un exemple de génération de phrases

5.2.3.3 Notre grammaire

Nous consacrons cette sous-section à la grammaire que nous avons écrite pour analyser et générer des suites contenant des marqueurs d'arguments de l'arabe. Les différents modules présentés ici contiennent des commentaires (reconnaissables grâce au point-virgule en début de ligne) qui facilitent la lecture de chaque module. Nous présentons les résultats obtenus grâce à cette grammaire dans la section (5.2.3.4).

5.2.3.3.1 Le script (`script`)

```
;;; Ce script charge les différents modules de la grammaire dans LKB.
(lkb-load-lisp (this-directory) "globaux.lsp")
(lkb-load-lisp (this-directory) "globaux-mrs.lsp")
```

```
(lkb-load-lisp (this-directory) "fonct-utilisateur.lsp")
(load-lkb-preferences (this-directory) "pref-util.lsp")
(read-tdl-type-files-aux
   (list (lkb-pathname (this-directory) "types.tdl")))
(read-tdl-lex-file-aux
   (lkb-pathname (this-directory) "lexique.tdl"))
(read-morph-file-aux
   (lkb-pathname (this-directory) "rflex.tdl"))

;;; Cette fonction se charge de vérifier toutes les entrées lexicales à la
fois.


(batch-check-lexicon)


(read-tdl-grammar-file-aux
   (lkb-pathname (this-directory) "regles.tdl"))
(read-tdl-start-file-aux
   (lkb-pathname (this-directory) "debut.tdl"))
(read-tdl-parse-node-file-aux
   (lkb-pathname (this-directory) "noeuds-analyse.tdl"))

;;; Cette fonction crée un index qui sert à la génération.


(index-for-generator)
```

5.2.3.3.2 Les types globaux (globaux.lsp)

```
;;; Les types globaux
;;; Types. Le type *top* est la racine de la hiérarchie.
(def-lkb-parameter *toptype* '*top*)
(def-lkb-parameter *string-type* 'chaine
   "Un type particulier - toutes les suites Lisp en sont des sous-types")
```

```
(def-lkb-parameter *orth-path* '(phon list first))
(def-lkb-parameter *list-tail* '(rest))
(def-lkb-parameter *list-head* '(first))
(def-lkb-parameter *diff-list-type* '*dlist*)
(def-lkb-parameter *diff-list-list* 'list)
(def-lkb-parameter *diff-list-last* 'last)

;;; Quelques paramètres. t : true.
(def-lkb-parameter *display-type-hierarchy-on-load* t)
(def-lkb-parameter *simple-tree-display* t)

;;; Analyse
(def-lkb-parameter *mother-feature* NIL
   "Le trait fournissant la mère dans une règle de grammaire")

(def-lkb-parameter *start-symbol* '(start)
   "Les analyses valides")

;;; Génération
(def-lkb-parameter *semantics-index-path* '(SYNSEM CONT INDEX))
```

5.2.3.3.3 Les fonctions sémantiques globales (globaux-mrs.lsp)

```
(in-package "MRS")
(defparameter *initial-semantics-path* `(,(vsym "SEM") ))
(defparameter *main-semantics-path* `(,(vsym "SEM") ,
                                      (vsym "RELS") ,(vsym "LIST")))
(defparameter *top-semantics-type* (vsym "RELATION"))
(defparameter *rel-name-path* `(,(vsym "PRED") ))
(defparameter *rel-handel-path* nil)
(defparameter *psoa-top-h-path* nil)
```

```
(defparameter *psoa-liszt-path* `(,(vsym "RELS") ,(vsym "LIST")))
(defparameter *psoa-rh-cons-path* nil)
(defparameter *ref-ind-type* (vsym "objet"))
```

5.2.3.3.4 Les fonctions principales (fonct-utilisateur.lsp)

```
(defun establish-linear-precedence (rule-fs)

;;; Cette fonction définit l'ordre linéaire des filles.
;;; À l'origine, elle ne permettait pas plus que 3 filles
;;; à la fois sous un même noeud.
;;; Nous l'avons modifiée pour supporter les structures
;;; du type : tête-fille1-fille2-fille3 (verbe-SN1-SN2-SN3).

(let* ((mother NIL)
(daughter1 (get-value-at-end-of rule-fs '(ARGS FIRST)))
(daughter2 (get-value-at-end-of rule-fs '(ARGS REST FIRST)))
(daughter3 (get-value-at-end-of rule-fs '(ARGS REST REST FIRST)))
(daughter4 (get-value-at-end-of rule-fs '(ARGS REST REST REST FIRST)))
)
(declare (ignore mother))
(unless daughter1
(cerror "Ignore it" "Rule without daughter"))
(append (list nil '(ARGS FIRST))
(if daughter2
               (list '(ARGS REST FIRST)))
           (if daughter3
(if daughter2
              (list '(ARGS REST REST FIRST))))
            (if daughter4
                (if daughter3
              (list '(ARGS REST REST REST FIRST))))
)))
```

```
(defun spelling-change-rule-p (rule)
  (let ((rule-type (type-of-fs (tdfs-indef (rule-full-fs rule)))))
    (when (or (eql rule-type 'mot)
              (subtype-p rule-type 'mot)
              (eql rule-type 'lexeme)
              (subtype-p rule-type 'lexeme))
     (let* ((mother (tdfs-indef (rule-full-fs rule)))
              (morth (existing-dag-at-end-of mother *orth-path*))
              (daughter (existing-dag-at-end-of mother '(ARGS FIRST)))
              (dorth (existing-dag-at-end-of daughter *orth-path*)))
     (not (eq morth dorth))))))


(defun redundancy-rule-p (rule)

;;; Cette fonction est utilisée pour empêcher l'analyseur d'essayer d'employer
;;; une règle qui n'est en fin de compte qu'une règle de redondance.

(declare (ignore rule))
  nil)


;;; Cette fonction affiche t (true) pour les types qui ne devraient
;;; pas figurer dans la hiérarchie. Leurs descendants ne seront pas
;;; affichés non plus.

(defun hide-in-type-hierarchy-p (type-name)
  (declare (ignore type-name))
   nil)

(defun make-orth-tdfs (orth)
(let ((indef (process-unifications
             (list (make-unification :lhs
```

```
                                        (create-path-from-feature-list
                                        *orth-path*)
                                        :rhs
                                        (make-u-value
                                        :type orth))))))
      (when indef
        (setf indef (create-wffs indef))
        (when indef
           (make-tdfs :indef indef)))))
```

5.2.3.3.5 Les préférences personnelles (pref-util.lsp)

```
;;; Ce fichier est créé par le système et est édité directement
;;; depuis la fenêtre trollet (menu Advanced- LKB Options).

(defparameter *BRACKETING-P* 'NIL)
(defparameter *COMPARISON-DEPENDENCIES-FONT-SIZE* '12)
(defparameter *COMPARISON-DISCRIMINANT-FONT-SIZE* '8)
(defparameter *COMPARISON-TREE-FONT-SIZE* '7)
(defparameter *DIALOG-FONT-SIZE* '12)
(defparameter *DISPLAY-TYPE-HIERARCHY-ON-LOAD* 'T)
(defparameter *FIRST-ONLY-P* 'NIL)
(defparameter *FRAGMENT-START-SYMBOLS* 'NIL)
(defparameter *FS-TITLE-FONT-SIZE* '12)
(defparameter *FS-TYPE-FONT-SIZE* '12)
(defparameter *GEN-FIRST-ONLY-P* 'NIL)
(defparameter *LEXDB-PARAMS* 'NIL)
; Nous avons augmenté la valeur suivante pour pouvoir générer plus de suites.
(defparameter *MAXIMUM-NUMBER-OF-EDGES* '5000)
(defparameter *PARSE-TREE-FONT-SIZE* '12)
(defparameter *SHOW-LEX-RULES* 'T)
(defparameter *SHOW-MORPHOLOGY* 'T)
(defparameter *SHOW-SPELLING-RULES* 'T)
```

```
(defparameter *START-SYMBOL* '(START))
(defparameter *TYPE-TREE-FONT-SIZE* '12)
```

5.2.3.3.6 Les types (types.tdl)

```
;;; Types standard. Ces types sont susceptibles d'être présents
;;; dans plusieurs grammaires, d'où la notation spéciale avec
;;; le double astérisque **. Le type qui représente la racine
;;; de la hiérarchie est *top* et il est défini dans le module globaux.lsp.

chaine := *top*.
*list* := *top*.
*ne-list* := *list* & [ FIRST *top*, REST *list* ].
*null* := *list*.
*dlist* := *top* & [ LIST *list*, LAST *list* ].
*null-dlist* := *dlist* & [ LIST #last, LAST #last ].
*ne-dlist* := *dlist* & [ LIST *ne-list* ].

;;; Types des valeurs des attributs

form := *top*.
enclitique := form.
libre := form.
bool := *top*.
+ := bool.
- := bool.
cas := *top*.
nom := cas.
acc := cas.
gen := cas.

atomique := *top*.
```

```
gre := atomique.
nbre := atomique.
pers := atomique.
fem := gre.
masc := gre.
sing := nbre.
duel := nbre.
plur := nbre.
1 := pers.
2 := pers.
3 := pers.

;;; Les accords : les combinaisons possibles des traits d'accord en arabe

accord := *top*.
accord-genre := accord & [ GRE gre ].
accord-masc := accord-genre & [ GRE masc ].
accord-fem := accord-genre & [ GRE fem ].

accord-pers := accord & [ PERS pers ].
accord-premiere := accord-pers & [ PERS 1 ].
accord-deuxieme := accord-pers & [ PERS 2 ].
accord-troisieme := accord-pers & [ PERS 3 ].

accord-nombre := accord & [ NBRE nbre ].
accord-sing := accord-nombre & [ NBRE sing ].
accord-duel := accord-nombre & [ NBRE duel ].
accord-pluriel := accord-nombre & [ NBRE plur ].

accord-partiel := accord-genre & accord-pers.
accord-masc-1 := accord-masc & accord-premiere.
accord-masc-2 := accord-masc & accord-deuxieme.
```

```
accord-masc-3 := accord-masc & accord-troisieme.
accord-fem-1 := accord-fem & accord-premiere.
accord-fem-2 := accord-fem & accord-deuxieme.
accord-fem-3 := accord-fem & accord-troisieme.

accord-complet := accord-partiel & accord-nombre.
accord-masc-1-sing := accord-masc-1 & accord-sing.
accord-masc-1-duel := accord-masc-1 & accord-duel.
accord-masc-1-pluriel := accord-masc-1 & accord-pluriel.

accord-masc-2-sing := accord-masc-2 & accord-sing.
accord-masc-2-duel := accord-masc-2 & accord-duel.
accord-masc-2-pluriel := accord-masc-2 & accord-pluriel.

accord-masc-3-sing := accord-masc-3 & accord-sing.
accord-masc-3-duel := accord-masc-3 & accord-duel.
accord-masc-3-pluriel := accord-masc-3 & accord-pluriel.

accord-fem-1-sing := accord-fem-1 & accord-sing.
accord-fem-1-duel := accord-fem-1 & accord-duel.
accord-fem-1-pluriel := accord-fem-1 & accord-pluriel.

accord-fem-2-sing := accord-fem-2 & accord-sing.
accord-fem-2-duel := accord-fem-2 & accord-duel.
accord-fem-2-pluriel := accord-fem-2 & accord-pluriel.

accord-fem-3-sing := accord-fem-3 & accord-sing.
accord-fem-3-duel := accord-fem-3 & accord-duel.
accord-fem-3-pluriel := accord-fem-3 & accord-pluriel.

;;; Types, attributs de base et parties du discours
```

```
signe := *top* &
                [ PHON *dlist*,
                  TETE pdd,
                  SPEC *list*,
                  SUJ *list*,
                  THEME *top*,
                  COMPS *list*,
                  SEM semantique,
                  ARGS *list* ].

tete-fonct := signe & pdd & [ ARGS < syntagme, syntagme > ].

type-inna := tete-fonct &
            [ ARGS < syntagme & [TETE nom-nom, TETE.CAS acc],
                     syntagme > ].
type-kaana := tete-fonct &
[ ARGS < syntagme & [TETE nom-nom, TETE.CAS nom],
         syntagme > ] .

; Les parties du discours (pdd)
pdd := *top* & [ MOD *list* ].
pdd-accord := pdd & [ ACCORD accord ].
nominal := pdd-accord & [ CAS cas ].
nom-nom := nominal & [MOD <>].
verbe := pdd-accord & [MOD <>].
det := nominal & [MOD <>].
preposition := pdd & [MOD < [ SPEC < signe >] >].
adjectif := nominal. adverbe := nominal.

;Quelques types sémantiques

semantique := *top* &
```

```
[ INDEX index,
  RELS *dlist* ].
index := *top* &
[ INSTLOC chaine ].
evenement := index.
objet := index.
relation := *top* &
[ PRED chaine, ARG0 index ].
arg1-relation := relation &
[ ARG1 index ].
arg1-2-relation := arg1-relation &
[ ARG2 index ].
arg1-2-3-relation := arg1-2-relation &
[ ARG3 index ].

;;; Types syntagmatiques

syntagme := signe & [ COMPS <> ].

; Les règles sont déclarées comme des contraintes sur les différents
; sous-types de syntagme.

regle-unaire := syntagme &
[ PHON #phon,
  SEM #cont,
  ARGS < [ PHON #phon, SEM #cont ] > ].

regle-binaire := syntagme &
[ PHON [LIST #ofront, LAST #otail],
  SEM.RELS [LIST #cfront, LAST #ctail ],
  ARGS < [ PHON [LIST #ofront, LAST #omiddle ],
         SEM.RELS [LIST #cfront, LAST #cmiddle ] ],
```

```
       [ PHON [LIST #omiddle, LAST #otail ],
         SEM.RELS [LIST #cmiddle, LAST #ctail ] ] > ].


regle-ternaire := syntagme &
[ PHON [LIST #ofront, LAST #otail],
  SEM.RELS [LIST #cfront, LAST #ctail ],
  ARGS < [ PHON [LIST #ofront, LAST #omiddle1 ],
           SEM.RELS [LIST #cfront, LAST #cmiddle1 ] ],
         [ PHON [LIST #omiddle1, LAST #omiddle2 ],
           SEM.RELS [LIST #cmiddle1, LAST #cmiddle2 ] ],
         [ PHON [LIST #omiddle2, LAST #otail ],
           SEM.RELS [LIST #cmiddle2, LAST #ctail ] ] > ].


regle-quaternaire := syntagme &
[ PHON [LIST #ofront, LAST #otail],
  SEM.RELS [LIST #cfront, LAST #ctail],
  ARGS < [ PHON [LIST #ofront, LAST #omiddle1 ],
           SEM.RELS [LIST #cfront, LAST #cmiddle1 ] ],
         [ PHON [LIST #omiddle1, LAST #omiddle2 ],
           SEM.RELS [LIST #cmiddle1, LAST #cmiddle2 ] ],
         [ PHON [LIST #omiddle2, LAST #omiddle3 ],
           SEM.RELS [LIST #cmiddle2, LAST #cmiddle3 ] ],
         [ PHON [LIST #omiddle3, LAST #otail ],
           SEM.RELS [LIST #cmiddle3, LAST #ctail ] ] > ].


tete-initiale := syntagme &
[ TETE #tete,
  SEM [ INDEX #index ],
  ARGS < [ TETE #tete, SEM [ INDEX #index ] ], ... > ].


tete-initiale-unaire := regle-unaire & tete-initiale.
tete-initiale-binaire := regle-binaire & tete-initiale.
```

```
tete-initiale-ternaire := regle-ternaire & tete-initiale.
tete-initiale-quaternaire := regle-quaternaire & tete-initiale.
tete-seconde-binaire := regle-binaire &
[ TETE #tete,
  SEM [ INDEX #index ],
  ARGS < signe, [ TETE #tete, SEM [ INDEX #index ] ] > ].


;;; Les items lexicaux


item-lexical := signe &
[ PHON [ LIST [ REST #rest ], LAST #rest ] ].


lexeme := item-lexical &
[ SEM [ INDEX #index, RELS <! [ARG0 #index] !> ] ].


lexeme-invariable := lexeme.


lexeme-det := lexeme-invariable &
[ TETE det, SPEC < >, SUJ < >, COMPS < >,
  SEM [ INDEX objet ] ].


lexeme-nom := lexeme &
[ TETE nom-nom,
  SPEC < [TETE det, SEM.INDEX #index ] >,
  COMPS < >,
  SEM [ INDEX objet & #index ] ].


forme-nominale := mot &
[ ARGS < lexeme-nom > ].
```

```
lexeme-verbal := lexeme &
[ TETE verbe,
  SUJ < syntagme & [TETE nom-nom, SPEC <>, SUJ <>] >,
  SEM [ INDEX evenement ] ].


forme-verbale := mot &
[ ARGS < lexeme-verbal > ].


verbe-intransitif := lexeme-verbal &
[ SUJ < [SEM.INDEX #arg1, TETE.CAS nom] >,
  COMPS <>,
  SEM.RELS <! arg1-relation & [ARG1 #arg1] !> ].


verbe-transitif-direct := lexeme-verbal &
[ SUJ < [SEM.INDEX #arg1, TETE.CAS nom] >,
  COMPS < syntagme & [TETE nom-nom,
                      SPEC <>,
                      SEM.INDEX #arg2,
                      TETE.CAS acc] >,
  SEM.RELS <! arg1-2-relation & [ARG1 #arg1, ARG2 #arg2] !> ].


verbe-transitif-indirect := lexeme-verbal &
[ SUJ < [SEM.INDEX #arg1, TETE.CAS nom] >,
  COMPS < syntagme & [TETE preposition,
                      SPEC <>,
                      SEM.INDEX #arg2 ] >,
  SEM.RELS <! arg1-2-relation & [ARG1 #arg1, ARG2 #arg2] !> ].


verbe-ditran-sp := lexeme-verbal &
[ SUJ < [SEM.INDEX #arg1, TETE.CAS nom] >,
  COMPS < syntagme & [TETE nom-nom,
                      SPEC <>,
```

```
                    SEM.INDEX #arg2,
                    TETE.CAS acc],
          syntagme & [TETE preposition,
                     SPEC <>,
                     SEM.INDEX #arg3] >,
  SEM.RELS <! arg1-2-3-relation &
              [ARG1 #arg1, ARG2 #arg2, ARG3 #arg3] !>].


verbe-ditran-sn := lexeme-verbal &
[ SUJ < [SEM.INDEX #arg1, TETE.CAS nom] >,
  COMPS < syntagme & [TETE nom-nom,
                    SPEC <>,
                    SEM.INDEX #arg2,
                    TETE.CAS acc],
         syntagme & [TETE nom-nom,
                    SPEC <>,
                    SEM.INDEX #arg3,
                    TETE.CAS acc] >,
  SEM.RELS <! arg1-2-3-relation &
                    [ARG1 #arg1, ARG2 #arg2, ARG3 #arg3] !>].


lexeme-prepositionnel := lexeme-invariable &
[ TETE preposition & [MOD < [SEM.INDEX #index] > ],
  SPEC <>,
  COMPS < syntagme & [TETE nom-nom,
                    PEC <>,
                    SEM.INDEX #arg1] >,
  SEM [INDEX #index,
       RELS <! arg1-relation & [ARG1 #arg1] !> ]].


;;; Les mots


mot := item-lexical &
```

```
[ TETE #tete,
  SPEC #spec,
  SUJ #suj,
  THEME #theme,
  COMPS #comps,
  SEM #cont,
  ARGS < lexeme & [TETE #tete,
                   SPEC #spec,
                   SUJ #suj,
                   THEME #theme,
                   COMPS #comps,
                   SEM #cont ] > ].


pronom := mot & [ SPEC <>, COMPS <> ].
pro-pers := pronom & [ FORM form ].
pro-demon := pronom.
pro-relatif := pronom.
pro-ind := pro-pers & [ FORM libre ].
marq-obj := pro-pers & [ FORM enclitique ].
proind-nom := pro-ind & [ TETE.CAS nom ].
proind-nnom := pro-ind & [ TETE.CAS acc ].


;;; L'accord des verbes


verbe-1ms := forme-verbale & [ THEME < [ TETE.ACCORD accord-masc-1-sing ] > ].
verbe-1fs := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-1-sing ] > ].
verbe-1md := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-1-duel ] > ].
verbe-1fd := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-1-duel ] > ].
verbe-1mp := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-1-pluriel ] > ].
verbe-1fp := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-1-pluriel ] > ].
verbe-2ms := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-2-sing ] > ].
verbe-2fs := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-2-sing ] > ].
verbe-2md := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-2-duel ] > ].
```

```
verbe-2fd := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-2-duel ] > ].
verbe-2mp := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-2-pluriel ] > ].
verbe-2fp := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-2-pluriel ] > ].
verbe-3ms := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-3-sing ] > ].
verbe-3fs := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-3-sing ] > ].
verbe-3md := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-3-duel ] > ].
verbe-3fd := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-3-duel ] > ].
verbe-3mp := forme-verbale & [ THEME < [TETE.ACCORD accord-masc-3-pluriel] > ].
verbe-3fp := forme-verbale & [ THEME < [TETE.ACCORD accord-fem-3-pluriel] > ].

;;; L'accord des nominaux. Tous les SN sont à la 3e personne.

nom-ms := forme-nominale & [ TETE.ACCORD accord-masc-3-sing ].
nom-fs := forme-nominale & [ TETE.ACCORD accord-fem-3-sing ].
nom-md := forme-nominale & [ TETE.ACCORD accord-masc-3-duel ].
nom-fd := forme-nominale & [ TETE.ACCORD accord-fem-3-duel ].
nom-mp := forme-nominale & [ TETE.ACCORD accord-masc-3-pluriel ].
nom-fp := forme-nominale & [ TETE.ACCORD accord-fem-3-pluriel ].
```

5.2.3.3.7 Les entrées lexicales (lexique.tdl)

```
;;; Le lexique de base. Ce lexique est très réduit, mais
;;; il est suffisant pour tester notre grammaire.

;;; --> Des verbes <--

;écrire

katab := verbe-transitif-direct &
[ PHON.LIST.FIRST "katab",
  SEM.RELS.LIST.FIRST.PRED "katab_rel" ].
```

```
; dormir

naam := verbe-intransitif &
[ PHON.LIST.FIRST "naam",
  SEM.RELS.LIST.FIRST.PRED "naam_rel" ].

; donner

wahab := verbe-ditran-sn &
[ PHON.LIST.FIRST "wahab",
  SEM.RELS.LIST.FIRST.PRED "wahab_rel" ].

; craindre/avoir peur

xaaf := verbe-transitif-indirect &
[ PHON.LIST.FIRST "xaaf",
  SEM.RELS.LIST.FIRST.PRED "xaafa_rel" ].

;;; --> Des noms. Le genre est lexical. <--

; garçon

walad := lexeme-nom &
[ TETE.ACCORD.GRE masc,
  PHON.LIST.FIRST "walad",
  SEM.RELS.LIST.FIRST.PRED "walad_rel" ].

; fille
```

```
bint := lexeme-nom &
[ TETE.ACCORD.GRE fem,
  PHON.LIST.FIRST "bint",
  SEM.RELS.LIST.FIRST.PRED "bint_rel" ].

; histoire

qissat := lexeme-nom &
[ TETE.ACCORD.GRE fem,
  PHON.LIST.FIRST "qissat",
  SEM.RELS.LIST.FIRST.PRED "qissat_rel" ].

;;; --> Un déterminant <--
; le/la/les : ce déterminant défini ne s'accorde pas avec le nom.

al := lexeme-det &
[ PHON.LIST.FIRST "al",
  SEM.RELS.LIST.FIRST.PRED "al_rel" ].

;;; --> des complémenteurs <--

inna := type-inna &
[ PHON.LIST.FIRST "inna",
  SEM.RELS.LIST.FIRST.PRED "inna_rel" ].

kaana := type-kaana &
[ PHON.LIST.FIRST "kaana",
  SEM.RELS.LIST.FIRST.PRED "kaana_rel" ].

;;; --> des prépositions <--
```

```
li := lexeme-prepositionnel &
[ PHON.LIST.FIRST "li",
  SEM.RELS.LIST.FIRST.PRED "li_rel" ].

min := lexeme-prepositionnel &
[ PHON.LIST.FIRST "min",
  SEM.RELS.LIST.FIRST.PRED "min_rel" ].

;;; --> Des marqueurs d'objet <--

hu1 := marq-obj &
[ PHON.LIST.FIRST "hu",
  TETE.CAS acc,
  TETE.ACCORD accord-masc-3-sing,
  SEM.RELS.LIST.FIRST.PRED "hu1_rel" ].

hu2 := marq-obj &
[ PHON.LIST.FIRST "hu",
  TETE.CAS gen,
  TETE.ACCORD accord-masc-3-sing,
  SEM.RELS.LIST.FIRST.PRED "hu2_rel" ].

haa1 := marq-obj &
[ PHON.LIST.FIRST "haa",
  TETE.CAS acc,
  TETE.ACCORD accord-fem-3-sing,
  SEM.RELS.LIST.FIRST.PRED "haa1_rel" ].

haa2 := marq-obj &
[ PHON.LIST.FIRST "haa",
  TETE.CAS gen,
  TETE.ACCORD accord-fem-3-sing,
  SEM.RELS.LIST.FIRST.PRED "haa2_rel" ].
```

```
humaa1-cl := marq-obj &
[ PHON.LIST.FIRST "humaa",
  TETE.CAS acc,
  TETE.ACCORD accord-masc-3-duel,
  SEM.RELS.LIST.FIRST.PRED "humaa1-cl_rel" ].

humaa2-cl := marq-obj &
[ PHON.LIST.FIRST "humaa",
  TETE.CAS gen,
  TETE.ACCORD accord-masc-3-duel,
  SEM.RELS.LIST.FIRST.PRED "humaa2-cl_rel" ].

hum1-cl := marq-obj &
[ PHON.LIST.FIRST "hum",
  TETE.CAS acc,
  TETE.ACCORD accord-masc-3-pluriel,
  SEM.RELS.LIST.FIRST.PRED "hum1-cl_rel" ].

hum2-cl := marq-obj &
[ PHON.LIST.FIRST "humaa",
  TETE.CAS gen,
  TETE.ACCORD accord-masc-3-pluriel,
  SEM.RELS.LIST.FIRST.PRED "hum2-cl_rel" ].

hunna1-cl := marq-obj &
[ PHON.LIST.FIRST "hum",
  TETE.CAS acc,
  TETE.ACCORD accord-fem-3-pluriel,
  SEM.RELS.LIST.FIRST.PRED "hunna1-cl_rel" ].

hunna2-cl := marq-obj &
[ PHON.LIST.FIRST "hunna",
```

```
  TETE.CAS gen,
  TETE.ACCORD accord-fem-3-pluriel,
  SEM.RELS.LIST.FIRST.PRED "hunna2-cl_rel" ].

ii := marq-obj &
[ PHON.LIST.FIRST "ii",
  TETE.CAS gen,
  TETE.ACCORD accord-premiere,
  SEM.RELS.LIST.FIRST.PRED "ii_rel" ].

nii := marq-obj &
[ PHON.LIST.FIRST "nii",
  TETE.CAS acc,
  TETE.ACCORD accord-premiere,
  SEM.RELS.LIST.FIRST.PRED "nii_rel" ].

ka1 := marq-obj &
[ PHON.LIST.FIRST "ka",
  TETE.CAS acc,
  TETE.ACCORD accord-masc-2-sing,
  SEM.RELS.LIST.FIRST.PRED "ka1_rel" ].

ka2 := marq-obj &
[ PHON.LIST.FIRST "ka",
  TETE.CAS gen,
  TETE.ACCORD accord-masc-2-sing,
  SEM.RELS.LIST.FIRST.PRED "ka2_rel" ].

ki1 := marq-obj &
[ PHON.LIST.FIRST "ki",
  TETE.CAS acc,
  TETE.ACCORD accord-fem-2-sing,
  SEM.RELS.LIST.FIRST.PRED "ki1_rel" ].
```

```
ki2 := marq-obj &
[ PHON.LIST.FIRST "ki",
  TETE.CAS gen,
  TETE.ACCORD accord-fem-2-sing,
  SEM.RELS.LIST.FIRST.PRED "ki2_rel" ].

kumaa1 := marq-obj &
[ PHON.LIST.FIRST "kumaa",
  TETE.CAS acc,
  TETE.ACCORD accord-masc-2-duel,
  SEM.RELS.LIST.FIRST.PRED "kumaa1_rel" ].

kumaa2 := marq-obj &
[ PHON.LIST.FIRST "kumaa",
  TETE.CAS gen,
  TETE.ACCORD accord-masc-2-duel,
  SEM.RELS.LIST.FIRST.PRED "kumaa2_rel" ].

kumaa3 := marq-obj &
[ PHON.LIST.FIRST "kumaa",
  TETE.CAS acc,
  TETE.ACCORD accord-fem-2-duel,
  SEM.RELS.LIST.FIRST.PRED "kumaa3_rel" ].

kumaa4 := marq-obj &
[ PHON.LIST.FIRST "kumaa",
  TETE.CAS gen,
  TETE.ACCORD accord-fem-2-duel,
  SEM.RELS.LIST.FIRST.PRED "kumaa4_rel" ].

kum1 := marq-obj &
[ PHON.LIST.FIRST "kum",
```

```
  TETE.CAS acc,
  TETE.ACCORD accord-masc-2-pluriel,
  SEM.RELS.LIST.FIRST.PRED "kum1_rel" ].

kum2 := marq-obj &
[ PHON.LIST.FIRST "kum",
  TETE.CAS gen,
  TETE.ACCORD accord-masc-2-pluriel,
  SEM.RELS.LIST.FIRST.PRED "kum2_rel" ].

kunna1 := marq-obj &
[ PHON.LIST.FIRST "kunna",
  TETE.CAS acc,
  TETE.ACCORD accord-fem-2-pluriel,
  SEM.RELS.LIST.FIRST.PRED "kunna1_rel" ].

kunna2 := marq-obj &
[ PHON.LIST.FIRST "kunna",
  TETE.CAS gen,
  TETE.ACCORD accord-fem-2-pluriel,
  SEM.RELS.LIST.FIRST.PRED "kunna2_rel" ].

;;; --> Des pronoms indépendants nominatifs <--

huwa := proind-nom &
[ PHON.LIST.FIRST "huwa",
  TETE.ACCORD accord-masc-3-sing,
  SEM.RELS.LIST.FIRST.PRED "huwa_rel" ].

hiya := proind-nom &
[ PHON.LIST.FIRST "hiya",
  TETE.ACCORD accord-fem-3-sing,
  SEM.RELS.LIST.FIRST.PRED "hiya_rel" ].
```

```
humaa1 := proind-nom &
[ PHON.LIST.FIRST "humaa",
  TETE.ACCORD accord-masc-3-duel,
  SEM.RELS.LIST.FIRST.PRED "humaa_rel" ].


humaa2 := proind-nom &
[ PHON.LIST.FIRST "humaa",
  TETE.ACCORD accord-fem-3-duel,
  SEM.RELS.LIST.FIRST.PRED "humaa_rel" ].


hum := proind-nom &
[ PHON.LIST.FIRST "hum",
  TETE.ACCORD accord-masc-3-pluriel,
  SEM.RELS.LIST.FIRST.PRED "hum_rel" ].


hunna := proind-nom &
[ PHON.LIST.FIRST "hunna",
  TETE.ACCORD accord-fem-3-pluriel,
  SEM.RELS.LIST.FIRST.PRED "hunna_rel" ].
```

5.2.3.3.8 Les règles flexionnelles (rflex.tdl)

```
;;; Règles de flexion des verbes perfectifs


rlexicale-verbe-1ms-per :=
%suffix (* tu) (naam nimtu) (xaaf xiftu)
verbe-1ms & [TETE.ACCORD accord-masc-1-sing].


rlexicale-verbe-1fs-per :=
%suffix (* tu) (naam nimtu) (xaaf xiftu)
verbe-1fs & [TETE.ACCORD accord-fem-1-sing].
```

```
rlexicale-verbe-1md-per :=
%suffix (* naa) (naam nimnaa) (xaaf xifnaa)
verbe-1md & [TETE.ACCORD accord-masc-1-duel].

rlexicale-verbe-1fd-per :=
%suffix (* naa) (naam nimnaa) (xaaf xifnaa)
verbe-1fd & [TETE.ACCORD accord-fem-1-duel].

rlexicale-verbe-1mp-per :=
%suffix (* naa) (naam nimnaa) (xaaf xifnaa)
verbe-1mp & [TETE.ACCORD accord-masc-1-pluriel].

rlexicale-verbe-1fp-per :=
%suffix (* naa) (naam nimnaa) (xaaf xifnaa)
verbe-1fp & [TETE.ACCORD accord-fem-1-pluriel].

rlexicale-verbe-2ms-per :=
%suffix (* ta) (naam nimta) (xaaf xifta)
verbe-2ms & [TETE.ACCORD accord-masc-2-sing].

rlexicale-verbe-2fs-per :=
%suffix (* ti) (naam nimti) (xaaf xifti)
verbe-2fs & [TETE.ACCORD accord-fem-2-sing].

rlexicale-verbe-2md-per :=
%suffix (* tumaa) (naam nimtumaa) (xaaf xiftumaa)
verbe-2md & [TETE.ACCORD accord-masc-2-duel].

rlexicale-verbe-2fd-per :=
%suffix (* tumaa) (naam nimtumaa) (xaaf xiftumaa)
verbe-2fd & [TETE.ACCORD accord-fem-2-duel].
```

```
rlexicale-verbe-2mp-per :=
%suffix (* tum) (naam nimtum) (xaaf xiftum)
verbe-2mp & [TETE.ACCORD accord-masc-2-pluriel].

rlexicale-verbe-2fp-per :=
%suffix (* tunna) (naam nimtunna) (xaaf xiftunna)
verbe-2fp & [TETE.ACCORD accord-fem-2-pluriel].

rlexicale-verbe-3ms-per :=
%suffix (* a)
verbe-3ms & [TETE.ACCORD accord-masc-3-sing].

rlexicale-verbe-3fs-per :=
%suffix (* at)
verbe-3fs & [TETE.ACCORD accord-fem-3-sing].

rlexicale-verbe-3md-per :=
%suffix (* aa)
verbe-3md & [TETE.ACCORD accord-masc-3-duel].

rlexicale-verbe-3fd-per :=
%suffix (* ataa)
verbe-3fd & [TETE.ACCORD accord-fem-3-duel].

rlexicale-verbe-3mp-per :=
%suffix (* uu)
verbe-3mp & [TETE.ACCORD accord-masc-3-pluriel].

rlexicale-verbe-3fp-per :=
%suffix (* na) (naam nimna) (xaaf xifna)
verbe-3fp & [TETE.ACCORD accord-fem-3-pluriel].
```

```
;;; Règles de flexion des verbes imperfectifs indicatifs.
;;; Le système ne permet pas d'avoir des préfixes et des suffixes
;;; pour une même forme et présente plusieurs autres lacunes qui
;;; nous poussent vers cette solution non élégante : nous avons consigné
;;; les formes obtenues après l'application des règles flexionnelles
;;; comme des formes irrégulières, alors qu'elles sont en réalité
;;; obtenues par des règles productives.

rlexicale-verbe-1ms-imper-ind :=
%prefix (* 9a) (katab 9aktubu) (naam 9anaamu) (wahab 9ahabu) (xaaf 9axaafu)
verbe-1ms & [TETE.ACCORD accord-masc-1-sing].

rlexicale-verbe-1fs-imper-ind :=
%prefix (* 9a) (katab 9aktubu) (naam 9anaamu) (wahab 9ahabu) (xaaf 9axaafu)
verbe-1fs & [TETE.ACCORD accord-fem-1-sing].

rlexicale-verbe-1md-imper-ind :=
%prefix (* na) (katab naktubu) (naam nanaamu) (wahab nahabu) (xaaf naxaafu)
verbe-1md & [TETE.ACCORD accord-masc-1-duel].

rlexicale-verbe-1fd-imper-ind :=
%prefix (* na) (katab naktubu) (naam nanaamu) (wahab nahabu) (xaaf naxaafu)
verbe-1fd & [TETE.ACCORD accord-fem-1-duel].

rlexicale-verbe-1mp-imper-ind :=
%prefix (* na) (katab naktubu) (naam nanaamu) (wahab nahabu) (xaaf naxaafu)
verbe-1mp& [TETE.ACCORD accord-masc-1-pluriel].

rlexicale-verbe-1fp-imper-ind :=
%prefix (* na) (katab naktubu) (naam nanaamu) (wahab nahabu) (xaaf naxaafu)
verbe-1fp & [TETE.ACCORD accord-fem-1-pluriel].
```

```
rlexicale-verbe-2ms-imper-ind :=
%prefix (* ta) (katab taktubu) (naam tanaamu) (wahab tahabu) (xaaf taxaafu)
verbe-2ms & [TETE.ACCORD accord-masc-2-sing].

rlexicale-verbe-2fs-imper-ind :=
%prefix (* ta) (katab taktubiina) (naam tanaamiina) (wahab tahabiina) (xaaf
taxaafiina)
verbe-2fs & [TETE.ACCORD accord-fem-2-sing].

rlexicale-verbe-2md-imper-ind :=
%prefix (* ta) (katab taktubaani) (naam tanaamaani) (wahab tahabaani) (xaaf
taxaafaani)
verbe-2md & [TETE.ACCORD accord-masc-2-duel].

rlexicale-verbe-2fd-imper-ind :=
%prefix (* ta) (katab taktubaani) (naam tanaamaani) (wahab tahabaani) (xaaf
taxaafaani)
verbe-2fd & [TETE.ACCORD accord-fem-2-duel].

rlexicale-verbe-2mp-imper-ind :=
%prefix (* ta) (katab taktubuuna) (naam tanaamuuna) (wahab tahabuuna) (xaaf
taxaafuuna)
verbe-2mp & [TETE.ACCORD accord-masc-2-pluriel].

rlexicale-verbe-2fp-imper-ind :=
%prefix (* ta) (katab taktubna) (naam tanamna) (wahab tahabna) (xaaf taxafna)
verbe-2fp & [TETE.ACCORD accord-fem-2-pluriel].

rlexicale-verbe-3ms-imper-ind :=
%prefix (* ya) (katab yaktubu) (naam yanaamu) (wahab yahabu) (xaaf yaxaafu)
verbe-3ms & [TETE.ACCORD accord-masc-3-sing].
```

```
rlexicale-verbe-3fs-imper-ind :=
%prefix (* ta) (katab taktubu) (naam tanaamu) (wahab tahabu) (xaaf taxaafu)
verbe-3fs & [TETE.ACCORD accord-fem-3-sing].

rlexicale-verbe-3md-imper-ind :=
%prefix (* ya) (katab yaktubaani) (naam yanaamaani) (wahab yahabaani) (xaaf
taxaafaani)
verbe-3md & [TETE.ACCORD accord-masc-3-duel].

rlexicale-verbe-3fd-imper-ind :=
%prefix (* ta) (katab taktubaani) (naam tanaamaani) (wahab tahabaani) (xaaf
taxaafaani)
verbe-3fd & [TETE.ACCORD accord-fem-3-duel].

rlexicale-verbe-3mp-imper-ind :=
%prefix (* ya) (katab yaktubuuna) (naam yanaamuuna) (wahab yahabuuna) (xaaf
yaxaafuuna)
verbe-3mp & [TETE.ACCORD accord-masc-3-pluriel].

rlexicale-verbe-3fp-imper-ind :=
%prefix (* ya) (katab yaktubna) (naam yanamna) (wahab yahabna) (xaaf yaxafna)
verbe-3fp & [TETE.ACCORD accord-fem-3-pluriel].

;;; Règles de flexion des noms. Ces règles s'appliquent
;;; à tous les mots de type lexeme-nom(inal).

rlexicale-nom-ms-nom :=
%suffix (* u)
nom-ms & [TETE.CAS nom].

rlexicale-nom-ms-acc :=
%suffix (* a)
nom-ms & [TETE.CAS acc].
```

```
rlexicale-nom-ms-gen :=
%suffix (* i)
nom-ms & [TETE.CAS gen].

rlexicale-nom-fs-nom :=
%suffix (* u)
nom-fs & [TETE.CAS nom].

rlexicale-nom-fs-acc :=
%suffix (* a)
nom-fs & [TETE.CAS acc].

rlexicale-nom-fs-gen :=
%suffix (* i)
nom-fs & [TETE.CAS gen].

rlexicale-duelm-sain-nom :=
%suffix (* aani)
nom-md & [TETE.CAS nom].

rlexicale-duelm-sain-acc :=
%suffix (* ayni)
nom-md & [TETE.CAS acc].

rlexicale-duelm-sain-gen :=
%suffix (* ayni)
nom-md & [TETE.CAS gen].

rlexicale-duelf-sain-nom :=
%suffix (* aani)
nom-fd & [TETE.CAS nom].
```

```
rlexicale-duelf-sain-acc :=
%suffix (* ayni)
nom-fd & [TETE.CAS acc].

rlexicale-duelf-sain-gen :=
%suffix (* ayni)
nom-fd & [TETE.CAS gen].

rlexicale-plurm-sain-nom :=
%suffix (* uuna) (walad 9awlaadu)
nom-mp & [TETE.CAS nom].

rlexicale-plurm-sain-acc :=
%suffix (* iina) (walad 9awlaada)
nom-mp & [TETE.CAS acc].

rlexicale-plurm-sain-gen :=
%suffix (* iina) (walad 9awlaadi)
nom-mp & [TETE.CAS gen].

rlexicale-plurf-sain-nom :=
%suffix (* aatu) (bint banaatu) (qissat qisasu)
nom-fp & [TETE.CAS nom].

rlexicale-plurf-sain-acc :=
%suffix (* aati) (bint banaati) (qissat qisasa)
nom-fp & [TETE.CAS acc].

rlexicale-plurf-sain-gen :=
%suffix (* aati) (bint banaati) (qissat qisasi)
nom-fp & [TETE.CAS gen].
```

```
;;; Les lexemes invariables. Pour les prépositions, les pronoms
;;; et certains noms non déclinables par les règles précédentes.

rl-lex-inv := mot & [ PHON #1,
                      ARGS < lexeme-invariable & [ PHON #1 ] > ].
```

5.2.3.3.9 Les règles syntaxiques (regles.tdl)

```
;;;; La tête du syntagme est un verbe intransitif

regle-tete-complement-0 := tete-initiale-unaire &
[ SUJ #suj,
  THEME #theme,
  ARGS < mot & [ SUJ #suj, THEME #theme, COMPS < > ] > ].

;;;; La tête est un verbe transitif direct à un seul complément.

regle-tete-complement-1 := tete-initiale-binaire &
[ SUJ #suj,
  THEME #theme,
  ARGS < mot &
             [ SUJ #suj, THEME #theme, COMPS < #1 > ],
         #1 > ].

;;;; La tête est un verbe transitif à deux compléments.

regle-tete-complement-2 := tete-initiale-ternaire &
[ SUJ #suj, THEME #theme, COMPS <>,
  ARGS < mot &
             [SUJ #suj,
```

```
            THEME #theme,
            COMPS < #1 &[TETE nom-nom], #2&[TETE nom-nom] > ],
        #1, #2 > ].

;;; La tête est un élément nominal qui sélectionne un déterminant.

regle-tete-specifieur := tete-seconde-binaire &
[ SPEC < >, COMPS #comps,
  ARGS < syntagme & #1 & [TETE det],
         syntagme & [ SPEC < #1 >,
                      TETE nom-nom,
                      COMPS #comps ] > ].

;;; Cette règle permet d'avoir l'ordre de surface SN+V+(O).

regle-tete-theme := tete-seconde-binaire &
[ THEME < >, SUJ < >, SPEC < >, COMPS #comps,
  ARGS < syntagme & #1 &
                        [TETE nom-nom,
                         TETE.CAS nom,
                         TETE.ACCORD #accord ],
         syntagme &
                     [TETE verbe,
                      TETE.ACCORD #accord,
                      THEME < #1 >,
                      COMPS #comps ] > ].

;;;; Cette règle permet d'avoir l'ordre de surface VS.

regle-tete-suj1 := tete-initiale-binaire &
[ SUJ < >, SPEC < >, THEME < >, COMPS < >,
```

```
  ARGS < mot &
              [TETE verbe,
               TETE.ACCORD.GRE #gre,
               TETE.ACCORD.NBRE sing,
               TETE.ACCORD.PERS 3],
        syntagme & [ TETE nom-nom,
                     TETE.CAS nom,
                     SPEC < >,
                     THEME < >,
                     SUJ < >,
                     TETE.ACCORD.GRE #gre ] > ].


;;;; Cette règle permet d'avoir l'ordre de surface VSO.

regle-tete-suj2 := tete-initiale-ternaire &
[ SUJ < >, SPEC < >, THEME < >, COMPS < >,
  ARGS < mot &
             [TETE verbe,
              TETE.ACCORD.GRE #gre,
              TETE.ACCORD.NBRE sing,
              TETE.ACCORD.PERS 3],
        syntagme & [ TETE nom-nom,
                     SPEC < >,
                     TETE.CAS nom,
                     TETE.ACCORD.GRE #gre],
        syntagme & [SPEC < >,
                    TETE nom-nom,
                    TETE.CAS acc ] > ].

;;;; Cette règle permet d'avoir l'ordre de surface VOS.

regle-tete-suj3 := tete-initiale-ternaire &
```

```
[ SUJ < >, SPEC < >, THEME < >, COMPS < >,
  ARGS < mot & [TETE verbe,
                TETE.ACCORD.GRE #gre,
                TETE.ACCORD.NBRE sing,
                TETE.ACCORD.PERS 3],
         syntagme & [SPEC < >,
                     TETE nom-nom,
                     TETE.CAS acc ],
         syntagme & [TETE nom-nom,
                     SPEC < >,
                     TETE.CAS nom,
                     TETE.ACCORD.GRE #gre] > ].


;;;; Cette règle permet d'avoir 4 filles (V-S-O-O).


regle-verbe-initial-2-complements := tete-initiale-quaternaire &
[ SUJ < >, SPEC < >, THEME < >, COMPS < >,
  ARGS < mot & [TETE verbe,
                TETE.ACCORD.GRE #gre,
                TETE.ACCORD.NBRE sing,
                TETE.ACCORD.PERS 3],
         syntagme & [TETE nom-nom,
                     SPEC < >,
                     TETE.CAS nom,
                     TETE.ACCORD.GRE #gre],
         syntagme & [SPEC < >,
                     TETE nom-nom,
                     TETE.CAS acc ],
         syntagme & [SPEC < >,
                     TETE nom-nom,
                     TETE.CAS acc ] > ].


;;;; Cette règle permet d'avoir un syntagme qui contient
```

```
;;; deux SN (le 2e est gén) ou un SN et un marq-obj (gén).

regle-tete-modifieur-1 := tete-initiale-binaire &
[ SPEC <>, COMPS <>, SUJ <>, THEME <>,
  ARGS < syntagme & #filletete & [TETE nom-nom,
                                  SPEC <>,
                                  COMPS <> ],
         syntagme & [TETE [MOD < #filletete > ],
                     TETE.CAS gen] > ].

;;;; Cette règle permet d'avoir un syntagme prépositionnel.

regle-tete-modifieur-2 := tete-initiale-binaire &
[ SPEC <>, COMPS <>,
  ARGS < mot & [TETE preposition ],
         syntagme & [TETE nom-nom,
                     TETE.CAS gen] > ].

;;;; Ces règles permettent d'avoir des phrases dont la tête est fonctionnelle.

regle-inna-1 := tete-initiale-ternaire &
[ SPEC <>, SUJ <>, THEME <>,
  ARGS < type-inna ,
         syntagme & [TETE nom-nom, TETE.CAS acc],
         syntagme & [TETE verbe] > ].

regle-inna-2 := tete-initiale-ternaire &
[ SPEC <>, SUJ <>, THEME <>,
  ARGS < type-inna ,
         syntagme & [TETE nom-nom, TETE.CAS acc],
         syntagme & [ARGS <mot & [TETE preposition],
                          syntagme & [TETE nom-nom, TETE.CAS gen] > ] > ].
```

```
regle-inna-3 := tete-initiale-ternaire &
[ SPEC <>, SUJ <>, THEME <>,
  ARGS < type-inna ,
         syntagme & [TETE nom-nom, TETE.CAS acc],
         syntagme & [TETE nom-nom, TETE.CAS nom] > ].
```

5.2.3.3.10 Les nœuds d'analyse (nœuds-analyse.tdl)

```
;;; Les nœuds syntaxiques

ph := syntagme & [ TETE verbe, SPEC <>, COMPS <>, SUJ <>, THEME <> ].
sv := syntagme & [ TETE verbe, SPEC <>, COMPS <> ].
v := [ TETE verbe ].
sn := syntagme & [ TETE nom-nom, SPEC <>, COMPS <> ].
n := [ TETE nom-nom, SPEC < [] > ].
det := [ TETE det ].
sp := syntagme & [ TETE preposition, SPEC <>, COMPS <>].
p := [ TETE preposition ].
```

5.2.3.3.11 Le symbole initial (debut.tdl)

```
;;; La structure bien formée est une projection quelconque du verbe
;;; (ce qui inclut les SV et les phrases).

start := syntagme & [ TETE verbe, SPEC <> ].
```

5.2.3.4 Les résultats

La grammaire présentée dans les paragraphes précédents est capable d'analyser (et de générer) des suites bien formées et de rejeter celles qui sont mal formées. Le domaine empirique ainsi couvert englobe la majorité des phénomènes empiriques observés dans les deux premiers chapitres ayant trait aux marqueurs d'arguments. Cette grammaire démontre également que notre

analyse des marqueurs d'arguments se justifie sur un plan computationnel, puisque le test de l'implémentation aboutit à un succès.

Les suites qui nous ont servi de tests de l'analyseur sont regroupés en deux groupes : celles impliquant des marqueurs de sujet (exclusivement) et celles impliquant des marqueurs d'objet.

5.2.3.4.1 L'analyse des marqueurs de sujet

Les suites contenant des marqueurs de sujet sont analysées et générées correctement que ce soit dans les structures à verbe initial, dans celles avec un SN thématisé, en présence d'un pronom indépendant nominatif ou en l'absence de tout sujet lexical. Le traitement est totalement lexicaliste, encodant les marqueurs de sujet comme faisant partie de l'entrée lexicale du verbe, cette entrée étant elle-même générée par des règles flexionnelles. Cette implémentation est le reflet de notre analyse linguistique de ces unités en tant qu'affixes flexionnels, mais également de l'analyse du rôle joué par les SN et les pronoms indépendants pré- et postverbaux.

Les structures à SN disloqué à gauche

Ces structures sont illustrées par les exemples suivants :

(5.57) a. *ʔal binta:ni katab -ata: ʔal- qissat -a*
les- filles.NOM écrire.PER -3FD l'- histoire -ACC
« Les deux filles, elles ont écrit l'histoire »

b. **ʔal binta:ni katab -at ʔal- qissat -a*
les- filles.NOM écrire.PER -3FS l'- histoire -ACC

c. **ʔal binta:ni katab -a: ʔal- qissat -a*
les- filles.NOM écrire.PER -3MD l'- histoire -ACC

L'accord entre le verbe et le thème (le SN disloqué) implique le nombre, le genre et la personne, comme le démontre l'agrammaticalité des exemples en (5.57b et c). L'analyseur LKB reconnaît et génère correctement les suites de type (5.57a) et rejette les suites de types (5.57b et c). Notons également que l'analyseur rejette les suites où le SN thème n'est pas au nominatif et celles où le SN objet n'est pas à l'accusatif. De même, il rejette les suites où la grille de sous-catégorisation du verbe n'est pas respectée (selon son appartenance au type *intransitif*, *transitif direct*, *transitif indirect*, *ditransitif-sn* ou *ditransitif-sp*).

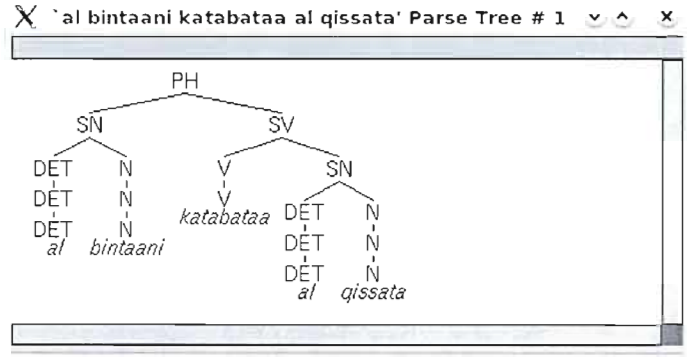


FIGURE 5.19: L'analyse réussie d'une structure dite SVO

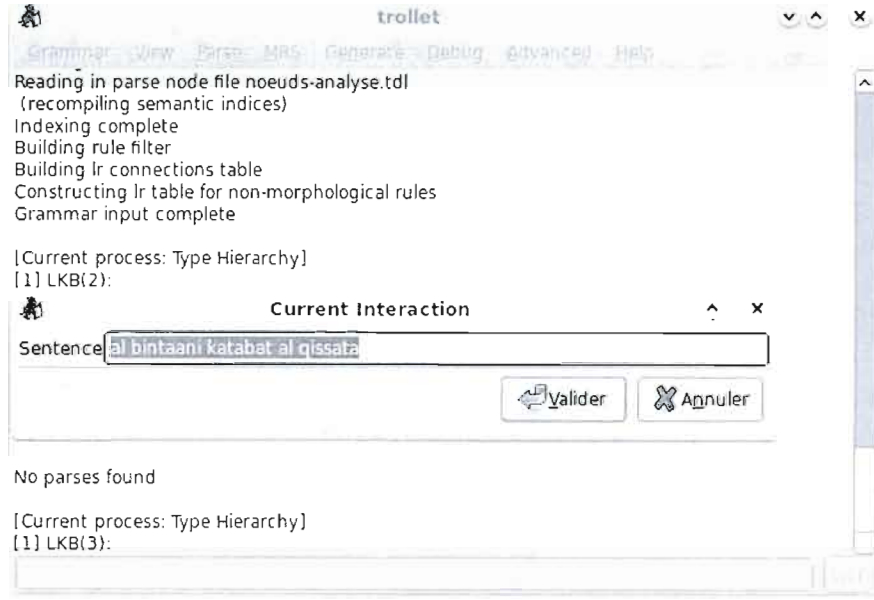


FIGURE 5.20: Échec de l'analyse d'une structure dite SVO mal formée (pas d'accord en nombre)

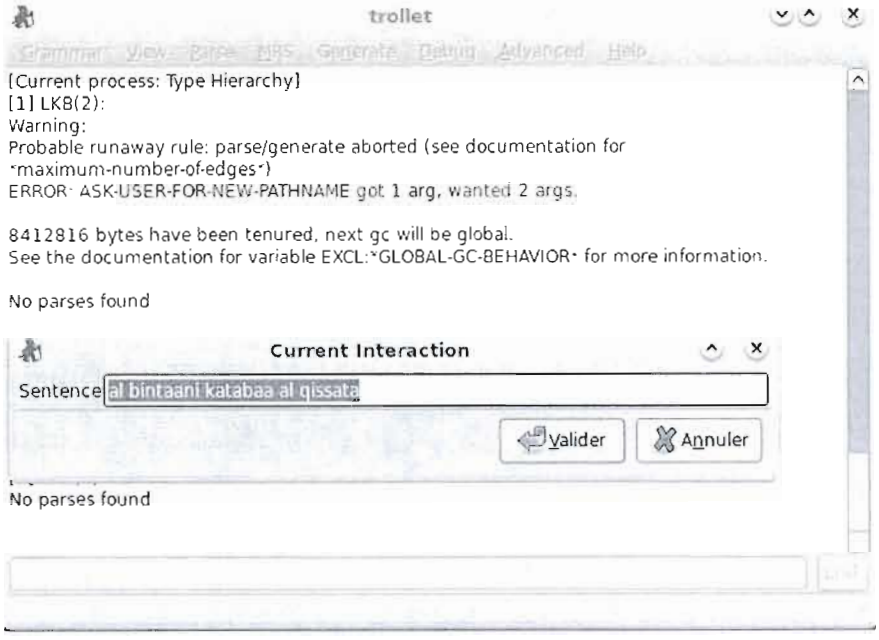


FIGURE 5.21: Échec de l'analyse d'une structure dite SVO mal formée (pas d'accord en genre)

Des résultats semblables sont obtenus lorsque le constituant préverbal est un pronom indépendant nominatif.

Les structures à verbe initial

Ces structures représentent les phénomènes appelés « asymétries de l'accord ». Un verbe initial ne s'accorde avec un SN sujet postposé qu'en genre et en personne. L'accord en nombre est exclu et engendre l'agrammaticalité. Ce comportement est correctement modélisé par notre grammaire. L'analyseur accepte et génère toutes les suites bien formées et rejette celles où le verbe s'accorde en nombre avec le sujet. Les exemples que nous lui soumettons sont les suivants :

(5.58) a. *na:m* *-a* *ʔal* *ʔawla:d* *-u*
dormir.PER -3M les- garçons -NOM
« Les garçons se sont couchés »

b. **na:m* *-u:* *ʔal* *ʔawla:d* *-u*
dormir.PER -3MP les- garçons -NOM

L'analyseur accepte et génère les phrases de type (5.58a) et rejette celles de type (5.58b). Les contraintes sur le cas du sujet (le nominatif) et sur la grille de sous-catégorisation du verbe sont également respectées.

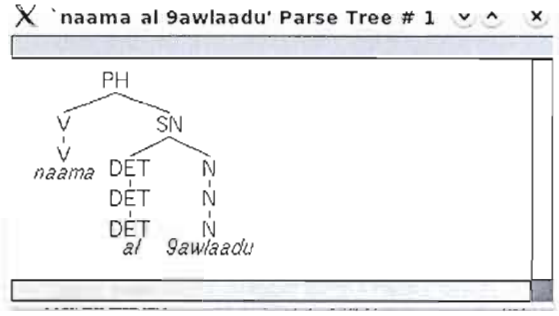


FIGURE 5.22: L'analyse d'une structure VS avec un sujet pluriel (type 5.58a)

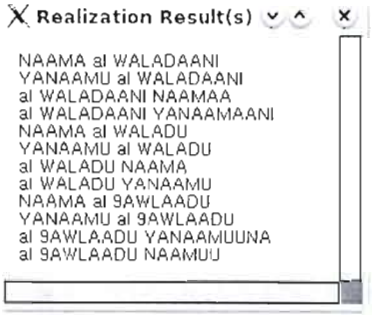


FIGURE 5.23: Génération à partir d'une structure VS (type 5.58a)

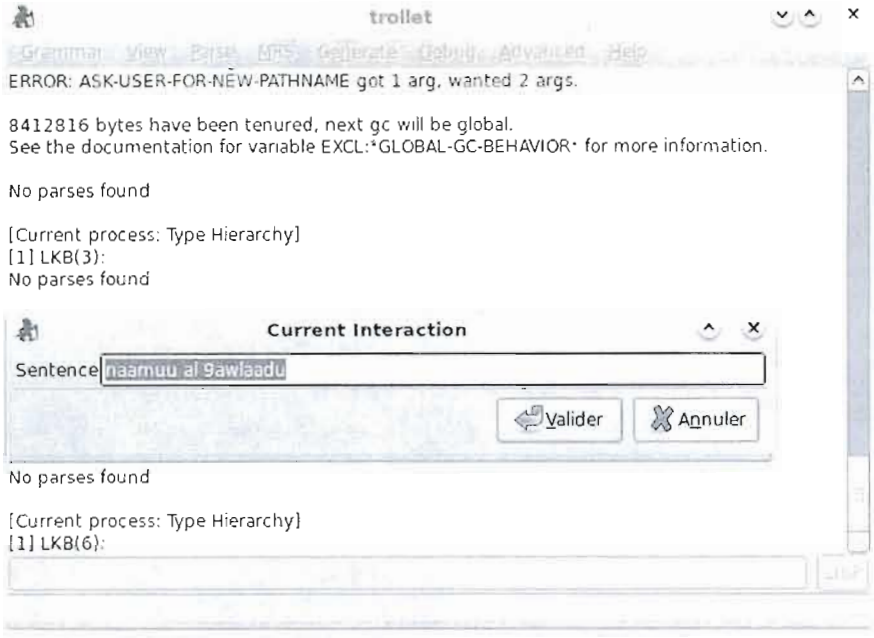


FIGURE 5.24: Échec de l'analyse d'une structure VS (type 5.58b)

L'absence d'un sujet lexical

L'arabe permet l'absence d'un sujet lexical (SN) dans les phrases avec un marqueur de sujet pronominal. Cette possibilité est illustrée dans l'exemple suivant :

(5.59) *katab* *-tu* *ʔal* *qissatajni*
écrire.PER -1S les- histoires.D.ACC
« J'ai écrit les deux histoires »

Les suites de ce type sont correctement analysées et générées par notre grammaire bien que le nœud soit étiqueté SV. Cela revient à la définition de cette catégorie, qui devrait, nous le pensons, être questionnée pour une langue VSO comme l'arabe.

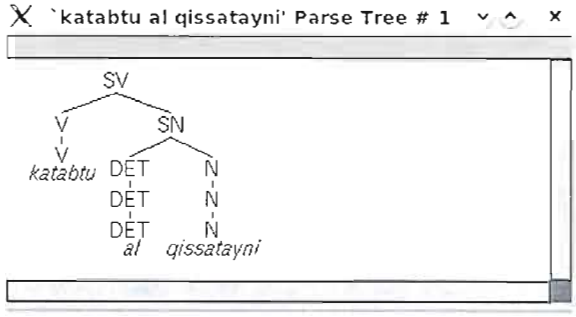


FIGURE 5.25: Analyse des structures sans sujet lexical

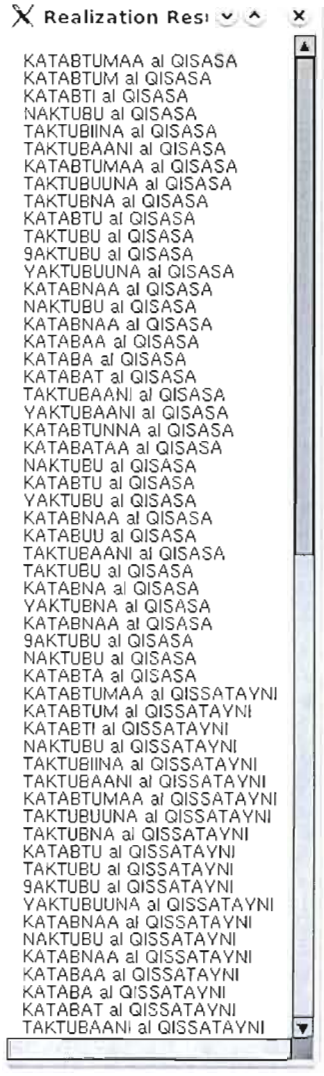


FIGURE 5.26: Génération de structures sans sujet lexical (une partie seulement est montrée)

Notons qu'ici les contraintes sur la grille de sous-catégorisation des verbes sont respectées également.

5.2.3.4.2 L'analyse des marqueurs d'objet

Les marqueurs d'objet peuvent avoir plusieurs hôtes en arabe : des verbes, des prépositions, des nominaux et certaines catégories fonctionnelles (comme les complémenteurs).

Compléments d'un verbe

Les phrases suivantes sont des exemples de structures où un marqueur d'objet est attaché à un verbe.

(5.60) a. *ʔal- walada:ni j- aktub -a: -ni -ha:*
les- garçons.D.NOM 3- écrire.IMP -MD -Ind -la
« Les deux garçons, ils l'écrivent »

b. *j- aktub -u -ha: ʔal- walada:ni*
3M- écrire.IMP -Ind -la les- garçons.D.NOM
« Les deux garçons l'écrivent »

c. *ʔal- bint -u wahab -at -ka -ha:*
la- fille -NOM donner.PER -3FS -te -la
« La fille, elle te l'a donnée »

d. **ʔal- bint -u wahab -at -ha: -ku*
la- fille -NOM donner.PER -3FS -la -te

Dans la phrase en (5.60a), le marqueur d'objet apparaît attaché au verbe avec un SN disloqué à gauche. L'arbre d'analyse de cette structure dans notre grammaire est représenté dans la figure (5.27).

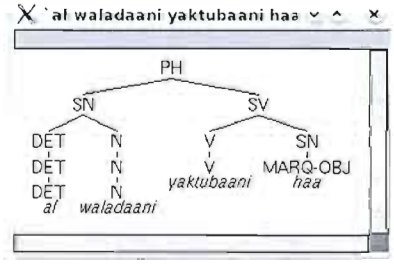


FIGURE 5.27: L'analyse assignée à la structure en (5.60a)

Dans la phrase en (5.60b), le marqueur d'objet sépare le verbe et le SN sujet. L'accord est partiel puisque le verbe est initial et un accord complet entraîne l'agrammaticalité de la phrase. L'analyse de cette structure par notre grammaire donne lieu à la représentation arborescente dans la figure (5.28).

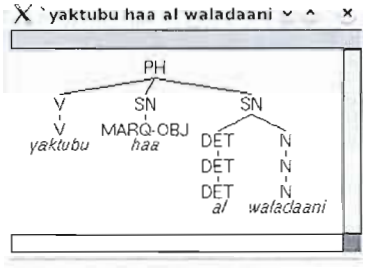


FIGURE 5.28: L'analyse assignée à la structure en (5.60b)

Les deux phrases en (5.60c,d) illustrent les contraintes sur l'attachement de deux marqueurs d'objet à un même verbe. L'ordre à respecter doit obéir à la contrainte de la personne ($1 < 2 < 3$) et à la contrainte thématique (bénéficiaire < thème). L'analyse de ces suites par notre grammaire est conforme à ces attentes : la figure (5.29) est une représentation de l'analyse réussie de la structure en (5.60c) et la figure (5.30) représente l'échec de l'analyse de la phrase agrammaticale en (5.60d).

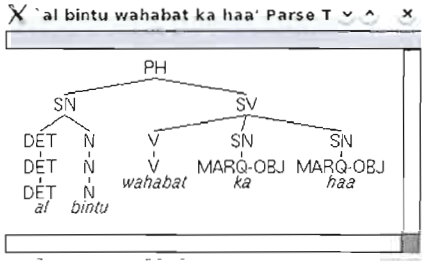


FIGURE 5.29: L'analyse assignée à la structure en (5.60c)

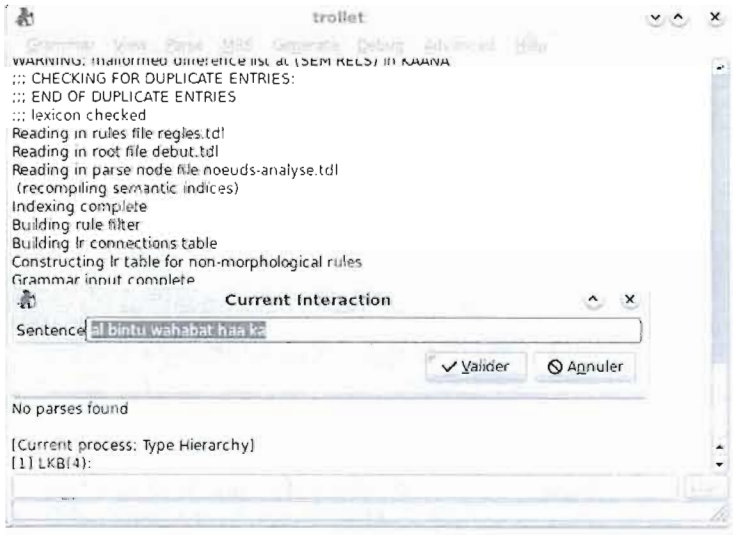


FIGURE 5.30: L'échec de l'analyse de la structure en (5.60d)

Complément d'un nominal

Les marqueurs d'objet peuvent être des compléments de nominaux, comme dans l'exemple suivant :

(5.61) *katab* *-a* *ʔal-* *walad* *-u* *qissat* *-a* *-hu*
écrire.PER -3M le- garçon -NOM histoire -ACC -son
« Le garçon a écrit son histoire »

L'analyse de cette structure donne lieu à la représentation arborescente de la figure (5.31).

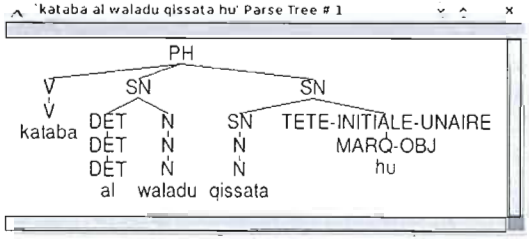


FIGURE 5.31: L'analyse d'une structure contenant un marqueur d'objet complément d'un nominal

Compléments d'une préposition

Les marqueurs d'objet peuvent être des compléments d'une préposition, comme l'illustre l'exemple suivant :

(5.62) *ʔal- walad -u xa:f -a min ha:*
le- garçon -NOM craindre.PER -3MS de -elle
« Le garçon, il a eu peur d'elle »

L'analyse de cette structure donne lieu à la représentation arborescente de la figure (5.32).

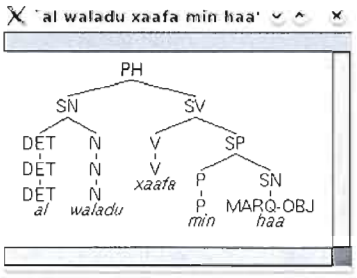


FIGURE 5.32: L'analyse d'une structure contenant un marqueur d'objet complément de préposition

Compléments d'une tête fonctionnelle

Notre grammaire implémente deux têtes fonctionnelles dont l'une peut être hôte des marqueurs de sujet et l'autre hôte des marqueurs d'objet. Cette implémentation est cependant provisoire et devrait être basée sur une étude plus détaillée de ces têtes. Néanmoins, notre grammaire est capable d'analyser trois sortes de structures où l'hôte du marqueur d'objet est un complémenteur de type *ʔinna* « que ». L'étiquetage des nœuds syntaxiques de ces structures explicite les règles associées à la création de chaque noeud.

(5.63) a. *ʔinna -ha: ʔal- bint -u*
que -elle la- fille -NOM
« C'est la fille »

b. *ʔinna -ha: katab -at ʔal- qissat -a*
que -elle écrire.PER -3FS l'- histoire -ACC
« Elle, elle a écrit l'histoire, certes »

c. *ʔinna -ha: min -kum*
que -elle de -vous
« Elle fait partie de [votre groupe de gens], certes »

L'analyse de ces structures nous renvoie les structures arborescentes représentées dans les figures (5.33), (5.34) et (5.35).

FIGURE 5.33: L'analyse assignée à la structure en (5.63a)

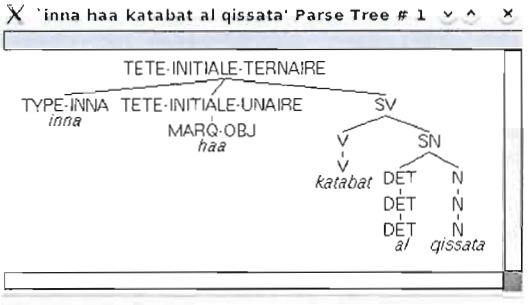


FIGURE 5.34: L'analyse assignée à la structure en (5.63b)

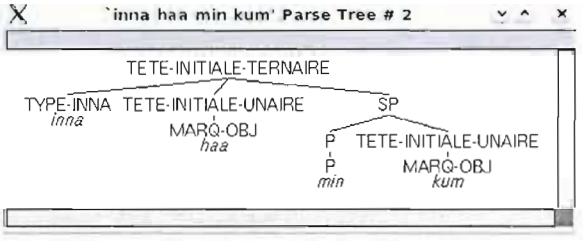


FIGURE 5.35: L'analyse assignée à la structure en (5.63c)

5.3 Conclusion

L'implémentation informatique d'une grammaire des marqueurs d'arguments de l'arabe nous a permis de valider nos choix méthodologiques ainsi que notre analyse tout en permettant de traiter une grande partie de nos données empiriques.

Sur le plan empirique, notre grammaire couvre la majorité des contextes où nous pouvons constater la présence de marqueurs d'arguments. Les différents contextes ainsi couverts englobent les ordres de surface attestés en arabe (essentiellement SVO et VSO), les structures sans sujet lexical, les phrases introduites par des complémenteurs, les SN avec annexion de nom ou de pronom, les SV à double complément (nominal ou pronominal), les structures avec pronom indépendant dans une phrase verbale et les SP avec un dépendant nominal ou pronominal, pour ne citer que ces exemples. Nous avons également implémenté les structures illustrant les asymétries de l'accord suivant l'ordre de surface et suivant la nature du composant antéposé ou postposé au verbe (le SN ou le pronom indépendant). L'ensemble de ces phénomènes est traité de façon à ce que toute suite acceptée par l'analyseur ou générée par ce dernier soit grammaticale et toute suite agrammaticale soit rejetée. La grammaticalité et l'agrammaticalité sont non seulement conformes à nos intuitions, mais encore pouvant avoir des équivalents dans notre corpus de vérification et dans les livres de grammaire que nous avons consultés.

Sur le plan de l'analyse, cette grammaire nous a également permis de tester la pertinence de l'analyse linguistique des marqueurs d'arguments dans un système de performance. Rappelons ici, de ce que nous avons avancé dans les chapitres III et IV, que nous traitons ces marqueurs comme appartenant à deux groupes hétérogènes d'un point de vue morphologique, des affixes et des clitiques postlexicaux. Notre grammaire reflète cette distinction en insérant les marqueurs de sujet comme faisant partie de leurs hôtes verbaux lors de la dérivation des « mots-verbes » à partir des lexèmes par le biais des règles lexicales flexionnelles appropriées. Notre grammaire permet également de rendre compte de l'absence du sujet lexical (SN) dans ces structures où nous traitons le marqueur de sujet comme un argument et non comme un simple marqueur d'accord.

Les marqueurs d'objet sont, par contre, traités comme des « mots » et sont donc insérés dans des nœuds syntaxiques autonomes. Ils sont également traités comme appartenant au type lexical *mot-invariable* pour empêcher les règles flexionnelles de changer leur forme phonologique en leur assignant des cas morphologiques. Tous leurs traits morphosyntaxiques ainsi que leur cas sont de nature lexicale et ils sont donc consignés dans leurs entrées lexicales.

Sur le plan théorique, cette implémentation est fidèle au modèle que nous avons adopté. Elle repose en fait sur une hiérarchie des types qui englobe non seulement les lexèmes, mais également les règles syntaxiques. Ces dernières sont conçues comme des contraintes sur des types particuliers de syntagmes et sont de ce fait intégrées dans la hiérarchie générale. Cette dernière, qui agit comme un cadre de définition pour toute la grammaire, repose sur l'idée suivante : nous pouvons représenter toute information linguistique grâce à la hiérarchie des types en imposant à ces types une seule opération formelle (l'unification) et un seul mode de représentation (les structures de traits typées). L'implémentation de notre grammaire permet ainsi de démontrer la pertinence de ces choix théoriques ainsi que la puissance et la flexibilité du formalisme.

En outre, et même si les règles flexionnelles de notre grammaire restent limitées par les contraintes qu'impose un système conçu à l'origine pour une morphologie strictement concaténative, celle de l'anglais, l'implémentation de ces règles nous a permis de modéliser plusieurs aspects productifs de la morphologie des verbes et des noms de l'arabe. Cette implémentation permet une grande économie dans les entrées lexicales. Cette économie est souhaitable pour des raisons à la fois informatiques et linguistiques. Sur un plan strictement informatique, le traitement de la grammaire est accéléré grâce aux règles flexionnelles et pourrait être un critère décisif dans l'implémentation d'un lexique à une plus grande échelle. Sur le plan linguistique, cette manière d'aborder la morphologie est une façon adéquate d'exprimer les généralisations dont abonde le lexique des langues naturelles. De plus, pour une langue à la morphologie riche comme l'arabe, cela devient une nécessité. Nous pouvons donc affirmer que notre choix d'employer le dispositif des règles lexicales se justifie sur le plan computationnel et sur le plan purement linguistique. En fait, une théorie comme celle de Koenig (1999), où aucune règle lexicale n'est employée, serait difficilement implémentable et impliquerait probablement une explosion du lexique et une perte de généralisation.

Malgré la couverture réduite de notre grammaire LKB, nous pensons qu'elle constitue un noyau solide au tour duquel nous pouvons bâtir une grammaire complète de l'arabe. Notre grammaire est en fait à l'image de notre sujet d'étude, un élément d'interface qui fait appel à un grand nombre de phénomènes linguistiques et permet d'en observer les interactions possibles.

CONCLUSION

Dans le cadre de cette thèse, nous souhaitions étudier les marqueurs d'arguments de l'arabe standard, à savoir les marqueurs de sujet et les marqueurs d'objet, dans le cadre de l'une des grammaires à base de contraintes, la grammaire syntagmatique endocentrique HPSG. Cette étude se voulait, à la fois, descriptive, analytique et formelle.

Dans une perspective purement descriptive, nous voulions établir plusieurs faits concernant les marqueurs d'arguments avant d'entreprendre l'analyse en tant que telle. Nous pensons, en fait, que cette étape est primordiale, voire nécessaire, pour tenir compte du plus grand nombre possible de données. Pour atteindre ce but, qualifions-le d'adéquation descriptive en reprenant un terme chomskyen mais dans un sens légèrement différent, nous nous sommes constitué un corpus de vérification. Ce corpus englobe des données textuelles et audio provenant de plusieurs sources qui reflètent les deux variétés de l'arabe standard, à savoir l'arabe classique et l'arabe moderne. Nous avons également consulté un grand nombre de livres de grammaire parce que nous pensons que ces derniers regorgent non seulement de données pertinentes, mais encore d'intuitions, d'analyses et d'explications fort instructives. C'est le cas, par exemple, du Livre de Sibawayhi et de Charh El Kaafiya de Astarabaadii. Nous nous sommes également servi de nos intuitions personnelles pour construire des exemples simplifiés imitant les structures rencontrées dans notre corpus.

Lors de ce travail descriptif, nous avons isolé les phénomènes qui impliquent les marqueurs d'arguments en les confrontant aux phénomènes impliquant les pronoms indépendants et les syntagmes nominaux. Nous avons également recensé les similitudes et les divergences entre les deux sous-groupes de ces marqueurs, à savoir les marqueurs de sujet et les marqueurs d'objet. Nous avons ainsi remarqué qu'il existe, dans la grammaire de l'arabe, un réseau de liens complexes entre ces différentes unités. D'une part, les marqueurs de sujet partagent avec les marqueurs d'objet certaines propriétés morphosyntaxiques (comme le sous-spécification du trait de nombre), catégorielles (l'étiquette pronominale est justifiée en prenant en compte les définitions fournies par Bresnan (2001)) et phonologiques (l'attachement et la sensibilité aux règles de l'assignation de l'accent). Ce partage de propriétés n'empêche pas l'apparition de plusieurs différences tant sur le plan syntaxique (le comportement vis-à-vis des structures à corroboration appositive, par exemple), que morphologique (les frontières du mot morphologique englobant les marqueurs de

sujet et excluant les marqueurs d'objet). D'autre part, les marqueurs d'arguments partagent avec les pronoms indépendants certaines propriétés, comme l'appartenance catégorielle et les similitudes morphophonologiques, mais ils s'en démarquent par leurs propriétés distributionnelles ainsi que sur le plan purement phonologique de l'attachement. En outre, les marqueurs d'arguments semblent partager avec les syntagmes nominaux certaines propriétés morphosyntaxiques compte tenu de leur argumentalité dans certains contextes et de leur rôle d'explétifs dans d'autres, mais les différences entre ces unités et ces SN sont évidentes ne serait-ce que sur le plan morphologique et sur le plan morphophonologique. Finalement, les syntagmes nominaux sont en distribution complémentaire avec certains pronoms indépendants. C'est le cas lorsque nous observons les occurrences des pronoms indépendants accusatifs, dont la position préverbale est obligatoire alors que les SN accusatifs peuvent occuper autant cette position que celle qui est postposée. Toutefois, la position postverbale accueille également les marqueurs d'objet, qui se trouvent ainsi en distribution complémentaire avec les pronoms indépendants accusatifs.

Ce réseau de liens, constitué des marqueurs d'arguments, des pronoms indépendants et des syntagmes nominaux, illustre la place occupée par ces marqueurs dans le système grammatical de l'arabe. Nous pouvons ainsi soutenir l'idée selon laquelle ils sont des éléments d'interface dont l'étude doit faire appel non seulement à la syntaxe, mais également à la phonologie et à la morphologie en plus des interfaces entre ces systèmes, à savoir la morphophonologie et la morphosyntaxe. C'est pour cette raison que le volet analytique de notre thèse était centré sur deux questions : le statut morphologique et le statut morphosyntaxique des marqueurs d'arguments. Le problème que nous voulions résoudre est de savoir dans quel module de la grammaire nos marqueurs doivent être étudiés ?

Pour répondre à cette question, nous avons consacré le chapitre III à la question morphosyntaxique et le chapitre IV à la question morphologique. La question morphosyntaxique s'énonce dans ces termes : si les marqueurs d'arguments entrent dans des relations distributionnelles (parfois, de complémentarité), peut-on les analyser comme des arguments ? La question morphologique s'énonce ainsi : si les marqueurs d'arguments ont cette propriété essentielle d'être attachés à des hôtes, sont-ils des affixes, faisant ainsi partie de leurs hôtes, ou des clitiques postlexicaux, attachés à leurs hôtes en une phase postlexicale ? Nous avons posé ces deux questions en gardant en tête notre hypothèse de départ, à savoir que le statut morphologique et le statut morphosyntaxique n'ont pas à avoir une relation un-à-un. En d'autres termes, nous sommes d'avis, tout comme Auger (1994), que les affixes n'ont pas à être forcément des marqueurs d'accord. Cette idée se vérifie-t-elle dans le cas des marqueurs d'arguments de l'arabe ?

Pour répondre à la première question, nous avons passé en revue les deux hypothèses principales défendues dans la littérature : l'hypothèse dite argumentale et celle dite flexionnelle. Nous avons démontré que ces deux hypothèses posent certains problèmes et n'embrassent pas le domaine empirique que nous avons défini dans le chapitre II. Nous avons opté pour une hypothèse qui reflète nos observations empiriques dans le chapitre III et nous avons donc soutenu l'idée selon laquelle les marqueurs de sujet ne doivent pas être traités sur un pied d'égalité avec les marqueurs d'objet lorsque l'on entreprend d'analyser leur statut morphosyntaxique. Les premiers regroupent en réalité deux sortes d'unités : des marqueurs d'accord et des pronoms. Les marqueurs d'accord sont exclusivement de troisième personne du singulier et se rencontrent dans les structures à verbe initial, à savoir en présence d'un sujet lexical (un SN nominatif postposé). Les pronoms sont les autres unités du paradigme, mais également les unités portant les traits 3S. Nous constatons ainsi qu'il existe une certaine ambiguïté fonctionnelle dans le paradigme des marqueurs de sujet de 3S, puisque ces derniers couvrent deux réalités différentes, celle de pronoms et celle de marqueurs d'accord. Les pronoms se distinguent des marqueurs d'accord par la spécification du nombre, ce trait étant sous-spécifié dans les unités du deuxième type.

Cette analyse a été confrontée à l'un des problèmes les plus débattus de la linguistique arabe, à savoir celui des asymétries de l'accord. Ce dernier a été présenté de manière systématique dans le chapitre III. Nous avons ainsi présenté toutes les données que nous avons cru pertinentes, nous avons classifié les approches proposées selon les courants d'analyse et nous avons exposé ce que nous croyons être une manière d'aborder la question tout en ayant à l'esprit toutes les données pertinentes. Ce travail nous a poussé à nous interroger sur le statut précis des structures appelées SVO. Selon notre analyse, ces dernières n'ont pas d'existence réelle en arabe si le « S » est considéré comme le sujet. Ce constituant nominatif préverbal, qu'il soit co-indicé avec un marqueur de sujet ou un marqueur d'objet, est disloqué. Sans que ce terme à saveur transformationnelle n'implique un déplacement syntaxique, il désigne dans notre analyse le fait qu'un constituant soit réalisé en périphérie de la phrase en maintenant des liens de coréférence avec les marqueurs d'arguments qui sont, eux, des arguments de la tête.

Cette analyse est corroborée par nos observations lors de l'écoute de notre corpus audio : chaque fois qu'un constituant nominatif préverbal est réalisé, il est séparé du reste de la phrase par une coupure intonatoire. Cette coupure étant le trait distinctif entre la dislocation et le redoublement selon la théorie sur laquelle nous nous basons, nous en concluons que les SN (et les pronoms indépendants) nominatifs préverbaux sont disloqués. À la lumière de cette approche, nous pouvons avancer l'idée selon laquelle les asymétries de l'accord ne sont, après tout, que

l'expression de deux relations d'accord différentes : un accord anaphorique, non local, et un accord grammatical, local de nature. Ces notions proviennent de l'ouvrage de Bresnan et Mchombo (1987) et l'application des tests proposés par ces auteurs pour distinguer ces deux sortes d'accord donne lieu au constat suivant : l'ordre dit SVO illustre un cas d'accord anaphorique en tous points semblable à l'accord anaphorique entre un objet préposé et le marqueur d'objet, l'ordre VSO, au contraire, est un contexte d'accord grammatical. Ainsi, le véritable accord entre le sujet et le verbe se réalise dans les structures où le verbe est initial et il est un accord partiel (avec une sous-spécification du nombre). Aussi peu naturel que cela puisse paraître, l'accord grammatical en arabe est caractérisé par une certaine pauvreté ou sous-spécification qui touche autant le système nominal (en y incluant l'accord adjectival) que le système verbal.

Nous avons également soulevé d'autres questions touchant la problématique des asymétries de l'accord, dont celle de la variation dialectale. Les dialectes arabes contemporains réalisent un accord riche dans l'ordre VSO et ce comportement constitue un défi à toutes les approches de la question. Nous avons proposé de traiter cette variation en faisant appel à l'évolution diachronique et en insistant sur les différences existant au sein même des dialectes étudiés.

La morphosyntaxe des marqueurs d'objet est moins problématique même si un auteur, Shlonsky (1997), propose une analyse en termes de projection d'accord AGR. Nous avons discuté de cette approche et nous avons démontré qu'elle ne peut rendre compte de plusieurs phénomènes observés. Notre analyse, sur ce point, fait la distinction entre des marqueurs d'objet argumentaux (essentiellement compléments de têtes lexicales) et de marqueurs d'objet non argumentaux (compléments de têtes fonctionnelles).

Soulever la question de la morphosyntaxe des marqueurs d'arguments était pour nous une occasion d'aborder plusieurs questions qui ont trait à la syntaxe de l'arabe. Ces marqueurs ont ainsi prouvé leur trait de caractère le plus imminent : on ne peut les analyser sans toucher au coeur même du système grammatical de cette langue.

La deuxième question à laquelle nous avons essayé de répondre dans notre thèse concernait l'aspect morphologique des marqueurs d'arguments : sont-ils des affixes ou des clitiques postlexicaux ? L'importance de poser cette question provient du fait que ces deux statuts morphologiques impliquent deux sortes différentes de traitement : un traitement lexical dans le cas des affixes et un traitement syntaxique dans le cas des clitiques.

Pour répondre à cette question, nous avons eu recours aux tests communément appelés les

tests de Zwicky ainsi qu'aux critères d'autres auteurs, dont Miller (1992). Nous avons passé en revue ces différents tests, nous avons testé certaines de leurs prédictions sur des données de l'arabe et nous les avons appliqués aux marqueurs d'arguments. Il s'en est suivi que ces unités sont hétérogènes sur le plan morphologique : les marqueurs de sujet sont compatibles avec l'étiquette d'affixes (flexionnels) alors que les marqueurs d'objet présentent à la fois les propriétés des mots et celles des affixes. Cela nous a permis de statuer, en ce qui concerne ces derniers, en proposant de les traiter comme clitiques, étant donné qu'un clitique est, au fond, une unité langagière dont les propriétés sont un amalgame de celles des mots et de celles des affixes. Ce statut intermédiaire a, à maintes reprises, été avancé comme étant le critère déterminant de ce qui est un clitique par des auteurs comme Zwicky (1977) et Miller et Monachesi (2003).

Le tableau suivant résume ces résultats :

	Morphosyntaxe	Morphologie
Marqueurs de sujet	arguments, marqueurs d'accord	affixes
Marqueurs d'objet	arguments, explétifs, thèmes	clitiques

TABLEAU 1: Le résultat de l'analyse des marqueurs d'arguments

Cela nous a permis de confirmer nos hypothèses énoncées dans l'introduction :

1. Le statut morphologique des marqueurs d'arguments doit être séparé de leur statut morphosyntaxique, les affixes n'étant pas forcément des marqueurs d'accord.
2. On ne peut traiter les marqueurs d'arguments de manière uniforme. Les marqueurs de sujet diffèrent des marqueurs d'objet sur plusieurs plans et doivent être analysés en prenant en compte ces différences. Les marqueurs de sujet sont des affixes flexionnels alors que les marqueurs d'objet sont des clitiques.
3. On ne peut traiter les marqueurs d'objet comme des affixes. À l'encontre de la majorité des analyses des clitiques dans le cadre de HPSG, qui tendent à dissoudre ces unités dans l'étiquette affixale, nous pensons qu'il y a lieu de postuler l'existence d'une catégorie autonome, celle des clitiques, différente de celle des affixes.
4. Les marqueurs de sujet sont des affixes flexionnels faisant partie des entrées lexicales des verbes qui en sont les hôtes.

Cette analyse des marqueurs d'arguments a été modélisée dans une théorie linguistique basée sur les contraintes, la grammaire syntagmatique endocentrique (HPSG). Cette théorie, dont nous

avons présenté les principes de base et des exemples d'application dans le premier chapitre, s'est avérée un cadre formel et théorique adéquat.

Le défi que présentait cette modélisation était essentiellement le fait que cette grammaire avait une longue tradition en ce qui a trait aux langues germaniques et romanes, mais peu a été fait dans les langues sémitiques. Toutefois, la flexibilité de ce cadre est telle que nous n'avons pas eu besoin de modifier la théorie pour rendre compte des données. Une seule opération, l'unification, et une seule structure de données, la structure de traits, étaient suffisantes pour modéliser le comportement des marqueurs d'arguments. Il va de soi que nous avons effectué certains choix théoriques, comme le recours aux règles lexicales, qui ne font pas l'unanimité, mais nous pensons que ces choix sont amplement motivés par la nature de la langue étudiée. Le lexique de l'arabe ne peut en fait être convenablement modélisé sans rendre compte de toutes les régularités (morphologiques essentiellement) qui le caractérisent. La redondance lexicale est donc éliminée grâce à l'adoption du dispositif des règles de redondance, ou règles lexicales.

Cette modélisation, unique en son genre en ce qui a trait à la langue étudiée, prouve que le modèle est assez puissant pour rendre compte de phénomènes langagiers issus de familles de langues aussi différentes que les langues sémitiques et les langues romanes grâce surtout à la flexibilité de l'approche lexicaliste. Cette dernière localise une grande partie des informations linguistiques pertinentes dans les entrées lexicales et dans la hiérarchie des types. Le nombre de règles strictement syntaxiques est ainsi réduit grâce à la richesse du lexique et grâce à la puissance du dispositif des structures de traits typées.

Le cinquième chapitre de notre thèse a été consacré à cette modélisation et à une implémentation informatique qui constituait un test permettant de mesurer l'efficacité de l'analyse et sa faisabilité dans un système de performance. Pour ce faire, nous avons écrit une grammaire de l'arabe et l'avons implémentée dans l'environnement de développement LKB. Cette implémentation a prouvé la faisabilité du modèle que nous avons proposé et sa conformité aux principes des grammaires à base de contraintes. Notre grammaire est ainsi capable d'analyser, de rejeter et de générer des phrases en partant d'un lexique réduit, de quelques règles syntaxiques modélisées sous forme de contraintes sur des types syntagmatiques et de quelques règles lexicales (de nature à la fois morphophonologique et morphosyntaxique).

Notre recherche sur les marqueurs d'arguments de l'arabe nous a donné l'occasion de toucher à plusieurs aspects de la grammaire de cette langue. Du fait de leurs propriétés linguistiques en tant

qu'éléments d'interface, ces marqueurs, et surtout les marqueurs de sujet, constituent un champ de recherches fertile et presque inépuisable. La description, l'analyse, la modélisation, ainsi que l'implémentation constituent, nous le croyons fermement, des étapes nécessaires pour avoir une compréhension adéquate de ces éléments. La présence de ces quatre étapes essentielles au sein d'un même travail constitue en elle-même une contribution non seulement à l'étude linguistique de ces unités, mais également à la théorie HPSG et aux grammaires à base de contraintes en général.

Cette recherche est également unique en son genre eu égard à deux aspects : le sujet étudié et le cadre théorique. En ce qui a trait au sujet étudié, aucun travail précédent n'a été entièrement consacré aux marqueurs d'arguments de l'arabe. Les travaux cités dans cette thèse portent généralement sur d'autres phénomènes et touchent à une partie ou à une autre de l'image complexe que présentent ces marqueurs. Ainsi, le fait de consacrer un travail de recherche de longue haleine à ces unités permet une meilleure compréhension de plusieurs phénomènes, dont, notamment, les asymétries de l'accord et certains aspects de la question de la variation dialectale arabe. Cela s'ajoute évidemment à l'étude pure et simple de ces marqueurs. Notons que cette étude apporte un grand lot de richesse grâce à la diversité des sources consultées et donc grâce à la richesse du domaine empirique que nous avons ainsi couvert par nos exemples.

Quant au cadre théorique, les recherches sur les phénomènes impliquant les langues sémitiques dans le cadre de la théorie HPSG sont plutôt peu nombreuses. Nous pensons donc que notre contribution est précieuse pour toute recherche future portant sur ces langues et s'inscrivant dans le cadre de cette théorie. De plus, ayant mené notre recherche sur les marqueurs d'arguments, qui constituent des éléments d'interface, nous avons eu l'occasion d'aborder plusieurs questions d'ordres théorique et formel qui peuvent servir de base à un travail de recherche beaucoup plus global. Une grammaire HPSG de l'arabe est donc possible en partant d'une grammaire HPSG des marqueurs d'arguments comme celle que nous proposons.

Pour terminer cette conclusion, nous rappelons les limites de notre travail ainsi que les perspectives que nous envisageons.

Notons, tout d'abord, que ce travail que nous voulions aussi collé que possible aux données empiriques pertinentes ne peut rendre compte de tous les contextes dans lesquels on emploie les marqueurs d'arguments. L'une des raisons de ce constat c'est que nous avons volontairement éliminé les contextes où les marqueurs de sujet sont attachés à des verbes dits défectifs, analysés par

certains comme des auxiliaires. Ces verbes, qui englobent *ka:na et ses soeurs* dans la terminologie traditionnelle, mais également les verbes inchoatifs, se déclinent comme les verbes non défectifs tout en affectant le sujet et le prédicat de la phrase équative ou de la phrase avec un SN disloqué à gauche. Ils assignent ainsi le cas nominatif au premier élément et le cas accusatif au deuxième (si ce dernier est nominal). Ils peuvent également être employés sans le premier SN, auquel cas le marqueur de sujet occupe cette position. En outre, ces verbes présentent les mêmes asymétries de l'accord que tout autre verbe de l'arabe (accord riche vs accord partiel). Le débat qui les entoure consiste à se demander si nous pouvons les traiter comme auxiliaires ou non. Nous avons évité ce débat parce qu'il ne fait pas partie de notre sujet d'étude et parce qu'il n'affecte en rien notre analyse des marqueurs de sujet. Ces derniers peuvent ainsi être analysés, globalement, dans la même ligne d'idées de ce que nous avons proposé dans notre thèse. Les marqueurs de sujet qui s'attachent aux verbes *défectifs* sont ainsi les mêmes affixes qui s'attachent aux verbes *normaux* et ils ont les mêmes statuts morphosyntaxiques : celui d'arguments, comme en (1a) et celui de marqueurs d'accord, comme en (1b).

(1) a. *bana:t* *-u* *-ka* *kun* **-na** *j-* *alʕab* *-na*
filles -NOM -tes être -3FS 3- jouer.IMP -FS
« Tes filles, elles étaient en train de jouer »

b. *ka:n* **-at** *bana:t* *-u* *-ka* *j-* *alʕab* *-na*
être -3F filles -NOM -tes 3- jouer.IMP -FS
« Tes filles étaient en train de jouer »

Le problème réside en fait dans l'expression du temps. Le verbe principal est à l'imperfectif et ce mode a généralement une valeur de présent ou de futur. L'emploi de *ka:na* force une interprétation au passé. L'expression du mode, de l'aspect et du temps semble donc s'amalgamer de nouveau aux marqueurs de sujet. Rappelons que cela s'observe avec l'imperfectif des autres verbes, puisque certains des marqueurs de sujet qui s'attachent aux verbes à cet aspect se trouvent concaténés à des marques d'aspect/mode et posent plus d'un problème quand nous voulons les isoler. Et si le problème était la morphologie de l'imperfectif justement ? Nous nous sommes aventuré sur ce sentier dans notre thèse et nous avons fait des choix motivés, mais l'expression de l'aspect et du temps en arabe mériterait beaucoup plus d'investigations et pourrait constituer un projet de recherche à part entière. Cette étude ne pourrait passer sous silence le rôle joué par les marqueurs de sujet et nous sommes convaincu que notre analyse constituerait un bon point de départ lors de recherches futures sur ce sujet.

Annexes

EXEMPLES CITÉS

Les exemples arabes cités dans la thèse sont regroupés ici par chapitre.

Chapitre II

(1) a. ***huwa*** *ʕalijj* *-un* *ʔibn* *-u* *muħammad* *-in*
lui Ali -NOM fils -NOM Mohammed -GÉN
« Il est Ali fils de Mohammed »

b. ***ʔanta*** *ʔakal* *-ta* *ʃaba:b* *-a* *-ka*
toi manger.PER -2MS jeunesse -ACC -ta
« Toi, tu as dilapidé ta jeunesse »

c. *ʔakal* *-ta* *ʃaba:b* *-a* *-ka*
manger.PER -2MS jeunesse -ACC -ta
« Tu as dilapidé ta jeunesse »

(2) a. *hal* *raʔaj* *-ta* ***-ha:*** *ʔ*
est-ce-que voir.PER -2MS -la
« L'as-tu vue ? »

b. *xif* *-tu* *min* ***-ha:***
avoir.peur.PER -1S de -elle
« J'ai eu peur d'elle »

(3) *hal* *raʔaj* ***-ta*** *-ha:* *ʔ*
est-ce-que voir.PER -2MS -la
« L'as-tu vue ? »

(4) *hum*$_i$ *daxal* *-u:*$_i$
eux entrer.PER -3MP
« Eux, ils sont entrés »

(5) a. *huwa*$_i$ *j*$_i$*-* *uħibb* *-u* *ha:*dh*a:*
lui 3MS- aimer.IMP -ind cei
« Lui, il aime ceci »

b. *hija*$_i$ *t*$_i$*-* *uħibb* *-u* *ha:*dh*a:*
elle 3FS- aimer.IMP -ind ceci
« Elle, elle aime ceci »

(6) *lam* *t-* *axruʁ* *-u:* → *ʔuxruʁ* *-u:!*
ne.pas 2 sortir.IMP 2MP sortir.IMPÉ -2MP

(7) a. *j-* *aktub* *-u* *sa:mi:* *risa:lat* *-a* *-n*
3M- écrire.IMP -Ind Sami lettre -ACC -Indé
« Sami écrit une lettre »

b. *t-* *aktub* *-u* *l-binta:ni* *risa:lat* *-a* *-n*
3F- écrire.IMP -Ind les-deux.filles lettre -ACC -Indé
« Les deux filles écrivent une lettre »

c. *t-* *aktub* *-u* *n-nis:ʔ-u* *risa:lat* *-a* *-n*
3F- écrire.IMP -Ind les-femmes-NOM lettre -ACC -Indé
« Les femmes écrivent une lettre »

d. *j-* *aktub* *-u* *l-ʔawla:d-u* *risa:lat* *-a* *-n*
3M- écrire.IMP -Ind les-garçons-NOM lettre -ACC -Indé
« Les garçons écrivent une lettre »

(8) a. *katab* *-a* *sa:mi:* *risa:lat* *-a* *-n*
écrire.PER -3M Sami lettre -ACC -Indé
« Sami a écrit une lettre »

b. *katab* *-at* *l-binta:ni* *risa:lat* *-a* *-n*
écrire.PER -3F les-deux.filles lettre -ACC -Indé
« Les deux filles ont écrit une lettre »

c. *katab* *-at* *n-nis:ʔ-u* *risa:lat* *-a* *-n*
écrire.PER -3F les-femmes-NOM lettre -ACC -Indé
« Les femmes écrivent une lettre »

d. *katab* *-a* *l-ʔawla:d-u* *risa:lat* *-a* *-n*
écrire.PER -3M les-garçons-NOM lettre -ACC -Indé
« Les garçons ont écrit une lettre »

(9) a. *huwa*$_i$ *katab* *-a*$_i$ *risa:lat* *-a* *-n*
lui écrire.PER -3MS lettre -ACC -Indé
« Lui, il a écrit une lettre »

b. *huwa*$_i$ *raʔaj* *-tu* *-hu*$_i$ *ʔamsi*
lui voir.PER -1S -le hier
« Lui, je l'ai vu hier »

c. *katab* *-a*$_i$ *huwa*$_i$ *risa:lat* *-a* *-n*
écrire.PER -3MS lui lettre -ACC -Indé
« Il a écrit une lettre, lui »

d. *raʔaj* *-tu* *-hu*$_i$ *huwa*$_i$ *ʔamsi*
voir.PER -1S -le lui hier
« Je l'ai vu hier, lui »

(10) ***hunna*** *ʔaxawa:t* *-u* *l-* *munðir* *-i*
elles sœurs -NOM le- Mundhir -GÉN
« Elles sont les sœurs d'El Mundhir »

(11) a. *ʔana:* ***huwa*** *l-* *fa:ʕil* *-u* (1re personne)
moi lui le- faiseur -NOM
« C'est moi le malfaiteur »

b. *ʔanta* ***huwa*** *l-* *fa:ʕil* *-u* (2e personne)
toi lui le- faiseur -NOM
« C'est toi le malfaiteur »

c. *zajd* *-un* ***huwa*** *l-* *fa:ʕil* *-u* (3e personne)
Zayd -NOM lui le- faiseur -NOM
« C'est Zayd le malfaiteur »

(12) a. ***huwa*** *l-* *kawn* *-u* *ħajj* *-u* *-n*
lui l'- univers -NOM vivant -NOM -Indé
« Le fait est que l'univers est un vivant »

b. ***hija*** *l-* *ʔumm* *-u* *madrasat* *-u* *-n*
elle la- mère -NOM école -NOM -Indé
« Le fait est que la mère est une [véritable] école »

(13) a. *qaraʔ* *-tu* *r-* *risaʔlat* *-a*
lire.PER -1S la- lettre -ACC
« J'ai lu la lettre »

b. *ʔar-* *risaʔlat* *-a* *qaraʔ* *-tu*
la- lettre -ACC lire.PER -1S
« C'est la lettre que j'ai lue (et non le roman) »

c. ***ʔijja:ha:*** *qaraʔ* *-tu*
elle(ACC) lire.PER -1S
« C'est elle que j'ai lue (et non une autre) »

d. **qaraʔ* *-tu* ***ʔijja:ha:***
lire.PER -1S elle(ACC)

(14) *qaraʔ* *-tu* ***-ha:***
lire.PER -1S -la
« Je l'ai lue »

(15) a. *wahab* *-tu* *-ka* ***-ha:***
donner.PER -1S -te -la
« Je te l'ai donnée »

b. *wahab* *-tu* *-ka* *ʔijja:ha:*
donner.PER -1S -te elle(ACC)
« Je te l'ai donnée »

(16) a. **raʔj* *-tu* *-hu$_i$* *s'adi:q* *-i:$_i$* (marqueur d'objet)
voir.PER -1S -le ami -mon

a'. *raʔj* *-tu* *-hu*
voir.PER -1S -le
« Je l'ai vu »

a". *raʔj* *-tu* *s'adi:q* *-i:*
voir.PER -1S ami -mon
« J'ai vu mon ami »

b. **naʤaħ* *-u:$_i$* *ʔas'diqa:ʔ* *-i:$_i$* (marqueur de sujet)
réussir.PER -3MP amis -mes

b'. *naʤaħ* *-u:*
réussir.PER -3MP
« Ils ont réussi »

b". *naʤaħ* *-a$_i$* *ʔas'diqa:ʔ* *-i:$_i$*
réussir.PER -3M amis -mes
« Mes amis ont réussi »

(17) a. *raʔaj* *-tu* *-hu$_i$* *huwa$_i$* (marqueur d'objet)
voir.PER -1S -le lui(NOM)
« Je l'ai vue, lui »

b. *raʤaʕ* *-a$_i$* *huwa$_i$* (marqueur de sujet)
revenir.PER -3MS lui(NOM)
« Il est revenu, lui »

(18) a. **raʔaj* *-tu* *-hu* *wa* *-ha:* (marqueur d'objet)
voir.PER -1S -le et -la

b. *raʔaj* *-tu* *-hu* *wa* *raʔaj* *-tu* *-ha:*
voir.PER -1S -le et voir.PER -1S -la
« Je l'ai vu et je l'ai vue »

c. **ʤiʔ* *-tu* *wa* *-ta* (marqueur de sujet)
venir.PER -1S et -2MS

d. *ʤiʔ* *-tu* *wa* *ʤiʔ* *-ta*
venir.PER -1S et venir.PER -2MS
« Je suis venu et toi aussi »

(19) a. **raʔaj -tu -**HU*** (marqueur d'objet focalisé)
voir.PER -1S -LE

b. *ʔijja:**hu** raʔaj -tu* (Pro.Ind.ACC focalisé)
lui(ACC) voir.PER -1S
« C'est lui que j'ai vu »

c. **nim -**TU*** (marqueur de sujet focalisé)
dormir.PER -1S

d. *ʔan:**a** nim -tu* (Pro.Ind.NOM focalisé)
moi dormir.PER -1S
« C'est moi qui me suis endormi »

(20) a. *katab -ta -**ha:***
écrire.PER -2MS -la
« Tu l'as écrite »

b. *kita:b -u -**ha:***
livre -NOM -son
« son livre »

c. *xif -tu min -**ha:***
craindre.PER -1S de -elle
« J'ai eu peur d'elle »

d. *ʔinna -**ha:** qari:bat -u -n*
que -elle proche -NOM -Indé
« Elle est proche, certes»

(21) a. *rakʒaʕ -**u:***
revenir.PER -3MP
« Ils sont revenus »

b. *j- arkʒiʕ -**u:** -na*
3- revenir.IMP -MP -Ind
« Ils reviennent »

c. *ʔirkʒiʕ -**u:***
revenir.IMPÉ -2MP
« Revenez ! »

(22) a. *katab -**tu***
écrire.PER -1S
« J'ai écrit »

b. *katab -**tu** -**ha:***
écrire.PER -1S -la
« Je l'ai écrite »

c. **katab* *-**ha:***
écrire.PER -la

d. **katab* *-**ha:*** *-**tu***
écrire.PER -la -1S

e. ***t-*** *aktub* *-**a:*** *-**ni*** *-**ha:***
2- écrire.IMP -MD -Ind -la
« Vous (deux) l'écrivez »

(23) a. *ʔal-* *muʕallim**u:na*** *j-* *uɣa:dir* *-**u:*** *-**na*** *l-* *qa:ʕat* *-a*
les- enseignants 3- quitter.IMP -MP-Ind la- salle -ACC
« Les enseignants, ils quittent la salle »

b. *ʔal-* *muʕallim**a:ni*** *j-* *uɣa:dir* *-**a:-ni*** *l-* *qa:ʕat* *-a*
les- deux.enseignants 3- quitter.IMP -MD-Ind la- salle -ACC
« Les deux enseignants, ils quittent la salle »

(24) a. *qa:bal* *-tu* *-**hu*** *nafs -a -hu* (marqueur d'objet ACC)
rencontrer.PER -1S -le âme-ACC-son
« Je l'ai rencontré, lui-même »

b. *farar* *-tu* *min* *-**hu*** *nafs -i -hi* (marqueur d'objet GÉN)
fuir.PER -1S de -lui âme-GÉN-son
« Je l'ai fui, lui-même »

c. **ʤa:ʔ* *-**u:*** *ʔanfus -u -hum* (marqueur de sujet)
venir.PER -3MP âmes-NOM-leurs

(25) *ʤa:ʔ* *-**u:*** ***hum**$_i$* *ʔanfus -u -hum*$_i$
venir.PER -3MP eux âmes-NOM-leurs
« Ils sont venus, eux-mêmes »

(26) a. *bajt* *-u* *-**ha:***
maison -NOM -sa
« sa maison »

a'. **bajt* *-u* *ʔ**ijja:ha:***
maison -NOM elle

b. *min* *-**ha:***
de -elle
« d'elle »

b'. **min* *ʔ**ijja:ha:***
de elle

c. *ʔanna* ***-ha:***
que -elle
« qu'elle »

c'. ***ʔanna*** ***ʔijja:ha:***
que elle

d. *laqi:* *-tu* ***-ha:***
trouver.PER -1S -la
« Je l'ai trouvée »

d' ***laqi:*** *-tu* ***ʔijja:ha:***
trouver.PER -1S elle

(27) a. *xil* *-tu* *r-* *raʤul* *-a* *lis's'* *-a* *-n*
croire.PER -1S l'- homme -ACC voleur -ACC -Indé
« J'ai confondu l'homme avec un voleur »

b. *xil* *-tu* *-hu* *lis's'* *-a* *-n*
croire.PER -1S -le voleur -ACC -Indé
« Je l'ai confondu avec un voleur »

b. ***xil*** *-tu* *-hu* *-hu*
croire.PER -1S -le -le

(28) a. *wahab* *-tu* *-ka* ***l- kita:b -a***
donner.PER -1S -te le-livre-ACC
« je t'ai donné le livre »

b. *wahab* *-tu* *-ka* ***-hu***
donner.PER -1S -te -le
« Je te l'ai donné »

c. *wahab* *-tu* *-ka* ***ʔijja:hu***
donner.PER -1S -te lui(ACC)
« Je te l'ai donné »

(29) a. *bajt* *-u* ***zajd -in***
maison -NOM Zayd-GÉN
« la maison de Zayd »

a'. *bajt* *-u* ***-hu***
maison -NOM -sa
« sa maison »

b. *ʕinda* ***zajd -in***
chez Zayd-GÉN
« chez Zayd »

b'. *ʔinda* ***-hu***
chez -lui
« chez lui »

c. *ʔanna* ***zajd -an***
que Zayd-ACC
« que Zayd »

c'. *ʔanna* ***-hu***
que -il
« qu'il »

d. *laqi:* *-tu* ***zajd -an***
trouver.PER -1S Zayd-ACC
« J'ai trouvé Zayd »

d'. *laqi:* *-tu* ***-hu***
trouver.PER -1S -le
« Je l'ai trouvé »

e. ***zajd -an*** *laqi:* *-tu*
Zayd-ACC trouver.PER -1S
« C'est Zayd que j'ai trouvé (et non Michel) »

e'. ***ʔijja:hu*** *laqi:* *-tu*
lui(ACC) trouver.PER -1S
« C'est lui que j'ai trouvé (et non elle) »

(30) a. **wahab* *-tu* ***-hu*** ***-ka***
donner.PER -1S -lui -te

b. *wahab* *-tu* ***-hu*** ***ʔijja:ka***
donner.PER -1S -lui toi
« Je t'ai donné à lui »

(31) a. *xil* *-tu* ***-ka*** ***lis's'-a-n***
croire.PER -1S -te voleur-ACC-Indé
« J'ai cru que tu étais un voleur "

b. **xil* *-tu* ***-ka*** ***-hu***
croire.PER -1S -te -le

c. *xil* *-tu* ***-ka*** ***ʔijja:hu***
croire.PER -1S -te lui
« J'ai cru que tu l'étais »

(32) *wi:n* *-hum* *l-wla:ad*
où -eux les-garçons
« Où sont les garçons ? »

(33) a. *fhim* *-t* *-ha:*$_i$ *la-* *l-* *mʕalme*$_i$
comprendre.PER -1S -la à- l'- enseignante
« J'ai compris l'enseignante »

b. **fhim* *-t* *-ha:*$_i$ *l-* *mʕalme*$_i$
comprendre.PER -1S -la l'- enseignante

(34) *ʔar-* *risa:lat* *-a* *qaraʔ* *-tu*
la- lettre -ACC lire.PER -1S
? « La lettre, j'ai lu »

(35) a. *katab* *-tu* *-ha:*
écrire.PER -1S -la
« Je l'ai écrite »

b. *bajt* *-u* *-ha:*
maison -NOM -sa
« sa maison »

c. *xif* *-tu* *min* *-ha:*
craindre.PER -1S de -elle
« J'ai eu peur d'elle »

d. *ka:tib* *-u* *l-* *qis's'at* *-i* *sa:rid* *-u* *-ha:*
écrivain -NOM le- roman -GÉN narrateur -NOM -son
« L'écrivain du roman est son narrateur »

e. *ʔinna* *-hu* *ʤa:ʔ* *-a* *-na:* *raʤul* *-u* *-n*
que -il venir.PER -3M -nous homme -NOM -Indé
« Le fait est qu'un homme est venu vers nous »

f. *ʔinna* *-ka* *t-* *uri:d* *-u* *l-* *ʕira:k* *-a*
que -toi 2MS- vouloir.IMP -Ind la- zizanie -ACC
« Tu veux la zizanie, certes »

(36) a. **katab* *-tu* *-ha:* *r-risa:lat-a* / *ʔijja:aha:*
écrire.PER -1S -la la-lettre-ACC / elle(ACC)

b. **raʔaj* *-tu* *-ha:* *raba:b -a* / *ʔijja:aha:*
voir.PER -1S -la Rabab-ACC / elle(ACC)

(37) a. **katab* *-tu* *-ha:* *risa:lat*$_i$ *-a* *-n*
écrire.PER -1S -la lettre -ACC -Indé
(Pour signifier : « J'ai écrit une lettre »)

b. **katab* *-tu* *-ha:*$_i$ *risa:lat*$_i$ *-a* *ʃukr* *-i* *-n*
écrire.PER 1S -la lettre -ACC remerciement -GÉN -Indé
(Pour signifier : « J'ai écrit une lettre de remerciement »)

(38) a. *ʔar* *-risa:lat*$_i$ *-u* *katab* *-tu* *-ha:*$_i$
la- lettre -NOM écrire.PER -1S -la
« La lettre, je l'ai écrite »

b. *hija*$_i$ *katab* *-tu* *-ha:*$_i$
elle(NOM) écrire.PER -1S -la
« Elle, je l'ai écrite »

c. *katab* *-tu* *-ha:*$_i$ *r-* *risa:lat*$_i$ *-u*
écrire.PER -1S -la la- lettre -NOM
« Je l'ai écrite, la lettre »

d. *katab* *-tu* *-ha:*$_i$ *hija*$_i$
écrire.PER -1S -la elle(NOM)
« Je l'ai écrite, elle»

(39) a. **ha:ða:* *bajt* *-u* *-ha:*$_i$ *raba:b -a*$_i$
celle-ci maison -NOM -sa Rabab-GÉN

b. *ha:ða:* *bajt* *-u* *-ha:*
celle-ci maison -NOM -sa
« Celle-ci est sa maison »

c. *ha:ða:* *bajt* *-u* *raba:b* *-a*
celle-ci maison Rabab -GÉN
« Celle-ci est la maison de Rabab »

(40) a. *raba:b -u*$_i$ *ʔaxawa:t* *-u* *-ha:*$_i$ *ʕaʃr* *-u* *-n*
Rabab-NOM sœurs -NOM -ses dix -NOM -Indé
« Rabab, ses sœurs sont au nombre de dix »

b. *hija*$_i$ *ʔaxawa:t* *-u* *-ha:*$_i$ *ʕaʃr* *-u* *-n*
elle sœurs -NOM -ses dix -NOM -Indé
« Elle, ses sœurs sont au nombre de dix »

(41) a. *ʔaxawa:t* *-u* *-ha:*$_i$ *hija*$_i$ *ʕaʃr* *-u* *-n*
sœurs -NOM -sa elle(NOM) dix -NOM -Indé
« Ses sœurs à elle sont au nombre de dix »

b. **ʔaxawa:t* *-u* *-ha:*$_i$ *raba:b -a*$_i$ *ʕaʃr* *-u* *-n*
sœurs -NOM -ses Rabab-GÉN dix -NOM -Indé

(42) a. *ʔinna* *-hu* *s'adiiq* *-i:*
que -lui ami -mon

« Il est mon ami, certes »

b. *ʔinna* *-hu* *j-* *uħibb* *-u* *l-* *muzaʔħ* *-a*
que -lui 3MS- aimer.IMP -Ind la- plaisanterie -ACC
« Lui, il aime plaisanter, certes »

c. *ʔinna* *-ha:* *ʕa:d* *-at* *l-* *fata:ta:ni*
que -elle revenir.PER -3F les- deux.filles.NOM
« Le fait est que les deux filles sont revenues »

(43) *ʔinna* *zajd* *-an* *s'adi:q* *-i:*
que Zayd -ACC ami -mon
« Zayd est mon ami, certes »

(44) *ʔinna* *zajd* *-an* *j-* *uħibb* *-u* *l-* *muza:ħ* *-a*
que Zayd -ACC 3MS- aimer.IMP -Ind la- plaisanterie -ACC
« Zayd, il aime plaisanter, certes »

(45) a. *ʔinna* *-hu* *ʕa:d* *-a* *l-* *fitjat* *-u*
que -il revenir.PER -3M les- jeunes.garçons -NOM
« Le fait est que les jeunes garçons sont revenus »

b. *ʔinna* *-ha:* *ʕu:d* *-at* *n-* *nisa:ʔ* *-u*
que -elle revenir.PER -3F les- femmes -NOM
« Le fait est que les femmes sont revenues »

(46) a. *zajd -un* *ʔinna* *-hu* *s'adi:q* *-i:*
Zayd-NOM que -lui ami -mon
« Zayd, il est mon ami, certes »

b. *ʔax-i:* *ʔinna* *-hu* *j-* *uħibb* *-u* *l-* *muza:ħ* *-a*
frère-mon que -lui 3MS- aimer.IMP -Ind la- plaisanterie -ACC
« Mon frère, lui, il a aime plaisanter »

(47) a. *katab* *-a* *l-* *walad* *-u* *risa:lat* *-a* *-n* (redoublement)
écrire.PER -3M le- garçon -NOM lettre -ACC -Indé
« Le garçon a écrit une lettre »

b. *katab* *-a* *huwa* *risa:lat* *-a* *-n* (dislocation)
écrire.PER -3MS lui lettre -ACC -Indé
« Il a écrit une lettre, lui »

(48) *saʔal* *-u:* *-ni:* *s'-* *s'afħ* *-a* *fa-* *s'afaħ* *-tu*
demander.PER -3MP -me le- pardon -ACC et- pardonner.PER -1S
« Ils m'ont demandé pardon et j'ai pardonné »

(49) a. *raba:b -u* *ʔaxawa:t* *-u* ***-ha:*** *ʕaʃr* *-u* *-n*
Raba-NOM sœurs -NOM -ses dix -NOM -Indé
« Rabab, ses sœurs sont au nombre de dix »

b. *hija* *ʔaxawa:t* *-u* ***-ha:*** *ʕaʃr* *-u* *-n*
elle sœurs -NOM -ses dix -NOM -Indé
« Elle, ses sœurs sont au nombre de dix »

(50) a. *ʔinna* *-ha:* *lajlat* *-u* *-n* *t'awi:lat* *-u* *-n*
que -elle nuit -NOM -Indé longue -NOM -Indé
« C'est une longue nuit, certes »

b. **t'awi:lat* *-u* *-n* *ʔinna* *-ha:* *lajlat* *-u* *-n*
longue -NOM -Indé que -elle nuit -NOM -Indé

(51) a. *za:r* ***-a*** *-na:* *l-wa:li:*
visiter.PER -3M -nous le-gouverneur
« Le gouvernenur nous a rendu visite »

b. *za:r* ***-a*** *-na:* *l-wa:lija:ni*
visiter.PER -3M -nous les-deux.gouverneurs
« Les deux gouverneurs nous ont rendu visite »

c. *za:r* ***-a*** *-na:* *l-wula:la:t -u*
visiter.PER -3M -nous les-gouverneurs-NOM
« Les gouverneurs nous ont rendu visite »

(52) a. *man*$_i$ *za:r* ***-a***$_i$ *-kum*
qui visiter.PER -3MS -vous
« Qui vous a rendu visite ? »

b. *man*$_i$ *za:r* ***-a:***$_i$ *-kum*
qui visiter.PER -3MD -vous
« Qui sont les deux personnes qui vous ont rendu visite ? »

c. *man*$_i$ *za:r* ***-u:***$_i$ *-kum*
qui visiter.PER -3MP -vous
« Qui sont les personnes qui vous ont rendu visite ? »

(53) a. *ʔal- wa:lija:ni* *za:r* *-a:* *-na:*
les- deux.gouverneurs.NOM visiter.PER -3MD -nous
« Les deux gouverneurs, ils nous ont rendu visite »

b. *huma:* *za:r* *-a:* *-na:*
eux.deux visiter.PER -3MD -nous
« Eux deux, ils nous ont rendu visite »

(54) a. *za:r* *-a:* *-na:* *l- wa:lija:ni*
visiter.PER -3MD -nous les- deux.gouvernenurs.NOM
« Ils nous ont rendu visite, les deux gouvernenurs »

b. *za:r* *-a:* *-na:* *huma:*
visiter.PER -3MD -nous eux.deux
« Ils nous ont rendu visite, eux deux »

(55) a. *ʤa:ʔ* *-a* *raʤul* *-u* *-n* (sujet indéfini)
arriver.PER -3M homme -NOM -Indé
« Un homme est arrivé »

b. *ʤa:ʔ* *-a* *r-* *raʤul* *-u* (sujet défini)
arriver.PER -3M l'- homme -NOM
« L'homme est arrivé »

(56) a. *ʔar-* *rajul*$_i$ *-u* *ʤa:ʔ* *-a*$_i$ (topique défini)
l'- homme -NOM venir.PER -3MS
« L'homme, il est arrivé »

b. **raʤul* *-u* *-n* *ʤa:ʔ* *-a*[42] (topique indéfini)
homme -NOM -indé venir.PER -3MS

c. *ʔar-* *raʤul*$_i$ *-u* *raʔj* *-tu* *-hu*$_i$ (topique objet défini)
l'- homme -NOM voir.PER -1S -le
« L'homme, je l'ai vu »

(57) a. *ʔal-* *falla:ħ* *-u:na* *ħalab* *-u:* *baqara:t* *-i* *-him*
les- agriculteurs -NOM traire.PER -3MP vaches -ACC -leurs
« Les agriculteurs, ils ont trait leurs vaches »

b. *ʔal-* *falla:ħ* *-u:na* *ħulib* *-at* *baqara:t* *-u* *-hum*
les- agriculteurs -NOM traire.PASSIF.PER -3F vaches -NOM -leurs
« Les agriculteurs, leurs vaches ont été traites »

(58) a. *ħalab* *-a* *l-* *falla:ħ* *-u:na* *l-* *baqara:t* *-i*
traire.PER -3M les- agriculteurs -NOM les- vaches -ACC
« Les agriculteurs ont trait les vaches »

b. *ħulib* *-at* *l-* *baqara:t* *-u* *min* *qibal* *-i* *l-*
traire.PASSIF.PER -3F les- vaches -NOM de la.part -GÉN les-
falla:ħ *-i:na*
agriculteurs -GÉN
« Les vaches ont été traites par les argriculteurs »

(59) *j-* *askun* *-u* *l-* *ʕira:q* *-a* *ʔumam* *-u* *-n* *muxtalifat* *-u*
3M- habiter.IMP -Ind l'- Irqa -ACC peuples -NOM -Indé différents -NOM
-n
-Indé
« Différents peuples habitent l'Iraq »

(60) a. *ðahab* *-a* *sa:ʕat -u -n* *mina* *l-* *lajl* *-i*
partir.PER -3M heure -NOM -Indé de la- nuit -GÉN
« Une fraction de la nuit s'écoula »

b. *qut'ʕ* *-a* *ʔajdi: -him*
trancher.PASSIF -3M mains -leurs
« Leurs mains furent tranchées »

c. *baʕaθ* *-a* *qurajʃ -un* ...
envoyer.PER -3M Quraych -NOM
« Quraych envoya ...»

(61) a. *ðahab* ***-at*** *sa:ʕat -u -n* *mina* *l-* *lajl* *-i*
partir.PER -3F heure -NOM -Indé de la- nuit -GÉN
« Une fraction de la nuit s'écoula »

b. *qut'ʕ* ***-at*** *ʔajdi: -him*
trancher.PASSIF -3F mains -leurs
« Leurs mains furent tranchées »

c. *baʕaθ* ***-at*** *qurajʃ -un* ...
envoyer.PER -3F Quraych -NOM
« Quraych envoya ... »

(62) a. *ʔaʃʤa:r* *-u* *-n* *ʕiʤa:m* *-u* *-n*
arbres -NOM -Indé grands -NOM -Indé
« de grands arbres »

b. *ʔaʃʤa:r* *-u* *-n* *ʕaʤi:mat* *-u* *-n*
arbres -NOM -Indé grande -NOM -Indé
« de grands arbres »

(63) a. *raʤul* *-u* *-n* *qa:dim -u -n* *ʔaxu: -hu*
homme -NOM -Indé arrivant -NOM -Indé frère -son
« un homme dont le père est en train d'arriver »

b. *raʤul* *-u* *-n* *qa:dim -u -n* *ʔaxawa: -hu*
homme -NOM -Indé arrivant -NOM.D -Indé frères.NOM.D -ses
« Un homme dont les deux frères sont en train d'arriver»

c. **raʤul* *-u* *-n* *qa:dim -a:ni* *ʔaxawa: -hu*
homme -NOM -Indé arrivants -NOM.D frères.NOM.D -ses

d. *raʤul* *-u* *-n* *qa:dim -u -n* ***ʔixwat -u -hu***
homme -NOM -Indé arrivant.S -NOM -Indé frères.P -NOM -ses
« un homme dont les frères sont en train d'arriver »

e. **raʤul* *-u* *-n* *qa:dim -u:na* *ʔixwat -u -hu*
homme -NOM -Indé arrivant.P -NOM-Indé frères.P -NOM -ses

(64) a. *raʤul* *-u* *-n* ***ʔaxu: -hu*** *qa:dim -u -n*
homme -NOM -Indé frère -son arrivant -NOM -Indé
« un homme dont le frère est en train d'arriver »

b. **raʤul* *-u* *-n* *ʔaxawa: -hu* *qa:dim -u -n*
homme -NOM -Indé frères.NOM.D -ses arrivant -NOM.D -Indé

c. *raʤul* *-u* *-n* *ʔaxawa: -hu* *qa:dim -a:ni*
homme -NOM -Indé frères.NOM.D -ses arrivants -NOM.D
« Un homme dont les deux frères sont en train d'arriver»

d. **raʤul* *-u* *-n* *ʔixwat -u -hu* *qa:dim -u -n*
homme -NOM -Indé frères.P -NOM -ses arrivant.S -NOM -Indé

e. *raʤul* *-u* *-n* *ʔixwat -u -hu* *qa:dim -u:na*
homme -NOM -Indé frères.P -NOM -ses arrivant.P -NOM-Indé
« un homme dont les frères sont en train d'arriver »

(65) a. *raʤul* *-u* *-n* *kari:mat -u -n* *ʔuxta: -hu*
homme -NOM -Indé généreuse.S -NOM -Indé sœurs.NOM.D -ses
« un homme dont les deux sœurs sont généreuses»

b. **raʤul* *-u* *-n* *kari:m -ata:ni* *ʔuxta: -hu*
homme -NOM -Indé généreuses -NOM.D sœurs.NOM.D -ses

c. *raʤul* *-u* *-n* *ʔuxta: -hu* *kari:m -aat:ni*
homme -NOM -Indé sœurs.NOM.D -ses généreuses -NOM.D
« un homme dont les deux sœurs sont généreuses »

(66) a. *ʔal-* *ʔawla:d* *-u* *daxal* ***-u:*** (MP)
les- garçons -NOM entrer.PER -3MP
« Les garçons, ils sont entrés »

b. *daxal* ***-a*** *l-* *ʔawla:d* *-u*
entrer.PER -3M les- garçons -NOM
« Les garçons sont entrés »

c. **ʔal-* *ʔawla:d* *-u* *daxal* ***-a***
les- garçons -NOM entrer.PER -3MS

d. **daxal* ***-u:*** *l-* *ʔawla:d* *-u*
entrer.PER -3MP les- garçons -NOM

(67) a. *ʔal- bana:t -u daxal* ***-na*** (FP)
les- filles -NOM entrer.PER -3FP
« Les filles, elles sont entrées »

b. *daxal* ***-at*** *l- ʔbana:t -u*
entrer.PER -3F les- filles -NOM
« Les filles sont entrées »

c. **ʔal- ʔbana:t -u daxal* ***-at***
les- filles -NOM entrer.PER -3FS

d. **daxal* ***-na*** *l- ʔbana:t -u*
entrer.PER -3FP les- filles -NOM

(68) a. *ʔal- walad -a:ni daxal -a:* (MD)
les- garçons -NOM.D entrer.PER -3MD
« Les deux garçons, ils sont entrés »

b. *daxal* ***-a*** *l- ʔwalad -a:ni*
entrer.PER -3M les- garçons -NOM.D
« Les garçons sont entrés »

c. **ʔal- ʔwalad -a:ni daxal* ***-a***
les- garçons -NOM.D entrer.PER -3MS

d. **daxal -a: l- ʔwalad -a:ni*
entrer.PER -3MP les- garçons -NOM.D

(69) a. *ʔal- bint -a:ni daxal* ***-ata:*** (FD)
les- filles -NOM.D entrer.PER -3FD
« Les deux filles, elles sont entrées »

b. *daxal* ***-at*** *l- ʔbint -a:ni*
entrer.PER -3F les- filles -NOM.D
« Les garçons sont entrés »

c. **ʔal- ʔbint -a:ni daxal* ***-at***
les- filles -NOM.D entrer.PER -3FS

d. **daxal* ***-ata:*** *l- ʔbint -a:ni entrer.PER -3FP*
les- filles -NOM.D

(70) a. *ʔal- walad -u daxal* ***-a*** (MS)
le- garçon -NOM entrer.PER -3MS
« Le garçon, il est entré »

b. *daxal* ***-a*** *l- walad -u*
entrer.PER -3M le- garçon -NOM
« Le garçon est entré »

(71) a. *ʔal- bint -u daxal* ***-at*** (FS)
la- fille -NOM entrer.PER -3FS
« La fille, elle est entrée »

b. *daxal* ***-at*** *l- bint -u*
entrer.PER -3F la- fille -NOM
« La fille est entrée »

(72) a. *hum daxal* ***-u:*** (MP)
eux entrer.PER -3MP
« Eux, ils sont entrés »

b. **daxal* ***-a*** *hum*
entrer.PER -3M eux

c. **hum daxal* ***-a***
eux entrer.PER -3MS

d. *daxal* ***-u:*** *hum*
entrer.PER -3MP eux
« Ils sont entrés, eux »

(73) a. *hunna daxal* ***-na*** (FP)
elles entrer.PER -3FP
« Elles, elles sont entrées »

b. **daxal* ***-at*** *hunna*
entrer.PER -3F elles

c. **hunna daxal* ***-at***
elles entrer.PER -3FS

d. *daxal* ***-na*** *hunna*
entrer.PER -3FP elles
« Elles sont entrées, elles »

(74) a. *huma: daxal* ***-a:*** (MD)
eux.D entrer.PER -3MD
« Eux deux, ils sont entrés »

b. **daxal* ***-a*** *huma:*
entrer.PER -3M eux.D

c. **huma: daxal* ***-a***
eux.D entrer.PER -3MS

d. *daxal* ***-a:*** *huma:*
entrer.PER -3MP eux.D
« Ils sont entrés, eux deux »

(75) a. *huma:* *daxal* *-ata:* (FD)
elles.D entrer.PER -3FD
« Elles deux, elles sont entrées »

b. **daxal* *-at* *huma:*
entrer.PER -3F elles.D

c. **huma:* *daxal* *-at*
elles.D entrer.PER -3FS

d. *daxal* *-ata:* *huma:*
entrer.PER -3FP elles.D
« Elles sont entrées, elles »

(76) a. *huwa* *daxal* *-a*
lui entrer.PER -3MS (MS)
« Lui, il est entré »

b. *daxal* *-a* *huwa*
entrer.PER -3M lui
« Il est entré, lui »

(77) a. *hija* *daxal* *-at* (FS)
elle entrer.PER -3FS
« Elle, elle est entrée »

b. *daxal* *-at* *hija*
entrer.PER -3F elle
« Elle est entrée, elle »

(78) *ʔiddaʕa:* *ʔaħmad* *-u* *ʔanna* *-hu* *ʤa:ʔ* *-a* *r-* *raʤul* *-u*
prétendre.PER.3M Ahmed -NOM que -il venir.PER -3M l'- homme -NOM
« Ahmed a prétendu que l'homme est arrivé »

(79) a. *hum(u)* *l-* *ʤun:d* *-u* (MP)
eux les- soldats -NOM
« Ce sont les soldats »

b. *hunna* *n-* *nisa:ʔ* *-u* (FP)
elles les- femmes -NOM
« Ce sont les femmes »

(80) a. *l-* *wla:d* *xurʤ* *-u:* (SN V)
les- garçons sortir.PER -3MP
« Les garçons sont sortis »

b. *xurʁ* *-u:* *l-* *wla:d* (V SN)
sortir.PER -3MP les- garçons
« Les garçons sont sortis »

(81) a. *wi:n* ***-hum***$_i$ *l- wla:d*$_i$
où -ils les- garçons
« Où sont les garçons ? »

b. *ma-* *ħla:* ***-ha:***$_i$ *tu:nis*$_i$
que- belle -elle Tunisie
« Que la Tunisie est belle ! »

(82) *dixl* ***-u:***$_i$ *l- wla:d*$_i$
entrer.PER -3MP les- garçons
« Les garçons sont entrés »

Chapitre IV

(83) a. *ʔinna* *d- dali:l -a* *j-* *aqu:d* *-u* *n- na:s -a* *ʔila:*
que le- guide -ACC 3MS- guider.IMP -Ind les- gens -ACC vers
l- qas'r -i
le- palais -GÉN
« Le guide, il dirige les gens vers le palais, certes »

b. *ʔinna* ***-hu*** *j-* *aqu:d* *-u* ***-hum*** *ʔilaj* ***-hi***
que -lui 3MS- guider.IMP -Ind -les vers- lui
« Lui, il les y dirige, certes »

(84) a. *kull* *-a* *jawm -i -n*
chaque -ACC jour -ACC -Indé
« chaque jour »

a'. **kull* *-a* ***-hu***
chaque -ACC -le

b. *kita:b* *-u* *t'- t'ibb -i*
livre -NOM la- médecine -GÉN
« le livre de médecine »

b'. **kita:b* *-u* ***-hu***
livre -NOM -le

c. *raʔaj* *-tu* *ʔumm* *-a* *zajd* *-in*
voir.PER -1S mère -ACC Zayd -GÉN
« J'ai vu la mère de Zayd »

c'. **raʔaj -tu -ha: zayd -in*
voir.PER -1S -la Zayd -GÉN

(85) a. **wahab -tu -hu -ka*
donner.PER -1S -le -te

b. **wahab -ta -hu: -ni:*
donner.PER -2MS -le -me

(86) a. *wahab -tu -ka -hu*
donner.PER -1S -te -le
« Je te l'ai donné »

b. *wahab -ta -ni: -hi*
donner.PER -2MS -me -le
« Tu me l'as donné »

(87) a. *ki/ta:/bu -/**kún/na*** (marqueur d'objet)
livre-NOM -votre(FP)
« Votre livre »

b. *ka/tab -**tún/na*** (marqueur de sujet)
écrire.PER -3FP
« Vous(F) avez écrit »

(88) a. *ki/tá:/bu -**hu*** (marqueur d'objet)
livre-NOM -son
« son livre »

b. *ká/ta/b-**u:*** (marqueur de sujet)
écrire.PER-3MS
« Ils ont écrit »

(89) a. **ħatta: -**hu***
jusque -lui

b. **ka -**hu***
comme -lui

(90) a. *ħatta: -ma*
jusque -quoi
« Jusqu'à quand ? »

b. *ka* *-ma:*
comme -cela
« Comme cela »

(91) a. *ħatta:* *l-* *masa:ʔ* *-i*
jusque le- soir -GÉN
« Jusqu'au soir »

b. *ka* *s-* *sajl* *-i*
comme le- fluide -GÉN
« comme le fluide »

(92) a. *kita:b* *-u* ***-ha:*** *wa* *qalam* *-u* ***-ha:*** (marqueur d'objet)
livre -NOM -son et crayon -NOM -son
« son livre et son crayon »

b. **kita:b* *-u* ***-ha:*** *wa* *qalam* *-u*
livre -NOM -son et crayon -NOM

c. **kita:b* *-u* *wa* *qalam* *-u* ***-ha:***
livre -NOM et crayon -NOM -son

(93) a. *ʔakal* ***-tu*** *wa* *ʃarib* ***-tu*** (marqueur de sujet)
manger.PER -1S et boire.PER -1S
« J'ai mangé et j'ai bu »

b. **ʔakal* ***-tu*** *wa* *ʃarib*
manger.PER -1S et boire.PER

c. **ʔakal* *wa* *ʃarib* ***-tu***
manger.PER et boire.PER -1S

(94) a. *maqtal* *-u* *wa* *ʤarħ* *-u* *l-* *ʕaʃra:t* *-i*
mort -NOM et blessure -NOM les- dizaines -GÉN
« La mort et la blessure de dizaines [de gens] »

b. *maqtal* *-u* *l-* *ʕaʃra:t*$_i$ *-i* *wa* *ʤarħ* *-u* ***-hum***$_i$
mort -NOM les- dizaines -GÉN et blessure -NOM -leur
« La mort et la blessure de dizaines [de gens] »

(95) a. *raʔaj* *-tu* ***-ha:***
voir.PER -1S -la
« Je l'ai vue »

b. *kita:b* *-u* ***-ha:***
livre -NOM -son
« son livre »

c. *min* ***-ha:***
de -elle
« d'elle »

d. *ʔinna* ***-ha:***
que -elle
« Elle est certes ... »

e. *t'awi:l* *-u* *ʃ-* *ʃaʕr* *-i* *ʤami:l* *-u* ***-hu***
long -NOM les- cheveux -GÉN beau -NOM -le
« aux cheveux longs et beaux »

Chapitre V

(96) *katab* *-a* *l-* *walad* *-u* *risa:lat* *-a* *-n*
écrire.PER -3M le- garçon -NOM lettre -ACC -Indé
« Le garçon a écrit une lettre »

(97) a. *ma:ða:* *faʕal* *-a* *l-* *ʔawla:d* *-u* *?*
que faire.PER -3M les- garçons -NOM
« Qu'ont fait les garçons ?»

b. *xaraʤ* *-u:*
sortir.PER -3MP
« Ils sont sortis. »

(98) a. *ʔal-* *walad* *-u* *katab* *-a* *risa:lat* *-a* *-n*
le- garçon -NOM écrire.PER -3MS lettre -ACC -Indé
« Le garçon, il a écrit une lettre »

b. *ʔal-* *walad* *-u* *raʔaj* *-tu* *-hu*
le- garçon -NOM voir.PER -1S -le
« Le garçon, je l'ai vu »

(99) a. *ʔinna* *-ha:* *daxal* *-at* *leila*
que -elle entrer.PER -3F Leila
« Le fait est que Leila est entrée »

b. *ʔinna* *-hu* *daxal* *-a* *l-* *ʔawla:d* *-u*
que -il entrer.PER -3M les- garçons -NOM
« Le fait est que les garçons sont entrés »

(100) *ʔinna* ***-hum*** *j-* *adxul* *-u:* *-na* *l-* *manzil* *-a*
que -eux 3- entrer.IMP -MP -Ind la- maison -ACC
« Eux, ils entrent dans la maison, certes »

(101) a. *wahab* *-ta* *-ni:* *-hi*
offrir.PER -2MS -me -le
« Tu me l'as offert »

b. *wahab* *-tu* *-ka* *-hu*
offrir.PER -1S -te -le
« Je te l'ai offert »

c. *wahab* *-tu* *-ha:* *-hu*
offrir.PER -1S -la -le
« Je le lui ai offert (à elle) »

(102) a. **wahab* *-ta* *-hu:* *-ni:*
offrir.PER -2MS -le -me

b. **wahab* *-tu* *-hu* *-ka*
offrir.PER -1S -le -te

c. **wahab* *-a* *-ka* *-ni:*
offrir.PER -3MS -te -me

(103) a. *daxal* *-a* *l-* *walad* *-u*
entrer.PER -3M le- garçon -NOM
« Le garçon est entré »

b. *daxal* *-at* *l-* *bint* *-u*
entrer.PER -3F la- fille -NOM
« La fille est entrée »

c. **daxal* *-a* *l-* *bint* *-u*
entrer.PER -3M la- fille -NOM

d. **daxal* *-at* *l-* *walad* *-u*
entrer.PER -3F le- garçon -NOM

(104) *daxal* *-at* *l-* *bana:t* *-u*
entrer.PER -3F les- filles -NOM
« Les filles sont entrées »

(105) *ʔal* *walad* *-u* *katab* *-a* *ʔal-* *qissat* *-a*
le garçon -NOM écrire.PER 3MS l'- histoire -ACC
« Le garçon, il a écrit l'histoire »

(106) **katab* *-u:* *ʔal* *awla:d* *-u* *ʔal-* *qissat* *-a*
écrire.PER -3MP les- garçons -NOM l'- histoire -ACC

(107) a. *ʔal binta:ni katab -ata: ʔal- qissat -a*
les- filles.NOM écrire.PER -3FD l'- histoire -ACC
« Les deux filles, elles ont écrit l'histoire »

b. **ʔal binta:ni katab -at ʔal- qissat -a*
les- filles.NOM écrire.PER -3FS l'- histoire -ACC

c. **ʔal binta:ni katab -a: ʔal- qissat -a*
les- filles.NOM écrire.PER -3MD l'- histoire -ACC

(108) a. *na:m -a ʔal ʔawla:d -u*
dormir.PER -3M les- garçons -NOM
« Les garçons se sont couchés »

b. **na:m -u: ʔal ʔawla:d -u*
dormir.PER -3MP les- garçons -NOM

(109) *katab -tu ʔal qissatajni*
écrire.PER -1S les- histoires.D.ACC
« J'ai écrit les deux histoires »

(110) a. *ʔal- walada:ni j- aktub -a: -ni -ha:*
les- garçons.D.NOM 3- écrire.IMP -MD -Ind -la
« Les deux garçons, ils l'écrivent »

b. *j- aktub -u -ha: ʔal- walada:ni*
3M- écrire.IMP -Ind -la les- garçons.D.NOM
« Les deux garçons l'écrivent »

c. *ʔal- bint -u wahab -at -ka -ha:*
la- fille -NOM donner.PER -3FS -te -la
« La fille, elle te l'a donnée »

d. **ʔal- bint -u wahab -at -ha: -ku*
la- fille -NOM donner.PER -3FS -la -te

(111) *katab -a ʔal- walad -u qissat -a -hu*
écrire.PER -3M le- garçon -NOM histoire -ACC -son
« Le garçon a écrit son histoire »

(112) *ʔal- walad -u xa:f -a min ha:*
le- garçon -NOM craindre.PER -3MS de -elle
« Le garçon, il a eu peur d'elle »

(113) a. *ʔinna -ha: ʔal- bint -u*
que -elle la- fille -NOM
« C'est la fille »

b. *ʔinna -ha: katab -at ʔal- qissat -a*
que -elle écrire.PER -3FS l'- histoire -ACC
« Elle, elle a écrit l'histoire, certes »

c. *ʔinna -ha: min -kum*
que -elle de -vous
« Elle fait partie de [votre groupe de gens], certes »

Conclusion

(114) a. *bana:t -u -ka kun* **-na** *j- alʕab -na*
filles -NOM -tes être -3FS 3- jouer.IMP -FS
« Tes filles, elles étaient en train de jouer »

b. *ka:n* **-at** *bana:t -u -ka j- alʕab -na*
être -3F filles -NOM -tes 3- jouer.IMP -FS
« Tes filles étaient en train de jouer »

PHRASES SOUMISES À L'ANALYSEUR

Ces phrases sont celles que nous avons soumises à l'analyseur de LKB et quelques-unes de celles qu'il a générées. La liste contient des phrases grammaticales et d'autres qui sont agrammaticales (marquées d'un *). Ces phrases sont regroupées dans le fichier `test.items`, qui permet de tester plusieurs phrases à la fois.

```
al bintaani katabataa al qissata.

*al bintaani katabat al qissata.

*al bintaani katabaa al qissata.

;;;;

naama al waladu.

yanaamu al waladu.

al waladu naama.

al waladu yanaamu.

naama al waladaani.

yanaamu al waladaani.

al waladaani naamaa.

al waladaani yanaamaani.
```

naama al 9awlaadu.

yanaamu al 9awlaadu.

al 9awlaadu naamuu.

al 9awlaadu yanaamuuna.

*naamuu al 9awlaadu.

*yanaamuuna al 9awlaadu.

*naamaa al waladaani.

*yanaamaani al waladaani.

;;;; Cette série peut être augmentée avec, comme complément,

;;;; « al qisasa » et « al qissatayni ». Ce sont d'ailleurs les

;;;; phrases que LKB génère à partir des items *katab* et *qissa*.

katabtu al qissata.

katabnaa al qissata.

katabta al qissata.

katabti al qissata.

katabtumaa al qissata.

katabtum al qissata.

katabtunna al qissata.

kataba al qissata.

katabat al qissata.

katabaa al qissata.

katabataa al qissata.

katabuu al qissata.

katabna al qissata.

9aktubu al qissata.

naktubu al qissata.

taktubu al qissata.

taktubiina al qissata.

taktubaani al qissata.

taktubuuna al qissata.

taktubna al qissata.

yaktubu al qissata.

taktubu al qissata.

yaktubaani al qissata.

taktubaani al qissata.

yaktubuuna al qissata.

yaktubna al qissata.

;;;;

al waladaani yaktubaani haa.

yaktubu haa al waladaani.

al bintu wahabat ka haa.

*al bintu wahabat haa ka.

kataba al waladu qissata hu.

al waladu xaafa min haa.

inna haa al bintu.

inna haa katabat al qissata.

inna haa min kum.

RÉFÉRENCES

Abeillé, Anne. 1993. *Les Nouvelles Syntaxes : grammaires d'unification et analyse du français.* Paris : Armand Collin.

Abeillé, Anne, Danièle Godard, Philip H. Miller et Ivan A. Sag. 1998. «French Bounded Dependencies». *In* Balari Sergio et Dini Luca (éditeurs), «Romance in HPSG», Stanford : CSLI Publications, pp. 1–54.

Abramson, Harvey et Veronica Dahl. 1989. *Logic Grammars.* Springer-Verlag.

Abu Nahleh, Lamis Yacoub. 1985. *A syntactic-Semantic Analysis of Object Clitics and Pronouns in Ramallah Palestinian Arabic.* Thèse de doctorat, University of Michigan.

Ackema, Peter et Ad Neeleman. 2003. «Context-Sensitive Spell-Out». *Natural Language & Linguistic Theory*, vol. 21, pp. 681–735.

Adger, David, Cécile De Cat et George Tsoulas. 2004. *Peripheries : Syntactic Edges and their Effects.* Studies in natural language and linguistic theory ; v. 59, Dordrecht : Kluwer Academic.

Akkal, Ahmed. 1996. «Word Order and Related Issues in Standard Arabic : A Minimalist Approach». *Recherches linguistiques / Linguistic Research*, vol. 1, no 1, pp. 1–33.

Alexopoulou, Theodora, Edit Doron et Caroline Heycock. 2004. «Broad Subjects and Clitic Left Dislocation». *In* David Adger, Cécile De Cat et George Tsoulas (éditeurs), «Peripheries : Syntactic edges and their effects», Dordrecht : Kluwer Academic Publishers, pp. 329–358.

Alexopoulou, Theodora et Dimitra Kolliakou. 2002. «On Linkhood, Topicalization and Clitic Left Dislocation». *Journal of Linguistics*, vol. 38, no 2, pp. 193–245.

Aloulou, Chafik. 2003. «Analyse syntaxique de l'arabe : le système MASPAR». *In* «Colloques TALN et RÉCITAL», VVF de Batz-sur-Mer.

Aloulou, Chafik, Lamia Hadrich Belguith et Abdelmajid Ben Hamadou. 2002. «MASPAR : A Multi-Agent System for Parsing Arabic». *IEEE International Conference on Systems, Man and Cybernetics*, vol. 7, p. 5.

Aloulou, Chafik, Lamia Hadrich Belguith, Ahmed Hadj Kacem et Hammami Mezghani Souha. 2003. «Implémentation du système MASPAR selon une approche multi-agent». *In* «Papers from IWPT2003, 8th International Workshop of Parsing Technologies», Nancy, France, p. 2.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge : Cambridge University Press.

———. 2005. *Aspects of the Theory of Clitics*. Oxford : Oxford University Press.

Aoun, Joseph. 1996. «Clitic-Doubled Arguments», university of Southern California.

Aoun, Joseph et Elabas Benmamoun. 1998. «Minimality, Reconstruction, and PF Movement». *Linguistic Inquiry*, vol. 29, no 4, pp. 569–597.

Aoun, Joseph, Elabbas Benmamoun et Dominique Sportiche. 1994. «Agreement, Word Order, and Conjugation in Some Varieties of Arabic». *Linguistic Inquiry*, vol. 25, no 2, pp. 195–220.

———. 1999. «Further Remarks on First Conjunct Agreement». *Linguistic Inquiry*, vol. 30, no 4, pp. 669–681.

Ariel, Mira. 2000. «The Development of Person Agreement Markers : From Pronouns to Higher Accessibility Markers». *In* Barlow Michael et Suzanne Kemmer (éditeurs), «Usage-Based Models of Language», Stanford : CSLI Publications, pp. 197–260.

Astarabaadii, RaDhy Eddiine. 1978. *SharH Al Kaafiya*. Lybie : Université de Qaar Younes, (grammairien du VIIIe siècle).

Auger, Julie. 1994. *Pronominal Clitics in Québec Colloquial French : A Morphological Analysis*. Thèse de doctorat, University of Pennsylvania, Philadelphia.

———. 1995. «Les clitiques pronominaux en français parlé informel : une approche morphologique». *Revue québécoise de linguistique*, vol. 24, no 1, pp. 21–60.

Auger, Julie et Philip H. Miller. 1994. «French Clitics». *Linguist List*, vol. 5, no 1088, pp. Disponible en ligne, URL : http ://listserv.linguistlist.org/cgi-bin/wa ?A2=ind9410a&L=linguist&P=3408.

Bahloul, Maher et Wayne Harbert. 1992. «Agreement Asymmetries in Arabic». *In* «The Eleventh West Coast Conference on Formal Linguistics», vol. 11, pp. 15–31.

Bahou, Younes, Lamia Hadrich Belguith, Chafik Aloulou et Abdelmajid Ben Hamadou. 2006. «Adaptation et implémentation des grammaires HPSG pour l'analyse de textes arabes non voyellés». *In* «Actes du 15e congrés francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle», Tours, France.

Baker, Mark C. 1988. *Incorporation : A Theory of Grammatical Function Changing*. Chicago : University of Chicago Press.

Barwise, Jon et John Perry. 1983. *Situations and Attitudes*. Cambridge, Mass. : MIT Press.

Beller, Charley. 2006. *The Role of Argument Structure in Arabic Impoverished Agreement Constructions*. Mémoire de maîtrise, University of Rochester.

Benmamoun, Elabbas. 1995. «The Conditions on pro and the ECP». *In* Raul Aranovich, William Byrne, Susanne Preuss et Martha Senturia (éditeurs), «Proceedings of WCCFL», Stanford : CSLI, pp. 173–188.

———. 2000a. «Agreement Asymmetries and the PF Interface». *In* Jacqueline Lecarme, Jean Lowenstamm et Ur Shlonsky (éditeurs), «Research in Afroasiatic Grammar», Amsterdam : John Benjamin Publishing Company.

———. 2000b. *The Feature Structure of Functional Catergories : a Comparative Study of Arabic Dialects*. Oxford studies in comparative syntax, New York : Oxford University Press.

Benmamoun, Elabbas et Heidi Lorimor. 2006. «Featureless Expressions : When Morphophonological Markers are Absent». *Linguistic Inquiry*, vol. 37, no 1, pp. 1–23.

Benveniste, Emile. 1966. *Problèmes de linguistique générale*. Paris : Gallimard.

Blachère, Régis et Maurice Gaudefroy-Demombines. 2004. *Grammaire de l'arabe classique*. Le Liban : Librairie américaine et orientale.

Bloch, Ariel. 1986. *Studies in Arabic Syntax and Semantics*. Wiesbaden : O. Harrasowitz.

Borer, Hagit. 1984. *Parametric Syntax : Case Studies in Semitic and Romance Languages*. Dordrecht : Foris Publications.

Borsley, Robert. 1987. «Subjects and Complements in HPSG». Rapport technique numéro csli-107-87, CSLI, Stanford.

Bouchard, Denis. 2002. *Adjectives, Number and Interfaces : Why languages vary ?* North-Holland : Elsevier.

Bresnan, Joan. 1982. *The Mental Representation of Grammatical Relations*. Cambridge, Mass. : MIT Press, (ed.).

———. 2001. «The Emergence of the Unmarked Pronoun». *In* Geraldine Legendre, Jane Grimshaw et Sten Vikner (éditeurs), «Optimality-theoretic Syntax», Massachusetts : MIT Press, pp. 113–142.

Bresnan, Joan et Sam A. Mchombo. 1987. «Topic, Pronoun and Agreement in Chichewa». *Language*, vol. 63, no 4, pp. 741–782.

Broselow, Ellen. 1983. «A lexical Treatment of Cairene Arabic Object Clitics». *Current Approaches to African Linguistics*, vol. 1, pp. 287–297, foris, Dordrecht.

Bulk, Andreas. 2004. «A Functional Account of Pronominal Clitics in Spoken Arabic». Camling.

Cantarino, Vicente. 1975. *Syntax of Modern Arabic Prose : the Expanded Sentence (Vol. II)*. Bloomington : Indiana University Press.

Cardinaletti, Anna et Lori Repetti. 2006. «Proclitic vs. Enclitic Pronouns in Northern Italian Dialects», manuscrit.

Cardinaletti, Anna et Michael Starke. 1994. «The Typology of Structural Deficiency : On the Three Grammatical Classes». *Working Papers in Linguistics*, vol. 4, no 2, pp. 41–109.

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures*. Cambridge, England : Cambridge University Press.

Carpenter, Bob et Gerald Penn. 1995. «Compiling Typed Attribute-Value Logic Grammars». *In* H Bunt et M Tomita (éditeurs), «Current issues in parsing technologies», Dordrecht : Kluwer, vol. 2.

———. 1999. «ALE : The Attribute Logic Engine - User's Guide». Rapport technique, Lucent Technologies et Universität Tübingen.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Linguistic inquiry monographs. 6, Cambridge, Mass. : MIT Press.

———. 1995. *The Minimalist Program*. Cambridge, Mass. : The MIT Press.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford : CSLI Publications.

Copestake, Ann et Dan Flickinger. 2000. «An Open-Source Grammar Development Environment and Broad-Coverage English Grammar Using HPSG». *In* «Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)», Athènes.

Copestake, Ann, Daniel Flickinger, Ivan A. Sag et Carl Pollard. 1999. «Minimal Recusion Semantics : An Introduction», manuscrit.

Crysmann, Berthold. 2000. «Syntactic Transparency of Pronominal Affixes». *In* Ronnie Cann, Claire Grover et Philip Miller (éditeurs), «Grammatical Interfaces in HPSG», Stanford : CSLI Publications, pp. 77–96.

Cummins, Sarah et Yves Roberge. 1994. «A Morphosyntactic Analysis of Romance Clitic Constructions». *In* M. Mazzola (éditeur), «Issues and theory in romance linguistics», Washington : Georgetown University Press, pp. 237–255.

Eid, Mushira. 1993. «Negation and Predicate Heads in Arabic». *In* Mushira Eid et Gregory Iverson (éditeurs), «Principles and predication : the analysis of natural language. Papers in honor of Gerald Sanders», Amsterdam : John Benjamins, vol. 98, pp. 135–152.

Eloussfourri, Ezzohra. 1998. *Les pronoms personnels de l'arabe : syntaxe interne et externe.* Thèse de doctorat, Université Paris VIII.

Everett, Daniel L. 1985. «Piraha Clitic-Doubling and the Parametrization of Nominal Clitics». *In* «GLOW», Dordrecht : Foris.

Fassi Fehri, Abdelkader. 1988. «Agreement in Arabic, Binding and Coherence». *In* Michael Barlow et Charles A. Ferguson (éditeurs), «Approaches, theories, descriptions», Chicago : The University of Chicago Press, pp. 107–158.

———. 1992. «Sous-spécification, accord et pronoms en arabe». *Revue québécoise de linguistique*, vol. 22, no 1, pp. 117–144.

———. 1993. *Issues in the Structure of Arabic Clauses and Words.* Dordrecht : Kluwer Academic Publishers.

———. 2000. «Distributing Features and Affixes in Arabic Subject Verb Agreement Paradigms». *In* Lecarme Jacqueline, Lowenstamm Jean et Shlonsky Ur (éditeurs), «Research in Afroasiatic Grammar», Amsterdam : John Benjamins Publishing Company, Research in Afroasiatic Grammar, pp. 79–100.

Fleisch, Henri. 1979. *Traité de philologie arabe.* Beyrouth : Dar El-Machreq Éditeurs.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum et Ivan Sag. 1985. *Generalized Phrase Structure Grammar.* Oxford : Basil Blackwell.

Ginzburg, Jonathan et Ivan Sag. 2000. *Interrogative Investigations : The Form, Meaning and Use of English Interrogative Constructions.* Stanford : CSLI Publications.

Givon, Talmy. 1976. «Topic, Pronoun and Grammatical Agreement». *In* Charles Li et Vicente Cantarino (éditeurs), «Subject and topic», New York : Academic Press, pp. 149–188.

———. 1983. *Topic Continuity in Discourse : A Quantitative Cross-Language Study.* Amsterdam : John Benjamins.

Gray, Louis. 1934. *Introduction to Comparative Semitic Linguistics.* Amsterdam : Philo Press.

Guerssel, Mohamed. 1985. «Some Notes on the Strucure of Berber», cambridge, Mass. : MIT.

———. 1995. «Berber Clitic Doubling and Syntactic Extraction». *Revue québécoise de linguistique*, vol. 24, no 1, pp. 111–133.

Habash, Nizar. 2007. «Arabic Morphological Representations for Machine Translation». *In* Abdelhadi Soudi, Antal Van Den Bosch et Neumann Günter (éditeurs), «Arabic Computational Morphology : Knowledge-Based and Empirical Methods», Dordrecht : Springer, vol. 38 de *Text, Speech and Language Technology*, pp. 263–286.

Hale, Kenneth. 1973. «Person Marking in Walbiri». *In* Stephen Anderson et Paul Kiparsky (éditeurs), «A Festschrift for Morris Halle», New York : Holt, Rinehart & Winston Inc., pp. 308–344.

Halpern, Aaron. 1995. *On the Placement and Morphology of Clitics.* Center for the Study of Language (CSLI).

Harbert, Wayne et Maher Bahloul. 2002. «Postverbal Subjects in Arabic and the Theory of Agreement». *In* Jamal Ouhalla et Ur Shlonsky (éditeurs), «Themes in Arabic and Hebrew Syntax», Dordrecht : Kluwer Academic Publishers, pp. 45–70.

Hassan, Abbas. 1987. *Ennahw el Waafii.* Le Caire : Daar el Ma9aarif.

Haysharii, al-Shaadhilii. 2003. *Al-Damiir : bunyatuh wa-dawruh fii al-jumlah.* Tunis : Jaami9at Manuubah Kulliyyat al-aadaab.

Heap, David. 1997. «Subject Pronoun Variation in the Northern Italian Geolinguistic Continuum». *QUADERNI d'italianistica*, vol. XVIII, no 2, pp. 235–250.

Heap, David et Yves Roberge. 2001. «Clitisation et théorie syntaxique, 1971-2001». *Revue québécoise de linguistique*, vol. 30, no 1, pp. 63–90.

Hoberman, Robert D. 1988. «Local and Long-Distance Spreading in Semitic Morphology». *Natural Language and Linguistic Theory*, vol. 6, no 4, pp. 541–549.

Jackendoff, Ray. 1975. «Morphological and Semantic Regularities in the Lexicon». *Language*, vol. 51, pp. 639–671.

Jaeggli, Oswaldo. 1982. *Topics in Romance Syntax*. Dordrecht : Foris.

Jamari, Abderrahim. 1993. *Clitic Phenomena in Arabic*. Thèse de doctorat, London University.

Jelinek, Eloise. 2002. «Agreement, Clitics and Focus in Egyptian Arabic». *In* Jamal Ouhalla et Ur Shlonsky (éditeurs), «Themes in Arabic and Hebrew syntax», Dordrecht : Kluwer, pp. 71–105.

Joshi, Aravind K. et Yves Schabes. 1997. «Tree-Adjoining Grammars». *In* G. Rozenberg et A. Salomaa (éditeurs), «Handbook of Formal Languages», Berlin, New York : Springer, vol. 3, pp. 69–124.

Kaiser, Georg A. 1994. «More About INFL-ection : The Acquisition of Clitic Pronouns in French». *In* Jürgen M. Meisel (éditeur), «Bilingual First Language Acquisition, French and German Grammatical Development», Amsterdam : Benjamins, pp. 131–159.

Kamp, Hans et Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht : Kluwer.

Kathol, Andreas. 1999. «Agreement and the Syntax-Morphology Interface in HPSG». *In* Robert Levine et Georgia Green (éditeurs), «Studies in Current Phrase Structure Grammar», Cambridge : Cambridge University Press, pp. 223–274.

Kay, Martin. 1979. «Functional Grammar». *In* «Proceedings of the 5th Meeting of the Berkeley Linguistics Society», .

Kayne, Richard S. 1975. *Syntaxe du français : le cycle transformationnel*. Paris : Seuil.

———. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass. : MIT Press.

———. 2000. *Parameters and Universals*. Oxford : Oxford University Press.

Keenan, Edward L. 1976. «Towards a Universal Definition of "Subject"». *In* Charles Li (éditeur), «Subject and Topic», New York : Academic Press, pp. 303–333.

Kenstowicz, Michael. 1989. «The Null Subject Parameter in Modern Arabic Dialects». *In* Osvaldo Jaeggli et Kenneth Safir (éditeurs), «The Null Subject Parameter», Dordrecht : Kluwer, pp. 1–44.

Klavans, Judith L. 1983. «The morphology of cliticization». *In* «Papers from the parasession on the interplay of phonology, morphology and syntax», Chicago : Chicago Linguistic Society, pp. 103–121.

———. 1985. «The Independence of Syntax and Phonology in Cliticization». *Language*, vol. 61, no 1, pp. 95–120.

———. 1995. *On Clitics and Cliticization : The Interaction of Morphology, Phonology, and Syntax*. New York : Garland.

Koenig, Jean-Pierre. 1999. *Lexical Relations*. CSLI Publications.

Krieger, Hans-Ulrich et John Nerbonne. 1993. «Feature-Based Inheritance Networks for Computational Lexicons». *In* Ted Briscoe, Valeria de Paiva et Ann Copestake (éditeurs), «Inheritance, Defaults and the Lexicon», Cambridge : Cambridge University Press, pp. 90–136.

Kupsc, Anna. 1999. «Clitic Climbing in Polish Verb Clusters : An HPSG Approach». *In* M. Kruijff Geert-Jan et T. Oehrle Richard (éditeurs), «Proceedings of Formal Grammar», Utrecht, The Netherlands, pp. 75–82.

———. 2000. *An HPSG Grammar of Polish Clitics*. Thèse de doctorat, Institute of Computer Science, Polish Academy of Sciences et Université Paris 7, Warszawa, Poland.

Labelle, Marie. 1985. «Caractère post-lexical de la cliticisation française». *Lingvisticae Investigationes*, vol. IX, no 1, pp. 83–96.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form : Topic, Focus, and the Mental Representations of Discourse Referents*. Angleterre : Cambridge University Press Cambridge.

Larcher, Pierre. 2003. *Le sytème verbal de l'arabe classique*. Publications de l'Université de Provence.

Letourneau, Mark S. 1993. «Case-Making and Binding of Subject Clitics in Arabic Complement Clauses». *In* Mushira Eid et Clive Holes (éditeurs), «Perspectives on Arabic Linguistics V : Papers from the Fifth Annual Symposium on Arabic Linguistics», Amsterdam : John Benjamins Publishing Company, pp. 261–291.

Li, Charles N. et Sandra A. Thompson. 1976. «Subject and Topic : A New Typology of Language». *In* Charles N. Li (éditeur), «Subject and Topic», New York : Academic Press, pp. 457–489.

Lumsden, John S. et Girma Haleform. 2003. «Verb Conjugations and the Strong Pronoun Declension in Standard Arabic». *In* Jacqueline Lecarme (éditeur), «Research in Afro-Asiatic Grammar II», Amsterdam : John Benjamins Publishing Company, pp. 305–337.

Mahfoudhi, Abdessatar. 2002. «Agreement Lost, Agreement Regained : a Minimalist Account of Word Order Variation in Arabic». *California Linguistic Notes*, vol. 27, no 2, pp. 1–28.

Manning, Christopher D. et Ivan A. Sag. 1999. «Dissociations between Argument Structure and Grammatical Relations». *In* Gert Webelhuth, Jean-Pierre Koenig et Andreas Kathol (éditeurs), «Lexical And Constructional Aspects of Linguistic Explanation», Stanford, CA : CSLI Publications, pp. 63–78.

Marusic, Franc. 2002. «*Aff-stem-ix : On the Nature of Discntinuous Affixes». *In* «The CUNY/SUNY/NYU Linguistics Mini-Conference dedicated to the memory of Jerrold J. Katz», NY University.

McCarthy, John. 1981. «A Prosodic Theory of Nonconcatenative Morphology». *Linguistic Inquiry*, vol. 12, pp. 373–418.

———. 1985. *Formal Problems in Semitic Phonology and Morphology*. New York : Garland.

Meillet, Antoine. 1912. *L'évolution des formes grammaticales*. Réimprimé dans (Meillet, 1958) Linguistique historique et linguistique générale. Paris : Champion.

Melnik, Nurit. 2007. «From "Hand-written" to Computationally Implemented HPSG Theories». *Research on Language & Computation*, vol. 5, no 2, pp. 199–236.

Miller, Philip H. 1992. *Clitics and Constituents in Phrase Structure Grammar*. New York : Garland.

Miller, Philip H. et Paola Monachesi. 2003. «Les pronoms clitiques dans les langues romanes». *In* Danièle Godard (éditeur), «Langues Romanes : problèmes de la phrase simple», Paris : Éditions du CNRS, pp. 67–123.

Miller, Philip H. et Ivan A. Sag. 1995. «Une analyse lexicaliste des affixes pronominaux en français». *Revue québécoise de linguistique*, vol. 24, no 1, pp. 135–171.

———. 1997. «French Clitic Movement without Clitics or Movement». *Natural Language and Linguistic Theory*, vol. 15, no 3, pp. 573–639.

Mohammad, Mohammad A. 1990. «The Problem of Subject-Verb Agreement in Arabic : Towards a Solution». *In* Mushira Eid (éditeur), «Perspectives on Arabic Linguistics», Amsterdam : John Benjamins Publishing Company, vol. 1, pp. 95–125.

———. 2000. *Word Order, Agreement, and Pronominalization in Standard and Palestinian Arabic.* Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, v. 181, Amsterdam : John Benjamins Publishing Company.

Monachesi, Paola. 1998a. «Decomposing Italian Clitics». *In* Balari Sergio et Dini Luca (éditeurs), «Romance in HPSG», Stanford : CSLI Publications, pp. 305–357.

———. 1998b. «On the Representation of Italian Clitics». *In* U. Kleinhenz (éditeur), «Interfaces in Phonology», Berlin : Akademie Verlag, pp. 83–101.

———. 1999a. «A Lexical Analysis of Italian Clitics». *In* «Proceedings of VEXTAL'99», Venezia : San Servolo, pp. 57–65.

———. 1999b. «The Morphosyntax of Romanian Cliticization». *In* P. Coppen, H. Van Halteren et L. Teunissen (éditeurs), «Proceedings of Computational Linguistics in The Netehrlands 1997», Amsterdam : Rodopi, pp. 99–118.

Moosally, Michelle Jamila. 1994. *Resumptive Pronouns in Modern Standard Arabic : a Head-Driven Phrase Structure Grammar Account.* Mémoire de maîtrise, University of Texas, Austin.

Moravcsik, Edith. 1974. «Object-Verb Agreement». *Working Papers in Language Universals*, vol. 15, pp. 25–140.

Moscati, Sabatino. 1964. *An introduction to the Comparative Grammar of the Semitic Languages : Phonology and Morphology.* Wiesbaden : O. Harrassowitz.

Mullen, Dana Shirley. 1986. *Issues in the Morphology and Phonology of Amharic : the Lexical Generation of Pronominal Clitics.* Thèse de doctorat, University of Ottawa.

Munn, Alan. 1999. «First Conjunct Agreement : Against a Clausal Analysis». *Linguistic Inquiry*, vol. 30, no 4, pp. 643–668.

Mutawa, A. M., Salah Alnajem et Fadi Alzhouri. 2007. «An HPSG Approach to Arabic Nominal Sentences». *Journal of the American Society for Information Science and Technology*, vol. 59, no 3, pp. 422–434.

N'Diaye-Correard, Geneviève. 2003. «Structure des propositions et système verbal en wolof». *Revue électronique internationale de sciences du langage SUDLANGUES*, , no 3, pp. 163–188.

Nevis, Joel Ashmore. 1988. *Finnish Particle Clitics and General Clitic Theory*. New York : Garland Publishing.

Noyer, Robert Rolf. 1992. *Features, Positions and Affixes in Autonomous Morphological Structure*. Thèse de doctorat, MIT.

———. 1997. *Features, Positions, and Affixes in Autonomous Morphological Structure*. Outstanding dissertations in linguistics, New York : Garland.

Penn, Gerald. 1999a. «A Generalized-Domain-Based Approach to Serbo-Croatian Second Position Clitic Placement». *In* Bouma Gosse, Erhard Hinrichs, Geert-Jan Krujiff et Richard Oehrle (éditeurs), «Constraints and Resources in Natural Language Syntax and Semantics», Stanford : CSLI Publications, pp. 119–136.

———. 1999b. «An RSRL Formalization of Serbo-Croatian Clitic Placement». *In* Valia Kordoni (éditeur), «Tuebingen Studies in Head-Driven Phrase Structure Grammar», Universität Tübingen : Arbeitspapiere des Sonderforschungsbereichs 340, Nr. 132, pp. 177–197.

Perlmutter, David. 1971. *Deep and Surface Structure Constraints in Syntax*. New-York : Holt, Rinehart & Winston.

Philippaki-Warburton, Irene, Spyridoula Varlokosta, Michalis Georgiafentis et George Kotzoglou. 2002. «On the Status of Clitics and their Doubles in Greek». *Working Papers in Linguistics*, vol. 6, pp. 57–84.

Pollard, Carl. 1996. «The Nature of Constraint-Based Grammar». *In* «Pacific Asia Conference on Language, Information, and Computation», Kyung Hee University, Seoul, Korea.

———. 1997. «Lectures on the Foundations of HPSG», manuscrit, Ohio State University.

Pollard, Carl et Ivan A. Sag. 1987. *Information-Based Syntax and Semantics, Vol. 1*. Stanford : Stanford University Press, CSLI.

———. 1994. *Head-Driven Phrase Structure Grammar*. Stanford : Stanford University Press, CSLI.

Prunet, Jean-François, Renée Béland et Ali Idrissi. 2000. «The Mental Representation of Semitic Words». *Linguistic Inquiry*, vol. 31, no 4, pp. 609–648.

Ramiro, Campos. 1986. *Inflectional Elements in Romance.* Thèse de doctorat, University of California, Los Angeles.

Ritter, Elizabeth. 1997. «Agreement in the Arabic Prefix Conjugation : Evidence for a Non-Linear Approach to Person, Number and Gender Features». *In* Leslie Blair, Christine Burns et Lorna Roswell (éditeurs), «Proceedings of the Canadian Linguistic Association Conference», Calgary, Department of Linguistics : University of Calgary, pp. 191–202.

Rivas, Alberto Mario. 1977. *A Theory of Clitics.* Thèse de doctorat, MIT.

Roberge, Yves. 1986. «Subject Doubling, Free Inversion, and Null Argument Languages». *Revue canadienne de linguistique*, vol. 31, pp. 55–79.

Sag, Ivan A. et Thomas Wasow. 1999. *Syntactic Theory : A Formal Introduction.* Stanford : CSLI Publications.

Sag, Ivan A., Thomas Wasow et Emily Bender. 2003. *Syntactic Theory : A Formal Introduction.* CSLI Publications.

Shieber, Stuart. 1986. *An Introduction to Unification-Based Approaches to Grammar.* Stanford University : CSLI Publications.

———. 1991. «Natural-Language Processing : Grammar Formalisms». *In* William Bright (éditeur), «The Oxford International Encyclopedia of Linguistics», New York, New York : Oxford University Press, pp. 61–64.

———. 1992. *Constraint-Based Grammar Formalisms.* MIT Press.

Shieber, Stuart, Hans Uszkoreit, Fernando Pereira, Jane Robinson et Mabry Tyson. 1983. «The Formalism and Implementation of PATR-II». *In* Barbara Grosz et Mark Stickel (éditeurs), «Research on Interactive Acquisition and Use of Knowledge», pp. 39–79.

Shlonsky, Ur. 1997. *Clause Structure and Word Order in Hebrew and Arabic : An Essay in Comparative Semitic Syntax.* New York · Oxford University Press.

Shulte-Nafeh, Martha. 2004. *Overt Subjet Pronouns in Cairene Arabic : Pragmatic and Syntactic Functions.* Thèse de doctorat, University of Arizona.

Sibawayhi. 1977. *Al kitaab.* Le Caire . Muhammad Abd al-Salaam Haaroun, (grammairien du VIIIe siècle).

Siegel, Melanie et Emily Bender. 2002. «Efficient Deep Processing of Japanese». *In* «Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization. Coling 2002 Post-Conference Workshop», Taipei, Taiwan.

Smets, Martine. 2000. *Unification-Based Paradigmatic Morphology : A Theoretically-Motivated Approach to Computational Morphology.* Thèse de doctorat, University of Illinois at Urbana-Champaign.

Soltan, Usama. 2004. «An Argument for AGREE and Multiple Spell-out : Standard Arabic Agreement Asymmetries Revisited». *In* «The Workshop on Minimalist Theorizing», Indiana University, Bloomington.

Sportiche, Daniel. 1983. *Structural Invariance and Symmetry in Syntax.* Thèse de doctorat, MIT.

Steedman, Mark. 2000. *The Syntactic Process.* Cambridge, Mass. : MIT Press.

Stump, Gregory. 1980. «An "Inflectional" Approach to French clitics». *Ohio State University Workig Papers in Linguistics*, vol. 24, pp. 1–54.

Suner, Margarita. 1992. «Subject Clitics in the Northern Italian Vernaculars and the Matching Hypothesis». *Natural Language and Linguistic Theory*, vol. 10, pp. 641–672.

Webelhuth, Gert, Jean-Pierre Koenig et Andrea Kathol. 1999. *Lexical and Constructional Aspects of Linguistic Explanation.* Stanford : CSLI Publications.

Wright, William. 1898. *A Grammar of the Arabic Language.* Cambridge : The University Press.

Zwicky, Arnold M. 1977. *On Clitics.* Bloomington : Indiana University Linguistic Club.

———. 1985a. «Cliticization versus Inflection : The Hidatsa Mood Markers». *International Journal of American Linguistics*, vol. 51, no 4, pp. 629–630.

———. 1985b. «Clitics and Particles». *Language*, vol. 61, no 2, pp. 283–305.

———. 1987. «Suppressing the Zs». *Journal of Linguistics*, vol. 23, pp. 133–148.

Zwicky, Arnold M. et Geoffrey Pullum. 1983. «Clicticization vs. Inflection : English n't». *Language*, vol. 59, no 3, pp. 502–513.