



SAPIENZA
UNIVERSITÀ DI ROMA

Vers une ontologie de la culture arabo-musulmane

Dipartimento di Istituto italiano di studi orientali (ISO)
CIVILTÀ DELL'ASIA E DELL'AFRICA

Candidate

Ouafae Nahli

Numéro de matricule

1644315

Tuteur

Prof. Giuliano Lancioni

Cotuteur

Dr. Simonetta Montemagni

Dr. Vito Pirelli

Anno Accademico 2017/2018

Vers une ontologie de la culture arabo-musulmane
Thèse de doctorat - Université de Rome

© 2017 Ouafae Nahli. All rights reserved

Email auteur : ouafaenahli@gmail.com

Table des matières

Table des matières	3
INTRODUCTION	13
LA LANGUE ARABE CARACTÉRISTIQUES MORPHOSYNTAXIQUES ET SÉMAN- TIQUES	29
1 Catégories grammaticales	31
2 Notions de dérivation	35
2.1 Principes de morphologie arabe	35
2.1.1 La racine	36
2.1.1.1 <i>ḥurūf az=ziyyādah</i> 'les lettres d'augment'	36
2.1.1.2 Racines tri-consonantiques et quadri-consonantiques	37
2.1.2 Le schème (ou le gabarit)	38
2.2 Les différents types de dérivation	38
2.3 Théorie des Matrices et Étymons	41
2.4 Théorie autosegmentale et représentations multilinéaires	42
2.5 Restrictions des cooccurrences	45
2.5.1 Introduction	45
2.5.2 Le Principe du Contour Obligatoire	46
2.5.2.1 Application de l'OCP dans la mélodie consonantique	47
2.5.2.1.1 Cas des racines avec gémination	47
2.5.2.1.2 Cas des consonnes homorganiques	49
2.5.2.2 Application de l'OCP dans la mélodie vocalique	52
3 Le système verbal	53

3.1	Étude morphologique du verbe arabe	53
3.1.1	Classification du verbe arabe	53
3.1.2	Analyse morphologique du verbe <i>māḍī</i> (accompli)	55
3.1.2.1	Conjugaison du verbe <i>māḍī</i>	55
3.1.2.2	Les désinences flexionnelles du <i>māḍī</i>	56
3.1.2.2.1	Analyse morphologique	56
3.1.2.2.2	Analyse morphosyntaxique	56
3.1.2.2.2.1	Cas du SN postverbal	56
3.1.2.2.2.2	Cas du SN préverbal (ordre SVO)	58
3.1.2.2.3	Conclusion	59
3.1.3	Analyse morphologique du verbe <i>muḍāriʿ</i> (inaccompli)	60
3.1.3.1	Conjugaison du verbe <i>muḍāriʿ</i>	60
3.1.3.2	Analyse des affixes du <i>muḍāriʿ</i>	61
3.1.3.2.1	Analyse des préfixes	61
3.1.3.2.2	Analyse des suffixes	62
3.1.3.2.3	Analyse de la partie invariable des affixes	63
3.1.3.2.3.1	Étude morphologique	63
3.1.3.2.3.2	Étude syntaxique	66
3.1.4	Conclusion	69
3.1.5	Pattern (ou mélodie) vocalique du <i>māḍī</i> et du <i>muḍāriʿ</i>	69
3.2	Études morphosémantiques du verbe arabe	71
3.2.1	Le verbe trilitère	71
3.2.1.1	Classes morphosyntaxiques et sémantiques	71
3.2.1.2	Les verbes de schème C ₁ aC ₂ uC ₃ a	71
3.2.1.2.1	Alternance vocalique	71
3.2.1.2.2	Valeurs sémantiques	72
3.2.1.2.3	Adjectif associé au verbe C ₁ aC ₂ uC ₃ a	73
3.2.1.3	Les verbes de schème C ₁ aC ₂ iC ₃ a	75
3.2.1.3.1	Alternance vocalique	75
3.2.1.3.2	Valeurs sémantiques	76
3.2.1.4	Les verbes de schème C ₁ aC ₂ aC ₃ a	78
3.2.1.4.1	Alternances vocaliques	78
3.2.1.4.2	Valeurs sémantiques	80
3.2.1.5	Formalisation de l'ablaut en langue arabe	82
3.2.2	Les verbes dérivés du verbe trilitère	83
3.2.2.1	Introduction	83

3.2.2.2	La forme IV : 'aC ₁ C ₂ aC ₃ -a	87
3.2.2.2.1	Valeur causative	87
3.2.2.2.2	Valeur privative	88
3.2.2.2.3	Valeur expositive	88
3.2.2.2.4	Valeur estimative	89
3.2.2.2.5	Valeur du verbe trilitère	89
3.2.2.2.6	Valeur adverbiale	90
3.2.2.3	La forme II : C ₁ aC ₂ C ₂ aC ₃ -a	90
3.2.2.3.1	Pluriactionnalité ou pluralité verbale	90
3.2.2.3.2	Causativité	92
3.2.2.3.3	Valeur dénomminative	94
3.2.2.4	La forme III : C ₁ āC ₂ aC ₃ -a	94
3.2.2.4.1	La participation	94
3.2.2.4.2	L'exagération	96
3.2.2.5	Les formes dérivées en t	97
3.2.2.5.1	La forme V : taC ₁ aC ₂ C ₂ aC ₃ -a	97
3.2.2.5.2	La forme VI : taC ₁ āC ₂ aC ₃ -a	97
3.2.2.5.3	La forme X : (i)staC ₁ C ₂ aC ₃ -a	98
3.2.2.6	Les formes VII et VIII : (i)nC ₁ aC ₂ aC ₃ -a et (i)C ₁ taC ₂ aC ₃ -a	99
3.2.2.7	Les autres formes dérivées	101
3.2.3	Le verbe quadrilitère et ses dérivés	102
4	Le système nominal	105
4.1	Introduction	105
4.1.1	Critères morphologiques	105
4.1.2	Critères sémantiques	106
4.1.3	Critères syntaxiques	107
4.2	Caractéristiques du système nominal	108
4.2.1	Dans la flexion	108
4.2.1.1	Le genre	108
4.2.1.2	Le nombre	110
4.2.1.2.1	Le singulier et le duel	110
4.2.1.2.2	Le pluriel sain	110
4.2.1.2.3	Le pluriel brisé	111
4.2.1.2.3.1	Le pluriel de paucité	111
4.2.1.2.3.2	Le pluriel de grand nombre	111

4.2.1.2.3.3	Le pluriel du pluriel	116
4.2.1.2.3.4	Le pluriel extrême	116
4.2.1.2.3.5	Le nom du pluriel	116
4.2.1.2.4	Le nom collectif	117
4.2.1.3	Système nominale vs système verbale	117
4.2.2	Dans la dérivation	118
4.2.2.1	Les noms primitifs	118
4.2.2.2	Les noms déverbaux	118
4.2.2.2.1	Le nom verbal	118
4.2.2.2.1.1	Le <i>maṣḍar</i> du verbe «nu »	119
4.2.2.2.1.2	Le <i>maṣḍar</i> des verbes dérivés	120
4.2.2.2.2	Le nom d'agent et le nom de patient	120
4.2.2.2.3	L'adjectif analogue	121
4.2.2.2.4	Le nom de lieu et le nom de temps	121
4.2.2.2.5	Le nom d'instrument	123

LE LEXIQUE ARABE

NAISSANCE, HISTOIRE, ÉVOLUTION

125

1	Tradition lexicographique arabe	127
1.1	Introduction	127
1.2	Collecte des données linguistiques	130
1.3	Les sources fiables	132
1.3.1	Les citations en poésie	132
1.3.2	Les citations en prose	136
1.4	Époque de la collecte	137
2	Lexiques de la langue arabe	141
2.1	Introduction	141
2.2	Lexiques onomasiologiques	144
2.2.1	Domaine lexicale	144
2.2.1.1	Le <i>ġarīb</i>	144
2.2.1.1.1	<i>ġarīb al=Qur'ān</i>	146
2.2.1.1.2	<i>ġarīb al=Ḥadīṭ</i>	147
2.2.1.1.3	Le <i>ġarīb</i> général	148
2.2.1.2	Les emprunts et les paroles arabisées	149

2.2.1.3	Néologisme et solécisme	151
2.2.1.4	Synonymie, homonymie et énantiosémie	154
2.2.1.4.1	La synonymie	155
2.2.1.4.1.1	<i>at=taṣḥīf</i> 'erreurs'.	155
2.2.1.4.1.2	Les procédés de la récolte des données	157
2.2.1.4.2	L'homonymie	158
2.2.1.4.3	L'énantiosémie	159
2.2.1.5	La métathèse	161
2.2.1.6	Les lettres et les particules	161
2.2.1.6.1	Les lettres	162
2.2.1.6.2	Les particules	165
2.2.2	Domaine de la flexion	166
2.2.2.1	Le singulier, le duel et le pluriel	166
2.2.2.2	Le masculin et le féminin	167
2.2.3	Domaine de la dérivation	168
2.2.3.1	<i>al=muṭallaṭāt</i> 'les triplets'	168
2.2.3.2	Les schèmes nominaux	169
2.2.3.3	Les schèmes verbaux	170
2.2.3.4	Les schèmes nominaux et verbaux	171
2.2.4	Domaine sémantique	172
2.2.4.1	Livres des proverbes	172
2.2.4.2	Thèmes sémantiques divers	172
2.2.5	Thésaurus à thèmes divers	174
2.3	Lexiques sémasiologiques	175
2.3.1	Ordre phonétique-permutatif	175
2.3.2	Ordre permutatif / alphabétique	179
2.3.3	Ordre alphabétique second le dernier radical de la racine	181

CRÉATION D'UNE RESSOURCE SÉMANTIQUE

al=qāmūs al=muḥīṭ

189

1	Étude de <i>al=qāmūs al=muḥīṭ</i>	191
1.1	L'introduction de <i>qāmūs</i>	192
1.1.1	La richesse des unités lexicales	193
1.1.2	L'exhaustivité associée à la simplification et à la condensation	194
1.1.3	Le contrôle des diacritiques	195

1.1.4	Le contrôle vocalique	196
1.2	Macrostructure du <i>qāmūs</i>	197
1.2.1	Organisation et dimensions	198
1.2.2	Critères lexicaux	200
1.2.3	Critères linguistiques	201
1.2.4	Critères idéologiques	202
1.3	Microstructure du <i>qāmūs</i>	203
1.3.1	Microstructure des entrées lexicales nominales	203
1.3.2	Microstructure des entrées lexicales verbales	205
1.3.3	Nature des définitions	209
1.3.3.1	La définition dérivative	210
1.3.3.2	La définition synonymique	210
1.3.3.3	La définition antonymique	210
1.3.3.4	La définition hyperonymique	211
1.4	Conclusions	212
2	Acquisition d'une ressource numérique de <i>qāmūs</i>	213
2.1	Procédures d'acquisition et de structuration	213
2.1.1	Acquisition et pré-traitement du texte	213
2.1.2	Encodage de niveau supérieur	216
2.1.2.1	Le format XML	216
2.1.2.2	Balisage de la structure du contexte	216
2.1.2.3	Conversion de <i>qāmūs</i> en format XML semi-structuré	216
2.2	Extractions des informations et enrichissement de <i>qāmūs</i>	220
2.2.1	Extraction de la famille dérivationnelle et de la racine	222
2.2.2	Structuration et enrichissement des entrées verbales	223
2.2.2.1	Extraction et étiquetage du lemme	223
2.2.2.1.1	Règles de microstructure	223
2.2.2.1.2	Patrons morphosyntaxiques	224
2.2.2.1.3	Cas particuliers	228
2.2.2.2	Extraction des informations flexionnelles et dérivationnelles	230
2.2.2.2.1	Patrons morphologiques et morphosyntaxiques	230
2.2.2.2.2	Patrons syntaxiques	230
2.2.2.2.3	Patrons linguistiques	233
2.2.3	Structuration et enrichissement des entrées nominales	236

2.2.3.1	Règles de microstructure	236
2.2.3.2	Patrons morphosyntaxiques	237
2.2.4	Structuration et enrichissement des entrées adjectivales	242
2.2.5	Traitement des définitions	244
2.2.5.1	Cas où la définition est constituée d'un seul mot	244
2.2.5.1.1	Cas des entrée lexicales verbales	244
2.2.5.1.2	Cas des entrée lexicales nominales	245
2.2.5.1.3	Cas particuliers	248
2.2.5.2	Cas où la définition est constituée par plusieurs paroles	249
2.2.5.2.1	Cas où la définition est un antonyme	249
2.2.5.2.2	Cas où la définition est explicative	252
2.2.6	Quelques résultats obtenus	253
2.3	Enrichissement sémantique de <i>qāmūs</i>	255
2.3.1	Procédures d'extractions automatiques	255
2.3.2	Procédures d'alignement	258
2.3.3	Ressources lexicales bilingues utilisées	259
2.3.3.1	«An Arabic-English Dictionary »	259
2.3.3.2	«A Dictionary of Modern Written Arabic »	260
2.3.3.3	Analyse et comparaison de AED et DMWA	261
2.3.4	Ressources ontologiques utilisées	262
2.3.4.1	Définition	262
2.3.4.2	WordNet	264
2.3.4.2.1	Organisation de la WordNet	264
2.3.4.2.2	WordNet de la langue arabe (Arabic WordNet)	269
2.3.4.3	SUMO (The Suggested Upper Merged Ontology)	269
2.3.4.4	Alignement de la PWN avec SUMO	271
2.3.5	Alignement de <i>qāmūs</i> avec les ressources ontologiques	273
2.3.5.1	Cas de correspondance entre de <i>qāmūs</i> , PWN et SUMO	273
2.3.5.1.1	Levée d'ambigüité	273
2.3.5.1.1.1	Cas où lemme et sens ont des traductions communes	274
2.3.5.1.1.2	Cas où lemme et sens n'ont pas des traductions communes	276
2.3.5.1.1.3	Cas où le lemme n'est pas traduit	278
2.3.5.1.2	Problèmes de désambigüisation	280
2.3.5.2	Cas de non correspondance entre <i>qāmūs</i> , PWN et SUMO	284

2.3.5.2.1	Problèmes stylistiques	284
2.3.5.2.2	Problèmes de culture et de religion	285
2.3.5.2.3	Problèmes morphosyntaxiques	286
2.3.5.2.4	Problèmes morphosémantiques	287
2.3.5.2.5	Problèmes ontologiques	287
2.4	Représentation standard de la ressource lexicale	289
2.4.1	LMF : Lexical Markup Framework	289
2.4.2	LEMON : LExicon Model for ONtologies	294
2.5	Éditeur LexO	296

DISSERTATION ET CONCLUSIONS **299**

Bibliographie **307**

Conventions

La transcription de l'alphabet arabe

consonne	transcription	consonne	transcription
ء	'	ض	<i>ḍ</i>
ب	<i>b</i>	ط	<i>ṭ</i>
ت	<i>t</i>	ظ	<i>ẓ</i>
ث	<i>ṯ</i>	ع	'
ج	<i>ǧ</i>	غ	<i>ǧ</i>
ح	<i>ḥ</i>	ف	<i>f</i>
خ	<i>ḫ</i>	ق	<i>q</i>
د	<i>d</i>	ك	<i>k</i>
ذ	<i>ḏ</i>	ل	<i>l</i>
ر	<i>r</i>	م	<i>m</i>
ز	<i>z</i>	ن	<i>n</i>
س	<i>s</i>	ه	<i>h</i>
ش	<i>š</i>	و	<i>w</i>
ص	<i>ṣ</i>	ي	<i>y</i>
ة	<i>ah / at</i>		

voyelle courte	transcription	voyelle longue	transcription
ا	<i>a</i>	آ - اَ - اِ	<i>ā</i>
و	<i>u</i>	وُ	<i>ū</i>
ي	<i>i</i>	يَ - يِ	<i>ī</i>

Les conventions suivies pour les gloses

On a suivi, dans la mesure du possible, les conventions données dans «The Leipzig Glossing Rules » ([Leipzig Glossing Rules, 2015](#)), dont les plus importants dans cette recherche sont :

Segmentable morphemes are separated by hyphens, both in the example and in the gloss. There must be exactly the same number of hyphens in the example and in the gloss.

Clitic boundaries are marked by an equals sign, both in the object language and in the gloss. ([Leipzig Glossing Rules, 2015, p. 2](#))

Vu les caractéristiques de l'écriture arabe, dans plusieurs cas, on a utilisé seulement la forme translittérée qui permet de diviser entre morphèmes, clitiques et paroles. Par commodité, le tableau suivant donne les abréviations les plus utilisées.

1	première personne
2	deuxième personne
3	troisième personne
ACC	accusatif
DEF	défini
DU	duel
F	féminin
GEN	génitif
IND	indicatif
INDF	indéfini

M	masculin
NOM	nominatif
OBJ	objet
PFV	perfectif
PL	pluriel
PST	passé
SBJ	sujet
SBJV	subjonctif
SG	singulier

Première partie

INTRODUCTION

Avant-propos

Cette recherche se situe au croisement entre plusieurs disciplines et, au fur et à mesure qu'on procède, on se trouve à faire face à différents aspects qui touchent la linguistique, la lexicographie, la numérisation et l'ontologie computationnelle.

Sur le front linguistique et lexicographique, quand il s'agit de la langue arabe, une des premières questions qu'on se trouve à affronter est la perspective synchronique et/ou diachronique. Cependant, on n'arrive pas à tracer une ligne temporelle, en parlant de langue classique ou moderne, parce que c'est une langue qui a fondu sa grammaire et son lexique dans la nuit des temps. Précisément, depuis le 6^e siècle, avec la naissance de la religion musulmane, l'arabe est préservé comme langue du Coran et utilisé par les musulmans, de manière quotidienne. Comme on ne peut, non plus, tracer une ligne géographique et parler de langue standard, lorsqu'on pense que c'est la langue officielle d'une large bande qui va de l'Afrique du Nord jusqu'au Moyen Orient.¹

Cependant, comme toutes les langues, la langue arabe a subi des changements au fil des siècles et elle en subit, également, le long de la vaste aire géographique où elle est pratiquée. Pour une enquête scientifique, on aurait besoin d'une étude comparée et approfondie, aussi bien synchronique que diachronique. L'ère du numérique ouvre la perspective à de fortes potentialités qu'une étude classique n'avait pas. Avoir accès à des enquêtes statistiques, au lieu du simple comptage, permet de maîtriser des ensembles de données de tailles considérables et, permet de cerner de manière plus précise les phénomènes complexes en question. La numérisation des lexiques arabes médiévaux et récents serait un premier pas, pour une utilisation plus fructueuse de celle que pouvait offrir l'enquête humaine sur les volumes en papier.

Une partie de cette recherche se penche sur les processus de numérisation et modélisation du lexique arabe en format numérique. Une des tâches les plus importantes est identifier les phénomènes linguistiques de surface qui permettent d'extraire l'information. Se situant à des niveaux hiérarchiques différents, on a dû procéder par étapes en considérant les aspects structuraux, morphologiques, syntaxiques et sémantiques qui sont sous-jacents dans le texte de nature lexicale.

L'enrichissement sémantique d'une ressource lexicale est devenu une tâche essentielle pour la nouvelle génération des dictionnaires qui doivent être accessibles et lisibles par les logiciels. Vu la prolifération des documents et le nombre énorme des pages web disponibles, on se trouve dans l'impossibilité d'administrer et catalo-

1. (Danks, 2010, page 38.)

guer les informations qui y circulent. Rédigées librement en langue naturelle, ils ne se conforment à aucune nomenclature et il en résulte que les résultats d'une recherche sur le web manquent de précision. Ces problèmes sont nés parce que les informations sont accessibles seulement pour les humains qui peuvent interpréter leurs contenus sémantiques, alors que les machines qui les contiennent en sont incapables parce que, d'une part, elles sont fondées sur des techniques statistiques, et de l'autre, elles fonctionnent isolées les unes des autres.²

Toutefois, si on arrive à expliciter les arguments que traitent les documents, on donne au logiciel (ou machine) la capacité de retrouver automatiquement toutes les données appartenant à la même catégorie et traitant le même objet. Les machines de la nouvelle génération pourraient « comprendre » et « raisonner », « structurer » et « sélectionner » les données qu'elles contiennent et elle pourraient alors interagir entre elles. C'est le principe du « web sémantique », dont le nom plus approprié est « le web des données » (linked data).

En d'autres termes, pour une meilleure manipulation et compréhension des documents électroniques, il faut transformer la connaissance humaine qui circule en désordre (dans le web) en une connaissance explicite et partagée.

Les ontologies, une des techniques les plus prometteuses, fournissent un point de départ pour atteindre ce but. En plus, elles peuvent être utilisées dans plusieurs autres domaines de l'intelligence artificielle, tels que le traitement automatique du langage naturel, dont l'objectif principal est d'exploiter la richesse de la langue naturelle afin de capter l'information portée par un texte. L'ontologie computationnelle est, également, utilisée dans la représentation et la gestion des connaissances ; l'indexation des documents ; la recherche de l'information et la traduction automatique qui permettrait de « comprendre » des textes en langues étrangères et exploiter des ressources qui, jusque maintenant, sont inaccessibles.

Connaissance lexicale versus connaissance ontologique Avant de procéder, il faut faire la distinction entre l'Ontologie philosophique et l'ontologie dans le domaine de l'Ingénierie des connaissances.

Dans le domaine philosophique, depuis qu'Aristote l'a définie, l'Ontologie a pour objet l'étude de l'être en tant qu'Être et les propriétés qui le définissent telles que l'existence, la possibilité, la durée et le devenir. En tant que telle, l'Ontologie est indépendante de la langue utilisée pour la décrire.

2. (Berners-Lee, 1999).

Dans le domaine de l'intelligence artificielle, l'ontologie computationnelle cherche à représenter, de manière formelle et opérationnelle, le/les sens d'un vocabulaire présent dans un texte donné et peut être parangonnée à une base de connaissances lexicales-sémantique :

Pour un domaine mettant en œuvre des connaissances complexes sur lesquelles on veut effectuer des traitements intelligents, le programme élaboré est un système à base de connaissances qui manipule une base de connaissances. Cette base répertorie, entre autres, et de la même façon que dans les systèmes d'information, les concepts du domaine hiérarchiquement organisés dans une « ontologie », nommée ainsi en faisant référence à ARISTOTE.³

En conclusion, dans les applications informatiques, une ontologie est un artéfact d'ingénierie qui sert surtout pour capturer l'information sémantique existante dans les documents pour assurer leur gestion de manière automatique. Elle décrit une certaine réalité, en utilisant un vocabulaire spécifique et un ensemble d'axiomes explicites concernant le sens des mots qui décrivent cette réalité. Les axiomes ont la forme d'une théorie logique de premier ordre, où les mots peuvent représenter les concepts et les relations.

Dans le cas où on arrive à structurer toutes les langues avec les mêmes principes et le même formalisme, on obtiendrait une ontologie qui serait un réseau sémantique dans une échelle gigantesque indispensable et servirait comme base de connaissances inter-langues pour : i) comprendre, interpréter et structurer les vocabulaires constituant un document ; ii) décrire, raisonner et cataloguer les documents que circulent dans le web ; iii) servir comme représentation pivot pour relier et faire communiquer entre eux différents systèmes hétérogènes (de langues et cultures différentes).

Dans ce cas, les ressources lexicales avec une hiérarchie sémantique peuvent servir de base dans la découverte et la conception d'une ontologie, et l'ontologie peut servir de fondement à la résolution sémantique du langage naturel.⁴ La conséquence directe qui en découle est le fait que la valeur informative d'un document (oral ou écrit) est véhiculée par les paroles qui le composent. Il en découle aussi qu'il est nécessaire de prendre en considération la langue utilisée dans les documents. En effet, le focus le plus important pour développer une ressource puissante et polyvalente est créer une synergie entre le savoir ontologique et le savoir lexicale.⁵

3. (Charlet et al. 2004).

4. (Prévot et al., 2010).

5. (Prévot et al., 2010).

L'intégration des connaissances lexicales et des connaissances ontologiques permettrait de faire le lien entre les concepts et les termes qui représentent leurs réalisations linguistiques.

Dans cette thèse, la parole «ontologie » désignera l'artéfact d'ingénierie qui formalise les concepts à travers les termes correspondant à leur(s) réalisation(s) linguistique(s).

Niveaux de représentation de la connaissance lexicale et de la connaissance ontologique La différence entre la connaissance lexicale et celle ontologique réside dans le niveau de la représentation de la connaissance. Le lexique met l'accent sur les propriétés et les relations linguistiques entre les mots, telles que la synonymie et l'antonymie. L'ontologie, par contre, se concentre sur les relations surtout hiérarchiques, telles que hyperonymie et l'hyponymie.⁶

Bien que le lexique est considéré le référentiel clé tenant une grande partie de l'information, il est destiné à un usage humain parce que l'information est sous-jacente. La division en entrées lexicales séparées empêche la perméabilité et, par conséquence, l'information et les relations paradigmatiques (synonymie, hyperonymie, méronymie, etc.) restent implicites. Prenant, comme exemple, les définitions des paroles *ṣaḥīfah*, *ṣafīḥah* et *waraqah* dans le dictionnaire de la langue arabe moderne, *mu'ğam al-luğah al-'arabiyyah al-mu'āširah*:⁷

صَحِيفَةٌ *ṣaḥīfah* 1- le journal [...]; 2- [le matériel] sur lequel on écrit, comme *waraq* ⁸ et similaire, et on appelle ce qui vient d'être écrit « *ṣaḥīfah maṣqūlah* » [...]; 3- le livre.

صَفِيحَةٌ *ṣafīḥah* [...] partie rectangulaire mince et grande d'un matériel, comme *waraq* ou le métal ou le verre ou le bois [...].

وَرَقَةٌ *waraqah* 1- une des feuilles qui constituent le livre et elle est constituée de deux pages (*ṣafḥah*) successives rétro verso; 2- la partie extrême verte, aplatie et mince qui sort de la tige (*sāq*) de la plante (*nabāt*), portée par un pétiole (*'unuq*), et elle produit le sucre (*sukkar*) grâce à la photosynthèse (*al=binā' ḍ=ḍaw'iyy*), et elle est souvent verte.

6. (Rastier, 2004).

7. (*mu'ğam al-luğah al-'arabiyyah al-mu'āširah*, 2008, respectivement en page 1272 - 1300 - 2427.)

8. *waraq* est le pluriel de *waraqah* 'feuille'

Le dictionnaire donne deux sens à la parole *waraqah*. C'est seulement en parcourant et en analysant *toutes* les définitions et après une analyse plus attentive, qu'on se rend compte qu'en fait, elle en a trois. En outre, on se rend compte qu'elle est liée avec d'autres paroles par plusieurs relations de synonymie et d'hyponymie mises en évidence dans le tableau 0.1.

Sens	partie de livre constituée de deux pages	partie de plante extrémité verte	matériel mince, grande
	matériel pour écrire	responsable de photosynthèse	-
Synonymes	صَحِيفَةٌ ; وَرَقَةٌ <i>waraqah ; ṣaḥīfah</i>	وَرَقَةٌ <i>waraqah</i>	صَفِيحَةٌ ; وَرَقَةٌ <i>waraqah ; ṣafīḥah</i>

Tab. 0.1 : Analyse des définitions de *waraqah*, *ṣaḥīfah* et *ṣafīḥah*

L'arrivée du support informatique a permis une véritable révolution dans le domaine de la lexicographie. La nouvelle génération des ressources lexicales se présente sous forme de bases de données où, grâce à une taxonomie organisée, les entrées lexicales sont interconnectées entre elles par un réseau de relations qui contribuent à leur signification. Ainsi, dans sa nouvelle conception, le lexique peut servir de base pour la conceptualisation des catégories caractéristiques d'une ontologie. On parle de lexique relationnel et, comme exemple, on peut citer WordNet de la langue anglaise (PWN)⁹ et FrameNet.¹⁰

La WordNet de la langue anglaise est le lexique numérisé le plus connu, dont les nœuds sont constitués par des ensembles de synonymes appelés synsets. Les synsets eux-mêmes sont liés entre eux et sont spécifiés par divers relations telles l'hyponymie et l'hyponymie. La figure 0.1 représente le résultat de recherche de la parole *leaf*, traduction de la parole *waraqah*, dans PWN.

En suivant l'organisation de PWN, le tableau 0.2 présente les synsets arabes contenant la parole *waraqah* et les relations hiérarchiques d'hyponymie-hyponymie qui les caractérisent. D'autres relations lexicales peuvent être ajoutées pour enrichir le lexique. Elles peuvent comprendre les relations d'agentivité (feuille-photosynthèse), d'instrumentalité (feuille-écrire) etc.

Toutefois, la construction manuelle de toute une ressource lexicale-sémantique se révèle une opération très couteuse et complexe et, de conséquence, il est d'usage

9. (Fellbaum, 1998).

10. (Fillmore et al. 2003).

ries existantes parce que le lexique omet (en général) toute référence aux catégories qui ne sont pas lexicalisées dans la langue qu'il traite. En fait, chaque langue présente des lacunes lexicales correspondantes à d'autres langues.¹¹ Par conséquent, il faut prendre en considération divers aspects comme les lacunes lexicales, les quasi-synonymes et le glissement lexical. Par exemple, le sens de la parole *faṭara* dans la langue arabe correspond aussi bien à :

فَطَرَ *faṭara*: prendre le petit-déjeuner ;

فَطَرَ *faṭara* : prendre le premier repas après une journée de jeûne.

Le premier sens n'est pas lexicalisé en français et exige une description de plusieurs mots . Par contre, il est lexicalisé dans la PWN, comme suite :

breakfast : eat an early morning meal ; "We breakfast at seven".

Le deuxième sens n'est lexicalisé dans aucune langue occidentale, parce qu'il est lié à des concepts strictement religieux et musulmans.

En plus, il est rare que les mots qui sont équivalents dans la traduction soient complètement identiques en sens. Les verbes arabes *lamaḥa* ; *ra'ā* ; *'abṣara* ; *šāhada* et *šāfa* peuvent être traduits en français, comme *apercevoir*, *distinguer*, *voir* et *constater*, alors que les significations de ces verbes varient :¹²

لَمَحَ *lamaḥa* entrevoir ; apercevoir à peine

رَأَى *ra'ā* voir ;

أَبْصَرَ *'abṣara* distinguer ou discerner par la vue

شَاهَدَ *šāhada* observer attentivement

شَافَ *šāfa* observer d'en haut ; voir

Dans la perspective d'une interconnexion et échange d'information entre documents de diverses langues et cultures, ces constatations mènent à penser qu'on aurait besoin d'une hiérarchie conceptuelle générique qui représente la signification du mot à un niveau plus générale, de sorte que les groupes des (quasi-)synonymes soient connectés

11. (Hirst, 2004).

12. (mu'ğam al-luğah al-'arabiyyah al-mu'āširah, 2008).

à un seul nœud : leur signification fondamentale.¹³ Dans ce cas, on obtiendrait une ontologie générique (dite aussi de haut niveau) où, les concepts sont génériques et décrivent des connaissances tels que l'état, l'action, l'espace et les composants. Toutes les langues naturelles *doivent être* en mesure de représenter les niveaux supérieurs de l'ontologie et couvrir au moins les concepts les plus saillants et importants de la vie quotidienne.

Si on prend le soin de mapper les ressources lexicales de diverses langues sur une même ontologie de haut de niveau, où chaque sens de mot est représenté comme un pointeur vers un concept ou une catégorie au sein d'une ontologie générale, le système, ainsi construit, peut agir comme une application inter-linguale, utile à toutes les applications qui exigent des ressources multilingues.¹⁴ Le projet EuroWordNet est né dans cette perspective, avec le but de créer une base de données multilingue construite à partir des WordNets de plusieurs langues européennes (néerlandais, italien, espagnol, allemand, français, tchèque etc.), structurés à la manière de WordNet anglais (PWN)¹⁵ et liés à un indice inter-lingual basé sur celui anglais. Par l'intermédiaire de cet indice, les langues sont interconnectées, de sorte qu'il est possible de passer d'un mot dans une langue à des mots similaires dans d'autres langues.

Un autre projet a été réalisée, en alignant PWN avec l'ontologie de haut niveau, SUMO (The Suggested Upper Merged Ontology), qui est une ontologie gratuite de propriété IEEE (Institute of Electrical and Electronics Engineers) et dont les ontologies de domaine constituent la plus grande ontologie publique, utilisées pour la recherche et pour les applications de recherche linguistique et de raisonnement.¹⁶ La PWN, jouant de pivot, permet de lier les autres ressources lexicales à SUMO.

Par exemple, les verbes *lamaḥa* ; *ra'ā* ; *'abšara* et *šāfa*, peuvent être traduits comme *to see* qui se trouve lié, dans SUMO, au concept Seeing, lui-même une sous-classe du concept Perception. Par contre, *šāhada* peut être traduit comme *to observe* qui est lié au concept Looking, lui même une sous-classe du concept Seeing.

Jusqu'à présent, une ontologie spécifique de la culture arabo-musulmane n'a jamais été réalisée et juste des fragments portant sur des aspects spécifiques ont été faites en suivant trois approches différentes.

13. (Hirst, 2004).

14. (Hirst, 2004).

15. (Fellbaum, 1998).

16. (Pease et al. 2002)

Projets d'ontologie spécifique de la langue arabe

Une ontologie arabe a été présentée comme une extension de l'ontologie SUMO, elle est formée de synsets spécifiques de la culture arabe (ArabicCulture.kif)¹⁷ mais, elle est très limitée et contient juste 70 termes. Ce travail présente, avant tout, certains problèmes typiques des ontologies annotées manuellement. D'un côté, elle contient un ensemble relativement aléatoire de données et, de l'autre, les annotations culturelles sont plutôt incomplètes. Par exemple, la parole *suḥūr*, définie dans SUMO comme « A light meal before starting a new day of Ramadan (before day break) », est codée comme une sous-classe de « Eating », sans aucune référence aux implications liées à la religion et à la culture. Par contre, la parole *fuṭūr* définie supra est absente.

Comme projet d'ontologie de « culture spécifique », on peut citer “Arabic Ontology Project” [AOP].¹⁸ Encore aux premiers stades, l'équipe AOP prévoit de construire une ontologie spécifique arabe de haut niveau qui sera connectée à PWN via un algorithme de correspondance de traduction : traduire les définitions arabes en anglais par les systèmes de traduction automatique disponibles gratuitement et trouver dans PWN les définitions plus proches et corriger manuellement les mauvais alignements.

Cependant, les ressources spécifiques d'une langue ont tendance à être d'un intérêt plutôt limité lorsqu'elles sont sans interconnexion avec les réseaux de connaissances universelles existantes. Un des aspects importants dans la conception d'une ontologie est l'intégration de l'information et l'interopérabilité entre divers systèmes. Or, même si deux systèmes adoptent le même vocabulaire, il n'y a aucune garantie qu'ils soient d'accord sur certaines informations à moins qu'ils s'engagent à la même conceptualisation.¹⁹ L'intégration de multiples ontologies locales peut ne pas fonctionner, surtout si elles sont concentrées seulement sur les relations conceptuelles pertinentes à un contexte spécifique. Par conséquent, il est plus facile se mettre d'accord sur une seule ontologie de haut niveau plutôt que de compter sur des accords basés sur l'intersection des différentes ontologies.²⁰ En conclusion, l'intérêt serait de construire un réseau de connaissances de la langue et culture arabe connecté directement aux ontologies préexistantes (par exemple, PWN et SUMO).

Bien que les ontologies sont souvent caractérisées comme universelles, c'est un fait que la plupart de celles actuelles ont été mises au point en prenant en considération surtout les langues et les civilités occidentales. Or il est difficile d'évaluer l'affirmation

17. Disponible sur le site :<https://github.com/ontologyportal/sumo/blob/master/ArabicCulture.kif>

18. (Jarrar, 2011).

19. (Guarino, 1998).

20. (Guarino, 1998).

qu'elles soient universelles sans essayer de les étendre à d'autres langues et cultures comme la civilisation arabo-islamique.

Objectifs de la thèse

Dans le cadre décrit ci-dessus, le projet vise à décrire les méthodologies permettant de développer un réseau de connaissance pour la culture arabo-islamique sur la base d'un processus d'extractions automatiques de données à partir du lexique arabe *al=qāmūs al=muḥīṭ* (*qāmūs*).

Le choix de *qāmūs* est justifié par le fait qu'il a un statut d'autorité dans le monde arabe, au point que la parole *qāmūs* [océan] a supplanté la parole *munǧid* 'dictionnaire'. Dans l'introduction, l'auteur, *al=fīrūz 'ābādī* (1329-1414), déclare qu'il a créé son ouvrage, en fusionnant plusieurs dictionnaires préexistants. Pour cette raison, il lui a donné le titre de *al=qāmūs al=muḥīṭ* 'l'océan universel'.

La contribution de *al=fīrūz 'ābādī*, à l'évolution du lexique arabe, est très importante à différents points de vue. Dans une tentative de rédiger un dictionnaire destiné à atteindre un large public, il fut l'un des premiers lexicographes à saisir l'essence du terme et présenter ses définitions et ses sens en mode succinct et condensé. L'objectif de l'auteur, en rédigeant son lexique, était l'exhaustivité. Dans le but d'avoir un "petit" dictionnaire, il a considérablement réduit le contenu original des dictionnaires de source et, il a abandonné la tradition de ses prédécesseurs en éliminant la plupart des citations coraniques, poésies, noms des grammairiens et des historiens.

Pour ce fait, le dictionnaire *qāmūs* se distingue des autres lexiques par plusieurs caractéristiques qui ont contribué au fait qu'il soit l'un des meilleurs. Il est bien structuré comme lexique et contient de courts éléments lexicaux. Malgré cela, il contient, aussi, les noms propres des personnes, des pays et des régions. On trouve aussi un grand nombre de noms de maladies, d'épées, des oiseaux, etc. L'auteur a donné aussi une grande importance aux noms des arbres, des végétaux et des herbes médicales, en expliquant leurs particularités.

Le projet prévoit divers étapes de travail et, avant tout, l'acquisition d'une version numérique de *qāmūs*. La mise au point d'algorithmes pour la codification partielle et automatique de la macrostructure lexicale et la conversion du lexique en format XML. D'autres algorithmes permettent l'identification de la microstructure lexicale et, l'annotation de chaque partie constituante de l'entrée lexicale, entre autres, le lemme, sa nature morphologique, ses définitions, etc.

En utilisant deux dictionnaires bilingues arabe-anglais,²¹ un système de recherche permet de trouver, de manière automatique et quand c'est possible, la traduction de chaque lemme, ce qui permet de le lier au synset correspondant dans PWN et au concept de SUMO à qui il pourrait faire référence. Une autre étape serait l'analyse de divers échantillons de lemmes pour détecter la validité des résultats.

Plan de la thèse

Après une introduction générale qui expose les problématiques ; les motivations et le but général de la thèse, cette dernière est composée de trois parties.

Afin de cerner, de manière rigoureuse, les contraintes linguistiques qui définissent la parole en langue arabe, la première partie est dédiée à une étude théorique de la langue arabe en soulignant divers aspects. L'objet de cette partie de la recherche décrit le système de construction des paroles, les constituants fondamentaux du lexique, et les divers aspects morphosyntaxiques et sémantiques qui les caractérisent. Elle se penche, également, sur les caractéristiques théoriques de l'organisation du lexique arabe. Ces notions auront une importance capitale puisqu'elles nous permettent de comprendre le matériel qu'on aura l'occasion de traiter et les méthodes qu'on pourra utiliser.

Le premier chapitre fait un rappel sur la catégorisation grammaticale de la parole, selon la grammaire traditionnelle. Le deuxième chapitre est consacré aux notions de schème et racine et approfondit le concept de la dérivation en langue arabe et son rôle dans l'enrichissement lexicographique et la création des paroles, comme il permet de comprendre les caractéristiques d'organisation du lexique arabe. Les autres chapitres se concentrent sur les caractéristiques morphologiques et sémantiques des deux grandes super-classes, celle des verbes et celle des noms.

La deuxième partie se concentre sur l'étude de la tradition de la lexicographie arabe et permet de tracer la naissance et l'évolution du lexique arabe. La présentation des principales écoles lexicographiques permet de saisir les aspects historiques, les faits, les motivations et les procédures de la collecte du matériel lexicographique. En d'autres termes, elle aide à comprendre le contexte historique, religieux et social dans lequel est né *qāmūs*. Le fait que chaque lexicographe a suivi les traces de ses prédécesseurs et a utilisé les dictionnaires publiés précédemment, permet une reconstruction généra-

21. (Salmoné, 1889) et (Wehr, 1971).

logique à travers laquelle, très souvent, on arrive à remonter, plusieurs siècles avant, aux origines des termes.

La troisième partie illustre les méthodologies et les différentes phases qui permettent la conversion de *qāmūs*, et généralement les dictionnaires arabes, dans un format numérisé. Elle permet de commenter les procédures suivies pour acquérir un lexique numérisé à partir des extractions d'informations du lexique.

Un premier chapitre est destiné à l'étude lexicographique de *qāmūs*. Il est indispensable de faire une analyse de l'introduction où l'auteur explique sa méthodologie et les abréviations qu'il a utilisées pour construire son dictionnaire. L'analyse de la macrostructure et de la microstructure du dictionnaire permet de définir et de mettre en évidence les régularités qui le caractérisent afin de les utiliser pour le marquage et la conversion en format XML.

Dans la tradition lexicographique arabe, la notion de l'entrée lexicale est très différente de celle qu'on a l'habitude d'étudier et de trouver dans un dictionnaire d'une langue indo-européenne. On peut la considérer comme une structure contenant, de manière implicite et condensée, des informations méta-textuelles et méta-lexicales. Le deuxième chapitre est destiné à décrire les procédures d'acquisition et de structuration du lexique en format numérisé XML et les mécanismes mis en place, pour automatiser les extractions lexicales. Grâce à des exemples extraits de *qāmūs*, les modalités d'extraction sont illustrées, en se basant sur des patrons identifiés dans le texte. Les informations terminologiques sont basées sur des patrons morphosyntaxiques, caractérisant la langue arabe, et des patrons syntaxiques, caractérisant le style de l'auteur. Les informations sémantiques, par contre, se basent sur les patrons morphosémantiques individualisés dans le lexique

Le troisième chapitre est dédié à décrire les procédures de projection des concepts et des relations sémantiques de la PWN et de SUMO vers la ressource numérisée de *qāmūs* et, enfin les mécanismes automatiques et manuels de désambiguïsation et de validation des résultats obtenus.

Un dernier chapitre se penche sur l'étude des modèles standard de représentation LMF (Lexical Markup Framework) et LEMON (LEXicon Model for ONtologies) et évalue leur compatibilité pour représenter le lexique de la langue arabe.

Au cours de cette recherche, plusieurs perplexités ont surgi concernant surtout la validité d'aligner un lexique arabe avec un lexique d'une langue différente comme l'anglais. Une conclusion générale permet d'illustrer le bilan général permettant de souligner surtout les aspects différents entre les deux langues qui, pris en considération, per-

mettraient de trouver des solutions pour rendre possible l'interopérabilité entre langues aussi différentes, comme l'arabe et l'anglais.

Ressources produites

En plus de la thèse qui trace les différentes phases du travail et explique les modalités suivies pour l'obtention d'une ressource lexico-sémantique, il sera mis à disposition, diverses ressources.

- Le dictionnaire, *qāmūs*, révisé et divisé en sections selon le troisième radical des racines. Les fichiers obtenus sont en format .TXT. Chaque fichier est divisé en chapitre, selon la première consonne des racines. Chaque fichier a été enrichi de marqueurs et de méta-données, qui permettent d'identifier les sections et les chapitres qui les constituent. À l'intérieur du chapitre, d'autres méta-données ajoutées permettent de le diviser en familles dérivationnelles (*root_family*) correspondant à l'ensemble de tous les vocables dérivant d'une racine. Enfin, les entrées lexicales sont segmentées et marquées, en respectant la microstructure de *qāmūs*.
- Il a été nécessaire d'avoir à disposition, également, un dictionnaire bilingue. Le seul qu'on a pu trouver à disposition, est «A Dictionary of Modern Written Arabic», compilé par Hans Wehr (1909-1981) et publié en 1961. La version téléchargée est numérisée en format Unicode et a été générée à l'aide d'un système OCR. De conséquence, il a été nécessaire d'effectuer plusieurs phases de corrections et de normalisation. La version obtenue est en format .TXT, divisée en trois fichiers contenant, respectivement, les verbes, les noms, et enfin les particules et les noms arabisés.
- Un exemple prototype de l'entrée lexicale, *kabura* 'être grand', créé manuellement en fonction des conventions de LMF «Lexical Markup Framework».
- Un fichier généré automatiquement en fonction des conventions de LMF et contenant 432 entrées lexicales verbales, extraites de *qāmūs* et révisées manuellement.
- En collaboration avec l'Université Sidi Mohamed Ben Abdellah (Maroc), Faculté des Sciences et Techniques Fès-Sais, il a été programmé un système d'extractions automatiques, qui sera disponible à peine seront finies les dernières révisions.

- En collaboration avec l'Istituto di Linguistica Computazionale du CNR de Pise, on a adopté LexO, un éditeur collaboratif pour les ressources multilingues lexicales et termino-ontologiques, qui sera disponible à peine terminés les derniers perfectionnements.²²

22. En utilisant un mot de passe, on peut le visiter à l'adresse :
<http://ditmao-dev.ilc.cnr.it:8082/Lexo4AIQuamus>

Deuxième partie

**LA LANGUE ARABE
CARACTÉRISTIQUES
MORPHOSYNTAXIQUES ET
SÉMANTIQUES**

To strive at higher mathematical formulas for linguistic meaning

while knowing nothing correctly of the shirt-sleeve

rudiments of language is to court disaster.

(Whorf, *Language, Thought, and Reality*, p. 223)

Chapitre 1

Catégories grammaticales

Pour organiser les parties du discours, la grammaire arabe classique suit une division tripartite, contenant trois catégories grammaticales ; le verbe, le nom et la particule, que al=*'astarabādī* (m. vers 686/1287) définit comme suite :

هَذَا اللَّفْظُ الدَّالُّ عَلَى مَعْنَى مُفْرَدٍ - أَعْنِي الْكَلِمَةَ - إِمَّا أَنْ يَدُلَّ عَلَى مَعْنَى فِي نَفْسِهِ، أَوْ عَلَى مَعْنَى لَا فِي نَفْسِهِ، الثَّانِي الْحَرْفُ، أَعْنِي الْكَلِمَةَ الدَّالَّةَ عَلَى مَعْنَى لَا فِي (نَفْسِهَا)، وَالْأَوَّلُ - أَيِ الْكَلِمَةُ الدَّالَّةَ عَلَى مَعْنَى فِي نَفْسِهَا - إِمَّا أَنْ تَقْتَرِنَ بِأَحَدِ الْأَزْمِنَةِ الثَّلَاثَةِ أَوْ لَا (الثَّانِي الْأِسْمُ، أَيِ الْكَلِمَةُ الدَّالَّةَ عَلَى مَعْنَى فِي نَفْسِهَا غَيْرِ مُقْتَرِنٍ بِأَحَدِ الْأَزْمِنَةِ الثَّلَاثَةِ) وَالْأَوَّلُ الْفِعْلُ، أَيِ الْكَلِمَةُ الدَّالَّةَ عَلَى مَعْنَى فِي نَفْسِهَا مُقْتَرِنٍ بِأَحَدِ الْأَزْمِنَةِ الثَّلَاثَةِ.

"ce terme qui indique un sens isolé - je veux dire la parole - soit il indique un sens intrinsèque, soit un sens non intrinsèque. Dans ce dernier cas, il s'agit de la particule. Par contre, les paroles qui indiquent un sens intrinsèque, elles peuvent être divisées en celles qui indiquent un sens corrélé à un des trois temps [passé, présent ou futur] et celle dont le sens n'est pas corrélé à un des trois temps. Dans le premier cas, il s'agit du verbe et, dans le deuxième cas, il s'agit du nom.¹

Le nom est donc une super-catégorie qui regroupe, à son tour, plusieurs catégories, entre autres, *ism al='ayn* 'le substantif', *al=mašdar* 'l'infinitif substantivé', *ism al=fā'il* 'le participe actif' ; *ism al=maf'ūl* 'le participe passif', *aš=šifah al=mušabbahah* 'l'adjectif

1. Traduction libre ; *'Astarabādī*, *Šarḥ al-kāfiyyah*, V. I, pp. 11-15.

analogue [au participe]', *aḍ=ḍamīr* 'le pronom', *ism al=išārah* 'le démonstratif', *al=ism l=mawṣūl* 'le pronom relatif' et *ism al=istifhām* 'les pronoms interrogatifs'.

Les grammairiens arabes ont proposé plusieurs définitions du nom. Certaines d'entre eux énumèrent les caractéristiques morphologiques, comme le fait d'être gouverné par une préposition et être au génitif. D'autres définitions citent les caractéristiques syntaxiques, comme le fait que le nom a la capacité d'agir comme *fā'il* 'agent' ou *maf'ūl* 'objet' d'un prédicat.

Versteegh fait noter que, dans certains cas, les termes *fā'il* et *maf'ūl* désignent les caractéristiques physiques des substances qui leur permettent d'être, respectivement, «active» et «passive».² Il ajoute :

My own interpretation is as follows: in speech nouns are the representatives of substances, and verbs of actions. Substances are the agents or objects of actions and thus, they are characterized by the fact that they are always either active or passive. Accordingly, nouns may be defined as active or passive, because they represent the substances. Activity and passivity as qualities of substances are known from Stoic doctrine [...] and the definition of a body as something which acts or is being acted upon was known in the Arab world from the translation of the *Placita Philosophorum* (277.18 kull *fā'il wa-kull munfa'il fa-huwa ġism*).³

Par contre, Versteegh identifie trois types de définitions concernant le verbe.⁴ Un premier type est morphologique, comme le fait que le verbe peut être combiné avec un suffixe pronominal, par exemple, le pronom =*tu* dans l'exemple 1. Un deuxième type de définition est syntaxique, par exemple, il est toujours utilisé comme prédicat, mais lui-même ne peut pas avoir un prédicat ou il peut être combiné avec un adverbe temporel, par exemple '*ams* 'hier' dans l'exemple 1.

(1)

katab=tu '*ams-an*
écrire.PFV=1S hiers-ACC
'j'ai écrit hier'

Et enfin, un type de définition sémantique, que visent à déterminer, non la nature linguistique du verbe, mais la nature de ce qui est signifié par le verbe. Selon ces

2. Versteegh, *Explanation*, note 2, p. 56.

3. Ibid., note 3, pp. 56-57.

4. Ibid., note 28, pp. 65-66.

définitions, outre les verbes proprement dits, la super-catégorie des verbes contient, entre autres, les «auxiliaires» (classe de *kāna* 'être') et les verbes épistémiques (classe *ẓanna* 'croire').

Enfin, la super-catégorie des particules comprend tout ce qui reste. La grammaire traditionnelle les divise en particules '*āmilah* 'opératives', c'est à dire celles qui ont une rection, et les particules *ġayr* '*āmilah* 'non opératives' qui n'ont pas de rection. Les particules '*āmilah* 'opératives' sont, elles mêmes, divisées en celles qui sont opératives sur le verbe et lui assignent une marque modale et celles qui le sont sur le noms, entre lesquelles on trouve les prépositions.

Chapitre 2

Notions de dérivation

2.1 Principes de morphologie arabe

Dans la tradition lexicographique occidentale, la structure classique du lexique suit l'ordre alphabétique. Mais, la structure de la langue arabe fait en sorte que le lexique, conçu selon l'ordre alphabétique, perd beaucoup d'informations qui lient les paroles dérivant de la même racine. En fait, les mots sont ramenés à la racine de laquelle ils dérivent, par exemple, les paroles citées en (1) sont quelques éléments dérivant de la racine *ktb* et les paroles citées en (2) dérivent de la racine *nḥr*.

(1) *kataba* 'écrire' ; *istaktaba* 'faire écrire pour soi' ; *maktabah* 'bibliothèque'.

(2) *naḥara* 'égorger' ; *'intaḥara* 'se suicider' ; *manḥar* 'gorge'.

Ainsi, la classification lexicale se fait en fonction des consonnes qui constituent les racines. Les paroles en (1) sont répertoriées sous la racine *ktb* qui, par ordre alphabétique, est classée avant la racine *nḥr*, sous laquelle sont regroupées les paroles en (2).

Du fait, la consultation du lexique arabe nécessite un minimum de connaissance de morphologie pour reconnaître les lettres ajoutées pour constituer une parole et pouvoir dégager la racine. Selon une tradition qui remonte au moyen âge, la langue arabe est constituée d'unités dont l'identité est définie par deux critères : i) un critère de nature lexicale : **la racine** et ii) un critère de nature grammaticale : **le schème** (ou le gabarit). Kouloughli définit les paroles arabes comme des unités ayant une propriété essentielle :

“Celle d’être analysables en deux strates distinctes. La première constituée uniquement de consonnes, ordonnées de façon stricte, s’appelle la racine du mot (en arabe *’aṣl*). La seconde, constituée de voyelles longues ou

brèves et parfois aussi de consonnes, s'appelle le gabarit (en arabe *ṣiġah*). Les mots qui ont même racine présentent en général une parenté sémantique plus ou moins marquée. Ceux qui ont même gabarit appartiennent en principe à la même classe de mots et subissent normalement les mêmes règles morphologiques (déclinaisons, conjugaisons, etc.).”¹

Ainsi, le mot appartient à deux systèmes distincts qui se complètent et dont chacun est véhiculaire d'une partie de sa définition. L'analyse d'un mot se présente second un schéma binaire dont quelques exemples sont illustrés dans le tableau 2.1.² On note que, verticalement, toutes les paroles véhiculent le même concept abstrait et présentent le même squelette consonantique ($C_1C_2C_3$) : la racine. Horizontalement, toutes les paroles véhiculent le même sens grammaticale et ont le même schème (ou gabarit), formé par les consonnes de la racine auxquelles s'intercalent un pattern vocalique et des fois l'ajout de lettres, appelées *ḥurūf az-zīyādah* 'les lettres d'augment', en position de préfixe, infixes et/ou suffixes.

vers un même schème					
Vers une même racine	' <i>abyaḍ</i> 'blanc'	' <i>aswad</i> 'noir'	' <i>azraq</i> 'bleu'	' $aC_1C_2aC_3$ (masculin)	Adjectif
	' <i>bayḍā'</i> 'blanche'	' <i>sawdā'</i> 'noire'	' <i>zarqā'</i> 'bleue'	$C_1aC_2C_3ā'$ (féminin)	de couleur
	' <i>bayyaḍa</i> 'blanchir'	' <i>sawwada</i> 'noircir'	' <i>zarrāqa</i> 'rendre bleu'	$C_1aC_2C_2aC_3a$	Verbe
	' <i>bayāḍ</i> 'blancheur'	' <i>sawād</i> 'noirceur'	' <i>zurqah</i> 'bleu'	$C_1aC_2āC_3 /$ $C_1uC_2C_3ah$	Substantif
	' <i>byḍ</i>	' <i>swd</i>	' <i>zrq</i>	$C_1C_2C_3$	Racine

Tab. 2.1 : schéma binaire d'analyse morphologique

2.1.1 La racine

2.1.1.1 *ḥurūf az-zīyādah* 'les lettres d'augment'

Ce sont les consonnes : ' - *m* - *n* - *s* - *t* - *l* - *h* - *y* - *w* et la lettre 'alif qui a pour rôle l'allongement de la voyelle /a/. Ces lettres peuvent contribuer à la dérivation comme affixes, et permettent de changer la valeur sémantique et créer de nouveaux lexèmes.

1. Kouloughli, 1994, p. 58.

2. Les exemples sont inspirés à ceux de Cantineau CANTINEAU, 1950, pp. 119-124.

Dans ce cas, elles font partie du schème de la nouvelle parole et peuvent être ajoutées en position :

- pré-radical : c'est le cas, par exemple, du préfixe *ma*, dont la présence indique le nom de lieu, par exemple, *maktabah* 'bibliothèque' dérivant de la racine *ktb* dont le sens général est *écrire*; et *maqbarah* 'cimetière' dérivant de la racine *qbr* dont le sens général est *enterrer*.
- interne : par exemple, l'infixe *ta* dans les verbes *intaḥara* 'se suicider', dérivé de *nhr* dont le sens général est *égorger*, et *ištarā* 'acheter pour soi' dérivé de *šry* dont le sens général est *acheter*. Dichy ajoute que le préfixe /t/ a, dans ces verbes, le sens du « morphème-écho » ou « retour du procès sur le sujet ». ³
- post-radical : par exemple, le suffixe *-iyy* dont l'ajout permet la formation de l'adjectif relationnel, ainsi de *'insān* 'homme', on peut obtenir *'insāniyy* 'humain' et *'insāniyyah* 'humanité'.

Elles contribuent aussi à la flexion des paroles, dans ce cas elles ne font pas partie du schème mais elles servent comme :

- préfixes ou suffixes qui permettent de marquer :
 - le temps, le mode et la personne des verbes, par exemple : *katab-a* 'il a écrit', *ya-katab-u* 'il écrit' et *ya-ktub-ūna* 'ils écrivent' ;
 - le genre et le nombre des noms et des adjectifs, comme dans le cas de *mu'allim* 'enseignant' ; *mu'allim-ah* 'enseignante' ; *mu'allim-ūna* 'enseignants' et *mu'allim-āt* 'enseignantes'. Le suffixe *āt* permet aussi d'obtenir le pluriel des choses inanimées avec un sens « d'abondance » ; c'est le cas de *maktab-āt* 'lieux de nombreux livres ; bibliothèques' et *maqbar-āt* 'lieux de nombreuses tombes ; cimetières'.
- enclitiques, c'est le cas des pronoms sujets, par exemple, *katab=tu* 'j'ai écrit' et des compléments, par exemple *katab=tu=hu* 'je l'ai écrit'.

2.1.1.2 Racines tri-consonantiques et quadri-consonantiques

La racine n'existe pas réellement comme parole mais, elle résulte d'une construction abstraite due à la soustraction des voyelles et éventuellement des augments affiaux qui contribuent à la formation des mots. Dans la majeure partie des cas, elle résulte

3. Dichy, 2002.

constituée de trois consonnes mais, généralement, elle peut être constituée au minimum de deux et au maximum de cinq consonnes. La racine est dotée d'un « concept sémantique et lexical », c'est à dire que les unités lexicales qui partagent la même racine sont liées entre elles d'un « sens général » et, constituent, ce qu'on peut appeler, une « famille dérivationnelle » qui forme un réseau de significations.

Selon la grammaire arabe, la racine est décrite comme *tarkīb* 'une combinaison' de consonnes.⁴ Elle est communément représentée au moyen de la racine *f'l*, de laquelle dérive le verbe *fa'ala* 'faire'. Les consonnes *f*, *'* et *l* correspondent respectivement à la première, seconde et troisième consonne radicale. Si la racine comporte quatre consonnes, la représentation traditionnelle n'ajoute pas une 4^e consonne quelconque à la racine *f'l*, mais se contente de dupliquer la 3^e consonne radicale, ce qui donne *f'll*.⁵

Par commodité, on peut aussi présenter la racine comme $C_1C_2C_3$ (et éventuellement C_4), où C_1 , C_2 , C_3 et C_4 correspondent respectivement aux consonnes radicales.

2.1.2 Le schème (ou le gabarit)

Le schème est un « patron général » dans lequel les consonnes radicales, les lettres d'augment et le pattern vocalique (timbre et longueur) prennent une position bien précise. Les paroles qui partagent le même gabarit appartiennent à la même famille morphologique et partagent certains traits grammaticaux et sémantiques, par exemple :

- toutes les paroles de schème $maC_1C_2aC_3$ ou $maC_1C_2aC_3ah$ sont des noms de lieu où se réalise l'action de schème $C_1aC_2aC_3a$: *maktabah* 'bibliothèque' et *maktab* 'bureau' sont des noms de lieux liés à l'action *kataba* 'écrire' ;
- toutes les paroles de schème $iC_1taC_2aC_3a$ sont des verbes qui partagent le sens de « procès réalisé au profit, au détriment, à partir du sujet ou encore en lui »⁶; par exemple : *intaḥara* 'se suicider' et *iṣṭarā* 'acheter pour soi'.

2.2 Les différents types de dérivation

L'étude des schèmes, des racines et leurs relations avec la sémantique et la grammaire relèvent des sciences de *al-'iṣṭiqāq* 'la dérivation'. On distingue trois types d'études en fonction de la parenté existant entre les paroles dérivant de la même racine ou dérivant de racines différentes.

4. Larcher, 2008.

5. Dichy, 2002.

6. Dichy, 2002.

***al=’ištiqāq al=’akbar* ’la très grande dérivation’**

Elle met en relation la valeur sémantique des mots et les caractéristiques phonétiques des radicaux qui les constituent. Elle s’occupe des mots ayant en commun deux consonnes radicales et différant par une radicale. C’est la différence qu’on peut constater par exemple entre les deux verbes *kašaṭa* et *qašaṭa* qui ont le même sens de ‘gratter’ et, entre les verbes *na’aqa* et *nahaqa* qui dénotent la même action de ‘crier’, avec une différence subtile consistant dans la nature du sujet humain (une personne) pour *na’aqa* et non humain (un animal) pour *nahaqa*.

***al=’ištiqāq al=kabīr* ’la grande dérivation’**

Elle permet d’établir les parentés sémantiques qui peuvent exister entre les mots ayant les mêmes radicales, mais dans un ordre différent et donc, entre les différentes racines composées des mêmes consonnes mais dans un ordre différent. Les différentes compositions des radicaux de la même racine s’appellent *taqālib*. Le tableau 2.2 montre quelques exemples, étudiés par ibn ġinnī (m. 392/1002), qui démontre que les différentes racines composées des lettres ġ, b et r dénotent un sens de *force* et d’*intensité*.⁷

Racine	Lemme	Sens
ġbr	’aġbara	forcer ; obliger
brġ	burġ	citadelle ; forteresse
rġb	raġaba	glorifier, craindre
ġrb	ġarbā’	ciel
bġr	buġrī	impressionnant

Tab. 2.2 : Exemples de parenté entre différentes racines

***al=’ištiqāq aṣ=ṣaġīr* ’la petite dérivation’**

Elle étudie la parenté sémantique existant entre les mots ayant les mêmes consonnes radicales dans un ordre fixe, par exemple, entre les paroles *katb* ‘écrire’, *kataba* ‘il a écrit’, *kātib* ‘écrivain’ et *maktūb* ‘écrit, destin’. La petite dérivation s’occupe de la formation de nouvelles paroles à partir d’autres préexistantes, à condition qu’il y ait un ajout dans le sens primaire.⁸

7. Ibn Ģinnī, *’al-Ḥaṣā’iṣ*, V. II, pp. 134-136 et *’aṣ-Ṣāliḥ*, 2009, pp. 190-194.

8. *’al-’Astarabāqī, Ṣarḥ aṣ-ṣāfiyyat*, Vol.I, page 83.

La base de la dérivation est le *maṣḍar*, comme le souligne Larcher :

Le *maṣḍar* est donc choisi comme base de la dérivation, parce qu'il est la forme sémantiquement la moins marquée du paradigme. Il dénote le seul procès, alors que la forme verbale qui en est immédiatement dérivée dénote, en plus du procès, le temps. C'est bien un choix, comme le prouve le fait que les grammairiens de *kūfah* (l'école concurrente de celle de *Baṣ-rah*) faisaient le choix contraire, dérivant non pas le verbe du *maṣḍar*, mais le *maṣḍar* du verbe. Beaucoup d'arabisants sont coufiotes sans le savoir, considérant le *maṣḍar*, sans doute sous l'influence des noms d'action de nos langues, comme déverbaux. Quoi qu'il en soit, une seconde différence apparaît ici. La dérivation telle que l'entendent les arabisants est d'abord morphologique [...]; le *ištiqāq* des grammairiens arabes est indissolublement formel et sémantique.⁹

Le verbe dérive du *maṣḍar* auquel s'ajoute la notion du temps (accompli/inaccompli). Dans ce cas, la grammaire arabe parle de *fi'l muğarrad* 'verbe nu', qui peut être trilitère (ou triconsonantique), lorsqu'il est formé à partir d'une racine à trois consonnes, ou quadrilitère (ou quadriconsonantique) lorsque sa racine est composée de quatre consonnes. Du verbe nu, peuvent être dérivés :

- les formes nominales associées au verbe, comme le participe actif, le participe passif, le nom du lieu, le nom du temps, le nom d'instrument, l'adjectif assimilé et l'élatif. L'adjectif assimilé est au verbe d'état ce que le participe actif est au verbe d'action. L'élatif, par contre, est lié soit au verbe d'état, soit à une partie des verbes d'action ;¹⁰
- les verbes dérivés, appelés traditionnellement '*af'āl mazīdah* 'verbes augmentés', ajoutent, à la signification basilaire du verbe, une notion additionnelle comme la causativité et l'intensité.

L'étude morphosémantique et la représentation lexicale, selon la grammaire classique, sont basées sur la petite dérivation qu'on appellera la dérivation "tout court". Toutefois, s'occupant de lister les paroles dérivant d'une même racine caractérisée par un ordre fixe de ses radicaux, la dérivation est un mode de représentation imparfait qui est incapable d'expliquer plusieurs aspects et de saisir les relations sous-jacentes qui peuvent

9. Larcher, 2008.

10. Larcher, 2008 et 'aṣ-Ṣāliḥ, 2009, p. 181.

exister entre schèmes différents ou entre racines différentes.¹¹ Différentes recherches contemporaines ont essayé de trouver des théories plus adéquates.¹²

2.3 Théorie des Matrices et Étymons

Le tableau 2.3 présente des exemples de paroles dérivant de racines constituées des mêmes consonnes radicales dans un ordre différent ou constituées par deux radicales identiques et différentes par la nature d'un radical. Malgré ceci, on note qu'elles dénotent toutes le même sens de *couper*.

Racine	Verbe	Sens
<i>btt</i>	<i>batta</i>	couper, retrancher en coupant
<i>tbb</i>	<i>tabba</i>	couper, retrancher en coupant
<i>btr</i>	<i>batara</i>	couper la queue à un animal
<i>btl</i>	<i>batala</i>	couper, retrancher
<i>btk</i>	<i>bataka</i>	couper, retrancher en coupant
<i>sbt</i>	<i>sabata</i>	couper, retrancher en coupant, raser

Tab. 2.3 : Exemples de verbes dérivant de racines différents

Ces exemples illustrent bien que la théorie de la racine trilitère, sur laquelle se basent les études classiques, n'est pas suffisante pour capturer ce genre de régularités. Comme le souligne Fournet :

... la doxa officielle de l'arabe prône une théorie de la racine à trois consonnes. En pratique, cette doxa empêche de percevoir la structure réelle du lexique, qui se trouve atomisé en une myriade de racines de même sens mais sans lien supposé.¹³

Bohas propose comme forme de base, de la morphologie et du lexique arabe, une unité constituée d'une combinaison non ordonnée de deux consonnes qu'il appelle étymon.¹⁴ Ainsi, l'étymon, qui représente les exemples illustrés dans le tableau précédant (2.3), est {b, t}. Au contraire de la racine trilitère dont l'ordre des consonnes est fixe, l'étymon est caractérisé de *réversibilité*. Bien que l'ordre des consonnes est inverse, les verbes

11. Boudelaa & Marslen-Wilson, 2001.

12. McCarthy, 1981; McCarthy, 1991; Bohas, 1991; Bohas & Dat, 2007.

13. Fournet, 2007.

14. Bohas & Dat, 2007

batta et *tabba* du tableau 2.3 font partie de la même famille caractérisée par l'étymon {b, t}, dont la valeur sémantique est *couper* et l'ordre des deux consonnes n'est pas important. Pour satisfaire les exigences de la morphologie arabe basée sur des schèmes à trois places consonantiques, ces bases biconsonantiques peuvent subir des processus d'élargissement :

i) par diffusion de la deuxième consonne : c'est le cas de *t* pour *batta* et de *b* pour *tabba*;

ii) par incrémentation, avec ou sans valeur sémantique ajoutée : c'est le cas de *sabata*, *batara*, *batala* et *bataka*, où on note l'ajout d'une consonne en position initiale, médiane ou finale, par rapport aux deux consonnes de l'étymon {b, t}. Les ajouts peuvent alors être des sonantes, des labiales, des gutturales ou des glides (*wāw* ou *yā'*).

2.4 Théorie autosegmentale et représentations multilinéaires

Un autre aspect important est illustré par les exemples du tableau 2.4. L'étude traditionnelle assume trois schèmes différents pour la voix active et trois autres schèmes pour la voix passive. Cependant, la comparaison, entre tous ces schèmes, met en évidence qu'ils présentent tous la même alternance vocalique a-a (ou ā/a) pour l'actif à l'opposé de u/i (ou ū/i) pour le passif.

voix active	exemple	voix passive	exemple
C ₁ aC ₂ aC ₃ -a	<i>naqala</i> 'il a déplacé'	C ₁ uC ₂ iC ₃ -a	<i>nuqila</i> 'il a été déplacé'
C ₁ āC ₂ aC ₃ -a	<i>hāğama</i> 'il a attaqué'	C ₁ ūC ₂ iC ₃ -a	<i>hūğima</i> 'il a été attaqué'
staC ₁ C ₂ aC ₃ -a	<i>stağrağa</i> 'il a extrait'	stuC ₁ C ₂ iC ₃ -a	<i>stuğriğa</i> 'il a été extrait'

Tab. 2.4 : Exemples de verbes de gabarits différents

Pour expliquer la structure des racines dans les langues sémitiques, en particulier, et la morphologie non concaténative, en générale, McCarthy propose la « théorie

autosegmentale multilinéaire », Dans le cadre de la théorie de la phonologie autosegmentale.¹⁵

Selon McCarthy, dans un système comme l'arabe, les voyelles et les consonnes apportent des informations bien distinctes à la signification de la parole. De conséquence, elles doivent être représentées séparément sur des plans différents.

En termes de la théorie autosegmentale, les consonnes de la racine et les voyelles constituent les *autosegments*. Chaque autosegment prend, dans le schème, une position précise, appelée *unité porteuse d'autosegment*. Inspiré des travaux de McCarthy, la figure 2.1 illustre un exemple de la représentation multilinéaire du verbe trilitère.

Exemple a: verbe *kataba*

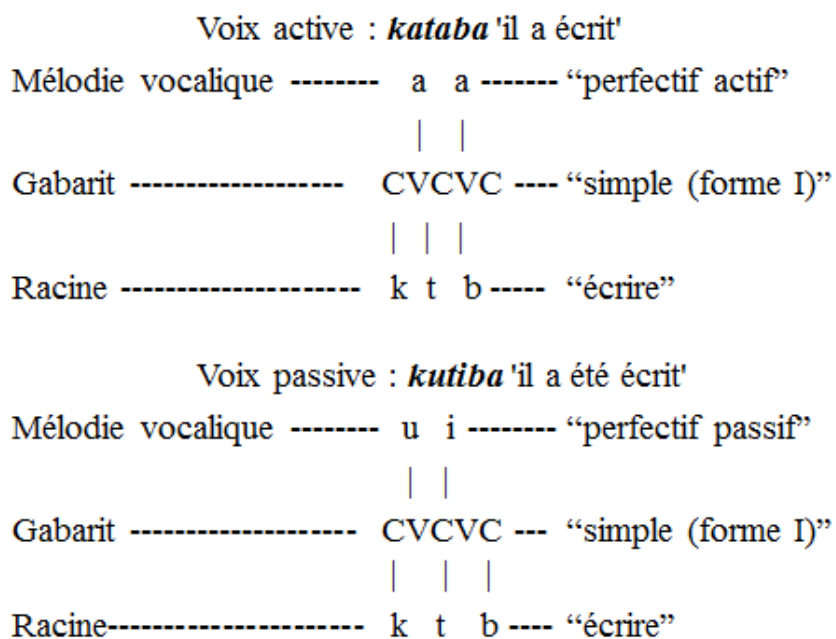


Fig. 2.1 : Exemple de représentation multilinéaire : verbe trilitère

La représentation multilinéaire se fait sur trois plans :

- la «mélodie vocalique » où on trouve les voyelles qui contribuent avec les informations morphosyntaxiques, par exemple, /a-a/ est la mélodie vocalique de l'accompli actif et /u-i/ de l'accompli passif ;

15. McCarthy, 1981; McCarthy, 1989; McCarthy, 1991

- le schème, appelé par McCarthy *squelette*, est le plan qui impose une structure syllabique et une prosodie fixe ; il détermine le nombre et l'organisation des consonnes et des voyelles pour former, dans ce cas, le verbe simple *kataba*;
- la «racine », porteuse du contenu sémantique, est constituée des consonnes radicales.

L'association des autosegments aux unités correspondantes suit des conventions, formalisées par Goldsmith,¹⁶ qu'on peut résumer comme suite :

- les autosegments de la racine sont associés aux positions C et les autosegments de la mélodie vocalique sont associés aux positions V ;
- les autosegments sont associés aux unités correspondantes sur le schème : i) de gauche à droite ; ii) un à un ;
- si le nombre des autosegments est inférieur au nombre d'unité, il y a propagation de l'autosegment le plus à droite sur les positions non encore pourvues ;
- les lignes d'association ne peuvent pas se croiser.

C'est toujours le schème qui prend en charge les variations permettant la génération des verbes dérivés, comme la gémination médiane, dans l'exemple b de la figure 2.2 ou l'allongement vocalique, dans exemple c de la figure 2.2.

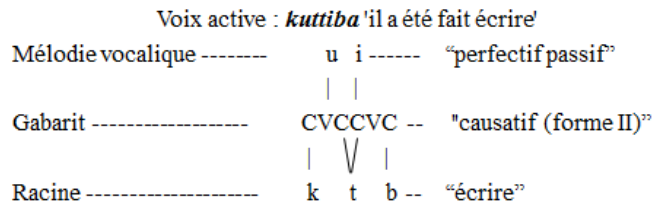
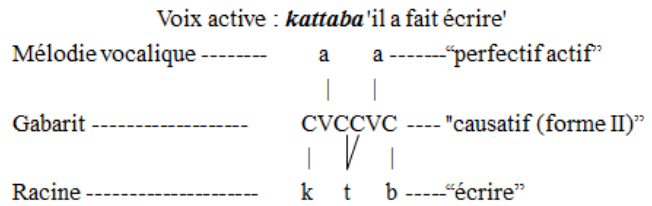
Selon McCarthy, à un certain niveau de la représentation, la racine, la mélodie vocalique et le gabarit représentent chacun un objet phonologique doué d'un sens et, pour ce fait, il propose de les considérer des morphèmes distincts. Une note très importante qu'il a ajoutée, en faveur de l'organisation multilinéaire à morphèmes distincts, est le fait que, de cette manière, le lexique de la langue arabe s'organise autour de la racine :

What is perhaps the most compelling argument for this characterization of the morpheme is the basic organization of the Arabic (and Semitic) lexicon around the consonantal root.¹⁷

16. Goldsmith, 1976

17. McCarthy, 1981, p. 377.

Exemple b: verbe *kattaba*



Exemple c: verbe *kātaba* 'il a écrit à qn'

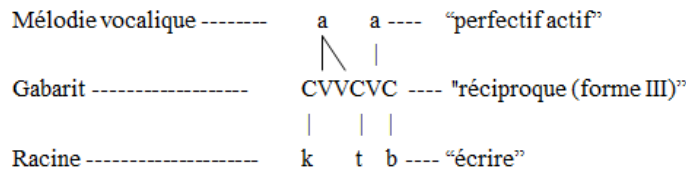


Fig. 2.2 : Exemples de représentations multilinéaires : verbes dérivés

2.5 Restrictions des cooccurrences

2.5.1 Introduction

Le grammairien arabe ibn ġinnī (m. 392/1002) fait noter que la combinaison des consonnes n'est pas aléatoire et qu'elle est régie par des contraintes dépendant de la nature, du point d'articulation et de la position des consonnes dans la racine.¹⁸ Il a noté que la cooccurrence de certaines consonnes homorganiques dans la même racine est impossible, c'est le cas des consonnes dorsales {*k*, *ġ*, *q*}. Par contre, dans d'autres cas, une racine affiche deux consonnes homorganiques, par exemple, les consonnes gutturales {*ħ*, *ġ*, *h*, *ʿ*, *ʾ*, *h*} mais, un des deux phénomènes suivants se produit :

- les deux consonnes ne sont pas adjacentes et occupent la première et la troisième position à l'intérieur de la racine ; par exemple la racine *hdʾ* de laquelle dérive *hada`a* 'se calmer' ;

18. Ibn Ġinnī, *sirr*, pp. 811-820 et *Suyūṭī, Muzhir*, V. I; pp. 240-247.

- les deux consonnes sont adjacentes mais on note, alors, un gradient de «force» qui impose la précedence de l'une sur l'autre. C'est le cas de la *hamza* qui précède toujours l'une des trois consonnes *h*, *ḥ* ou *h*, par exemple : la racine *'hl* de laquelle dérive la parole *'ahl* 'personne avec qui on a un lien de parenté'.

Il a noté aussi que les consonnes identiques, bien qu'elles soient homorganiques, peuvent coexister dans la même racine mais toujours en deuxième et troisième positions, par exemple, le verbe *madda* 'il a allongé' de la racine *mdd*. En langue arabe, seules quatre paroles présentent des consonnes identiques en première et deuxième position : *dadān* 'jeu, amusement'; *babbān* 'chose unique'; *babbah* 'gros' et *babr* 'léopard, panthère d'Afrique'.

Dans les études récentes, la formalisation des restrictions de cooccurrence des consonnes dans les racines arabes a profité de la notation de la phonologie autosegmentale, où les objets phonologiques sont séparés et représentés sur des niveaux autosegmentaux différents. Les niveaux autonomes et distincts, des consonnes et des voyelles, a facilité l'étude de leurs distributions et de leurs représentations, définissant ainsi les contraintes qui régissent la bonne formation des représentations phonologiques.

2.5.2 Le Principe du Contour Obligatoire

En 1973, Leben a proposé que les séquences d'autosegments soient soumises à une condition de bonne formation, appelée le **Principe du Contour Obligatoire**, dorénavant indiqué comme OCP (**O**bligatory **C**ontour **P**rinciple).¹⁹ Généralement, l'OCP est une règle de la Grammaire Universelle qui prohibe que les éléments identiques soient adjacents et, stipule que les consonnes (ou les voyelles) qui occupent deux positions voisines doivent être différentes :

OCP_{total} pour toute paire d'autosegments adjacents a et b ==> **a # b** où le symbole # veut dire différent.

Dans le cas des consonnes identiques considérées homorganiques ayant une **similarité totale**, l'OCP est dit **total**.

19. Voir entre autres McCarthy, 1981, p. 383; McCarthy, 1989, p. 71; Pierrehumbert, 1993 et Frisch et al., 2004, p. 213.

2.5.2.1 Application de l'OCP dans la mélodie consonantique

2.5.2.1.1 Cas des racines avec gémiation Selon l'OCP, qui stipule que deux autosegments adjacents doivent être différents, les racines comme *mdd* ne devraient pas exister, parce qu'elles contiennent deux autosegments C₂ et C₃ identiques et, de conséquence, elles violent le principe de l'OCP.

Selon les conventions proposées par Goldsmith, et discutées dans la section précédente, les autosegments s'associent aux unités porteuses du schème, selon les principes qui stipulent que le remplissage se fait **un à un**, en suivant la direction **gauche-droite** et, dans le cas où **il n'y a plus** de matériel pour remplir toutes les unités porteuses, il y a **propagation du dernier autosegment**.²⁰

De conséquence, la racine d'un verbe comme *madda* ne peut être que bilitère (ou biconsonantique), *md*, avec propagation du dernier autosegment *d* vers la droite, comme illustré dans la figure 2.3.

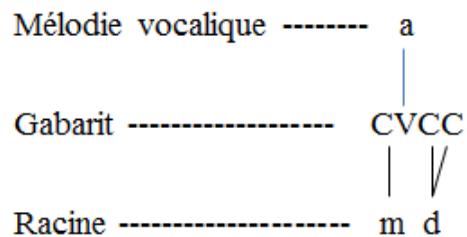


Fig. 2.3 : Représentation multilinéaire du verbe *madd-a*

Les mêmes conventions permettent, aussi, d'expliquer la non existence de racine de type C₁C₁R où R peut être n'importe quelle consonne différente de C₁. Selon les principes de propagation, la consonne C₁ ne peut se propager sur la position médiane alors qu'il y a **encore** un autosegment R qui doit être associé.

Cependant, les mêmes conventions se trouvent en difficulté pour expliquer les verbes avec gémiation médiane, comme le verbe causatif/intensif C₁VC₂C₂VC₃, où C₂ ne peut se propager, vu qu'il y a encore l'autosegment C₃ à associer.

Lowenstamm fait noter que la gémiation de C₂ est un phénomène dérivationnel. Il propose, dans ce cas, que le matériel dérivationnel doit constituer une tête autonome et responsable des phénomènes de dérivation.²¹ Ainsi, pour expliquer la morphologie arabe, il propose un schème général qui contient deux sites morphologiques qu'il appelle [DS **CV**] (DS : Derivational Syllable), et peut être présenté comme suite :

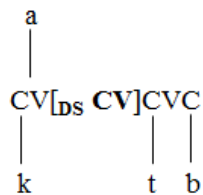
20. Lowenstamm et Prunet, 1986, p. 184.

21. Lowenstamm, 2003

$[_{DS} CV]CV[_{DS} CV]CVCV$

Un site $[_{DS} CV]$, localisé au début du schème, contrôle la génération d'un préfixe dérivationnel. Le deuxième $[_{DS} CV]$, localisé au milieu du schème, contrôle la génération d'un infixe dérivationnel. La génération d'une parole se fait en deux phases distinctes, comme expliqué dans la figure 2.4.

PREMIÈRE PHASE



DEUXIÈME PHASE

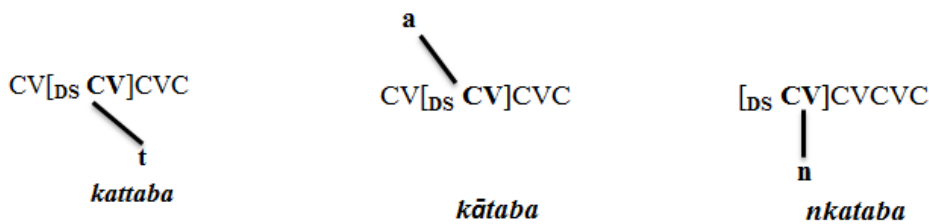


Fig. 2.4 : phases de formation des verbes dérivés

1. Dans la première phase, l'association des consonnes de la racine et de la mélodie vocalique aux positions C et V du schème. Cette phase suit les conventions d'associations de Goldsmith (1976), citées ci-dessus. Dans le cas du verbe simple, le phénomène se ferme à ce stade et, aucun site dérivationnel n'est activé. Mais, dans le cas d'un verbe dérivé, l'une des deux têtes est activée et donne naissance à la deuxième phase.
2. Selon le schème du verbe dérivé, il y a identification de la syllabe dérivationnelle désignée comme $[_{DS} CV]$. La directionnalité du processus, dans cette phase, est imposée par la tête dérivationnelle : il y a propagation de la deuxième consonne radicale *t*, pour la formation du verbe *kattaba* 'il a fait écrire' ou de la voyelle *a*, pour former le verbe *kātaba* 'Il a écrit à'. Par contre il y a ajout d'un nouveau matériel dérivationnel, *n*, pour former le verbe *nkataba* 'il a été écrit'.

2.5.2.1.2 Cas des consonnes homorganiques D'autres recherches ultérieures ont étendu les résultats obtenus au cas des consonnes homorganiques.²² Un autre Principe du Contour Obligatoire, appelé OCP-Place, a été introduit et contrôle le lieu de l'articulation. En d'autres termes, il permet de contrôler la cooccurrence des consonnes en fonction de leurs caractères phonologiques et défavorise la combinaison des consonnes homorganiques dans la même racine.

Ainsi, selon l'OCP-Place, la cooccurrence des consonnes, appartenant à la même classe naturelle, est interdite. Pour appliquer ces restrictions, les consonnes arabes sont divisées en classes naturelles,²³ illustrées dans le tableau 2.5.

Classe naturelle	Consonnes
labiales	b - f - m
coronales obstruantes	ṭ - ḍ - t - d - ṭ - ḍ - s - z - ṣ - ẓ - š
dorsales	k - ġ - q - ḥ - ğ
gutturales	ḥ - ġ - ḥ - ' - h - '
coronales sonantes	l - r - n

Tab. 2.5 : Classes naturelles des consonnes arabes

Cependant, Janet Pierrehumbert fait noter que les racines contenant des consonnes homorganiques peuvent exister mais, le pourcentage des cas où les consonnes en question sont adjacentes est inférieur à celui où elles ne le sont pas.²⁴ Elle propose que l'OCP-Place concerne les consonnes qui sont perçues comme étant similaires, et que la force de l'effet augmente avec la similarité perçue.

L'indice de similarité est variable selon le mode d'articulation (ou l'ensemble des propriétés phonétiques) des consonnes.²⁵ À titre d'exemple, le tableau 2.6 indique les propriétés phonologiques et l'indice de similarité entre les consonnes labiales {b, f, m}.²⁶

On note que l'indice de similarité est, évidemment, maximal (=1.00) dans le cas d'une paire de consonnes identiques, par exemple, {b, b}.

La similitude entre /b/ et /f/ est plus grande qu'entre /b/ et /m/. De conséquence, l'indice de similitude de la paire {b, f} est égale à 0.67 ; tandis que celui de la paire {b, m} est 0.33 ; et on peut faire le même raisonnement concernant la paire {f, m}.

22. McCarthy, 1986; McCarthy, 1989; Pierrehumbert, 1993; Frisch et al., 2004.

23. Frisch et al., 2004 et Frisch et al., 1997.

24. Pierrehumbert, 1993.

25. Pour une étude approfondie, voir entre autres Frisch et al., 2004; Pierrehumbert, 1993

26. Les valeurs sont extraites de Pierrehumbert, 1993.

Propriétés phonologiques			
Propriétés	b	f	m
labiale	+	+	+
obstruante	+	+	-
continue	-	+	-
sourde	-	+	-
voisée	+	-	+
occlusive	+	-	+
sonante	-	-	+
nasale	-	-	+

Indice de similarité entre	b	f	m
b	1.00	0.67	0.33
f		1.00	0.33
m			1.00

Tab. 2.6 : Propriétés phonologiques et indice de similarité entre les consonnes {b, f, m}

Sachant que l'effet de l'OCP-Place se manifeste lorsque l'indice de similarité est au dessus de 0.33 et plus l'indice de similarité augmente, plus l'effet de la contrainte est plus fort, on peut tirer plusieurs conclusions :

1. Avec une similarité totale (indice = 1.00), l'OCP-Place interdit la cooccurrence des consonnes identiques. La contrainte de l'OCP-Place subsume celle de l'OCP total, autrement dit, l'OCP total est le cas maximum de l'application de l'OCP-Place.
2. Toutes les labiales sont soumises à un effet OCP et ne peuvent se combiner entre elles. Mais, l'effet est plus fort pour les obstruantes {b, f}, dont l'indice de similarité est 0.67, que pour les paires {b, m} et {f, m} dont l'indice est 0,33. En effet, il n'y a pas de racines contenant une paire de consonnes labiales en position adjacente. Par contre, il y a 17 racines qui les contiennent en position non adjacente, dont 16 contiennent {m} combinée avec une labiale obstruante {b} ou {f} et une seule racine avec les deux obstruantes {b, f}.²⁷
3. La contrainte de l'OCP-Place n'est pas catégorique comme l'OCP totale, mais elle a un effet graduel en fonction de l'indice de similarité entre les consonnes homorganiques.

Dans d'autres expériences, on a pu calculer l'indice de cooccurrence, O/E, obtenu à partir du nombre de paires de consonnes existant réellement dans un dictionnaire de l'arabe (O) en fonction du nombre présumé (E) dans le cas où les consonnes homorganiques se seraient combinées de manière arbitraire. Une valeur O/E inférieure à 1 indique qu'il y a moins de racines observées que celles présumées et dans ce cas, on

27. Frisch et al., 2004

peut conclure qu'il existe une restriction de cooccurrence affectant les consonnes de la paire étudiée. Si O/E est supérieur à 1, le nombre de paires observées est supérieur au nombre présumé. Si O/E est égal à 1, le nombre de paires observées est exactement ce qu'on s'attend par cooccurrence aléatoire.

Dans d'autres expériences, on a calculé l'indice de cooccurrence des consonnes identiques et des consonnes homorganiques non identiques en position adjacente (C_1C_2 ou C_2C_3) et non adjacente (C_1C_3).²⁸ Leurs données sont basées sur les racines contenues dans le dictionnaire bilingue *A Dictionary of Modern Written Arabic* écrit par Wehr.²⁹ Les résultats de l'expérience sont illustrés dans le tableau 2.7.

Consonnes	adjacentes		non adjacentes	
	identiques	homorganiques non-identiques	identiques	homorganiques non-identiques
Labiales	0.00	0.00	0.05	0.42
Coronal Obstruantes	0.03	0.33	0.46	0.70
Vélaires	0.00	0.05	0.24	0.36
Gutturales	0.00	0.08	0.05	0.43
Coronal Sonantes	0.00	0.09	0.08	0.94

Tab. 2.7 : Indice de cooccurrence des consonnes arabes

L'analyse de ces résultats fait noter que l'indice de cooccurrence, en positions adjacentes, des consonnes identiques est presque toujours nul, en plus, il est inférieur à celui des consonnes homorganiques non identiques. Ceci confirme que les restrictions sur la cooccurrence des consonnes identiques sont plus fortes que celles qui s'appliquent sur les consonnes homorganiques non identiques.

On note aussi qu'il y a un effet de distance, dans le sens que, concernant une même paire de consonnes, l'indice est plus petit en position adjacente et donc les restrictions sont plus fortes dans le cas où les consonnes similaires occupent des positions adjacentes (C_1C_2 ou C_2C_3).

En comparaison, l'indice de la même paire, en positions non adjacentes, est plus grand, de conséquence, le pourcentage des racines contenant des consonnes similaires en position non adjacente (C_1C_3) est plus grand.

28. Frisch et al., 1997.

29. Wehr, 1971

2.5.2.2 Application de l'OCP dans la mélodie vocalique

Des études concernant la mélodie vocalique permettent de constater que les contraintes de cooccurrence s'étendent aussi aux voyelles.³⁰

Par exemple, le tableau 2.8 illustre la fréquence des verbes en fonction de la mélodie vocalique à partir d'une analyse faite par McCarthy³¹ sur les racines présentes dans le dictionnaire «*A Dictionary of Modern Written Arabic*».³²

On note que la première voyelle est toujours basse, *a*, suivie d'une autre voyelle *a* ou d'une voyelle haute *i* ou *u*.

Dans toute la langue arabe, on ne trouve jamais une mélodie contenant, en positions adjacentes, les deux voyelles hautes *i* et *u*.

Mélodie		Fréquence	Exemples		Traduction
Accompli	Inaccompli		Accompli	Inaccompli	
a-a	a-u	1029	<i>katab-a</i>	<i>y-aktub-u</i>	'écrire'
	a-i	842	<i>ḍarab-a</i>	<i>y-aḍrib-u</i>	'frapper'
	a-a	436	<i>fa'a-la</i>	<i>y-af'al-u</i>	'faire'
a-i	a-a	518	<i>šarib-a</i>	<i>y-ašrab-u</i>	'boire'
a-u	a-u	191	<i>balud-a</i>	<i>y-ablud-u</i>	'être stupide'

Tab. 2.8 : Fréquence des verbes trilitères

Cependant, l'accompli passif est le seul schème qui enfreint la contrainte de la similarité entre les voyelles hautes avec une mélodie contenant *u* suivie de *i*. En effet, pour tous les verbes :

- le schème de l'accompli passif est $C_1uC_2iC_3-a$, comme *kutib-a* 'il a été écrit' ; *ḍurib-a* 'il a été frappé' et *šurib-a* 'il a été bu' ;
- le schème de l'inaccompli passif est $y-uC_1C_2aC_1-u$, comme *y-uktab-u* ; *y-uḍrab-u* et *y-ušrab-u*.

Un autre aspect important qu'on note dans le tableau 2.8 est l'alternance de la deuxième voyelle de la mélodie. Ce phénomène apophonique, ou ablaut, véhicule l'opposition aspectuelle entre l'accompli et l'inaccompli et divise le verbe arabe en classes morphosyntaxiques qui seront étudiées dans le chapitre suivant.

30. Lahrouchi, 2011.

31. McCarthy, 1994, page 207.

32. Wehr, 1971.

Chapitre 3

Le système verbal

3.1 Étude morphologique du verbe arabe

3.1.1 Classification du verbe arabe

En fonction du nombre des lettres qui le constituent

Dans la grammaire arabe, on parle de *fi'l muğarrad* 'verbe nu', lorsque le verbe est constitué seulement des consonnes de la racine. La grammaire arabisante occidentale lui associe parfois l'appellation de verbe «primitif » ou «de base » et le désigne, aussi, comme verbe de «forme I ». Le verbe *muğarrad* peut être trilitère, lorsqu'il est formé d'une racine à trois consonnes ou quadrilitère, lorsque sa racine est composée de quatre consonnes.

En ajoutant des *hurūf az=ziyyādah* 'lettres d'augment' au verbe de base, on obtient d'autres verbes dérivés, qu'on appelle, traditionnellement, '*af'āl mazīdah* 'verbes augmentés'.

En fonction de la nature des lettres qui le constituent

Prenant en considération, seulement, la nature des consonnes radicales, la grammaire arabe divise le verbe en deux grandes classes.

Le verbe est dit *mu'tall* 'malade', dans le cas où l'une des consonnes radicales est une lettre déficiente, la *wāw* ou la *yā*'.

Dans le cas contraire, le verbe est *ṣaḥīḥ* 'sain' et peut être *muḍa* 'af'géméné', dans le cas où la deuxième et la troisième consonne radicales sont identiques ; *mahmūz* lorsqu'une de ses radicales est une *hamza*; ou *sālim* 'régulier' dans le cas où il ne contient pas la *hamza* et il n'est pas géméné.

Les temps verbaux

Le verbe arabe a trois temps en fonction du moment de l'énonciation :¹

- *al=fi' l al=māḍī*: lorsqu'il exprime une situation ou une action qui a eu lieu avant le moment de l'énonciation, et donc il exprime le temps « passé »;
- *al=fi' l al= muḍāri*: exprimant le temps présent ou future et donc le temps « non passé »;
- *al=fi' l al='amr*: exprimant l'ordre de faire quelque chose dans le future ; le verbe est impératif.

Les formes du verbe arabe sont désignées en littérature occidentale par plusieurs noms : « accompli » et « inaccompli », « achevé » et « inachevé », « perfectif » et « imperfectif ». Plusieurs auteurs adoptent une autre appellation basée sur les traits morphologiques du verbe : « verbe suffixé » pour l'accompli dans lequel les affixes flexionnels suivent la base verbale et « verbe affixé » pour l'inaccompli parce que, en plus des suffixes, il est caractérisé par la présence d'un préfixe qui précède la base. Toutefois, comme le fait noter Dichy :

Les termes d'accompli et inaccompli, d'achevé et d'inachevé ou de perfectif et imperfectif, pourtant reçus, comportent un problème, celui de désigner la forme morphologique par un terme dénotant l'une des valeurs aspectuo-temporelles que cette forme peut prendre en contexte. D'où le choix de désigner ces paradigmes verbaux par leur nom arabe [...]²

Dans cette partie de notre étude, on suivra l'appellation de Dichy, en désignant les paradigmes verbaux par leur nom arabe *māḍī* 'accompli', *muḍāri* 'inaccompli'.

La flexion verbale

Elle se fait en fonction de :

- la personne : la première (1), la deuxième (2) et la troisième (3) ;
- le genre : masculin (M) et féminin (F) ;
- le nombre : singulier (S), duel (D) et pluriel (P).

1. Grand'henry, 2000, pp 28-29.

2. Dichy- 2007, note 4 page 315.

3.1.2 Analyse morphologique du verbe *māḍī* (accompli)

3.1.2.1 Conjugaison du verbe *māḍī*

Le tableau 3.1 illustre les formes fléchies à la 2MS et à la 3MS au *māḍī* de quelques exemples de verbes de base et dérivés.

racine	3MS	2MS	base verbale ou «stem »
<i>ktb</i>	<i>katab-a</i> 'il a écrit'	<i>katab-ta</i> 'tu as écrit'	<i>katab</i>
	<i>istaktab-a</i> 'il a fait écrire pour soi'	<i>istaktab-ta</i> 'tu as fait écrire pour toi'	<i>istaktab</i>
<i>fʿl</i>	<i>faʿal-a</i> 'il a fait'	<i>faʿal-ta</i> 'tu as fait'	<i>faʿal</i>
	<i>infaʿal-a</i> 'il s'est ému'	<i>infaʿal-ta</i> 'tu t'es ému'	<i>infaʿal</i>
	<i>iftaʿal-a</i> 'il a fait semblant'	<i>iftaʿal-ta</i> 'tu as fait semblant'	<i>iftaʿal</i>
<i>trğm</i>	<i>tarğam-a</i> 'il a traduit'	<i>tarğam-ta</i> 'tu as traduit'	<i>tarğam</i>

Tab. 3.1 : Analyse de formes verbales *māḍī*

L'analyse de ces exemples révèle l'existence d'une base verbale (appelé aussi «stem »), commune à toutes les formes fléchies du même verbe. Dans le cas du verbe primitif, la base verbale est formée des consonnes de la racine et de leurs voyelles, par exemple *katab* et *faʿal*. Dans le cas du verbe dérivé, elle contient aussi les voyelles et les consonnes ajoutées par le schème du verbe dérivé, par exemple *istaktab* et *iftaʿal*. La flexion au *māḍī* est la même pour tous les verbes arabes et se fait en ajoutant, à la base verbale, les mêmes «désinences » qui sont variables et véhiculent des informations concernant les traits morphosyntaxiques : personne, genre et nombre. Dans l'exemple donné, la «désinence » de la 3MS est *a* et celle de la 2MS est *ta*. Toutefois, elles sont identiques pour les formes de tous les verbes conjugués à la même personne. Pour une étude morphologique, on utilise, traditionnellement, la table de conjugaison du verbe *faʿala* "faire". La base verbale, dans ce cas, est *faʿal*, à laquelle s'attachent

les « désinences » qui varient en fonction de la personne ; du genre et du nombre et sont rapportées en caractères rouges dans le tableau 3.2.

Mode	1	2M	2F	3M	3F
S	<i>fa'al-tu</i>	<i>fa'al-ta</i>	<i>fa'al-ti</i>	<i>fa'ala</i>	<i>fa'alat</i>
D	<i>fa'al-nā</i>	<i>fa'al-tumā</i>	<i>fa'al-tumā</i>	<i>fa'alā</i>	<i>fa'al-tā</i>
P	<i>fa'al-nā</i>	<i>fa'al-tum</i>	<i>fa'al-tunna</i>	<i>fa'al-ū</i>	<i>fa'al-na</i>

Tab. 3.2 : Conjugaison au *māḍī* du verbe *fa'ala*

Cependant, certaines formes fléchies ne sont pas discriminées en fonction du genre, notamment, *fa'al-tu* (1S), *fa'al-nā* (1P) et *fa'al-tumā* (2D) sont les mêmes pour le féminin et le masculin. On note aussi un effacement du duel de la 1^{ère} personne qui a seulement deux formes, *fa'al-tu* (1S) et *fa'al-nā* (1P).

3.1.2.2 Les désinences flexionnelles du *māḍī*

3.1.2.2.1 Analyse morphologique Elles sont caractérisées surtout par le fait qu'elles sont très sélectives. Elles s'attachent uniquement à la catégorie des verbes et font partie intégrante avec la base verbale pour former le verbe fléchi.

Une comparaison des désinences flexionnelles du *māḍī* avec les pronoms nominatifs indépendants, dans le tableau 3.3, met en évidence une relation morphophonologique entre plusieurs de ces désinences et les pronoms équivalents.

Du fait, on note que les pronoms indépendants de la 2^e personne sont formés par les désinences du *māḍī* précédées d'un support 'an.

Par contre, bien que les désinences de la 1^{ère} et celles de la 3^e personne ne semblent pas être directement liées aux pronoms équivalents, on note un rapprochement sur la base de leurs constitutions morphophonologiques.

3.1.2.2.2 Analyse morphosyntaxique Une analyse morphosyntaxique permet de comprendre, encore mieux, la nature des désinences du *māḍī*.

En langue arabe, le syntagme nominale (SN) peut être postverbal (ordre VSO) ou préverbal (ordre SVO).

3.1.2.2.2.1 Cas du SN postverbal En analysant les exemples 2, 3, 4 et 5, on note que l'accord est partiel.

La désinence verbale exprime seulement le genre et la personne en excluant le nombre, et elle est toujours la même quelque soit le nombre :

Personne	Numéro	Genre	désinences du <i>māḍī</i>	Pronoms indépendants
1	S	M/F	- <i>tu</i>	' <i>anā</i>
	D/P		- <i>nā</i>	<i>naḥnu</i>
2	S	M	- <i>ta</i>	' <i>anta</i>
		F	- <i>ti</i>	' <i>anti</i>
	D	M/F	- <i>tumā</i>	' <i>antumā</i>
	P	M	- <i>tum</i>	' <i>antum</i>
		F	- <i>tunna</i>	' <i>antunna</i>
3	S	M	- <i>a</i>	<i>huwa</i>
		F	- <i>at</i>	<i>hiya</i>
	D	M	- <i>ā</i>	<i>humā</i>
		F	- <i>tā</i>	<i>humā</i>
	P	M	- <i>ū</i>	<i>hum</i>
		F	- <i>na</i>	<i>hunna</i>

Tab. 3.3 : Comparaison morphèmes et pronoms indépendants

- *a*, dans le cas où le SN est masculin, dans l'exemple 2, où il est singulier, et dans l'exemple 4, où il est pluriel ;

- *at*, dans le cas où le sujet est féminin, singulier dans l'exemple 3 ou pluriel dans l'exemple 5.

(2)

katab-a *al=walad-u*
 écrire.PFV-3MS DEF=garçon-NOM
 'le garçon a écrit'

(3)

katab-at *al=bint-u*
 écrire.PFV-3FS DEF=filles-NOM
 'la fille a écrit'

(4)

katab-a *al='awlād-u*
 écrire.PFV-3MS DEF=garçons-NOM
 'les garçons ont écrit'

(5)

katab-at *al=banāt-u*
 écrire.PFV-3FS DEF=filles-NOM
 'les filles ont écrit'

3.1.2.2.2 Cas du SN préverbal (ordre SVO) L'accord est complet en genre, personne et nombre et les désinences du singulier sont différentes de celles du pluriel.

Concernant le singulier, les désinences sont *a* et *at*, pour exprimer respectivement la 3MS et la 3FS, dans les exemples 6 et 7.

(6)

al=walad-u *katab-a*
 DEF=garçon-NOM écrire.PFV-3MS
 'le garçon a écrit'

(7)

al=bint-u *katab-at*
 DEF=fille-NOM écrire.PFV-3FS
 'la fille a écrit'

Par contre, pour marquer respectivement la 3MP et la 3FP, dans les exemples 8a, 9a, on utilise les désinences *ū* et *na*.

Le fait que les phrases sont agrammaticales,³ dans les exemples 8b, 9b, prouve que les désinences, *ū* et *na*, se réalisent seulement dans le cas où le SN est pré-verbal (ordre SVO).

En comparaison, la grammaticalité des exemples 2, 3, 4 et 5 prouve que les désinences *a* et *at* n'ont pas le même comportement que les autres désinences.

(8)

a <i>al='awlād-u</i> <i>katab-ū</i> DEF=garçons-NOM écrire.PFV-3MP 'les garçons ont écrit'	b * <i>katab-ū</i> <i>al='awlād-u</i> * écrire.PFV-3MP DEF=garçons-NOM
---	--

(9)

a <i>al=banāt-u</i> <i>katab-na</i> DEF=filles-NOM écrire.PFV-3FP 'les filles ont écrit'	b * <i>katab-na</i> <i>al=banāt-u</i> *écrire.PFV-3FP DEF=filles-NOM
---	--

Il est intéressant de noter qu'en dialecte marocain, la flexion verbale est plus simple. Dans les exemples 10a et 10b, le SN est singulier masculin et le verbe est toujours *ktb*.

Par contre, dans les exemples 11a et 11b, où le sujet est singulier féminin, le verbe est *ktbat*.

Dans le cas où le SN est pluriel, il y a neutralisation du genre. Le verbe est toujours *ktbū*, dans le cas des exemples 12a et 12b, où le sujet est masculin, et des exemples

3. Le signe * indique l'agrammaticalité de la phrase.

13a et 13b, où il est féminin. En plus, la désinence du pluriel *ū* ne change pas, quelque soit la position et le genre du SN, et ceci prouve, qu'en dialecte marocain, elle est devenue un simple marqueur d'accord du nombre.

(10)

- | | |
|--|--|
| <p>a <i>ktb</i> <i>l=wld</i>
 écrire._{PFV}-3MS DEF=garçon-NOM
 'le garçon a écrit'</p> | <p>b <i>l=wld</i> <i>ktb</i>
 DEF=garçon-NOM écrire._{PFV}-3MS
 'le garçon a écrit'</p> |
|--|--|

(11)

- | | |
|--|--|
| <p>a <i>ktbat</i> <i>l=bnt</i>
 écrire._{PFV}-3FS DEF=fille-NOM
 'la fille a écrit'</p> | <p>b <i>l=bnt</i> <i>ktbat</i>
 DEF=fille-NOM écrire._{PFV}-3FS
 'la fille a écrit'</p> |
|--|--|

(12)

- | | |
|--|--|
| <p>a <i>ktb-ū</i> <i>l=wlād</i>
 écrire._{PFV}-3MP DEF=garçons-NOM
 'les garçons ont écrit'</p> | <p>b <i>l=wlād</i> <i>ktb-ū</i>
 DEF=garçons-NOM écrire._{PFV}-3MP
 'les garçons ont écrit'</p> |
|--|--|

(13)

- | | |
|---|---|
| <p>a <i>ktb-ū</i> <i>al=bnāt</i>
 écrire._{PFV}-3FP DEF=filles-NOM
 'les filles ont écrit'</p> | <p>b <i>al=bnāt</i> <i>ktb-ū</i>
 DEF=filles-NOM écrire._{PFV}-3FP
 'les filles ont écrit'</p> |
|---|---|

3.1.2.2.3 Conclusion En arabe standard, les désinences du verbe *māḍī* ont les propriétés des affixes et résultent, vraisemblablement, de l'incorporation des pronoms nominatifs indépendants au verbe. En plus, elles ont un comportement mixte.

En ce qui concerne la première et la deuxième personne, la structure morphologique du *māḍī* peut être synthétisée comme :

Base verbale = Suffixe₁- Suffixe₂

Le Suffixe₁ est un pronom qui peut être :

- *nā* marqueur de 1D/P : *fa'al=nā* 'nous avons fait' ;
- *t* pour toutes les autres formes d'où l'utilité du Suffixe₂ pour la désambiguïsation selon personne, genre et /ou nombre :
 - pour les formes du singulier, le Suffixe₂ est une voyelle marquant la personne et/ou le genre :
 - fa'al=t-u* 'j'ai fait' ;
 - fa'al=t-a* 'tu (masculin) as fait' ;

fa'al=t-i 'tu (féminin) as fait' ;

- pour les formes du pluriel, le Suffixe₂ est *um* ou *unna*:

fa'al=t-um 'vous (masculin) avez fait' ;

fa'al=t-unna 'vous (féminin) avez fait' ;

- le duel s'obtient, en ajoutant le suffixe *ā* à la forme du pluriel masculin :

fa'al=t-um-ā 'vous (deux) avez fait'.

En ce qui concerne la troisième personne, au singulier, la structure morphologique du *māḍī* peut être synthétisée comme suite :

Base verbale - Suffixe₁ - Suffixe₂

À la 3SM, le Suffixe₁ est juste une voyelle *a* : *fa'al-a* 'il a fait'. De tout le système verbal, c'est la forme la moins marquée, ce qui explique le fait qu'elle représente le verbe dans le lexique. Pour marquer le féminin, il y a ajout du suffixe₂, *t*: *fa'al-a-t* 'elle a fait'.

Pour les autres formes (duel et pluriel), la structure peut être synthétisée comme :

Base verbale = Suffixe₃

Le fait que ces formes se réalisent seulement dans le cas où le sujet est préverbal, prouve que le Suffixe₃ est un pronom marqueur du nombre :

- *ā* pour marquer le duel : *fa'al=ā* 'ils (duel masculin) ont fait' ;
- *ū* pour le pluriel masculin : *fa'al=ū* 'ils ont fait' ;
- *na* pour marquer le pluriel féminin : *fa'al=na* 'elles ont fait'.

Pour marquer le duel féminin, on note que le suffixe *a-t* subsiste, parce que le pronom *ā* est identique aussi bien au masculin qu'au féminin. La structure résulte plus complexe :

Base verbale - Suffixe₁ - Suffixe₂ = Suffixe₃

fa'al-a-t=ā 'elles (duel féminin) ont fait'.

3.1.3 Analyse morphologique du verbe *muḍāri'* (inaccompli)

3.1.3.1 Conjugaison du verbe *muḍāri'*

Le *muḍāri'* a trois « modes », l'indicatif, le subjonctif et l'apocope, dont les conjugaisons sont illustrées dans le tableau 3.4. En général, toutes les formes se caractérisent par la présence d'une base verbale, *af'al*, précédée d'un préfixe et suivie d'un suffixe. En plus, on peut noter que :

- il n'y a pas distinction de genre pour la 1^{ère} personne et pour le duel la 2^e personne ;
- il n'y a pas distinction entre le duel et le pluriel de la 1^{ère} personne ;
- les formes de la 2M et de la 3F sont identiques au singulier et au duel ;
- les trois modes du *muḏāri'* se différencient par leurs désinences.

		1	2M	2F	3M	3F
S	Indicatif	'af'alu	taf'alu	taf'alīna	yaf'alu	taf'alu
	Subjonctif	'af'ala	taf'ala	taf'alī	yaf'ala	taf'ala
	Apocopé	'af'al	taf'al	taf'alī	yaf'al	taf'al
D	Indicatif	naf'alu	taf'alāni	taf'alāni	yaf'alāni	taf'alāni
	Subjonctif	naf'ala	taf'alā	taf'alā	yaf'alā	taf'alā
	Apocopé	naf'al	taf'alā	taf'alā	yaf'alā	taf'alā
P	Indicatif	naf'alu	taf'alūna	taf'alna	yaf'alūna	yaf'alna
	Subjonctif	naf'ala	taf'alū	taf'alna	yaf'alū	yaf'alna
	Apocopé	naf'al	taf'alū	taf'alna	yaf'alū	yaf'alna

Tab. 3.4 : Conjugaison du verbe *fa'ala* au *muḏāri'*

3.1.3.2 Analyse des affixes du *muḏāri'*

Définir l'identité des préfixes et des désinences qui contribuent à la formation du *muḏāri'* pose beaucoup de problèmes. Des études récentes considèrent les préfixes et les suffixes du *muḏāri'* comme des circonfixes.⁴ Mais une étude plus attentive démontre que la morphologie du système verbale arabe est beaucoup plus complexe.

Cette complexité se manifeste déjà dans le fait que certains traits flexionnels du verbe s'expriment par la combinaison vocalique du schème. Pour comprendre la nature réelle et le rôle des affixes qui contribuent à la flexion verbale, il est nécessaire de comprendre d'abord leurs composants et leur rôle syntaxique.

3.1.3.2.1 Analyse des préfixes

Une première remarque est le fait qu'au *muḏāri'*, la consonne C₁ perd la voyelle qui la suivait et se trouve précédée d'une voyelle, qu'on appellera V₀ dans cette étude.

4. Noyer, 1992; Habash, 2007

Quelque soit le verbe, V_0 est toujours /a/ à la voix active, par exemple, *y-aktub-u* 'il écrit' ; *y-aḍrib-u* 'il frappe' et *y-ašrab-u* 'il boit'. Par contre, c'est une voyelle /u/ à la voix passive : *y-uktab-u*; *y-uḍrab-u* et *y-ušrab-u*.

Cette caractéristique renforce l'idée que V_0 ne fait pas partie du préfixe mais de la mélodie vocalique et qu'elle marque la voix verbale.

La grammaire traditionnelle considère les préfixes comme caractéristiques du *muḍāri'*; cette considération est uniquement morphologique limitée au fait qu'ils existent seulement au *muḍāri'*. Elle affirme, d'autre part, qu'ils varient en fonction des traits morphosyntaxiques (personne, nombre et genre) comme le démontre le tableau 3.5.

Dans le cas de la 3SF et la 3DF, le préfixe est *t*, utilisé lorsque les formes du masculin et du féminin sont homonymes, comme il sera démontré plus avant et, il ne doit pas être confondu avec celui de la deuxième personne.

Personne	Genre	Singulier	Duel	Pluriel
1	M	ʾ	<i>n</i>	<i>n</i>
	F			
2	M	<i>t</i>	<i>t</i>	<i>t</i>
	F	<i>t</i>	<i>t</i>	<i>t</i>
3	M	<i>y</i>	<i>y</i>	<i>y</i>
	F	<i>t</i>	<i>t</i>	<i>y</i>

Tab. 3.5 : Préfixes du *muḍāri'*

3.1.3.2.2 Analyse des suffixes La grammaire traditionnelle divise les formes verbales en trois groupes en fonction du suffixe.

1. Un premier groupe est illustré dans le tableau 3.6. On note que le suffixe est une voyelle *u* à l'indicatif (In), *a* au subjonctif (Su), et il perd cette voyelle à l'apocopé (Ap).

Mode	1 S	1P	2MS	3MS	3FS
In	ʾ-af'al- <i>u</i>	<i>n</i> -af'al- <i>u</i>	<i>t</i> -af'al- <i>u</i>	<i>y</i> -af'al- <i>u</i>	<i>t</i> -af'al- <i>u</i>
Su	ʾ-af'al- <i>a</i>	<i>n</i> -af'al- <i>a</i>	<i>t</i> -af'al- <i>a</i>	<i>y</i> -af'al- <i>a</i>	<i>t</i> -af'al- <i>a</i>
Ap	ʾ-af'al	<i>n</i> -af'al	<i>t</i> -af'al	<i>y</i> -af'al	<i>t</i> -af'al

Tab. 3.6 : Premier groupe de formes au *muḍāri'*

2. Un deuxième groupe, dans le tableau 3.7, se caractérise par une désinence moins simple. La comparaison des suffixes entre indicatif, subjonctif et apocopé permet d'isoler deux parties :

- une partie invariable : *ī* pour la 2FS ; *ā* pour le duel et *ū* pour le pluriel masculin.
- une partie, plus externe, peut être *na* ou *ni* à l'indicatif, et subir une élision au subjonctif et à l'apocope.

Mode	2FS	2M/FD	3FD	3MD	2MP	3MP
In	<i>t-af'al=ī-na</i>	<i>t-af'al=ā-ni</i>	<i>t-af'al=ā-ni</i>	<i>y-af'al=ā-ni</i>	<i>t-af'al=ū-na</i>	<i>y-af'al=ū-na</i>
Su	<i>t-af'al=ī</i>	<i>t-af'al=ā</i>	<i>t-af'al=ā</i>	<i>y-af'al=ā</i>	<i>t-af'al=ū</i>	<i>y-af'al=ū</i>
Ap	<i>t-af'al=ī</i>	<i>t-af'al=ā</i>	<i>t-af'al=ā</i>	<i>y-af'al=ā</i>	<i>t-af'al=ū</i>	<i>y-af'al=ū</i>

Tab. 3.7 : Deuxième groupe de formes au *muḍāri'*

3. Un troisième groupe est constitué par les formes fléchies du pluriel féminin, caractérisé par une désinence *na* qui est invariable pour l'indicatif, le subjonctif et l'apocopé : *t-af'al=na* (2FP) et *y-af'al=na* (3FP).

3.1.3.2.3 Analyse de la partie invariable des affixes

3.1.3.2.3.1 Étude morphologique Les affixes variables du *muḍāri'* sont des suffixes qui peuvent être :

- une voyelle, dans le cas du premier groupe ;
- la désinence *na* ou *ni*, concernant le deuxième groupe.

Ils seront analysés en comparaison avec le système nominale, dans le paragraphe 4.2.1.3.

Par contre, les affixes invariables sont les parties qui ne subissent aucune variation à l'indicatif, au subjonctif et à l'apocopé. Concernant le premier groupe, la partie invariable est constituée seulement du préfixe, tandis que le suffixe est une voyelle variable. Par contre, la partie invariable, du deuxième et du troisième groupe, est constituée du préfixe et d'une partie du suffixe.

Le tableau 3.8 illustre la comparaison entre les pronoms indépendants, les suffixes du *māḍī* et les affixes invariables du *muḍāri'*, Préfixes et Suffixe₁, chacun accompagné des traits d'accords qu'il marque.

LA LANGUE ARABE
CARACTÉRISTIQUES MORPHOSYNTAXIQUES ET SÉMANTIQUES

Pr	N	G	Pronoms	<i>māḍī</i>	Affixes invariables <i>muḍāri</i> '			
				Suffixes	Préfixe	Traits	Suffixe ₁	Traits
1	S		'anā	-tu	'-	P+N	-	-
	D	-	naḥnu	-nā	n-	P+N	-	-
	P							
2	S	M	'anta	-ta	t-	P	-	-
		F	'anti	-ti	t-	P	-ī	G
	D	M	'antumā	-tumā	t-	P	-ā	N
		F						
	P	M	'antum	-tum	t-	P	-ū	G+N
		F	'antunna	-tunna	t-	P	-na	G+N
3	S	M	huwa	-a	y-	P	-	-
		F	hiya	-at	t-	P+G	-	-
	D	M	humā	-ā	y-	P	-ā	N
		F	humā	-tā	t-	P+G	-ā	N
	P	M	hum	-ū	y-	P	-ū	G+N
		F	hunna	-na	y-	P	-na	G+N

M : masculin ; F : féminin

S : singulier ; D : duel ; P : Pluriel

Pr : personne ; G : genre ; N : nombre

In : indicatif ; Su : subjonctif ; Ap : apocopé

Tab. 3.8 : Analyse comparative des pronoms indépendants ; des suffixes du *māḍī* et des affixes invariables du *muḍāri*'.

Affixes invariables de la 1^{ère} personne Dans toutes les formes fléchies de la 1^{ère} personne, la désinence consiste seulement dans la voyelle qui change en fonction du mode, et par conséquent, elle n'est pas prise en considération dans cette étude.

La partie invariable est constituée du préfixe qui marque les deux traits d'accord essentielles, la personne et le nombre. L'analyse comparative permet de souligner un rapprochement morphologique entre le pronom indépendant 'anā de la 1S et le préfixe ' du *muḍāri*'. Au pluriel, le rapprochement est encore majeur entre le pronom indépendant naḥnu, le suffixe -nā du *māḍī* et le préfixe n du *muḍāri*'.

Affixes invariables de la 2^{ème} personne En confrontant les pronoms et les formes verbales de la 2P, quelque soient le nombre et le genre, on note qu'ils ont, en commun, la consonne *t*, qui se réalise comme préfixe au *muḍāri'*. On peut conclure que le préfixe *t* du *muḍāri'* est la marque de la 2P. Au masculin singulier, le suffixe est uniquement la voyelle qui change en fonction du mode. En supposant que la partie invariable du suffixe est par défaut nulle au singulier, l'ajout des autres suffixes permet de marquer le genre et/ou le nombre pour obtenir les autres formes :

- *ī* pour marquer le genre féminin singulier ; et on note qu'il est très voisin à la voyelle *i* spécifique du pronom *'anti* et du suffixe *-ti* du *māḍī* de la 2FS ;
- *ā* qui marque le nombre duel. En comparant, on note qu'il représente la même marque du duel pour les prénoms, le *muḍāri'* et le *māḍī*.
- *ū* qui marque le nombre pluriel et le genre masculin. En comparaison, il est très voisin de la voyelle *u* qui caractérise le pronom *'antum* et le suffixe *-tum* du *māḍī*.
- *na* qui marque le nombre pluriel et le genre féminin. L'analyse comparative permet de souligner que *na* marque le pluriel féminin aussi bien dans le pronom *'antunna* que dans les formes verbales du *māḍī* (avec le suffixe *-nna*) et du *muḍāri'* (avec le suffixe *-na*).

Affixes invariables de la 3^{ème} personne On peut noter deux préfixes qui s'alternent : la *y* et la *t*. Le préfixe *y* est présent dans toutes les formes de la 3M mais c'est surtout sa présence à la 3FP qui prouve qu'il représente le préfixe par défaut de la troisième personne. Par contre, le préfixe *t* se réalisent seulement dans certaines formes du féminin. Si on fait les mêmes considérations prises dans le cas de la 2P, autrement dit que la partie invariable du suffixe *muḍāri'* est, par défaut, nulle au singulier, on peut déduire que les autres formes sont obtenues en ajoutant des suffixes marquant le genre et/ou le nombre :

- *ā* qui marque le duel mais indépendamment du genre ;
- *ū* qui marque le nombre pluriel et le genre masculin ;
- *na* qui marque le nombre pluriel et le genre féminin.

La réalisation du préfixe *t* se fait dans le cas où le suffixe ne spécifie pas le genre. C'est le cas du singulier où le suffixe est justement nul et, du duel où le suffixe *ā* ne marque pas le genre. Dans ces deux cas, pour distinguer entre le masculin et le féminin, il y

a changement du préfixe de défaut *y* (qui marque en général la troisième personne) avec le préfixe spécifique *t* qui marque le féminin de la 3P.

3.1.3.2.3.2 Étude syntaxique Les exemples, 14, 15, 16, 17, 18 et 19, présentent des cas où le syntagme nominal (SN) est post-verbal. Quelque soit le nombre (singulier, duel ou pluriel), on note que le verbe est *y-aktub-u* lorsque le sujet est masculin, et *t-aktub-u* avec un sujet féminin. Le changement du préfixe permet de marquer le genre. Avec un sujet masculin, le préfixe est *y* dans la forme *y-aktub-u*, par contre, avec un sujet féminin, le préfixe est *t* dans le cas de *t-aktub-u*. Dans tous les cas, le suffixe du verbe est formé uniquement de la voyelle *u* qui marque le mode de l'indicatif. Comme dans le cas du *māḍī*, l'accord est donc partiel. Les affixes du verbe marquent seulement le genre et la personne, par contre la marque du nombre est neutralisée. Le sujet, lui même, indique le nombre, parce qu'il suit le verbe.

(14)

y-aktub-u *l=walad-u*
 3-écrire_{.IPFV-IND.PRS} DEF=garçon-NOM
 'le garçon écrit'

(15)

y-aktub-u *l=walad-ā-ni*
 3-écrire_{.IPFV-IND.PRS} DEF=garçons-DUEL-NOM
 'les deux garçons écrivent'

(16)

y-aktub-u *l='awlād-u*
 3-écrire_{.IPFV-IND.PRS} DEF=garçons-NOM
 'les garçons écrivent'

(17)

t-aktub-u *l=bint-u*
 3F-écrire_{.IPFV-IND.PRS} DEF=fille-NOM
 'la fille écrit'

(18)

t-aktub-u *l=bint-ā-ni*
 3F-écrire_{.IPFV-IND.PRS} DEF=filles-DUEL-NOM
 'les deux filles écrivent'

(19)

t-aktub-u *l=banāt-u*
 3F-écrire_{.IPFV-IND.PRS} DEF=filles-NOM
 'les filles écrivent'

Les exemples, 20, 21, 22, 23, 24 et 25, sont des cas où le SN est pré-verbal.

(20)

al=walad-u *y-aktub-u*
DEF=garçon-NOM 3-écrire._{.IPFV}-IND.PRS
'le garçon écrit'

(21)

al=bint-u *t-aktub-u*
DEF=filles-NOM 3F-écrire._{.IPFV}-IND.PRS
'la fille écrit'

(22)

al=walad-ā-ni *y-aktub=ā-ni*
DEF=garçons-DUEL-NOM 3-écrire._{.IPFV}-DUEL-IND.PRS
'les deux garçons écrivent'

(23)

al=bint-ā-ni *t-aktub=ā-ni*
DEF=filles-DUEL-NOM 3F-écrire._{.IPFV}-DUEL-IND.PRS
'les deux filles écrivent'

(24)

al='awlād-u *y-aktub=ū-na*
DEF=garçons-NOM 3-écrire._{.IPFV}-PL-IND.PRS
'les garçons écrivent'

(25)

al=banāt-u *y-aktub=na*
DEF=filles-NOM 3-écrire._{.IPFV}-F.PL
'les filles écrivent'

On note que l'accord est complet en genre, personne et nombre.

Le préfixe est *y* lorsque le sujet est masculin, dans les exemples 20, 22 et 24.

Cependant, on le trouve, également, dans la forme du féminin pluriel, dans l'exemple 25 qui est différente de celle du pluriel masculin, de l'exemple 24.

Par contre, le préfixe est la marque du féminin *t*, lorsque le féminin a une forme identique à celle du masculin, c'est le cas du singulier, exemple 21 et du duel, exemple 23, en les comparant, respectivement, avec les exemples 20 et 22.

En ce qui concerne le suffixe, il contient :

- la marque modale de l'indicatif : *u* au singulier, *na/ni* au duel et au pluriel masculin ;
- la désinence est composée seulement de la voyelle *u*, dans le cas où le sujet est singulier : *y-aktub-u* ou *t-aktub-u*;

- dans le cas d'un sujet duel ou pluriel, on note que le suffixe contient la marque du nombre : *ā* pour le duel, *ū* pour le pluriel masculin et *na* pour le pluriel féminin.
- les suffixes, *ā*, *ū* et *na*, se réalisent pour marquer le nombre uniquement dans le cas où le SN précède le verbe et ils sont considérés des pronoms (*damīr mutṭaṣīl*).

Conclusion La structure morphologique du *muḍāri'* peut être synthétisée comme suite :

Préfixe - Base verbale = Suffixe₁ - Suffixe₂

Le Préfixe et la Base verbale constituent les deux parties stables du *muḍāri'*. Par contre, l'interaction entre Préfixe, Suffixe₁ et Suffixe₂ peut avoir comme conséquence la réalisation ou non des suffixes.

La partie suffixale la plus externe, **Suffixe₂**, contribue à la réalisation des traits temps et mode. Le **Préfixe** et le **Suffixe₁** fonctionnent comme un « circumfixe » qui marquent les traits d'accord (personne, nombre, et genre). Cependant, comme le fait noter Noyer :

The Arabic tense/agreement "morpheme" is not simply a circumfix or simulfix that splits into two pieces, encircling the stem. Rather, the prefix and suffix are realized separately and interactively.⁵

De conséquence, on ne peut pas parler réellement d'un vrai circumfixe, puisque les parties d'un circumfixe ne se réalisent jamais l'une sans l'autre ; alors que le Préfixe du *muḍāri'* a la capacité de se réaliser indépendamment du Suffixe₁.

La neutralisation du singulier a pour conséquence l'effacement du Suffixe₁. Dans ce cas, la seule partie restante est le Suffixe₂ dont la réalisation marque l'indicatif et le subjonctif et sa non-réalisation marque l'apocopé.

Le Suffixe₁ se réalise pour marquer le genre de la 2FS (*ī*), le duel (*ā*) et le pluriel (*ū* du masculin). Dans ce cas, le Suffixe₂ est *ni* ou son allomorphe *na*, dont la réalisation marque l'indicatif et son élision marque le subjonctif et l'apocopé.⁶

Le Suffixe₁ se réalise aussi pour marquer le pluriel féminin (*na*). Dans ce cas, l'homographie entre le Suffixe₁, marqueur du pluriel féminin, et le Suffixe₂, marqueur de mode, a pour conséquence l'effacement de ce dernier.

Les exceptions, à ce schéma général, est la preuve de l'interaction entre ces affixes pour la réalisation morphologique du *muḍāri'*. Il suffit de noter l'effacement du Suffixe₁ dans la forme de la 1P, *n-af'al-u*, où le Préfixe marque à la fois la personne et le nombre.

5. Noyer, 1992, p. 39.

6. Benmamoun, 1999.

Les formes de la troisième personne au féminin sont, aussi, une preuve de cette interaction. La réalisation du singulier, *t-af'al-u*, et du duel *t-af'al=ā-ni*, se fait grâce au préfixe *t*, qui marque justement le féminin, parce que la partie suffixale, marquant le nombre, est commune avec les formes du masculin, *y-af'al-u* et *y-af'al=ā-ni*. Par contre, le pluriel, *y-af'al=na*, se réalise grâce au préfixe de défaut, *y*, parce que la désinence *na* est spécifique du pluriel féminin.

3.1.4 Conclusion

Le verbe arabe peut avoir deux formes, le *māḍī*, appelé aussi le passé ou encore l'accompli, et le *muḍāri'*, appelé aussi le non passé ou l'inaccompli. L'opposition de ces deux formes verbales se manifeste, aussi, dans leur structure morphologique, comme le montrent les exemples reportés dans le tableau 3.9.⁷

	Pers.	Morphème	base verbale	Morphème	
<i>māḍī</i>	2 ^{ème}		<i>katab</i>	<i>t-a</i>	'tu as écrit'
<i>muḍāri'</i>	2 ^{ème}	<i>t</i>	<i>aktub</i>	<i>-u</i>	'tu écris'

Tab. 3.9 : *māḍī* vs *muḍāri'*

Le *māḍī* est construit grâce à la postposition du morphème de la personne par rapport à la base verbale (dans l'exemple du tableau, il s'agit du morphème *t* de la 2P), on parle de « conjugaison suffixée » (suffix-conjugation).⁸ En ce qui concerne le *muḍāri'*, par contre, on note une antéposition du morphème de la personne et pour ceci on l'appelle aussi « conjugaison préfixée » (« préfix-conjugation »).

3.1.5 Pattern (ou mélodie) vocalique du *māḍī* et du *muḍāri'*

Comme il a été déjà expliqué, dans le paragraphe 3.1.3.2.1, le pattern vocalique, dans le cas du verbe trilitère, est constitué de deux voyelles (V_1 - V_2) au *māḍī* où V_1 suit la première consonne radicale et V_2 suit la deuxième consonne radicale. Au *muḍāri'*, on constate qu'il y a élision de V_1 et en contrepartie il y a ajout d'une voyelle V_0 , précédant la première consonne radicale. Dans ce cas, le pattern vocalique est également constitué de deux voyelles : (V_0 - V_2). Cependant il faut noter que V_1 est toujours identique à V_0 et, toutes les deux, constituent la voyelle marquant la voix active, et dans ce cas, V_1 et V_0 sont *a*, ou la voix passive, dans ce cas, V_1 et V_0 est *u*.

7. Les exemples sont discutés dans Hamzé, 1999, page 30.

8. Moscati, 1964.

Dans le cas des verbes quadrilitères et des verbes dérivés, certains exemples sont reportés dans le tableau 3.10.

Pattern		Verbe	Racine	<i>māḍī</i>	<i>muḍāri</i> '	Traduction
<i>māḍī</i>	<i>muḍāri</i> '					
a-a	u-a-i	quadrilitère	<i>dḥrġ</i>	<i>daḥraġ-a</i>	<i>y-udaḥriġ-u</i>	'faire rouler'
		dérivé	<i>ksr</i>	<i>kassar-a</i>	<i>y-ukassīr-u</i>	'broyer'
			<i>ktb</i>	<i>kātab-a</i>	<i>y-ukātib-u</i>	'écrire à'
			<i>šrb</i>	<i>ʾašrab-a</i>	<i>y-ušrib-u</i>	'faire boire'
a-a	a-a-i	dérivé	<i>ktb</i>	<i>nkatab-a</i>	<i>y-ankatib-u</i>	'être écrit'
			<i>ktb</i>	<i>ktatab-a</i>	<i>y-aktatib-u</i>	'se faire écrire'
			<i>ktb</i>	<i>staktab-a</i>	<i>y-astaktib-u</i>	'dicter'
			<i>dḥrġ</i>	<i>ndaḥraġ-a</i>	<i>y-andaḥriġ-u</i>	'rouler'
a-a-a	a-a-a-a	dérivé	<i>ḍrb</i>	<i>taḍārab-a</i>	<i>y-ataḍārab-u</i>	'se battre'
			<i>ksr</i>	<i>takassar-a</i>	<i>y-atakassar-u</i>	'se rompre'
			<i>dḥrġ</i>	<i>tadaḥraġ-a</i>	<i>y-atadaḥraġ-u</i>	'rouler'

Tab. 3.10 : Exemples de verbes non trilitères.

On note qu'au *muḍāri*' , il y a seulement ajout de V_0 . Donc, si le pattern vocalique du *māḍī* est constitué de (n) voyelles, il est constitué de $V_0 + (n)$ voyelles, au *muḍāri*' . On distingue trois groupes de flexion, selon la variation du pattern vocalique.

Un premier groupe est constitué par le verbe quadrilitère «nu » caractérisé par V_0 qui est /u/. Dans ce cas, V_0 ne marque pas la voix mais, le fait que, bien qu'il ne soit pas dérivé, le verbe est constitué de quatre consonnes radicales ; ce qui a des implications sémantiques importantes. Par analogie les verbes dérivés constitués, également, de quatre lettres suivent cette règle.

Les deux autres groupes sont constitués seulement par des verbes dérivés. Ils sont caractérisés par $V_0 = a$, marquant la voix active.

Un autre facteur lié à la longueur du pattern vocalique divise ces groupes :

- au *māḍī*, le pattern vocalique est constitué de deux voyelles : a-a ; il y a variation de la dernière voyelle en /i/ au *muḍāri*' : V_0 -a-i ;
- au *māḍī*, le pattern vocalique est constitué de trois voyelles : a-a-a ; on note qu'il est plus stable et ne subit aucune variation au *muḍāri*' : V_0 -a-a-a.

3.2 Études morphosémantiques du verbe arabe

3.2.1 Le verbe trilitère

3.2.1.1 Classes morphosyntaxiques et sémantiques

Comme déjà expliqué dans le chapitre précédent, et selon tableau 3.11, le schème général du verbe trilitère est $C_1V_1C_2V_2C_3-V_3$ à l'accompli et $y-V_0C_1C_2V_2C_3-V_3$ à l'inaccompli.

Voix active				
accompli	inaccompli	alternance vocalique	Voix passive	
$C_1aC_2aC_3-a$	$y-aC_1C_2uC_3-u$	a-u	accompli	$C_1uC_2iC_3-a$
	$y-aC_1C_2iC_3-u$	a-u	inaccompli	$y-uC_1C_2aC_3-u$
	$y-aC_1C_2aC_3-u$	a-a	alternance vocalique	i-a
$C_1aC_2iC_3-a$	$y-aC_1C_2aC_3-u$	i-a		
$C_1aC_2uC_3-a$	$y-aC_1aC_2uC_3-u$	u-u		

Tab. 3.11 : Alternance de V_2 du verbe triconsonnantique

Le préfixe et la troisième voyelle V_3 sont des affixes flexionnels variables en fonction de la personne, du genre et du mode.

La première voyelle du schème (V_0 ou V_1) est toujours *a*, marquant la diathèse active, ou *u* marquant la diathèse passive.

Par contre, la voyelle médiane V_2 joue plusieurs rôles. D'une part, **au sein du même verbe**, V_2 est véhiculaire d'informations aspectuelles et représente la voyelle distinctive dont l'alternance marque le contraste entre l'accompli et l'inaccompli. De l'autre, en comparant **la forme active de l'accompli de divers verbes**, V_2 peut être /a/ ; /u/ ou /i/. Cette variation, véhiculaire d'informations sémantiques, permet la division des verbes trilitères en trois classes qu'on essaiera d'approfondir dans cette section.

3.2.1.2 Les verbes de schème $C_1aC_2uC_3a$

3.2.1.2.1 Alternance vocalique

Cette classe est caractérisée par une voyelle médiane *u*, à l'accompli et à l'inaccompli. Cette classe est constituée exclusivement de

verbes d'états que Dichy définit comme suite :

“L'aspect 'état' exprime la stabilité de la situation référentielle représentée, c'est-à-dire qu'il réfère à une absence de mouvement et de changement. L'état décrit un procès qui n'est borné ni à gauche (en son commencement) ni à droite (le terme n'est pas spécifié), et ne fait référence à aucun agent.⁹

L'absence de l'alternance vocalique souligne la stabilité de la situation exprimée par le verbe et l'absence d'action. Très important à noter est le fait qu'on ne trouve jamais un verbe dont C₂ ou C₃ est une *yā'*, à part deux verbes que al-'astarabādī (m. vers 686/1287) a essayé d'expliquer, en affirmant qu'ils représentent deux exemples qui prouvent l'importance du schème verbale :¹⁰

- *hayu'a - yahyu'u* 'être beau' : selon les règles de flexion, la présence de la *yā'* aurait dû changer la voyelle *u* en *i*. De conséquence le verbe aurait dû être **hā'a - yahī'u*¹¹ et le schème aurait dû changer en C₁aC₂iC₃a - yaC₁C₂aC₃a. Au voisinage de la *yā'*, la voyelle *u* change toujours en *i*, sauf dans ce cas où elle reste intacte, justement pour préserver intact le schème qui indique que le verbe *hayu'a* est un verbe d'état.
- *bahuwa - yabhū* 'être beau' : al-'astarabādī affirme que la racine est *bhy* et le verbe aurait dû être **bahuya*. La présence de la *yā'* aurait apporté au verbe des changements phonétiques de **bahuya* en **bahiya* et ; le schème de l'accompli aurait dû changer de C₁aC₂uC₃a en C₁aC₂iC₃a. Mais, pour le préserver, au lieu de changer la voyelle, c'est la consonne radicale *yā'* qui s'est transformée en *wāw*. En langue arabe, selon les règles phonétiques, la *wāw* peut changer en *yā'* mais jamais le contraire, sauf dans le cas du verbe *bahuwa*.

3.2.1.2.2 Valeurs sémantiques Sibawayhi décrit les verbes de schème C₁aC₂uC₃a comme verbes exprimant les qualités qui sont dans les choses :

*al=ḥisāl al=latī takūnu fī l='asyā'i.*¹²

Pour Wright, ils expriment un état permanent et une qualité naturellement inhérente :

9. Dichy, 2002, p. 326.

10. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I ; pp.76-80.

11. le signe * indique que la forme est inexistante.

12. *sībawayhi*, 'al-kitāb, Vol. IV- p. 28.

u indicates a permanent state, or a naturally inherent quality.¹³

Par contre, Dichy a proposé de substituer la « qualité permanente » avec la notion de « état caractéristique » :

L'état caractéristique - par opposition à l'état non caractéristique - décrit une propriété ou un attribut de l'objet, posé, selon le point de vue, comme dominant, intrinsèque, inhérent ou inscrit dans la nature de celui-ci. Cet état peut être acquis à l'issue d'un évènement ou d'un processus.¹⁴

3.2.1.2.3 Adjectif associé au verbe C₁aC₂uC₃a Pour le fait que ce sont des verbes d'état, ils sont intransitifs et ne peuvent être conjugués ni au passif ni à l'impératif. En plus, ils ne peuvent avoir aucun des deux participes, actif et passif. Ils sont associés, dans le lexique, à un adjectif qui exprime l'état du sujet, et que, traditionnellement, on appelle *šifah mušabbaha* 'adjectif assimilé/analogue'. Le tableau 3.12 montre quelques exemples de verbes d'états et les adjectifs qui leur sont corrélés. Il faut noter que le schème de l'adjectif «analogue » n'est pas formel et peut être C₁aC₂aC₃; C₁aC₂ĪC₃ ou C₁aC₂C₃.

Qualité	Verbe	Adjectif	Schème de l'adjectif
beauté - laideur	<i>ḥasuna</i> 'être beau'	<i>ḥasan</i> 'beau'	C ₁ aC ₂ aC ₃
	<i>ḡamula</i> 'être beau'	<i>ḡamīl</i> 'beau'	C ₁ aC ₂ ĪC ₃
grandeur - petitesse	<i>ʿazuma</i> 'être énorme'	<i>ʿazīm</i> 'énorme'	C ₁ aC ₂ ĪC ₃
courage - lâcheté	<i>ḍaʿufa</i> 'être faible/lâche'	<i>ḍaʿīf</i> 'faible/lâche'	
effort physique - mental	<i>šaʿuba</i> 'être difficile'	<i>šaʿb</i> 'difficile'	C ₁ aC ₂ C ₃

Tab. 3.12 : Exemples de verbes C₁aC₂uC₃a et adjectifs apparentés

Si l'adjectif «analogue » est X, le verbe est traduit par «être ou devenir X ». La différence entre l'utilisation du verbe ou de l'adjectif est subtile. Pour ces verbes, l'accompli

13. Wright, 1896, Vol. I, pag. 30.

14. Dichy- 2007, pp. 327-338.

correspond au présent actuel du français et le moment de l'énonciation coïncide avec le passage de la frontière entre «non être X » et «être X ».

Dans l'exemple 26, l'utilisation du verbe *ḥasun-a* 'être beau' sous-entend le passage de «non être beau » avant l'énonciation à «être beau » au moment de l'énonciation.

Par contre, dans l'exemple 27, l'utilisation de l'adjectif *ḥasan* 'beau' exprime seulement la constatation de la beauté de Zaid.

(26)

ḥasun-a *Zaid-un*
 être.beau_{PFV-3MS} Zaid-NOM
 'Zaid est beau [maintenant]'

(27)

Zaid-un *ḥasan-un*
 Zaid-NOM beau-NOM
 'Zaid est beau'

Pour exprimer le passé, on peut utiliser le verbe, comme dans l'exemple 28, mais il faut l'introduction d'un circonstanciel à valeur temporelle, par exemple, *al=bāriḥat-a* 'hier'. Par contre, en faisant recours à l'adjectif, dans l'exemple 29, on a besoin du verbe "auxiliaire" *kāna* 'être'.

(28)

ḥasun-a *Zaid-un* *al=bāriḥat-a*
 être.beau_{PER-3MS} Zaid-NOM DEF=hier-ACC
 'Zaid était beau hier'

(29)

kān-a *Zaid-un* *ḥasan-an* (*al=bāriḥat-a*)
 être_{PFV-3MS} Zaid-NOM beau-ACC (DEF=hier-ACC)
 'Zaid était beau (hier)'

Très intéressante la discussion que fait al-'astarabādī (m. vers 686/1287) à propos de l'adjectif «assimilé », dérivant d'un verbe intransitif qui n'indique pas une action et contient le sens de «permanence »:

Puisque l'adjectif assimilé n'est pas fait pour « advenir dans le temps », il n'est pas fait non plus pour «perdurer dans tous les temps » parce que «advenir » et «perdurer » sont deux conditions de l'adjectif mais il ne contient pas d'indication sur aucun des deux.¹⁵

15. 'Astarabādī, *Šarḥ al-kāfiyyah*, V. II, pp. 745-746.

Selon *al=ʿastarabādī*, l'adjectif *ḥasan* 'beau' indique seulement le fait de posséder la qualité *ḥusn* 'beauté' et, ne contient aucune indication du temps, ni permanente ni transitoire. Il ajoute que, apriori, ce type d'adjectif existe dans tous les temps, jusqu'à ce qu'il y ait un indice qui caractérise un temps précis, comme dans l'exemple 30.

(30)

kān-a ḥasan-an fa=qabuḥ-a
 être.PFV-3MS beau-ACC puis=s'enlaidir.PER-3MS
 'Il était beau, puis il s'est enlaidi / il est *devenu* laid'

Dans le cas où le contexte insinue l'incidence et la localisation dans le future, on sous-entend la «volonté» et «l'effort» du sujet pour obtenir la qualité exprimée par le verbe. Dans ce cas, on peut utiliser :

- l'inaccompli du verbe, associé à un circonstanciel à valeur temporelle du future ; c'est le cas de l'exemple 31;
- un participe actif, de paradigme $C_1\bar{a}C_2iC_3$, qui, généralement, est associé à un verbe d'action, comme dans l'exemple 32.

(31)

huwa yaḥsun-u ḡad-an
 lui être.beau.IPFV demain-ACC
 'il [sera] beau demain'

(32)

huwa ḥāsin-un ḡad-an
 lui beau-NOM demain-ACC
 'il [sera] beau demain'

3.2.1.3 Les verbes de schème $C_1aC_2iC_3a$

3.2.1.3.1 Alternance vocalique L'alternance formelle de ce groupe est i-a.

Cependant, quelques verbes ont l'alternance i-i, que les sciences traditionnelles ont essayé d'expliquer en cherchant *al=ʿillah* 'la cause'.¹⁶ Ces verbes à alternance i-i sont juste une trentaine dont la plupart ont une *wāw* comme C_1 ou C_2 .

¹⁶ Pour une dissertation plus détaillée sur les causes des changements ou *ʿillah* voir, entre autres, Ibn Ğinnī, *ʿal-Ḥaṣāʾiṣ*, V.I, pp. 48-96 et Dichy, 2002, pp. 317-319.

On peut aussi tracer une autre perspective de classement, en distinguant deux grandes catégories sémantiques de ces verbes. La première catégorie, dont quelques exemples sont illustrés dans le tableau 3.13, inclue les verbes qui expriment un état ou l'acquisition d'un état, généralement, une particularité physique ou psychique (défaut ou qualité) ou une couleur.

verbe		Adjectif		Sens
accompli	inaccompli	masculin	féminin	
<i>ğayida</i>	<i>yağyadu</i>	'ağyad	ğaydā'	'avoir le cou long' (signe de beauté)
'awiša	ya'wašu	'a'waš	'awšā'	'être difficile/incompréhensible'
<i>baliha</i>	<i>yablahu</i>	'ablah	balhā'	'être idiot'
<i>ħađira</i>	<i>yaħđaru</i>	'aħđar	ħađrā'	'devenir vert'

Tab. 3.13 : Exemples de verbes associés à un adjectif d'état

Ils sont associés à un adjectif «assimilé» de schème formel : 'aC₁C₂aC₃ dont le féminin est C₁aC₂C₃ā'. Il faut noter que, dans le cas où ces verbes contiennent une *wāw* ou *yā'* comme radicale, elle est toujours en position médiane C₂, et ils n'appliquent pas les règles de transformation phonologiques habituelles pour préserver l'intégrité du schème (comme le cas des verbes C₁aC₂uC₃a).

La deuxième catégorie incluent les verbes moyens définis, selon les considérations de Benveniste :

Dans l'actif, les verbes dénotent un procès qui s'accomplit à partir du sujet et hors de lui. Dans le moyen, qui est la diathèse à définir par opposition, le verbe indique un procès dont le sujet est le siège ; le sujet est intérieur au procès. [...] Ici le sujet est le lieu du procès, même si ce procès, [...], demande un objet ; le sujet est centre en même temps qu'acteur du procès ; il accomplit quelque chose qui s'accomplit en lui, naître, dormir, gésir, imaginer, croire, etc. Il est bien intérieur au procès dont il est l'agent.²²

En fonction de leur agentivité, ces verbes peuvent avoir un adjectif 'assimilé' ou un nom d'agent. Dans le cas où l'agentivité est nulle, les verbes ont aussi bien l'adjectif 'assimilé' que le nom d'agent. Par exemple :

مَرِيضٌ marīḍa 'être ou tomber malade' dont l'adjectif est *marīḍ*, et le participe actif est *māriḍ*;

22. Benveniste, 1950-1966, (p. 172).

غَرِقَ *ġariqa* 'se noyer ou mourir noyé' qui se caractérise par un adjectif, *ġarīq* et un participe actif *ġāriq*.

L'adjectif 'assimilé', dans l'exemple 34, correspond à un état résultatif engendré par l'évènement ou il coïncide avec la dernière étape du processus décrit par le verbe. Par contraste, dans l'exemple 35, le participe actif a une valeur de progressif à terme spécifié.²³

(34)

huwa ġarīq-un
lui noyé-NOM
'il est noyé'

(35)

huwa ġāriq-un
lui noyé-NOM
'il est [en train] de se noyer'

Si l'agentivité est partielle, le sujet a un contrôle limité sur le processus. Le verbe est transitif indirect ou intransitif et a seulement un nom d'agent. C'est le cas, par exemple, de *daḥika min* 'rire de ; se moquer de' dont le nom d'agent est *dāḥik*.

Lorsque l'agentivité est entière, le sujet a le plein contrôle du processus ou de l'évènement qui, peut être transitif direct ou indirect, et comme l'explique bien Dichy :²⁴

L'autre grande partie des verbes de gabarit *fa'ila* est constituée de verbes également moyens (ou à diathèse interne), d'agentivité entière et de régime syntaxique transitif direct ou indirect. Dans cette configuration, le procès comporte un agent autonome contrôlant effectivement (du fait de l'agentivité entière) l'évènement ou le processus décrit par le verbe et un « patient » (du fait de la transitivité). Il englobe en outre d'une manière ou d'une autre l'agent ou se déroule au moins partiellement en lui. Ces verbes ont un participe actif, et n'admettent pas d'adjectif 'assimilé'. Exemples, *šariba* 'boire' ; *raḍiya* 'être satisfait de [...]'; *laḥiqa* 'suivre, poursuivre' [...]

3.2.1.4 Les verbes de schème C₁aC₂aC₃a

3.2.1.4.1 Alternances vocaliques Le schème C₁aC₂aC₃a peuvent avoir trois alternances : a-a ; a-i ou a-u.

23. Dichy- 2007, p. 347.

24. Dichy- 2007, p. 349.

Cependant, l'alternance a-a, dont le tableau 3.14 donne quelques exemples, se note seulement pour les verbes dont C₂ ou C₃ est une consonne gutturale, appelées par les grammairiens arabes *ḥurūf al-ḥalq* 'les consonnes de l'arrière-bouche'. Ces consonnes sont six : ʾ, *h*, *ḥ*, ʿ, *ǧ*, *ḫ* et elles sont toutes des consonnes postérieures fricatives, à part la *hamzah* qui est occlusive.

Racine	Accompli	Inaccompli	Traduction
<i>sʾl</i>	<i>saʾala</i>	<i>yasʾalu</i>	'demander'
<i>qrʾ</i>	<i>qaraʾa</i>	<i>yaqraʾu</i>	'lire'

Tab. 3.14 : Exemples de verbes à alternance a-a

Selon la grammaire traditionnelle, la classe d'alternance a-a n'est pas formelle (*laysat bi-ʾaṣl*), mais dérive de la transformation de la classe d'alternance a-i ou a-u dans le but «d'alléger et harmoniser les sons »²⁵.

Par conséquent, les alternance formelles des verbes C₁aC₂aC₃a sont a-i ou a-u. En plus, certains verbes peuvent avoir, en même temps, les deux alternances, comme le montrent les exemples du tableau 3.15.

Accompli	Inaccompli	Sens
<i>nafara</i>	<i>yanfuru - yanfiru</i>	's'enfuir ; saillir ; abominer'
<i>šatama</i>	<i>yaštumu - yaštimu</i>	'blasphémer'
<i>nasala</i>	<i>yansulu - yansilu</i>	'engendrer'

Tab. 3.15 : Exemples de verbes à alternances a-i et a-u

Cette dispersion morphologique n'est pas systématique et elle reste sujette à la fréquence de l'usage.

Mais, on peut détecter quelques corrélats formels, illustrés par les exemples du tableau 3.16.

Dans le cas où la 2^e ou la 3^e consonne radicale est une *wāw*, le verbe fait partie de la classe a-u.

Par contre, dans le cas où la 2^e ou la 3^e consonne radicale est une *yāʾ*, il fait partie de la classe a-i.

25. ʾibn-yaʾīṣ, *al-mufaṣṣal*, Vol. IV, pp. 428-429 et ʾal-ʾAstarabādī, *Šarḥ aš-šāfiyyat*, Vol. I, pp. 118-125.

Classe	Racine	Accompli	Inaccompli	Maṣḍar	Sens
a-u	<i>nwh</i>	<i>nāḥa</i>	<i>yanūḥu</i>	<i>nawḥ</i>	'pleurer ; gémir'
	<i>qwl</i>	<i>qāla</i>	<i>yaqūlu</i>	<i>qawl</i>	'dire'
	<i>ḡzw</i>	<i>ḡazā</i>	<i>yaḡzū</i>	<i>ḡazw</i>	'infester ; assaillir ; conquérir'
a-i	<i>byʿ</i>	<i>bāʿa</i>	<i>yabīʿu</i>	<i>bayʿ</i>	'vendre'
	<i>rmy</i>	<i>ramā</i>	<i>yarmī</i>	<i>ramy</i>	'éjecter' ; 'jeter' ; 'lancer'

Tab. 3.16 : Exemples de verbes dont la 2^e ou la 3^e consonne radicale est faible

3.2.1.4.2 Valeurs sémantiques On note une forte dispersion sémantique des verbes C₁aC₂aC₃a, et dans ce cas aussi, on peut seulement essayer d'identifier et formaliser certains corrélats.

Tous les verbes actifs qui, selon la définition de Benveniste, *s'accomplissement à partir du sujet et hors de lui*, sont de schème C₁aC₂aC₃a et sont liés, justement, à un participe actif de schème C₁āC₂iC₃. Dans le cas où ils sont transitifs, ils sont liés, également, à un participe passif, de schème maC₁C₂ūC₃, comme le montrent les exemples du tableau 3.17.

Classe	verbe	Participe		Sens
		actif	passif	
a-u	<i>madda -yamuddu</i>	<i>mādd</i>	<i>mamdūd</i>	'tendre ; étendre'
	<i>qāla - yaqūlu</i>	<i>qā'il</i>	<i>maqūl</i>	'dire'
a-i	<i>ḍaraba - yaḍribu</i>	<i>ḍārib</i>	<i>maḍrūb</i>	'frapper'
	<i>bāʿa - yabīʿu</i>	<i>bā'iʿ</i>	<i>mabyūʿ</i>	'vendre'
a-a	<i>ḥaḍaʿa - yaḥḍaʿu</i>	<i>ḥāḍiʿ</i>	<i>maḥḍūʿ</i>	'tromper ; duper'
	<i>saḥara - yaṣḥaru</i>	<i>sāḥir</i>	<i>mashūr</i>	'charmer ; ensorceler'

Tab. 3.17 : Exemples de verbes à diathèse externe

Toutefois, pas tous les verbes C₁aC₂aC₃a sont des verbes d'action ; on peut trouver aussi des verbes moyens, comme *ḍahaba* 's'en aller' ; *daḥala* 'entrer'.

Tous les verbes d'état de schème C₁aC₂aC₃a qui sont associés à **un adjectif «as-similé»**, sont des verbes dont la 2^e consonne radicale est une *wāw* ou une *yāʿ* (verbes concaves) ; ou des verbes dont la 2^e et la 3^e consonnes radicales sont identiques (verbes géminés). Le tableau 3.18 illustre quelques exemples.

Verbes	Classe	Exemples	Adjectif «assimilé »	Sens
Concaves	a-i	<i>ḡāḡa</i>	<i>ḡayyiq</i>	'devenir étroit'
		<i>lāna</i>	<i>layyin</i>	'devenir souple/tendre'
	a-u	<i>jāda</i>	<i>jayyid</i>	'devenir excellent'
		<i>sāda</i>	<i>sayyid</i>	'devenir le chef'
Géminés	a-i	<i>xaffa</i>	<i>xaff</i>	'devenir léger'
		<i>qalla</i>	<i>qalīl</i>	'devenir peu nombreux'

Tab. 3.18 : exemples de verbes d'état C₁aC₂aC₃a

Toutefois, comme le précise Dichy, pas tous les verbes concaves ou géminés de schème C₁aC₂aC₃a sont des verbes d'état.²⁶

Certains verbes sont de la classe C₁aC₂aC₃a - yaC₁aC₂uC₃a (alternance a-u), dans un contexte de défi et de compétition, pour indiquer la supériorité du sujet sur le concurrent qui, dans ce cas, occupe la position du complément. Comme le fait noter al-'astarabādī, même si le verbe est d'un autre schème, il peut être ramené à la classe a-u, lorsque le contexte insinue «la victoire après la compétition ».²⁷

L'exemple typique, de ce dernier cas, est le verbe *karuma - yakrumu* 'être généreux', de l'exemple 36, qui fait partie de la classe u-u.

(36)

karum-a *Zaid-un*
être.généreux-3SG Zaid-NOM
'Zaid est généreux'

Ce verbe peut avoir la variance *karama - yakrumu*, avec le sens «d'être plus généreux que ». La variance vocalique, de *karuma* à *karama*, permet une extension de «l'être généreux » à l'action de « compéter à être plus généreux ». Cette extension comporte aussi une variation syntactique, comme illustré dans l'exemple 37. Le verbe *karama* est transitif, avec le sujet « concurrent vainqueur » (*Zaid*) et le complément «second concurrent perdant » (*'aliyy*).

(37)

karam-a *Zaid-un* *'aliyy-an*
être.plus.généreux.que.PAST-3SG Zaid-NOM Ali-ACC
'Zaid a été **plus** généreux **que** Ali [après que ce dernier l'a défié]'

26. Dichy- 2007, pag. 354.

27. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I, pp. 70-71.

On ne doit pas confondre la signification complexe de *karama* avec le comparatif qui consiste, juste, à donner l'assertion que *Zaid est plus généreux de Ali* et s'obtient en utilisant l'adjectif superlatif de schème 'aC₁C₂aC₃, illustré dans l'exemple 38.

(38)

Zaid-un 'akram-u min 'aliyy-in
Zaid-NOM être.plus.généreux-NOM de Ali-GEN
'Zaid est plus généreux que Ali'

Cependant, on peut seulement noter l'existence de verbes de classe a-u, qui ne relèvent pas le sens « vaincre un défi », comme, *kataba - yaktubu* 'écrire'.

3.2.1.5 Formalisation de l'ablaut en langue arabe

Les verbes qui ont la voyelle médiane *u* à l'accompli sont formels et ont la même voyelle *u* à l'inaccompli, pour cause sémantique.

Les verbes qui ont la voyelle médiane *i*, à l'accompli, sont formels et ont la voyelle *a* à l'inaccompli. Il y a interférence phonologique qui impose une alternance a-a, dans le cas des verbes qui contiennent une consonne gutturale en C₂ ou C₃. Par contre, les trente verbes d'alternance i-i ne font pas une classe représentative et doivent être considérés comme des exceptions.

Le fait d'avoir une voyelle médiane *a* à l'accompli ne permet pas de prédire la voyelle de l'inaccompli, qui peut être *u* ou *i*. L'appartenance d'un verbe à une classe ou à une autre reste imprévisible comme par exemple *kataba - yaktubu* 'écrire' et *ḍaraba - yaḍrib-u* 'frapper'. Cette ambiguïté ne permet pas la formalisation des phénomènes apophoniques et il a toujours été nécessaire d'indiquer l'alternance vocalique dans le dictionnaire. Dans les dictionnaires arabes, la convention est d'indiquer la forme de l'inaccompli à côté de celle de l'accompli (par exemple : *kataba - yaktubu*). Dans un dictionnaire plurilingue ou numérisé, on indique la forme canonique de l'accompli, en ajoutant à côté la voyelle de l'inaccompli (par exemple : *kataba - u*).

En 1996, M. Guerssel et J. Lowenstamm ont été les premiers à proposer une solution de caractère applicatif qui permet de fournir les formes de l'inaccompli par la seule dérivation apophonique et sans qu'elles soient enregistrées dans le lexique.²⁸

Pour annuler l'ambiguïté des verbes en *a*, ils proposent de considérer la voyelle *a* des verbes de type *ḍaraba - yaḍrib-u* comme un zéro /ø/ sous-jacent. Pour le fait que la vocalisation de C₂ ne peut être nulle, il y aura automatiquement propagation de la première voyelle qui est toujours *a*. De cette manière, la dérivation de l'inaccompli, à partir

28. Guerssel et Lowenstamm, 1996.

de l'accompli, devient entièrement prévisible. Le "chemin apophonique" de caractère applicatif se dégage comme suite :

Chemin apophonique $\emptyset \Rightarrow i \Rightarrow a \Rightarrow u \Rightarrow u$

Ainsi, la voyelle de l'accompli, au niveau sous-jacent, donne toujours une et une seule voyelle, pour la forme inaccompli, comme le montre la figure 3.1, extraite du travail de M. Guerssel et J. Lowenstamm.

Une remarque fondamentale est la propriété formelle frappante des déclarations apophoniques, à savoir que chaque voyelle de sortie d'une apophonie est l'entrée pour une autre. D'autres travaux sur d'autres langues (le guèze, le kabyle berbère et l'allemand) ont montré que le chemin apophonique proposé pourrait être universel.

<p>a.</p> <p>1. $\emptyset \rightarrow i$</p> <p>2. $i \rightarrow a$</p> <p>3. $a \rightarrow u$</p> <p>4. $u \rightarrow u$</p>	<p>b.</p> <p><i>ḍar</i><u>\emptyset</u><i>b</i> <i>ya</i><u>δ</u><i>ri</i><u>b</u></p> <p><i>la</i><u>δ</u><i>is</i> <i>ya</i><u>δ</u><i>l</i><u>δ</u><i>bas</i></p> <p><i>ka</i><u>δ</u><i>tab</i> <i>ya</i><u>δ</u><i>kt</i><u>δ</u><i>ub</i></p> <p><i>ka</i><u>δ</u><i>bu</i><u>δ</u><i>r</i> <i>ya</i><u>δ</u><i>ku</i><u>δ</u><i>bu</i><u>δ</u><i>r</i></p>
---	---

Fig. 3.1 : Chemin apophonique proposé par Guerssel&Lowenstamm

3.2.2 Les verbes dérivés du verbe trilitère

3.2.2.1 Introduction

Selon al-'Astarabādī, la dérivation doit servir pour «ajouter » un sens à celui pré-existant;²⁹ tandis que Larcher précise, dans son article sur les verbes dérivés :

Pour les grammairiens arabes, comme, à leur suite, pour les grammairiens arabisants, les formes usuelles du verbe trilitère, dites « augmentées » par les premiers et « dérivées » par les seconds, ne constituent pas un ensemble inorganisé, mais au contraire un système structuré, tant sur le plan morphologique que sémantique et syntaxique.³⁰

29. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol. I, page 83.

30. Larcher, 1999.

La langue arabe utilise quatorze schèmes verbaux dérivant du verbe de base trilitère, que les grammairiens arabisants énumèrent de II à XV (le verbe trilitère étant la forme I). Le tableau 3.19 présente la fréquence des verbes arabes, existants dans le dictionnaire «*A Dictionary of Modern Written Arabic*»,³¹ calculée par John McCarthy et Alan Prince.³² On note que la forme IX est très rare, tandis que les formes de XI à XV sont presque inexistantes en Arabe moderne.

Forme	Schème	Fréquence
I	C ₁ aC ₂ vC ₃ -a	2569
II	C ₁ aC ₂ C ₂ aC ₃ -a	1398
III	C ₁ āC ₂ aC ₃ -a	463
IV	'aC ₁ C ₂ aC ₃ -a	951
V	taC ₁ aC ₂ C ₂ aC ₃ -a	1025
VI	taC ₁ āC ₂ aC ₃ -a	394
VII	(i)nC ₁ aC ₂ aC ₃ -a	260
VIII	(i)C ₁ taC ₂ aC ₃ -a	621
IX	(i)C ₁ aC ₂ aC ₃ C ₃ -a	18
X	(i)staC ₁ C ₂ aC ₃ -a	389
XI	(i)C ₁ C ₂ āC ₃ C ₃ -a	2
XII	(i)C ₁ C ₂ awC ₂ C ₃ -a	7
XIII	(i)C ₁ C ₂ awwaC ₃ -a	0
XIV	(i)C ₁ C ₂ anC ₃ aC ₃ -a	2
XV	(i)C ₁ C ₂ anC ₃ ā	0

Tab. 3.19 : Fréquence des verbes dérivant du verbe trilitère

En ce qui concerne l'Arabe Classique, on peut noter que al-'astarabādī a dédié un chapitre pour chaque forme dérivée de II à X, par contre, il a consacré seulement quelques lignes pour les formes XI à XV, en parlant des « autres formes dérivées » (*bāqī aṣ-ṣiyāg*),³³ ce qui fait suggérer qu'ils n'étaient pas aussi fréquents que les autres.

Aspects morphologiques des verbes dérivés La dérivation verbale se fait grâce à l'un des trois processus morphologiques, la duplication du deuxième ou du troisième

31. Wehr, 1971.

32. McCarthy & Prince, 1990, p. 33; Danks, 2010, p. 44

33. 'al-'Astarabādī, *Šarḥ aṣ-ṣāfiyyat*, Vol. I page 112.

radical ; ou l'augment de la quantité de la première voyelle ou l'adjonction d'un affixe juste avant (préfixe) ou juste après (infixe) le premier radical .

Dans le cas de la duplication du deuxième radical, on obtient $C_1aC_2C_2aC_3-a$ (la forme II), par exemple, *farraḥa* 'rendre qn. heureux' de *faraha* 'être heureux' et *kassara* 'casser en petits morceaux' de *kasara* 'casser'.

La forme III, $C_1āC_2aC_2a$; s'obtient grâce à l'allongement de la première voyelle, par exemple, du verbe *kataba* 'écrire' on obtient *kātaba* 'écrire à qn.'

L'adjonction du préfixe 'a permet d'obtenir la forme IV, $'aC_1C_2aC_2a$, c'est le cas de 'aḡlasi 'faire assoir qn.' dérivant du verbe ḡlasi 's'asseoir' et 'abḥala 'trouver qn. avare' de baḥula 'être avare'.

L'ajout du préfixe ta permet d'obtenir $taC_1aC_2C_2aC_3-a$ et $taC_1āC_2aC_2a$; respectivement forme V et forme VI. On peut noter que ces deux formes sont liés, morphologiquement et respectivement, à $C_1aC_2C_2aC_3-a$ et $C_1āC_2aC_2a$. Ainsi du verbe *kasara* 'casser', on obtient *kassara* 'casser en petits morceaux' et *takassara* 'être cassé en petits morceaux'. Par contre, du verbe *kataba* 'écrire', on obtient *kātaba* 'écrire à qn.' et *takātaba* 'correspondre'.

L'ajout d'un autre préfixe, *sta*, permet d'obtenir la forme X, $(i)staC_1C_2aC_2a$. Larcher fait noter qu'il est possible de construire, en diachronie, une relation entre $'aC_1C_2aC_2a$ et $(i)staC_1C_2aC_2a$.³⁴ En plus, en analysant quelques exemples, on note une relation sémantique entre ces deux formes, par exemple, du verbe de base, *salima* 'être sain et sauf', dérivent *aslama* 'consigner qc. intacte' et *istaslama* 'se rendre'.

La langue arabe a deux autres formes dérivées, la forme VIII, $(i)C_1taC_2aC_3-a$, également «une forme en t» mais, l'adjonction de *ta* se fait comme infixe. Enfin, la forme VII, $(i)nC_1aC_2aC_2a$, qui s'obtient par l'ajout du préfixe *n*. En analysant quelques exemples, on peut déduire qu'il y a une parenté sémantique entre ces deux formes et la forme I, de laquelle ils dérivent directement ; du verbe 'azala 'séparer qn. ou qc.', on peut obtenir *i'tazala* 'rester à l'écart ; abandonner' et *in'azala* 'être séparé'.

En général, toutes les formes verbales, citées ci dessus, dérivent d'un verbe de base, et dans ce cas, on peut parler de dérivation inter-catégorielle.

Selon Larcher, on peut les classer en deux groupes, présentés dans le tableau 3.20. Le premier groupe est constitué des verbes qui dérivent directement de la forme de base, appelés *les formes augmentées basiques*.

34. Larcher, 2003, page 45.

Le deuxième groupe contient les *formes surdérivées* qui dérivent d'une forme déjà dérivée en ajoutant le préfixe *ta*.³⁵

Forme I	Verbes dérivés		Verbes surdérivés	
	forme	schème	forme	schème
C ₁ aC ₂ vC ₃ -a	II	C ₁ aC ₂ C ₂ aC ₃ -a	V	taC ₁ aC ₂ C ₂ aC ₃ -a
	III	C ₁ āC ₂ aC ₃ -a	VI	taC ₁ āC ₂ aC ₃ -a
	IV	'aC ₁ C ₂ aC ₃ -a	X	(i)staC ₁ C ₂ aC ₃ -a
	VII	(i)nC ₁ aC ₂ aC ₃ -a		-
	VIII	(i)C ₁ taC ₂ aC ₃ -a		-

Tab. 3.20 : Paradigmes des verbes dérivés et surdérivés

Le seul cas de dérivation trans-catégorielle se présente avec la forme IX qui se forme grâce à la duplication du troisième radical, (i)C₁C₂aC₃C₃-a. La forme IX dérive d'un adjectif, exclusivement, de couleur ou de handicap et aboutit à un verbe inchoatif dont le sens est *devenir+adjectif*. Par exemple :

إِبْيَضَ *ibyaḍḍa* 'devenir blanc' dérive de l'adjectif 'abyaḍ 'blanc'

إِعْوَجَّ *i'waġġa* 'devenir courbé' dérive de l'adjectif 'a'waġ 'courbé'.

Conclusion À part la forme (i)C₁C₂aC₃C₃-a, qui est dénomminative, tous les autres verbes dérivés présentent des parentés morphologiques et sémantiques.

Les sections suivantes permettent d'approfondir les relations entre verbe de bases et formes dérivés. La classification ne suivra pas la numérisation occidentale mais, plutôt, une vision sémantique :

- les verbes dérivés directement du verbe de base : 'aC₁C₂aC₃-a ; C₁aC₂C₂aC₃-a et C₁āC₂aC₂a ;
- les verbes surdérivés des trois formes précédentes ;
- les formes qui dérivent directement du verbe trilitère ;
- les autres formes qui sont rares.

35. Larcher, 2003, page 45.

3.2.2.2 La forme IV : 'aC₁C₂aC₃-a

3.2.2.2.1 Valeur causative En comparant le sens des verbes de base, dans les exemples 39a et 40a, avec celui des verbes de schème 'aC₁C₂aC₃-a, dans les exemples 39b et 40b, on peut conclure que ces derniers sont causatifs.

(39)

- | | | |
|---|------------------------------------|------------------------------------|
| <p>a <i>ğalas-a</i>
savoir_{.PAST-3SGM}
'Zaid s'est assis'</p> | <p><i>zayd-un</i>
Zaid-NOM</p> | |
| <p>b 'ağlas-a
faire.s'assoir_{.PAST-3SGM}
'Omar a fait assoir Zaid'</p> | <p>'omar-un
Omar-NOM</p> | <p><i>zayd-an</i>
Zaid-ACC</p> |

(40)

- | | | |
|--|------------------------------------|--|
| <p>a 'alim-a
savoir_{.PAST-3SGM}
'Zaid connaît la nouvelle'</p> | <p><i>zayd-un</i>
Zaid-NOM</p> | <p><i>al=ğabar-a</i>
DEF=nouvelle-ACC</p> |
| <p>b 'a lam-a
faire.savoir_{.PAST-3SGM}
'Omar a fait connaître [à] Zaid la nouvelle'</p> | <p>'omar-un
Omar-NOM</p> | <p><i>zaid-an</i>
Zaid-ACC</p> <p><i>al=ğabar-a</i>
DEF=nouvelle-ACC</p> |

La causation a été définie par al='astarabādī :³⁶

Sache que la valeur la plus fréquente de 'af'ala est de rendre transitif le verbe trilitère. Cela consiste à faire de ce qui est le sujet du verbe intransitif l'objet d'un *faire* virtuel, cet objet restant le sujet du procès de base, car 'adhābtu zaidan signifie ġa 'altu zaydan dāhiban [j'ai fait partir Zaid]: Zaid est l'objet de la valeur *faire* marquée par la *hamza* et le sujet de *partir* comme il l'était dans *dahaba zaidun* [Zaid est parti].³⁷

Elle a été définie, également, par Alain Christol :³⁸

La diathèse causative change le rôle associé à un actant. À partir d'une proposition organisée autour d'un verbe V1 et de ses actants A (B, C), elle introduit un actant extérieur X, actant non requis par V1 et dont le référent

36. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol. I ; p. 86.

37. Traduction de Pierre Larcher Larcher, 2003, Page 64.

38. CHRISTOL, 2004.

dispose d'un pouvoir suffisant pour faire agir un être ou modifier l'état d'un objet ; cet actant, dit causateur (angl. causer) a un rôle d'initiateur, sans être nécessairement un véritable agent. La relation entre X et l'actance (A + V1) est assurée par un opérateur causatif (V2) dont les actants sont X et la proposition [A + V1].

Dans le cas de la forme 'aC₁C₂aC₃-a, l'ajout du préfixe 'a donne au verbe un sens de *ġa'* 'faire/rendre'. On obtient ainsi un verbe causatif, qui introduit un nouveau participant, le causateur, avec le rôle syntaxique du sujet. Du point de vue sémantique, le causateur est l'initiateur de l'action exprimée par le verbe de base. Le deuxième élément, le causataire est initialement sujet de la construction non-dérivée. Dans la construction causative, le causataire devient objet mais, en fait, il est toujours sujet de l'action de base. Selon al-'astarabādī, le causataire devient accusatif parce qu'il est objet du concept du *ġa'* 'faire ; causer' et donc il est *maġ'ūl* qu'on peut traduire, justement, par «causataire ». ³⁹

3.2.2.2 Valeur privative Pour expliquer cette valeur, ibn ġinī donne l'exemple de 'a *ġama l-kitāba* 'il a explicité/éclairci le livre'. Le verbe 'a *ġama* a le sens de 'expliciter et éclairer le texte [en mettant les points diacritiques]'. Par contre, la racine 'ġm, de laquelle il dérive, a un sens primaire de 'uġmah et de 'ibhām 'obscurité'. Le verbe dérivé 'a *ġama* a le sens de *salb* 'privation' et *nafy* 'suppression', c'est à dire qu'il sert pour supprimer l'*obscurité* exprimée par le substantif 'uġmah 'obscurité'. ⁴⁰

Par contre, al-'astarabādī ajoute :

['af'ala sert] pour la suppression, dans le sens que tu l'utilises quand tu enlèves au complément d'objet de 'af'ala ['aC₁C₂aC₃-a] la chose de laquelle ce dernier a été dérivé, comme, 'a *škay-tu-hu*, c'est à dire, 'a *zal-tu* 'j'ai éliminé' *šakwā-hu* 'sa plainte'. ⁴¹

3.2.2.3 Valeur expositive Dans l'exemple 41a, le verbe de base, *qatala*, a le sens de 'tuer'. Dans le cas du verbe dérivé, 'a *qatala*, de l'exemple 41b, le sujet (Zaid) «a exposé » l'objet (Omar) à être tué [par quelqu'un d'autre], qu'il soit vraiment tué ou pas.

39. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol.I, pp. 86-87 ; 96.

40. Ibn Ğinnī, *sirr*, V. I, p. 37.

41. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol. I ; p. 91.

(41)

a *qatal-a* *zayd-un* *'omar-an*
tuer._{PAST-3SGM} Zaid-NOM Omar-ACC
'Zaid a tué Omar'

b *'aqtal-a* *zayd-un* *'omar-an*
'Zayd a exposé Omar à être tué [par quelqu'un d'autre]'

Dans ce cas, al=*'astarabāḡī* parle de *ta'riḡ* 'exposition' de l'objet du verbe dérivé à l'action du verbe de base qu'il la subisse vraiment ou pas.⁴² Selon Larcher, cette valeur correspond au verbe factif en français.⁴³

3.2.2.2.4 Valeur estimative Dans le cas où *'aC₁C₂aC₃-a* dérive d'un verbe statif, elle a le sens de *trouver* l'objet dans l'état exprimé par le verbe de base, comme illustré par les exemples du tableau 3.21.

verbe de base trilitère	verbe dérivé de forme IV
<i>baḡul-a</i>	<i>'abḡal-a=hu</i>
être.avare. _{PAST-3SGM}	trouver.avare. _{PAST-3SGM} =lui
'il est avare'	'il le trouve avare'
<i>ḡabun-a</i>	<i>'aḡban-a-hu</i>
être.lâche. _{PAST-3SGM}	trouver.lâche. _{PAST-3SGM} =lui
'il est lâche'	'il le trouve lâche'

Tab. 3.21 : Exemple de verbes de forme IV avec valeur estimative

3.2.2.2.5 Valeur du verbe trilitère C'est le cas de plusieurs verbes, par exemple :

- *ḡazana* et *'aḡzana* dont le sens est 'attrister ; rendre quelqu'un triste' ;
- *saru'a* et *'asra'a* dont le sens est 'être rapide ; faire rapidement'.

Cependant, al=*'astarabāḡī* fait noter que *'aḡzana* dérive d'un autre verbe, *ḡazina* 'être triste' et qu'il transmet le sens de 'devenir triste'. Par contre, *ḡazana* est un verbe d'action dont le sens est 'transmettre la tristesse'.

42. 'al-'Astarabāḡī, *Šarḡ aš-šāfiyyat*, Vol. I ; p. 88.

43. Larcher, 2003, Page 66.

Il ajoute que *saru* 'a est intransitif et transmet le sens 'être rapide' comme état inné, tandis que *'asra* 'a ne contient pas le sens inné.⁴⁴

En plus, *'asra* 'a est toujours lié à la préposition *fī* 'à' et a le sens de 'être rapide à [faire quelque chose]; faire rapidement; se hâter', qui le rapproche à une classe de verbes 'aC₁C₂aC₃-a que Larcher appelle adverbiaux.⁴⁵

3.2.2.2.6 Valeur adverbiale Ces verbes ont le sens de «faire + adverbe », dont quelques exemples sont illustrés dans le tableau 3.22.

verbe de base		verbe dérivé	
<i>kaṭur-a</i>	'être nombreux'	<i>'akṭar-a</i>	'faire beaucoup'
<i>qall-a</i>	'être peu'	<i>'aqall-a</i>	'faire peu'
<i>ḥasun-a</i>	'être beau'	<i>'aḥsan-a</i>	'faire bien'

Tab. 3.22 : Exemples de verbes avec ajout d'une valeur adverbiale

Dans certains cas, le verbe de schème 'aC₁C₂aC₃-a dérive directement d'un nom de lieu ou de temps, dont certains exemples sont illustrés dans le tableau 3.23.

Adverbe		Verbe dérivé	
<i>ṣabāḥ</i>	'matin'	<i>'aṣbāḥ-a</i>	'être/devenir le matin'
<i>masā</i>	'soir'	<i>'amsā</i>	'être/devenir le soir'

Tab. 3.23 : Exemples de verbes dérivant d'un nom de temps ou de lieu

3.2.2.3 La forme II : C₁aC₂C₂aC₃-a

3.2.2.3.1 Pluriactionnalité ou pluralité verbale Selon Peter Lasersohn, les marqueurs verbaux de la pluriactionnalité s'attachent au verbe, pour indiquer une multiplicité d'actions qui peut impliquer plusieurs participants ou une dispersion spatiale (spatial scattering) ou une répétition dans le temps (temporal iteration).⁴⁶

La grammaire traditionnelle, aussi, fait ces distinctions. Par exemple, *al-'astarabādī* fait noter que la plupart des verbes, de schème C₁aC₂C₂aC₃-a, ont un sens de *multiplier le sujet de l'action* (*li=taktīri fā 'ili=hi*).⁴⁷

44. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol. I; p. 87.

45. Larcher, 2003, p. 72.

46. Lasersohn, 1995, Chapitre 13, pages 238-266

47. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol. I; pages 92-93.

En utilisant le verbe *māta*, dérivant de la racine *mwt*, dans l'exemple 42a, le sens est tout simplement 'mourir'. Dans l'exemple 42b, par contre, l'utilisation du verbe *mawwata* souligne la multiplicité du sujet.

(42)

<p>a <i>māt-at</i> <i>al='ibil-u</i> mourir_{.PAST-3SGF} les-chameaux-NOM 'les chameaux sont morts'</p>	<p>b <i>mawwat-at</i> <i>al='ibil-u</i> mourir_{.PAST-3SGF} les-chameaux-NOM 'beaucoup de chameaux sont morts'</p>
--	---

Cependant, des exemples qu'il cite, on déduit que la multiplicité ne concerne pas seulement le sujet.

Dans les exemples 43a, 'j'ai égorgé la brebis', et 44a, 'j'ai fermé la porte', il n'y a aucun sens de multiplicité. Du fait, on ne peut pas utiliser la forme II et, les phrases 43b et 44b ne sont pas correctes.

(43)

<p>a <i>ḡabaḥ=tu</i> <i>š=šāt-a</i> égorger_{.PAST=1SG} la-brebis-ACC 'j'ai égorgé la brebis'</p>	<p>b *<i>ḡabbaḥ=tu</i> <i>š=šāt-a</i></p>
--	--

(44)

<p>a <i>ḡalaq=tu</i> <i>al=bāb-a</i> fermer_{.PAST=1SG} la-porte-ACC 'j'ai fermé la porte'</p>	<p>b *<i>ḡallaq=tu</i> <i>al=bāb-a</i></p>
---	---

Par contre, dans les exemples, 45a, 'j'ai égorgé les brebis', et 46a, 'j'ai fermé les portes', l'objet est pluriel. Dans ce cas, l'utilisation de la forme II, dans 45b et 46b, dénote la multiplicité de l'objet.

(45)

<p>a <i>ḡabaḥ=tu</i> <i>l=ḡanam-a</i> égorger_{.PAST=1SG} les-brebis-ACC 'j'ai égorgé les brebis'</p>	<p>b <i>ḡabbaḥ=tu</i> <i>l=ḡanam-a</i> égorger_{.PAST=1SG} les-brebis-ACC 'j'ai égorgé beaucoup de brebis'</p>
--	---

(46)

<p>a <i>ḡalaq=tu</i> <i>al='abwāb-a</i> fermer_{.PAST=1SG} les-portes-ACC 'j'ai fermé les portes'</p>	<p>b <i>ḡallaq=tu</i> <i>al='abwāb-a</i> fermer_{.PAST=1SG} les-portes-ACC 'j'ai fermé beaucoup de portes'</p>
--	---

Cusic souligne deux types de répétition de l'action : i) la répétition interne à l'évènement qui est unique et se déroule en *une seule occasion* mais il est constitué de

phases internes répétées ; ii) la répétition externe à l'évènement qui est complet mais se répète en *plusieurs occasions*.⁴⁸

En arabe, les verbes de forme II peuvent dénoter la répétition mais, sans indication si elle est interne ou externe. Cependant, al-'astarabādī interprète le verbe *ǧawwala* comme *taktīr* 'augmentation' ou 'répétition' du verbe *ǧāla* 'déambuler' et donc *ǧawwala* signifie 'déambuler plusieurs fois ; vagabonder'.

Il explique aussi, dans le cas du Qur'ān, qu'on n'a pas utilisé le verbe *nazala* 'descendre', parce qu'il n'a pas été révélé en un seul coup ; mais on utilise le verbe *nazzala*, qui signifie 'révéler les versets du Qur'ān, l'un après l'autre'.⁴⁹

3.2.2.3.2 Causativité Comme déjà vu pour 'aC₁C₂aC₃-a, le causatif ajoute un participant et fait augmenter la valence verbale.

En analysant les exemples 47a et 47b, le verbe *farraḥa* 'rendre heureux' est causatif transitif et dérive du verbe d'état intransitif, *fariḥa* 'être heureux'.

(47)

a	<i>fariḥa</i>		<i>zaid-un</i>
	être.heureux _{PAST-3SGM}		Zaid-NOM
	'Zaid est heureux'		
b	<i>farraḥa</i>	'omar-un	<i>zaid-an</i>
	rendre.heureux _{PAST-3SGM}	Omar-NOM	Zaid-ACC
	'Zaid a rendu Omar heureux'		

Par contre, d'un verbe bivalent comme 'alima 'savoir', illustré dans l'exemple 48a, on obtient le verbe trivalent 'allama 'faire savoir ; enseigner', de l'exemple 48b.

(48)

a	'alim-a		<i>zayd-un</i>	<i>al=ḥabar-a</i>
	savoir _{PAST-3SGM}		Zaid-NOM	la=nouvelle-ACC
	Zaid a su la nouvelle			
b	'allam-a	'omar-un	<i>zaid-an</i>	<i>al=ǧabr-a</i>
	enseigner _{PAST-3SGM}	Omar-NOM	Zaid-ACC	la=algèbre-ACC
	'Omar a enseigné [à] Omar l'algèbre			

Dans les exemples 47b et 48b, le causateur, Omar, a le rôle syntaxique du sujet. Par contre, le sujet de la construction non dérivée, Zaid, devient causataire, destitué au rôle syntaxique d'objet.

48. Voisin&Vittrant, 2012.

49. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, Vol. I ; pages 92-93.

Différence entre 'aC₁C₂aC₃-a et C₁aC₂C₂aC₃-a Selon la théorie de la « motivation iconique » proposée par Haiman,⁵⁰ lorsque plusieurs constructions causatives sont en concurrence, c'est la construction la plus compacte qui sera utilisée pour marquer la causation la plus directe et effective. Ainsi, plus la distance de séparation augmente entre les expressions linguistiques, plus les concepts que ces dernières représentent sont indépendants entre eux.

Haiman établit une échelle où le degré de la séparation linguistique varie en fonction des limites entre les mots et les morphèmes :

- a- X#A#Y où # est le séparateur entre mots différents ;
- b. X#Y ;
- c. X+Y où + représente la limite entre morphèmes différents au sein du même mot ;
- d. Z où les morphèmes non identifiables.

Dans les représentations a et b, où le symbole # souligne la limite entre deux mots, les deux morphèmes X et Y représentent deux paroles séparées. On parle d'une représentation analytique.

Cependant, dans la représentation a, les deux paroles X et Y sont séparées par une autre parole A, alors que, dans la représentation b, elles sont adjacentes. De conséquence, la distance entre X et Y est plus grande dans a que dans b.

Dans la représentation c, le symbole + représente la limite entre les morphèmes X et Y au sein du même mot, on parle dans ce cas de représentation synthétique.

La représentation d est également synthétique mais encore plus compact parce qu'on ne peut pas identifier les morphèmes qui composent la parole Z.

La forme 'aC₁C₂aC₃-a peut être considérée de type c, où le morphème de la causativité, 'a, est bien distinct. Par contre, C₁aC₂C₂aC₃-a peut être considérée une forme « plus compacte ». Selon cette distinction, on s'attend que la causation transmise par 'aC₁C₂aC₃-a soit moins effective que celle transmise par C₁aC₂C₂aC₃-a.

En effet, les verbes comme 'aqtala 'faire tuer' 'abā'a 'faire vendre' et 'asqā 'faire boire' expriment surtout *at-ta'riḍ* 'l'exposition' à l'action, dans le sens que l'objet est exposé à l'action, mais il peut la subir ou no et le résultat peut ne pas advenir.⁵¹

Par contre, la forme C₁aC₂C₂aC₃-a est complexe ; en plus de la causativité, elle dénote la répétition et *at-tašdīd* 'l'intensification' de l'action et, un résultat effectivement avéré et incontestable.

50. Haiman, 1983

51. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I, p. 88.

Un autre aspect, qui diffère les deux formes, est illustré par la paire de verbes *'al-lama* et *'a'lama*, tous les deux causatifs avec le sens de 'faire savoir'. Cependant, il y a une grande différence entre les deux, comme expliqué magistralement par az-Zabīdī (m. 1205/1790) :⁵²

[...] وَهُوَ صَرِيحٌ فِي أَنَّ التَّعْلِيمَ وَالْإِعْلَامَ شَيْءٌ وَاحِدٌ، [...] إِلَّا أَنَّ الْإِعْلَامَ إِحْتَصَّ بِمَا كَانَ بِإِخْبَارٍ سَرِيعٍ، وَالتَّعْلِيمَ إِحْتَصَّ بِمَا يَكُونُ بِتَكَرُّرٍ وَتَكْثِيرٍ

[...] Il a raison [al-Fīrūz'abādī] que *at-ta'lim* [maṣdar de *'allama*] et *al-'i'lām* [maṣdar de *'a'lama*] est la même chose [...]. Cependant, *al-'i'lām* est spécifique de ce qui est une information rapide et *at-ta'lim* est spécifique à [l'information] qui est de manière répétitive et intensive.

Donc, *'allama* dénote l'intensité et la répétition de l'action, par contre, *'a'lama* dénote la rapidité de l'action. L'intensité de l'action dénotée par les deux verbes se répercute sur le résultat causé.

Le résultat de *'a'lama* 'informer', dans l'exemple 49, est *'alima* 'savoir'.

Par contre, la conséquence de *'allama* 'enseigner', dans l'exemple 50, est *ta'allama* 'apprendre'.

(49)

'a lam=tu=hu *l=ḥabar-a* *fa='alim-a*
 informer.PAST=1SG=lui la=nouvelle-ACC donc=savoir.PAST-3SGM
 'je l'ai informé [de] la nouvelle donc il sait'

(50)

'allam=tu=hu *l=kitābat-a* *fa=ta'allam-a*
 enseigner.PAST=1SG=lui la=algèbre-ACC donc=apprendre.PAST-3SGM
 'je lui ai enseigné l'écriture donc il a appris'

3.2.2.3.3 Valeur dénominate C₁aC₂C₂aC₃-a peut être dénominate. Dans ce cas, elle dénote le sens de traiter, supprimer ou collecter l'objet duquel elle dérive.

Les exemples du tableau 3.24 en illustrent quelques exemples.

3.2.2.4 La forme III : C₁āC₂aC₃-a

3.2.2.4.1 La participation Les exemples 51a, 51b et 51c, illustrent certaines différences entre le verbe *ǧalasa* 's'asseoir' et le verbe dérivé, *ǧālasa*.

52. Zabīdī, Tāǧ al-'arūs, V. 33 ; p. 128.

taxique, est le sujet, Zaid. Par contre l'autre agent, Omar, est accusatif pour souligner sa participation implicite.⁵³ Dans ce cas, le *mušāarak* 'participant' (Omar) ne coïncide pas avec le complément d'objet (*al=ḥabl* 'la corde') qui subit l'action du verbe de base. De conséquence, le verbe *ḡāḍaba* est bi-transitif.

Dans le cas où le verbe de base est transitif, le verbe $C_1\bar{a}C_2aC_3$ -a acquiert deux arguments : le *maf'ūl* 'complément d'objet' de l'action de base et le *mušāarak* 'le participant'.

(52)

- (a) *ḡāḍab-a* *zaid-un* *al=ḥabl-a*
 tirer._{PAST-3SGM} Zaid-NOM la=corde-ACC
 'Zaid a tiré la corde'
- (b) *ḡāḍab-a* *zaid-un* *'omar-an* *al=ḥabl-a*
 tirer.avec._{PAST-3SGM} Zaid-NOM Omar-ACC la=corde-ACC
 'Zaid tirait avec Omar la corde [entre eux]'

Dans l'exemple 53b, le verbe *ḍārabā* désigne, également, la *mušāarakah* 'participation' de deux agents dans l'évènement, où Zaid est sujet et Omar est l'agent accusatif. Cependant, Omar est, en même temps, *maḍrūb* 'frappé' et *mušāarak* 'participant'. Dans ce cas, la valence verbale reste la même que celle du verbe de base, *ḍaraba*, parce que les deux compléments coïncident.

(53)

- a *ḍarab-a* *zayd-un* *'omar-an*
 frapper._{PAST-3SGM} Zaid-NOM Omar-ACC
 'Zaid a frappé Omar'
- b *ḍārab-a* *zayd-un* *'omar-an*
 combattre.contre._{PAST-3SGM} Zaid-NOM Omar-ACC
 échanger.des.coups.avec._{PAST-3SGM} /
 'Zaid s'est battu contre Omar' ou 'Zaid a échangé des coups avec Omar'

3.2.2.4.2 L'exagération Dans certains cas, $C_1\bar{a}C_2aC_3$ -a dénote le même sens du verbe de base, cependant, elle ajoute le sens d'effort et de volonté du sujet que *al=astarabādī* appelle *mubālaḡah* 'exagération'.⁵⁴

53. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I ; p. 96-97.

54. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I, p. 99.

Cette valeur emphatique se note dans des verbes, comme *ṭālaba* 'exiger' et *tāba* 'a 'poursuivre' qui dérivent, respectivement, de *ṭalaba* 'demander' et *tabi* 'a 'suivre'.

3.2.2.5 Les formes dérivées en t

3.2.2.5.1 La forme V : taC₁aC₂C₂aC₃-a Principalement, taC₁aC₂C₂aC₃-a exprime le résultat de l'action C₁aC₂C₂aC₃-a et, sur le plan syntaxique, la valence du verbe diminue d'un actant.

Le sujet de taC₁aC₂C₂aC₃-a est l'objet qui a subi l'action de C₁aC₂C₂aC₃-a. C'est le cas, par exemple, du verbe intransitif *taqattā* 'a 'être découpé' qui exprime le résultat du verbe transitif *qattā* 'a 'découper quelque chose'.

En grammaire classique, on dit que le verbe taC₁aC₂C₂aC₃-a est *muṭāwi* 'qui se laisse influencer/flexible', défini par al='astarabādī :⁵⁵

Le *muṭāwi* ne signifie pas intransitif, comme on croit. Dans leur terminologie [i.e. celle des grammairiens], la *muṭāwa* 'ah est le fait de subir un effet et de recevoir l'effet de l'acte, que [le verbe exprimant] l'effet subi soit transitif, comme '*allamtu-hu l-fiqha fa-ta* 'allama-hu 'je lui ai enseigné la jurisprudence et il l'a apprise', c'est à dire il a reçu l'enseignement, car l'enseignement est l'effet exercé et l'apprentissage est l'effet subi et la réception de cet effet ; et ce verbe est transitif comme on voit ; ou qu'il soit intransitif comme *kasartu-hu fa-nkasara* 'je l'ai cassé et il s'est trouvé cassé', c'est à dire il a subi l'effet de la casse.⁵⁶

Il ajoute que le *muṭāwi* est, en réalité, l'objet qui est devenu sujet. Dans l'exemple, '*allam=tu zaid-an fa-ta* 'allama 'j'ai enseigné Zaid et il a appris', le vrai *muṭāwi* de '*allama* est le sujet de *ta* 'allama, c'est à dire Zaid. Le verbe *ta* 'allama, par contre, est appelé *muṭāwi* par extension.

Dans certains cas, le sens de taC₁aC₂C₂aC₃-a se rapproche de celui du verbe trilitère, comme c'est le cas de '*alima* 'savoir', '*allama* 'enseigner', '*ta* 'allama 'apprendre et donc savoir'.

3.2.2.5.2 La forme VI : taC₁āC₂aC₃-a Le verbe *ḡāḡab-a*, de schème C₁āC₂aC₃-a, dans l'exemple 52b, dénote la réciprocité implicite où la relation est asymétrique entre le sujet X et l'objet Y.

55. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I, p. 103.

56. Traduction Larcher Larcher, 2003, p. 75.

Par contre, la construction avec le verbe *tağāḍaba*, de schème $taC_1\bar{a}C_2aC_3-a$, dans l'exemple 54, exprime la réciprocité explicite où X et Y sont tous les deux sujets.

(54)

tağāḍab-a *zaid-un* *wa=ʿomar-un* *al=ḥabl-a*
 se tirer._{PAST-3SGM} Zaid-NOM Omar-NOM la=corde-ACC
 'Zaid et Omar se tiraient [entre eux] la corde'

Selon *al=ʿastarabādī*, avec le verbe $C_1\bar{a}C_2aC_3-a$, la séparation entre la fonction du sujet et celle de l'objet est morphologique et la pluralité est sémantique (implicite). Par contre, avec le verbe $taC_1\bar{a}C_2aC_3-a$, la pluralité est morphologique mais, sémantiquement, chacun des actants est sujet et objet, en même temps.⁵⁷

Dans certains cas, $taC_1\bar{a}C_2aC_3-a$ a le sens du *muṭāwiʿ* de $C_1\bar{a}C_2aC_3-a$, lorsque cette dernière assume un sens causatif comme *tabāʿad-a* 's'éloigner' qui dérive de *bāʿad-a* 'éloigner'. Dans d'autres cas, elle peut avoir un sens de simulation, par exemple, *tamāraḍa* 'faire semblant d'être malade' dérivant de *marīḍa* 'être malade'.

3.2.2.5.3 La forme X : (i)staC₁C₂aC₃-a Elle peut être le *muṭāwiʿ* de $'aC_1C_2aC_3-a$. Par exemple, le verbe *istaslama* 'se rendre' qui dérive de *'aslama* 'consigner intact quelqu'un ou quelque chose'.

Dans d'autres cas, il peut acquérir le sens de 'demander de faire'. Ce sens peut être sujet d'une interprétation selon que l'objet est animé ou inanimé.

Dans l'exemple 55a où le premier objet est animé (Omar), le verbe *istaktaba* a le sens de 'demander à Omar d'écrire'.

Mais, Dans l'exemple 55b, où l'objet est inanimé (la lettre), *istaktaba* a le sens de 'se faire écrire la lettre'. Avec un sujet inanimé, (i)staC₁C₂aC₃-a a le sens de 'se faire faire [par quelqu'un d'autre]'.

(55)

a *istaktab=tu* *ʿomar-an* *ar=rīsālat-a*
 demander.écrire._{PAST=1SGM} Omar-ACC la=lettre-ACC
 'j'ai demandé à Omar d'écrire la lettre'

b *istaktab=tu* *ar=rīsālat-a* [*min ṭaraf-i ʿomar-in*]
 se.faire.écrire._{PAST-1SGM} la=lettre-ACC [de part-GEN Omar-GEN]
 'J'ai fait écrire la lettre [de la part de Omar]'

57. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I, p. 101-102.

Les exemples 56, et 57 illustrent les utilisations et les sens des verbes *'aħraġa* et *istaħraġa* dérivant du verbe *ħaraġa* 'sortir'.

(56)

'aħraġ=tu *al=watid-a*
faire.sortir._{PAST=1SGM} le=piquet-ACC
'J'ai fait sortir le piquet'

(57)

a *istaħraġ=tu* *al-watid-a* [*min ʔaraf-i Zayd-in*]
faire.sortir._{PAST=1SGM} le=piquet-ACC [de part-GEN Zaid-GEN]
'J'ai fait sortir le piquet [par Zayd]'

b *istaħraġ=tu* *al-watid-a*
faire.sortir._{PAST=1SGM} le=piquet-ACC
'J'ai fait [l'effort de faire] sortir le piquet'

c *istaħraġ=tu* *zaid-an*
demander.sortir._{PAST=1SGM} Zaid-ACC
'J'ai demandé à Zaid de sortir'

Dans l'exemple 56, le verbe *'aħraġa*, dérivant du verbe *ħaraġa* 'sortir', dénote le sens causatif 'faire sortir'. Dans l'exemple 57a, le verbe *istaħraġa* a le sens de 'faire faire sortir quelque chose [par quelqu'un d'autre]'. Dans l'exemple 57b, le sujet de *istaħraġa* est le même que celui de *'aħraġa* 'faire sortir'. Le verbe *istaħraġa*, dans ce cas, dénote le sens de 'se faire faire quelque chose [par soi même]' et exprime l'effort du sujet à compléter l'action. Cette interprétation est juste, dans le cas où l'objet est inanimé. Par contre, avec un objet animé, dans l'exemple 57c, le verbe *istaħraġa* a un sens unique 'demander à Zaid de sortir'.

3.2.2.6 Les formes VII et VIII : (i)nC₁aC₂aC₃-a et (i)C₁taC₂aC₃-a

Les deux formes dérivent directement de la forme trilitère et tous les deux expriment un sens *muʔāwi'* du verbe trilitère.

Cependant, il y a des contraintes morphologiques qui limitent la formation du verbe (i)nC₁aC₂aC₃-a et laissent la place seulement à (i)C₁taC₂aC₃-a. Ainsi, lorsque le premier radical est une des consonnes {*n; m; l; r; w*}, pour éviter le redoublement du premier radical avec le préfixe *n*, le verbe trilitère a seulement, (i)C₁taC₂aC₃-a comme *muʔāwi'*. C'est le cas, par exemple, des verbes *la'am-a* 'réparer; guérir' et *waṣal-a*

'arriver' qui ne peuvent donner **inla'am-a* et **inwaṣal-a*, mais, ils ont, respectivement, *ilta'am-a* 'être réparé ; guéri [blessure]' et *ittaṣal-a* 'être lié, se lier'.⁵⁸

D'autres contraintes plutôt sémantiques font la différence entre les deux formes :⁵⁹

- (i)nC₁aC₂aC₃-a est utilisé avec un sujet inanimé et (i)C₁taC₂aC₃-a avec un sujet animé. Dans ce cas, (i)nC₁aC₂aC₃-a est intransitif et a une valeur réflexive directe ; par contre, (i)C₁taC₂aC₃-a est transitif, avec un sens réfléchi moyen. Par exemple :

kašaf-a 'révéler' : *inkašaf-a* 'être révélé' et *iktašaf-a* 'découvrir' ;

qasam-a 'diviser ; partager' : *inqasam-a* 'être divisé en...' et *iqtasam-a* 'se partager [quelque chose]'.

- L'action (i)nC₁aC₂aC₃-a est faible, même avec un sujet animé. Par exemple :

inḥadar-a 'décliner ; déchoir' ; *inḥaraf-a* 'dévier ; se détourner de la bonne voie' et *inbaṭaḥ-a* 'se vautrer'.

Par contre, (i)C₁taC₂aC₃-a contient un sens de « s'élever » :

irtaqā 'monter ; progresser' ; *irtafa'-a* 'se lever'.

- Le verbe trilitère, *šawā* 'rôtir', de l'exemple 58, a le même le sujet du verbe *ištawā*, dans l'exemple 59. Dans ce cas, le sens du verbe (i)C₁taC₂aC₃-a est *faire pour soi* et le sujet est bénéficiaire de l'action de base.

(58)

šaway=tu *l=laḥm-a*
 rôtir_{PAST=1SGM} la=viande-ACC
 'J'ai rôti la viande'

(59)

ištaway=tu *l=laḥm-a*
 rôtir_{PAST=1SGM} la=viande-ACC
 'j'ai rôti la viande « pour moi »'

Par contre, dans l'exemple 60, le sujet du verbe *inšawā* est l'objet qui a subi l'action exprimée par le verbe de base, *šawā* 'rôtir'. Dans ce cas, le verbe dérivé, (i)nC₁aC₂aC₃-a, est le *muṭāwi'* du trilitère.

(60)

inšawā *l=laḥm-u*
rôtir_{PAST-3SGM} la=viande-NOM
'La viande est rôtie'

3.2.2.7 Les autres formes dérivées

Les formes IX, (i)C₁C₂aC₃C₃-a, et XI, (i)C₁C₂āC₃C₃-a dérivent toujours d'un adjectif, de schème 'aC₁C₂aC₃ qui exprime une couleur ou un défaut physique, et ont le sens inchoatif *devenir+adjectif*. Cependant, on ne trouve jamais la cooccurrence des deux formes. La forme IX se trouve en arabe classique. Par contre, la forme XI est, actuellement, la variante qu'on trouve dans le dialecte marocain.⁶⁰ Ci-dessous le tableau 3.25 en illustre quelques exemples.

Racine	Adjectif	Verbe dérivé utilisé en	
		arabe classique	dialecte marocain
'wr	'a'war	(i)'warra	'wār
	'borgne'	'devenir borgne'	
ḥmr	'aḥmar	(i)ḥmarra	ḥmār
	'rouge'	'devenir rouge'	

Tab. 3.25 : Exemples de verbes (i)C₁C₂aC₃C₃-a et (i)C₁C₂āC₃C₃-a

Il est important de noter que, dans l'arabe classique, l'adjectif 'a'war est lié à un verbe d'état 'awira 'être borgne'.

La forme XII, (i)C₁C₂awC₂aC₃-a, est caractérisée par la répétition de la deuxième consonne radicale (C₂) et l'insertion entre les deux C₂ de la consonne /w/. Elle peut ajouter un sens d'exagération à celui de la forme I, par exemple :

- *i'šawšaba* 'devenir riche d'herbe abondante [terre]' de 'ašiba 'devenir herbeux' ;
- *iḡdawdana* 'devenir très luxuriant [herbe]' ;
- *iḥdawdaba* 'devenir bosselé [route]'.

58. Larcher, 2003, p. 99.

59. Ibid., pp. 100-101.

60. Larcher, 2003, p. 120.

Trois autres formes sont très rares ou presque inexistantes. Selon al-'astarabādī, ce sont des formes « improvisées » et ne cite que quelques exemples :⁶¹

- (i)C₁C₂awwaC₃-a : *i'lawwaṭa* 's'élever' ; *iġlawwada* et *iḥrawwaṭa* 'se hâter' ;
- (i)C₁C₂anC₃aC₃-a : *iq'ansasa* 'avoir la poitrine saillante et le dos rentré' ;⁶²
- (i)C₁C₂anC₃ā : *iġrandā* 's'élever'.

3.2.3 Le verbe quadrilitère et ses dérivés

Le tableau 3.26 illustre quelques exemples de verbes quadrilitères et de leurs dérivés. En plus, il illustre leurs fréquences calculées par J. McCarthy et A. Prince, en se basant sur les données qui se trouvent dans le dictionnaire *A Dictionary of Modern Written Arabic*.⁶³

Forme	Schème	Fréquence	Exemples	
QI	C ₁ aC ₂ C ₃ aC ₄ -a	296	<i>daḥraġ-a</i>	'rouler quelque chose'
				'sur soi-même'
			<i>tamtam-a</i>	'murmurer ; bégayer'
QII	taC ₁ aC ₂ C ₃ aC ₄ -a	111	<i>tadaḥraġ-a</i>	'se rouler sur soi-même'
QIII	(i)C ₁ C ₂ anC ₃ aC ₄ -a	1	(i) <i>ḥranġam-a</i>	's'assembler'
QIV	(i)C ₁ C ₂ aC ₃ aC ₄ C ₄ -a	8	(i) <i>ṭma'anna</i>	'se tranquilliser'
			(i) <i>ḍmaḥalla</i>	'disparaître'

Tab. 3.26 : Fréquences et exemples du verbe quadrilitère et ses dérivés

On note que le verbe quadrilitère *muġarrad* 'non augmenté' a un schème unique C₁aC₂C₃aC₄-a (présentée comme *fa'lala* en grammaire classique) et a une fréquence (296) beaucoup plus faible que celle du verbe trilitère (2569 ; voir tableau 3.19).

Une grande partie des verbes quadrilitères est caractérisée par la duplication des deux premières radicales (C₁aC₂C₁aC₂-a), comme *tamtam-a* 'murmurer ; bégayer' et *ṭaṭṭaq-a* 'faire le bruit tac tac'.

La forme dérivée, taC₁aC₂C₃aC₄-a, est suffisamment fréquente, toujours intransitive et exprime le *muṭāwi'* de C₁aC₂C₃aC₄-a.

61. 'al-'Astarabādī, *Šarḥ aš-šāfiyyat*, V. I, p. 112-113.

62. Larcher, 2003, p. 121

63. McCarthy & Prince, 1990, p. 34; Danks, 2010, p. 44

Les deux autres formes (i)C₁C₂anC₃aC₄-a et (i)C₁C₂aC₃aC₄C₄-a sont plutôt rares mais il faut noter que les deux verbes donnés dans le tableau, *iṭma`anna* 'se tranquilliser' et *iḍmaḥalla* 'disparaître' sont très utilisés.

Chapitre 4

Le système nominal

4.1 Introduction

Les sections suivantes ne peuvent être exhaustives vue la vastité du sujet mais elles sont destinées à donner un aperçu sur le système nominal.

La définition du nom repose sur plusieurs critères, morphologiques, syntaxiques et sémantiques.¹

La catégorie des «noms » comprend les noms communs, les noms propres, les adjectifs, les démonstratifs, les pronoms, le numéraux, le noms d'action, et les participes actif et passif.

Les critères morphologiques, syntaxiques et sémantiques sont, également, utilisés pour différencier entre les sous-classes.² Cependant, les dissemblances entre les sous-classes nominales ne sont pas situées au même niveau que celles qui séparent le nom, le verbe et la particule.³

4.1.1 Critères morphologiques

Comme démontré dans le tableau 4.1, le nom est compatible avec :

- l'article défini, le préfixe *al*, invariable pour le genre et le nombre ;
- la désinence casuelle qui se présente comme une voyelle finale d'inflexion ;

1. Owens, 1980, pp. 125-128.

2. Owens, 1980, p. 130

3. Hamzé, 1994, p. 97.

- la marque de *tanwīn* 'indéfinitude', sous forme d'une consonne *n*, qui s'ajoute après la voyelle casuelle.⁴

Cas	désinence	Exemple	
		'la fille'	'une fille'
Nominatif	u	<i>al=bint-u</i> DEF=fille-NON	<i>bint-u-n</i> fille-NON-INDF
Accusatif	a	<i>al=bint-a</i> DEF=fille-ACC	<i>bint-a-n</i> fille-ACC-INDF
Génitif	i	<i>al=bint-i</i> DEF=fille-GEN	<i>bint-i-n</i> fille-GEN-INDF

Tab. 4.1 : Compatibilité morphologique du nom

Le nom est également compatible avec :

- l'infixe *-ay-* qui permet la formation du diminutif, par exemple, de la parole *bint* 'fille' dérive *bunayt* 'petite fille' ;
- Le suffixe *-iyy* qui permet la formation de l'adjectif relatif et des noms abstraits qui indiquent la qualité exprimée par l'adjectif. Par exemple, de la parole, '*insān* 'homme' ; on peut avoir, '*insāniyy* 'humain' et '*insāniyyah* 'humanité'.

4.1.2 Critères sémantiques

Le nom ne contient aucun sens du temps (passé, future ou présent) et, pour ce fait, au contraire du verbe, il n'est pas compatible avec les pronoms nominatifs et les affixes de flexion qui marquent le temps.

En outre, le verbe ne peut être que *musnad*, ou *muḥbar bihi* 'prédicat'.

Par contre, le nom peut être *musnad* '*ilayh*, ou *muḥbar* '*anh* 'prédicande', représentant la « chose » dont on parle. C'est le cas de *al=walad-u* 'le garçon', dans l'exemple 61, et de *al=asad-u* 'le lion', dans l'exemple 62.

Le nom peut être, aussi, *musnad* ou *ḥabar* 'prédicat', représentant l'information qu'on donne ; par exemple *ḥayawān-u-n* 'un animale' de l'exemple 62.

4. La langue arabe ne possède pas un article indéfini.

(61)

qabbal-a *al=walad-u* *al=bint-a*
embrasser._{PAST-3SGM} DEF=garçon-NON DEF=fille-ACC
'Le garçon a embrassé la fille'

(62)

al='asad-u *ḥayawān-u-n*
DEF=lion-NOM animal-NOM-INDF
'Le lion [est] un animal'

4.1.3 Critères syntaxiques

Dans une phrase verbale, le nom peut occuper la position du sujet et celle de l'objet, respectivement, *al=walad-u* 'le garçon' et *al=bint-a* 'la fille', dans l'exemple 61.

Par contre, dans une phrase nominale, il peut occuper aussi bien la position du *mubtada* 'prédicande' que du *ḥabar* 'prédicat', respectivement *al='asad-u* 'le lion' et *ḥayawān-u-n* 'un animale', dans l'exemple 62.

Le nom peut être concomitant avec le vocatif *yā* 'oh!', par exemple, *yā bint-u* 'Oh fille!'; les particules qui gouvernent le génitif, par exemple, la préposition *li=*, dans l'exemple 63; l'adjectif qualificatif (*na't*), le cas de *ḡamīl*, dans l'exemple 64; et il peut être remplacé par un pronom, par exemple, le suffixe *hu*, dans l'exemple 65.

(63)

li=l=bint-i *li=bint-i-n*
pour=DEF=fille-GEN pour=fille-GEN-INDF
'pour la fille' 'pour une fille'

(64)

al=bint-u *al=ḡamīl-at-u*
DEF=fille-NOM DEF=beau-F-NOM
'la belle fille'

(65)

katab=tu=hu
écrire._{PAST=1S.NOM=3SGM.ACC}
'Je l'ai écrit'

Il peut être en état d'annexion (*al-'iḏāfah*). Dans ce cas, le nom en état d'annexion (*muḏāf* 'annexé') ne peut porter ni l'article défini ni la marque de l'indétermination (*at-tanwīn*). Il peut occuper, aussi, la position du complément (*muḏāf 'ilayh*; 'annexé à lui') et, dans ce cas, il est toujours au génitif.

La détermination du nom en état d'annexion (*muḍāf* 'annexé') dépend de celle du complément du nom. Dans l'exemple 66, où le complément est déterminé, le nom, auquel il est annexé, est également déterminé mais sans article. Par contre, dans l'exemple 67, où le complément n'est pas déterminé, le nom n'est pas déterminé mais simplement spécifié.

(66)

kitāb-u *al=bint-i*
livre-NOM DEF=filie-GEN
'le livre [de] la fille'

(67)

kitāb-u *bint-i-n*
livre-NOM fille-GEN-INDF
'un livre d'une fille'

4.2 Caractéristiques du système nominal

4.2.1 Dans la flexion

Les caractéristiques flexionnelles du système nominal sont le genre, le nombre, la détermination et le cas.

Le genre et le nombre sont les facteurs essentiels pour déterminer l'accord entre le verbe et son sujet.

Tandis que le genre, le nombre, la détermination et le cas sont les facteurs essentiels pour déterminer l'accord entre les noms et leurs modificateurs.

4.2.1.1 Le genre

Il y a deux genres en arabe, le masculin et le féminin. Le masculin n'a pas de marque spécifique. Dans le cas des entités sexuées, c'est la forme masculine non marquée qui sera présente dans le dictionnaire. Le féminin a trois marques sous formes de suffixes dont quelques exemples sont illustrés dans le tableau 4.2:

- le suffixe le plus utilisé est *at*, dont la consonne *t* est appelée *tā'* *marbūṭah* .(s)
En pause, la *tā'* *marbūṭah* n'est pas prononcée ou prononcée comme *h*, par contre elle est prononcée *t* lorsque la parole est déclinée. Dans le cas des entités sexuées, le suffixe s'ajoute au masculin, par exemple, *ḡamīl* 'beau' et *ḡamīl-ah*

Cas	Ind	Det	Ind/Det	Ind	Det
Pause	<i>ġamīl-a</i> <i>ġamīl-ah</i>	<i>al-ġmīl-a</i> <i>al-ġmīl-ah</i>	<i>dunyā</i>	<i>ṣaḥrāʾ</i>	<i>aṣ-ṣaḥrāʾ</i>
Nom	<i>ġamīl-at-u-n</i>	<i>al-ġmīl-at-u</i>	<i>dunyā</i>	<i>ṣaḥrāʾ-u</i>	<i>aṣ-ṣaḥrāʾ-u</i>
Acc	<i>ġamīl-at-a-n</i>	<i>al-ġmīl-at-a</i>	<i>dunyā</i>	<i>ṣaḥrāʾ-a</i>	<i>aṣ-ṣaḥrāʾ-a</i>
Gén	<i>ġamīl-at-i-n</i>	<i>al-ġmīl-at-i</i>	<i>dunyā</i>	<i>ṣaḥrāʾ-a</i>	<i>aṣ-ṣaḥrāʾ-i</i>
Sens	'belle'		'monde'	'désert'	
	Nom	nominatif	Acc	Accusatif	Gen Génitif
	Ind	Indéterminé	Det	Déterminé	

Tab. 4.2 : Flexion des noms et des adjectifs féminins

'belle'. Les noms qui se terminent par *-at* (désigné aussi comme *-ah*) sont appelés tout simplement *muʿannaṭ* 'féminin' ;

- les deux autres suffixes sont *ā*, par exemple, *dunyā* 'monde' et *kubrā* 'la plus grande' ; et *āʾ*, par exemple, *kibriyā* 'orgueil', *ṣaḥrāʾ* 'désert' et *ṣaqrā* 'blonde'. Les noms qui se terminent par *-āʾ* ou *-ā* sont appelés, respectivement, *al=mamdūd* et *al=maqṣūr*. Lorsque les mots féminins qui se terminent par *ā* ou *āʾ* sont indéterminés, ils ne prennent ni le suffixe *n* de l'indéfinitude, ni le suffixe *i* du génitif. On parle de déclinaison diptota parce que le nom est compatible seulement avec les suffixes, *u*, pour marquer le nominatif et *a*, pour marquer l'accusatif et le génitif.

Cependant, plusieurs noms masculins se terminent, aussi, avec un suffixe féminin ; par exemple, *ḥalīf-ah* 'khalife', *fatā* 'jeune homme' et *duʿāʾ* 'prière'. C'est le cas aussi de certains noms propres masculins, par exemple *ḥamz-ah* et *ʿusām-ah*. En outre, la marque *-ah* peut être ajoutée à des noms (ou adjectifs) masculins pour dénoter un « haut degrés » d'activité ou exprimer *al-mubālaġah* 'l'exagération', par exemple, *raḥḥāl-ah* 'grand voyageur' et *allam-ah* 'grand savant'.

Par contre, certains noms ou adjectifs spécifiquement féminins peuvent ne pas avoir la marque du féminin, par exemple, *ʿumm* 'mère', *bint* 'fille', *ṭāliq* 'répudiée' et *ḥāmīl* 'enceinte'.

Le troisième aspect problématique est ce qu'on appelle *al=muʿannaṭ as-samāʾī*, 'lit. le féminin selon l'usage', qui concerne les noms désignant des choses inanimées ou des parties du corps et, sont féminins sans marqueur, par exemple, *šams* 'soleil', *nār* 'feu', *ḥarb* 'guerre', *ʿarḍ* 'terre'. C'est le cas, aussi, de la plupart des parties doubles du corps, comme, *ʿayn* 'œil' et *yad* 'main'

Il existe en plus des schèmes qui sont utilisés à la fois pour le masculin et le féminin, tels que *faʿīl* qui indique le participe passif, comme *qatīl* 'assassiné/assassinée' et *faʿūl* qui indique le participe actif, comme *ṣabūr* 'patient/patiente'; et des mots qui peuvent être traités à la fois comme masculins et féminins, par exemple, 'unuq' 'cou' et *lisān* 'langue'.

4.2.1.2 Le nombre

4.2.1.2.1 Le singulier et le duel Le singulier est la forme la moins marquée, par exemple, *muʿallim* 'un enseignant' et *muʿallim-ah* 'une enseignante'.

Par contre, on trouve une marque spécifique pour le duel sous forme d'un suffixe invariable en fonction du genre mais variable en fonction du cas, comme illustré par les exemples du tableau 4.3.

Cas	indéterminé ou déterminé		d'annexion	
	masculin	féminin	masculin	féminin
nominatif	<i>muʿallim-āni</i>	<i>muʿallim-at-āni</i>	<i>muʿallim-ā</i>	<i>muʿallim-at-ā</i>
accusatif ou génitif	<i>muʿallim-ayni</i>	<i>muʿallim-at-ayni</i>	<i>muʿallim-ay</i>	<i>muʿallim-at-ay</i>

Tab. 4.3 : Exemples de noms au duel

4.2.1.2.2 Le pluriel sain Le pluriel se réfère à un nombre de trois ou plus. Le pluriel sain, appelé aussi pluriel externe, s'obtient grâce à l'ajout d'un suffixe, comme illustré par les exemples du tableau 4.4.

Pluriel	masculin		féminin	
	indéterminé ou déterminé	d'annexion	indéterminé	déterminé ou cas d'annexion
nominatif	<i>muʿallim-ūna</i>	<i>muʿallim-ū</i>	<i>muʿallim-āt-u-n</i>	<i>muʿallim-āt-u</i>
accusatif ou génitif	<i>muʿallim-īna</i>	<i>muʿallim-ī</i>	<i>muʿallim-āt-i-n</i>	<i>muʿallim-āt-i</i>

Tab. 4.4 : Exemples du pluriel sain

Le pluriel sain masculin est presque exclusif aux noms (substantifs et adjectifs) qui se réfèrent à des entités humaines de sexe masculin. On l'obtient grâce à l'ajout d'un

suffixe variable, en fonction du cas. En plus, il est caractérisé par une partie finale, *na*, qu'il perd en état d'annexion. Par contre, le pluriel sain féminin s'obtient grâce à l'ajout du suffixe *-āt* et s'applique à :

- tous les noms (substantif ; adjectif et noms propres) qui se terminent par un suffixe féminin même si, sémantiquement, ils sont masculins ; par exemple, du nom propre masculin *'usām-ah* on obtient *'usām-āt*;
- tous les noms féminins (substantif ; adjectif et noms propres) même s'ils ne se terminent pas par un suffixe féminin ; par exemple, du nom propre féminin *marīam* on obtient *marīam-āt*;
- les diminutifs des noms masculins désignant des entités non humaines : *durayhim* 'petit dirham' donne le pluriel *durayhim-āt* 'quelques dirhams'.

4.2.1.2.3 Le pluriel brisé Une autre forme de pluriel, appelé brisé ou interne, se construit en impliquant le changement du pattern vocalique interne et/ou l'ajout d'un affixe. On peut distinguer divers types de pluriel brisé.

4.2.1.2.3.1 Le pluriel de paucité Appelé *ǧam' al=qillah*, il désigne un pluriel de petit nombre, compris entre trois et dix, dont les schèmes les plus importants sont reportés dans le tableau 4.5.⁵

Schèmes	Exemples		
	singulier	pluriel	signification
'aC ₁ C ₂ uC ₃	<i>nafs</i>	' <i>anfus</i>	'âme'
'aC ₁ C ₂ āC ₃	<i>bāb</i> (racine <i>bwb</i>)	' <i>abwāb</i>	'porte'
'aC ₁ C ₂ iC ₃ -ah	' <i>amūd</i>	' <i>a'mid-ah</i>	'colonne'
C ₁ iC ₂ C ₃ -ah	<i>fatā</i> (racine <i>fty</i>)	<i>fityah</i>	'jeune homme'

Tab. 4.5 : Schèmes de pluriel de paucité

4.2.1.2.3.2 Le pluriel de grand nombre En arabe, *ǧam' al=kaṭrah*, il désigne un nombre entre trois et l'infini dont les schèmes les plus répondus sont :⁶

5. El-Dahdah, 1999, p. 59. et Ben Braham, 1897, pp. 12-13 et 16-29.

6. El-Dahdah, 1999, p. 59. et Ben Braham, 1897, pp. 13-15 et 30-89.

1. **C₁uC₂C₃** est régulier avec les adjectifs de couleur ou de handicap physique, dont le schème singulier est 'aC₁C₂aC₃ au masculin et C₁aC₂C₃ā' au féminin, par exemple, les deux adjectifs 'aḥmar, 'rouge' au masculin, et ḥamrā' 'rouge' au féminin ; ont le pluriel ḥumr.
2. **C₁uC₂uC₃** est le pluriel des :
 - adjectifs masculins de schème C₁aC₂ūC, par exemple :
ṣabūr - ṣubur 'patient' ;⁷
 - substantifs dont la voyelle médiane est longue au singulier ayant un des schèmes : C₁aC₂āC₃; C₁iC₂āC₃; C₁uC₂āC₃; C₁aC₂ūC₃ ou C₁aC₂īC₃, comme :
'amūd - 'umud 'colonne' ;
kitāb - kutub 'livre' ;
qaḍīb - quḍub 'bâton mince ; barre'.
3. **C₁uC₂aC₃** est le pluriel des :
 - adjectifs féminins de schème C₁uC₂C₃ā : kubrā - kubar 'plus grande' ;
 - substantifs de schème C₁uC₂C₃ah : ġurfah - ġuraf 'chambre'.
4. **C₁iC₂aC₃** est le pluriel des substantifs de schème C₁iC₂C₃ah :
qiṭ'ah - qiṭa 'partie'.
5. **C₁uC₂aC₃-ah** est, généralement, le pluriel des participes actifs désignant des êtres humains, de schème C₁āC₂iC₃, dont C₃ est une wāw ou une yā'. Le pluriel qui se forme, alors, est de schème C₁uC₂awah ou C₁uC₂ayah. Pour raisons phonologiques, le groupe, awa ou aya, se transforme en ā, et le suffixe tā' marbūṭah (ة) est prononcée comme une consonne t. Le schème finale est, de conséquence, C₁uC₂āt, par exemple : qāḍ-in - quḍāt 'juge'.
6. **C₁aC₂aC₃-ah** est, généralement, le pluriel des participes actifs désignant des êtres humains, de schème C₁āC₂iC₃, dont C₃ n'est pas une wāw ou une yā' :
sāḥir - saḥar-ah 'sorcier'.
7. **C₁aC₂C₃ā** est le pluriel des adjectifs désignant la souffrance ou la mort et de schème :
C₁aC₂īC₃: qatīl - qatīlā 'assassiné' et ġarīḥ - ġarḥā 'blessé' ;

7. Les exemples sont représentés selon l'ordre : singulier - pluriel 'traduction'.

$C_1\bar{a}C_2iC_3$: *hālik* - *halkā* 'périssable ; mortel' ;
 $C_1aC_2iC_3$: *zamin* - *zamnā* 'malade ; impotent' ;
 $'aC_1C_2aC_3$: *'ahmaq* - *ḥamqā* 'malade ; impotent' ;
 $C_1aC_2C_3\bar{a}n$: *'aṣṣān* - *'aṣṣā* 'assoiffé'.

8. $C_1iC_2aC_3\text{-ah}$ est surtout le pluriel de substantif de schème $C_1uC_2C_3$:
durġ - *diraġah* 'tiroir'.
9. $C_1uC_2C_2aC_3$ et $C_1uC_2C_2\bar{a}C_3$ sont des pluriels d'adjectifs de schème $C_1\bar{a}C_2iC_3$ dont C_3 est une consonne saine : *rāki'* - *rukka* 'agenouillé' et *qāri'* - *qurrā* 'lecteur'.
10. $C_1iC_2C_2\bar{a}C_3$ est le pluriel des :
 adjectifs et substantifs de schème $C_1aC_2C_3$ ou $C_1aC_2C_3ah$: *kalb* - *kilāb* 'chien' et *ša'b* - *ši'āb* 'difficile' ;
 substantifs de schème $C_1aC_2aC_3$ ou $C_1aC_2aC_3ah$: *ġamal* - *ġimāl* 'chameau' et *raqabah* - *riqāb* 'cou' ;
 substantifs de schème $C_1iC_2C_3$: *dī'b* - *dī'āb* 'loup' ;
 substantifs de schème $C_1uC_2C_3$: *rumḥ* - *rimāḥ* 'flèche' ;
 adjectifs de schème $C_1aC_2\bar{i}C_3$ et son féminin $C_1aC_2\bar{i}C_3ah$: *karīm* 'généreux' et *karīmah* 'généreuse' ont le pluriel commun *kirām*.
11. $C_1uC_2\bar{u}C_3$ est le pluriel des substantifs de schème :
 $C_1aC_2iC_3$: *kabid* - *kubūd* (foie) ;
 $C_1vC_2C_3$ dont la première voyelle peut être /a/ ; /i/ ou /u/ :⁸
ka'b - *ku'ūb* 'talon' et *ḍirs* - *ḍurūs* 'molaire'.
12. $C_1uC_2aC_3\bar{a}$ est le pluriel des adjectifs masculins désignant des êtres humains et dont le singulier est de schème :
 $C_1aC_2\bar{i}C_3$: *karīm* - *kuramā* 'généreux' ;
 $C_1\bar{a}C_2iC_3$: *ṣāliḥ* - *ṣulaḥā* 'vertueux'.
13. $C_1uC_2C_3\bar{a}n$ est le pluriel des substantifs dont le singulier est de schème :
 $C_1aC_2C_3$: *zahr* - *zahrān* 'dos' ;

8. v = a à condition que C_2 soit différent de *wāw* ($C_2\# wāw$).
 v = u à condition que $C_2\# wāw$, $C_3 \# yā$ et $C_2\# C_3$

$C_1aC_2aC_3$: *ḥamal - ḥumlān* 'agneau' ;

$C_1aC_2īC_3$: *qaḍīb - quḍbān* 'bâton'.

14. $C_1iC_2C_3ān$ est le pluriel des substantifs dont le singulier est de schème :

$C_1uC_2āC_3$: *ḡulām - ḡilmān* 'garçon' ;

$C_1uC_2aC_3$: *ṣurad - ṣirdān* 'pie ; grièche' ;

$C_1uC_2C_3$ et $C_1aC_2C_3$ dont C_2 est une *wāw* qui devient d'allongement en fonction de la voyelle qui la précède, donc le schème est $C_1ūC_3$, par exemple, *ḥūt - ḥītān* 'poissons' et $C_1āC_3$, par exemple, *nār - nīrān* 'feu'.

15. ' $aC_1C_2iC_3ā$ ' est le pluriel des substantifs et des adjectifs de schèmes $C_1aC_2īC_3$ dont C_3 est une *wāw* ou une *yā* : *nabiyy - 'anbiyā* 'prophète' ; ou C_2 et C_3 sont identiques : '*azīz - 'a'izzā* 'cher'.

Cependant on trouve des exceptions à cette règle, tel que *ṣaḍīq - 'aṣḍiqā* 'ami'.

16. $C_1awāC_2iC_3$ est le pluriel des :

- substantifs dont le singulier est de schème :

$C_1āC_2iC_3$: *ḥāḡib- ḥawāḡib* 'sourcil' ;

$C_1āC_2iC_3ah$: *ḡāriḥah - ḡawāriḥ* 'organe ; partie du corps' ;

$C_1āC_2aC_3$: *ḥātām - ḥawātim* 'sceau ; cachet' ;

$C_1awC_2aC_3$: *ḡawhar- ḡawāhir* 'quiddité ; entité' ;

$C_1awC_2aC_3ah$: *ṣawma 'ah - ṣawāmi'* 'minaret'.

- adjectifs féminins de schèmes :

$C_1āC_2iC_3ah$: *kāḍibah - kawāḍib* 'menteuse' ;

$C_1āC_2iC_3$: *ṭāliq - ṭawāliq* 'répudiée'.

17. $C_1aC_2āiC_3$ est le pluriel des substantifs féminins avec ou sans préfixe féminin (-ah), dont l'avant dernière lettre est de prolongation ('*alif* ; *wāw* ou *yā*) ou, mieux dire que, la dernière voyelle est longue : *šamāl - šamā'il* 'vent du nord' ; *ṣaḥīfah - ṣaḥā'if* 'page de livre' ; '*aḡūz - 'aḡā'iz* 'vieille femme'.

18. $C_1aC_2āC_3ī$ dont la voyelle finale (ī) est stable quand le substantif est déterminé mais, subit une élision en cas d'indétermination. Dans ce dernier cas le substantif prend *in* invariable quelque soit le cas. Elle s'emploie régulièrement avec les substantifs dont le singulier est :

$C_1aC_2C_3\text{āt}$: *mawmāt - al=mawāmī* (dans le cas indéterminé *mawām-in*) 'désert' ;

$C_1iC_2C_3\text{āt}$: *si'lāt - as-sa'ālī (sa'āl-in)* 'sorcière' ;

$C_1iC_2C_3\text{iyah}$: *hibriyat - al=habārī (habār-in)* 'flocon' ;

$C_1aC_2C_3\text{uwah}$: '*arquwah - al='arāqī ('arāq-in)* 'anse'.

19. $C_1aC_2\bar{a}C_3\bar{a}$ s'emploie avec les adjectifs dont le singulier masculin est $C_1aC_2C_3\bar{a}$ et le féminin est $C_1aC_2C_3\bar{a}$:

sakrān 'ivre' dont le féminin est *sakrā* et le pluriel dans les deux cas est *sakārā*.

20. $C_1aC_2\bar{a}C_3\bar{i}$ et $C_1aC_2\bar{a}C_3\bar{a}$ s'emploient concurremment et régulièrement avec :

- les substantif et les adjectifs féminin dont le singulier est $C_1aC_2C_3\bar{a}$:

ṣaḥrā' (désert) a deux pluriels *ṣaḥārā* et *aṣ-ṣaḥārī* ;

'*adrā*' (vierge) a deux pluriels '*aḍārā* et *al='aḍārī* ;

- les substantifs et les adjectifs qui se terminent avec la voyelle \bar{a} final :

ḥublā (enceinte) a deux pluriels *ḥabālā* et *al=ḥabālī*.

21. $C_1aC_2\bar{a}C_3\text{iiyy}$ est le pluriel des substantifs de schème $C_1vC_2C_3\text{iiyy}$ dont *v* peut être /*a*/ ; /*u*/ ou /*i*/ : *kursiyy - karāsiyy* 'chaise'

22. $C_1aC_2\bar{a}C_3iC_4$ s'emploie avec :

- les quadrilitères : *ḡa'far - ḡa'āfir* 'rivière' ;

- les quintilitères en supprimant la cinquième radicale : *safarḡal - safāriḡ* (coing).

23. Les schèmes analogues à $C_1aC_2\bar{a}C_3iC_4$, caractérisés surtout par le pattern vocalique *a-ā-i*, et indépendamment de la nature des consonnes qui les constituent. Par exemple, $C_1ay\bar{a}C_2iC_4$; $C_1aw\bar{a}C_2iC_4$; $maC_1\bar{a}C_2iC_4$; ' $aC_1\bar{a}C_2iC_4$; etc. Ils s'appliquent aux noms (substantifs et adjectifs) dérivant d'une racine trilitère et ayant un ou trois augments :

de la racine *fḍl*, '*afḍal - 'afāḍil* 'meilleur' ;

de la racine *sḡd*, *masḡid - masāḡid* 'mosquée'.

Ce qu'on peut noter est la multiplicité et l'imprévisibilité des schèmes. En outre, une parole peut avoir plusieurs pluriels brisés qu'on utilise indifféremment.⁹, comme par exemple, *naḡm* (étoile) a deux pluriels, '*anḡum* (schème de pluriel de paucité) et *nuḡūm*

9. Ben Braham, 1897, pp. 13-15.

(schème de pluriel de grand nombre) et *lisān* (langue) en a trois *'alsun* et *'alsinah* (schèmes de pluriel de paucité) et *lusun* (schème de pluriel de grand nombre).

4.2.1.2.3.3 Le pluriel du pluriel Un autre caractéristique de la langue arabe est le fait qu'il est possible de pluraliser le pluriel, on obtient alors, *ǧam' al=ǧam'*. Ainsi on peut obtenir un nouveau pluriel sain ou brisé, appelé le «pluriel du pluriel », par exemple :

bayt 'maison', *buyūt* 'maisons' on obtient *buyūt-āt* 'plusieurs maisons';
raǧul 'homme', *riǧāl* 'hommes' on obtient *riǧāl-āt* 'plusieurs hommes';
'afḍal 'meilleur', *'afāḍil* 'meilleurs' on obtient *'afāḍil-ūn* 'meilleurs'.

Il faut noter, dans ce cas, que le suffixe *āt* ne marque pas le genre féminin mais la grande pluralité. Autrement dit, *riǧāl-āt* 'plusieurs hommes' reste masculin sur le plan sémantique et syntaxique.

4.2.1.2.3.4 Le pluriel extrême On ne peut pluraliser que les pluriels qui ont un schème analogue au singulier. De conséquence tous les schèmes de petite pluralité et les schèmes de grande pluralité numérisés de 1 jusqu'à 15 (voir ci-dessus) admettent un nouveau pluriel.

Par contre, les pluriels numérisés de 16 jusqu'à 23 (voir ci-dessus) n'admettent pas d'être pluraliser parce qu'ils ont une voyelle longue *ā* après la deuxième consonne, et donc ils ne sont analogues à aucun singulier. On les appelle *muntaḥā l=ǧumū'* (litt. paroxysme/summum des pluriels) communément appelés les pluriels extrêmes, dans le sens qu'ils ne peuvent être pluraliser.

4.2.1.2.3.5 Le nom du pluriel Appelé *ism al=ǧam'*, il représente un pluriel qui dérivent d'une racine différente de celle dont dérive le singulier, comme les exemples du le tableau 4.6.

Singulier	Racine	Pluriel	Racine
(i)mra'-ah	mr'	nīsā'	ns'
'femme'		'femmes'	
faras	frs	ḥayl	ḥyl
'cheval'		'chevaux'	

Tab. 4.6 : Exemples de noms du pluriel et leurs singuliers

4.2.1.2.4 Le nom collectif Dans le cas où *ism al=ğins* 'nom collectif' représente un ensemble d'individus distincts (*ism al=ğins al=ğama* ʾ), par exemple *samak* 'poissons', on peut dériver le nom de l'unité (*ism al=waḥdah*), en ajoutant le suffixe (*at/ah*), *samak-ah* 'un poisson'.

Par contre, dans le cas où il désigne «une masse » non divisible en individus (*ism al=ğins al=ifrādī*), par exemple, *zayt* 'huile', il n'admet pas de singulier.

4.2.1.3 Système nominale vs système verbale

Le tableau 4.7 permet une comparaison entre les affixes (préfixes et suffixes) utilisés dans la conjugaison du *muḍāriʿ* et les suffixes utilisés dans la flexion du nom.

G	N	<i>muḍāriʿ</i>					nom			
		préfixe	suffixe				suffixe			
		In	Su	Ap	variable	N	Ac	G	variable	
M	S	y	u	a	-	-	u	a	i	n
	D	y	ā	ā	ā	ni	ā	ay	ay	ni
	P	y	ū	ū	ū	na	ū	ī	ī	na
F	S	t	u	a	-	-	at-u	at-a	at-i	n
	D	t	ā	ā	ā	ni	at-ā	at-ay	at-ay	ni
	P	y	na	na	na	-	āt-u	āt-i	āt-i	n

Spécifique au *muḍāriʿ* In : indicatif Su : subjonctif Ap : apocopé
 Spécifique au nom N : nominatif Ac : accusatif G : génitif
 Commun G : genre M : masculin F : féminin
 N : nombre S : singulier D : duel P : pluriel

Tab. 4.7 : Parties analogues entre système verbal et système nominale

Les affixes du *muḍāriʿ* et du nom présentent une analogie morphologique qui explique l'appellation arabe du *muḍāriʿ* qui signifie 'le ressemblant [au nom]'. Elle explique, aussi, le fait que l'indicatif (du verbe) et le nominatif (du nom) sont, tous les deux, appelés *marfūʿ* (du nom de la voyelle *u*: *rafʿ*); le subjonctif (du verbe) et l'accusatif (du nom) sont appelés *maṣṣūb* (du nom de la voyelle *a*: *naṣṣb*).

On note aussi une analogie de la partie variable, dont la présence marque l'indicatif du *muḍāriʿ* (*ni/na*) et la détermination du nom (*n/ni/na*), par contre, son absence marque le subjonctif et l'apocopé du verbe et l'indétermination du nom.

La différence morphologique se note surtout entre le *muḍāri'* apocopé et le nom génitif. Le premier est caractérisé, au singulier, par l'élision de la voyelle finale, tandis que le deuxième se caractérise par une voyelle finale *i*. Tandis que le verbe ne peut jamais avoir une voyelle *i*, le nom ne peut jamais se terminer par une consonne muette.

4.2.2 Dans la dérivation

Le nom *muğarrad* 'sans augment' est composé d'un nombre de consonnes radicales, compris entre trois et cinq, comme *kalb* 'chien'; *sarmad* 'éternité' et *safarğal* 'coing'. Cependant la plupart des noms provient de racines trilitères ou quadrilitères et ceux dérivant de racines quinquilitères sont rares et improductifs.

Par contre le nom *mazīd* 'avec augment' (dérivé) est composé de quatre à sept consonnes, par exemple, *ğazāl* 'gazelle'; *za'farān* 'safran' et *ħandaqūqā* 'Melilotus' (nom d'une plante).¹⁰

Un nombre très limité de noms dérivent de racines bilitères qui ont été normalisées en ajoutant généralement une consonne *wāw* ou *yā'*. Ce sont des mots archaïques indiquant des relations familiales comme *'aħ* 'frère' de la racine *'ħw*; des parties du corps, par exemple *yad* 'main' de la racine *ydy* ou des concepts physiques ou sociaux fondamentaux comme *mā'* 'eau' de la racine *mwh* et *dū* 'possesseur [de]'.¹¹

Une autre classification divise les noms selon qu'ils résultent de la dérivation à partir d'un verbe (noms déverbaux) ou non (noms primitifs)

4.2.2.1 Les noms primitifs

Beaucoup de noms, attestés dans le lexique arabe, ne dérivent pas du verbe. On les appelle les «noms primitifs», par exemple, *'arḍ* 'terre', *sarmad* 'éternité' et *'uqrub* 'scorpion'.

4.2.2.2 Les noms déverbaux

4.2.2.2.1 Le nom verbal *al=maşḍar* 'nom verbal' dénote le même contenu sémantique du verbe, en d'autres termes, il exprime une action ou un état. Cependant, il ne contient aucune indication de temps, mode ou personne.

Comme déjà annoncé dans le chapitre de la dérivation 2, l'école de *Başrah* considère le *maşḍar* la base de la dérivation de laquelle dérive le verbe, parce qu'il repré-

10. Ben Braham, 1897, p. 2

11. Ryding, 2005, p. 92.

sente la forme la moins marquée sémantiquement et dénote seulement le procès. Par contre, les grammairiens de *kūfah* faisaient le choix contraire.¹²

Dans cette recherche, le but est de donner un aperçu général des schèmes des noms déverbaux sans rentrer dans les détails si le *maşḍar* est déverbal et donc dérive du verbe ou le contraire.

4.2.2.2.1.1 Le *maşḍar* du verbe «nu » Le verbe trilitère peut avoir avoir plusieurs *maşḍar*, dont certains sont *qiyāsī* 'formels' et d'autres sont *samā'ī* 'non formels'. Le tableau 4.8 présente les schèmes formels qui ont un contenu sémantique spécifique.

Schème <i>maşḍar</i>	Contenu sémantique	Exemples	Traductions
C ₁ iC ₂ āC ₃ ah	métier	<i>tiġārah</i>	'commercer'
C ₁ aC ₂ aC ₃ ān	trouble - agitation - anomalie de	<i>ġalayān</i>	'bouillonner'
	fonctionnement d'un organe	<i>ḥafaqān</i>	'palpitation'
C ₁ uC ₂ āC ₃	bruit - son	<i>şurāḥ</i>	'hurler'
	maladie	<i>zukām</i>	'avoir le rhume'
C ₁ aC ₂ īC ₃	bruit - son	<i>ṭanīn</i>	'bourdonner'
	cheminement - changement	<i>raḥīl</i>	'partir'
C ₁ iC ₂ āC ₃	abstinence - refus	' <i>ibā</i> '	'refuser'
C ₁ uC ₂ C ₃ ah	couleur	<i>ḥumrah</i>	'être rouge'

Tab. 4.8 : *maşḍar* formels selon le contenu sémantique du verbe trilitère.

Par contre, le tableau 4.9 présente certains schèmes formels du *maşḍar*, selon le schème du verbe qui leur est associé.¹³

Concernant le verbe quadrilitère, lorsque il est constitué de quatre consonnes différentes, le *maşḍar* est C₁aC₂C₃aC₄ah, par exemple, *tarġama* 'il a traduit' dont le *maşḍar* est *tarġamah* 'traduire'.

Dans le cas où verbe quadrilitère est caractérisé par une réduplication de C₁ en C₃ et de C₂ en C₄, le *maşḍar* est C₁iC₂C₃āC₄; par exemple, *waswasa* 'il s'est méfié' dont le *maşḍar* est *wiswās* 'se méfier'.

12. Larcher, 2008.

13. El-Dahdah, 1999, p. 74.

Schème du verbe	Schème <i>maṣḍar</i>	Exemples	Traductions
C ₁ aC ₂ uC ₃ a	C ₁ uC ₂ ūC ₃ ah	<i>suhūlah</i>	'être/devenir facile'
	C ₁ aC ₂ āC ₃ ah	<i>faṣāḥah</i>	'être/devenir éloquent'
	C ₁ aC ₂ aC ₃	<i>karam</i>	'être/devenir généreux'
C ₁ aC ₂ iC ₃ a intransitif	C ₁ aC ₂ aC ₃	<i>faraḥ</i>	'être joyeux'
C ₁ aC ₂ aC ₃ a intransitif	C ₁ uC ₂ ūC ₃	<i>qu'ūd</i>	's'asseoir'
C ₁ aC ₂ iC ₃ a transitif	C ₁ aC ₂ C ₃	<i>fahm</i>	'comprendre'
C ₁ aC ₂ aC ₃ a transitif	C ₁ aC ₂ C ₃	<i>kasb</i>	'gagner'

Tab. 4.9 : Noms verbaux formels selon le schème du verbe trilitère.

4.2.2.1.2 Le *maṣḍar* des verbes dérivés Les verbes dérivés ont, généralement, un (ou deux) *maṣḍar* qui sont formels et reportés dans le tableau 4.10.

Schème verbal	<i>maṣḍar</i>	Exemple	Traduction
C ₁ aC ₂ C ₂ aC ₃ a	taC ₁ C ₂ īC ₃	<i>taqdim</i>	'présenter'
	taC ₁ C ₂ iC ₃ ah	<i>tahni'ah</i>	'féliciter'
C ₁ āC ₂ aC ₃ a	muC ₁ āC ₂ aC ₃ ah	<i>munāḍalah</i>	'lutter'
	C ₁ iC ₂ āC ₃	<i>niḍāl</i>	'lutter'
'aC ₁ C ₂ aC ₃ a	'iC ₁ C ₂ āC ₃	<i>iḥsān</i>	'faire la charité'
taC ₁ aC ₂ C ₂ aC ₃ a	taC ₁ aC ₂ C ₂ uC ₃	<i>taqaddum</i>	'avancer' ; 'évoluer'
taC ₁ āC ₂ aC ₃ a	taC ₁ āC ₂ uC ₃	<i>tašāruk</i>	'participer' ; 'cotiser'
taC ₁ aC ₂ C ₃ aC ₄ a	taC ₁ aC ₂ C ₃ uC ₄	<i>tadahruġ</i>	'rouler'
(i)staC ₁ C ₂ aC ₃ a	(i)stiC ₁ C ₂ āC ₃	<i>(i)stiqbāl</i>	'accueillir'
(i)nC ₁ C ₂ aC ₃ a	(i)nC ₁ iC ₂ āC ₃	<i>(i)nkisār</i>	'se casser'
(i)C ₁ taC ₂ aC ₃ a	(i)C ₁ tiC ₂ āC ₃	<i>(i)ftidā'</i>	'racheter'
(i)C ₁ C ₂ aC ₃ C ₃ a	(i)C ₁ C ₂ iC ₃ āC ₃	<i>(i)ḥmirār</i>	'devenir rouge'
(i)C ₁ C ₂ anC ₃ C ₄ a	(i)C ₁ C ₂ inC ₃ āC ₄	<i>(i)ḥringām</i>	'se réunir'

Tab. 4.10 : Schèmes formels des *maṣḍar* des verbes dérivés

4.2.2.2.2 Le nom d'agent et le nom de patient Lorsque le verbe désigne une action, donc de paradigme C₁aC₂aC₃a ou C₁aC₂iC₃a, il est associé à un nom d'agent (ou participe actif), désignant l'agent qui fait l'action. Dans le cas des verbes transitifs,

on peut dériver aussi un nom de patient (ou participe passif) qui désigne le patient qui subit l'action ou le résultat de cette action.

Dans le cas du verbe trilitère, le participe actif est $C_1\bar{a}C_2iC_3$ et le participe passif est $maC_1C_2\bar{u}C_3$. Par exemple le verbe *ḍaraba* 'frapper' est associé à *ḍārib* qui désigne l'agent 'celui qui a frappé' et *maḍrūb* qui désigne le patient 'celui qui a été frappé'.

Pour les verbes $(i)C_1C_2aC_3C_3a$ intransitifs et désignant une couleur ou un handicap, le nom d'agent est $muC_1C_2aC_3C_3$. Par exemple le verbe $(i)ḥmarra$ (devenir rouge) est associé au nom d'agent *muḥmarr*.

Pour les autres verbes dérivés et quadrilitères, les deux noms ont le même stem¹⁴ du verbe ; mais ils sont caractérisés par l'ajout d'un préfixe *mu* et l'avant dernière voyelle est /i/ dans le cas du nom d'agent et /a/ dans le cas du nom du patient.

Par exemple, le verbe *kassara* 'fracasser', de schème $C_1aC_2C_2aC_3-a$, est associé au nom d'agent *mukassir*, de schème $muC_1aC_2C_2iC_3$, et le nom du patient est *mukassar*, de schème $muC_1aC_2C_2aC_3$.

4.2.2.2.3 L'adjectif analogue Les verbes qui dénotent un état ne sont pas associés aux deux participes actif et passif. Mais, ils sont associés à une *ṣifah muṣabbahah* 'adjectif analogue [au participe]'.
L'adjectif « analogue » exprime l'état du sujet. Si l'adjectif est X, le verbe est traduit par « être ou devenir X ». Le tableau 4.11 illustre les schèmes formels de l'adjectif selon le verbe auquel il est associé.¹⁵

4.2.2.2.4 Le nom de lieu et le nom de temps Le verbe peut être associé, également, à *ism al-makān* 'nom du lieu' qui désigne le lieu où se produit l'action exprimée par le verbe.

Dans le cas où le sens se prête à une interprétation temporelle, il peut être associé, aussi, à *ism az-zamān* 'nom du temps' qui indique le moment quand s'est produite l'action. Le tableau 4.12 illustre les schèmes formels des noms de lieu ou de temps, selon le schème verbe auquel ils sont associés.

Les verbes de schèmes $C_1aC_2aC_3a - yaC_1C_2uC_3u$, sont associés normalement à $maC_1C_2aC_3$, cependant, certains verbes très fréquents font exception et sont associés à $maC_1C_2iC_3$,¹⁶ comme :

14. Le stem verbal est la forme du verbe sans les affixes de flexion.

15. El-Dahdah, 1999, p. 84.

16. El-Dahdah, 1999, p. 91.

LA LANGUE ARABE
 CARACTÉRISTIQUES MORPHOSYNTAXIQUES ET SÉMANTIQUES

Schème		Exemples		Traduction
du verbe	de l'adjectif	verbe 'être/devenir + adjectif'	adjectif	
C ₁ aC ₂ uC ₃ a	C ₁ aC ₂ C ₃	<i>ḍaḥuma</i>	<i>ḍaḥm</i>	'corpulent ; gros'
	C ₁ uC ₂ C ₃	<i>ṣaluba</i>	<i>ṣulb</i>	'dur ; rigide'
	C ₁ aC ₂ aC ₃	<i>baṭula</i>	<i>baṭal</i>	'héroïque ; intrépide'
	C ₁ aC ₂ āC ₃	<i>ḡabuna</i>	<i>ḡabān</i>	'poltron ; craintif'
	C ₁ uC ₂ āC ₃	<i>šaḡu'a</i>	<i>šuḡā'</i>	'valeuroux ; preux'
	C ₁ aC ₂ īC ₃	<i>šarufa</i>	<i>šarīf</i>	'honnête ; honorable'
C ₁ aC ₂ iC ₃ a	'aC ₁ C ₂ aC ₃	<i>sawida</i>	<i>'aswad</i>	'noir'
	C ₁ aC ₂ iC ₃	<i>fariḥa</i>	<i>fariḥ</i>	'heureux ; content'
	C ₁ aC ₂ C ₃ ān	<i>ḡaḍiba</i>	<i>ḡaḍbān</i>	'exaspéré'
	C ₁ uC ₂ C ₃ ān	<i>'ariya</i>	<i>'uryān</i>	'découvert ; dénué'
C ₁ aC ₂ aC ₃ a	C ₁ āC ₂ iC ₃	<i>ṭahara</i>	<i>ṭāhir</i>	'honnête ; vertueux'
	maC ₁ C ₂ ūC ₃	<i>ḥamada</i>	<i>maḥmūd</i>	'louable ; méritoire'

Tab. 4.11 : Schèmes formels de l'adjectif selon le schème verbal

marḡib 'moment du coucher du soleil ; Ouest ; occident' ;
mašriq 'moment du lever du soleil ; levant ; Est ; orient' ;
maṭli' 'lieu du lever du soleil ; commencement d'une chose ; entrée en matière' ;
masḡid 'mosquée' ;
maskin 'habitation'.

Schème verbal		Nom de lieu ou du temps	Exemples	
accompli	inaccompli			
C ₁ aC ₂ aC ₃ a	yaC ₁ C ₂ uC ₃ u	maC ₁ C ₂ aC ₃	<i>raqada - yarqudu</i> 'gésir ; pleuter'	<i>marqad</i> 'lit'
C ₁ aC ₂ aC ₃ a	yaC ₁ C ₂ iC ₃ u	maC ₁ C ₂ iC ₃	<i>ḡalasa - yaḡlisu</i> 's'asseoir'	<i>maḡlis</i> 'séance ; club'
C ₁ aC ₂ iC ₃ a			<i>qadima - yaqdimu</i> 'arriver'	<i>maqdim</i> 'arrivée'

Tab. 4.12 : Schèmes formels des noms de lieu et de temps

On peut dériver le nom de lieu, également, à partir d'un nom. Dans ce cas, il désigne la multiplicité de l'objet désignée par le nom dans un lieu. Par exemple :

'asad 'lion' - *ma'sadah* 'lieu infesté où vivent beaucoup de lions'

kalb 'chien' - *maklabah* 'lieu infesté où vivent beaucoup de chiens'.

4.2.2.2.5 Le nom d'instrument *ism al-'ālah* désigne l'instrument qui sert pour exécuter l'action exprimée par le verbe, et il est dérivé selon l'un des trois schèmes illustrés dans le tableau 4.13.

Schème du nom d'instrument	Exemples	Verbe
miC ₁ C ₂ aC ₃	<i>mibrad</i> 'lime'	<i>barada</i> 'limer'
miC ₁ C ₂ aC ₃ ah	<i>miknasah</i> 'balai'	<i>kanasa</i> 'balayer'
miC ₁ C ₂ āC ₃	<i>miftāḥ</i> 'clé'	<i>fataḥa</i> 'ouvrir'

Tab. 4.13 : Schèmes formels du nom d'instrument

Troisième partie

LE LEXIQUE ARABE NAISSANCE, HISTOIRE, ÉVOLUTION

Chapitre 1

Tradition lexicographique arabe

1.1 Introduction

La propagation de l'islam et les conquêtes islamiques ont eu pour conséquence la formation d'une nouvelle population mixte et hétérogène et, la naissance d'une nouvelle langue de communication. D'un côté, les arabes se sont trouvés à simplifier leur langue sur plusieurs fronts pour faciliter la communication avec les peuples conquis. De l'autre, la nouvelle langue s'enrichissait d'emprunts linguistiques acquis des peuples étrangers.¹ Les érudits arabes appelaient ces changements *lahn* 'solécisme ; faute', et les considéraient comme corruptions qui peuvent affecter divers niveaux :²

1. liés à la prononciation des lettres, comme par exemple, prononcer *ḥukk* au lieu de *ḥuqq* 'coffret' ;³
2. liés à la sémantique et l'utilisation inappropriée des paroles, comme par exemple, la fameuse anecdote qu'on raconte sur le non moins fameux gouverneur de la Mecque et de l'Iran, Ḥālid ibn 'Abdallah al-Qasrī (m. 126/743), qui, assoiffé, a demandé de l'eau en disant : *aṭ'imūnī mā'an* 'nourrissez-moi de l'eau' ;⁴
3. liés à la dérivation et à l'utilisation de schèmes incorrects, par exemple, la parole *bazīm* au lieu de 'ibzīm 'l'agrafe ; la boucle de ceinture'⁵ ; *maqwad* au lieu de *miqwad* 'bride ; pièce de harnais' ;⁶

1. Naṣṣār, 1988, Vol. I, p. 19

2. Pour un étude approfondie Zubaydī, *Lahn*.

3. Zubaydī, *Lahn*, p. 115.

4. Baalbaki, 2014, p. 4. et Naṣṣār, 1988, V.I, p. 20

5. Zubaydī, *Lahn*, p. 73.

6. Le schème maC1C2aC3 ne peut être utilisé dans le cas où C2 est une wāw (Zubaydī, *Lahn*, p. 122).

4. liés à la syntaxe et en particulier à *al=i'rāb* 'système de déclinaison nominale'. Une des plus significatives anecdotes est liée à la lecture du Qur'ān concernant le Verset 3 de la *Sūrah 9* :⁷

أَنَّ اللَّهَ بَرِيءٌ مِنَ الْمُشْرِكِينَ وَرَسُولُهُ

'anna l=lāha barī'un mina l=mušrikīna wa=rasūlu=hu

'que Allah dissout [les traités] avec les idolâtres et son Messenger [fais de même]'.
'que Allah dissout [les traités] avec les idolâtres et son Messenger'

Le grammairien 'abū l='aswad ad=du'alī (m. 69/688) a entendu une personne lire le Verset mais en utilisant le génitif, *wa-rasūlihi*, au lieu du nominatif *wa-rasūluhu*, ce qui change complètement le sens :

أَنَّ اللَّهَ بَرِيءٌ مِنَ الْمُشْرِكِينَ وَرَسُولِهِ

'anna l=lāha barī'un mina l=mušrikīna wa=rasūli=hi

'que Allah dissout [les traités] avec les idolâtres et son Messenger' !

Pour limiter les fautes de lecture, ad=du'alī s'est ingénié à mettre les bases de la vocalisation et de la syntaxe.

Une autre histoire est liée à la transmission du *ḥadīṭ*.⁸ Le transmetteur de *ḥadīṭ* (*muḥaddit*), ḥammād bnu salamah (m. 167/784) dictait un jour le *ḥadīṭ*:

لَيْسَ مِنْ أَصْحَابِي إِلَّا مَنْ لَوْ شِئْتُ لَأَخَذْتُ عَلَيْهِ لَيْسَ أَبَا الدَّرْدَاءِ

laysa min 'aṣḥāb=i'illā man law ši'tu la='aḥad=tu 'alay=hi laysa 'abā d=dardā'i

'Il n'y a pas de compagnon à moi à qui je ne trouverais pas de défaut, si je veux, sauf 'abū d-dardā'i'.

Pensant à tort que *laysa* est la « particule » négative qui requiert que son nom soit nominatif, sībawayhi (m. 180/796) qui l'écoutait, a commis l'erreur de corrigé *laysa 'abū l-dardā'i*.⁹ L'erreur commise par sībawayhi l'a poussé à étudier la langue et devenir la figure la plus influente de la tradition grammaticale, en écrivant son fameux livre *al=kitāb* et en mettant les bases solides de la grammaire arabe.

7. Baalbaki, 2014, p. 4.

8. Le *ḥadīṭ* est le recueil qui comprend l'ensemble des traditions relatives aux actes et aux paroles du Prophète Muḥammad.

9. Baalbaki, 2014, p. 4.

Comme on s'attend, le *lahn* se notait surtout dans les couches inférieures de la société, qu'on appelle 'āmmah 'peuple, gens', facilement accusées d'ignorance et donc d'erreur linguistique. Cependant, plusieurs ouvrages accusent aussi les érudits, appelés ḥāṣṣah 'élite', tels les juristes et même les grammairiens et les écrivains.¹⁰

Certains auteurs, comme Zubaydī (m.379/989), affirmaient que le *lahn* était inconnu dans la ḡāhiliyyah 'temps préislamique' et dans ṣaḍr al-'islām 'le début de l'Islam'. La corruption de la langue a eu lieu, seulement, après la conquête islamique et le mélange des arabes avec d'autres nations.¹¹ Les érudits jugeaient mal ces changements. Par exemple, les grammairiens, ibn 'aṣḥūr (m. 663/1263 ou 669/1270) considérait la nouvelle langue un « parler » qui n'appartient à aucune langue et, 'abū ḥayyān (m. 745/1344) considérait tous les emprunts 'a ḡamiyy 'étranger'.¹²

Afin de préserver le Qur'ān et la langue du Qur'ān de toute corruption, les premiers érudits ont mis les bases pour une étude approfondie de la langue arabe. En effet, toutes les premières activités linguistiques sont liées à l'étude du Qur'ān et les lectures coraniques (*qirā'āt*). Une des premières et significatives activités qui prouvent le lien étroit entre les études linguistiques et les études religieuses est l'effort d'une chaîne ininterrompue d'érudits, pour améliorer l'écriture arabe afin de minimiser l'erreur dans la récitation et la lecture du Qur'ān.

Le grammairien 'abū l-aswad al-du'alī (m. 69/688) a mis les premières bases de la syntaxe. En plus, pour marquer la dernière voyelle de la déclinaison, il a adopté un système de *nuqaṭ al-'i'rāb* 'points de déclinaison', qu'il a jouté en couleur différente dans le Qur'ān.¹³ Ses élèves, naṣr bnu 'āṣim (m. 89/708) et yaḥyā bnu ya'mar (m. 129/746), ont adopté son système, et ont ajouté dans le Qur'ān, en couleur identique du teste, d'autres points, *nuqaṭ al-'i'ḡām* 'points diacritiques', pour différencier entre les lettres qui ont la même graphie.¹⁴

Pour différencier entre les point de la vocalisation et les point consonantiques, al-ḥalīl bnu 'aḥmad (m. 175/791), élève de maymūn al-'aqrān (m. ?), lui-même élève de ad-du'alī, a affiné le système des voyelles courtes, tel qu'on le connaît actuellement. En plus, il a introduit d'autres améliorations, en ajoutant des signes qui marquent la *hamza*, la *šaddah* 'signe de la gémation' et les voyelles longues.¹⁵

Ainsi, l'activité linguistique a commencé au deuxième/huitième siècle et, voit émer-

10. Zubaydī, *Lahn*, pp. 62-63 et Ḡāḥiṣ, *al-Bayān wa-l-tabyīn*, p. 137.

11. Zubaydī, *Lahn*, p. 59.

12. Suyūṭī, *lqṭirāḥ*, p. 36.

13. 'ad-Dānī; al-Muḥkam, pp.3-4; al-Ḡumahī, *Ṭabaqāt*, V. I, p. 12 et Faramāwī, 1978, pp. 59-71.

14. Faramāwī, 1978, pp.72-79.

15. Faramāwī, 1978, pp. 93-100.

ger deux disciplines (*'ilm*) : la science de la grammaire, *'ilm an-naḥw*, qui traite les questions liées à la syntaxe, la morphologie, la morpho-phonologie ; et la science de la langue, *'ilm al-luġah*, qui s'occupe des questions philologiques et lexicographiques, liées à la collecte des données linguistiques et aux significations des mots. Avant de devenir des sciences indépendantes, les frontières entre *'ilm al-luġah* et *'ilm an-naḥw* étaient floues, parce qu'ils sont nés étroitement liés à l'étude du Qur'ān, aux *qirā'āt* 'lectures coraniques' et à d'autres enquêtes religieuses, tels que le *Ḥadīth*, le *fiqh* 'jurisprudence' et le *tafsīr* 'exégèse'.¹⁶ Au fur et mesure, les études se sont spécialisées et quelques siècles plus tard, al-baġdādī (m. 1093/1682) énumère déjà six disciplines : *al-luġah* 'la lexicographie', *aṣ-ṣarf* 'la morphologie', *an-naḥw* 'la syntaxe', *al-ma'ānī* 'la sémantique', *al-bayān* 'science des significations et science des tropes' et *al-badī'* 'science de l'embellissement [de l'expression]'.¹⁷

Depuis les premières prémices des activités linguistiques, philologues, lexicographes et grammairiens ont commencé à faire la collecte de données fiables afin de déterminer l'usage approprié. Plusieurs auteurs étaient, à la fois, grammairiens et lexicographes et partageaient une grande partie de leurs données et de leurs sujets.¹⁸ En outre, ces premiers travaux ont fourni une grande partie de la matière première que les auteurs postérieurs ont inclus dans leurs lexiques. De conséquence, l'histoire et les étapes de ces premières activités ont conditionné la lexicographie arabe et permettent de mieux comprendre son évolution.

1.2 Collecte des données linguistiques

Sans tenir compte du développement naturel du langage et du fait que l'usage est constamment sujet au changement, les *naḥwiyyūn* 'grammairiens' et les *luġawiyyūn* 'philologues, lexicographes' étaient convaincus qu'il existait une Langue Arabe saine de *lahn* et préservée de toute corruption. Ainsi, ils ont décidé, à l'unanimité, que la langue parlée par les bédouins, qui vivaient dans le désert et loin des cités, est la forme la plus fiable et la « plus pure » de la langue arabe. En d'autres termes, elle présentait *kalām al-'arab* 'la langue des Arabes'. Conscients de la nécessité de préserver ce qu'ils considéraient l'usage correct, ils la prenaient comme *šāhid* 'témoin ; preuve' et la considéraient un matériel attesté, parce que recueilli auprès des bédouins les plus éloquents. Ainsi, pour décrire les caractéristiques morphologiques et syntaxiques et pour

16. Baalbaki, 2014, p. 2 et Naṣṣār, 1988, V. I, p. 26.

17. Baġdādī, *Ḥizānat*, V. I, p. 5.

18. Baalbaki, 2014, p. 60.

définir les sens lexicaux des mots, les œuvres grammaticales et lexicales abondent de *šawāhid* (pl. de *šāhid* 'témoin').¹⁹

Les *šawāhid* représentent les citations prises du corpus de *ši'r* 'poésie', *kalām al= 'arab* 'prose', *Qur'ān* et *Hadīth* et ont une fonction testimoniale pour justifier et garantir l'attestation et la description du bon usage et de la norme.

Pour le choix des bédouins et des tribus, les philologues suivaient des critères solides. Ils prenaient en considération l'éloquence de l'informateur bédouin qui doit être capable de générer la rime et d'utiliser une forme de langue arabe caractérisée par un degré élevé de précision dans le choix des mots. Une autre considération très significative est la prédisposition naturelle et innée (*salīqah wa=saġiyyah*) de l'éloquence de l'informateur bédouin qui devait adhérer instinctivement à son usage dialectal et résister à toute tentative de le changer. Cette caractéristique le mettait à l'abri, aussi bien de la corruption des langues étrangères que de l'influence des grammairiens, dont le discours était considéré artificiel et prétentieux. Un autre critère surprenant, il devait être illettré, ce qui prouvait davantage que sa *faṣāḥah* 'éloquence' est due à une prédisposition naturelle et non à une connaissance acquise.²⁰ En outre, le bon informateur bédouin, digne de confiance, devait provenir d'une tribu située loin des frontières avec les pays étrangers et sans contact avec les non-arabes.

Pour ces raisons, les tribus sont prises en considération, non seulement pour leur éloquence, mais aussi pour leur situation géographique isolée, telle que, les tribus qays, tamīm et 'asad et dans une moindre mesure huḍayl et des parties de kinānah et de ṭayyi'. Par contre, pour les mêmes raisons, plusieurs autres tribus ont été exclues ; laḥm et ġuḍām, parce qu'elles étaient voisines du peuple égyptien ; namir voisine des grecques ; bakr voisine au royaume nabatéen et persan. Les tribus de 'abdu l=qays et 'azd 'umān ont été, également, exclues parce qu'elles se mêlaient avec le peuple indien et persan, par contre le peuple du Yémen a été exclu car il se mêlait avec les Indiens et les Abyssins. On voit exclure les gens de quḍā'ah et ġassān, parce qu'ils sont chrétiens et priaient en langue non arabe, en plus ils divisaient leurs frontières avec *bilād aš=šām* 'pays du Levant'. Très significatives l'exclusion de banī ḥanīfah et des habitants de al=yamāmah, ṭaqīf et aṭ=ṭā'if, parce qu'ils se sont mêlés avec les commerçants étrangers et l'exclusion des cités du ḥiġāz parce qu'ils fréquentaient d'autres pays.²¹

En contradiction avec ceci, les érudits considéraient, à l'unanimité, que les qu-

19. Baalbaki, 2014, p. 6 et 'aṣ-Ṣāliḥ, 2009, p. 23.

20. Baalbaki, 2014, pp. 11-15; Ibn Ġinnī, 'al-Ḥaṣā'is, pp. 76 - 80.

21. Suyūfī, Iqtirāḥ, pp. 47-48.

rayšyyūn sont les Arabes les plus éloquents, bien que qurayš est une tribu commerçante et donc en contact avec les étrangers, en plus d'être le lieu de rencontre des différentes tribus, en particulier pendant la saison de pèlerinage à la Mecque. Les savants justifient cette décision par le fait que les qurayšyyūn étaient capables de sélectionner et d'adopter les meilleurs dialectes, justement parce qu'ils sont en contact continu avec les autres tribus. En réalité, le fait que Dieu ait choisi le Prophète Moḥammad parmi les qurayšyyūn et la position religieuse de qurayš constituent la raison principale derrière l'attribution à qurayš du degré le plus élevé de *faṣāḥah* parmi les Arabes.²² Cependant, il est tout à fait révélateur le fait que qurayš ne soit pas mentionnée parmi les tribus dont dépendaient les philologues dans la collecte de leurs données.²³

1.3 Les sources fiables

Baḡdādī (m. 1093/1682) a consacré son œuvre, *ḥizānatu al='adabi wa=lubbu lubābi lisāni al='arabi*, pour expliquer les *šawāhid* 'témoignages ; preuves' utilisés par al='astarābādī (m. 686/1287 ou 688/1289) dans son œuvre *al=kāfiyyah*. Il spécifie que le matériel considéré comme source fiable (*mā yaṣiḥḥu l='istišhāda bi=hi*) peut être *ši'run wa=ḡayruh* 'poésie et autres (la prose)'. Il ajoute que, pour les sciences de *al=lughah* 'lexicographie', *aṣ=ṣarf* 'morphologie' et *an=naḥw* 'syntaxe', on doit faire appel seulement à *kalām al='arab* 'la langue des arabes'. Par contre, pour les sciences de *al=ma'ānī* 'sémantique', *al=bayān* 'sciences des significations et des tropes' et *al=badī'* 'sciences de l'embellissement de l'expression', on peut faire appel aussi aux *muwalladīn* (les poètes et les littéraires contemporains à la période de témoignage, voir ci-dessous), parce que ces trois dernières disciplines s'occupent des significations qui sont liées à la raison (*al='aql*) et, dans ce cas, il n'y a pas de différence entre un arabe et un étranger.²⁴

1.3.1 Les citations en poésie

Vu les particularités de la poésie, les érudits ont estimé que la poésie devait être classée en fonction de son époque. L'un des auteurs les plus anciens à adopter une classification « temporelle » est ibn sallām al=ḡumaḥī (m. 232/846).²⁵ Il a classifié les poètes en trois groupes : *'ahlu l=ḡāhiliyyah* qui ont vécu seulement pendant la *ḡāhiliyyah* 'période avant l'Islam' et dont la poésie était le recueil de la sagesse et de la

22. Suyūṭī, Muzhir, V. I, pp. 210-211

23. Baalbaki, 2014, p. 9.

24. Baḡdādī, *Ḥizānat*, V. I, p. 5.

25. Baalbaki, 2014, p. 33.

science des arabes ; *al=muḥaḍramūn* 'mixtes ; mélangés', qui ont vécu pendant la période chevauchant la *ḡāhiliyyah* et l'islam et enfin *'ahlu l= 'islām* 'les gens de l'islam' qui ont vécu seulement après la venue de l'islam.²⁶ Quelques siècles plus tard, baḡdādī (m. 1093/1682) reconnaît quatre catégories de poètes :²⁷

1. *al=ḡāhiliyyūn* sont les poètes qui ont vécu avant l'islam et appartiennent à la période préislamique la *ḡāhiliyyah*, comme imru' u l-qays et *al= 'a 'šā*.
2. *al=muḥaḍramūn* sont les poètes qui ont vécu avant et au début de l'islam, comme labīd bnu rabī'ah (m. 40/660) et ḥassān bnu ṭābit (m. 40/660).
3. *al=mutaqaddimūn*, appelés aussi *'islāmiyyūn*, sont les poètes qui ont vécu au début de la période islamique *ṣadr al= 'islām*, comme *al=farazdaq* (m. 110/728) et ḡarrīr (m. 110/728).
4. *al=muwalladūn*, appelés aussi *muḥdaṭūn*, sont les poètes qui ont vécu après la période *ṣadr al= 'islām*, comme baššār bnu burd (m. 167/784) et 'abū nuwās (m. 198/814).

Baḡdādī (m. 1093/1682) spécifie qu'on peut prendre à témoin toute la poésie de la première et de la deuxième catégorie et, il est légitime de prendre en considération la troisième mais, jamais la quatrième catégorie.²⁸ Ainsi, on ne peut que constater l'exclusion, du corps des *šawāhid*, de la poésie de poètes fameux et célèbres, comme 'abū nuwās (m. 198/814), ibn ar=rūmī (m. 283/896), *al=buḥturī* (m. 284/898), *al=mutanabbī* (m. 354/965) et 'abū l= 'alā' *al=ma'arrī* (m. 449/1057), parce qu'ils étaient considérés *muwalladīn*.²⁹

Toutefois, on note un débat et une évolution au cours des siècles concernant l'utilisation de la poésie des *muwalladūn*. Le lecteur du Qur'ān, 'abū 'amr bnu al= 'alā' (m. 154/770), considéré le fondateur de la grammaire arabe de l'école de baṣra, ne citait jamais *al=farazdaq* (m. 110/728) et ḡarrīr (m. 110/728) parce que, contemporains à lui, il les considérait *muwalladīn*. Alors que, Baḡdādī (m. 1093/1682), les considérait une autorité dans l'histoire de la poésie arabe et jugeait leurs poésie digne d'être *šawāhid* parce que, selon lui, elle appartenait à la période classique.

26. *al-Ġumahī, Ṭabaqāt*, V. I, p. 23-24.

27. *Baḡdādī, Hizānat*, V. I, pp. 5-6.

28. *Baḡdādī, Hizānat*, V. I, p. 6.

29. Cependant, les lexicographes étaient plus tolérants que les grammairiens, par exemple, dans *Lisān al= 'Arab*, ibn manzūr (m.711/1311) a utilisé vingt six *šawāhid* tirés de la poésie des *muwalladūn*: trois de Abū Nuwās, huit de Abū Tammām, un de Buḥturī, onze de Mutanabbī, trois de Ma'arrī. Il reste quand même à noter qu'il a utilisé des centaines de *šawāhid* tirés de la poésie des *ḡāhiliyyūn* et des *'islāmiyyūn* (Baalbaki, 2014, pp. 35-36).

Ces contradictions sont bien expliquées par ibn rašīq (m. 463/1071) dans son œuvre, *al=ʿumdatu fī maḥāsini š-šīʿri wa-ʿādābihi wa-naqdihi*. Dans *bāb fī al=qudamāʾi wal-muḥdaṭīn* 'chapitre sur les classiques et les modernes', il explique que tout poète est *muwallad* dans son époque. Selon lui, les érudits, comme ʿabū ʿamr bnu al=ʿalā (m. 154/770), citaient les poètes qui les ont précédés et excluaient al=farazdaq (m. 110/728) et ġarrīr (m. 110/728), parce qu'ils les considéraient *muwalladīn* et n'avaient pas confiance en leurs poésies. Du passage, *kullu wāḥidin min=hum yaḍhabu fī ʿahli ʿašri=hi hādā l=maḍhabi* 'chacun d'eux [grammairien/philologue] applique cette doctrine aux gens de son époque', on comprend que le terme *muwallad* est réservé au néologisme, que les *muwalladūn* sont les contemporains du linguiste ou du lexicographe et, que ces derniers se méfiaient du néologisme qui caractérisait les nouveaux poètes.³⁰

De sa part, ibn ġinnī (m. 392/1002), dans son œuvre *al=ḥaṣāʾiṣ*, donne un exemple de la poésie de son contemporain al=mutanabbī (m. 354/965) et ajoute :

وَلَا تَسْتَنْكِرْ ذِكْرَ هَذَا وَإِنْ كَانَ مُوَلَّدًا، [...] فَإِنَّ أَلْمَعَانِي يَتَنَاهَبُهَا أَلْمُوَلَّدُونَ كَمَا
 يَتَنَاهَبُهَا أَلْمُتَقَدِّمُونَ

'ne blâmes pas le fait de mentionner cet homme même s'il est *muwallad*
 [...] car les *muwalladūn* concourent [pour arriver aux] significations comme
 le font les *mutaqaddimūn*'.³¹

Dans son livre, *al=ʿumdatu fī maḥāsini aš-šīʿri wa-ʿādābihi wa-naqdihi*, dans *bāb al=maʿānī al=muḥdaṭah* 'le chapitre des significations qui [n'existaient pas et] ont été créées' (le néologisme), l'auteur Ibn Rašīq (m. 463/1071) lui donne raison. Il ajoute que, si on contemple bien, on verra que les *islāmiyyūn* ont ajouté, dans leurs poésies, des significations qui ne se trouvaient pas dans celles des *qudamāʾ* (les ancêtres) et des *mutaqaddimūn*. Il fait noter qu'il y a des néologismes et des innovations merveilleuses dans les poésies des poètes de la catégorie de Ġarrīr (m. 110/728) et de al=Farazdaq (m. 110/728) qu'on trouve rarement dans celles des ancêtres. Ibn Rašīq a consacré tout le chapitre aux poésies dites «modernes». En les comparant avec les poètes «classiques», il fait noter que les nouveaux ont excellé et ont inventé des sens qui ne se trouvaient pas avant : *lam yusbaq ʿilay=hi* (personne ne l'a précédé [à exprimer ce sens]); *lam yatanāwal=hu ʿaḥadun* (personne ne l'a utilisé [le sens]); *mā ʿalim=tu ʿanna=hu subiqa ilay=hi* (je ne connais personne qui a précédé [à exprimer ce sens]).³² Il ajoute que les *maqamāt* 'histoires en prose et rythme', les temps et les pays peuvent

30. Ibn Rašīq, 'Umdah, V.I, pp. 137 - 142

31. Ibn Ġinnī, 'al-Ḥaṣāʾiṣ, V. I, p. 24.

32. Ibn Rašīq, 'Umdah, V. I, pp. 985-1000.

changer, et de conséquence, ce qui est admirable dans un temps ne l'est pas dans un autre et ce qui est appréciable par les gens d'un pays ne l'est pas par d'autres.³³

En conclusion, les savants de la première phase, comme 'abū 'amr bnu al-'alā' et sibawayhi, faisaient la collecte des *šawāhid* strictement auprès des bédouins, ils étaient conservateurs et, afin de préserver la langue arabe de toute corruption, ils excluaient tout phénomène de néologisme par méfiance. Au cours des siècles, la spécialisation des sciences linguistiques a ouvert d'autres horizons d'études et a permis la naissance de savants illuminés, comme ibn ġinnī et ibn rašīq, qui étaient plus sensibles à l'évolution de la langue arabe.

Cependant, toutes les œuvres, anciennes et plus récentes, concordent que le but sacré dans l'étude de la poésie est comprendre le Qur'ān,³⁴ comme l'affirmait ibn 'abbās (m. 68/687) et l'a reporté ibn rašīq (m. 463/1071) :

وَكَانَ ابْنُ عَبَّاسٍ يَقُولُ، إِذَا قَرَأْتُمْ شَيْئًا مِنْ كِتَابِ اللَّهِ فَلَمْ تَعْرِفُوهُ، فَاطْلُبُوهُ فِي
 أَشْعَارِ الْعَرَبِ، فَإِنَّ الشَّعْرَ دِيْوَانُ الْعَرَبِ

'Et ibn al-'abbās disait : si vous lisez quelque chose dans le Livre de Allah et vous ne la connaissez pas, cherchez le, donc, dans les poésies des arabes, car la poésie est, certainement, l'encyclopédie (*dīwān*) des arabes'.³⁵

On note que ibn 'abbās dit *walam ta 'rifūh* 'vous ne la connaissez pas', du verbe 'arafa 'savoir ; connaître', et ceci met le point sur un des intérêts principaux des savants. Ils ont noté que le Qur'ān et le Ḥadīth abondaient de mots inconnus et rares et de constructions peu courantes, qu'ils ont appelé *ġarīb* 'étrange, étranger'. De conséquence, ils les cherchaient dans la poésie et autres, dans le but de les comprendre et d'être capables de les expliquer.³⁶

Ibn fāris (m. 395/1004), aussi, confirme que la poésie est l'encyclopédie des arabes, grâce à laquelle on apprend les descendances et on connaît les exploits et, à partir de laquelle on comprend la langue. Il ajoute que la poésie constitue une *ḥuġġah* 'preuve ; fonte autorisée', pour expliquer ce qui est compliqué et *ġarīb* dans Qur'ān et dans le Ḥadīth.³⁷

La recherche des mots *ġarīb* et des constructions peu courantes a eu une grande influence sur l'évolution de la lexicographie arabe. La grande majorité du matériel *ġarīb*

33. Ibid., V. I, p. 141.

34. Baalbaki, 2014, p. 43.

35. Ibn Rašīq, 'Umdah, V. I, p. 27.

36. Baalbaki, 2014, p. 43.

37. Ibn Fāris, Ṣāhibī, p. 212 et Baalbaki, 2014, p. 33.

est constitué de la poésie de type *rağaz*, caractérisée par une structure métrique facile et flexible parce qu'elle était composée par les chameliers et utilisée dans les chansons populaires.³⁸

1.3.2 Les citations en prose

La prose, appelée *kalām al=‘arab* 'la langue des arabes', est constituée par un corpus formé de proverbes et d'expressions idiomatiques. En plus de *kalām al=‘arab*, le Qur'ān et le *Ḥadīṭ* ont été pris en considération comme *šawāhid*, parce qu'ils étaient conçus également à une époque antérieure. Cependant, ils ont posé des problèmes méthodologiques pour les philologues et les grammairiens.³⁹

En ce qui concerne le Qur'ān, on trouve plusieurs *qirā'āt* 'récitations, lectures', car il a été transmis de manière orale grâce aux *ḥamalāt al=Qur'ān* 'transmetteurs du Coran'. Comme le précisent les ressources, il y a eu plusieurs variations dans la récitations et la lecture du Qur'ān, selon des critères phonétiques et dialectaux mais, aussi, selon certains points de vue différents de *'i'rāb*. Ibn muğāhid (m. 324/ 936) est un des premiers à analyser les *qirā'āt*. Il a attesté celles qui seront canoniques et les a appelées les *qirā'āt as=sab* 'les sept lectures' et, il a considéré les autres *šāḍḍah* 'anomal ; bizarre ; aberrant ; déconcertant'. Plusieurs savants, après lui, ont expliqué et justifié les *qirā'āt as=sab* canoniques.⁴⁰

Mais, comme dans le cas de la poésie, on note un changement d'opinion qui évolue avec le progrès des disciplines linguistiques. L'érudit, ibn ġinnī (m. 392/1002), a analysé les *qirā'āt aš=šāḍḍah* 'les lectures non canoniques'. Dans son introduction, il justifie le but de son œuvre en disant :⁴¹

وَلَعَلَّهُ، أَوْ كَثِيرًا مِنْهُ، مُسَاوٍ فِي الْفَصَاحَةِ لِالْمُجْتَمَعِ عَلَيْهِ [...] لَكِنَّ غَرَضًا مِنْهُ أَنْ
نُرِي وَجَهَ قُوَّةِ مَا يُسَمَّى آلَانَ شَاذًا وَأَنَّه ضَارِبٌ فِي صِحَّةِ الرِّوَايَةِ بِحَيْرَانِهِ
Il se peut qu'elles [les *qirā'āt aš=šāḍḍah*], ou la plupart d'entre elles, égalisent en éloquence celles qui ont été attestées. [...] Mais notre but est de démontrer l'importance de ce qui est appelé, aujourd'hui, *šāḍḍ* 'aberrant ; déconcertant' et prouver qu'il égalise à raconter l'histoire comme les autres.

Ainsi, il «légalise» (*wa=kullu dālika ġā'izun ḥasanun*) les différentes vocalisations dues aux différentes prononciations dialectales, comme *ẓulumāt - ẓilimāt - ẓulamāt- ẓulmāt*

38. Baalbaki, 2014, pp. 44-45.

39. Baalbaki, 2014, pp. 29-30.

40. Ibn Ġinnī, 'al-Muḥtasab, V. I, pp.8-10.

41. Ibid., V. I, p. 32.

'obscurités'.⁴² Il explique les aspects phonétiques qui justifient les différentes lectures, par exemple, lire *yaḥtaṭṭif* ou encore *yihittif* au lieu de *yaḥtaṭṭif* 'dérober ; enlever'.⁴³

En outre, il donne, comme *ḥujjah* 'preuve ; témoignage', une poésie de al=mutanabbī, son contemporain. Il se justifie que, pour expliquer le Qur'ān, il n'y a aucune raison de ne pas citer al=mutanabbī, bien qu'il soit *muḥdaṭ* et ajoute :⁴⁴

فَأَنَّ الْمَعَانِي لَا يَرْفَعُهَا تَقَدُّمٌ وَلَا يُزِيرِي بِهَا تَأَخُّرٌ

'Surement que les significations ne sont ni exhaussées par l'ancienneté (ou le classicisme) ni abaissées par le modernisme'.

En ce qui concerne le Ḥadīṭ, les érudits ne le considéraient pas valable, comme *šawāhid*, pour trois raisons : i) la transmission était orale ; ii) les *ruwwāt al=Ḥadīṭ* 'transmetteurs du Ḥadīṭ' permettaient que la transmission soit selon le sens de ce qu'a dit le Prophète et non textuelle (*ḡawwazū an=naqla bi=l=ma'nā*) et enfin iii) certains transmetteurs n'étaient même pas des locuteurs natifs.⁴⁵ Cependant, dès les premières œuvres, on note une divergence entre grammairiens et lexicographes.

Le but des grammairiens était de formaliser les lois grammaticales et syntaxiques, en étudiant la pratique linguiste des locuteurs bédouins et ils ne basaient pas leurs justifications sur le Ḥadīṭ. À titre d'exemple, dans son livre de grammaire, le *kitāb*, sībawayhi (m. 180/796) a cité comme *šawāhid*, 1050 poésies, 447 versets coraniques, 350 expressions idiomatiques, 41 proverbes et seulement 7 ou 8 Ḥadīṭs.

Le but des lexicographes, par contre, était de faire la récolte d'un corpus de paroles utilisées, et ils n'avaient aucune scrupule d'utiliser le Ḥadīṭ, pour prouver l'existence d'un mot ou en clarifier le sens. Par exemple, al=ḥalīl (m. 175/791), contemporain et maître de sībawayhi, cite 428 ḥadīṭ dans son lexique *kitāb al='ayn*. Pour une comparaison diachronique, ibn durayd (m. 321/933) cite 508 Ḥadīṭ dans *ḡamharat al=lughah*, et enfin le lexique *al=nihāyah fī ḡarīb al=Ḥadīṭ* de ibn al='aṭīr (m. 606/1210) est entièrement dédié aux lexèmes du genre *ḡarīb* dans le Ḥadīṭ.⁴⁶

1.4 Époque de la collecte

Depuis le début du deuxième/huitième siècle jusqu'au milieu du troisième/neuvième siècle, environ, les premiers érudits ont contribué à la collecte (*ḡam'*) de *kalām al='arab*,

42. Ibid., V. I, p. 56.

43. Ibid., V. I, p. 59.

44. Ibid., V. I, p. 231.

45. *Suyūṭī, Iqtirāḥ*, pp. 43-46.

46. *Baalbaki, 2014*, pp. 29-30.

constituant un corpus précieux qui est à la base de toutes les études linguistiques arabes. Plusieurs, d'entre eux, ont fait le voyage et ont séjourné dans le désert, pour apprendre la langue de la bouche des bédouins. L'exemple le plus significatif est l'auteur du premier lexique, *al=ʿayn*, *al=ḥalīl bnu ʿaḥmad* (m. 175/791), qui a séjourné dans le désert auprès des tribus *ḥiḡāz*, *naḡd* et *tihāma*.⁴⁷

On note aussi la migration de plusieurs bédouins du désert vers *al=ʿamṣār* 'les villes ; zones urbaines', où plusieurs d'entre eux sont devenus eux-mêmes philologues et professeurs de langue arabe. Cependant, plusieurs d'entre eux furent accusés d'avoir perdu leur *faṣāḥah* 'éloquence'. À cause du long séjour dans les cités et loin du désert, leur parler s'était corrompu.⁴⁸

De surcroît, pour satisfaire les besoins des érudits, certains informateurs bédouins ont commencé à inventer des poésies.⁴⁹ Découvrant ces falsifications, les érudits les dénonçaient. Par exemple, dans l'introduction de son lexique *kitāb al=ʿayn*, *al=ḥalīl* a mis en garde contre les *naḥārīr* (pluriel de *naḥīr* 'rusé ; roué ; malin') et contre les paroles *muwalladah* et *mubtadaʿah* 'inventées' qui ne doivent pas être incluses dans le discours des arabes.⁵⁰

Le phénomène de falsification a pris tellement d'ampleur que plusieurs érudits ont dédié un chapitre entier pour le dénoncer, par exemple, *suyūṭī* (m. 911/1505) a inclus, dans *muzhir*, un chapitre intitulé *bāb maʿrifati l-maṣnūʿ* (chapitre de la détection du faux).⁵¹ D'autres ont écrit des œuvres entières pour aider à distinguer entre l'original et le faux, par exemple, *ṭabaqāt fuḡūl al=ṣuʿarāʿ* écrite par *ibn sallām al=ḡumaḥī* (m. 232/846).

À cause de ces événements, les érudits ont commencé à adopter des méthodes critiques. Pour démontrer la crédibilité du matériel qu'ils citent, ils ont commencé à citer les bédouins qu'ils ont entendus.⁵²

La méthode de citer les émetteurs, appelée *al=ʿisnād*, était déjà utilisée par les érudits du *Ḥadīṭ* (les *muḥaddīṭūn*) et constitue le pivot pour valider la fiabilité d'un transmetteur de *Ḥadīṭ*. Cependant, les philologues ne suivaient pas de manière rigoureuse

47. Ibid., p. 17.

48. Naṣṣār, 1988, V. I, p. 23 et Baalbaki, 2014, pp. 18-21.

49. Comme, par exemple, l'histoire que raconte *al=ḡumaḥī* (m. 232/846) concernant le neveu d'un poète bédouin qui, terminé la poésie de son grand-père, en inventait d'autres en les attribuant toujours à son aïeul (*al=ḡumaḥī*, *Ṭabaqāt*, V. I, pp. 47-48) ; ou l'histoire de 'Abān *al=Lāḥiqī* (m. 200/815) qui a inventé une ligne de poésie pour répondre à la demande de *Sībawayhi* s'il connaît un *ṣāhid* qui peut illustrer le rôle syntaxique de l'adjectif de schème C₁aC₂iC₃ (*faʿil*) (*Suyūṭī*, *Muzhir*, V. I, p. 180).

50. Baalbaki, 2014, p. 28.

51. *Suyūṭī*, *Muzhir*, V. I, pp. 171-183.

52. *Suyūṭī*, *Muzhir*, V. I, p. 141.

les règles du *'isnād*, comme les érudits du Ḥadīṭ.⁵³ Le *muḥaddīṭ* doit assurer une chaîne ininterrompue de transmetteurs, pour garantir la validité de la transmission. Le philologue, par contre, peut accepter ce que atteste même une seule personne à condition qu'elle soit crédible.⁵⁴

En conclusion, l'objectif principale des lexicographes était la formation d'un corpus de la langue arabe, comme elle a été pratiquée à l'époque de la *ḡāhiliyyah* et au début de l'Islam. Jugeant que la langue des bédouins était pure et non contaminée, ils ont commencé à la prendre comme exemple, pour expliquer et formaliser la langue arabe. Mais, l'immigration et le séjour prolongé des bédouins dans les zones urbaines et le manque de méthodes philologiques solides ont eu pour conséquence que les lexicographes ont commencé à douter de la *faṣāḥah* et la fiabilité des transmetteurs bédouins. Le doute a commencé à immerger dans les lexiques sous plusieurs formes, comme par exemple, la phrase *laysa bi-ṭabt* (il n'est pas attesté) qu'utilise souvent Ibn Durayd (m. 321/933) dans son lexique *al-ḡamharah*.

Dans le cas de la poésie, les critères appliqués, pour l'admission d'une poésie comme *ṣāhid*, étaient très stricts. Comme indiqué ci-dessus, déjà au deuxième siècle, on ne mentionnait aucune poésie *muwalladah* 'contemporaine' dans les ouvrages grammaticaux et peu dans les lexiques.

Dans le cas de la prose, depuis la fin du deuxième/huitième siècle, les érudits ne prenaient plus directement des arabes sédentaires. Par contre, la collecte directe des bédouins qui vivaient dans le désert continuait. Par exemple, dans l'introduction de *tahḏīb al-luḡah*, al-'Azharī (m. 370/981) affirme que la collecte des données de son lexique est le résultat de la transmission directe de bédouins auprès desquels il a été prisonnier.⁵⁵ Dans l'introduction du lexique *aṣ-ṣaḥāḥ*, al-ḡawharī (m. 400/1010) affirme qu'il n'a introduit une parole qu'après avoir consulté les arabes *al-'āribah* 'qui parlent arabe de manière éloquent', *fī diyārihim bi-l-bādiyyah* 'qui habitent dans le désert'.⁵⁶

Mais à la fin du quatrième/dixième siècle, on ne trouve que des témoignages sporadiques de la transmission directe même des bédouins du désert.⁵⁷

53. Baalbaki, 2014, p. 24.

54. Suyūṭī, Muzhir, V. I, pp. 113-137.

55. al-'Azharī, Tahḏīb, V. I, pp. 6-7.

56. Ḡawharī, Ṣaḥāḥ, V. I, p. 33.

57. Baalbaki, 2014, p. 32.

Chapitre 2

Lexiques de la langue arabe

2.1 Introduction

La cause principale de l'apparition des études linguistiques est la volonté de comprendre et d'expliquer le texte coranique et l'interprétation du Ḥadīṭ. Les prémices de ces activités scientifiques ont vu le jour durant la vie du Prophète qui expliquait ce qui n'était pas clair dans le Qur'ān et ceci explique le lien très étroit entre les études linguistiques et l'exégèse coranique.¹ Après la mort du Prophète en l'an 10/632, ces activités ont continué avec les *ṣaḥābah* 'Compagnons [du Prophète]' dont le plus fameux est ibn al-'abbās (m. 68/687) qui tenait un *maǧlis* 'assemblée' pour répondre aux questions des gens. Le *maǧlis* était tellement affolé que ibn al-'abbās recevait, d'abord, ceux qui voulaient lui poser des questions sur le Qur'ān et ses *ḥurūf* 'paroles ou modes de lecture possibles'. Il rencontrait ensuite les personnes qui cherchaient des réponses à des questions relatives à : i) *tafsīr al-Qur'ān* 'l'interprétation du Qur'ān', ii) *al-ḥalāl* 'le licite' et *al-ḥarām* 'l'interdit' ; iii) *al-fiqh* 'la jurisprudence' et iv) *al-farā'id* 'les obligations religieuses'. Enfin, il recevait les personnes qui ont des demandes relatives à : i) *al-'arabiyyah* 'la langue arabe', *aš-ši'r* 'la poésie' et iii) *al-ǧarīb* 'les usages linguistiques étranges et peu communs'.²

L'adhésion en masse à l'islam des non Arabes a eu pour conséquence la naissance d'une langue de communication, entre Arabes et non Arabes. La nouvelle langue était jugée corrompue et non arabe parce que, dans ses évolutions, elle s'éloignait de la langue sacrée du Qur'ān. L'un des aspects de la corruption linguistique touchait justement le lexique. On peut citer, par exemple, l'utilisation non correcte des paroles, la naissance de nouveaux mots et l'emprunt de paroles des langues étrangères.

1. Naṣṣār, 1988, V. I, p. 27.

2. Baalbaki, 2014, p. 39.

Les érudits ont commencé à craindre que le Qur'ān et le Ḥadīṭ ne deviennent incompréhensibles par suite de la disparition de la langue dans laquelle le premier a été révélé et le deuxième a été transmis. Ainsi ils ont commencé à faire la collecte de *kalām al=ʿarab*, en prose et poésie, auprès des bédouins. La période de la collecte, appelée *ʿuṣūr al=iḥtiḡāḡ*, a commencé au milieu du deuxième/huitième siècle avec des érudits tels que ʿabdullāh bnu ʿabī ʾiṣḡāq (m. 117/735) et ʿabū ʿamr bnu al=ʿalāʾ (m. 154/770). Elle a continué au milieu du troisième/neuvième siècle avec des savants comme al=māzinī (m. 249/863), ʿabū ḥātim as=siḡistānī (m. 255/869). La collecte a pris fin, généralement au quatrième siècle, lorsque les érudits ont estimé que les bédouins ont perdu leur éloquence et leurs parler ne correspond plus à celui de l'époque de la révélation.

Un autre aspect important, toujours lié aux activités de l'exégèse et des lectures coraniques, est le fait que les érudits se sont aperçus qu'il y a des aspects linguistiques, dans le Qur'ān, qui sont différents de ce qu'ils ont décidé, à l'unanimité, comme Langue Arabe. En effet, le Qur'ān contient beaucoup de paroles étranges et des utilisations syntaxiques peu communes appelées *ḡarīb* 'étrange, étranger'. En plus, ils se sont aperçus que la plupart du *ḡarīb* dans le Qur'ān provient des variantes dialectales de différentes tribus et même des langues étrangères différentes de la langue arabe.

Tenir la liste des paroles *ḡarīb* qui se trouvent dans le Qur'ān et noter leurs origines tribales ou étrangères est, peut être, la première activité lexicale, connue comme *ḡarīb al=Qur'ān*. Durant ces premiers pas, les érudits dictaient à leurs disciples, sans ordre précis, les paroles jugées *ḡarīb* dans le *Qur'ān* et prenaient note de ce qu'ils entendaient auprès des bédouins également sans ordre précis.³

En évoluant, l'activité linguistique s'est enrichie, au point que al=ḥalīl (m. 175/791) a eu l'idée d'ordonner les données lexicales et composer le premier lexique général, qu'il a nommé *kitāb al=ʿayn*. Dans la même époque, d'autres lexiques sont apparus mais, traitant d'un thème sémantique unique concernant, par exemple, les insectes ou les chevaux ou le corps humain, etc. Ainsi, deux types de lexiques ont pris naissance :⁴

1. le lexique de type onomasiologique, appelé *al=mubawwab*, qui traite un domaine thématique ou un genre spécifique, et dans lequel c'est le *maʿnā* 'sens' qui conduit au *lafẓ* 'signe' ;
2. le lexique de type sémasiologique, appelé *al=muḡannas*, dont l'objectif est répertorier tous les éléments lexicaux de la langue, et dans lequel les lemmes sont

3. Naṣṣār, 1988, V. I, p. 26 et p. 29

4. Baalbaki, 2014, pp. 46-48.

uniformément classés selon des caractéristiques liées au *lafz* 'signe' et non au *ma'nā* 'sens'.

Le fait que le lexique *mubawwab* 'onomasiologique' et le lexique *muğannas* 'sémasiologique' ont coexisté, pendant une longue période, prouve que chacun d'eux avait un but différent dans la tradition lexicale arabe. Un autre fait important est que certains auteurs, comme ibn sīda (m. 458/1066), ont écrit des œuvres des deux types. Ibn sīda décrit son lexique sémasiologique, *al=muḥkam*, destiné à guider le chercheur à l'endroit exact où un mot est enregistré. Par contre, il précise que son lexique onomasiologique, *al=muḥaṣṣaṣ*, réunit les différents synonymes dans le cas où *al=musammā* 'celui qui est nommé' (le signifié) a plusieurs *ism* 'nom' (le signifiant) et où *al=mawṣūf* 'le qualifié' a, pour le décrire, plus de *ṣifah* 'adjectif'. Ainsi, il recommande son lexique onomasiologique, surtout, à *al=ḥaṭīb* 'l'orateur' et à *aš=šā'ir* 'le poète', puisqu'il leur donne la possibilité de choisir, parmi les différents synonymes, ceux qui correspondent à leurs besoins.⁵

Les lexiques *mubawwab* ont fourni une grande partie de la matière première pour les lexiques *muğannas* postérieurs. Par exemple, pour écrire son lexique, *maqāyīs al=luğah*, ibn fāris (m. 395/1004) déclare avoir utilisé le matériel de cinq sources, dont trois *mubawwab* et deux *muğannas*.⁶ Pour écrire le fameux lexique *lisān al='arab*, ibn an-nadīm (m. 380/990) mentionne vingt et un *mubawwab* écrits, principalement, par des auteurs du troisième/neuvième siècle.⁷ Les lexiques *muğannas* abondent de proverbes que les auteurs ont tiré surtout des monographies traitant *al='amṭāl* 'les proverbes'. Par exemple, on trouve 70 proverbes dans *al='ayn* de al=ḥalīl (m. 175/791), 315 dans *al=ğamharah* de ibn durayd (m. 321/933) et 2278 proverbes dans *lisān al='arab* de ibn manzūr (m. 711/1311).⁸ Mais, on assiste aussi au phénomène inverse où le matériel d'un lexique *muğannas* constitue la base pour rédiger un lexique spécialisé *mubawwab*. Par exemple, les 357 fragments reconstruits du *kitāb al='i'tiqāb*, de 'abū turāb (m. à la fin du troisième/neuvième siècle), sont constitués de matériel tirée, surtout, du *tahqīb al=luğah*, lexique *muğannas* écrit par al='azharī (m. 370/981). Et *kitāb al='i'tiqāb*, lui même, constitue une des sources utilisées par ibn manzūr (m. 711/1311), pour écrire son *muğannas*, *lisān al='arab*.⁹ Pour rédiger son lexique *mubawwab*, *al=Muhassaṣ*, ibn sīda (m. 458/1066) a cité, comme source deux célèbres lexiques *muğannas*, *al='ayn* de

5. Ibn Sīda, *Muḥaṣṣaṣ*, V. I, p. 10 et Baalbaki, 2014, p. 48.

6. Baalbaki, 2014, p. 48.

7. Ibid., p. 86.

8. Ibid., p. 102.

9. Ibid., p. 229.

al=ḥalīl bnu 'aḥmad (m. 175/791) et *ḡamharat al=luḡah* de ibn durayd (m. 321/933).¹⁰

Pour ce fait, il est inéluctable d'étudier les différents types de dictionnaires, pour avoir une idée générale et complète sur la lexicographie arabe. La prochaine section est dédiée à l'étude non exhaustive, mais générale, des dictionnaires *mubawwab* et la section suivante est dédiée aux dictionnaires *muḡannas*.

2.2 Lexiques onomasiologiques

La grande partie des travaux de type *mubawwab* est une *risālah* 'courte monographie', sous forme d'une récolte de paroles restreintes à un sujet spécifique.

Le thème peut être lexicale traitant de : *al=ḡarīb* 'étrange' ; *al=mu'arrab* 'les paroles arabisées' ; *al=muwallad* ou *al=muḥdaṭ* 'les nouvelles paroles' (néologisme) ; *al=lahn* 'les erreurs ou les variantes dialectales' (solécisme) ; *al=mutarādif* 'les synonymes' ; *al=muštarak* 'les homonymes' ; *al='aḡdād* 'l'énantiosémie' ; *ḡurūf al=mu'ḡam* 'les lettres' ou *ḡurūf al=ma'ānī* 'les particules'.

On trouve, aussi, des monographies qui traitent un sujet linguistique lié à *aṣ=ṣarf* 'la flexion nominale et la conjugaison verbale' ou *al='ištiqāq* 'la dérivation'.

D'autres monographies traitent un sujet sémantique lié aux animaux, aux plantes ou autres. On trouve aussi des lexiques *mubawwab* multi-thématiques, qui sont de vrais thésaurus subdivisés en section dont chacune traite un thème différent.

2.2.1 Domaine lexicale

2.2.1.1 Le ḡarīb

La définition du *ḡarīb* n'est pas facile à saisir selon les lexiques et les corpus qui nous sont parvenus. Selon 'abū sulaymān al=ḥaṭṭābī (m. 388/998) :¹¹

الْغَرِيبُ مِنَ الْكَلَامِ إِثْمًا هُوَ الْغَامِضُ الْبَعِيدُ مِنَ الْفَهْمِ، كَالْغَرِيبِ مِنَ النَّاسِ، إِثْمًا
هُوَ الْبَعِيدُ عَنِ الْوَطَنِ الْمُنْقَطِعُ عَنِ الْأَهْلِ [...] ثُمَّ إِنَّ الْغَرِيبَ مِنَ الْكَلَامِ يُقَالُ بِهِ عَلَى
وَجْهَيْنِ : أَحَدُهُمَا أَنْ يُرَادَ بِهِ بَعِيدُ الْمَعْنَى غَامِضُهُ، لَا يَتَنَاوَلُهُ الْفَهْمُ إِلَّا عَنِ بُعْدِ
وَمُعَانَاةِ فِكْرٍ، وَالْوَجْهُ الْآخَرُ أَنْ يُرَادَ بِهِ كَلَامٌ مَن بَعْدَتْ بِهِ الدَّارُ وَنَأَى بِهِ الْمَحَلُّ مِنْ
شَوَاطِدِ قَبَائِلِ الْعَرَبِ، فَإِذَا وَقَعَتْ إِلَيْنَا الْكَلِمَةُ مِنْ لُغَاتِهِمْ اسْتَعْرَبْنَاهَا

10. Ibid., p. 135.

11. *Ḥaṭṭābī, Ḡarīb*, V. I, pp. 70-71.

Le *ġarīb*, dans la langue, est ce qui est obscur et loin de la compréhension, comme le *ġarīb* pour les gens, c'est la personne qui est loin de son pays, séparée de ses gens [...]. En plus, le *ġarīb* dans la langue sont deux types : l'un des deux concerne celui dont la signification est loin et obscure, la compréhension ne le saisit qu'après une profonde et fatigante réflexion, et l'autre type concerne le parler des tribus arabes excentriques (*šawāḍḍ*), dont la maison est loin et la résidence est isolée, et quand on entend une parole de leurs parler, nous la trouvons étrange.

Dans *tuḥfat al=ʿarīb*, ʿabū ḥayyān précise que, dans le Qurʾān, il existe deux types de mots. Le premier type est connu par *al=ʿāmmah* 'les gens communs' et *al=ḥāṣṣah* 'l'élite'. Il donne les exemples de *samā* 'ciel', *ʿarḍ* 'terre', *fawq* 'haut' et *taḥt* 'bas'. Tandis que les mots du second type sont connus seulement de ceux qui ont une connaissance profonde de la langue. Il ajoute que les auteurs ont écrit des livres qu'ils ont appelé *ġarīb al=Qurʾān* et qui traitent le second type.¹²

Dans le chapitre consacré à la définition de *ġarīb*, as=suyūṭī ajoute les termes qu'il définit *mutaqāribah* 'ayant une signification voisine' à *ġarīb*: *nādir* (pl. *nawādir*) 'rare', *ḥūšiyy* 'parole obscure'¹³ ou *waḥṣī* 'inconnu; grossier; barbare', *šāḍḍ* 'excentrique; bizarre; extravagant; aberrant' et *šārid* 'erratique'. Il ajoute qu'ils sont tous des antonymes de *faṣṭḥ* (éloquent).¹⁴

Baalbaki précise que toutes ces paroles sont synonymes. Mais, il ajoute que les paroles étranges et rares qui se trouvent dans le Qurʾān et le Ḥadīṭ sont désignés seulement par le terme *ġarīb*, jamais par un autre terme comme *nādir* ou *šārid*, et certainement pas par *ḥūšiyy* ou *sāḍḍ*. En plus, il essaie d'identifier les critères qui ont permis aux érudits de classer, comme *ġarīb*, une parole ou une construction linguistique :¹⁵

1. Mots rarement utilisés, par exemple, *kurbuq* ou *kurbaq* 'boutique', ou moins fréquents qu'un autre mot de la même racine et largement utilisé, par exemple, *šarafah* par rapport à son synonyme plus commun *šaraf* 'honneur' ;
2. collections de synonymes, par exemple, *iʿranfaza*, *iḥraʿabba*, *iġraʿanna*, *ġasaʿa* et *taraza* 'raidir ou mourir de froid' ;

12. Abū Ḥayyān, *Tuḥfa*, p. 40 et Baalbaki, 2014, p. 70.

13. *ḥūšiyy* : *al=ġāmiḍ mina l=kalām* 'parole obscure' (*Firūzābādī, Qāmūs*, p. 591).

14. Suyūṭī, *Muzhir*, V. I, pp. 233-240

15. Baalbaki, 2014, pp. 37-84..

3. les mots qui se prononcent différemment dans divers dialectes, par exemple, *ǧilḏah*, *ǧulḏah* et *ǧalḏah* 'rugosité ; rudesse' ;
4. les formes anormales qui constituent une exception à une norme généralement applicable, par exemple, les deux formes *ru'yah* 'voir' et *ǧurdah* 'dénuder' sont les seuls noms verbaux qui expriment la manière (*maṣḏar naw'* ou *hay'ah*) qui ne sont pas de schème C₁ iC₂C₃ah ;
5. expressions idiomatiques qui ne contiennent pas de mots *ǧarīb*, par exemple, *iḥṭalaṭa l-laylu bi-l-turābi* 'la nuit s'est mêlée au sable', dont le sens que 'l'obscurité s'est intensifiée ou que l'affaire est confuse' ;
6. phénomènes syntaxiques rares, par exemple, l'utilisation du pronom relatif *mā* avant *lā an-nāhiya* (*lā* de la prohibition), comme dans l'expression, *hāḏā mā lā turidhu* 'c'est ce que je vous interdis de vouloir ou c'est ce que vous ne devriez pas vouloir'.

Toutefois, il ajoute que, dans certains cas, les entrées étiquetées comme *ǧarīb* ne donnent pas l'impression de l'être, comme *waḏīr'* 'humble', *wasīṭ'* 'moyen' et *raḏīr'* 'éminent'. Les critères ne sont pas toujours clairs, exception faite, dans le cas du Qur'ān, où l'inclusion des paroles parmi le *ǧarīb*, bien qu'elles ne semblent pas être obscures ou rares, a pour but de clarifier leur sens dans un contexte coranique.¹⁶

Comme il a été mentionné précédemment, l'intérêt des érudits s'est révélé depuis les premiers pas de l'activité linguistique pour comprendre *ǧarīb al=Qur'ān* et *ǧarīb al=Ḥadīṭ*. Les érudits se sont rendus compte très vite que, dans le texte coranique, d'abord, et puis dans le corpus du *Ḥadīṭ*, existent des variantes étrangères au dialecte de *Qurayš* et provenant des dialectes d'autres tribus arabes.¹⁷ Ceci les a amenés à faire la récolte d'un autre type de corpus *ǧarīb*, appelé aussi *an-nawādir* (pl. de *nādir*), où sont enregistrées les variantes dialectales et leurs origines.

2.2.1.1.1 *ǧarīb al=Qur'ān* La recherche et la récolte de *ǧarīb al=Qur'ān* ont commencé à une époque très précoce, dans la moitié du premier/septième siècle. Depuis, plusieurs œuvres ont été écrites et on peut les différencier selon l'ordre des entrées :¹⁸

- Les livres dont les entrées suivent l'ordre des *suwar* (pl. de *sūrah*) et des versets coraniques, comme *tafsīr ǧarīb al=Qur'ān* écrit par ibn qutaybah (m. 276/889). Ils

16. Ibid. pp. 38, 40, 71.

17. Naṣṣār, 1988, pp. 26-27.

18. Pour une liste plus complète et une étude plus approfondie voir Naṣṣār, 1988, pp. 33-42 et Baalbaki, 2014, p. 64-71

peuvent servir l'utilisateur qui connaît l'endroit exact où le mot se trouve dans le Qur'ān.

- Les livres qui suivent un ordre alphabétique. Dans son *kitāb* intitulé *ġarīb al=Qur'ān* ou *nuzhat al=qulūb*, muḥammad bnu 'uzayr as=siġistānī (m. 330/941) a introduit un arrangement alphabétique, mais seulement selon la première lettre du mot. En plus les mots sont énumérés sous la forme fléchis comme ils sont utilisés dans le Qur'ān. Dans *kitāb al=ġarībayn fī l=Qur'ān wa=l=Ḥadīṭ*, 'abū 'ubayd al=harawī (m. 401/1011) est le premier à utiliser un ordre alphabétique complet, sur la base du premier radical. Il a été suivi par al=rāġib al='iṣfahānī (m. 502/1108) dans son livre *al=mufradāt fī ġarīb al=Qur'ān*.

2.2.1.1.2 ġarīb al=Ḥadīṭ Il y a deux différences majeures entre *ġarīb al=Qur'ān* et *ġarīb al=Ḥadīṭ*. La première est le fait que l'activité sur *ġarīb al=Ḥadīṭ* est plus tardive, par exemple, le premier livre est attribué à 'ubaydah ma'mar bnu l=muṭannā (m. 209/824) ou à son contemporain 'abū 'adnān 'abdu r=raḥmān bnu 'abd al='a'lā as=sulamī.¹⁹

La deuxième différence et la plus importante concerne la nature des mots inclus dans chacun d'eux. Dans le *ġarīb al=Qur'ān*, comme il a été mentionné précédemment, on trouve également des mots couramment utilisés, dont on ne s'attendrait pas à ce qu'ils soient classés comme *ġarīb*. Le but de les inclure est expliquer leurs significations dans un contexte coranique précis. En revanche, le *ġarīb al=Ḥadīṭ* contient, presque exclusivement, des mots peu communs qui nécessitent une explication. Toutefois, on trouve, aussi, des mots communs utilisés métaphoriquement, comme c'est le cas de *raqabah* 'cou' qui se réfère à l'individu dans la phrase *ka='annamā 'a'taqa raqabatan* (litt. 'comme s'il a libéré un cou' qui signifie par synecdoque 'comme s'il a libéré un esclave').²⁰

Les œuvres traitant *ġarīb al=Ḥadīṭ* peuvent être également classés en fonction de l'ordre suivi pour énumérer les entrées.²¹

Le premier ordre suivi a été selon *al='isnād*, basé sur la longue chaîne de transmission d'autorité. Il a été introduit par 'abū 'ubayd al=qāsim bnu sallām (m. 224/838) dans *ġarīb al=Ḥadīṭ*. L'auteur explique brièvement les mots qu'il considère comme *ġarīb*, en commençant par le *Ḥadīṭ* prophétique, suivie par celui des *Ṣaḥābah* 'Compagnons du Prophète', et des *Tābi'ūn* 'Successesurs'.

19. Baalbaki, 2014, p. 72 et Naṣṣār, 1988, pp. 42-43, 52.

20. Baalbaki, 2014, pp. 81-82.

21. Pour une liste plus complète et une étude plus approfondie voir 'aṣ-Ṣāliḥ, 2009, pp. 42-52.

Auteur de *al=ġarībayn fī l=Qur'ān wa=l=Ḥadīṭ*, 'abū 'ubayd al=harawī (m. 401/1011) est le premier à utiliser l'ordre alphabétique classique qui sera appliqué, systématiquement, par les auteurs postérieurs.²²

Il est important, à l'occasion, de mentionner aussi *an=nihāyah fī ġarīb al=Ḥadīṭ wa=l='aṭar*, considérée l'œuvre la plus complète de son genre, elle se caractérise par une augmentation considérable du nombre de *Ḥadīṭ*. L'auteur, ibn al='aṭīr (m. 606/1210), explique que l'ampleur de *nihāyah* est dû au fait qu'à son époque, la connaissance de l'arabe s'est détériorée, en la comparant à ce qu'elle était environ quatre siècles plus tôt quand les premiers livres ont été écrits. De conséquence, il a trouvé nécessaire d'expliquer même les questions qui étaient considérées banales.²³

Il faut noter aussi que ibn manzūr (m.711/1311) affirme qu'il a utilisé et intégrée entièrement *an-nihāyah* pour écrire *lisān al='arab*.

2.2.1.1.3 Le ġarīb général Des fois, les auteurs enregistreraient tout ce qu'ils trouvaient de matériel *ġarīb*, dans les quatre sources essentielles, Qur'ān, Ḥadīṭ, poésie et prose. Ainsi, on trouve dans ces livres les mots rarement utilisés, les variantes dialectales, les formes et les expressions anormales, ainsi que les synonymes, les homonymes, les 'aḍḍād 'mots ayany deux sens contradictoires', les paroles caractérisées par *al='ibdāl* 'substitution de consonnes et semi-voyelles', etc.²⁴ Le premier livre du genre, intitulé *kitāb an=nawādir*, est attribué à l'un des premiers savants, 'abū 'amr bnu al='alā' (m. 154/770).²⁵ Un autre livre qu'on peut citer est *kitāb al=ġīm*, également, connu avec le titre *kitāb an-nawādir*. L'auteur, 'abū 'amr aš=šaybānī (m. 206/821), s'intéressait principalement au matériel dialectal et à la poésie des diverses tribus. Il a recueilli son matériel *ġarīb* directement des bédouins et des poètes, surtout de *raġaz*. Le plus important à noter est le fait que *kitāb al=ġīm* est le premier lexique ordonné selon l'ordre alphabétique (selon la première consonne seulement). En outre, il a été rédigé presque au même moment que la rédaction de *kitāb al='ayn*, par al=ḥalīl (m. 175/791), qui est ordonné selon l'ordre phonétique.²⁶

Les livres *nawādir* ont imprégné les lexiques postérieurs dans la microstructure des éléments lexicaux, par exemple, la présentation du verbe, en citant le *māḍī* 'accompli', suivi, de manière casuelle, par le *muḍāri* 'inaccompli', le *mašdar* et l'adjectif ; et

22. Baalbaki, 2014, p. 78 et Naṣṣār, 1988, p. 52.

23. Ibn al-Aṭīr, *Nihāya*, p. 5.

24. Naṣṣār, 1988, p. 88.

25. Pour une liste complète des *kutub an-nawādir* et une vision plus générale de leur évolution, voir Naṣṣār, 1988, pp. 109-118 et Baalbaki, 2014, pp. 84-99.

26. Baalbaki, 2014, p. 330 ; 332-337.

la présentation du nom au singulier suivi, également de manière casuelle, par le/les pluriels correspondants. Ainsi, ils nous fournissent une idée sur les premières étapes de l'évolution lexicale.²⁷

2.2.1.2 Les emprunts et les paroles arabisées

Les emprunts des langues étrangères sont appelés en deux modes, *mu'arrab* 'arabisé' et *daḥīl* 'intrus'; as=suyūṭī (m. 911/1505) reporte la définition de 'abū ḥayyān al=ġarnāṭī (m. 745/1344), qui distingue trois types d'emprunts :

1. une catégorie que les arabes utilisent en modifiant phonétiquement la parole et en changeant sa morphologie, selon un schème attesté dans leur langue. Ce dernier phénomène est appelé '*ilḥāq* 'affiliation'. La parole est dite *mulḥaqah* 'affiliée', et elle est considérée *mu'arrabah* 'arabisée' et entre à faire partie de la langue arabe ;
2. une catégorie dont les paroles sont changées phonétiquement mais elles ne sont pas affiliées à un schème précis ;
3. la dernière catégorie comprend les paroles utilisées sans modification, ni phonétique ni morphologique. Dans ce cas, elles sont considérées *daḥīl* 'intrus', dans le sens qu'elles restent étrangères à la langue arabe.²⁸

Dans son traité, intitulé *risālah fī taḥqīq ta'rīb al=kalimah al=a'ġamiyyah*, ibn kamāl bāšā (m. 940/1534) identifie deux types de changements que peut subir une parole étrangère, utilisée par les arabes, pour qu'elle soit conforme à un schème attestée : i) changement phonétique et ii) changement morphologique (*'ilḥāq*) . Selon lui, la combinaison de ces deux facteurs permet d'identifier quatre types de paroles étrangères. En plus des trois catégories définies ci-dessus, il a ajouté la catégorie des paroles qui ont subi une modification morphologique sans changement phonétique.²⁹ Cependant, cette subdivision reste théorique, car on note que les érudits utilisent les deux termes *mu'arrab* 'arabisé' et *daḥīl* 'intrus', de manière indifférente. Par exemple, as=suyūṭī (m. 911/1505) reporte la définition de al=ġarnāṭī qui distingue trois types différents. Mais deux lignes avant, il reporte la définition de al=ġawālīqī (m. 540/1145), selon laquelle, les deux termes sont synonymes (*wa=yuṭlaqu 'alā l=mu'arrabi daḥīlun*).³⁰

27. Naṣṣār, 1988, p. 118.

28. Suyūṭī, Muzhir, V. I, pp. 260-270; Naṣṣār, 1988, pp. 58-59

29. Ibn Kamāl Bāšā, *Risāla*, pp. 47-48.

30. Suyūṭī, Muzhir, V. I, p. 269

Depuis les premières étapes des sciences linguistiques, grammairiens et lexicographes ont essayé d'identifier les caractéristiques phonétiques et morphologiques, pour reconnaître ce qui peut être intrus et étranger, à la langue arabe. Il suffit de citer al=ḥalīl (m. 175/791) qui a défini les principes phonologiques permettant de reconnaître si une parole est arabe ou non.

Sur la base de ses travaux et d'autres, as=suyūṭī (m. 911/1505) identifie plusieurs critères pour identifier les mots non-arabes :³¹

- les paroles qui ne respectent pas les schèmes arabes, par exemple, *ḥurāsān*, dont le schème C₁uC₂āC₃ān n'existe pas en arabe ;
- les paroles qui commencent par la consonne *n* suivi de *r*, par exemple, *narġis* 'parfum' ;
- les paroles qui se terminent par la consonne *z* précédée par *d*, par exemple, *handasah* 'géométrie' qui provient de la parole persane 'āb 'andaz, mais elle a subi un changement parce que les arabes ne prononcent pas la *d* suivi de *z*;³²
- les paroles contenant les deux consonnes *ṣ* et *ğ*, comme le terme *ğaşṣ* 'gypse', ou les deux consonnes *ğ* et *q*, comme le terme *manğanīq* 'catapulte' qui dérive du persan *man jah nīk*;³³
- les paroles formées de quatre ou cinq consonnes et ne contiennent pas une des consonnes fluides (*ḥurūf al=dulq*) qui sont {b, r, f, l, m, n}.

Cette étude, aussi précise, est certainement liée à la question bien épineuse, à savoir si le Qur'ān contient des mots *mu'arrabah* 'emprunts étrangers'. Sur cette question, les érudits se sont divisés. Plusieurs d'entre eux, comme 'abū 'ubaydah (m. 209/824), reconnaissaient que la révélation du Qur'ān a été faite exclusivement en langue arabe et déniaient la présence de paroles étrangères.³⁴

D'autres, comme ibn 'abbās (m. 68/687), ont reconnu l'existence, dans le Qur'ān, des emprunts de langue étrangère. En effet, dans son livre *al=lugāt fī l=Qur'ān*, ibn 'abbās a énuméré vingt paroles qui se trouvent aussi dans d'autres langues, par exemple, *al=yamm* 'la mer' dérivant de *Suryāniyyah* 'syriaque' ; *miškāt* 'niche' de *ḥabašīyyah* 'éthiopien' ; *al=firdaws* 'le paradis' et *aṣ-ṣirāṭ* 'chemin ; voie' du *yūnāniyyah* 'grec'.³⁵

31. Suyūṭī, Muzhir, p. 270.

32. Fīrūzābādī, Qāmūs, p. 582.

33. Fīrūzābādī, Qāmūs, p. 872.

34. Baalbaki, 2014, p. 163.

35. Suyūṭī, Muzhir, p. 268.

Un troisième groupe, comme 'abū 'ubayd (m. 224/839), reconnaît que ces paroles sont d'origine étrangères mais, après être arabisées, elles sont entrées faire partie de la langue arabe.³⁶

Cependant, l'intérêt pour l'étude des paroles étrangères dans le Qur'ān s'est atténué, avec le temps. Après l'œuvre de ibn 'abbās (m. 68/687), il faut attendre as=suyūfī (m. 911/1505) qui écrit deux monographies à ce propos.³⁷

Par contre, *al=mu'arrab min al=kalām al='a'ġamī 'alā ħurūf al=mu'ġam*, écrit par al=ġawālīqī (m. 540/1145), est le premier lexique dédié entièrement aux emprunts étrangers existants dans la langue arabe, généralement et non strictement liés au Qur'ān. Il est considéré le plus important du genre, contenant 743 mots répartis sur vingt-six chapitres et ordonnés seulement selon la première lettre.³⁸

Certains lexiques pluri-thématique ont dédié quelques chapitres aux paroles étrangères. Le plus significatif est celui de aṭ=ṭa'ālibī (m. 429/1039), qui a consacré un chapitre, *bāb fiqh al=luġah wa='asrār al='arabiyyah*, à la *muwāzanah* 'comparaison' entre l'Arabe et le Persan. Le plus frappant de cette étude est le fait que aṭ=ṭa'ālibī a divisé le chapitre en sous-chapitres qui traitent :³⁹

- les paroles existantes seulement en Arabe et ne se trouvent pas dans le Persan ; on note qu'elles appartiennent exclusivement au domaine religieux et sont citées dans le Qur'ān ou dans le Ḥadīṭ ;
- les paroles existantes, dans la même forme, dans la langue arabe et dans la langue persane ;
- les paroles qui se trouvent seulement dans le Persan, ce qui a obligé les arabes à les utiliser avec ou sans modification. L'auteur a subdivisé ce dernier sous-chapitre, selon le domaine sémantique, tel que les ustensiles de cuisine, le pain, les plats de nourriture, les gâteaux, les boissons ; les épices, les habits, les bijoux, les fleurs et les parfums.

2.2.1.3 Néologisme et solécisme

Le néologisme, *al=muwallad*, a été traité en deux modes différents selon que la nouvelle parole est '*āmmiyyah* 'commune' et a une signification commune ou elle est '*islāmiyyah* 'islamique' et donc a une signification religieuse.

36. Ġawālīqī, *Mu'arrab*, p. 53 et Baalbaki, 2014, pp. 163-164.

37. Baalbaki, 2014, p. 164.

38. Pour une étude plus approfondie voir Naṣṣār, 1988, pp. 69-78 et Baalbaki, 2014, pp. 161-169.

39. Ṭa'ālibī, *Fiqh*, pp. 314-316.

Avec l'avenue de l'Islam, on assiste à plusieurs phénomènes linguistiques concernant la sphère religieuse. On observe des glissements sémantiques, par exemple, la parole *muḥaḍram* 'mélangé' a un nouveau sens et désigne les personnes qui ont vécu dans la période chevauchant entre la *ḡāhiliyyah* 'période avant l'Islam' et la naissance de l'Islam. On observe, aussi, la création de nouvelles paroles dérivant de celles existantes, par exemple, la parole *munāfiq* 'hypocrite ; sournois' dérive du verbe *nāfaqa*, spécifique de la gerboise qui entre dans *an-nāfiqā* 'terrier de la gerboise'. L'étude de ce type de néologisme concernait *al=fiqh* 'les études religieuses'. Par contre, dans le domaine linguistique et/ou lexical, on peut citer *kitāb al=zīnah fī al=kalimāt al=islāmiyyah al=arabiyyah* écrit par 'abū ḥātim ar=rāzī (m. 322/934). Il discute, dans un chapitre, les significations des mots et les attributs islamiques de Dieu comme *aṣ=ṣamad* 'l'Éternel', *al=ḥakīm* 'le Sage', etc. Dans un autre chapitre, il traite les notions islamiques spécifiques, telles que *ḥalq* 'création', *qadar* 'destin', 'arš' 'trône', et *qiyāmah* 'résurrection'.

Dans la plupart de ces lexiques, on note que les érudits se basent sur les ouvrages de *fiqh* et qu'ils abordent surtout l'aspect religieux, sans s'occuper de celui linguistique, par exemple, sans mentionner le sens commun quand il existe.⁴⁰

Dans son œuvre, *al=muzhir fī 'ulūmi al=luḡah wa=anwā'i=hā*; *as=suyūṭī* a dédié un chapitre, *ma'rifat al=alfāz al=islāmiyyah*, qui traite les nouvelles paroles islamiques.⁴¹ Par contre, il traite les phénomènes linguistiques concernant la vie commune dans un autre chapitre, intitulé *ma'rifat al=muwallad*, et où il identifie deux types de paroles qu'il appelle *muwalladah* 'hybride ; qui n'est de race pure' :

1. La naissance de nouvelles paroles, par exemple, *qaḥṭiyy* 'une personne qui mange sans rien laisser' qui dérive de *qaḥṭ* 'disette ; manque de vivres' ; *tabaḡdada* 'une personne d'origine de *baḡdād*' (dans le sens réel ou fictif) qui dérive de *baḡdād* 'Bagdad' ; *buhṛān*, utilisé par les médecins, pour désigner 'la crise de délire durant une maladie fébrile' et dérive de *bāḥūrā* 'hautes températures en Août'. L'auteur donne plusieurs exemples, en disant, à plusieurs reprises, que la parole est d'usage commun et n'est pas arabe ('*āmmiyyah laysat 'arabiyyah*), ou qu'elle ne fait pas partie de la langue des bédouins (*laysat min kalām 'ahl al=bādiyyah*) ou encore qu'elle fait partie du langage des villes et non de la campagne (*min kalām al=ḥāḍirah dūna al=bādiyyah*).⁴²
2. L'utilisation d'une parole existante avec une nouvelle signification, par exemple,

40. Naṣṣār, 1988, pp. 54-57.

41. Suyūṭī, Muzhir, V. I, pp. 294-303.

42. Suyūṭī, Muzhir, V. I, pp. 304-310.

tanazzuh dont le sens original est 'l'éloignement de toute calamité' et vient utilisé avec le nouveau sens 'promenade dans les jardins'.⁴³

Dans le même chapitre, as=suyūṭī note aussi l'utilisation des paroles mais avec :⁴⁴

- changement de la vocalisation, par exemple, *šam* ' au lieu de *šama* ' 'cire' ; *suffl* au lieu de *siff* 'dessous ; bas' et aussi 'ulw au lieu de 'ilw 'dessus ; haut' ;
- changement de consonnes, par exemple, *na'aq* au lieu de *naġaq* 'cri du corbeau'.
- variation phonétique, comme la réalisation de la *hamza* dans la nouvelle parole 'a'zab au lieu de 'azab 'célibataire' ou son changement en wāw ou yā' dans wākala au lieu de 'ākala 'manger avec quelqu'un' ;

Toutefois, as=suyūṭī n'utilise pas, dans ces cas, le terme *muwallad*, mais considère les paroles comme *lahn* 'solécisme' et les introduit avec *taqūlu al= 'āmmah* 'la populace dit'. En effet, *lahn al= 'āmmah* 'solécisme de la populace' ne faisait pas, non plus, partie de la langue arabe et apparaît très peu dans le matériel lexical.⁴⁵ Toute déviation, des normes puristes, était considérée incorrecte et inacceptable par les grammairiens et les philologues. Afin de standardiser la langue arabe, ils se sont donnés l'objectif d'enregistrer ces déviations, les dénoncer et les corriger en faisant référence au *faṣīḥ* 'éloquent'. Plusieurs savants, du deuxième/huitième et troisième/neuvième siècle, ont contribué avec des monographies traitant *lahn al= 'āmmah*, dont le contenu se concentrait principalement sur les changements vocaliques et consonantiques et sur les modèles verbaux et nominaux. Dans plusieurs de ces ouvrages, l'auteur précise qu'il dénonce aussi bien le *lahn* des 'āmmah 'roturier, gens du peuple' mais aussi des *ḥāṣṣah* 'l'élite'.⁴⁶

Un des ouvrages sur le *lahn*, qui a eu un impact sur la philologie et la lexicographie, est 'iṣlāḥ al=manṭiq de ibn as=sikkīt (m. 244/858). Dix chapitres, de 'iṣlāḥ, traitent les trois types que la 'āmmah utilise de manière incorrecte, à savoir, les erreurs dans les voyelles, les changements consonantiques et les changements de sens. Il convient de noter que ibn as=sikkīt appelle *faṣīḥ* ce qu'il considère le représentant de la norme mais il inclut, aussi, le matériel dialectal, comme variantes moins acceptables. Un autre livre, aussi important, est *al=faṣīḥ* écrit par aṭ=ṭa'lab (m. 291/904). L'auteur affirme, dans son introduction, qu'il dénonçait le *lahn* qui se produit dans la langue parlée et dans les textes écrits. Trente chapitres traitent, presque exclusivement, les schèmes

43. firūzābādī, par exemple, le considère une erreur grave (*Firūzābādī, Qāmūs*, p. 1254).

44. *Suyūṭī, Muzhir*, V. I, pp. 311-321.

45. *Baalbaki, 2014*, p. 162.

46. *Zubaydī, Lahn*, pp. 62-63

morphologiques des verbes et des noms et les différences de sens entre les schèmes considérés, par exemple, la différence entre *naqiha* 'comprendre' et *naqaha* 'entrer en convalescence'; *humūlah* 'fardeaux' et *hamūlah* 'chameaux, chevaux, ou autres, qui portent des charges'.⁴⁷

À partir du quatrième/dixième siècle, on note une attitude différente vers l'erreur linguistique. Par exemple, ibn hišām al=laḥmī (m. 577/1181) défend la langue parlée de la 'āmmah 'populace'. Il affirme qu'elle ne doit pas être accusée de *lahn*, même si elle n'est pas *faṣīḥah* 'éloquente', tant que son usage est conforme à un dialecte attesté (*luḡah masmū'ah*). On note aussi des études sur les différences régionales et des notes sur l'évolution diachronique, 'islāḥ *lahn al='āmmah bi=l='andalus*, écrit par az=zubaydī (m. 379/989) et, *taṭqīf al=lisān wa=taḥqīq al=ḡanān*, de ibn makkī al=ṣiqillī (m. 501/1107). Les deux auteurs font noter que les erreurs commises par les gens de l'Andalousie ('*ahl al=maḡrib*) sont différentes, en les comparant à celles commises par les gens de l'Orient ('*ahl al=mašriq*), ou à celles mentionnées dans les livres précédents (*kutub al=mutaqaddimīn*).⁴⁸

2.2.1.4 Synonymie, homonymie et énantiosémie

Le grammairien, sībawayhi (m. 180/796), a défini trois types de relations sémantiques entre les paroles. La relation la plus courante se trouve entre des paroles différentes dans la forme et dans la signification ('*ihṭilāf al=lafẓayn li=ihṭilāf al=ma'nayayn*). Une deuxième relation, appelée *tarāduf* 'synonymie', lie les paroles *mutarādifah* 'synonymes' qui ont la même signification mais sont différentes dans la forme ('*ihṭilāf al=lafẓayn wa=l=ma'nā wāḥid*). La dernière relation est '*ištirāk* 'polysémie' qui lie les paroles, *muštarakah* 'polysémiques, homonymes', ayant la même forme mais avec des sens différents ('*ittifāq al=lafẓayn wa=ihṭilāf al=ma'nayayn*).

Deux autres grammairiens, quṭrub (m. 206/821) et al=mubarrad (m. 285/898), font noter l'existence de paroles qui concordent en forme et donc elles sont homonymes mais, elles expriment deux significations contradictoires. En arabe, ces paroles sont appelées al='*aḡdād* 'les contraires'. Par exemple, la parole *ḡawn* exprime aussi bien le "noir" et le "blanc".⁴⁹ C'est le cas de l'énantiosémie, que Cohen désigne comme « homonymie antonymique ».⁵⁰

47. Baalbaki, 2014, p. 174-175.

48. Ibid., pp. 180-182.

49. Baalbaki, 2014, p. 188-190; Quṭrub, 'Aḡdād, p. 70.

50. Cohen, 1961.

Certains érudits appelés *šū ūbiyyūn*, connus pour exalter la culture persane sur la culture arabe, profitaient du fait que la polysémie, la synonymie et l'énantiosémie sont des cas particuliers de l'ambiguïté et affirmaient que l'arabe conduit à l'obscurité et à la confusion.⁵¹

D'autres niaient l'existence des paroles *muštarakah* et *mutarādifah*, en confirmant qu'il y a toujours une variance dans le sens des deux synonymes. Par exemple, les paroles *sayf*, *muhannad* et *husām* sont considérées synonymes ; mais le terme *sayf* 'épée' exprime l'objet, par contre, les termes *muhannad* et *husām*, sont des adjectifs qui désignent des qualités différentes de l'épée. Le terme *muhannad* désigne 'l'épée fabriquée en Inde ou forgée dans le métal indien' et *husām* désigne 'l'épée bien affilée et tranchante'.

Un autre exemple intéressant est celui des verbes *qa'ada* et *ğalasa*, considérés synonymes avec le sens 's'asseoir'. Cependant, *qa'ada* a le sens de 's'asseoir' dans un mouvement du haut vers le bas, comme dans l'exemple, *qāma tumma qa'ada* 'il s'est levé puis il s'est assis'. Par contre, *ğalasa* a le sens de 's'asseoir' dans un mouvement du bas vers le haut, comme dans l'exemple, *kāna muḏṭağī'an tumma ġalasa* 'il était allongé puis il s'est assis'.⁵²

2.2.1.4.1 La synonymie Le phénomène de la synonymie est très riche en langue arabe. Selon *quṭrub* (m. 206/821), elle prouve la grande expressivité des arabes qui parlent une langue remarquablement poétique et, ne se limitant pas à une seule parole, ont besoin d'extension.⁵³ Toutefois, dans son introduction à l'édition de *al=furūq al=luğawiyah*, écrit par *al=askarī* (m. 382/993), l'éditeur *muḥammad ibrīm salīm* distingue diverses origines.⁵⁴

2.2.1.4.1.1 at=taṣḥīf 'erreurs'. Dans la plupart des cas, la différence entre une lettre et une autre est très subtile et peut être signalée par l'absence ou la présence d'un point. En absence des points, la lettre est dite *muhmalah* 'non diacritisée ; sans points',⁵⁵ c'est le cas de la lettre ح 'h'. En présence des points, la lettre est dite *mu'ğamah* 'diacritisée',⁵⁶ c'est le cas des lettres ج 'ğ' et ح 'h'. La position et le nombre des

51. Baalbaki, 2014, p. 190.

52. *Suyūṭī*, *Muzhir*, V. I ; p. 404.

53. *Quṭrub*, 'Aḏḏād, p. 69.

54. 'Askarī, *Furūq*, pp. 17-19.

55. *muhmalah* dérive du verbe 'ahmala 'laisser sans soin ; laisser sans assistance'.

56. *mu'ğamah* dérive du verbe 'a'ğama 'enlever l'ambiguïté' (*Ibn Ğinnī*, *sirr*, V. I ; p. 37).

points consonantiques sont, également, très importantes. C'est le cas, par exemple, des lettres ب, ت, et ث respectivement *b*, *t* et *ṭ* et ج, ح et خ respectivement *ǧ*, *ḥ* et *ḫ*.

Le tableau 2.1 illustre les lettres, de l'alphabet arabe, en fonction qu'elles soient *muhmalah* ou *mu'ğamah*.

lettres <i>muhmalah</i> vs lettres <i>mu'ğamah</i>		
<i>ḥā'</i>	<i>ḥ</i>	ح
<i>ḫā'</i>	<i>ḫ</i>	خ
<i>ǧīm</i>	<i>ǧ</i>	ج
<i>dāl</i>	<i>d</i>	د
<i>ḍāl</i>	<i>ḍ</i>	ذ
<i>rā'</i>	<i>r</i>	ر
<i>zāy</i>	<i>z</i>	ز
<i>sīn</i>	<i>s</i>	س
<i>šīn</i>	<i>š</i>	ش
<i>ṣād</i>	<i>ṣ</i>	ص
<i>ḍād</i>	<i>ḍ</i>	ض
<i>ṭā'</i>	<i>ṭ</i>	ط
<i>ẓā'</i>	<i>ẓ</i>	ظ
<i>'ayn</i>	'	ع
<i>ğayn</i>	<i>ğ</i>	غ

lettres uniquement <i>muhmalah</i>		
<i>hamzah</i>	'	ء
<i>kāf</i>	<i>k</i>	ك
<i>lām</i>	<i>l</i>	ل
<i>mīm</i>	<i>m</i>	م
<i>wāw</i>	<i>w</i>	و

lettres uniquement <i>mu'ğamah</i>		
<i>bā'</i>	<i>b</i>	ب
<i>tā'</i>	<i>t</i>	ت
<i>ṭā'</i>	<i>ṭ</i>	ث
<i>nūn</i>	<i>n</i>	ن
<i>yā'</i>	<i>y</i>	ي
<i>fā'</i>	<i>f</i>	ف
<i>qāf</i>	<i>q</i>	ق

Tab. 2.1 : Nom et symbole des consonnes *muhmalah* et des consonnes *mu'ğamah*

Une faute de reproduction du texte, de lecture ou d'interprétation peut conduire à des erreurs qui, avec le temps, sont supposées entrées à faire partie du lexique. Ainsi

on trouve, par exemple, les deux paroles نَاصَّ *nāṣa* et نَاَضَّ *nāḍa* avec le même sens 'se mouvoir'.

2.2.1.4.1.2 Les procédés de la récolte des données Les paroles peuvent être différentes d'un dialecte à un autre, mais les philologues ne mentionnaient pas l'origine de la parole. En époques successives, les lexicographes ont regroupé les paroles comme synonymes et non comme variantes dialectales.

Les premiers philologues ne s'intéressaient pas ou ne mentionnaient pas la fréquence de l'usage. Ainsi, pour enregistrer une parole, il suffisait qu'elle soit utilisée, dans une des quatre sources attestées sans se préoccuper de sa fréquence.

Les phénomènes de *tawlīd* 'néologisme' et *ta'rib* 'emprunt de langue étrangère' ont permis la naissance de nouvelles paroles coexistantes avec une parole arabe désignant le même sens. Dans ce cas, les philologues annotaient la parole, sans mentionner si elle est 'arabiyyah' 'arabe'; *muwalladah* 'nouvelle' ou *mu'arrabah* 'arabisée d'origine étrangère'.

Les philologues ne faisaient pas la différence entre l'utilisation réelle et celle figurative. Ainsi l'adjectif *husām* 'affilée' est utilisé, par extension, pour désigner *sayf* 'épée'.

Dans un lexique *mubawwab*, qui traite les paroles *mutarādifah* 'synonymes', on trouve enregistrés tous les synonymes pour chaque mot, au contraire des lexiques *muğannas* qui ne citent habituellement qu'un seul ou expliquent le mot sans mentionner aucun synonyme.⁵⁷

Le premier lexique *mubawwab* existant, traitant le *mutarādif*, est *mā ḥtalafat 'alfāzu=huwa=ttafaqat ma'ānī=hi* de 'aṣma'ī (m. 216/831) et dont le matériel n'est ni thématique ni d'ordre alphabétique. Les lexiques, qui l'ont suivi, ont connu un développement majeur aussi bien en quantité des synonymes traités qu'en qualité et en disposition du matériel.⁵⁸ Par exemple, ibn as-sikkīt (m. 244/858) a disposé son livre, *al=alfāz*, en 148 chapitres, dont chacun contient les mots et les expressions appartenant à un champ sémantique spécifique, tels que la richesse et la fertilité, la pauvreté et la stérilité, avarice, maladie, vin, couleur, etc.

Le livre, *al=alfāz al=kitābiyyah* de al=hamadānī (m. 320/932), est intéressant dans la mesure où l'auteur ne s'est pas limité seulement aux mots synonymes, mais il a inclus les expressions idiomatiques qui peuvent être utilisées alternativement pour exprimer un certain sens. Dans la même direction, on trouve, aussi, *'alfāz al=kitābah wa-l-ta'bīr* où, en plus des lexèmes simples, l'auteur, ibn al-marzubān al=bāḥiṭ (m. 330/941), a

57. Baalbaki, 2014, p. 205.

58. Pour plus de détail voir Baalbaki, 2014, pp. 204-209.

inclus les styles d'expressions. Chaque expression est suivie de phrases qui peuvent la remplacer. Le livre est divisé en vingt-huit thèmes, entre autres, les félicitations, intercessions, condoléances, remerciements, excuses, etc. chaque thème comprend un texte modèle qui se compose d'un certain nombre d'expressions.

Lors du quatrième/dixième siècle, période au cours de laquelle la validité de la synonymie a été très débattue, on trouve un autre type de livres intitulés *al=furūq* 'les différences'. L'objectif de ces œuvres est de démontrer que deux mots ne peuvent avoir la même signification mais, expriment deux concepts différents. On peut citer *al=furūq wa=man' al=tarāduf* de *al=ḥakīm al=tirmiḍī* (m. vers 320/932) et *al=furūq al=luġawiyyah* de 'abū hilāl al='askarī (m. après 395/1005). Dans les deux œuvres, le matériel se présente sous forme de paires (ou plus) de paroles prétendues synonymes. L'auteur explique et clarifie les différences et les nuances qui les séparent. Par exemple, la différence entre la parole *muḥāğġah* 'argumentation' qui signifie *établir la vérité et la défendre grâce à une ḥuğġah* 'preuve' et la parole *muğādalah* 'dispute' qui se réfère à *la controverse et la discussion mais dans l'intérêt de vaincre son adversaire*.⁵⁹

2.2.1.4.2 L'homonymie Selon divers ressources, *al='aṣma'ī* (m. 216/831) est le premier à utiliser le terme 'aġnās, dans son livre *kitāb al='aġnās*, pour désigner les mots homonymes polysémiques.⁶⁰

Les livres peuvent être divisés en deux types, ceux qui traitent la polysémie comme thème général et ceux qui en examine un aspect spécifique.⁶¹

Du premier type, on peut citer *mā ttafaqa lafzu=hu wa=ḥtalafa ma'nā=hu* de 'abū l='amaytal (m. 240/854). Il contient 273 paroles *muštarakah*, énumérées sans ordre spécifique et dont les différentes significations sont brièvement expliquées. Un autre lexique est *al=munağġad fī l=luġah*, de kurā' an=naml (m. 310/922), qui contient un total de 884 paroles classées en six chapitres thématiques : les parties du corps humain, les animaux, les oiseaux, les armes, le ciel et la terre. Chaque mot est cité sous son sens le plus commun, accompagné des autres significations. Du même genre, on peut citer également, *mā ttafaqa lafzu=hu wa=ḥtalafa ma'nā=hu*, de ibn aš-šaġarī (m. 542/1148), avec 1670 mots *muštarak* classés par ordre alphabétique, selon la première lettre du mot et indépendamment de sa racine.

Des lexiques spécialisés dans un genre sémantique précis, on peut citer, *mā ttafaqa lafzu=hu wa=ḥtalafa ma'nā=hu mina l=Qur'ān al=mağīd* de *al=mubarrad* (m. 285/898).

59. Ibid., pp. 209-211.

60. Ibid., pp. 200-201.

61. Pour plus de détails voir Baalbaki, 2014, pp. 200-204

Il traite spécifiquement les mots coraniques qui, dans le contexte coranique, ont un sens différent de celui plus commun.

2.2.1.4.3 L'énantiosémie Aux critiques des *šū ūbiyyūn* qui affirment que la langue arabe est sujette à la confusion et l'ambigüité, un des plus intéressantes réfutations est celle de 'abū bakr bnu al-'anbārī (m. 328/940). Dans son œuvre, *kitāb al-'aḍḍād*, il explique que l'ambigüité, due aux 'aḍḍād, est la même que celle posée par tous les mots ayant des sens différents, même s'ils ne sont pas opposés (*mutaḍḍān*). Il ajoute que le sens d'une parole est définie, seulement, en fonction de ce qui la précède et de ce qui la suit ('*illā bi=mā yataḡaddamu l=ḡarfa wa=yata 'aḡaru ba 'da=hu*).⁶² En plus, il donne plusieurs explications de la formation des 'aḍḍād, en langue arabe, par exemple :

- si un mot a deux significations contradictoires, l'une d'entre elles est le 'aṣl 'l'original', tandis que l'autre est le résultat du 'ittisā' 'extension';⁶³
- les 'aḍḍād ne peuvent pas se vérifier dans un même moment précis et dans un même dialecte. Mais chaque sens a été utilisé par une tribu différente (*ḡayy*), avant que l'interaction ne provoque l'occurrence des deux sens dans le même dialecte.⁶⁴

Le premier livre traitant les 'aḍḍād est attribué à *ḡuṭrub* (m. 206/821) et s'intitule, justement, *al-'aḍḍād*, où l'auteur énumère 218 mots pour lesquels il apporte des *šawāhid* principalement du Qur'ān et de la poésie. L'œuvre, citée plus haut, *kitāb al-'aḍḍād* de al-'anbārī (m. 328/940), présente une réfutation en faveur de l'existence des 'aḍḍād et comprend 357 mots.⁶⁵ Les deux œuvres, entre autres, ont leurs articles classés de manière aléatoire ce qui rend très difficile leur consultation.

Le classement alphabétique, bien que partiel selon la première consonne radicale, a été introduit par 'abū ḡayyib al-luḡawī (m. 351/962), dans *al-'aḍḍād fī kalām al-'arab*.

Une dernière note sur cet argument, est le fait que plusieurs auteurs citent un grand nombre de paroles comme 'aḍḍād, alors qu'elles ne le sont pas effectivement. Désignées par Cohen, les «faux 'aḍḍād», elle peuvent être rejetées pour divers motifs, entre autres :⁶⁶

62. Ibn al-Anbārī, 'Aḍḍād, pp. 3-4

63. Ibid., p. 8.

64. Ibid., p. 11.

65. Pour plus de détails voir Baalbaki, 2014, p. 194

66. Cohen, 1961, pp. 7-9.

- le manque de rigueur de la part des lexicographes et la confusion entre schèmes différents. Par exemple, *ḥaḍama* cité pour désigner 'trancher' et 'être tranché'. Il s'agit en fait, à l'intérieur d'une même racine, d'une opposition morphologiquement marquée entre *ḥaḍima*, à valeur interne, et *ḥaḍama*, à valeur externe ;
- l'altération des points diacritiques (*taṣḥīf*), comme le fait de considérer la parole *ḡābir* comme *ḡidd* avec les sens opposés de «qui est passé » et «qui reste » alors que ce dernier sens est lié, selon Cohen, à la parole 'ābir. Il faut noter que la *ḡayn* (غ) et la 'ayn (ع) diffèrent d'un point ;
- le type de préposition utilisée avec les verbes, par exemple, le verbe *raḡiba* dont le sens change en fonction de la préposition qui la suit. Dans la construction, *رَغِبَ إِلَى raḡiba 'ilā*, le verbe signifie 'désirer ardemment'. Par contre, dans la construction, *رَغِبَ عَنْ raḡiba 'an*, il signifie 'éprouver de l'aversion pour' ;
- des raisons stylistiques telles que l'ironie ou l'euphémisme. Dans son œuvre, ibn al-'anbarī a introduit certaines paroles avec la phrase, 'ce qui ressemble aux 'aḍḍād dans l'ironie' *مِمَّا يُشْبِهُ الْأَضْدَادَ قَوْلُهُمْ فِي الْإِسْتِهْرَاءِ*⁶⁷. Ce qui prouve qu'il était conscient que les mots, pris en considération, sont définis à la fois dans un sens propre et dans un sens métaphorique.

Il nous est impossible de juger la validité des 'aḍḍād dans les lexiques qui nous sont parvenus, et on ne peut pas déterminer l'extension ni temporelle ni spatiale du corpus qu'on a à disposition. D'un côté, selon les procédés qu'ils ont adoptés pour la récolte des données, les lexicographes ne prenaient pas en considération l'évolution de la langue. Du fait, ils font attester les significations contemporaines (au moment de la récolte) dans la poésie ancienne. De l'autre, bien qu'ils reconnaissaient l'existence des divers dialectes, ils ne mentionnaient pas la source dialectale, ni les différents niveaux de l'usage (parlé, écrit, populaire...)⁶⁸.

67. Ibn al-Anbārī, 'Aḍḍād, pp. 257-258.

68. Cohen, 1961.

2.2.1.5 La métathèse

Les phénomènes *'ibdāl* 'échange ; substitution' et *qalb* 'inversion ; changement' sont deux autres aspects qui ont attiré l'intérêt des philologues.

Selon as=suyūṭī, le *qalb* est le phénomène de métathèse, comme dans le cas de, *ḡabaḍa* et *ḡaḍaba* 'fasciner ; charmer ; séduire'.⁶⁹

Selon les ressources, *al='ibdāl* n'est pas une substitution intentionnelle, mais résulte des différences dialectales qui affectent une lettre d'un certain mot, en donnant naissance à des mots prononcés différemment mais, ils ont la même signification.⁷⁰

Dans l'introduction de son œuvre, *al='ibdāl wa=l=mu'āqabah wa=n=naḗā'ir*, l'auteur, az=zaḡḡāḡī (m. 337/949), affirme que *al='ibdāl* est synonyme de *mu'āqabah* et *naḗā'ir*. Dans les notes 2 et 3, l'éditeur, 'izz ad=dīn at=tanūḥī, ajoute deux autres synonymes de *'ibdāl*, *ta'āqub* et *i'tiqāb*, et spécifie que les couplets des lettres (de deux ou plus) qui subissent le phénomène de *'ibdāl* sont appelés *al='abdāl*. Par contre, le terme *naḗā'ir* 'paire ; semblables' désigne le couplet des paroles qui naissent du phénomène de la substitution.

La substitution se fait entre deux consonnes similaires dont les points d'articulation sont très proches. Dans une même parole, on peut noter la substitution d'une seule lettre, par exemple, *qaḍama* et *qaṭama* 'croquer ; grignoter' ; de deux lettres, par exemple, *saḥaqa* et *sahaka* 'broyer' ; ou de trois lettres, par exemple, *dara'a* et *ṭala'a* 'émerger ; se manifester'.⁷¹

On peut citer aussi, *kitāb al='ibdāl*, de 'abū ṭ=ṭayyib al=luḡawī (m. 351/962). Considéré une œuvre importante dans son genre, parce que, d'une part, l'auteur a classé les chapitres dans un ordre alphabétique strict. De l'autre, il a essayé d'énumérer une liste très complète de matériel *'ibdāl*, à tel point qu'il est décrit dans la tradition arabe comme le lexique le plus complet sur ce sujet.⁷²

2.2.1.6 Les lettres et les particules

La parole *ḥar*, dont le pluriel est *ḥurūf*, est très polysémique. En effet, elle peut signifier 'dialecte' et 'parole', et elle a été utilisée comme terme technique, dans plusieurs domaines linguistiques, de la phonétique jusqu'à la syntaxe.⁷³

69. Suyūṭī, Muzhir, V. I, pp. 476-482

70. Baalbaki, 2014, p. 227.

71. Zaḡḡāḡī, 'Ibdāl, pp. 1-2.

72. Baalbaki, 2014, pp. 229-230.

73. Zaḡḡāḡī, 'Ibdāl, note 1, p.1. Pour les sens de la parole *ḥarf* voir aussi Ibn Ḡinnī, sirr, pp. 13-17.

En phonétique, les lettres sont appelées *ḥurūf al=mu'ğam* ou *ḥurūf al=hiğā* 'lettres de l'alphabet', en s'inspirant à la parole *ḥarf* qui est le *maşdar* du verbe *ḥarafa* 's'incliner ; se détourner d'une direction donnée'.⁷⁴

Selon ibn ğinnī, le *şawt* 'onde sonore' est un '*arağ* 'incident' qui sort avec *nafas* 'air expirée', de manière continue et sans déviation. Dans le cas où le *şawt* rencontre un *maqta* 'lit. lieu de coupure', il est dévié et infléchi de sa continuité et ceci produit un *ğars* 'son' particulier. Les *maqāti* 'lieux de coupure' sont seize et se trouvent dans la gorge, la bouche et les lèvres.⁷⁵ Cependant, il y a trois '*aşwāt* qui ne rencontrent aucune déviation, sortent continus et longs et se terminent avec l'air expirée ; ce sont les voyelles longues ; *ā - ū - ī* ; respectivement appelées '*alif, wāw* et *yā*', dans la grammaire traditionnelle.

On trouve un deuxième terme technique, *ḥurūf al=ma'ānī*. L'appellation est née du terme *ḥarf*, dont le sens est 'bordure ; extrémité ; limite' et regroupe l'ensemble des particules qui ne sont ni un verbe ni un substantif. Dans la tradition arabe, le verbe et le nom ont un sens autonome et complet. Par contre, la particule n'en a pas et représente seulement une partie extrême et non essentielle du sens d'où son appellation *ḥarf*.⁷⁶

Pour le fait que la parole *ḥarf* est polysémique, comme on vient de voir, les travaux lexicaux qui se penchent sur les *ḥurūf* peuvent être distribués selon deux domaines distincts, dont l'un se penche sur les lettres de l'alphabet, dans ce cas, on parlera des *ḥurūf* (tout court) et l'autre est spécialisé dans l'étude des particules, et dans ce cas, on parlera des *ḥurūf al=ma'ānī* .

2.2.1.6.1 Les lettres Les *ḥurūf* sont utilisés, aussi bien, en phonétique, en morphologie et en syntaxe.

En phonétique Ce sont les unités phonologiques, nommées *ḥurūf al=mu'ğam* ou *ḥurūf al=hiğā* , et constituent les 29 lettres de l'alphabet qui regroupent :

- 28 consonnes qui sont illustrées dans le tableau 2.1. Elles sont produites grâce à la déviation du son par un *maqta* 'coupure', et peuvent être *mutaḥarrikah* 'vocalisées' ou *sākinah* 'non vocalisées' ;⁷⁷

74. Ibn Ğinnī, *sirr*, pp. 6-7 et Fleisch, 1958, pp. 78-79.

75. 'al-'Astarabādī, *Şarḥ aš-şāfiyyat*, V. III, pp. 250-254

76. Murādī, *Ġanā*, pp. 23-24.

77. Ibn Ğinnī, *sirr*, pp. 6-7 et Fleisch, 1958, p. 79.

- 3 voyelles longues, appelées *ḥurūf al=madd wa=l=’istiṭālah* ‘lettres de prolongation et de continuité’ : la *’alif*, la *wāw* et la *yā’*. Elles sont produites grâce à l’extension de la sortie du son qui coule continu et prolongé, sans *maqta’* ‘coupure’ et ne peuvent être que *sākinah* ‘non vocalisées’.⁷⁸

Les deux *ḥurūf*, *wāw* et *yā’*, peuvent être consonnes, comme dans le cas de قَوْم *qawm* ‘gens, groupes de personnes’ et يَوْم *yawm* ‘jour ; journée’. Elles jouent aussi le rôle de voyelles longues, respectivement, /ū/ et /ī/, par exemple, dans les paroles أَقُومُ *’aqūmu* ‘je me lève’ et أَبِيعُ *’abī’u* ‘je vends’.

Cette confusion entre les deux rôles, consonantique et vocalique, a pour conséquence que le compte des lettres arabes «ne tourne pas» lorsqu’on parle de 29 lettres où la *wāw* et la *yā’* sont communes aux consonnes et aux voyelles longues.

Une autre confusion se fait entre la *’alif* /ā/, écrite ا, et la *hamza*, lorsqu’elle a la *’alif* comme support, écrite أ. Dans plusieurs sources, même lexicales, les deux lettres sont communément appelées *’alif*.

Par contre, les voyelles courtes ne font pas partie de l’alphabet, ce sont les trois *ḥarakāt*. Elles sont écrites comme diacritiques. Les voyelles /a/ et /u/ sont écrites au dessus de la consonne et la voyelle /i/ au dessous, par exemple, ب *ba* - بُ *bu* - بِ *bi*.

En morphologie Les lettres se divisent en deux groupes, d’une part, les consonnes qui ne peuvent être que radicales et, de l’autre, l’ensemble des lettres qui peuvent être radicales mais peuvent être aussi des lettres d’augment, c’est à dire qu’elles peuvent être ajoutées, pour obtenir une forme fléchie ou une nouvelle parole dérivée.

Dans une parole, les lettres faisant partie de la racine sont appelées *ḥurūf ’ašliyyah*, et les lettres d’augment sont appelées *ḥurūf zā’idah* ou *mazīdah*. Les *ḥurūf* qui peuvent être *mazīdah* sont les consonnes { ’ - t - s - l - m - n - w - y } et les trois voyelles longues, ā, ū et ī.

En syntaxe Certains *ḥurūf al=ma’ānī* ‘les particules’ sont monolittères : bi - ta - fa - ka - la - li - wa. Pour le fait que les voyelles courtes ne sont pas prises en considération

78. Ibn Ğinnī, *sirr*, pp. 7-8 et Fleisch, 1958, pp. 80-84.

dans les travaux lexicaux, ces particules sont étudiées comme des lettres isolées : b - t - f - k - l - w.

La polysémie de la parole *ḥarf* a créé une grande confusion qui se reflète dans les lexiques *mubawwab* spécialisés dans l'étude des *ḥurūf*. Par exemple, dans *risālah fi l=ḥurūfi l='arabiyyah*, l'auteur, al-naḍr bnu šumayl (m. 203/819), énumère les lettres de l'alphabet et spécifie leurs différents types d'utilisation ('*awḡuh/wuḡūh*), qui chevauchent les niveaux phonologique, morphologique et syntaxique et peuvent aller de un (le cas de *tā*' et *ḥā*') à vingt-deux rôles (comme c'est le cas de la '*alif*').⁷⁹

Une des sources principales, et la plus complète du genre, est *sirru šinā 'ati l='i 'rābi*, écrite par ibn ḡinnī (m. 392/1002). Dans son introduction, ibn ḡinnī spécifie que son domaine d'intérêt est celui des lettres *mufradah* 'isolées'.⁸⁰ Son livre est constitué de 29 chapitres, dont chacun est dédié à une lettre de l'alphabet.

Au début de chaque chapitre, l'auteur présente les caractéristiques phonétiques de la lettre et spécifie si elle est *maḡhūrah* 'voisée' ou *mahmūсах* 'sourde'. Il étudie, ensuite, les cas où elle est sujette au phénomène de '*ibdāl* 'substitution', par exemple, la lettre š peut être remplacée par s, c'est le cas lorsqu'on dit *masdūh*, au lieu de *mašdūh*, 'ahuri ; étonné' .⁸¹ Il se penche ensuite sur le rôle morphologique et spécifie si la lettre est seulement '*ašliyyah* ou, elle peut être aussi *zā'idah*. Enfin, une étude détaillée des aspects syntaxiques et sémantiques, dans le cas où la lettre est aussi une particule. L'auteur, ibn ḡinnī, mentionne les rôles que peut jouer la lettre, sans faire de distinction si ce rôle est flexionnel, dérivationnel, lexicale ou sémantique. Par exemple, dans le chapitre de la lettre *nūn*, il mentionne qu'elle est ajoutée, dans *naḍribu* 'nous frappons' ; dans *infa 'ala* 's'agiter' et *nifriḡah* 'faible trouillard'.⁸² Mais il ne spécifie pas que, dans *naḍribu*, la *nūn* est le préfixe de la 1^e personne pluriel au *mudāri* 'inaccompli' du verbe *ḍaraba* 'frapper' ; ou que c'est une lettre d'augment qui permet de dériver le verbe *infa 'ala* de *fa 'ala* 'faire' et l'adjectif *nifriḡah* de la racine *frḡ*. Dans le chapitre de la lettre *bā*' , après une introduction très intéressante pour définir les particules composées d'une seule consonne, suivie d'un commentaire général sur les prépositions, il donne les sens que peut avoir la préposition *bī*, en fonction de son utilisation.⁸³

D'autres auteurs se sont concentrés sur une lettre particulière. Une des lettres qui a tenu beaucoup d'attention est la *hamza*. Des études liées à la prononciation de la *hamza* (*taḡqīq* ou *nabr*) et à son allègement (*taḡfīf* ou *tashīl*) ainsi qu'à sa substitu-

79. Baalbaki, 2014, p. 213.

80. Ibn ḡinnī, *sirr*, p. 5.

81. Ibn ḡinnī, *sirr*, p. 199.

82. Ibid., p. 445.

83. Ibid., pp. 120-135.

tion (*'ibdāl*) par *wāw* ou *yā'* font le sujet de plusieurs ouvrages. Dans ce cas aussi, les auteurs mentionnent le rôle de la *hamza* sans faire une différence entre les niveaux d'étude. Par exemple, 'abū bakr bnu al='anbārī (m. 328/940) a limité son travail, *šarḥ al='alifāt*, à la *hamza* initiale divisée en trois types, selon que le mot en question est un verbe, un nom ou une particule. Il reconnaît six variétés de *hamza*, qu'il appelle toujours 'alif : 'alif al='aṣl (*hamza* radicale : rôle morphologique); 'alif al=qaṭ' (*hamza* disjonctif : rôle phonétique); alif al=waṣl (*hamza* conjonctif : rôle phonétique); 'alif al='istifhām (*hamza* interrogatif : rôle lexicale); 'alif al=muḥbir 'an nafsīhi (*hamza* de la première personne singulière, par exemple dans 'aktubu 'j'écris' : rôle flexionnel) et 'alif mā lam yusamma fā'ilu=hu (*hamza* dans les formes passives 'uf'ila et 'uftu'ila: rôle dérivationnel).⁸⁴

D'autres lettres, ayant divers rôles qui chevauchent la morphologie et la syntaxe, ont été sujet d'étude dans des ouvrages spécifiques. Par exemple, sur la lettre *lām* on a, entre autres, *al=lāmāt*, de 'abū l=ḥasan al=harawī (m. 415/1025), qui comprend trente-quatre types de *lām* selon son utilisation. On peut citer, aussi, *al=fuṣūl al=mufīdah fī l=wāw al=mazīdah*, de ḥalīl bnu kaykaldī al='alā'ī (m. 761/1359), qui traite des rôles de la lettre *wāw*.

2.2.1.6.2 Les particules Les *ḥurūf al=ma'ānī* sont des particules dont le sens se complète lorsqu'elles sont annexées à un nom ou un verbe, et peuvent être :

- spécifiques du nom, par exemple, *ḥurūf al=ğarr* 'les particules du génitif' ;
- spécifiques du verbe, comme *ḥurūf an=naṣb* 'lit. les particules de l'accusatif' ou *ḥurūf al=ğazm* 'les particules de l'apocopé' ;
- communes au verbe et au nom, comme *ḥurūf al='aṭf* 'les conjonctions'.

En outre, la particule peut être constituée d'une consonne, dans ce cas, elle est préfixée à la parole qui la suit. Elle peut être composée de plus d'une consonne (de deux à quatre) et, dans ce cas, c'est une parole isolée.

Les œuvres qui ont traité *ḥurūf al=ma'ānī*, sont surtout de nature syntaxique et peuvent être classées en trois groupes selon le type de l'arrangement considéré.⁸⁵

Un premier groupe de lexiques ignore l'arrangement alphabétique, par exemple *ḥurūf al=ma'ānī* de az=zağğāğī (m. 337/949) et *al='uzhiyyah fī 'ilm al=ḥurūf* de 'abū l=ḥasan al=harawī (m. 415/1025).

84. Baalbaki, 2014, p. 217.

85. Baalbaki, 2014, pp. 213-216.

Un deuxième groupe comprend les œuvres, dans lesquelles les particules sont classées par ordre alphabétique, indépendamment du nombre de leurs radicaux, par exemple, *raṣf al=mabānī fī šarḥ ḥurūf al=maʿānī*, écrit par al=mālaqī (m. 702/1302) et *maṣābiḥ al=maḡānī fī ḥurūf al=maʿānī* de ibn nūr al=dīn al=mawzaʿī (m. 825/1422).

Dans le troisième groupe, les particules sont divisées en fonction du nombre de leurs lettres, et chaque groupe est classé par ordre alphabétique, on trouve *maʿānī l=ḥurūf* (ou *manāzil al=ḥurūf*), écrit par ar=rummānī (m. 384/994) ; *ḡawāhir al=ʿadab fī maʿrifat kalām al=ʿarab* de al=ʿirbilī (m. 741/1340) et *al=ḡanā d=dānī fī ḥurūfi l=maʿānī* écrit par al=Murādī (m. 749/1348).

2.2.2 Domaine de la flexion

Dans les domaines de *aṣ=ṣarf* 'la flexion' que *al=ʿistiḡāq* 'la dérivation', certains aspects sont problématiques et engendrent une grande confusion qui contribue à la propagation du *lahn*. Encore une fois, grammairiens et philologues ont expliqué et énuméré l'usage correct et normalisé et ont dénoncé les usages retenus incorrects.

En ce qui concerne la conjugaison du verbe, les erreurs concernent surtout la voyelle médiane du verbe trilitère à l'inaccompli qui n'est pas formelle et suit les règles *samāʿī* 'selon l'usage'. Cependant, on trouve un seul ouvrage consacré à ce sujet, *buḡyat al=ʿamāl fī maʿrifat mustaqbalāt al=ʿafʿāl* de ʿabū ḡaʿfar al=lablī (m. 691/1291).⁸⁶

En ce qui concerne la flexion des noms et des adjectifs, on trouve des monographies traitant des sujets spécifiques, tel que la flexion nominale en nombre (le singulier, duel et pluriel), ou en genre (féminin et masculin).

2.2.2.1 Le singulier, le duel et le pluriel

Généralement, les lexiques qui traitent ce sujet énumèrent les noms qui manquent d'une ou de deux formes fléchies. En d'autres termes, ils énumèrent les noms qui ont seulement le singulier, comme *al=ʿaraq* 'la sueur' et *al=yamm* 'la mer' ;⁸⁷ ou le duel, comme *labbay-ka*, qui peut être le duel du *maṣdar labb* 'se tenir ferme ; persister', mais on l'utilise pour signifier 'me voici pour te servir une fois après l'autre' ;⁸⁸ ou le pluriel, par exemple, *ḡalābīs* 'quelque chose sans ordre, chaotique'.⁸⁹

Ils traitent aussi les noms dont une des formes a acquis un sens sémantique autonome, par exemple, *al=malawān* 'le jour et la nuit' dont le singulier est *al=malā* 'le

86. Baalbaki, 2014, p. 261.

87. Suyūfī, Muzhir, V. II, p. 201.

88. Pour plus de détails, voir Fīrūzābādī, Qāmūs, p. 133.

89. Suyūfī, Muzhir, V. II, p. 197.

désert' ; *al=ḥağarān* 'lit. les deux pierres' pour désigner l'or et l'argent et *al='aswadān* 'lit. les deux noirs', pour désigner les dattes et l'eau.⁹⁰

Deux livres, généralement représentatifs, sont *al=muṭannā* de 'abū ṭ=ṭayyib al=luğawī (m. 351/962) et *al=muṭannā wa=l=mukannā wa=l=mubannā wa=l=mu'āḥā wa=l=mušbah wa=l=munḥa*, écrit par ibn al=sikkīt (m. 244/858) et reporté par as=suyūṭī dans son œuvre, *al=muzhir*;^{91, 92}

2.2.2.2 Le masculin et le féminin

Plusieurs œuvres, appelées *al=Mudakkar wa=l=mu'annaṭ* 'le singulier et le féminin', traitent les aspects problématiques liés au genre et sont écrits par al=farrā' (m. 207/822), 'abū l=ḥātim as=siğistānī (m. 255/869), al=mubarrad (m. 285/898), niṭṭawayhi (m. 323/935), 'abū bakr bnū al='anbārī (m. 328/940) ; ibn at=tustarī (m. 361/972), ibn ḡinnī (m. 392/1002) et ibn fāris (m. 395/1004). D'autres érudits ont rédigé des ouvrages concernant un sujet spécifique, lié au masculin et au féminin, comme par exemple, le livre *mā yuḍakkar wa-mā yu'annaṭ min al='insān wa=l=libās*, de 'abū mūsā l=ḥāmiḍ (m. 305/918), qui traite exclusivement des parties du corps humain et des vêtements.⁹³

Concernant les noms *mamdūd* et *maqṣūr*, la confusion résulte de l'élision, dans certains dialectes, de la *hamza* du préfixe féminin *-ā'*, ce qui le confond avec l'autre marque *-ā* et conduit à la coalescence, en une seule forme, de deux mots avec des significations différentes. Par exemple, l'élision de la *hamza* finale de la parole *ḡadā'* 'profit' aboutit à une forme identique au mot *ḡadā'* 'don' qui est *maqṣūr*.

Un des critères utiles pour déterminer si la parole doit être *maqṣūr* ou *mamdūd* est la première voyelle. Les mots commençant par la voyelle *a* sont *mamdūd*, mais deviennent *maqṣūr* s'ils commencent par la voyelle *i*, comme première voyelle, par exemple, *balā'* et *bilā'* 'drame ; évènement tragique' et *qarā'* et *qirā'* 'hospitalité ; hébergement'.

Les monographies, spécialisées dans ces problématiques, ont généralement distribué leur matériel, selon la coexistence du *maqṣūr* et son homologue *mamdūd* et dans ce cas, on énumère les mots dont les significations sont différentes et ceux qui ont la même signification. Enfin, on trouve les mots qui sont exclusivement *maqṣūr*, et ceux qui sont exclusivement *mamdūd*.⁹⁴

90. Ibid., V, II, p. 173.

91. Ibid., V, II, pp. 173, 193.

92. Baalbaki, 2014, p. 237.

93. Pour plus de détails, voir Baalbaki, 2014, pp. 239-241.

94. Pour plus de détails : Baalbaki, 2014, pp. 241-247.

En ce qui concerne le classement, les premiers lexiques, rédigés durant le troisième/neuvième siècle, ne suivaient aucun ordre. Le premier lexique, classé alphabétiquement selon la première lettre du mot, est *al=maqṣūr wa=l=mamdūd*, de ibn wallād (m. 332/944), contenant 1168 mots.

Le travail le plus important dans le genre est *al=maqṣūr wa=l=mamdūd*, écrit par 'abū 'alī al=qālī (m. 356/967), avec 1544 mots, classés selon la première lettre du mot dans un ordre phonétique différent de celui de Ḥalīl.⁹⁵

2.2.3 Domaine de la dérivation

2.2.3.1 *al=muṭallaṭāt* 'les triplets'

Un intérêt important a été consacré aux paroles trilitères qui diffèrent, généralement, par une seule voyelle et sont appelées dans la tradition arabe les triplets (*muṭallaṭah* pl. *muṭallaṭāt*), parce que la voyelle peut être *fathah* (a) dans un mot, *kasrah* (i) dans un autre et *ḍammah* (u) dans le troisième. Dans le cas des triplets des substantifs, la voyelle concernée est la première, et dans le cas des verbes, la voyelle concernée est la médiane.

Le sens du triplet peut être identique, comme dans le cas de *ḡaṭwah*, *ḡiṭwah* et *ḡuṭwah* 'tumulus',⁹⁶ et *baḡaḡa*, *baḡiḡa* et *baḡuḡa* 'être heureux ; se réjouir'.⁹⁷ Mais dans la plupart des cas, les paroles ont des significations différentes, par exemple, *ḡallah* 'un installation ; un séjour', *ḡillah* 'direction ; objectif' et *ḡullah* 'veston',⁹⁸ et *ḡalama* 'rêver', *ḡalima* 'être gâté ou perforé ; [le cuir par exemple]', et *ḡaluma* 'devenir indulgent'.⁹⁹

Le premier lexicographe à écrire un livre sur le sujet est Quṭrub (m. 206/821), et il l'a intitulé, justement, *al=muṭallaṭāt*. Il a énuméré sans ordre trente-deux triplets, tous de significations différentes, dont trente triplets sont nominaux et deux sont verbaux.

Le premier livre complet est *al=muṭallaṭāt* de al=baṭalyawsī (m. 521/1127). Il contient 833 triplets dont 695 sont constitués de paroles ayant des significations différentes et 138 ayant le même sens.

En incluant le matériel qu'il a pris des deux lexiques, *taḡdīb al=luḡah*, de al='azharī (m. 370/981), et *ḡamharat al=luḡah*, de ibn durayd (m. 321/933), al=baṭalyawsī a rédigé

95. L'ordre des lettres est : ' , h , ' , ḡ , ḡ , q , k , ḡ , ṣ , y , l , r , n , ṭ , d , t , ṣ , z , s , z , ḡ , ṭ , f , b , m , w. Baalbaki, 2014, p. 246

96. Baṭalyawsī, *Muṭallaṭ* , V. I, p. 393.

97. Ibid., p. 351.

98. Ibid., p. 457.

99. Baalbaki, 2014, p. 249.

un second livre, *'ikmāl al=i'lām bi=taṭlīl al=kalām*, où le nombre des triplets est 2294, dont 163 sont composés de mots qui ont la même signification.

Fīrūzābādī (m. 817/1415), dans *ad=durar al=mubattata fī l=ġurar al=muṭallaṭah*, s'est limité, seulement, aux triplets, dont les mots ont la même signification et il en a recensé 453, qu'il a rangé dans l'ordre alphabétique complet, en commençant par la première consonne du mot et sans prendre en considération les racines.

En comparant le contenu des trois lexiques, on note que, dans *al=muṭallaṭāt* de quṭrub, le nombre très réduit (32 triplets) était restreint seulement aux triplets dont les mots avaient des significations différentes. Le premier à inclure des triplets dont les mots partagent le même sens est *al=baṭalyawsī* (m. 521/1127) qui en a recensé 163. Par contre, fīrūzābādī (m. 817/1415) a recensé 453 triplets. Ce phénomène est dû, d'une part, au recensement progressif du matériel dialectal, trouvé dans les livres sur *ġarīb* et, de l'autre, à l'expansion géographique de la langue arabe. En effet, il semble que la plupart des triplets sont des variantes dialectales du même mot, comme par exemple, *rašwah*, *rišwah* et *rušwah* 'pot-de-vin'.¹⁰⁰

2.2.3.2 Les schèmes nominaux

La tradition arabe s'est occupée des schèmes, aussi bien sur le plan sémantique que lexicale. Dans son *al=kitāb*, sībawayhi (m. 180/796) a recensé 308 schèmes pour les noms. Mais, au cours des siècles, le recensement, selon as-suyūtī, est arrivé à 1210 schèmes nominaux.¹⁰¹

Cependant, un nombre restreint d'ouvrages lexicaux ont traité les modèles nominaux.¹⁰² Dans un court traité, intitulé *kitāb an=nayrūz*, ibn fāris (m. 395/1004) énumère et explique quarante-trois mots de schème *fay'ūl*, par exemple, *nayrūz* 'le nouveau jour de l'an persan' et *bayrūt*. Par contre, aṣ=ṣaġānī (m. 650/1252) a rédigé deux monographies, dont une, intitulée *yaf'ūl*, recueille 41 substantifs de schème *yaf'ūl*, par exemple *ya'fūr* 'gazelle' et *yanbū'* 'fontaine' et, l'autre, intitulée *mā banathu l-'arab 'alā fa'āli*, comprend 130 mots de schème *fa'āli*.

Certaines monographies sont dédiées aux schèmes nominaux appartenant au paradigme verbale. Entre autres, *al=muqtaḍab min kalām al='arab*, de ibn ġinnī (m. 392/1002), traite du schème *maf'ūl* 'participe passif' des verbes trilitères, dont le second radical est une *wāw* ou une *yā'*, tels que *mašūb* 'mélange' et *mašīd* 'construit' dérivant, respectivement, des racines *šwb* et *šyb*.

100. Ibid, p. 248-251.

101. *Suyūtī, Muzhir*, V. II, p. 4.

102. Pour plus de détails voir *Baalbaki, 2014*, pp. 252-253.

Deux monographies de aṣ=ṣaġānī (m. 650/1252), *naq'at al=ṣadyān fī mā ġā'a 'alā l=fa'alān*, comprend 211 *maṣdar* 'noms verbaux' de schème *fa'alān*, comme *dawabān* 'fusion' ; tandis que la deuxième, *al=infi'āl*, contient 437 verbes de schème *infa'ala* et leurs noms verbaux de schèmes *infi'āl*, comme *insadaḥa* 's'allonger ; s'étendre'.

Considéré le premier lexique dont les éléments lexicaux sont ordonnés selon le dernier radical, *at-taqfiyah* de al=bandanīġī (m. 284/897) est une œuvre très significative dans l'histoire de la lexicographie arabe. Il faut noter, cependant, que al=bandanīġī n'a pas pris en considération la première et la deuxième consonnes radicales. Dans l'introduction, l'éditeur, ḥalīl 'ibrāhīm al=aṭiyyah, explique que le titre, *at-taqfiyah*, se réfère à *qawāfi*, pluriel de *qāfiyah* 'dernière partie de quelque chose',¹⁰³ et que le classement se base sur la dernière lettre de la parole. En plus, la *qāfiyah* de la poésie désigne la dernière parole du deuxième hémistiche du vers poétique.¹⁰⁴ L'auteur, al=bandanīġī, a introduit un deuxième critère de classification qu'il a appelé *al='afā'il*. Ces derniers ne représentent pas un schème particulier mais tous les schèmes qui peuvent se substituer l'un à l'autre pour former la rime dans une poésie.¹⁰⁵ Il a donc regroupé toutes les paroles qui peuvent être utilisées dans le même poème, comme *qāfiyah*. Par exemple ; *ġalīz*, *taqyīz* et *ġuḥūz*, respectivement, de schème *fa'il*, *taf'il* et *fu'ūl* font partie de la même *qāfiyah*.¹⁰⁶

2.2.3.3 Les schèmes verbaux

Une autre confusion, également liée aux différences dialectales, concerne le schème trilitère, *fa'ala* ou *fa'ila* ou *fa'ula*, et le schème du verbe dérivé, *'af'ala*. Dans certains cas, ils peuvent avoir la même signification, par exemple, *baraqa* et *'abraqa* signifient tous les deux 'menacer'. Dans d'autres cas, ils ont deux significations différentes, par exemple *tāba* 'revenir' et *'atāba* 'récompenser'.¹⁰⁷

Toujours dans l'objectif d'identifier l'usage correct de celui qui ne l'est pas, plusieurs érudits ont rédigé des lexiques où, généralement, ils font la recension de *fa'ala* et *'af'ala*, qui ont la même signification, ceux qui ont des sens différents ainsi que les cas où on trouve seulement *fa'ala* ou *'af'ala*.

On peut citer, entre autres, deux livres avec le même nom *fa'altu wa-'af'altu*, un écrit par 'abū ḥātim as=siġistānī (m. 255/869), sans ordre ni distinction entre les deux schèmes. Le deuxième est écrit par az=zaġġāġ (m. 311/923), qui sépare les deux

103. Ibn Manẓūr, *Lisān*, V. 15, p. 193

104. Fīrūzābādī, *Qāmūs*, p. 1326

105. Bandanīġī, *Taqfiya*, pp. 22-24.

106. Baalbaki, 2014, pp. 370-372.

107. Baalbaki, 2014, p. 254

types *fa'ala* et *'af'ala*, et organise le matériel par ordre alphabétique, selon la première consonne radicale seulement.

Trois livres, intitulés tous *kitāb al-'af'āl*, sont dédiés aux verbes, dont un est écrit par ibn al-qūṭīyah (m. 367/977) et traite particulièrement les schèmes *fa'ala* et *'af'ala*; les deux autres, par contre, sont rédigés par as-saraqustī (m. après 400/1010) et ibn al-qatṭā' (515/1121) et traitent tous les schèmes verbaux.¹⁰⁸

2.2.3.4 Les schèmes nominaux et verbaux

Innovateur dans le domaine, 'iṣḥāq bnu 'ibrāhīm al-fārābī (m. 350/961) est le premier lexicographe à écrire une œuvre complète sur les schèmes nominaux et verbaux. Il est aussi le premier à adopter le système de rime, utilisé dans *at-taqfiyah* par al-bandanīgī (m. 284/897). En plus, il a idéalisé et perfectionné son classement de manière aussi parfaite qu'il sera adopté, plus tard, par de nombreux lexicographes à partir de son neveu al-ğawharī (m. 400/1010). Son lexique, *dīwān al-'adab*, est méticuleusement divisé en sections dont chacune est appelée *kitāb*. Chaque *kitāb* est divisé en une sous-section dédiée aux noms, suivie d'une autre pour les verbes. À son tour, chaque sous-section est divisée en chapitres classés par ordre alphabétique, selon la dernière consonne radicale. Dans *at-taqfiya*, al-bandanīgī (m. 284/897) a pris en considération, seulement, le dernier radical ; al-fārābī, par contre, a pris en considération également le premier et puis le deuxième radical. Pour un classement ultérieur des mots, il a pris en considération plusieurs autres critères ; les formes trilitères précèdent celles augmentées, les voyelles suivent l'ordre où *sukūn* 'absence de voyelle' est classé avant la *fathah*, suivie par la *ḍammah* et enfin la *kasrah*. Ainsi on trouve les schèmes dans l'ordre, *fa'ala*, *fa'ula* et *fa'ila*, et on trouve aussi *fa'l* avant *fa'al*.¹⁰⁹

Enfin on peut citer également, *šamsu al-'ulūmi wa-dawā'u kalāmi al-'arabi mina al-kulūmi*, de našwān bnu sa'īd al-ḥimyarī (m. 573/1178), qui représente le travail le plus étendu du genre. En ce qui concerne le classement, našwān a appliqué l'ordre alphabétique complet aux racines. Dans l'introduction générale, Il explique qu'il avait l'intention d'organiser son lexique d'une manière qui le protège contre les erreurs des diacritiques et des voyelles. Ainsi, il a divisé son lexique en sections (*kutub*) de la 'alif jusqu'à la *yā'*, selon le premier radical, puis il a arrangé chaque *kitāb* en chapitres ('*abwāb*), qui sont basés sur le second radical, mais qui tiennent compte des autres radicaux dans l'arrangement.¹¹⁰

108. Ibid., pp. 254-260.

109. Ibid., pp. 262-263.

110. Ibid., pp. 265-266.

2.2.4 Domaine sémantique

2.2.4.1 Livres des proverbes

Les proverbes étaient considérés un type spécial de *kalām al-ʿarab*, dignes d'être enregistrés et d'être utilisés comme *šawāhid*. Ils sont fréquemment cités dans les lexiques surtout pour expliquer les mots *ġarīb*. Les philologues s'intéressaient à leurs sens et à leurs significations métaphoriques, les citaient comme ils les entendaient, sans aucun changement et même s'ils contenaient des paroles nouvelles reconnues comme *lahn*.¹¹¹ Les premiers livres consacrés aux proverbes sont apparus au premier/septième siècle et depuis ils ont continué à prospérer. Dans ces monographies, les proverbes sont rangés sous forme de lexique, d'abord sans ordre alphabétique (depuis 168/784 jusqu'à 328/940 environ) et puis, classés alphabétiquement, partiellement ou entièrement. Certaines œuvres contiennent, en plus des proverbes proprement dits, ceux produits dans la poésie, dans le Qur'ān et dans le Ḥadīṭ. D'autres incluent les paroles sages de différents genres, ou proverbes *muwallad* 'post-classique' et 'āmī 'vernaculaire'.¹¹²

2.2.4.2 Thèmes sémantiques divers

Lors de *ġam' al-luġah* 'collecte des données', les philologues enregistraient, dans de courtes monographies à thème unique, l'usage *faṣīḥ* des bédouins concernant, par exemple, les plantes, les animaux, l'être humain, etc. Étant donné que ces œuvres appartiennent, pour la plupart, au deuxième/huitième et troisième/neuvième siècles, elles ne sont pas ordonnées. En plus, elles contiennent un taux élevé de mots *ġarīb*, et elles sont enrichies par des *šawāhid*, tirés principalement de la poésie.¹¹³

Les travaux sur les plantes traitent les noms d'arbres et d'herbages, les références à leurs caractéristiques et utilisation, les méthodes de leur culture, et leurs habitats. Particulièrement les monographies, concernant *an-naḥl* 'le palmier', traitent les aspects spécifiques des palmiers, tels que fruits, branches, hauteurs, types, défauts, etc.

Basée sur des travaux antérieurs et sur les informations fournies par les bédouins, la monographie, *an-nabāt* de 'abū ḥanīfah ad-dīnawārī (m. 282/895), est considérée l'œuvre la plus complète du genre. Elle représente une encyclopédie botanique qui contient, non seulement, des descriptions détaillées les plantes et leurs utilisations, mais aussi, différents aspects de la tradition bédouine liée aux plantes. Concernant le

111. Baalbaki, 2014, p. 102.

112. Pour une vision plus générale, voir Baalbaki, 2014, pp. 100-104.

113. Baalbaki, 2014, pp. 132-133.

classement, ad-dīnawarī, dans son introduction, affirme qu'il préfère l'ordre selon la première consonne, plutôt que la dernière. La déclaration de ad-Dīnawarī fait penser que l'ordre alphabétique, basé sur la lettre finale du mot, existait déjà, bien qu'on ne connaît aucun lexique du troisième siècle, qui soit rangé sur la base des lettres finales.¹¹⁴

Les travaux sur les animaux traitent, entre autres, *al=ʿibil* 'les chameaux', *al=wuḥūš* 'les animaux sauvages', *an=naḥl* 'les abeilles'. Mais le thème privilégié concerne *al=ḥayl* 'les chevaux' et, peut être concentré sur des chevaux particuliers et la poésie dans laquelle ils ont été mentionnés ou inclure aussi les thèmes relatifs aux parties du corps, hennissements, couleurs, imperfections, démarche, etc.¹¹⁵

Les travaux concernant l'être humain sont très liés à ceux sur les animaux. On distingue les monographies, intitulées *kutub al=farq*, qui étudient la différence entre les êtres humains et les animaux et où sont différenciés les termes utilisés exclusivement pour l'homme et ceux utilisés pour les animaux. Par exemple les termes *'anf*, *ḥaṭm*, *ḥurṭūm*, *'irnīn* et *marsin*, sont utilisés pour désigner le nez humain et, en contrepartie on trouve *nuḥarah* pour celui de l'âne ; *ḥurṭūm* et *ḥaṭm* pour celui du chameau ; *fintīṣah* 'le mufle de la vache et le groin du cochon' ; *ḥarṭamah* et *'awtamah* 'le museau du chien' et *qirṭimatān* 'le bec du pigeon'.¹¹⁶

D'autres monographies, appelées généralement *kutub ḥalq al=ʿinsān*, peuvent être considérées une récolte impressionnante des vocabulaires liés aux parties du corps humain : la tête, le cuir chevelu, les oreilles, les cheveux, la barbe, le visage, le front, les sourcils, et se terminant par les organes génitaux, les hanches, jambes et pieds. On y trouve aussi des termes correspondant aux fonctions physiologiques comme la grossesse, la procréation et les différentes étapes de la vie humaine de la naissance à la vieillesse humaine. Dans ce domaine, *ḥalq al=ʿinsān*, écrit par le médecin saʿīd bnu ḥibat allāh bnu al=ḥusayn (m. 495/1101), n'appartient pas strictement au genre, mais c'est un texte médical, de cinquante chapitres, qui traite principalement des organes sexuels, des rapports sexuels, de la grossesse, des caractéristiques du fœtus, des soins infirmiers, de l'éducation des enfants, ainsi que de certaines questions philosophiques et métaphysiques, comme la différence entre *rūḥ* et *nafs*, tous deux utilisés indifféremment pour désigner l'âme et l'esprit.¹¹⁷

114. Baalbaki, 2014, p. 138-140.

115. Naṣṣār, 1988, pp. 102-105 et Baalbaki, 2014, pp. 144-148.

116. Baalbaki, 2014, p. 149

117. Naṣṣār, 1988, pp. 105-108 et Baalbaki, 2014, pp. 148-156.

2.2.5 Trésaurus à thèmes divers

Ils s'approchent beaucoup des œuvres sémasiologiques mais, ils sont constitués de sections divisées en chapitres dont, chacun est dédié à un thème spécifique. Les thèmes généraux peuvent concerner le corps humain, les vêtements, les aliments, le vin, les ustensiles, les maladies, les demeures, les armes, la guerre, les chameaux, les chevaux, etc. Tandis que, les thèmes linguistiques se concentrent sur les schèmes nominaux et verbaux, les *'aḍḍād*, les mots arabisés, le masculin et le féminin, les variantes dialectales, etc.

Le premier travail existant dans le genre est *al=garīb al=muṣannaf*, écrit par 'abū 'ubayd al=qāsīm bnu sallām (m. 224/838). Son arrangement est bien élaboré, au point que certaines sections du livre ont été distribuées sous forme de monographies indépendantes. La nature exhaustive du livre en a fait une des bases essentielles pour les auteurs postérieurs. Constitué de vingt-six sections, dont chacune, appelée *kitāb*, traite un thème sémantique général ou un sujet linguistique. Bien qu'il contienne, selon az=zubaydī, un total de 17970 entrées, le livre se caractérise par sa brièveté dans l'explication des mots et par la relative rareté des *šawāhid*.¹¹⁸

Une autre œuvre aussi célèbre, *'adab al=kātib* de ibn qutayba (m. 276/889), est divisée en quatre sections traitant de divers thèmes. La première section, *kitāb al=ma'rifa* 'Le livre de la Connaissance', contient certains chapitres traitant des mots et des expressions mal utilisés par les écrivains, et d'autres traitant des thèmes généraux. La deuxième section, *kitāb taqwīm al=yad* 'Le livre de l'amendement de l'écrit', traite des questions orthographiques. La troisième section, *kitāb taqwīm al=lisān* 'Le livre de l'amendement du discours', traite essentiellement de *lahn al='āmmah* 'erreurs de la populace'. La dernière section, *kitāb al='abniyah* 'Le livre des schèmes', traite principalement des schèmes verbaux et nominaux, en plus d'autres thèmes comme, la métathèse et les mots arabisés.¹¹⁹

Le lexique le plus complet dans le genre, qui nous est parvenu, est *al=muḥaṣṣaṣ* écrit par ibn sīdah (m. 458/1066). Il est tellement exhaustif qu'il a, probablement, mis fin aux travaux encyclopédiques de cette envergure. En effet, l'auteur a rassemblé, intégralement, un grand nombre d'ouvrages antérieurs et contemporains, des lexiques de type *mubawwab* et *muğannas* ainsi que divers livres sur la grammaire et la philologie.¹²⁰

118. Baalbaki, 2014, pp. 268-270.

119. Ibid., p. 270-271

120. Naṣṣār, 1988, pp. 169-170 et Baalbaki, 2014, pp. 275-278.

2.3 Lexiques sémasiologiques

Les lexiques *muğannas* 'sémasiologiques' visent à épuiser le corpus linguistique de l'arabe, c'est-à-dire l'inclusion de toutes les racines de la langue et la plus grande partie des dérivées de ces racines. Traditionnellement, ils se distinguent selon le critère de l'arrangement du matériel qu'ils comprennent et, on distingue trois écoles principales. La première école suit le fameux ordre phonétique-permutatif, où l'ordre des éléments suit le classement des lettres radicales selon leur lieu d'articulation. La seconde école est née à cause de la complexité du système phonétique, de conséquence, certains auteurs ont suivi l'ordre alphabétique, selon le premier radical de la parole ou au moins selon la première lettre indépendamment qu'elle soit radicale ou non radicale (préfixe). Toutefois, l'adoption de *al=fārābī* (m.350/961) du classement, selon le dernier radical, appelé communément le système de rime, a donné naissance à la troisième école.

2.3.1 Ordre phonétique-permutatif

Les auteurs qui ont suivi ce système de classement ont eu des objectifs différents et ont donné de grandes contributions à l'évolution du lexique arabe. Le premier lexique qui a suivi l'ordre phonétique-permutatif, *kitāb al='ayn*, est également le premier lexique *muğannas* en absolu.¹²¹ En plus, l'auteur de *al='ayn*, *al=ḥalil* (m. 175/791), a pu faire le recensement complet des racines de la langue arabe, grâce à un système, appelé *at=taqālib*, qui lui a permis de contrôler toutes les permutations possibles des consonnes radicales.

En écrivant *al=bāri' fī l=luğah*, *al=qālī* (m. 356/967) a innové un système pour contrôler les erreurs orthographiques. Dans *tahqīb al=luğah*, *al='azharī* (m. 370/981) a essayé de contrôler les erreurs lexicales, commises dans les lexiques précédents. Dans *al=muḥīṭ fī l=luğah*, *ibn 'abbād* (m. 385/995) a tenté de compléter ce qui manquait de *ḡarīb*. Et enfin, *ibn sīda* (m. 458/1066) a voulu réunir les données éparses dans les divers livres précédents dans son lexique, *al=muḥkam wa=l=muḥīṭ al='a'zam*.¹²²

***kitāb al='ayn* de *ibn 'ahmad ibn 'amr al=ḥalil* (m. 175/791)**

Spécialiste de la phonétique, de la prosodie et de la lexicographie, *al=ḥalil* (m. 175/791) est considéré le premier lexicographe à vouloir faire une recension complète

121. *Suyūṭī, Muzhir*, V. I, p. 76

122. *'aṣ-Ṣāliḥ*, 2009, V. I, p. 365

de la langue. En composant son livre *kitāb al-‘ayn* (du nom de la dix-huitième lettre de l’alphabet), il a essayé de rationaliser la pratique empirique de la lexicographie.

En faisant référence au calcul des arrangements et des combinaisons, il a pu énumérer, de manière exhaustive, toutes les racines de la langue arabe, qu’elles soient *musta‘malah* ‘utilisées’ ou *muhmalah* ‘non utilisées’. En se basant sur le fait que les lettres de l’alphabet arabe sont 29 et que les racines sont composées au minimum de 2 et au maximum de 5 lettres, il a pu calculer le nombre de permutations des radicaux qui composent les racines. Ces permutations, appelées *taqālib*, lui ont permis de calculer le nombre des combinaisons non répétitives des racines arabes. Ainsi une racine bilitère ne peut avoir que deux *taqālib* ou combinaisons : C_1C_2 et C_2C_1 . Une racine trilitère peut donner six combinaisons : $C_1C_2C_3 - C_1C_3C_2 - C_2C_1C_3 - C_2C_3C_1 - C_3C_1C_2$ et $C_3C_2C_1$. Pour les racines quadrilitère et quinquilitère, on peut avoir, respectivement, vingt-quatre et cent vingt combinaisons possibles.¹²³

L’existence réelle des racines est contrôlée par les règles d’incompatibilités des phonèmes de la racine et les restrictions sur la cooccurrence de certaines consonnes dans la même racine.¹²⁴ Les mêmes limites phonologiques permettent aussi de contrôler si une parole est réellement arabe ou fruit du néologisme.¹²⁵ Les racines sont ensuite reportées dans le lexique et, pour les ordonner, al-ḥalil a utilisé deux critères.

1. un classement phonétique : le lexique comprend vingt-neuf chapitres, dont chacun porte le nom de la lettre à laquelle il est dédié. Les chapitres sont classés selon l’ordre des sons, des plus postérieurs (glottales, laryngales, pharyngales) aux plus antérieurs (labiales) :

‘, ḥ, h, ḥ, ġ, q, k, ğ, š, ḍ, š, s, z, ṭ, d, t, ḏ, ṭ, r, l, n, f, b, m, w, ‘*alif*, y, *hamza*.¹²⁶

2. La longueur de la racine : les racines sont ordonnées, selon le nombre des radicaux qui les forment : bilitère, trilitère, quadrilitère et quinquilitère. Ainsi, chaque chapitre est divisé en sections (‘*abwāb* pl. de *bāb*). Chaque section (*bāb*) est dédiée à un type de racines de deux, trois, quatre et enfin de cinq consonnes. En outre, elle comprend toutes les permutations possibles des consonnes d’une combinaison donnée, qu’elles soient *musta‘malah* ‘utilisées’ ou *muhmalah* ‘non utilisées’.

123. Baalbaki, 2014, p. 294

124. Concernant les restrictions sur la cooccurrence des consonnes, voir la section 2.5.

125. Par exemple, aucune racine arabe ne commence par /n/ suivi par /r/.

126. Baalbaki, 2014, p. 290; Naṣṣār, 1988, V. I, p. 176

Par exemple, pour chercher le verbe *kataba*, il faut chercher sa racine *ktb* qui se trouve dans le chapitre *ḥarf al=kāf* 'la lettre k' ; dans la section concernant la combinaison de *k* et *t* avec *b* (*bāb al=kāf wa=t=tā' wa=l=bā' ma'a=humā*), qui contient toutes les combinaisons, *ktb*, *kbt*, *bkt*, *tkb* et *btk* desquelles les combinaisons *ktb*, *kbt* et *btk* sont utilisées et accompagnées des lemmes qui en dérivent et leurs gloses.¹²⁷

al=bāri' fī l=luḡah de 'abū 'alī al=qālī (m. 356/967)

Le classement phonétique a été imité, successivement, par divers lexicographes, entre autres, *al=qālī* (m. 356/967) dans son lexique, *al=bāri'*. Cependant, *al=qālī* a adopté un ordre phonétique légèrement différent de celui adopté par *al=ḥalīl* et qui est probablement, le suivant :¹²⁸

ʾ, h, ʿ, ḥ, ġ, ḫ, q, k, ḏ, ġ, š, y, l, r, n, ṭ, d, t, ṣ, z, s, ẓ, ḏ, ṭ, f, b, m, w

Le processus de collecte des données, *ġam' al=luḡah*, étant terminé, *al=qālī* et les autres lexicographes, après lui, ont dû se baser sur le matériel des philologues antérieurs. En effet, *al=bāri'* est considéré le premier lexique qui a rassemblé les documents dispersés dans plusieurs sources, entre autres, *al='ayn*.¹²⁹

En analysant et en confrontant divers matériel, *al=qālī* a pu porter des améliorations aux travaux antérieurs. En notant la gravité du *taṣḥīf* 'erreurs orthographiques', il a innové un système permettant le contrôle orthographique. Ainsi, pour le contrôle des diacritiques, il précisait, dans le cas nécessaire, si la lettre est *mu'ġamah* 'avec des points diacritiques' ou *muhmalah* 'sans points diacritiques'. Par contre, pour contrôler les voyelles, il tient à indiquer la voyelle ou le schème de la parole ; par exemple :¹³⁰

وَيَقَالُ مَاءٌ سَجَسَ بِفَتْحِ السَّيْنِ وَسُكُونِ الْجِيمِ [...]
 وَسَجِسَ عَلَى مِثَالِ فَعِيلٍ [...]

on dit eau saġs, en vocalisant la lettre s avec la voyelle a et la lettre ġ sans voyelle [...]
 et saġīs de schème fa'il [...]

Ces innovations ont permis aux lexicographes postérieurs, *al=ġawharī* (m. 400/1010) et *al=fīrūzābādī* (m. 817/1415), de contrôler l'orthographe et la vocalisation du texte dans leurs lexiques, respectivement, *aṣ=ṣaḥāḥ* et *al=qāmūs al=muhīṭ*.

127. Ḥalīl, 'Ayn, V. 5, pp. 341-342

128. Baalbaki, 2014, p. 305.

129. Ibid., pp. 304-310.

130. Naṣṣār, 1988, pp. 250-251; Baalbaki, 2014, pp. 309-310.

tahdīb al=luġah de al=’azharī (m. 370/981)

Au quatrième/dixième siècle, al=’azharī (m. 370/981) a rédigé *tahdīb al=luġah* ‘correction/amendement de la langue’. En effet, dans son introduction, il déclare avoir recensé les données de nombreux lexiques *mubawwab* et de deux *muġannas*, *al=’ayn* et *al=ġamharah*. Il affirme que son objectif est de porter les corrections aux ressources antérieures qui ont subi les erreurs de *taṣḥīf* et de s’assurer, personnellement, de la validité du matériel auprès des bédouins arabes. Pour ce dernier fait, al=’azharī est l’un des derniers lexicographes qui déclarent avoir recueilli les données directement auprès des bédouins.¹³¹

al=muḥīṭ fī l=luġah de aṣ=ṣāḥib bnu ‘abbād (m. 385/995)

Bien que aṣ=ṣāḥib donnait de brèves définitions avec très peu de *ṣawāhid*, *al=muḥīṭ* est considéré le lexique le plus riche de son siècle, grâce à l’effort de son auteur, qui s’est singularisé à énumérer, pour la première fois, une grande quantité de données linguistiques.¹³²

al=muḥkam wa=l=muḥīṭ al=’a ‘zam de ibn sīdah (m. 458/1066)

Voulant que son lexique soit exhaustif, ibn sīdah a incorporé le matériel de plusieurs sources *mubawwab* et ouvrages grammaticaux, en plus des lexiques *muġannas* antérieures, *al=’ayn*, *al=ġamharah* et *al=bāri’*. Toutefois, il a essayé d’être bref dans ses explications et de limiter le nombre des *ṣawāhid* et l’arrangement qu’il a suivi est identique à celui dans *al=’ayn*.

En plus, ibn sīdah est considéré le premier lexicographe à exposer un plan incluant les normes à suivre dans la présentation et l’agencement interne au sein des lemmes. Dans son introduction, il fait noter qu’il est superflu de mentionner certaines formes ou schèmes qui sont réguliers ou formels (*qiyāsī*, *muṭṭarid*) et, souligne l’importance d’organiser les éléments qui composent l’élément lexical. Ainsi, pour les verbes, il indique le verbe non-augmenté avant ceux qui en sont dérivés ; et pour chaque verbe, il énumère, successivement, *al=māḍī* ‘l’accompli’, *al=muḍāri’* ‘l’inaccompli’, *al=maṣdar* ‘le nom verbal’ et enfin *aṣ-ṣifah* ‘l’adjectif’. Pour les noms, il mentionne le singulier avant le pluriel et le pluriel de paucité avant le pluriel de pluralité. Bien qu’il n’ait pas toujours

131. Naṣṣār, 1988, pp. 260-261; Baalbaki, 2014, pp. 312-313.

132. Baalbaki, 2014, pp. 319-322.

appliqué les règles établies par lui-même, on peut lui reconnaître l'effort, du moins au niveau théorique, d'établir un modèle systématique pour lister les données.¹³³

2.3.2 Ordre permutatif / alphabétique

Plusieurs lexiques *mubawwab* ont été classés par ordre alphabétique dont le premier est *kitāb al-ǧīm* de aš-Šaybānī (m. 206/821), suivi par beaucoup d'autres.¹³⁴ Par contre, trois grands lexiques *muǧannas* appartiennent à cette école, dont le premier à suivre l'ordre alphabétique complet est *ǧamharat al=luǧah* de Ibn Durayd (m. 321/933) suivi par *maqāyīs al=luǧah* et *Muǧmal al=luǧah* tous les deux écrits par Ibn Fāris (m. 395/1004).

ǧamharat al=luǧah de ibn durayd (m. 321/933)

Les auteurs antérieurs donnaient une grande importance au matériel *ǧarīb* 'étrange, étranger' et, l'incluaient avec le matériel fréquent sans faire de distinction entre les deux. Dans l'introduction de *ǧamharat al=luǧah*, Ibn Durayd déclare que son objectif est se concentrer surtout sur ce qui est réellement attesté, en excluant le *waḥṣī* 'mot grossier', le *mustankar* 'répréhensible'.

Un autre caractéristique de ibn durayd est le fait qu'il est le premier à remplacer la définition d'un élément lexical par la parole *ma'rūf*, pour dire que la parole est bien connue et n'inclue aucune définition.¹³⁵ Cependant, l'intérêt de l'auteur pour les dialectes, les paroles arabisées et l'étymologie des noms propres, a eu pour conséquence que *al=ǧamharah* abonde d'articles lexicaux qui peuvent être considérés rares et obsolètes. Pour cette raison, *al=ʿazharī* (m. 370/981) a accusé Ibn Durayd de falsifier l'arabe et d'introduire des néologismes. Toutefois, on note que ibn durayd, lui-même, exprime son propre doute sur l'authenticité de certains éléments lexicaux, en les introduisant par des expressions telles que, *laysa bi=ṭabt* 'pas certain'; *lā aḥuqqu=hu* 'je ne peux pas confirmer son exactitude' ou *lā adrī mā ṣiḥḥatu=hu* 'je ne connais pas son exactitude'.¹³⁶

Dédiant son lexique au *ǧumhūr* 'la foule; la masse de la population', il a adopté l'ordre alphabétique pour faciliter et simplifier l'utilisation de son lexique. Cependant, il

133. Naṣṣār, 1988, pp. 287-303; Baalbaki, 2014, pp. 322-329.

134. On trouve, par exemple, *al=muǧarrad fī ǧarīb kalām al=ʿarab wa-luǧātihā* de Kurāʿan=Naml (m. 310/922); *al=maqṣūr wa-l-mamdūd* de Ibn Wallād (m. 332/944) et *dīwān al=ʿadab* de al=Fārābī (m. 350/961).

135. Baalbaki, 2014, p. 345.

136. Naṣṣār, 1988, V. II, pp. 336-339.

a gardé la permutation des consonnes radicales et la longueur de la racine, comme critères de classification. Unique dans son classement, il a divisé le matériel en sections, selon la longueur des racines, bilitères, trilitères etc. Chaque section est ensuite ordonnée selon l'ordre alphabétique du premier, du second et enfin du troisième radical. Par exemple, on trouve le verbe *kataba* en cherchant la racine *ktb* qui se trouve (avec les autres possibles combinaisons) dans le chapitre de la lettre *bā*, section *btk*. L'arrangement unique de *al=ǧamharah* peut donc être qualifié d'alphabétique-permutatif.¹³⁷

***maqāyīs al=luǧah* et *muǧmal al=luǧah* de ibn fāris (m. 395/1004)**

L'œuvre, *maqāyīs al=luǧah*, est basée sur deux notions philologiques de ibn fāris. La première est d'explorer systématiquement la relation entre la racine et le sens. Une approche qu'ibn fāris appelle *qiyās* et, qui exprime le sens de base ou la signification fondamentale de la racine. Par exemple, le *qiyās* de la racine *ǧhf* est 'balayer quelque chose dans son intégralité'. L'auteur affirme qu'il traite exclusivement les *maqāyīs* (pl. de *qiyās*) corrects de la langue arabe. Par conséquent, il a exclu les mots qui résultent du '*ibdāl* 'substitution'; du *qalb* 'métathèse'; les mots arabisés; les particules; les noms de lieux et les noms de plantes. La deuxième notion est le *naḥt*, que l'auteur définit comme le mélange de deux mots ou le mélange d'un mot et d'un affixe, pour former une seule parole. La parole composée a deux caractéristiques, d'un côté, elle dérive d'une racine qui contient plus de trois radicaux et de l'autre, elle contient une partie de sens de chacun des deux paroles (ou racines) desquelles elle dérive.¹³⁸

Par contre, le deuxième lexique, et comme l'indique son nom *muǧmal al=luǧah* 'résumé/synopsis de la langue', est basé sur la concision et la facilité d'utilisation. L'auteur limite son matériel au seul matériel *ṣaḥīḥ* 'correct' et n'inclue le matériel *ǧarīb* que dans le cas où il est fréquemment utilisé. En outre, il énumère les lemmes avec des définitions brèves, en évitant les longues citations, et en réduisant les *šawāhid*.¹³⁹

'*asās al=balāǧah* de az=zamaḥṣarī (m. 538/1144)

Les deux auteurs, ibn durayd et ibn fāris, ont suivi l'ordre alphabétique mais de manière non systématique. Le premier a organisé certains de ses chapitres, selon les schèmes plutôt que sur la succession des radicaux. Le second a énuméré les quadrilitères et quinquilitères sans ordre alphabétique.

137. Baalbaki, 2014, p. 338.

138. Ibid., pp. 351-353.

139. Ibid., pp. 354-356.

Dans *'asās al=balāghah*, az-zamaḥṣārī, par contre, a suivi l'ordre alphabétique de manière constante et sans considérer les *taqālib* et la longueur des racines. Un autre caractéristique pour laquelle *'asās al=balāghah* est considéré un travail innovateur est l'objectif de son auteur. En effet, la contribution majeure de az=zamaḥṣārī est la distinction, dans ses définitions, entre le *mağāz* 'métaphorique' et le *ḥaqīqī* 'réel'. Ainsi la plupart des lemmes sont divisés en deux parties. La première définition représente l'usage réel, alors que la seconde est normalement introduite par l'expression *wa=min al=mağāz* ou *wa=min al=musta'ār* 'appartenant au métaphorique' et comprend les significations métaphoriques.

En plus, az=zamaḥṣārī cite les éléments lexicaux, non comme unités isolées, mais utilisées dans des contextes qui permettent d'expliquer leurs sens, aussi bien réel que métaphorique. En cela, il diffère significativement des autres auteurs qui citaient normalement les éléments lexicaux isolés, parfois enrichi par les *šawāhid*.

Il faut aussi noter que, selon az=zamaḥṣārī, *al=faṣāḥah* autrement dit l'éloquence et l'usage correct de la langue n'est pas liée, exclusivement aux époques antérieures. En effet, il cite des *šawāhid* attribués à des poètes, considérés *muwalladūn*, parce qu'ils ont vécu après l'époque de la récolte des données, comme 'abū nuwās (m. 198/814) et al=mutanabbī (m. 354/965).¹⁴⁰

2.3.3 Ordre alphabétique second le dernier radical de la racine

Dans les ressources, les philologues considèrent al=ğawharī (m. 400/1010) comme l'innovateur de l'ordre selon la rime, c'est à dire selon la dernière consonne de la racine, bien que cet ordre existait avant lui.

En effet, *at=taqfiyah fī=l=luğah*, de bandanīğī (m. 284/897), est considéré le premier lexique existant, où le matériel est rangé selon le dernier radical mais, bandanīğī ne prenait pas en considération les autres radicaux. En plus, comme il a déjà été mentionné, le classement de *at=taqfiyah* est également basé sur *al='afā'īl*, c'est à dire les schèmes morphologiques des mots qui peuvent former la rime dans une poésie.

Dans *dīwān al='adab*, al=fārābī (m. 350/961), aussi, a adopté le classement selon la rime, en prenant en considération, également, le premier et le second radical. Ainsi il a appliqué le système de rime en entier (le dernier, le premier et enfin le deuxième radical). Mais il a gardé le critère des *'afā'īl* 'schèmes morphologiques'.

En créant son lexique, *tāğ al=luğah wa-ṣaḥāḥ al='arabiyyah* 'la couronne de la langue et le valide de l'arabe', al=ğawharī (m. vers 400/1010) est le premier lexico-

140. Naṣṣār, 1988, V. II, pp. 550-567; Baalbaki, 2014, pp. 356-363.

graphe à appliquer le système de rime en entier, en ignorant les autres critères de classification utilisés dans les œuvres antérieures, comme *at=taqālib* 'les permutations', *al='afā'il* 'les schèmes morphologiques' et la longueur de la racine.¹⁴¹

Parmi les dictionnaires qui ont adopté les principes de ce classement, on trouve les plus importants :

- *al='ubāb al=zāhir wa=l=lubāb al=fāhir* 'le torrent resplendissant et la pulpe somptueuse' de *aş=şağānī* (m. 650/1252)
- *lisān al='arab* 'la langue des arabes' de *ibn al=manzūr* (m. 1311);
- *al=qāmūs al=muḥīṭ* 'l'océan universel' de *al=fīrūz'abādī* ou *fīrūzābādī* (m. 1415);
- *tāğ al='arūs min ġawāhir al=qāmūs* 'la couronne de la mariée [faite] des bijoux de *al=qāmūs*' de *az=zabīdī* (m. 1791).

***tāğ al=luğah wa-şaḥāḥ al='arabiyyah* de *al=ğawharī* (m. vers 400/1010)**

Communément appelé *aş=şaḥāḥ* ou *aş=şiḥāḥ* 'le valide', le lexique *tāğ al=luğah wa-şaḥāḥ al='arabiyyah* a été composé au X^e siècle par *al=ğawharī* (m. 400/1010). Dans son introduction, l'auteur donne une explication au titre de son lexique et affirme qu'il sera limité strictement aux formes d'un parler idéal, confronté et vérifié auprès des bédouins arabes. Selon cette affirmation, *al=ğawharī* est l'un des rares philologues qui ont recensé leurs données, directement des bédouins arabes à cette période tardive. Toutefois, bien qu'il s'est limité uniquement au matériel correct, il enregistre aussi les emprunts étrangers.¹⁴²

L'approche d'éliminer tout ce qu'il ne pouvait pas établir, personnellement, a eu, comme conséquence, la réduction considérable de son lexique. En effet, *aş=şaḥāḥ* contient seulement 5618 racines comparées aux 9273 racines dans *lisān al='arab* et 11978 dans *tāğ al='arūs*.¹⁴³

Il est constitué de vingt-huit *'abwāb*, pluriel de *bāb* 'section', classés selon le dernier radical, de la *hamzah* à la *yā*. Donc une section donnée contient les racines qui finissent par le même radical. Chaque section est divisée en *fuṣūl*, pluriel *faṣl* 'chapitre', classés selon le premier radical. Un chapitre donné contient donc les racines commençant par le même radical et se terminant par le même radical. Une racine donnée est reportée à l'intérieur du *bāb* 'section' de son dernier radical, dans le *faṣl* 'chapitre' de

141. Baalbaki, 2014, pp. 363-366

142. Naşşār, 1988, V. II, pp. 391-392.

143. Baalbaki, 2014, p. 377.

son premier radical et, ordonnée en suivant l'ordre alphabétique interne du chapitre en fonction du deuxième radical. Par exemple, les deux verbes *kataba* 'écrire' et *kaḍaba* 'mentir' dérivent, respectivement, de la racine *ktb* et de la racine *kḍb*:

- finissant par la même lettre, les deux racines se trouvent dans la même section : *bāb al=bā* 'section de la lettre *b*' ;
- commençant par la même lettre elles se trouvent dans le même chapitre : *faṣl al=kāf* 'chapitre de la lettre *k*' ;
- à l'intérieur du chapitre, les deux racines sont ordonnées selon le deuxième radicale, donc les paroles dérivant de la racine *ktb* sont répertoriées avant celles qui dérivent de la racine *kḍb*.

al-'ubāb az=zāḥir wa=l=lubāb al=fāḥir de aṣ=ṣaḡānī (m. 650/1252)

Dans la tradition lexicographique, *al-'ubāb* est considéré l'un des lexiques arabes les plus complets.¹⁴⁴ En effet, l'objectif de *aṣ-ṣaḡānī* était de faire la collecte des données éparses dans divers ressources, tout en gardant un œil critique sur ce qui est correct et ce qui ne l'est pas, que ça soit un lemme, une définition ou un *ṣāhid*. De conséquence, il a enrichi et a porté d'énormes extensions au contenu de son matériel. Une autre caractéristique de *al-'ubāb* est le fait que la plupart des racines contiennent aussi leur sens fondamentale, suivant les *maqāyīs* de *ibn fāris*. Par exemple, le sens fondamentale de la racine *bk'* est 'la diminution de la chose et sa pénurie'.¹⁴⁵

La plus grande contribution de *aṣ-ṣaḡānī* constitue une étape majeure dans l'arrangement et dans la microstructure des éléments lexicaux. En effet et pour la première fois dans l'histoire de la lexicographie arabe, on note qu'il a classé les schèmes en distinguant entre le verbe de base et les verbes dérivés. Ainsi il a mis d'abord le verbe de base suivi des noms qui en dérivent et, ensuite les verbes dérivés avec les noms qui en dérivent.¹⁴⁶

lisān al='arab de ibn manẓūr (m. 711/1311)

Avec les 9273 racines qu'il contient, *lisān al-'arab* est le deuxième lexique arabe complet, juste après *tāǧ al-'arūs* qui contient 11978 racines. Il peut être décrit comme un dictionnaire encyclopédique qui contient une vaste quantité de matériel lié à divers

144. Baalbaki, 2014, p. 383.

145. Naṣṣār, 1988, V. II, p. 424.

146. Ibid., p. 423

champs, tels que les lectures coraniques, l'exégèse, Ḥadīṭ, plantes, animaux, grammaire, morphologie, mots arabisés, solécisme, noms propres, etc. En plus de cela, il contient un chapitre dédié aux lettres et un autre à la phonétique.¹⁴⁷

Dans son introduction, ibn manzūr déclare que les lexiques, qui l'ont précédé, se sont concentrés sur un des deux objectifs ; *al=ğam* 'recueillir le matériel' ou *al=waḍ* 'ordonner le matériel'. Il déclare que *at=tahdīb* de *al=azharī* (m. 370/981) et *al=muhkam* de *ibn sīda* (m. 458/1066) sont incomparables, pour leur exhaustivité mais ils sont difficiles à utiliser. Par contre, il loue *aş=şahāh* de *al=ğawharī* (m. 400/1010) pour la simplicité de son arrangement mais le critique pour sa concision. Ainsi, il a mûri le but de réunir le matériel de *at=tahdīb* et *al=muhkam*, ordonné selon le classement de *aş=şahāh*. Il a incorporé, également, *at=tanbīh wa=l=īdāh* 'ammā waqa 'a fī ş=şahāh, connu également comme, *al=ḥawāšī* ou *al='amālī* de *ibn barrī* (m. 582/1187), parce qu'il a porté les corrections à certaines erreurs de *al=ğawharī*.¹⁴⁸ En plus, voulant incorporer également le corpus du Ḥadīṭ, il a intégré l'un des plus grands lexiques *mubawwab* du genre, *an=niḥāyah fī ġarīb al=Ḥadīṭ wa=l='aṭar* de *ibn al='aṭīr* (m. 606/1210).

L'auteur, *ibn manzūr*, ne réclame aucune originalité dans son travail, au contraire, il déclare que la responsabilité du matériel, qu'il soit bon ou défectueux, relève de l'auteur original. Les éléments lexicaux, les significations et les *şawāhid*, ont été réunis dans *lisān*, intégralement à partir des cinq ressources déclarées.¹⁴⁹

Pour l'arrangement des racines, les cinq ressources sont représentants des trois modèles utilisés dans la tradition lexicale. Les lexiques *at=tahdīb* et *al=muhkam* suivent la permutation phonétique, *an=niḥāyah fī ġarīb al=Ḥadīṭ* suit le classement alphabétique ; par contre, *aş=şahāh* et *al=ḥawāšī* sont ordonnés selon la rime. Cependant, *ibn manzūr* a choisi le système de rime, dans presque tous les détails, parce qu'il est simple à utiliser, comme il l'a déclaré dans son introduction.¹⁵⁰

***al=qāmūs al=muhīṭ* de *al=fīrūz'abādī* (m. 817/1415)**

Dans son introduction, *al=fīrūz'abādī* (ou *fīrūzābādī*) déclare qu'il a nommé son lexique, *al=qāmūs al=muhīṭ* (désormais *qāmūs*) 'l'océan universel', parce qu'il est la récolte de plusieurs dictionnaires. Il critique *al=ğawharī* parce que, selon lui, *fāta=hu nişfu l=luġati 'aw 'aktaru* 'il a oublié la moitié de la langue ou plus'.¹⁵¹ Pour écrire son lexique, *al=fīrūz'abādī* a pris *aş=şihāh*, de *al=ğawharī*, comme base et l'a enrichi du ma-

147. Baalbaki, 2014, p. 387.

148. Naşşār, 1988, p. 451; Baalbaki, 2014, pp. 378-379; 386.

149. Baalbaki, 2014, p. 386.

150. Ibid., p. 389.

151. Fīrūzābādī, Qāmūs, Introduction, p. 27.

tériel manquant. Pour souligner le fait que son lexique comble les lacunes de *aṣ=ṣihāḥ*, il a marqué, en rouge, ce que *al=ğawharī* avait ignoré et que lui a ajouté, en intégrant le matériel de *al=‘ubāb* de *aṣ=ṣağānī* (m. 650/1252) et de *al=muḥkam* de *ibn sīda* (m. 458/1066).

L’objectif de l’auteur, en rédigeant son lexique, était l’exhaustivité. Dans le but d’avoir un “petit” dictionnaire, il a éliminé la plupart des citations coraniques, poésies, noms des grammairiens et des historiens. Bien que *qāmūs* est simplifié et condensé, il contient aussi les noms propres des personnes, des pays et des régions. On trouve aussi un grand nombre de noms de maladies, d’épées, des oiseaux, des guerres saintes. L’auteur a donné, aussi, une grande importance aux noms des arbres, des végétaux et des herbes médicales, en expliquant leurs particularités.

Il se distingue des autres lexiques par plusieurs caractéristiques qui ont contribué au fait qu’il soit l’un des meilleurs. En effet, pour la première fois dans l’histoire de la lexicographie arabe, on note l’ordre dans la distribution et la microstructure des éléments lexicaux.¹⁵² Ainsi, on note que l’auteur mentionne, généralement, le lemme suivi de tous les sens qui le caractérisent avant de passer à un autre élément, ce qui lui évite de le répéter.¹⁵³

Cependant, la concision de l’auteur a incité *az=zabīdī* (m. 1205/1790), presque quatre siècle après, à écrire son lexique, *tāğ al=‘arūs min ġawāhir al=qāmūs*, pour le compléter et l’expliquer, donnant ainsi naissance au lexique le plus complet de la langue arabe.

***tāğ al=‘arūs min ġawāhir al=qāmūs* de *az=zabīdī* (m. 1205/1790)**

Considéré une véritable œuvre encyclopédique, *tāğ al=‘arūs* est le lexique *muğanas* le plus complet de toute la tradition, suivi de *lisān al=‘arab*. Le tableau 2.2 illustre une comparaison entre les deux lexiques.¹⁵⁴

Lexique	auteur	racine	trilitères	quadrilitères	quadrilitères
<i>lisān</i>	ibn manzūr (m. 711/1311)	9273	6538	2548	187
<i>tāğ</i>	<i>az=zabīdī</i> (m. 1205/1790)	11978	7597	4081	300

Tab. 2.2 : Comparaison entre *tāğ al=‘arūs* et *lisān al=‘arab*

152. Naṣṣār, 1988, p. 463.

153. La discussion en détail du matériel sera faite dans le chapitre suivant.

154. Baalbaki, 2014, p. 401.

Le travail de az=zabīdī est, également, très semblable à celui de ibn manzūr. Comme lui, il visait l'exhaustivité, en fusionnant le matériel d'un grand nombre d'œuvres antérieures, et attribue la responsabilité du matériel, bon ou défectueux, à l'auteur original. Cependant, ibn manzūr a réuni, tout simplement, le matériel de cinq lexiques, tandis que az=zabīdī se base uniquement sur *al=qāmūs* de fīrūzābādī et le complète. Il explique, dans son introduction, que *al=qāmūs* est le meilleur lexique, mais ses explications concises doivent être développées et, dans ce but, il utilise le matériel dérivé de 116 sources qu'il a énuméré.¹⁵⁵

Il cite minutieusement et en entier le texte de *qāmūs* qu'il met entre crochets, et l'interrompt chaque fois qu'il doit être expliqué, développé ou modifié. Principalement, il ajoute ce que fīrūzābādī a négligé par souci de concision, comme les noms des auteurs auxquelles les données sont attribuées, les *šawāhid* et leurs émetteurs. Il corrige les erreurs d'explication, de vocalisation et de *tašhīf*, et souvent, il conteste à fīrūzābādī ce qu'il attribuait comme erreurs à az=zabīdī, l'auteur de *aṣ=ṣaḥāḥ*. La plupart des lemmes se terminent par une section appelée *mustadrak* 'supplément', contenant les racines manquantes dans *al=qāmūs* et les éléments lexicaux que fīrūzābādī n'a pas mentionné.

Un autre lexique que az=zabīdī a rédigé est *at=takmilah wa=d=dayl wa=ṣ=ṣilah li=mā fāta ṣāhib al=qāmūs mina l=luḡah*, où il explique que, tout comme fīrūzābādī avait écrit, dans *al=qāmūs*, en encre rouge le matériel laissé par al=ḡawharī, il a fait de même avec ce que fīrūzābādī avait négligé, avec l'intention de réfuter la croyance répandue que ce dernier couvrait toute la langue.¹⁵⁶

Conclusion En adoptant la même conception, suivie par son propre auteur et par az=zabīdī, on peut considérer *qāmūs* comme une base sur laquelle peuvent être construits les autres lexiques, antérieurs et postérieurs.

Situé à mi-chemin entre l'œuvre puriste, *aṣ=ṣaḥāḥ*, et l'œuvre encyclopédique, *tāğ al='arūs*, le dictionnaire «de poche», *qāmūs*, a suscité et suscite encore l'intérêt de la communauté des lexicographes, en particulier, et des linguistes, en général. En le comparant avec les autres, *qāmūs* est un lexique succinct et condensé, plus «facile» à manipuler pour construire une ressource qui, grâce aux techniques modernes, peut être alimentée, en suivant l'histoire de la lexicographie arabe.

Pour un tel objectif, l'accès au *qāmūs* sous forme électronique est une condition décisive. La version numérique doit être structurée, selon les critères linguistiques et

155. Ibid., p. 397.

156. Ibid., p. 400.

extralinguistiques et en respectant la macrostructure de *qāmūs* et la microstructure de son contenu. L'expérience dont rend compte la troisième partie de la thèse, décrit les caractéristiques lexicographiques de *qāmūs* et les divers phases pour le convertir en une ressource numérisée et structurée.

Quatrième partie

**CRÉATION D'UNE RESSOURCE
SÉMANTIQUE**

al=qāmūs al=muḥīṭ

Chapitre 1

Étude de *al=qāmūs al=muḥīṭ*

L'auteur, firūzābādī¹, est né l'an 726/1326 à kārizīn, comme il le déclare lui-même dans son lexique :²

وَكَارِزِينَ بَلَدٌ بِفَارِسٍ، مِنْهُ مُحَمَّدٌ ابْنُ الْحَسَنِ مُفْرِي الْحَرَامِ، وَبِهِ وُلِدْتُ

”et Kārizīn : pays [qui se trouve] en Perse ; d'où vient Moḥammad bnu I=Ḥasan lecteur de al=Ḥaram (Mecque) ; et où je suis né”

À l'âge de huit ans, il est parti à Šīrāz pour finir ses études et, depuis il a passé une grande partie de sa vie à voyager pour étudier et enseigner. Enfin, il s'est établi à Zabīd, au Yemen, où le roi 'ismā'īl l'a nommé *qāḍī* et où il est mort à Zabīd en l'an 817/1414. Il justifie le titre de son dictionnaire *al=qāmūs al=muḥīṭ* dans l'introduction :³

وَأَسَمَيْتُهُ الْقَامُوسَ الْمَحِيطَ لِأَنَّهُ الْبَحْرُ الْأَعْظَمُ

”et je l'ai appelé *al=qāmūs al=muḥīṭ* parce que c'est l'océan le plus grand”

John Haywood commente le titre et la valeur du *qāmūs* de la manière suivante :⁴

”To Majd al-Din Muhammad ibn Yacqub firūzābādī (726/1326-817/1414) fell the honour of unwittingly coining the modern Arabic word for dictionary, by calling his dictionary “*al-Qāmūs al-Muḥīṭ*” (the Surrounding Ocean). The word “*Qāmūs*”, thanks to the wide currency of the dictionary, and others based on it in several languages, came to mean a dictionary. Whether or not the word is a corruption of the Greek “*Okeanos*”, it occurs in the *Ḥadīṭ*,

1. Le nom complet est 'abū ṭ=ṭāhir maḡdu d=dīn ibn 'ibrāhīm al=firūz' abādī

2. Firūzābādī, *Qāmūs*, p. 522

3. Firūzābādī, *Qāmūs*, p. 27.

4. Haywood, *Arabic Lexicography*, pp. 83-85.

and is mentioned in many of the dictionaries, as meaning the main, middle, or deepest part of the sea.

[...] it can safely be said that no native teacher of Arabic - at any rate at the secondary school level and above - would be without his own copy. Whereas the “*Lisān*” is for the library, the “*Qāmūs*” is for the study and the school.”

Effectivement, le lexique de *fīrūzābādī* est devenu tellement populaire que la parole *qāmūs* a supplanté la parole *mu‘ğam* pour désigner le dictionnaire. Dans *al=mu‘ğam al=wasīṭ*, dictionnaire de langue arabe moderne, on trouve trois sens de *qāmūs*.⁵

- **الْبَحْرُ الْعَظِيمُ** : "la mer immense" ;
- **عَلَى مُعْجَمِ الْفِيْرُوْزَابَادِي** : "titre du dictionnaire de *Fīrūzābādī*" ;
- **كُلُّ مُعْجَمٍ لُغَوِيٍّ، عَلَى اتِّسَاعٍ** : "tout dictionnaire lexicographique, par extension".

1.1 L'introduction de *qāmūs*

La préface commence, d'abord, par les éloges à Allah, Donneur de la langue, qui a envoyé son prophète Moḥammad, le plus éloquent des orateurs. L'auteur ajoute, ensuite, que le devoir des lexicographes et des philologues est de se concentrer et d'étudier la langue arabe, parce que c'est la langue dans laquelle a été écrite la *šarī‘ah* 'recueil de lois islamiques'. Lui-même, dit-il, est expert de lexicographie et a déjà commencé un dictionnaire, supérieur à toute autre œuvre antérieure et intitulé :

الَّلَامِعُ الْمَعْلَمُ الْعُجَابُ الْجَامِعُ بَيْنَ الْمُحْكَمِ وَالْعُبَابِ

"le splendide, le repère, le magnifique, le collecteur de [deux dictionnaires] *al=muḥkam* et *al=‘ubāb*".

Cependant, il a estimé qu'il serait composé en soixante volumes et donc il serait inaccessible aux étudiants.

5. *al=mu‘ğam al=wasīṭ*, 2004, p. 758.

Comme il a été sollicité à composer un lexique avec la même qualité mais de petit volume, il a produit le présent *qāmūs*. Pour que le travail soit complet et conçu dans la brièveté et la précision, il a pensé à éliminer les *šawāhid* et le superflus. Ainsi, il a pu résumer trente volumes de sa grande œuvre en un volume du présent dictionnaire. Mais, assure-t-il, il a inclus l'ensemble du contenu, sous forme plus brève, des deux sources principales, *al=muḥkam* et *al='ubāb*, qu'il juge les meilleurs.

L'auteur fait remarquer que, malgré la popularité bien méritée du *aṣ=šaḥāḥ*, son auteur a omis la moitié, ou plus, de la langue, soit en omettant complètement les racines, soit en négligeant de mentionner les significations rares. Pour faire valoir la supériorité de son *qāmūs*, il a décidé de prendre *aṣ=šaḥāḥ* comme base autour de laquelle il a construit son lexique en écrivant, en encre noire, ce qui existait déjà dans *aṣ=šaḥāḥ* et, en rouge, ce qui manquait dans *aṣ=šaḥāḥ* et qu'il a ajouté.

En comparant les deux dictionnaires, *az=zabīdī* fait noter que, effectivement, *qāmūs* comprend 20 000 *māddah* "unités lexicales" qui ne figurent pas dans *aṣ=šaḥāḥ*.⁶ Dans son introduction, *fīrūzābādī* fait les éloges de son propre lexique en exaltant plusieurs aspects.

1.1.1 La richesse des unités lexicales

L'auteur déclare que son lexique est la récolte de deux milles dictionnaires, bien qu'il n'en mentionne aucun, à part les trois déjà cités, *al=muḥkam*, *al='ubāb* et *aṣ=šaḥāḥ*. En fait, une comparaison entre le matériel du *qāmūs* et les deux lexiques, *al=muḥkam*, et *al='ubāb*, démontre que ce sont les deux principales ressources qu'il a utilisé.⁷

Les ajouts des nouveaux éléments lexicaux ou des nouvelles significations sont sporadiques, dont la plupart sont des noms propres de tout type (Compagnons du Prophète, savants de *Ḥadīṭ*, poètes, noms de lieux et plantes ...). On estime, par exemple, qu'ils sont trois fois plus fréquents dans *qāmūs* que dans *lisān* et vingt fois plus que dans *aṣ=šaḥāḥ*.⁸ En outre, l'auteur a listé un grand nombre des termes médicaux qu'il semble avoir enregistrés au Yémen, où il a vécu les dernières vingt années de sa vie.

6. Baalbaki, 2014, p. 393.

7. Naṣṣār, 1988, pp. 459-465.

8. Baalbaki, 2014, p. 393.

1.1.2 L'exhaustivité associée à la simplification et à la condensation

L'auteur adopte différentes stratégies pour inclure, dans un ouvrage de petite taille, la grande majorité des éléments lexicaux et des significations contenues dans les lexiques qu'il a utilisés. Pour éviter les enchainements des filières citationnelles, il écarte les *šawāhid* et évite de citer les auteurs précédents qu'ils soient lexicographes, grammairiens, poètes ou autres.⁹

Mais, le génie de *fīrūzābādī* réside dans sa capacité de résumer en gardant les parties essentielles des significations et en éliminant les répétitions. Par souci d'économie, il adopte de manière systématique des règles qu'il explique dans sa traduction :

1. il ne mentionne pas le pluriel du nom d'agent (schème $C_1\bar{a}C_2iC_3$) des verbes dont C_2 est *wāw* ou *yā'* lorsque le dit pluriel est régulier et donc de schème $C_1\bar{a}C_3ah$. Par exemple, *qādah* est le pluriel *qā'id* "chef; leader" dérivant du verbe *qāda* "diriger; mener" dont la racine est *qwd*. Par contre, il le mentionne lorsque C_2 se réalise, comme c'est le cas de *ḡawalah* qui est le pluriel de *ḡā'il* "déambulateur" dérivant de *ḡāla* "déambuler; se promener" de la racine *ḡwl*.
2. Il ne mentionne pas la forme du féminin mais il ajoute après le masculin l'indication *wa=hiya bi=hā'* "et elle [est] avec *hā'*".¹⁰
3. Dans le cas où le verbe est de schème $C_1aC_2aC_3a$, il contrôle la voyelle du *muḍāri'* "inaccompli" de trois manières :
 - a) il ne mentionne pas le *muḍāri'*, dans le cas où il est de schème $yaC_1C_2uC_3u$ (comme *kataba-yaktubu* "écrire");
 - b) il mentionne le *muḍāri'* mais sans spécifier son schème dans le cas où ce dernier est $yaC_1C_2iC_3u$ (comme *ḍaraba-yaḍribu* "frapper");
 - c) autrement il mentionne le *muḍāri'* en spécifiant son ou ses schèmes.¹¹
4. L'utilisation des abréviations, dont le tableau 1.1 illustre leurs significations. Dans ce cas, *fīrūzābādī* est considéré un véritable pionnier dans l'histoire de la lexicographie.¹²

9. aš=šidyāq; 1982, p. 130 et Baalbaki, 2014, p. 394.

10. À l'époque le suffixe féminin *tā' marbūṭah* était communément appelé *hā'*.

11. *Fīrūzābādī, Qāmūs*, p. 28.

12. Haywood, *Arabic Lexicography*, p. 86.

abréviation	signification	abréviation	signification
ع	مَوْضِعٌ "lieu"	ج	جَمْعٌ "pluriel"
د	بَلَدٌ "pays"	ج ج	جَمْعُ الْجُمُوعِ "pluriel du pluriel"
ة	قَرْيَةٌ "village"	م	مَعْرُوفٌ "connu"

Tab. 1.1 : Les abréviations utilisées par firūzābādī et leurs significations

1.1.3 Le contrôle des diacritiques

Pour prévenir les erreurs dues au *taṣḥīf*, firūzābādī spécifie le nom de la consonne en précisant qu'elle est *muhmalah* 'sans points', ou *mu'ḡamah* 'avec points'.¹³ Dans les cas plus ambigus, il précise même le nombre des points et leurs positions. Le tableau 1.2 donne quelques indications utilisées par Fīrūzābādī. La combinaison des informations concernant le nombre et la position des points, permet d'identifier sans ambiguïté la lettre concernée.

Par exemple, en précisant مَثْنَاءٌ فَوْقُ "avec deux points en haut", on ne peut se tromper qu'il s'agit de (تَ), la lettre *tā*'. Par contre l'indication مَثْنَاءٌ تَحْتِيَّةٌ "avec deux points en bas", il s'agit, sûrement, de la lettre (يَ), la lettre *yā*'.

Indications générales	مُهْمَلَةٌ	"sans point"
	مُعْجَمَةٌ	"avec point"
Nombres des points	مُوَحَّدَةٌ	"avec un point"
	مَثْنَاءٌ	"avec deux points"
	مُثَلَاثَةٌ	"avec trois points"
Position des points	فَوْقُ	"au dessus [de la lettre]"
	تَحْتِيَّةٌ	"en bas [de la lettre]"

Tab. 1.2 : Exemples d'indications sur le contrôle des diacritiques

13. Pour plus de détails sur les lettres *muhmalah* et *mu'ḡamah* voir le tableau 2.1

1.1.4 Le contrôle vocalique

Les voyelles courtes sont également des diacritiques que l'auteur contrôle. Dans certains cas, il fait suivre l'entrée lexicale d'un exemple plus connu qui lui est identique concernant la **position des voyelles**. Dans d'autres cas, il utilise des indications qui permettent de spécifier la vocalisation de la **parole trilitère**, en donnant même une idée sur sa fréquence.¹⁴ Le tableau 1.3 résume les cas possibles :

- la parole est fréquente, de schème $C_1V_1C_2C_3$: la voyelle V_1 est précisée بِالضَّمِّ "avec la voyelle u" ; بِالْفَتْحِ "avec la voyelle a" ou بِالْكَسْرِ "avec la voyelle i" :
- la parole est certifiée avec les trois voyelles, on dit qu'elle est مُثَلَّثَةٌ "avec les trois voyelles" et son schème peut être $C_1uC_2C_3$ et $C_1aC_2C_3$ et $C_1iC_2C_3$.
- La parole est de schème $C_1aC_2aC_3$, dans ce cas, on dit qu'elle est مُحَرَّكَةٌ ou بِالْتَّحْرِيكِ "avec la voyelle a".

Patrons stylistiques			Nom trilitère
fréquent	moins fréquent	rare	Schème
بِالضَّمِّ	وَيُضَمُّ	وَقَدْ يُضَمُّ	$C_1uC_2C_3$
بِالْفَتْحِ	وَيُفْتَحُ	وَقَدْ يُفْتَحُ	$C_1aC_2C_3$
بِالْكَسْرِ	وَيُكْسَرُ	وَقَدْ يُكْسَرُ	$C_1iC_2C_3$
مُحَرَّكَةٌ	وَيُحَرَّكُ	وَقَدْ يُحَرَّكُ	$C_1aC_2aC_3$
بِالْتَّحْرِيكِ			
مُثَلَّثَةٌ			$C_1uC_2C_3 - C_1aC_2C_3 - C_1iC_2C_3$

Tab. 1.3 : Exemples d'indications pour le contrôle vocalique des noms.

Lorsque le nom est moins fréquent, l'auteur utilise le passif, avec le sens que *certaines vocalisent* avec la voyelle indiquée : وَيُضَمُّ ; وَيُفْتَحُ ; وَيُكْسَرُ ou يُحَرَّكُ. Pour indiquer

14. Naṣṣār, 1988, V. II ; pp. 473-475

la rareté du lemme, il utilise la particule وَقَدْ "peut-être" devant le passif, ce qui exprime le doute : *il se peut que certains vocalisent avec la voyelle indiquée.*

Concernant les verbes trilitère, généralement, il est accompagné d'un verbe plus fréquent et connu qui sert de modèle ou la voyelle médiane est spécifiée selon les indications résumées dans le tableau 1.4.

Patrons stylistiques	Verbe trilitère	
	Voyelle médiane	Schème
بِالضَّمِّ	u	C ₁ aC ₂ uC ₃ a
بِالْفَتْحِ	a	C ₁ aC ₂ aC ₃ a
بِالْكَسْرِ	i	C ₁ aC ₂ iC ₃ a

Tab. 1.4 : Exemples d'indications sur le contrôle vocalique des verbes.

1.2 Macrostructure du *qāmūs*

Le dictionnaire est décrit comme une structure binaire caractérisée par l'imbrication de sa macrostructure et de sa microstructure. Dans un dictionnaire moderne, la macrostructure est l'ensemble des lemmes à définir.

Par contre, dans un dictionnaire appartenant à une époque antérieure, comme c'est le cas de *qāmūs*, la parole à définir, dorénavant, appelée la parole-vedette, peut être le lemme ou une de ses formes fléchies, comme elle peut être accompagnée d'affixes (l'article défini, préposition, conjonction et/ou pronoms).

L'ensemble formé par la parole-vedette, la définition et les éventuelles autres informations qui l'accompagnent constitue une entrée lexicale ou un article.

L'agencement de la parole-vedette, de la définition et des autres informations caractérise ce qu'on appelle la microstructure. Comme son nom l'indique, la microstructure s'occupe de l'organisation interne des unités lexicales.

Cependant, il faut noter que la notion de l'entrée lexicale, dans la tradition lexicographique arabe, est très différente de celle qu'on a l'habitude d'étudier et de trouver dans un dictionnaire d'une langue indo-européenne.

On peut la considérer comme une structure contenant, de manière implicite, des objets méta-textuels et méta-lexicaux, parmi lesquels on peut distinguer la parole-vedette ; la catégorie grammaticales comme le verbe, nom, adjectif, particule ; les formes fléchies,

généralement ceux qui sont irréguliers ; la définition ; les forme dérivés liées à la parole-vedette, les informations sémantiques ou stylistiques.

1.2.1 Organisation et dimensions

Fīrūzābādī a écrit *qāmūs* selon l'ordre de la rime perfectionné par al=ğawharī. De ce fait, *qāmūs* est composé de *'abwāb* (pluriel de *bāb* "section") dont chacune est dédiée à une consonne correspondant à la troisième consonne radicale (C_3). Les sections de *qāmūs* sont vingt-huit (28) dont :

- vingt-six sections sont dédiées aux consonnes *ṣaḥīḥah* "fortes" de la hamzah à la *hā'* ;
- une section dédiée aux deux consonnes *mu'tallah* "faibles" la *wāw* et la *yā'* tout en séparant entre les deux comme il a été annoncé par l'auteur dans l'introduction ;
- une section dédiée à la *'alif* où l'auteur traite des particularités grammaticales d'un certain nombre de lettres et de particules qui se terminent par une *'alif* long.

Chaque section est composée de *fuṣūl* (pluriel de *faṣl* "chapitre"), classés selon la première consonne radicale (C_1). À l'intérieur du même chapitre, les racines sont classées selon la deuxième consonne radicale (C_2).

Le tableau 1.5 montre le nombre des chapitres contenus dans chaque section. On note que, certaines sections sont composées de 28 chapitres. Par contre, d'autres sections sont composées de moins de 28 chapitres, ce qui prouve qu'il y a manque des racines qui résultent de la combinaison de C_3 et de C_1 .

Comme c'est déjà expliqué dans la sous-section 2.5.2, les éléments identiques ne peuvent pas être adjacents, selon le principe OCP (**O**bligatory **C**ontour **P**rinciple). Par conséquence, les consonnes (appelées les autosegments) qui occupent deux positions voisines doivent être différentes.

OCP_{total} pour toute paire d'autosegments adjacents a et b : **a # b** où le symbole # indique que a est différent de b.

Par contre, dans le paragraphe 2.5.2.1.2, on a expliqué que, dans la plupart des cas, la non cooccurrence des consonnes est due aux restrictions imposées par l'OCP-Place, qui interdit la combinaison des consonnes similaires dans une même racine.

Sections		Nombre chapitres existants	Chapitres manquants
ء	<i>hamzah</i>	28	
ب	<i>b</i>	28	
ت	<i>t</i>	28	
ث	<i>ṭ</i>	25	<i>ḏ - s - ṣ</i>
ج	<i>ǧ</i>	28	
ح	<i>ḥ</i>	23	<i>ḥ - ṣ - ' - ḡ - h</i>
خ	<i>ḫ</i>	26	<i>ḥ - ḡ</i>
د	<i>d</i>	27	<i>ṣ</i>
ذ	<i>ḏ</i>	24	<i>ṭ - ḏ - ṣ - y</i>
ر	<i>r</i>	28	
ز	<i>z</i>	24	<i>ṭ - ḏ - ṣ - y</i>
س	<i>s</i>	25	<i>ṭ - z - ṣ</i>
ش	<i>š</i>	25	<i>s - š - ḏ</i>
ص	<i>ṣ</i>	21	<i>ṭ - ḏ - z - s - ḏ - ṭ - ṣ</i>
ض	<i>ḏ</i>	21	<i>ṭ - ḏ - z - s - š - ṭ - ṣ</i>
ط	<i>ṭ</i>	27	<i>t</i>
ظ	<i>ṣ</i>	18	<i>t - ṭ - ḏ - z - s - š - ḏ - ṭ - ṣ - h</i>
ع	'	26	<i>ḥ - ḡ</i>
غ	<i>ḡ</i>	23	<i>ḥ - ḫ - ' - q - y</i>
ف	<i>f</i>	27	<i>m</i>
ق	<i>q</i>	26	<i>ṣ - k</i>
ك	<i>k</i>	26	<i>ṣ - q</i>
ل	<i>l</i>	28	
م	<i>m</i>	28	
ن	<i>n</i>	28	
هـ	<i>h</i>	25	<i>ḥ - ṣ - ḡ</i>
و - ي	<i>w et y</i>	28	
ا	<i>'alif</i>	-	-

Tab. 1.5 : Sections et chapitres du *qāmūs*

En reprenant l'exemple des lettres labiales {b, f, m}, selon les propriétés phonologiques et l'indice de similarité, illustrées dans le tableau 2.6, on s'attend que la cooccurrence de la paire {b, f} soit moins probables que celle des paires {b, m} et {f, m}.¹⁵

Une étude détaillée dans *qāmūs* donne les résultats reportés dans le tableau 1.6. On note, effectivement, qu'il existe seulement deux racines *bwb* et *frb* qui combinent

15. Voir entre autres Frisch et al., 2004 et Pierrehumbert, 1993

les consonnes *b* et *f*, de la paire {*b*, *f*}. En plus, ils sont en positions non adjacentes et les paroles-vedettes sont, généralement, des noms propres (NP). Concernant les paires {*b*, *m*} et {*f*, *m*}, le nombre des racines est plus grand mais, plusieurs d'entre elles sont quadrilitères.

Combinaison	{ <i>b</i> , <i>b</i> }	{ <i>b</i> , <i>f</i> }	{ <i>f</i> , <i>b</i> }	{ <i>f</i> , <i>f</i> }
Nombre de racines	1 (<i>bwb</i>)	-	1 (<i>frb</i>)	-
Nombre de parole-vedette	38	3 NP	10 NP	12

Combinaison	{ <i>m</i> , <i>b</i> }	{ <i>b</i> , <i>m</i> }	{ <i>m</i> , <i>f</i> }	{ <i>f</i> , <i>m</i> }	{ <i>m</i> , <i>m</i> }
Nombre de racines	-	33	-	15	-
Nombre de parole-vedette	3 NP	181	0	127	4

Tab. 1.6 : Cooccurrence des lettres labiales dans *qāmūs*

1.2.2 Critères lexicaux

Pour composer un lexique facile à utiliser, l'auteur a déclaré, dans son introduction, qu'il a l'intention d'éliminer les *šawāhid*. Cependant, on note qu'il en fait recours bien que de manière parcimonieuse. Par exemple, dans la section de la lettre *b*, il cite 11 *šawāhid* extraits du Qur'ān et 6 *šawāhid* poétiques.

Pour une même famille dérivationnelle, autrement dit, les vocables construits sur les consonnes de la même racine, on trouve répertoriés d'abord le verbe trilitère et les noms qui en dérivent, et puis les verbes dérivés et les formes nominales qui en dérivent. Cependant, une analyse plus détaillée montre que l'auteur n'a pas suivi systématiquement ce classement.

Par exemple, la figure 1.1 reporte la famille dérivationnelle de la racine *ḍrb*,¹⁶. En analysant la succession des unités lexicales, on note une confusion dans l'ordre avec lequel l'auteur les a présentées.

Il a traité, d'abord, le verbe trilitère *ḍaraba* et les noms qui en dérivent *ḍarb*, *ḍarīb*, *ḍurayb*, *miḍrab*, et *maḍrab*; suivis du verbe dérivé *iḍṭaraba*; de nouveau le verbe de base *ḍaraba* et d'autres noms qui en dérivent : *ḍarībah*, *ḍariba*, *yaḍribu*, et *maḍrab*.

L'auteur retourne à exposer les verbes dérivés *istaḍraba*; *ḍāraba*; *ḍarraba* et *'aḍraba* entre lesquels sont intercalés une expression idiomatique (proverbe) et une citation coranique.

Il faut noter aussi que, dans le cas où la parole-vedette est polysémique, elle est répétée pour chaque signification. Par exemple, le verbe *ḍaraba* est répété treize fois,

16. *Fīrūzābādī, Qāmūs*, pp. 107-108.

toujours dans un contexte bien défini sous forme d'un exemple court. La polysémie, justement, est à l'origine de la nature complexe et/ou répétitive des entrées lexicales.

ضَرْبَهُ يَضْرِبُهُ، وَضَرْبُهُ، وَهُوَ ضَارِبٌ وَضَرْيَبٌ وَضَرْوَبٌ وَضَرْيَبٌ وَمِضْرَبٌ كَثِيرُهُ، وَمَضْرُوبٌ وَضَرْيَبٌ، وَالْمِضْرَابُ وَالْمِضْرَابُ مَا ضَرَبَ بِهِ. **وَضَرْبَتُ** يَدُهُ، كَكَرْمٍ: جَادَ ضَرْبُهَا. **وَضَرْبَتِ** الطَّيْرُ تَضْرِبُ: دَهَبَتْ تَبْتَعِي الرِّزْقَ. **وَضَرْبَ** عَلَى يَدَيْهِ: أَمْسَكَ. **وَضَرْبَ** فِي الْأَرْضِ ضَرْبًا وَضَرْبَانًا: خَرَجَ تَاجِرًا أَوْ غَازِيًا، أَوْ أَسْرَعَ، أَوْ دَهَبَ. **وَضَرْبَ** بِنَفْسِهِ الْأَرْضَ: أَقَامَ، كَأَضْرَبَ، ضَيْدٌ، **وَضَرْبَ** الْفَحْلُ ضِرَابًا: نَكَحَ. **وَضَرْبَتِ** النَّاقَةَ: شَالَتْ بِذَنْبِهَا **فَضْرِبَتِ** فَرَجَهَا فَشَتَّتْ، وَهِيَ ضَارِبٌ وَضَارِبَةٌ. **وَضَرْبَ** الشَّيْءِ الشَّيْءَ بِالشَّيْءِ: خَلَطَهُ، كَضَرْبَتَهُ. **وَضَرْبَ** فِي الْمَاءِ: سَبَخَ، وَلَذَعَ، وَتَحَرَّكَ، وَطَالَ، وَأَعْرَضَ، وَأَشَارَ. **وَضَرْبَ** الذَّهْرُ بَيْنَنَا: بَعَدَ. **وَضَرْبَ** بِنَفْسِهِ الْأَرْضَ: جَبَنَ، وَخَافَ. **وَضَرْبَ** الزَّمَانَ: مَضَى. **الضَّرْبُ**: الْمِثْلُ، وَالرَّجُلُ الْمَاضِي النَّدْبُ، وَالخَفِيفُ اللَّحْمِ، وَالصَّنْفُ مِنَ الشَّيْءِ، كَالضَّرِيْبِ وَالْمَضْرُوبِ، وَالْمَطْرُ الْخَفِيفُ، وَالْعَسَلُ الْأَبْيَضُ، وَبِالتَّحْرِيكِ أَشْهَرُ. **وَالضَّرْبُ** مِنَ بَيْتِ الشَّعْرِ: آخِرُهُ. **وَالضَّرِيْبُ**: الرَّأْسُ، وَالْمَوْكَلُ بِالْقِدَاحِ أَوْ الَّذِي يَضْرِبُ بِهَا، كَالضَّرَابِ، وَالْفِدْحُ الثَّالِثُ، وَاللَّبَنُ يُحْلَبُ مِنْ عِدَّةٍ لِقَاحٍ فِي إِنْاءٍ، وَالنَّصِيبُ، وَالْبَطِينُ مِنَ النَّاسِ، وَالتَّلْجُ، وَالجَلِيدُ، وَالصَّقِيعُ، وَرَدِيءُ الحَمَضِ أَوْ مَا تَكَسَّرَ مِنْهُ. **وَكَرْبِيرٌ**: ضَرْيَبٌ بِنُ تَقِيرُ فِي ن ق ر. **وَالْمِضْرَابُ**: الْفَسْطَاطُ الْعَظِيمُ. **وَبَفَتْحِ المِيمِ**: الْعَظْمُ الَّذِي فِيهِ الْمُخَّ. **وَالضَّرْبُ**: تَحَرَّكَ، وَمَاجَ، كَتَضْرَبَ، وَطَالَ مَعَ رَخَاوَةٍ، وَاخْتَلَّ، وَاكْتَسَبَ، وَسَأَلَ أَنْ يُضْرَبَ لَهُ. **وَالضَّرْبُ** الْقَوْمُ: ضَارِبُوا، كَتَضَارِبُوا. **وَالضَّرْبُ** خَيْلُهُمْ: اخْتَلَفَتْ كَلِمَتُهُمْ. **وَالضَّرِيْبَةُ**: الطَّبِيعَةُ، وَالسَّنِيفُ وَخَدُهُ، كَالْمَضْرَبِ وَالْمَضْرَبَةِ وَتَكَسَّرَ رَأُوهُمَا، وَالْقِطْعَةُ مِنَ الْقَطَنِ، وَالرَّجُلُ الْمَضْرُوبُ بِالسَّيْفِ، وَوَادٍ يَنْفَعُ فِي ذَاتِ عَرَقٍ، وَوَادَةٌ الضَّرَائِبِ الَّتِي تُؤَخِّدُ فِي الْجَزْيَةِ وَنَحْوِهَا، وَغَلَّةُ الْعَبْدِ. **وَضَرْبٌ**، كَفَرَحَ: ضَرْبَةُ الْبُرْدِ. **وَالضَّرَابُ**: الْمَكَانُ الْمُطْمَئِنُّ بِهِ شَجَرٌ، وَالْقِطْعَةُ الْغَلِيظَةُ تُسْتَطِيلُ فِي السَّهْلِ، وَاللَّيْلُ الْمُظْلِمُ، وَالنَّاقَةُ تُضْرَبُ حَالِيَتِهَا، وَشِبْهُ الرِّخْبَةِ فِي الْوَادِي، ج. ضَوَارِبٌ. وَهُوَ **يَضْرِبُ** الْمَجْدَ: يَكْتَسِبُهُ، وَيَطْلُبُهُ. **وَالضَّرْبُ** الْغَسَلُ: ابْتِيضٌ، وَغَلْظٌ. **وَالضَّرْبُ** النَّاقَةُ: اشْتَهَتْ الْفَحْلَ. **وَضَرَابِيَّةٌ**، كَقَرَّاسِيَّةٍ: كَوْرَةٌ بِمِصْرَ مِنَ الْخَوْفِ. **وَضَارِبٌ** لَهُ: اتَّجَرَ فِي مَالِهِ وَهِيَ الْقِرَاضُ. **وَضَارِبُ السَّلْمِ**: ع بِالْيَمَامَةِ. وَمَا يُعْرَفُ لَهُ "مَضْرَبٌ عَسَلَةٌ" أَي: أَصْلٌ وَلَا قَوْمٌ وَلَا أَبٌ وَلَا شَرَفٌ. **وَضَرْبَانَا** عَلَى أذَانِهِمْ: مَنْعَانَهُمْ أَنْ يَسْمَعُوا. وَجَاءَ **مُضْطَرِبُ** الْعِنَانِ: مُنْهَزِمًا، مُنْفَرِدًا. **وَضَرْبٌ** تَضْرِبِيًّا: تَعَرَّضَ لِلتَّلْجِ، وَشَرِبَ الضَّرِيْبَ. **وَضَرْبَتُ** عَيْنُهُ: غَارَتْ. **وَأَضْرَبَ** الْقَوْمُ: وَقَعَ عَلَيْهِمُ الصَّقِيعُ. **وَأَضْرَبَ** السَّمُومَ الْمَاءَ: انْتَشَفَهُ الْأَرْضَ. **وَأَضْرَبَ** الْخَبْرَ: نَضِجَ. **وَضَارِبُهُ** فَضْرَبَهُ، كَتَضْرَبَهُ: غَلَبَهُ فِي الضَّرْبِ.

Fig. 1.1 : La famille dérivationnelle de la racine ḍrb dans qāmūs

1.2.3 Critères linguistiques

On ne peut s'empêcher de noter l'intérêt de l'auteur pour les noms des plantes ; les maladies et leurs remèdes, comme le démontre l'exemple 1.¹⁷

Exemple 1

أَيُّاقُوتٌ مِنَ الْجَوَاهِرِ: مَ، مُعَرَّبٌ أَجْوَدُهُ الْأَحْمَرُ الرُّمَّانِيُّ نَافِعٌ لِلْوَسْوَاسِ وَضَعِيفِ
الْقَلْبِ شَرْبًا وَجُمُودِ الدَّمِ تَعْلِيْقًا.

"le=rubis [faisant partie] des bijoux ; [c'est] connu, [c'est] un emprunt, [dont] le meilleur est [celui] rouge grenade, bénéfique pour l'inquiétude et la faiblesse du cœur en le buvant et pour la coagulation du sang en le portant accroché."

17. Firūzābādī, Qāmūs, p. 163.

Note Dans cette partie de la thèse, les exemples d'entrées lexicales, prises de *qāmūs*, sont accompagnées de leurs traductions. Quand c'est nécessaire, on a appliqué certaines règles : **1)** les conventions suivies sont celles données dans «The Leipzig Glossing Rules »¹⁸ afin de donner une dimension à la complexité du traitement de la langue arabe ; **2)** la parole-vedette est écrite en gras et en italique et certaines fois, elle n'est pas traduite ; **3)** pour rendre la traduction plus fluide et compréhensible, on a ajouté du texte entre crochets pour souligner qu'il n'existe pas dans le texte arabe.

Également, on ne peut s'empêcher de remarquer que *qāmūs* abonde des noms propres et des toponymes.¹⁹ La recension de la section de la lettre *b*, illustrée dans le tableau 1.7, révèle 282 noms propres de personnes ou de tribus et 520 de noms géographiques.

Total unités lexicales de la sections <i>bā'</i>		5064
Noms propres	personne	246
	groupe - tribu	36
Toponymes	montagne - ville	39
	château - citadelle - maison	18
	Pays - région - ville - village	380
	Rivières - puits	83

Tab. 1.7 : Noms propres et toponymes dans la sections *bā'*

1.2.4 Critères idéologiques

Comme déjà annoncé plus haut, *fīrūzābādī* a construit son lexique en prenant comme base, *aṣ=ṣaḥāḥ* écrit par *al=ḡawharī*. Ce dernier croyait dans les valeurs de précision et de rigueur, par conséquence, *aṣ=ṣaḥāḥ* est un dictionnaire d'un arabe idéal qui ne contient que des lemmes certifiées par les *ṣawāhid*.

Par contre, selon le contenu de son lexique, la position de *fīrūzābādī* n'était pas puriste et *qāmūs* abonde de solécismes, des paroles arabisées, des emprunts et des variantes dialectales. Bien qu'on lui doit le fait que son lexique donne une idée plus exacte de la langue réelle, on ne peut s'empêcher de lui reprocher le fait qu'il ne différenciait pas la forme classique des autres.²⁰

18. Leipzig Glossing Rules, 2015.

19. *aṣ=ṣidyāq*; 1982, p. 349

20. *aṣ=ṣidyāq*; 1982, p. 131-132.

Pour remédier à ce manque de précision, az=zabīdī a construit son lexique *tāğ* *al='arūs min ġawāhir al=qāmūs* (dorénavant *tāğ*) sur la base du *qāmūs*. La figure 1.2 montre l'exemple d'une entrée lexicale dans *tāğ* et permet de mettre en évidence ce qui est manquant dans *qāmūs*. L'auteur de *tāğ* citait le texte existant dans *qāmūs* entre parenthèse, et l'enrichissait en citant d'un côté l'origine de la variante lexicale et de l'autre le lexique où elle a été mentionnée et son auteur.²¹

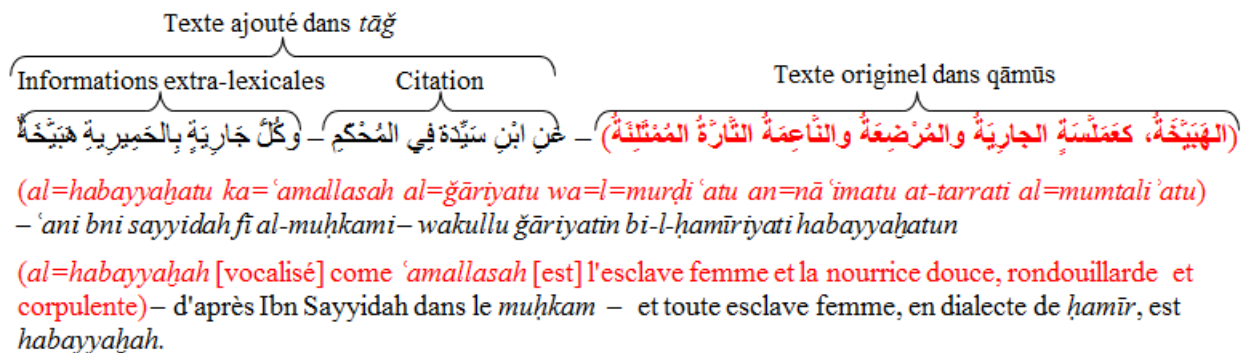


Fig. 1.2 : Exemple d'entrée lexicale dans *tāğ*

1.3 Microstructure du *qāmūs*

1.3.1 Microstructure des entrées lexicales nominales

Les exemples 2 et 3, dans les figures 1.3 et 1.4 ci-dessous, illustrent des entrées lexicales dont la parole-vedette est un nom. On note que la microstructure est composée de blocs dont chacun fournit des informations précises. La composition des entrées n'est pas uniforme parce qu'ils ne sont pas composés systématiquement de tous les blocs d'informations. Cependant, au maximum, on peut identifier les cinq blocs suivants :

1. La parole-vedette, étant un nom, est introduite par l'article, *ad=ḡanb* "le péché" dans l'exemple 2 et *al='azab* "le célibataire" dans l'exemple 3 ;
2. Le contrôle de la vocalisation suit directement la parole à contrôler. Dans l'exemple 3, la parole-vedette est suivie par la parole *muḥarrakatan* ", autrement dit, la première et la deuxième consonne radicale sont vocalisées avec voyelle *a*".

21. Zabīdī, *Tāğ al-'arūs*, Vol. VII ; pp. 367-368.

3. Les informations lexicales fournissent la ou les définitions. Dans l'exemple 2, la définition est un synonyme (*al= 'itm* "le péché"). Par contre dans l'exemple 3, la définition est une phrase libre (*man lā 'ahla lahu* "celui qui n'a pas de famille")
4. Les informations flexionnelles, concernant éventuellement le féminin précédé par *wa=hiya* "et elle" et/ou la/les formes du pluriel brisé annoncées par *ǧ*, abréviation de *ǧam* "pluriel", et/ou *ǧǧ*, abréviation de *ǧam al=ǧumū* "pluriel du pluriel".
5. Les variations lexicales introduites par la conjonction *ka* "comme". Dans l'exemple 3, le mot vedette *'azab* est **comme** *mi'zāb* et *'azīb* et peuvent être considérés variantes fréquentes, par contre *'a'zab* est une variante moins fréquente.

Exemple 2

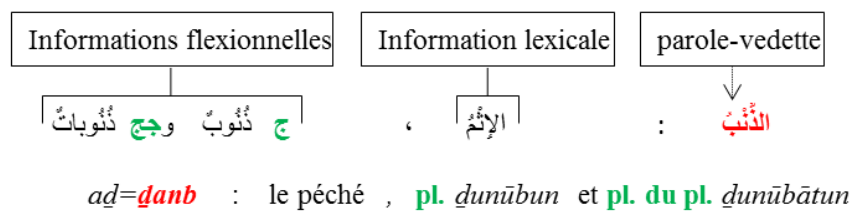


Fig. 1.3 : Exemple 2 de microstructure d'entrées lexicales nominales

Exemple 3

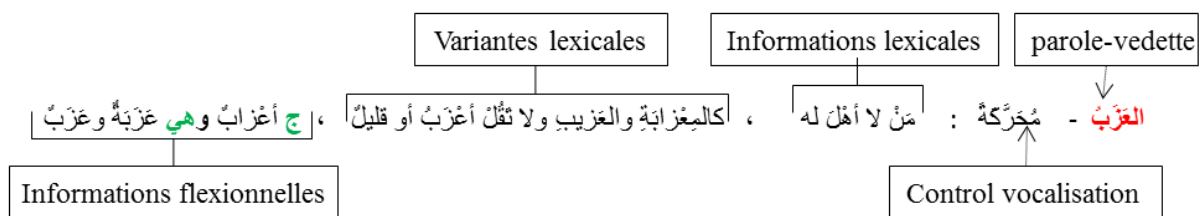


Fig. 1.4 : Exemple 3 de microstructure d'entrées lexicales nominales

Les cinq blocs identifiés ne suivent pas un ordre systématique et la présentation et le style ne sont pas uniformes. La parole-vedette peut être incluse dans un exemple syntagmatique qui la spécialise, et la définition, aussi, est un syntagme, comme c'est le cas de l'exemple 4.

Exemple 4

حَوَالِبُ الْبُئْرِ وَالْعَيْنِ: مَنَابِعُ مَائِهَا

”*ḥawālibu* du puits et de la fontaine : les sources de leurs eaux”

Dans d'autres cas, la parole-vedette peut manquer. Dans l'exemple 5, la première entrée lexicale illustre la parole-vedette *al=muḥqif* "le suivant". Par contre, dans la deuxième entrée, la parole-vedette manque. L'auteur commence directement avec le contrôle vocalique, *bi=fathī al-qāfi* "en vocalisant la lettre *qāf* avec la voyelle a".

Exemple 5

الْمُخَقِبُ : الْمُرْدُفُ
بِفَتْحِ الْقَافِ : التَّغْلِبُ

”*al=muḥqif* : le suivant”

”en vocalisant la lettre *qāf* avec la voyelle a : le renard”

Dans ce cas, l'auteur entend dire que la parole-vedette, en question, a le même schème de celle qui la précède, *al=muḥqif* "le suivant". Mais, il faut changer la voyelle de la lettre *q* en *a*, ainsi on obtient *al=muḥqaf* "le renard".

En commençant une entrée lexicale directement avec le contrôle vocalique, signifie qu'il *suffit* de récupérer la parole-vedette qui précède et faire le changement indiqué pour obtenir la parole-vedette en question.

Enfin, on ne peut s'empêcher de noter, également, la présence d'informations métalexicales contenant des traits polémiques contre *al=ğawharī* dont le but de diminuer la position scientifique de son dictionnaire *aş=şahāḥ* et justifier la position et le choix de *fīrūzābādī*. Ainsi on trouve assez souvent des énoncés comme *wahima* ou *ğalaṭa al=ğawharī* dans le sens que ce dernier "croyait ou s'imaginait ou se trempait".

1.3.2 Microstructure des entrées lexicales verbales

Quand il s'agit des verbes, la parole-vedette est colloquée à l'intérieur d'une phrase simple, généralement, forgée par l'auteur.

Le choix des exemples simples permet de déduire les propriétés morphosyntaxiques, la catégorie grammaticale et les informations syntaxiques comme la transitivité ou l'intransitivité.

Les exemples 6 et 7 représentent des entrées lexicales illustrant certaines significations du verbe *ḥaraba* qui est polysémique.

Exemple 6

خَرَبَ فُلَانٌ : صَارَ لَيْصًا

"quelqu'un **ḥaraba** : il est devenu un voleur."

Exemple 7

خَرَبَهُ : صَرَبَ خُرْبَتَهُ، وَثَقَبَهُ، أَوْ شَقَّهُ

"**ḥaraba=hu** : il a frappé sa **ḥurbah**, et il l'a troué, et il l'a fendillé."

Le tableau 1.8 confronte l'analyse morphosyntaxique des paroles-vedettes, *ḥarab-a* et *ḥarab-a=hu*, des exemples 6 et 7.

	Exemple 6	Exemple 7
Parole vedette :	ḥarab-a	ḥarab-a=hu
Analyse morphosyntaxique :	<i>ḥarab</i> .PFV -a .3MS	<i>ḥarab</i> .PFV -a .3MS = <i>hu</i> M.ACC

Tab. 1.8 : Analyse morphosyntaxique des paroles-vedettes des exemples 6 et 7

Dans l'exemple 6, le verbe *ḥaraba* est intransitif avec le sens de "devenir voleur".

Par contre, dans l'exemple 7, le verbe *ḥaraba* se présente suffixé au pronom accusatif *hu*, et donc il est transitif avec les sens de "frapper sa *ḥurbah*" (la *ḥurbah* est le sommet de la hanche) ; "le trouer" et "le fendiller".

Une complication, concernant l'exemple 7, est l'interprétation des sens en fonction du trait [± humain] du complément d'objet, comme démontré dans le tableau 1.9.

Parole-vedette	complément d'objet	le sens du verbe ḥaraba
	<i>hu</i> M.ACC se réfère à	
ḥaraba=hu	<i>šaḥṣ</i> "une personne"	"frapper le sommet de la hanche" [de la personne en question]
	<i>šay</i> ' "quelque chose"	"trouer/fendiller" [quelque chose]

Tab. 1.9 : Sens du verbe *ḥaraba* selon la nature du complément d'objet.

Dans le cas où le complément d'objet est humain, la parole-vedette *ḥaraba=hu* se réfère au verbe *ḥaraba* avec le sens de "frapper" la *ḥurbah* qui est "le sommet de la hanche" [de la personne qui est complément d'objet].

En supposant que le deuxième et le troisième sens se réfèrent toujours à la *ḥurbah* "sommet de la hanche" qui est un nom féminin, le pronom accusatif aurait dû être féminin : *taqaba=hā* "la trouer" et *šaqqā=hā* "la fendiller". Or, le pronom accusatif est masculin *hu*, on peut conclure qu'il se réfère à un complément d'objet masculin, *šay*'

- d'autres informations flexionnelles concernant, cette fois, les pluriels brisés de l'adjectif introduits par *ğ*, abréviation de *ğam* ' "pluriel".

Comme dans le cas des entrées nominales, cette microstructure n'est pas systématique et peut être très variable. Les exemples 9, 10 et 11 illustrent la variabilité de l'ordre dans la microstructure des entrées verbales.

Dans l'exemple 9, de la figure 1.6, la parole-vedette est le verbe *kabura* "il a grandi" suivi de :

- *maşdar* à l'accusatif, *kibaran*, *kabran* et *kabāratan*;
- l'information lexicale sous forme d'un verbe antonyme, *naqīḍ ṣağura* "contraire de *il est petit*";
- les adjectifs introduits par *fa=huwa* "donc il est";
- les informations flexionnelles concernant l'adjectif, le féminin introduit par *wa=hiya* "et elle est" et les pluriels introduits par *ğ* "pl.".

Exemple 9

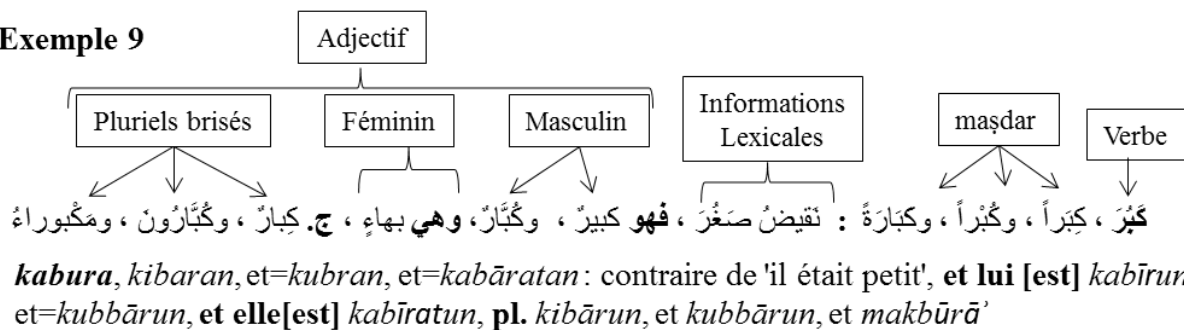
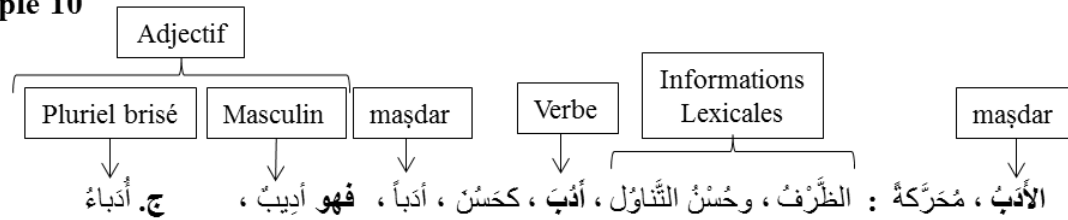


Fig. 1.6 : Entrée lexicale *kabura*

Dans l'exemple 10, dans la figure 1.7, l'entrée lexicale est le *maşdar*, *al='adab*, suivi par :

- les informations lexicales "la politesse et la bonne conduite" ;
- du verbe *'aduba* "il est poli" ;
- son *maşdar* à l'accusatif *'adaban* ;
- l'adjectif *'adīb* "poli" et son pluriel brisé *'udabā'*.

Exemple 10



al=ʿadabu, avec la voyelle a: la politesse, et la bonne conduite, *ʿaduba*, comme *ḥasuna*, *ʿadaban*, et lui [est] *ʿadībun*, pl. *ʿudabāʾ*

Fig. 1.7 : Entrée lexicale : *al=ʿadab*

Par contre, dans l'exemple 11, dans la figure 1.8, la parole-vedette est l'adjectif, *al=ḡašīb*, suivi par :

- les informations lexicales "rude grossier laid en toute chose, et mal éduqué dans le manger";
- le verbe correspondant *ḡašuba* "il est rude; il est grossier; il est laid" qui est introduit par la particule *qaḍ* qui donne le sens de la réalisation;
- le *maṣḍar* à l'accusatif *ḡuṣūbatan*.

Exemple 11



al=ḡašīb: rude grossier laid en toute chose, et mal éduqué dans le manger, *waqaḍ ḡašuba*, comme *karuma*, *ḡuṣūbatan*

Fig. 1.8 : Entrée lexicale : *al=ḡašīb*

1.3.3 Nature des définitions

L'objectif de réaliser un lexique, caractérisé par l'exhaustivité et la brièveté, a poussé *fīrūzābādī* à suivre des principes économiques qu'on retrouve également dans la nature courte des différents type de définitions qu'il a adoptées.

1.3.3.1 La définition dérivative

C'est le cas des exemples 9, 10 et 11, cités supra. L'auteur donne la définition de la parole-vedette que ça soit le verbe, le *maṣḍar* ou l'adjectif et se contente juste de mentionner les autres paroles qui ont une relation dérivationnelle avec elle sans les définir. Dans ce cas, on se base sur les connaissances morphosémantiques qui existent entre les paroles ayant des relations dérivationnelles entre elles.

1.3.3.2 La définition synonymique

Elle est constituée par des synonymes ayant la même catégorie grammaticale que la parole-vedette. Dans l'exemple 10, dans la figure 1.7, le *maṣḍar* al='adab est défini par un autre *maṣḍar* qui lui est synonyme, az=zarf "politesse". Dans l'exemple 11, dans la figure 1.8, la définition de l'adjectif al=ğašīb est constituée des adjectifs synonymes : al=ħašin "rude" ; al=ğalīz "grossier" ; al=bašī "laid" et as=sayyi "mal éduqué".

1.3.3.3 La définition antonymique

La parole vedette est défini en utilisant un contraire introduit par des opérateurs comme naqīdu, 'aksu ou diddu, qui ont tous le sens "contraire de".

Très important à noter que l'opérateur didd "contraire" est utilisé pour signaler la relation antonymique entre la parole-vedette et sa définition et, dans ce cas, il se trouve entre les deux. C'est le cas de l'exemple 9 où le verbe kabura "il est grand" est défini comme naqīdu ṣağura "contraire de il est petit".

Dans l'exemple 12, également, la parole-vedette al=ħaṭ'u "l'erreur", est diddu aṣ=ṣawābi "le contraire de l'exactitude". Il faut noter que la parole-vedette al=ħaṭ'u "l'erreur" est accompagnée de deux variantes al=ħaṭa'u et al=ħaṭā'u.

Exemple 12

الْخَطَأُ وَالْخَطَأُ وَالْخَطَأُ : ضِدُّ الصَّوَابِ

"le=ħaṭ'u et=le=ħaṭa'u et=le=ħaṭā'u : le contraire de l'exactitude"

Par contre, dans l'exemple 13, on note que l'opérateur diddun "contraire", ne se trouve pas mis entre la parole-vedette et le/les sens, mais à la fin de la définition entière.

Exemple 13

السَّاقِبُ : الْقَرِيبُ وَالْبَعِيدُ ضِدُّ

"le=sāqibu : le proche et le loin ; contraire.

Dans ce cas, la parole-vedette *as=sāqibu* désigne, en même temps, *al=qarību* "le prochain" et *al=ba'īdu* "le lointain". C'est le cas de l'énantiosémie où la parole-vedette désigne deux significations contraires entre elles. L'opérateur, **ḍiddun** "contraire", est mis à la fin pour signaler que la relation d'antonymie est justement entre les deux sens.

1.3.3.4 La définition hyperonymique

La parole-vedette est définie par un terme prototypique plus générale. De conséquence, entre la parole définie et sa définition existe une relation du genre et d'espèce, où le défini est l'hyponyme et la définition est l'hyperonyme. Dans les exemples 14 et 15, *al=uḥṭubān* et *al=andalīb* 'le rossignol' sont définis comme *ṭā'ir* "un oiseau".

Exemple 14

وَأُحْطَبَانُ : طَائِرٌ

"*al=uḥṭubānu* : [un] oiseau"

Exemple 15

الْعَنْدَلِيبُ : طَائِرٌ يُقَالُ لَهُ الْهَزَاؤُ يُصَوِّتُ أَلْوَانًا

"*al=andalību* : [un] oiseau, on l'appelle [aussi] le *hazāz*, il pousse divers cris"

Dans les exemples 16 et 17, *al=tabiyy* et *al=ḡanīb* sont définis comme *tamr* "[des] dattes".

Exemple 16

الْتَّبِيُّ وَيُكْسَرُ : تَمْرٌ كَالشَّهْرِيزِ

"le *tabiyy* et certains disent *tibiyy* : [un genre de] dattes comme le *šīhrīz*".²³

Exemple 17

الْحَنِيبُ : تَمْرٌ جَيِّدٌ

"*al=ḡanību* : [un genre de] dattes excellent"

Les exemples du tableau 1.10 permettent de comparer les caractéristiques morpho-syntaxiques de la parole-vedette et du sens qui consentent de distinguer entre une définition synonymique et celle hyperonymique/hyponymique.

- Dans le cas où le cas où la parole-vedette, concernant le lemme X, et le sens Y sont tous définis et nominatifs, donc ils commencent par l'article défini *al=* et

23. *šīhrīz* est un genre de dattes. Zabīdī, Tāḡ al-'arūs, p. 58.

Parole-vedette	Relation	Sens	Relation
<i>al=ǧašīb-u</i>	[est]	<i>al=ḥašīn-u</i> "le=rude"	synonymie
<i>al=ǧanīb-u</i>	[est un genre de]	<i>tamr-un</i> dattes	hyponymie
<i>al=ʿandalīb-u</i>	[est un]	<i>ṭāʿir-un</i> oiseau	

Tab. 1.10 : Distinction entre définition synonymique et antonymique

finissent par la voyelle *u*, on peut conclure que : X [est] Y, en d'autres termes, X et Y sont synonymes.

- Dans le cas où la parole-vedette, concernant le lemme X, est définie au nominatif, donc accompagnée de l'article *al=* et se termine par la voyelle *u*, par contre, le sens Y est nominatif indéfini finissant par la désinence *un*, on peut conclure que : X [est un genre de] Y ou X [est un] Y, donc X est un hyponyme de Y.

1.4 Conclusions

De manière générale, les dictionnaires arabes, appartenant à l'époque médiévale, sont caractérisés par une grande densité d'information où plusieurs plans d'analyse se mêlent et se recourent. Ceci explique la difficulté de séparer facilement entre entrée lexicale, définition et citations, et d'établir une structure.

Cependant, la structuration de *qāmūs*, bien qu'elle soit très complexe, est suffisamment récurrente et récursive. Grâce à des exemples simples, l'auteur montre les collocations et les environnements syntaxiques qui permettent d'identifier la catégorie grammaticale et les propriétés syntaxiques. Mais les indications restent implicites et doivent être analysées et, des fois, déduites afin de les rendre explicites.

Afin d'obtenir un format exploitable pour le traitement automatique, une phase de révision et de segmentation du texte a été nécessaire. Cette phase est longue et fastidieuse mais elle est fondamentale car elle permet, à la fin, d'économiser le temps pour effectuer les tâches successives de conversion, d'analyse et d'annotation à divers niveaux morphologique, syntaxique et sémantique.

Chapitre 2

Acquisition d'une ressource numérique de *qāmūs*

2.1 Procédures d'acquisition et de structuration

2.1.1 Acquisition et pré-traitement du texte

Comme première étape, on a téléchargé une version numérique de *qāmūs*, en format .doc, dont la figure 2.1 en illustre un fragment.



Fig. 2.1 : Fragment de *qāmūs* en format .doc

Cependant, contenant des erreurs, il a été nécessaire la révision et la correction de

l'œuvre entière selon une édition de référence.¹

Le texte en format .doc a l'avantage de visualiser les formatages (fonte, mise en gras, etc.), et un lecteur humain peut en déduire les différentes parties. Ainsi, on distingue le titre de la section *bāb al=bā* "section de la lettre b" ; le chapitre *faṣl al=hamzah* "chapitre de la lettre hamza" et chaque point gras représente le début d'une famille dérivationnelle, en d'autres termes, la liste de tous les vocables dérivant d'une même racine. Cependant, pour un traitement automatique, étant une propriété de la Microsoft Word, il ne peut être exploitable par d'autres logiciels.

De conséquence, il est nécessaire de convertir le texte en format .TXT, avec encodage Unicode (UTF-8) qui couvre tous les caractères de la langue arabe. Le choix du format .TXT est dû au fait qu'il a l'avantage d'une haute portabilité et peut être accessible, indifféremment, à tous les logiciels. Cette phase de conversion est appelée encodage de niveau zéro. La figure 2.2 illustre le texte de la figure 2.1 converti en format .TXT. On note que le texte perd toutes les méta-informations rendues par le formatage et la division en sections, chapitres, etc, parce que le format .TXT ne prévoit pas l'encodage des informations macro-structurelles. Pour ce fait, il est appelé *format de texte brut*.² De conséquence, le texte, appelé «*document plat*», est une version très partielle qui se présente sous forme linéaire, en d'autres termes, une simple séquence de caractères où toutes les informations sont au même niveau.

باب الباء
فصل الهمزة
الأب: الكلاء، أو المرعى، أو ما أنبتت الأرض، والخضر، و د باليمن، وبالكسر: باليمن وأب للسير ينب ويؤب أباً وأبيبا وأياباً وأتاباً وأتابية: ثهيأ، كائتب، و- إلى وطنه أباً وإتابية وأتابية: اشتاق، و- يذة إلى سيقه: رذها ليشله. وهو في أبيه: في جهازه. وأب آية: قصد فصدده. وأبت أبابته ويكسر استقامت طريقته والأبواب الماء والسراب وبالضم معظم السيل وأب: هزم بخملة لا مكتوبة فيها، و- الشيء: حرّكه. وآبة: اسم، وبه سميت آبة العليا والسفلى: قرأتان بلحج، وبالضم: د باقرية. وأبت: صاح. وثابت به: تعجب، وتبحج. وأبى، كحى: نهر بين الكوفة وقصر بني مقاتل، يُنسب إلى أبي بن الصامغان من ملوك النبط، ونهر بواسط العراق، ويتر بالمدينة، أو هي أنا بالنون مخففة، كهنا.
الإتب، بالكسر، والمثبته، كمنسنة: برذ يشق، فتلبيسه المرأة من غير جيب ولا كمين، والبقيرة، ودرع المرأة، وما قصر من الثياب فنصفت الساق، أو سزاويل بلا رجلين، أو قميص بلا كمين، ج: أتاب وإتاب وأتوب. وأتب الثوب ثابياً: صبر إتاباً وثابت به وأتب لسنه وأتبه إياه ثابياً: ألينته إياه. وأتب الشعير، بالكسر: قسره. والثائب: الاستعداد، والتصلب، وأن تجعل جمال الفوس في صدرك، وتخرج منكبتك منها. ورجل مؤتب الظفر، كمنظم: موعجة.
المثب، كمنبر: المشمل، والأرض السهلة، والجذول، وما ارتفع من الأرض، والمائب جمع، و ع أو جبل كان فيه صدقاته صلى الله عليه وسلم. والأتب، محركة: شجر، مخفف الأتاب.
الأذب، محركة: الظرف، وحسن التناول، أذب، كحسن، أذبا فهو أذيب، ج: أذباء. وأذبه: علمه، فتأذب واستأذب. والأذبة، بالضم، والمادبة والمادبة: طعام صنع لدعوة أو عرس. وأذب البلاد إيداباً: ملاًها عدلاً. والأذب، بالفتح: العجب، كالأذبة بالضم، ومصدر: أذبه يأذبه: دعاه إلى طعامه، كاذبه إيداباً، وأذب يأذب أذبا، محركة: عمل مادبة (وأذبة). وأذب البحر: كثرة ما به. وأذبي، كعربي: جبل.

Les deux premières paroles-vedettes, *al'abbu* et *wa'abba*, ont été encadrées pour faciliter leur visualisation.

Fig. 2.2 : Fragment *qāmūs* en format .TXT

Un autre aspect, plus complexe à considérer, est la structure linguistique qui détient, sous forme sous-jacente et implicite, les informations morphologiques, syntaxiques et

1. Fīrūzābādī, Qāmūs.
2. Lenci et al, 2012, p. 66..

sémantiques. En analysant les exemples encadrés dans la figure 2.2, le lecteur humain peut déduire que :

- la première parole encadrée, *أَبُّ* *al='abbu*, i) est la parole-vedette, ii) est un nom parce qu'elle est précédée de l'article définitif, iii) le lemme en question est *'abb*;
- la deuxième parole encadrée, *وَأَبُّ* *wa='abba*, i) est la parole-vedette, ii) elle peut être segmentée en *wa* qui est une conjonction "et" et le verbe *'abba*, iii) le lemme en question est *'abba*;
- les deux lemmes *'abb* et *'abba* dérivent de *'bb*, première racine listée dans *bāb al=bā'* "section de la lettre *b*" correspondant à la troisième consonne radicale ($C_3 = b$); *faṣl al=hamzah* "chapitre de la lettre *hamza*", correspondant à la première consonne radicale ($C_1 = hamza$). Et enfin, le lecteur humain peut déduire, des paroles de la famille dérivationnelle, que la deuxième consonne radicale est, également, la lettre *b* ($C_2 = b$).

Toutes ces informations représentent la clé pour accéder au contenu du texte. Afin de les rendre accessibles au logiciel et disponibles pour d'autres recherches et études, il est nécessaire qu'elles soient explicites et analysables de manière automatique. De conséquence, il est fondamentale d'étudier et diviser chaque famille dérivationnelle en unités lexicales, comme il est nécessaire d'identifier et de marquer les portions de texte jugées pertinentes et de rendre explicites leurs fonctions.

En conclusion, pour faciliter l'automatisation des procédures d'élaborations du texte, il est nécessaire d'enrichir le texte d'informations qui permettent à un programme informatique d'avoir accès : i) aux informations concernant la structure et l'articulation du texte en différentes parties et ; ii) à la structure linguistique sous-jacente qui regarde les informations morphologiques, syntaxiques, sémantiques etc. Ces démarches font partie de ce qu'on appelle l'encodage de niveau supérieure, réalisable en utilisant un langage de balisage (langage de mark-up), dont le plus important est le langage XML (eXtensible Mark-up Language).

2.1.2 Encodage de niveau supérieur

2.1.2.1 Le format XML

Le marquage en XML permet d'utiliser des ressources informatiques capables d'augmenter les potentialités de l'élaboration et de l'analyse des textes.³ Le format XML est également un *format de texte brut*, mais il permet d'enrichir le texte avec l'ajout d'étiquettes de balisage (ou tag) qui marquent et identifient les blocs qui le constituent. On parle de *document semi-structuré*.

2.1.2.2 Balisage de la structure du contexte

La figure 2.3, illustre les méta-informations qui ont été ajoutées dans le texte brut permettant d'identifier le titre, l'auteur, la date de production et la nature du texte en sujet. N'ayant pas une date précise concernant la production de *qāmūs*, on a déclaré la date entre la naissance et la mort de l'auteur. Grâce à ces méta-informations, il a été possible de les identifier et les marquer dans la version XML après la conversion.



Fig. 2.3 : Balisage des méta-informations

2.1.2.3 Conversion de *qāmūs* en format XML semi-structuré

Pour un format XML complet des autres informations, il est nécessaire d'enrichir le *document plat* de *qāmūs*, en ajoutant des marqueurs permettant d'identifier, automatiquement, les différents segments qui le constituent. L'analyse minutieuse des conventions lexicographiques a permis d'identifier plusieurs niveaux de structure qui ont induit à adopter les démarches illustrées dans les chapitres suivants.

3. Lenci et al, 2012, pp. 62-69

Un programme développé ad hoc permet d'ajouter, automatiquement, dans le texte de *qāmūs*, des indicateurs identifiant les portions qui constituent la macrostructure. La figure 2.4 illustre un fragment du texte qui en résulte après cette phase.

1 باب الباء

2 فصل الهمزة

@ الأَبُّ: الكَلأ، أو المرعى، أو ما أنبتت الأرض، والخضر، و د باليمن **وبالكسر**: ة باليمن **وَأَبٌ** للسَّيْرِ يَنْبُ وَيُؤَبُّ أَبًا وَأَبِيًّا وَأَبَابًا وَأَبَابًا وَأَبَابَةً: تَهَيَّأَ، كَانَتْتَبَّ **و-** إلى وطنه أَبًا وَأَبَابَةً وَأَبَابَةً: اشْتَأَقَ **و-** يَدُهُ إِلَى سَيْفِهِ: رَدَّهَا لَيْسَلَهُ **و** هو في **أَبَابِهِ**: في جهازه **وَأَبٌ** أَبَةٌ: قَصَدَ قَصْدَهُ **وَأَبْتُ** أَبَاتِهِ وَيَكْسُرُ اسْتَقَامَتْ طَرِيقَتُهُ **وَالأَبَابُ**: المَاءُ، والسراب **وبالضم**: مُعْظَمُ السَّيْلِ **وَأَبٌ**: هَزَمَ بِحَمَلَةٍ لَا مَكْدُوبَةَ فِيهَا **و-** الشَّيْءَ: حَرَكَهُ **وَأَبَةٌ**: اسْمٌ وَبِهِ سُمِّيَتْ أَبَةُ الْعُلَيَّا وَالسُّفْلَى قَرِيْنَانِ بَلَّحَ **وبالضم**: د بِأَفْرِيْقِيَّةٍ **وَأَبْتُ**: صَاحَ **وَتَأَبَّبَ** بِهِ: تَعَجَّبَ، وَتَبَجَّحَ **وَأَبِي**، كَحَتَّى: نَهَرَ بَيْنَ الْكُوفَةِ وَقَصْرِ بَنِي مُقَاتِلَ، يُنْسَبُ إِلَى أَبِي بِنِ الصَّامِغَانِ مِنْ مُلُوكِ النَّبَطِ، وَنَهْرٌ بِوِاسِطِ الْعِرَاقِ، وَيُنْرُّ بِالْمَدِيْنَةِ، أَوْ هِيَ أَنَا بِالنُّونِ مُخَفَّفَةً، كَهُنَا.

@ الإِنْتُبُ، بالكسر، والمنتبئة، كمكنسة: بُرْدٌ يُشَقُّ فَنَلْبَسُهُ الْمَرْأَةُ مِنْ غَيْرِ جَيْبٍ وَلَا كَمِيْنٍ، وَالْبَقِيْرَةُ، وَدِرْعُ الْمَرْأَةِ، وَمَا قَصُرَ مِنَ الثِّيَابِ فَتَصِفُ السَّاقَ، أَوْ سِرَاوِيْلَ بِلَا رَجْلِيْنِ، أَوْ قَمِيصٌ بِلَا كَمِيْنِ، ج. أَنَابَ وَإِتَابَ وَأَثُوبٌ **أَنْتَبَ** الثُّوبُ تَأْتِيْبِيًّا: صَيَّرَ إِثْبًا **تَأْتَبَ** بِهِ، وَأَنْتَبَ: لَبَسَهُ. **أَنْتَبَهُ** إِيَاه تَأْتِيْبِيًّا: أَلْبَسَهُ إِيَاه **إِنْتَبَ** الشَّعِيْرَ، بِالْكَسْرِ: قَشَرَهُ **التَّأْتَبُ**: الْاسْتِعْدَادُ، وَالتَّصَلُّبُ، وَأَنْ تَجْعَلَ حِمَالَ الْقَوْسِ فِي صَدْرِكَ **وَتُخْرِجُ مِنْكَ مِنْهَا** رَجُلٌ **مُؤْتَبٌ** الظَّفِرَ، كَمُعْظَمٍ: مُعْوَجَّةٌ.

Fig. 2.4 : Conventions structurelles de *qāmūs*

Le marquage du texte, avec l'ajout des indicateurs, a été effectué sur tout le lexique :

- l'indicateur *1* marque la parole *bāb* "section" qui doit être au début de la ligne et suivie uniquement du nom d'une des lettres de l'alphabet à laquelle elle correspond ;
- l'indicateur *2* marque *faṣl* correspondant à "chapitre". Dans ce cas aussi, la parole *faṣl* doit être au début de la ligne, suivie uniquement du nom d'une des lettres de l'alphabet à laquelle elle correspond ;
- l'indicateur @ marque le début d'une famille dérivationnelle et le retour à la ligne en marque la fin.

À ce stade, le même programme permet, également, la conversion du texte en format XML, comme illustré dans la figure 2.5.

On note que le format XML est un langage de balisage qui permet de segmenter le texte en parties strictement délimitées et déclarées :⁴

4. Lenci et al, 2012, p. 72.

- la première ligne commence par `?>xml` et précise qu'il s'agit en effet d'un document en .XML, la version utilisée est 0.1 et l'encodage est UTF-8 ;
- chaque segment de texte est compris entre deux balises délimitée par des chevrons : une balise ouvrante `<>` et une balise fermante `< / >` ;
- chaque balise *déclare la fonction abstraite* de la portion du texte à laquelle elle se réfère, section, chapitre, Root_Family. On dit que la structure de XML est déclarative ;
- une unité structurale du texte peut contenir d'autres unités et ainsi de suite. Par exemple, une section contient plusieurs chapitres et chaque chapitre contient plusieurs familles dérivationnelles. La structure de XML est, également, hiérarchique.

```

<?xml version="1.0" encoding="UTF-8"?>
<Document>
<section value="الباء">
  <chapitre value="الهمزة">
    <root_family>الأب: الكلاً، أو المرعى، أو ما أنبتت الأرض، والخضر، و د باليمن، وبالكسر: ة باليمن. وأب للسير يئب ويؤب أباً وأبيبا وأباً وأباً وأبابة: تهيأ، كاتب، و إلى وطنه أباً وإبابة وأبابة: اشتاق، و يده إلى سيفه: رذها ليشله. وهو في أبابه: في جهازه. وأب أبه: قصد قمده. وأبت أبابته ويكسر استقامت طريقته والأباب: الماء والسراب وبالضم: معظم السيل. وأب هزم يحملة لا مكذوبة فيها، و الشيء: حزكة. وأبة: اسم ويه سميت أبة العليا والسفلى قريتان يلحج، وبالضم: د بإفريقية. وأبب: صاح. وتأبب به: تعجب، وتبجح. وأبى، كحلى: نهر بين الكوفة وقصر بني مقاتل، ينسب إلى أبى بن الضامغان من ملوك النبط، ونهر بواسط العراق، وينز ، الإتب، بالكسر، والميتية، كمكنسة: برد يشق، فتليسه المرأة من غير جيب ولا كمين والبقيرة، ودرع المرأة، وما قصر من الثياب فنصف الساق، أو سزاويل بلا رجلين، أو قميص بلا كمين، ج: آتاب وإتاب وآتوب. وآتب الثوب تأتياً: صير إتاباً وتأتب به وآتب ليسة وآتبه إياه تأتياً: ألبسه إياه. وإتب الشعر، بالكسر: قشرة. والتأتب: الاستعداد، والتصلب، وأن تجعل جمال القوس في صدرك، وتخرج منكبيك منها. ورجل مؤتب الظهر، كمعظم: معوجه
    </root_family>
  </chapitre>
</section>
</Document>

```

Fig. 2.5 : Fragment de *qāmūs* en XML

Pour une structuration plus fine, il a été nécessaire d'effectuer un opération de pré-traitement en utilisant la ponctuation. Comme démontré dans la figure 2.4, à l'intérieur du même bloc constituant une famille dérivationnelle, le point correspond à la fin d'une entrée lexicale. Le programme permet, en trouvant le point, de retourner à la ligne et d'ajouter un indicateur \$ qui marque le début de l'entrée lexicale. Ainsi, on a pu diviser la famille dérivationnelle en entrées lexicales.

La figure 2.6 illustre une portion de *qāmūs* enrichi par l'ajout des différents marqueurs et délimiteurs. Le début de chaque entrée est marqué par l'indicateur \$ et sa fin est marquée par le retour de la ligne. Il faut noter que le marqueur @ a un double rôle,



Fig. 2.6 : Enrichissement de qāmūs par ajout d'indicateurs

d'une part, il marque le début de la famille dérivationnelle et, de l'autre, il indique une entrée lexicale, comme le marqueur \$.

L'ajout des marqueurs permet la conversion, en format XML, du texte. La figure 2.7 illustre le résultat obtenu après la conversion du fragment de la figure 2.6. La structure hiérarchisé aide à mieux visualiser la macrostructure :

- section dont la valeur (value) correspond au titre du *bāb* dans *qāmūs*;
- chapitre dont la valeur correspond au titre du *faṣl* dans *qāmūs*;
- la famille dérivationnelle (root_family) correspondant à l'ensemble de tous les vocables dérivant d'une racine dont C_1 est égale à la valeur du chapitre et C_3 à la valeur de la section ;
- les entrées lexicales (lexical_entry) dont chacune contient exactement la fraction du *document plat* de *qāmūs* (plain-text) qui lui correspond.

```

<section value="الباء">
  <chapitre value="الهمزة">
    <root_family>
      <lexical_entry>
        <plain-text>أَبٌ لِلسَّيْرِ يَتَّبِ وَيُؤَبُّ أَبًا وَأَبِيًّا وَأَبَابًا وَأَبَابًا وَأَبَابَةً: تَهَيَّأَ، كَانَتْبُ</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبٌ إِلَى وَطْنِهِ أَبًا وَإِبَابَةً وَأَبَابَةً: اشْتَقَّ</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبٌ يَدُهُ إِلَى سَيْفِهِ: رَدَّهَا لَيْسَلَةً</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبٌ أَبَاهُ: قَصَدَ قَصْدَهُ</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبٌ: هَزَمَ</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبٌ الشَّيْءَ: حَرَّكَهُ</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبُّبٌ: صَاحَ</plain-text>
      </lexical_entry>
      <lexical_entry>
        <plain-text>تَأَبَّبَ بِهِ: تَعَجَّبَ، وَتَبَجَّحَ</plain-text>
      </lexical_entry>
    </root_family>
  </chapitre>
</section>

```

Fig. 2.7 : Texte *qāmūs* semi-structuré en XML

2.2 Extractions des informations et enrichissement de *qāmūs*

La recherche d'informations est une étape essentielle dont le but est l'extraction des connaissances spécifiques d'un texte. Avant toute démarche, il faut focaliser les ressources à disposition et les buts visés : quel type de texte on est en train d'étudier ? Quelles types d'informations on cherche et on veut extraire ? Peut-on situer le niveau structurelle où chercher l'information ? Dans quelle mesure peut-on automati-

ser la fouille et l'extraction ?

Dans le contexte de cette recherche, le texte en question est un dictionnaire, sous forme d'un document semi-structuré XML. La structure du dictionnaire, en section, chapitre et entrées lexicales, a été balisée et définie sous forme d'une arborescence logique. Les informations qui se trouvent dans le contenu, par contre, sont encore implicites à plusieurs niveaux. Cependant, étant donné que c'est un dictionnaire, il suit des règles pour la présentation de ses articles.

Le but est expliciter le contenu du dictionnaire et l'enrichir de connaissances à partir des indications fournies par le dictionnaire, lui-même. En d'autres paroles, dans chaque entrée lexicale, il faut :

1. identifier la parole-vedette ;
2. trouver les procédures automatiques et les moins coûteuses (possiblement) pour en déduire le lemme ;
3. identifier les informations qui le caractérisent et l'accompagnent :
 - a) la racine de laquelle il dérive ;
 - b) sa catégorie grammaticale ;
 - c) les informations flexionnelles ;
 - d) les informations dérivationnelles ;
 - e) trouver les procédures automatiques et les moins coûteuses (possiblement) pour identifier le/les sens qui l'accompagnent ;
 - f) la relation lexicale entre la parole-vedette et le/les sens.

Une approche basée sur des règles structurelles et linguistiques a été conçue pour l'extraction automatique des informations pertinentes. La plateforme a été implémentée de manière à être appliquée sur tout le texte. Cependant, pour contrôler la cohérence des règles structurelles et linguistiques mises au point, elles ont été appliquées, dans une première phase, à *bāb al-bā'* la section de la lettre b' et les résultats des extractions ont été corrigés manuellement. Les sous-sections suivantes illustrent les outils et les démarches qui ont été adoptés.

2.2.1 Extraction de la famille dérivationnelle et de la racine

La structure en XML de *qāmūs* permet de limiter la requête à des sous-parties bien définies. Ainsi, en précisant les contraintes structurelles, on peut formuler des recherches sur le contenu et extraire des informations de manière plus précise.

Dans le but d'extraire la racine, on focalise la recherche sur la valeur (value) de la section (*bāb*) et du chapitre (*faṣl*) qui correspondent, respectivement, au nom de la troisième et de la première consonne. Par contre, la deuxième consonne radicale est implicite, car elle n'est pas nommée dans le texte. Par conséquent, elle doit être déduite selon de l'ordre alphabétique de la langue arabe.

Ainsi, on arrive à identifier les trois consonnes radicales (C₁- C₂- C₃) qui caractérisent la famille dérivationnelle. Dans le fichier XML, une nouvelle balise, *Root_family*, est ajoutée automatiquement pour préciser la racine à laquelle appartient chaque entrée lexicale, comme illustré dans la figure 2.8.

```

<section value="الباء">
  <chapter value="الهمزة">
    <root_family C1="الهمزة" C2="الباء" C3="الباء">
      <lexical_entry>
        <plain-text>أَبٌ لِلسَّيْرِ يَنْبُ وَيُؤَبُّ أَبًا وَأَبِيًّا وَأَبَابًا وَأَبَابًا وَأَبَابَةً: تَهَيَّأَ، كَانَتْ بَ </plain-text>
        <root>ب ب ء</root>
      </lexical_entry>
      <lexical_entry>
        <plain-text>أَبٌ إِلَى وَطْنِهِ أَبًا وَإِبَابَةً وَأَبَابَةً: اشْتَأَقَ </plain-text>
        <root>ب ب ء</root>
      </lexical_entry>
    </root_family>
  </chapter>
</section>

```

Fig. 2.8 : Extraction et marquage de la racine et de la famille dérivationnelle

Dans les phases successives, lors de l'extraction de la parole-vedette, les formes fléchies et les paroles dérivées qui l'accompagnent, il a été mis à point un contrôle automatique qui permet de s'assurer qu'elles contiennent les consonnes radicales caractéristiques de la famille dérivationnelle dont elles font partie.

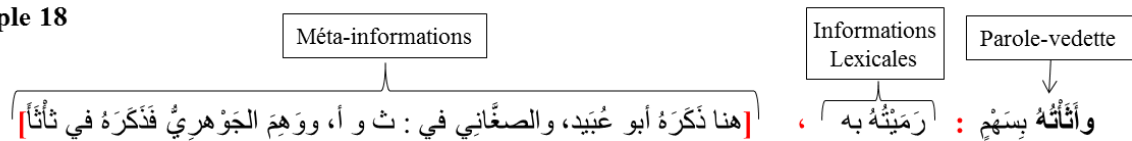
2.2.2 Structuration et enrichissement des entrées verbales

2.2.2.1 Extraction et étiquetage du lemme

2.2.2.1.1 Règles de microstructure L'entrée lexicale est l'unité minimale, au niveau de laquelle on opère pour l'extraction du lemme et des informations qui le concerne. Plus sa structure est bien segmentée et plus la récupération des informations pertinentes est meilleure.

L'entrée lexicale est constituée, de manière générale, d'objets méta-textuels qui ont été contrôlés et identifiés. Afin de limiter les différents blocs et éléments qui la composent, il a été nécessaire d'effectuer une opération de prétraitement en utilisant la ponctuation, comme illustré dans l'exemple 18 de la figure 2.9.

Exemple 18



wa=*'ata*'=*tu*=*hu* avec une flèche : je la lui ai tirée , [c'est ici que l'a listé 'abū 'ubayd, et *aṣ*=*ṣaḡnī* dans *ṭw* , et *al*=*ḡawharī* s'est trompé et l'a listé dans *ṭ't*]

Fig. 2.9 : Marquage de la microstructure de *qāmūs*

- les deux points permettent la localisation des deux plus importantes parties de l'entrée lexicale : à droite des deux points, le bloc contenant la parole-vedette, identifiée comme la première parole de l'entrée lexicale ; et à gauche des deux points le bloc contenant les sens ;
- la virgule permet de séparer entre les éléments dans un même bloc ;
- les crochets permettent d'identifier les méta-informations qui ne seront pas prises en considération dans cette recherche.

L'ajout des crochets et la cohérence de la ponctuation (les deux points et la virgule) ont été contrôlés, manuellement, sur la section de la lettre b qui sert de corpus d'apprentissage.

Nota Dans le reste de cette thèse, il faut rappeler que les crochets, **en rouge**, qui se trouvent dans le texte arabe et dans la traduction, sont ajoutés pour marquer les méta-informations qui ne sont pas prises en considération dans cette recherche. Par contre les crochets qui se trouve seulement dans la traduction soulignent le

texte qui n'existe pas dans le texte arabe mais ajouté afin de rendre la traduction plus fluide et compréhensible.

2.2.2.1.2 Patrons morphosyntaxiques Cependant, l'analyse morphosyntaxique, dans le tableau 2.1, révèle que la parole-vedette est complexe, constituée d'une forme fléchie incluse dans une phrase simple.

Parole-vedette	: <i>wa=ʾaṭaʾ=tu=hu (bi=sahm-i-n)</i>
Analyse	: <i>et=tirer_{PFV}=je=lui (avec=flèche-GEN-INDF)</i>
Traduction	: "je lui ai tiré (une flèche)"

Tab. 2.1 : Analyse morphosyntaxique de la parole-vedette de l'exemple 18

Par conséquence, on a eu recours à la formalisation de règles morphologiques et morphosyntaxiques qui permettent d'identifier et enlever les clitiques et les affixes de flexion.

Pour les langues occidentales, le procédé d'enlever les clitiques et les affixes, pour obtenir la racine ou le stem d'un mot, est appelé désuffixation ou stemming. Il est, également, appelé racinisation ; appellation qui peut causer certaines confusions concernant la langue arabe. Justement, la racine ou stem d'un mot, en langue occidentale, correspond à la partie du mot restante, après avoir supprimer tous les préfixe(s) et les suffixe(s). Par exemple, «chant » est le stem ou la partie commune des paroles "chanter", "chanteront", "chanteur", "chanteuses".

En langue arabe, la racine et le stem représentent deux concepts très divers. Par exemple, la figure 2.2 présente l'analyse de certaines formes fléchies, la racine et les stems correspondants au verbe *kataba* "écrire".

Exemple	<i>wa=katab=ta=hu</i>	<i>wa=ta-ktub-u=hu</i>
Analyse morphosyntaxique	<i>et=écrire_{PFV}=tu=le</i> 'et tu l'as écrit'	<i>et=2-écrire_{IPFV}-IND.PRS=le</i> 'et tu l'écris '
Stem (base)	<i>katab</i>	<i>ktub</i>
Racine	<i>ktb</i>	
Lemme	<i>kataba</i>	

Tab. 2.2 : Stem et racine du verbe *kataba* "écrire"

En langue arabe, la racine est discontinue et correspond à la partie consonantique commune à toutes les paroles d'une famille dérivationnelle et leurs formes fléchies (voir chapitre 2.1). Le procédé pour la récupérer, dans cette recherche, fait recours à des patrons structuraux qui ont été déjà expliqués dans la sous-section 2.2.1.

Le stem, par contre, correspond à la base commune aux formes fléchies d'une même parole (voir l'étude morphologique 3.1 et 4.2). Concernant les verbes, on peut avoir plusieurs stems. Dans le cas du verbe régulier, par exemple *kataba*, on distingue deux stems différents, **katab** pour l'accompli et **ktub** pour l'inaccompli. Le procédé d'extraire le stem, très utilisé pour la récupération d'information (Information Retrieval : IR), est appelé, dans ce cas, « light stemming ». ⁵ Dans le cadre de l'actuelle recherche, il a été nécessaire d'implémenter un algorithme de « light stemming », basé sur les règles morphologiques et morphosyntaxiques écrites sous formes d'expressions régulières.

Une expression régulière, appelée communément regex (contraction de regular expression), est une chaîne de caractères (ou une formule) écrite dans un langage ayant une signification particulière pour le programme qui va l'interpréter. Elle permet de représenter des modèles (ou des *patrons*) de chaînes de caractères qui servent à décrire, de manière concise, les phénomènes linguistiques qu'on veut sélectionner ou chercher. Le principe est identifier le contexte et les délimiteurs de l'information considérée pertinente, afin de constituer la ou les expressions régulières qui permettent de l'identifier. Lorsqu'une portion du texte correspond à une expression régulière, elle est mémorisée, extraite et enfin balisée.

L'algorithme est mis au point de manière à être appliqué à tout le lexique *qāmūs* et à la langue arabe en générale. Dans ce but, les règles morphosyntaxiques implémentées sont générales, afin d'identifier toute forme fléchie ou complexe et en extraire le stem. Comme exemple, le tableau 2.3 suivant illustre les patrons des regex qui identifient une forme fléchie verbale.

Exemple 1	<i>wa=katab=ta=hu</i>
Analyse morphosyntaxique	Conj= katab .PFV=Pron.NOM=Pron.Acc
regex	Conj? +base .PFV+Suff.NOM+Suff.ACC?
Exemple 2	<i>wa=ta-ktub-u=hu</i>
Analyse morphosyntaxique	Conj=Pref- ktub .IPFV-Suff.NOM=Pron.Acc
regex	Conj? +Pref+base .IPFV+Suff.NOM+Suff.ACC?

Tab. 2.3 : Regex identifiant les stem verbaux

5. Larkey et al., 2007 et Zeroual et Lakhouaja, 2014

Dans la syntaxe des regex, dans le cas où un élément est aléatoire, il est suivi d'un point d'interrogation. Par exemple, la regex de l'exemple 1 spécifie que la base verbale à l'accompli est **obligatoirement** suivi d'un suffixe nominatif. Par contre, la présence d'une conjonction et d'un suffixe accusatif est **aléatoire**.

Le procédé se déroule en plusieurs phases. Après avoir identifié la parole, l'algorithme permet d'extraire la base verbale en enlevant les affixes de flexion et les éventuels clitiques. Ainsi, par exemple, on obtient le stem **katab** ou **ktub** à partir, respectivement, de la forme *wa=katab=ta=hu* ou *wa=ta-ktub-u=hu*. Une phase de recodage permet ensuite de créer le lemme **kataba**, à partir du stem, en changeant, éventuellement, les voyelles internes et en ajoutant le suffixe *a*. Dans ce but, il a été nécessaire d'établir une liste explicite :

- des préfixes et des suffixes de flexion, y compris les pronoms nominatifs, spécifiques et communs à tous les verbes ;
- des pronoms accusatifs compatibles avec les verbes ;
- des proclitiques compatibles avec les verbes, par exemple, la conjonction *wa* "et".

Il a été nécessaire, également, d'établir la liste des schèmes des bases verbales, pour identifier la nature du verbe, trilitère ou quadrilitère (forme I) ou dérivé (forme II, III, IV, etc.). Le tableau 2.4, ci-dessous, donne quelques exemples des regex qui permettent d'identifier la nature de la base verbale.

La première voyelle est **aléatoire** pour tous les verbes et donc suivie de ?. La voyelle médiane du verbe trilitère est **obligatoire** car elle a une valeur sémantique. Et enfin, le signe du doublement est obligatoire pour tous les verbes, parce qu'il a une valeur consonantique (dans la syntaxe des regex, il est présenté par le signe ~).

verbe	regex	exemple	lemme
trilitère	$C_1(a) ?C_2aC_3$	<i>katab</i>	<i>katab-a</i> "écrire"
	$C_1(a) ?C_2uC_3$	<i>'adub</i>	<i>'adub-a</i> "être poli"
	$C_1(a) ?C_2iC_3$	<i>'arib</i>	<i>'arib-a</i> "avoir besoin"
	$C_1(a) ?C_2\sim$	<i>'abb</i>	<i>'abb-a</i> "se préparer"
	$C_1(a) ?AC_3$	<i>tāb</i>	<i>tāb-a</i> "se repentir"
dérivé II	$C_1(a) ?C_2\sim (a) ?C_3$	<i>'annab</i>	<i>'annab-a</i> "réprimander"

Tab. 2.4 : Exemples de regex identifiant les différents formes des verbes

Une des difficultés qu'on peut trouver est due à l'obligation de présence de certains diacritiques, comme la voyelle médiane du trilitère, le signe de redoublement et le suf-

fixe flexionnel de la 3PSM. Dans les textes arabes, on néglige de mettre les diacritiques et les voyelles et, dans ce cas, la regex n'identifie pas la parole-vedette. On a dû vérifier toutes les entrées lexicales verbales de la section de la lettre b pour contrôler que les règles sont adéquates et correctes.

Comme le démontre la figure 2.10 qui illustre l'exemple d'un verbe quadrilitère et certains verbes qui en dérivent, on arrive à ajouter, de manière automatique, les balises qui enrichissent le texte.

```

<lexical_entry>
  <plain_text>فَتَحْظَرِبَ السَّقَاءَ: مَلَأَهُ، فَتَحْظَرِبَ</plain_text>
  <lemma>حَظَرِبَ</lemma>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <root_information>Q</root_information>
  <morphosyntactic_information>Transitif</morphosyntactic_information>
  <sense>مَلَأَ</sense>
  <root>ح ظ ر ب</root>
</lexical_entry>
<lexical_entry>
  <plain_text>تَحْظَرِبَ: امْتَلَأَ عِدَاوَةً أَوْ طَعَامًا وَغَيْرَهُ</plain_text>
  <lemma>تَحْظَرِبَ</lemma>
  <pos>Verb</pos>
  <morphological_information>II</morphological_information>
  <root_information>Q</root_information>
  <morphosyntactic_information>Intransitif</morphosyntactic_information>
  <sense>امْتَلَأَ عِدَاوَةً أَوْ طَعَامًا وَغَيْرَهُ</sense>
  <root>ح ظ ر ب</root>
</lexical_entry>

```

Fig. 2.10 : Structuration des entrées lexicales du verbe *ḥazraba* et ses dérivés

Ainsi on arrive à spécifier :

- la catégorie grammaticale (POS : contraction de Part Of Speech) spécifiant que c'est un verbe, du fait que les suffixes de flexion sont spécifiques du verbe,
- les informations morphologiques (morphological_information) : I (forme de base) ; II ; III ; IV, ... XV (forme dérivée) ;
- les informations concernant la racine (root_information) : T (trilitère) / Q (quadrilitère) ;
- les informations morphosyntaxiques (morphosyntactic_information) : transitif / intransitif. Dans le cas où le verbe a un pronom accusatif ou est suivi d'un nom

accusatif (complément d'objet direct), le lemme acquiert, automatiquement, le statut du verbe transitif. Par contre, tous les verbes trilitères ayant une voyelle médiane u, acquièrent, automatiquement, le statut intransitif. Pour ceux qui n'ont aucun indicateur textuel, les informations morphosyntaxiques doivent être ajoutées manuellement.

2.2.2.1.3 Cas particuliers Dans certains cas, l'auteur ne mentionne pas explicitement le verbe et se limite à préciser un verbe plus connu qui sert de modèle sur lequel construire le verbe. C'est le cas de la deuxième entrée lexicale de l'exemple 19.

Exemple 19a

'*aruba* [...] : il est devenu sage [...]'

أُرِبَ [...] : عَقَلَ [...]

Exemple 19b

'comme=*fariḥa*: il est flexible, et il a besoin'

كَفَرِحَ : دَرَبَ، وَاحْتَاَجَ

Dans la première entrée, exemple 19a, le verbe est explicitement mentionné, '*aruba* "devenir sage". Par contre, dans l'entrée de l'exemple 19b, qui suit directement la première, le verbe modèle est en position initiale de l'entrée lexicale, introduit par la conjonction *ka* "comme" qui joue le rôle d'un marqueur morphosyntaxique.

Ce style, très utilisé par l'auteur, permet de contrôler les schèmes du verbe à l'accompli et à l'inaccompli. Le système, devant une parole initiale commençant par *ka* ou *k*, lance les procédures suivantes :

- il contrôle que la première consonne radicale n'est pas la lettre k, comme dans le verbe *kataba* "écrire". Dans le cas où elle ne l'est pas, il enlève la *k* ou *ka*, considérée conjonction "comme" ;
- il récupère le verbe modèle restant et change les consonnes avec celles de la racine qui caractérise la famille dérivationnelle qu'il traite.

Dans le cas de l'exemple 19b, le système construit automatiquement, le verbe '*ariba*, en changeant les consonnes radicales *frḥ*, du verbe modèle *fariḥa*, avec les consonnes radicales '*rb*, caractérisant la famille dérivationnelle traitée, et identifiée par l'étiquette <root>. La figure 2.11 montre les résultats obtenus.

Pour les deux entrées lexicales, le système ajoute, de manière automatique, une balise identifiant le *muḍāri'*. Concernant le premier verbe, '*aruba*, c'est un verbe d'état

```

<lexical_entry>
<plain_text>أَرْبٌ إِزْبًا كَصَغْرٍ صِغْرًا، وَأَرَابَةٌ، كَكْرَامَةٍ: عَقْلٌ، فَهُوَ أَرِيبٌ وَأَرِيبٌ</plain_text>
<lemma>أَرْبٌ</lemma>
<imperfective>يَأْرِبُ</imperfective>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>T</root_information>
<morphosyntactic_information>Intransitif</morphosyntactic_information>
<masdar>إِزْبٌ</masdar>
<masdar>أَرَابَةٌ</masdar>
<adjective>أَرِيبٌ</adjective>
<adjective>أَرِبٌ</adjective>
<sense>عَقْلٌ</sense>
<root>ر ب ء</root>
</lexical_entry>
<lexical_entry>
<plain_text>كَفَّرَخَ: ذَرِبَ، وَاجْتَنَجَ</plain_text>
<lemma>أَرْبٌ</lemma>
<imperfective>يَأْرِبُ</imperfective>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>T</root_information>
<sense>ذَرِبٌ</sense>
<sense>اجْتَنَجٌ</sense>
<root>ر ب ء</root>
</lexical_entry>

```

Fig. 2.11 : Structuration de l'entrée lexicale du verbe 'ariba

dont le *muḍāri'* est toujours de schème $yaC_1C_2uC_3u$, *ya'rubu*. Concernant le deuxième lemme 'ariba, le *muḍāri'* est construit selon le verbe modèle *fariḥa* dont le *muḍāri'* est *yafraḥu*, et de conséquence le *muḍāri'* est *ya'rabu*.

Pour un contrôle majeur, les verbes modèles, utilisés dans la section de la lettre *b*, ont été identifiés et ils sont listés dans le tableau 2.5 suivant.

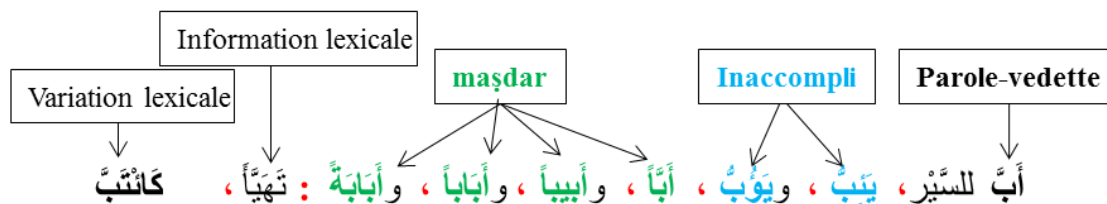
Verbe modèle	
<i>māḍī</i>	<i>muḍāri'</i>
<i>fariḥa</i> "se réjouir"	<i>yafraḥu</i>
<i>sami'a</i> "écouter"	<i>yasma'u</i>
<i>naṣara</i> "soutenir"	<i>yaṣuru</i>
<i>mana'a</i> "écouter"	<i>yamna'u</i>
<i>ḍaraba</i> "frapper"	<i>yaḍribu</i>
<i>karuma</i> "être généreux"	<i>yakrumu</i>

Tab. 2.5 : Verbes modèles détectés dans la section de la lettre *b*

2.2.2.2 Extraction des informations flexionnelles et dérivationnelles

Dans la figure 2.12, l'exemple 20 illustre l'entrée lexicale du verbe trilitère 'abba "se préparer". On note que la parole-vedette à l'accompli est suivi, directement, par deux formes fléchies du *muḍāri'* et que les *maṣḍar* qui l'accompagnent sont indéfinis et à l'accusatif.

Exemple 20



'abba pour marcher, *ya 'ibbu*, et=*ya 'ubbu*, '*abb-an*, et=*'abīb-an*, et=*'abāb-an*, et=*'abābat-an* : il s'est préparé, *ka='i tabba*

Fig. 2.12 : Entrée lexicale du verbe 'abba

2.2.2.2.1 Patrons morphologiques et morphosyntaxiques Les procédures qui permettent d'identifier les formes fléchies du *muḍāri'*, sont les mêmes, déjà expliquées dans le paragraphe précédent 2.2.2.1.2. L'algorithme identifie les formes du *muḍāri'* qui suivent directement la parole-vedette, selon le patron général de la regex, illustré dans le tableau 2.6. Cependant, dans ce cas, l'algorithme enlève seulement les conjonctions et les suffixes accusatifs, dans le cas où il les trouve. La forme restante, *Pref+base.IPfV+Suff.NOM*, est sélectionnée et marquée.

Exemple forme fléchie <i>muḍāri'</i>	: <i>wa=ta-ktub-u=hu</i>
Analyse morphosyntaxique	: <i>Conj=Pref-ktub.IPfV-Suff.NOM=Suff.ACC</i>
Regex	: <i>Conj ?+Pref+base.IPfV+Suff.NOM+Suff.ACC?</i>
Forme <i>muḍāri'</i> sélectionnée	: <i>Pref+base.IPfV+Suff.NOM</i>
Résultat	: <i>ta-ktub-u</i>

Tab. 2.6 : Patron général de la regex identifiant le *muḍāri'*

2.2.2.2.2 Patrons syntaxiques En général, dans la langue arabe, la construction, du verbe suivi de son *maṣḍar* à l'accusatif, est très utilisée. Sur le plan syntaxique, le

maṣḍar est indéfini à l'accusatif et il est appelé *maf'ūl muṭlaq*. Sur le plan sémantique, cette construction peut servir à :⁶

- *ta'kīd*: corroborer le procès désigné par le verbe, comme dans l'exemple :
ḡalas=tu ḡulūs-an "je me suis *effectivement* assis" ;
- *bayān al=naw'* : spécifier la manière dont est avenue l'action, par exemple :
ḡalas=tu ḡilsat-an "je me suis assis *d'une certaine manière*" ;
- *bayān al='adad* : quantifier combien de fois est avenue l'action, par exemple :
ḡalas=tu ḡalsat-an "je me suis assis *une fois*".

Cette construction est bien respectée dans *qāmūs* (et dans tous les dictionnaires de langue arabe), et on peut l'exploiter pour pouvoir détecter et extraire le/les *maṣḍar*.

Toute parole qui suit une parole-vedette verbale, détectée indéfinie, à l'accusatif finissant avec le suffixe *an*, est soumise au contrôle. Dans le cas où elle contient les mêmes consonnes radicales et donc appartient à la même famille dérivationnelle de la parole-vedette, elle est mémorisée, traitée pour enlever le suffixe *an* et, enfin, balisée avec l'étiquette qui l'identifie comme *maṣḍar*. La figure 2.13 illustre les résultats des extractions des informations concernant l'entrée lexicale *'abba*, enrichie grâce à l'identification et le balisage de :

- le lemme : أَبَّ 'abba;
- les formes fléchies du *muḍāri'* (imperfective) : يَأْبُ *ya'ubbu* et يَأْبُ *ya'ibbu*;
- la catégorie grammaticale (POS) : Verb ;
- informations morphologiques (*morphological_information*) : I - verbe de base ;
- Informations sur la racine de laquelle il dérive (*root_information*) : T (trilitère) ;
- Information morphosyntaxique (*morphosyntactic_information*) : Intransitif ;
- les quattres *maṣḍar* qui l'accompagnent ;

6. Les exemples sont pris de [Larcher, 1991](#).

- le sens corrélé au lemme : تَهَيَّأَ "préparer";
- les consonnes de la racine de laquelle il dérive (root) : ب ب ء (' b b).

```

<lexical_entry>
<plain_text>أَبٌ لِلسَّيْرِ يَتَّبُ وَيُؤَبُّ أَبًا وَأَبِيًّا وَأَبَابًا وَأَبَابًا وَأَبَابَةً: تَهَيَّأَ، كَانَتْبُ</plain_text>
<lemma>أَبٌ</lemma>
<imperfective>يَتَّبُ</imperfective>
<imperfective>يُؤَبُّ</imperfective>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>I</root_information>
<morphosyntactic_information>Intransitif</morphosyntactic_information>
<masdar>أَبٌ</masdar>
<masdar>أَبِيْبُ</masdar>
<masdar>أَبَابُ</masdar>
<masdar>أَبَابُ</masdar>
<masdar>أَبَابَةٌ</masdar>
<sense>تَهَيَّأَ</sense>
<sense>كَانَتْبُ</sense>
<root>ب ب ء</root>
</lexical_entry>

```

Fig. 2.13 : Structuration des entrées lexicales du verbe 'abba

Concernant les verbes quadrilitères et les verbes dérivés, la forme du *muḍāri'* est formelle et peut être construite et ajoutée automatiquement. La figure 2.14 montre le résultat d'extraction des informations concernant la parole-vedette جَرَجَبَهُ *garğaba=hu*:

- le lemme est : جَرَجَبَ *garğaba* "manger";
- la forme du *muḍāri'* construite automatiquement : يُجَرِّجِبُ *yuğarğibu*;

- la catégorie grammaticale (POS) : Verb ;
- les informations morphologiques (morphological_information) : I - verbe de base ;
- les informations sur la racine de laquelle il dérive (root_information) : Q (quadrilittère) ;
- les informations morphosyntaxique (morphosyntactic_information) : Transitif ;
- le sens corrélé au lemme : أَكَلَ "manger" ;
- les consonnes de la racine de laquelle il dérive (root) : ج ر ب (ğ r ğ b).

```

<lexical_entry>
  <plain_text>أَكَلَ: أَكَلَةٌ</plain_text>
  <lemma>أَكَلَ</lemma>
  <imperfective>يُكْرِبُ</imperfective>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <root_information>Q</root_information>
  <morphosyntactic_information>Transitif</morphosyntactic_information>
  <sense>أَكَلَ</sense>
  <root>ج ر ب</root>
</lexical_entry>

```

Fig. 2.14 : Entrée lexicale du verbe *ğarğaba*

2.2.2.2.3 Patrons linguistiques Une phase également importante est l'étude des régularités linguistiques et stylistiques qui permettent de repérer les expressions (ou patrons) linguistiques récurrentes, comme illustré dans les exemples suivants.

Dans la figure 2.15, l'exemple 21 illustre l'entrée lexicale de la parole-vedette *wa=ṣaḥība=hu*. Après les deux *maṣdar* qui l'accompagnent et le sens *عَاشِرُهُ* "il l'a accompagné", l'auteur a listé un adjectif *صَاحِبٌ ṣaḥībun* "compagnon ; copain",

”la chamelle **ḡaḍabat** : son lait a diminué, **donc=elle [est]** ḡāḍibatun et=ḡāḍūbun, pl. ḡawāḍibu et=ḡiḍābun, comme niyAmun”

Indépendamment de leur position dans l’entrée lexicale, le repérage des patrons linguistiques *wa=huwa / fa=huwa / wa=hiya / fa=hiya*, permet au système de :

- mémoriser les paroles qui suivent ;
- procéder au contrôle qu’elles contiennent les consonnes radicales caractéristiques de la famille dérivationnelle ;
- dans le cas positif, ajouter, automatiquement, le/les balises qui identifient le/les adjectifs.

Dans le cas où l’adjectif est, éventuellement, suivi de l’indicateur du pluriel, ḡ, le système déclenche les procédures de :

- mémoriser les paroles qui suivent et vérifier qu’elles contiennent les consonnes radicales de la famille dérivationnelle ;
- dans le cas positif, ajouter, automatiquement, la/les balises qui identifient le/les pluriels de l’adjectif.

Les figures 2.16, 2.17 et 2.18 montrent les résultats obtenus des entrées lexicales des exemples 21, 22 et 23.

```
<lexical_entry>
<plain_text>صَحْبٌ</plain_text>
<lemma>صَحْبٌ</lemma>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>T</root_information>
<morphosyntactic_information>Transitif</morphosyntactic_information>
<masdar>صَحْبَانَةٌ</masdar>
<masdar>صُحْبَانٌ</masdar>
<adjective>صَاحِبٌ</adjective>
<plural>أَصْحَابٌ</plural>
<plural>أَصْحَابِيٌّ</plural>
<plural>صُحْبَانٌ</plural>
<plural>صَحَابٌ</plural>
<plural>صَحَابَةٌ</plural>
<plural>صَحَابَةٌ</plural>
<plural>صَحْبٌ</plural>
<sense>عاشِرٌ</sense>
<root>ص ح ب</root>
</lexical_entry>
```

Fig. 2.16 : Structuration de l’entrée lexicale du verbe *ṣahiba*

```

<lexical_entry>
  <plain_text>"جَشَبَ الطَّعَامَ، كَتَصَرَ وَسَمِعَ، فَهُوَ جَشِبٌ وَجَشِبٌ وَمِجْشَابٌ وَمِجْشَابٌ وَجَشِيبٌ وَمِجْشُوبٌ، أَي: غَلِيظٌ، أَوْ "يَلَأُ أَدْمٌ"</plain_text>
  <lemma>جَشَبَ</lemma>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <adjective>جَشِبٌ</adjective>
  <adjective>جَشِبٌ</adjective>
  <adjective>مِجْشَابٌ</adjective>
  <adjective>جَشِيبٌ</adjective>
  <adjective>مِجْشُوبٌ</adjective>
  <sense>غَلِظٌ</sense>
  <sense>يَلَأُ أَدْمٌ</sense>
  <root>ج ش ب</root>
</lexical_entry>

```

Fig. 2.17 : Structuration de l'entrée lexicale du verbe *ḡašaba*

```

<lexical_entry>
  <plain_text>جَذَبَتِ النَّاقَةَ: قَلَّ لَبْنُهَا، فَهِيَ جَاذِبٌ وَجَاذِيَةٌ وَجَذُوبٌ، ج. جَوَادِبٌ وَجَذَابٌ، كَنِيَامٍ</plain_text>
  <lemma>جَذَبَ</lemma>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <morphosyntactic_information>Intransitif</morphosyntactic_information>
  <adjective>جَاذِبٌ</adjective>
  <adjective>جَاذِيَةٌ</adjective>
  <adjective>جَذُوبٌ</adjective>
  <sense>قَلَّ لَبْنُهَا</sense>
  <root>ج ذ ب</root>
</lexical_entry>

```

Fig. 2.18 : Structuration de l'entrée lexicale du verbe *ḡaḍaba*

2.2.3 Structuration et enrichissement des entrées nominales

2.2.3.1 Règles de microstructure

Généralement, la parole-vedette nominale est identifiée comme première parole de l'entrée lexicale, accompagnée de l'article défini (*al=*) ou d'un proclitique complexe *wa=al*, formé de la conjonction *wa* "et" et de l'article défini *al* "le", comme le montre l'exemple 24 suivant.

Exemple 24

وَالْأَبَابُ : الْمَاءُ، وَالسَّرَابُ

"**et=le= 'abāb** : l'eau, et le mirage"

Cependant, dans plusieurs cas, comme l'exemple 25, la parole-vedette est un adjectif, bien que qu'elle se localise en position initiale. De conséquence, les règles de position

structurelle, concernant les noms, sont ambigus et exigent un contrôle manuel des résultats obtenus.

Exemple 25

وَالْجَشِيبُ : الْحَسَنُ، الْعَلِيظُ [...]]

”et=le=*ġašīb* : le brutal, le grossier [...]”

2.2.3.2 Patrons morphosyntaxiques

Pour appliquer les procédures de « light Stemming », il a été nécessaire, dans ce cas aussi, de faire une liste complète et explicite des proclitiques simples : article défini, prépositions, conjonction, et des proclitiques composés par conjonction et/ou préposition et/ou article défini. La regex permettant d’identifier les noms (et les adjectifs) correspond à la règle générale dans le tableau 2.7. Éventuellement, le nom est précédé au maximum, par le proclitique complexe formé par conjonction+préposition+article défini. La conjonction et la préposition peuvent manquer et sont accompagnées par (?). La présence de l’article défini, par contre, est déterminante pour la sélection de la parole-vedette nominale. La voyelle de déclinaison peut manquer dans le texte arabe et du fait, elle est suivie, dans la regex par (?). Pour obtenir le lemme, il est nécessaire d’enlever les proclitiques et la dernière voyelle dont le rôle est syntaxique.

Parole-vedette :	wa=al='abāb-u.NOM
Regex :	Conj?+Prep?+Def+Nom+(v)?
Procédures :	« light stemming »
Lemme :	'abāb

Tab. 2.7 : Patron général de la regex identifiant les noms

L’exemple 26, ci-dessous, a été choisi pour sa complexité, et permet d’illustrer d’autres procédures, mises à point pour analyser et enrichir certaines entrées nominales qui nécessitent plusieurs phases de traitement.

Exemple 26

وَالثَّلْبُ ، بِالْكَسْرِ : الْجَمَلُ تَكَسَّرَتْ أَنْيَابُهُ هَرَمًا وَتَنَاطَرَ هَلْبُ ذَنْبِهِ، جَ أَثْلَابٌ وَثَلْبَةٌ وَهِيَ بِهَاءٍ، وَالشَّيْخُ، وَالْبَعِيرُ لَمْ يُلْقَحْ، وَصَحَائِي [أَوْ هُوَ بِالتَّاءِ وَتَقَدَّمَ]

وَكَيْفِ : الْمَسْتَلَمُ مِنَ الرِّمَاحِ

وَبِالتَّخْرِيكِ : التَّقْبُضُ

L'exemple est constitué par l'enchaînement de trois entrées lexicales, de la même famille dérivationnelle, où la première est suivie de deux autres dont la parole-vedette n'est pas explicite. La première entrée lexicale contient plusieurs informations condensées et sous-jacentes.

Première entrée lexicale de l'exemple 26

وَالثَّلْبُ ، بِالْكَسْرِ : الْجَمَلُ تَكَسَّرَتْ أَنْيَابُهُ هَرَمًا وَتَنَاقَرَ هَلْبُ ذَنْبِهِ ، جَ أَثْلَابٌ وَثَلْبَةٌ وَهِيَ بِهَاءٍ ، وَالشُّيْخُ ، وَالْبَعِيرُ لَمْ يُلْقَحْ ، وَصَحَابِيٌّ [أَوْ هُوَ بِالتَّاءِ وَتَقَدَّمَ]

"**et=le=tilb**, avec la voyelle i : le chameau qui a perdu ses canines et les poils de sa queue par vieillesse, **pl.** 'aṭlāb, et *tilabat*, **et elle est** avec *hā'*, et le vieux, et le mulet qui n'a pas été fécondé, et un Compagnon [ou ce dernier est avec la *tā'* et il a déjà été mentionné]".

La position en tête permet d'identifier la parole-vedette, *وَالثَّلْبُ* "et=le=tilb", et les patrons morphosyntaxiques permettent d'obtenir le lemme *tilb*. La parole-vedette est suivie du contrôle vocalique *bi=l=kasri* qui indique que la première consonne est vocalisée 'avec la voyelle i' .

Le premier sens, "le chameau qui a perdu ses canines et les poils de sa queue par vieillesse", est suivi d'une structure commençant par le marqueur *ǧ* "pl.". Comme déjà expliqué dans le paragraphe 2.2.2.2.3, en trouvant le marqueur linguistique *ǧ*, le système déclenche les procédures pour contrôler que les paroles qui suivent contiennent les consonnes radicales du lemme *et*, dans le cas positif, faire le balisage automatique que ce sont des pluriels.

Un autre marqueur *وَهِيَ بِهَاءٍ* "et elle [est] avec un *hā'*" déclenche d'autres procédures qui permettent de construire le féminin en ajoutant le suffixe féminin ة :

Lemme masculin : *ثَلْبٌ tilb* —> féminin : *ثَلْبَةٌ tilb-ah*

De l'autre, on note que le dernier sens, *wa=ṣaḥābiyyun* "et un compagnon [du Prophète]" est suivi de données méta-lexicales qui ont été mises entre crochets, afin que le système n'en tient pas compte. Ce genre d'opération est faite également manuelle. En

outre, il s'agit d'un nom propre de personne et les procédures permettant de l'identifier exigent des patrons sémantiques qui seront expliqués plus loin (voir 2.2.5).

Dans la deuxième entrée lexicale, l'auteur ne mentionne pas la parole-vedette. Il commence par le syntagme **وَكَتِيفٍ** *wa=ka=katif-in* "et=comme=katif" dont la parole *katif* "épaule" n'a aucune valeur sémantique ou lexicale. Elle sert, seulement, de modèle de vocalisation afin de construire le lemme.

Deuxième entrée lexicale de l'exemple 26

وَكَتِيفٍ : الْمُسْتَلَمُّ مِنَ الرَّمَاحِ

"et=comme=katif-in : les flèches émoussés".

En général, les paroles-modèles doivent être fréquentes et bien connues, afin d'éviter des erreurs de vocalisation. Le système met en jeu les processus résumés dans le tableau 2.8.

Procédures		Exemple
Regex	conj? ka +(modèle)+(in)?	<i>wa=ka=katif-in</i> .GEN.INDEF
récupérer et mémoriser	le modèle : C ₁ V ₁ C ₂ V ₂ C ₃	<i>katif</i>
	la racine : C ₁ C ₂ C ₃	<i>t / b</i>
remplacer c par C pour générer	le lemme : C ₁ V ₁ C ₂ V ₂ C ₃	<i>talib</i>

Tab. 2.8 : Regex et phases de construction du lemme manquant

La structure linguistique correspondante à la regex **doit commencer** par la préposition **ka** "comme", suivi d'une parole-modèle au génitif. La voyelle du génitif indéfini (*in*) peut ne pas être digitalisée et donc absente dans le texte ; pour ce fait elle est suivie par un point d'interrogation dans la regex. La structure, *ka=modèle-in*, peut éventuellement être précédée par une conjonction (généralement *wa* "et"). Le lemme de l'entrée lexicale est construit après avoir :

1. récupéré et mémorisé la parole-modèle **كَتِيفٍ** *katif* et les consonnes radicales de la famille dérivationnelle en question ;
2. changé les consonnes de la parole-modèle avec celles de la racine. Ainsi à partir

de كَتِف *katif*, on obtient le lemme تَلِب *"talib"*

Dans la troisième entrée lexicale, la parole-vedette, non plus, n'est pas explicite.

Troisième entrée lexicale de l'exemple 26

وَبِالتَّحْرِيكِ : التَّقْبُضُ

"**et=avec=la=voyelle=a** : la contraction".

L'auteur spécifie seulement le contrôle vocalique, وَبِالتَّحْرِيكِ 'en vocalisant avec la voyelle a'. Selon les indications dans l'introduction de *qāmūs*, l'auteur a expliqué que lorsqu'il ne mentionne pas la parole-vedette, le contrôle vocalique constitue le changement que subit la parole-vedette qui précède. Dans ce cas, les deux premières voyelles de l'entrée lexicale précédente doivent être changées en a. Le système doit donc récupérer le lemme qui précède, *talib*, et procéder aux changements nécessaires pour construire le lemme, de la troisième entrée, *talab*.

Il a été nécessaire de formaliser les règles de contrôle vocalique et les autres règles qui permettent de construire le lemme et que l'auteur a expliqué dans son introduction. Le tableau 2.9 illustre les directives nécessaires pour les regex responsables des opérations. Tenant compte des caractéristiques du texte arabe, les voyelles peuvent ne pas être présentes dans le lemme qui précède. Justement, dans la regex, elles sont précédées d'un point d'interrogation. L'algorithme, par contre, met la voyelle indiquée à la place d'une autre, que cette dernière soit digitalisée dans le texte ou pas.

Lemme précédent	Patron linguistique	Lemme construit
C ₁ v ₁ ?C ₂ v ₂ ?C ₃	<i>bi=d=dammi</i>	C ₁ uC ₂ v ₂ ?C ₃
	<i>bi=l=fathī</i>	C ₁ aC ₂ v ₂ ?C ₃
	<i>bi=l=kasri</i>	C ₁ iC ₂ v ₂ ?C ₃
	<i>muḥarrakah</i> ou <i>bi=t=tahrīk</i>	C ₁ aC ₂ aC ₃

Tab. 2.9 : Patrons pour la construction du lemme avec variation des voyelles

La figure 2.19 illustre les résultats des extractions des trois entrées lexicales étudiées dans l'exemple 26.

Une complexité à signaler est dû au fait que les deux premiers lemmes, *tilb* et *talib*, sont des adjectif et le troisième lemme, *talab*, est un nom. La complication est due à l'absence de marqueurs ou d'informations textuelles qui peuvent l'indiquer. Dans ce cas, le problème ne peut être résolu automatiquement et exige une révision manuelle des résultats obtenus.


```

<lexical_entry>
  <plain-text>الثَلْبُ، بالكسر : الجَمَلُ تَكْسَرَتْ أَنْيَابُهُ هَرَمًا وَتَنَاطَرَ هُلْبُ ذَنَبِهِ، ج. أَثْلَابٌ وَثَلْبَةٌ، كَقَرْدَةٍ، وَهِيَ بِهَاءٍ
  وَالشَّيْخُ، وَالْبَعِيرُ لَمْ يُنْفَخِ، وَصَحَابِيٌّ
  </plain-text> [أَوْ هُوَ بِالثَّاءِ، وَتَقَدَّمَ]
  <lemma>ثَلْبٌ</lemma>
  <pos>ADJ</pos>
  <plural>أَثْلَابٌ</plural>
  <plural>ثَلْبَةٌ</plural>
  <feminine>وَهِيَ بِهَاءٍ</feminine>
  <sense>الجَمَلُ تَكْسَرَتْ أَنْيَابُهُ هَرَمًا وَتَنَاطَرَ هُلْبُ ذَنَبِهِ</sense>
  <sense>شَيْخٌ</sense>
  <sense>الْبَعِيرُ لَمْ يُنْفَخِ</sense>
  <sense>صَحَابِيٌّ</sense>
  <root>ث ل ب</root>
</lexical_entry>
<lexical_entry>
  <plain-text>كَتِفٌ،</plain-text> كَتِفٌ: الْمُتَثَلِّمُ مِنَ الرَّمَاحِ
  <lemma>ثَلْبٌ</lemma>
  <pos>ADJ</pos>
  <sense>مُتَثَلِّمٌ</sense>
  <root>ث ل ب</root>
</lexical_entry>
<lexical_entry>
  <plain-text>بِالتَّخْرِيكِ: التَّقْبِضُ</plain-text>
  <lemma>ثَلْبٌ</lemma>
  <pos>Noun</pos>
  <sense>تَقْبِضٌ</sense>
  <root>ث ل ب</root>
</lexical_entry>

```

Fig. 2.19 : Structuration des entrées lexicales *tilb*, *ṭalib* et *ṭalab* .

D'autres patrons linguistiques ont été identifiés et permettent également de générer le lemme à partir de celui qui le précède.

Dans certains cas, l'entrée lexicale commence par le syntagme *bi=hā'* "avec [le suffixe féminin] *tā'*" ou le syntagme, *bi=l=lām* "avec [l'article défini] *al'*". Le tableau 2.10 illustre les directives pour construire le lemme.

Dans le cas où le système trouve, en tête de l'entrée lexicale, le syntagme *bi=hā'* "avec *tā'*", il récupère le lemme qui précède, et il lui ajoute le suffixe féminin, *tā'*, pour générer le nouveau lemme. Par contre, s'il trouve le syntagme *bi=l=lām* "avec *al'*", il génère le nouveau lemme en ajoutant l'article défini, *al*, au lemme précédent.

Lemme précédent	Patron linguistique	Lemme généré
$C_1v_1?C_2v_2?C_3$	<i>bi=hā'</i> "avec le suffixe féminin"	$C_1v_1?C_2v_2?C_3$ ah
	<i>bi=l=lām</i> "avec l'article défini"	al = $C_1v_1?C_2v_2?C_3$

Tab. 2.10 : Patrons pour la construction du lemme avec ajout d'affixes

L'exemple 27 illustre une autre structure, également, complexe. Dans ce cas, aussi, l'auteur a eu recours à des conventions lexicographiques qui peuvent être identifiées et utilisées comme patrons linguistiques.

Exemple 27

وَالْكُفَّةُ ، بِالضَّمِّ ، أَوْ الْقَهْمَةِ ، أَوْ آدْهُمَةَ ، أَوْ غُبْرَةَ مُشْرَبَةً سَوَادًا ، [أَوْ حَاصَّ بِالْإِيلِ] ،
وَالْفِعْلُ كَكَرَّمَ وَفَرِحَ ، وَهُوَ أَكْهَبُ وَكَأْهَبُ

"**et=la=kuhbah**, avec la voyelle u : la couleur trouble, ou la noirceur, ou une poussière imprégnée de noirceur [ou il concerne seulement les chameaux], **et le verbe** [est] comme=*karuma* et=*fariḥa*, **et il** [est] '*akhabu* et=*kāhibuN*".

Après la parole-vedette *wa=al=kuhbah* et ses sens qui expriment des couleurs, "la couleur trouble, ou la noirceur", l'auteur introduit le verbe dont le sens est 'être trouble, ou être noir ou être imprégné de noirceur', en respectant des conventions lexicographiques, qu'on trouve d'ailleurs dans plusieurs lexiques médiévaux. Le verbe est introduit par une structure commençant par *وَالْفِعْلُ* *wa=al=fi'lu* "et le verbe [est]", suivi par un verbe modèle introduit par la conjonction *ka* "comme", ce qui déclenche les procédures déjà expliquées dans 2.2.2.1.3. Dans d'autres cas, la structure *wa=al=fi'lu* est suivie, directement, par le verbe lui-même. Étant un verbe exprimant une couleur, il est accompagné par les adjectifs correspondants introduits par *وَهُوَ* "et il [est]", ce qui déclenche les procédures expliquées dans 2.2.2.2.3, pour extraire l'adjectif. La figure 2.20 montre les résultats des procédures mises en jeu pour les extractions et la structuration de l'exemple 27.

2.2.4 Structuration et enrichissement des entrées adjectivales

Dans les cas étudiés précédemment, la parole-vedette, nom ou adjectif, est localisée en tête de l'entrée lexicale, et elle est toujours accompagnée de l'article défini, ce qui nécessite les révisions manuelles pour distinguer les noms et les adjectifs.

```

<lexical_entry>
<plain-text></plain-text>
<lemma></lemma>
<POS>Noun</POS>
<Verb></Verb>
<Verb></Verb>
<Adjective></Adjective>
<Adjective></Adjective>
<sense></sense>
<sense></sense>
<sense></sense>
<root></root>
</lexical_entry>

```

Fig. 2.20 : Structuration de l'entrée lexicale *al=kuhbah*

Dans certains cas, l'adjectif est introduit dans un exemple simple sous forme de syntagme nominale, comme le montre l'exemple 28.

Exemple 28

إِبِلٌ مُخْبَعَةٌ ، بِالْفَتْحِ : كَثِيرَةٌ ، أَوْ سَمِينَةٌ ، حَسَنَةٌ كُلُّ مَنْ رَأَاهَا قَالَ مَا أَحْسَنَهَا

”[des] chameaux *muḥabḥabatun*, avec la voyelle a : nombreuses, ou grasses, belles, toute personne qui les voit s'exclame qu'elles sont belles”

L'analyse morphosyntaxique du syntagme est illustré dans le tableau 2.11. On note que les deux composants du syntagme sont indéfinis et finissent avec le suffixe du *tanwīn* nominatif, *un*, et l'adjectif occupe la deuxième position.

<i>'ibil-u-n</i>	<i>muḥabḥab-at-u-n</i>
chameaux-NOM-INDEF	<i>muḥabḥab</i> -FEM-NOM-INDEF
substantif-NOM-INDEF	adjectif-FEM-NOM-INDEF

Tab. 2.11 : Analyse morphosyntaxique du syntagme de l'exemple 28

L'absence de l'article, au début de l'entrée lexicale, déclenche les procédures qui permettent de : i) sélectionner la deuxième parole de l'entrée ; ii) appliquer les procédures de light stemming et enlever le préfixe féminin et le diacritique *un* (dans le cas où il est digitalisé) ; iii) contrôler que la parole qui reste contient les consonnes radicales de la famille dérivationnelle ; en enfin baliser le lemme comme adjectif (ADJ).

La figure 2.21 illustre les résultats obtenus de la structuration de l'exemple 28.

```

<lexical_entry>
<plain-text>إِبِلٌ مُخَبَّبَةٌ، (بِالْفَتْحِ) : كَثِيرَةٌ، أَوْ سَمِينَةٌ، حَسَنَةٌ
[كَلَّ مَنْ رَأَاهَا قَالَ "مَا أَحْسَنُهَا"]</plain-text>
<lemma>مُخَبَّبٌ</lemma>
<pos>ADJ</pos>
<sense>كَثِيرٌ</sense>
<sense>سَمِينٌ</sense>
<sense>حَسَنٌ</sense>
<root>خ ب ب</root>
</lexical_entry>

```

Fig. 2.21 : Structuration de l'entrée lexicale de l'adjectif *muḥabḥb*

2.2.5 Traitement des définitions

En se basant sur la nature de la définition, déjà expliquée dans la section 1.3.3, on arrive, dans plusieurs cas, à déduire la nature qui peut lier le sens au lemme de l'entrée lexicale.

2.2.5.1 Cas où la définition est constituée d'un seul mot

2.2.5.1.1 Cas des entrées lexicales verbales L'exemple 29 illustre le cas où le bloc des définitions, qui se trouve après les deux points, est constitué de deux paroles séparées par une virgule et donc correspondent à deux sens différents, 'āb-a=hu 'il l'a désapprouvé' et wa=lām-a=hu 'et il l'a blâmé'. Pour récupérer les lemmes correspondants, 'āb-a 'désapprouver' et lām-a 'blâmer', une phase de « light stemming » est nécessaire.⁷.

Exemple 29

أَشَبَ فُلَانًا، يَأْشِبُهُ، وَيَأْشِبُهُ : عَابَهُ، وَلَا مَهُ

'ašaba fulān-an, ya`šib-u=hu, wa=ya`šab-u=hu : 'āb-a=hu, wa=lām-a=hu

"[II] 'ašaba quelqu'un, ya`šibuhu, waya`šabuhu : il l'a désapprouvé, et il l'a blâmé

Comme illustré dans la figure 2.22, lors de la génération du format XML, le tag <sense> explicite la définition comme elle est formulée dans *qāmūs*. Les définitions, étant composées d'une seule parole, permettent de déduire que les lemmes générés, 'āb-a et lām-a, sont synonymes du lemme de l'entrée lexicale 'ašaba. Un nouveau type de tag

7. Concernant cette phase, voir 2.2.2.1.2.

<synonym> est ajouté automatiquement pour expliciter la relation entre le sens et le lemme de l'entrée lexicale.

```
<lexical_entry>
  <plain_text>أَشَبَ فلاناً، يَأشِبُه ويَأشُبُه: عابِه ، ولامه</plain_text>
  <lemma>أَشَبَ</lemma>
  <imperfective>يَأشِبُ</imperfective>
  <imperfective>يَأشُبُ</imperfective>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <root_information>T</root_information>
  <morphosyntactic_information>Transitive</morphosyntactic_information>
  <external_links />
  <sense att="عابِه">
  <synonym att="عَاب">
  </synonym>
  </sense>
  <sense att="ولامه">
  <synonym att="لَام">
  </synonym>
  </sense>
  <root>ء ش ب</root>
</lexical_entry>
```

Fig. 2.22 : Traitement des définitions synonymes du verbe 'ašaba de l'exemple 29.

2.2.5.1.2 Cas des entrée lexicales nominales La parole-vedette, *al=ġanāb-u*, dans l'exemple 30, est considérée un nom car elle est définie avec l'article et se termine avec la voyelle de déclinaison *u* du nominatif. Les sens qui l'accompagnent sont constitués de paroles isolées et séparées entre elles par des virgules.

Exemple 30

الْجَنَابُ : الْفِنَاءُ، وَالرَّحْلُ، وَالنَّاحِيَةُ، وَجَبَلٌ، وَعَلَمٌ

"*al=ġanābu* : le patio, et le baggage, et la direction, et [le nom de] une montagne, et un nom propre".

Les trois premières définitions sont constituées de paroles isolées, également, définies qui se terminent avec *u* et correspondent à des sens communs. On peut interpréter l'entrée lexicale :

الْجَنَابُ *al=ġanāb-u* : الْفِنَاءُ *al=finā'-u*, وَالرَّحْلُ *wa=r=raḥl-u*, وَالتَّاحِيَةُ *wa=n=nāḥiyat-u*

الْجَنَابُ *al=ġanāb-u* : [c'est] "le patio", et [c'est] "le bagage", et [c'est] "la direction".

Dans ce cas, la parole-vedette et les sens doivent subir les procédures de «light stemming» pour obtenir les lemmes correspondants.

Constituée d'une seule parole, le lemme généré de chaque définition est considéré synonyme du lemme extrait de l'entrée lexicale. La figure 2.23 représente le résultat des extractions des sens communs de l'exemple 30.

```
<lexical_entry>
<plain-text>الْجَنَابُ: الْفِنَاءُ، وَالرَّحْلُ، وَالتَّاحِيَةُ، وَجِبِلٌّ، وَعَلَمٌ</plain-text>
<lemma>جَنَاب</lemma>
<pos>Noun</pos>
<sense att="الْفِنَاءُ">
<synonym att="فِنَاء">
</synonym>
</sense>
<sense att="الرَّحْلُ">
<synonym att="رَحْل">
</synonym>
</sense>
<sense att="التَّاحِيَةُ">
<synonym att="تَّاحِيَةُ">
</synonym>
</sense>
<root>ج ن ب</root>
</lexical_entry>
```

Fig. 2.23 : Extraction des sens communs de l'exemple 30.

Par contre, les deux dernières définitions sont indéfinies et se terminent avec le suffixe du *tanwīn* nominatif, *un*.

La définition, وَجَبَلٌ, *wa=ġabal-un* "et une montagne", ne peut pas signifier "et [c'est] la montagne", mais elle sous-tend "et [c'est le nom propre de] une montagne".

La dernière définition, وَعَلَمٌ, *wa='alam-un*, est plus explicite, parce que la parole '*alam* signifie 'nom propre de personne'. De conséquence, elle sous-tend "et [c'est] un nom propre de personne".

Dans le cas où la définition est indéfinie et se termine avec le suffixe du *tanwīn* nominatif, *un*, elle indique un nom propre. Les procédure « light stemming » de la parole-vedette consistent, dans ce cas, à éliminer seulement les éventuelles conjonctions, prépositions et la dernière voyelle qui marque le cas nominatif. Le lemme de l'entrée lexicale doit garder l'article défini : *al=ġanāb*. En plus, le système doit créer des entrées lexicales différentes pour les noms communs où le lemme est *ġanāb*, et pour les noms propres, où le lemme est *al=ġanāb*.

Il faut noter que la structuration serait meilleure si on explicite le concept auquel peut être lié chaque nom propre. Dans ce but, les concepts correspondants ont été récupérés d'une ontologie existante, SUMO (The Suggested Upper Merged Ontology) qu'on définira, plus en détail, dans la section suivante 2.3. La figure 2.24 est la représentation des noms propres extraits de l'exemple 30.

```

<lexical_entry>
<plain-text>الْجَنَابُ: الْفِنَاءُ، وَالرَّحْلُ، وَالنَّاجِيَةُ، وَجَبَلٌ، وَعَلَمٌ</plain-text>
<lemma>الْجَنَاب</lemma>
<POS>ProperName</POS>
<sense>جَبَل</sense>
<Concept>Mountain@</Concept>
<root>ج ن ب</root>
</lexical_entry>
<lexical_entry>
<plain-text>الْجَنَابُ: الْفِنَاءُ، وَالرَّحْلُ، وَالنَّاجِيَةُ، وَجَبَلٌ، وَعَلَمٌ</plain-text>
<lemma>الْجَنَاب</lemma>
<POS>ProperName</POS>
<sense>عَلَم</sense>
<Concept>Man@</Concept>
<root>ج ن ب</root>
</lexical_entry>

```

Fig. 2.24 : Enrichissement sémantique des noms propres de l'exemple 30

Les sens *ġabal* et '*alam* sont accompagnés d'une étiquette spécifiant, respectivement, les concepts, «Mountain » et «Man », pour indiquer que le premier est le nom

propre d'une montagne et le deuxième celui d'une personne.

Nota Dans l'ontologie SUMO, le symbole @ signifie que le lemme identifié est une instance et donc un individu faisant partie du concept correspondant.

L'étude de *qāmūs* révèle qu'il est très riche de noms propres de personne, cheval, régions, pays, montagnes, fleuves, châteaux et citadelles, etc. En relevant les patrons linguistiques qui permettent de les identifier, on arrive à enrichir plusieurs lemmes avec des informations sémantiques plus détaillées. Cette phase de la recherche sera expliquée, avec plus de détails, dans la section suivante 2.3.

2.2.5.1.3 Cas particuliers Dans certains cas, le sens est introduit par la particule *ka* "comme" qui indique que c'est une variante lexicale. Le système, dans ce cas, lance les procédures où il contrôle que la première consonne radicale n'est pas la lettre *k*, comme dans le verbe *kataba* "écrire". Dans le cas où elle ne l'est pas, il enlève la particule *k* ou *ka*, considérée conjonction "comme". Après avoir contrôlé que la parole restante contient les mêmes radicales du lemme, il crée l'étiquette pour spécifier que le sens est une variante synonyme du lemme. La figure 2.25 en illustre un exemple où le lemme de l'entrée lexicale *rabba* a deux définitions, *ṭayyab-a=hu* et *ka=rabbab-a=hu*.

```

<lexical_entry>
<plain_text>رَبُّ الدَّهْنِ: طَيِّبُهُ، كَرَبِّيهِ</plain_text>
<lemma>رَبُّ</lemma>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<morphosyntactic_information>Transitive</morphosyntactic_information>
<sense att="طَيِّبُهُ">
<synonym att="طَيِّب">
</synonym>
</sense>
<sense att="كَرَبِّيهِ">
<synonym att="رَبَّب">
</synonym>
</sense>
<root>ر ب ب</root>
</lexical_entry>

```

Fig. 2.25 : Cas particulier des variantes lexicales

- la première définition, *ṭayyab-a=hu* 'il l'a aromatisé', dont le lemme *ṭayyaba* 'aromatiser' est synonyme du lemme de l'entrée lexicale *rabb-a*;

- le deuxième définition, *ka=rabbab-a=hu* qui est précédé de la conjonction *ka* et dont le lemme *rabbaba* est une variante considérée synonyme du lemme de l'entrée lexicale *rabba*.

2.2.5.2 Cas où la définition est constituée par plusieurs paroles

2.2.5.2.1 Cas où la définition est un antonyme En analysant l'exemple 31, on note que la définition est constituée de la parole *ḍiddu* qui signifie 'contraire' et du verbe *ḡamada* 'geler'. On peut conclure que le lemme *ḡāba* 'fondre' est antonyme du verbe *ḡamada* 'geler'.

Exemple 31

ذَابَ ذُوبًا وَذُوبَانًا : ضِدُّ جَمَدٍ

"*ḡāba* ḡawban waḡawabānan : **contraire** [de] geler

Dans le cas où le pattern linguistique *ḍiddu* ou son synonyme '*aksu*, signifiant 'contraire', se trouve juste après les deux points ou juste après une virgule, le système doit créer une balise spécifiant la relation antonymique entre le lemme et le sens. Cette règle est applicable à toutes les entrées lexicales, puisque les patterns linguistiques *ḍiddu* et '*aksu* sont compatibles aussi bien avec les verbes qu'avec les noms et les adjectif. La figure 2.26 illustre les résultats des extractions de l'exemple 31, où le sens *ḡamada* 'geler' est spécifié antonyme du lemme *ḡāba* 'fondre'.

```
<lexical_entry>
  <plain_text>ذَابَ ذُوبًا وَذُوبَانًا، مَحْرَكَةً: ضِدُّ جَمَدٍ</plain_text>
  <lemma>ذَابَ</lemma>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <masdar>ذُوبَ</masdar>
  <masdar>ذُوبَانَ</masdar>
  <sense att="ضِدُّ جَمَدٍ">
    <antonym att="جَمَدٍ">
  </antonym>
  </sense>
</lexical_entry>
```

Fig. 2.26 : Cas de relation antonymique entre lemme et sens

Cependant, dans d'autre cas, comme dans l'exemple 32, on ne trouve pas la parole *ḍiddu* 'contraire' juste après les deux points, mais en fin de ligne et après deux sens différents séparés entre eux par une virgule.

Exemple 32

خَشَبَهُ يَخْشِبُهُ : خَلَطَهُ، وَانْتَقَاهُ، ضِدُّ

"*ḥašabahu* ḥšibuhu : il l'a mélangé, et il l'a trié, **contraire**

Dans ce cas, la relation antonymique n'est pas entre lemme et sens mais entre les sens qui précèdent la parole *ḍiddu*.

Le lemme *ḥašaba* est synonyme du premier sens *ḥalaṭa* 'mélanger' et du deuxième sens *intaqā* 'trier'. Par contre les deux sens *ḥalaṭa* 'mélanger' et *intaqā* 'trier' sont antonymes.

En conclusion, dans le cas où la parole *ḍiddu* 'contraire' se trouve en fin de ligne et après deux sens différents séparés entre eux par une virgule. Le système ne doit pas la prendre en considération. La figure 2.27 montre le résultats des extractions de l'exemple 32.

```
<lexical_entry>
  <plain_text>خَشَبَهُ يَخْشِبُهُ : خَلَطَهُ، وَانْتَقَاهُ، ضِدُّ</plain_text>
  <lemma>خَشَبَ</lemma>
  <imperfective>يَخْشِبُ</imperfective>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <morphosyntactic_information>Transitive</morphosyntactic_information>
  <sense att="خَلَطَهُ">
  <synonym att="خَلَطَ">
  </synonym>
  </sense>
  <sense att="وانْتَقَاهُ">
  <synonym att="انْتَقَى">
  </synonym>
  </sense>
  <root>خ ش ب</root>
</lexical_entry>
```

Fig. 2.27 : Cas de relation antonymique entre les sens de l'entrée lexicale

Dans le cas des verbes, on identifie un autre type de structure où la définition est sous forme de négation du sens antonymique, comme dans l'exemple 33.

Exemple 33

رَتَّبَ رُتُوبًا : ثَبَّتَ، وَلَمْ يَتَحَرَّكَ، كَثَّرَتَّبَ

”*rataba rutūban rutūban* : il s’est fermé, et il n’a pas bougé, comme *tarat-taba*

Le premier sens *tabata* ‘il s’est fermé’ est synonyme puisque c’est une parole isolé. Par contre le deuxième sens *lam yataḥarrak* ‘il n’a pas bougé’ est constitué de la particule négative *lam* qui signifie ‘ne...pas’, et qui est compatible uniquement avec le verbe qui, en sa présence, va à l’apocopé *yataḥarrak*. Le troisième sens est précédé de la conjonction *ka* ‘comme’ et peut être considéré une variante synonyme.

Concernant le deuxième sens, en détectant la particule négative *lam*, le système mémorise seulement la forme fléchie *yataḥarrak* qui la suit et génère le lemme correspondant *taḥarraka*, après une phase de « light stemming ». Enfin, il ajoute, automatiquement, l’étiquette qui spécifie la relation antonymique. La figure 2.28 présente le résultat des extractions de l’exemple 33.

```
<lexical_entry>
<plain_text>رَتَّبَ رُتُوبًا : ثَبَّتَ، وَلَمْ يَتَحَرَّكَ، كَثَّرَتَّبَ</plain_text>
<lemma>رَتَّبَ</lemma>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<masdar>رُتُوب</masdar>
<sense att="ثَبَّت">
<synonym att="ثَبَّت">
</synonym>
</sense>
<sense att="وَلَمْ يَتَحَرَّكَ">
<antonym att="تَحَرَّكَ">
</antonym>
</sense>
<sense att="كَثَّرَتَّب">
<synonym att="رَتَّب">
</synonym>
</sense>
<root>ر ت ب</root>
</lexical_entry>
```

Fig. 2.28 : Cas de la relation antonymique entre verbes

Dans le cas des adjectifs, l'exemple 34 présente le cas où la définition, *ġayru muḥ-kamin* 'non élaboré', commence par la particule négative *ġayru* 'non' compatible seulement avec les formes nominales.

Exemple 34

أَمْرٌ مُعْتَلِبٌ بِالْكَسْرِ : غَيْرٌ مُحْكَمٌ

"[une] chose *mu'atlabun* : non élaboré

La particule négative *ġayru* 'non' constitue un pattern linguistique dont la détection permet au système, après l'avoir enlevée, de créer une étiquette spécifiant que la parole qui la suit est antonyme du lemme de l'entrée lexicale. La figure montre le résultat des extractions de l'exemple 34.

```
<lexical_entry>
  <plain_text>أَمْرٌ مُعْتَلِبٌ، بِالْكَسْرِ : غَيْرٌ مُحْكَمٌ</plain_text>
  <lemma>مُعْتَلِبٌ</lemma>
  <pos>ADJ</pos>
  <lemma_translate att="مُعْتَلِب" value="null" />
  <sense att="غَيْرٌ مُحْكَم" >
    <antonym att="مُحْكَم" >
      </antonym>
    </sense>
  <root>ع ث ل ب</root>
</lexical_entry>
```

Fig. 2.29 : Cas particulier de la relation antonymique entre adjectifs

2.2.5.2.2 Cas où la définition est explicative Dans certains cas, la définition est détaillée ou elle ne contient aucune structure linguistique qui permet d'automatiser les extractions. De conséquence, l'identification de la relation entre lemme et sens exige un travail manuel.

Dans l'exemple 35, la relation entre le lemme *ġabġaba* et le sens *sāḥa* 'se déverser' peut être considérée synonymique.

Exemple 35

جَبَجَبَ : سَاحَ فِي الْأَرْضِ

"il *ġabġaba* : il s'est déversé par terre"

Dans l'exemple 36, le lemme *ḡar'aba* 'boire quelque chose complètement/entièrement' peut être considéré un troponyme du verbe *šariba* 'boire'.

Exemple 36

جَزَعَبَ الْمَاءَ : شَرِبَهُ جَيِّدًا

"il *ḡar'aba* l'eau : il l'a bu complètement"

Dans l'exemple 37, le nom *'itb* peut être 1) un vêtement court à mi-jambe ; 2) un pantalon court ou 3) une chemise sans manches. Dans les trois cas, le lemme *'itb* peut être considéré hyponyme de : 1) vêtement ; 2) pantalon et 3) chemise.

Exemple 37

أَلْتَبُ ... : ... ، مَا قَصَرَ مِنَ التِّيَابِ فَتَصَفَّ السَّاقَ ، أَوْ سَرَاوِيلُ بِلَا رِجْلَيْنِ ، أَوْ قَمِيصٌ بِلَا كُمَيْنِ

"il *al='itbu* ... : ... , les vêtements courts à mi-jambe, ou les pantalons sans jambes, ou une chemise sans manches"

Les exemples 38 et 39, par contre, sont seulement explicatives.

Exemple 38

أَخْلَبَ : وَلَدَتْ إِيْلَهُ إِنَاثًا

"[un homme] *'ahlaba* : ses chamelles ont donné naissance à des femelles.

Exemple 39

أَجْلَبَ : وَلَدَتْ إِيْلَهُ ذُكُورًا

"[un homme] *'aḡlaba* : ses chamelles ont donné naissance à des mâles.

Dans l'exemple 38, on dit qu'un homme *'ahlaba* lorsque ses chamelles donnent naissance à des femelles. Il est nécessaire, dans ce cas, rappeler que le verbe *'ahlaba* est la forme IV dérivant du verbe *ḡalaba* 'traire'. De la racine *ḡlb*, dérive, également, la parole *ḡalīb* 'lait'. Le verbe *'ahlaba* peut avoir le sens de 'traire beaucoup de lait'

Par contre, dans l'exemple 39, on dit que «quelqu'un *'aḡlaba*», lorsque ses chamelles donnent naissance à des mâles. La forme IV *'aḡlaba* est la forme IV dérivée du verbe *ḡalaba* 'apporter'. Le lemme *'aḡlaba* peut être compris comme 'faire porter [le lait]'

2.2.6 Quelques résultats obtenus

Grâce aux patrons identifiés et les regex construites, on a pu identifier 4912 entrées lexicales extraites de la section de la lettre *bā'*, réparties selon leurs catégories grammaticales, comme le démontre le tableau 2.12. Le tableau 2.13, par contre, donne un

aperçu général sur les verbes extraits dans la section de la lettre *bā'*. On a pu contrôler, manuellement, 479 noms, 511 verbes et 200 adjectifs. De conséquence, il convient de signaler que les chiffres, présentés dans les différents tableaux, sont de nature indicatives, jusqu'à ce que le contrôle manuel ne soit terminé.

POS	nombre total
Verbes	1314
Noms	2149
Adjectifs	368
Noms propres	966
Noms de genre et d'espèce	165

Tab. 2.12 : Résultats des extractions de la section de la lettre *bā'*

Verbe	dérivant de racine trilitère									
	base	dérivées								
	I	II	III	IV	V	VI	VII	VIII	X	XV
Intransitif	449	43	10	139	61	20		37	24	6
Transitif	203	58	28	52	20	1	4	7	17	-
Total	652	101	38	191	81	20	15	44	41	6

Verbe	dérivant de racine quadrilitère		
	base	dérivées	
	I	II	IV
Intransitif	38	20	19
Transitif	38	1	-
Total	76	29	19

Tab. 2.13 : Statistiques des verbes contenus dans la lettre *bā'*

En plus, on a pu identifier 93 phrases idiomatiques, comme l'exemple 40.

Exemple 40

مِنْ شُبِّ إِلَى دُبِّ [...] : مِنَ الشَّبَابِ إِلَى أَنْ دَبَّ عَلَى الْعَصَا.

min šubba 'ilā dubba : de la jeunesse jusqu'à ce qu'il rampe avec un bâton

Comme déjà dit, selon l'organisation du lexique d'époque antérieure, l'entrée lexicale fait référence à une parole-vedette qui peut être le lemme ou une de ses formes fléchies. Plusieurs entrées lexicales, dans ce cas, font référence au même lemme, avec des significations différentes. Lors de la construction de la version « moderne » de *qāmūs*, une ultérieure phase de contrôle doit suivre pour réunir tous les sens à un seul lemme.

2.3 Enrichissement sémantique de *qāmūs*

2.3.1 Procédures d'extractions automatiques

Le lexique *qāmūs* est très riche de noms propres de personne, cheval, régions, pays, montagnes, fleuves, châteaux et citadelles, etc. Comme déjà expliqué dans le paragraphe 2.2.5.1.2, en relevant les structures linguistiques qui permettent de les identifier, le système est capable, automatiquement, de leur créer une entrée lexicale et les baliser comme noms propres. En plus, en identifiant les concepts auxquels ils peuvent appartenir, le système est capable de créer une balise spécifiant le concept auquel peut être liée le lemme de l'entrée lexicale.

Les tableaux 2.14 et 2.15 illustrent les structures relevées dans la section de la lettre *bā'*, pour identifier, respectivement, les noms propres de personne et les noms propres d'animal, régions, ect.

La structure linguistique peut être composée d'une seule parole **indéfinie au nominatif** qui termine avec le suffixe *un*, ce qui sous-tend [c'est le nom propre de]. Par exemple :

رَجُلٌ *rağul-un* sous-tend "[c'est le nom propre de] un homme".

Elle peut être sous forme d'un **syntagme** combinant une parole définie qui se termine avec le suffixe *u* du nominatif, et une autre indéfinie qui se termine avec le *tanwīn* du génitif, *in*. Par exemple :

إِسْمُ رَجُلٍ *ism-u rağul-in* sous-tend "[c'est] le nom [propre] d'un homme".

صَحَابِيٌّ	"un Compagnon"	عَلَمٌ	"un nom propre"
صَحَابِيُّونَ	"des Compagnons"	إِسْمٌ	"un nom [propre]"
رَجُلٌ	"un homme"	أَسْمَاءٌ	"des noms [propres]"
إِمْرَأَةٌ	"une femme"	مُحَدِّثُونَ	"des narrateurs de ḥadīṭ"
إِسْمُ إِمْرَأَةٍ	"le nom d'une femme"	مُحَدِّثٌ	"un narrateur de ḥadīṭ"
إِسْمُ رَجُلٍ	"le nom d'un homme"	شَاعِرٌ	"un poète"
جَدُّ ...	"le grand-père de ..."	أَخْبَارِيٌّ	'un narrateur'

Tab. 2.14 : Structures linguistiques identifiant les noms propres de personne

Patrons sémantiques	Nom Propre de	Patrons sémantiques	Nom Propre de
ع - مَوْضِعٌ	lieu	مَحَلَّةٌ - حَيٌّ	quartier
د - بَلَدٌ	pays	وَادٍ	fleuve
ة - قَرْيَةٌ	village	بَيْرٌ - مَاءٌ	source/puits
جَبَلٌ	montagne	نَجْمٌ	étoile
قَبَائِلٌ - قَبِيلَةٌ	tribu	قَصْرٌ	château
بُطُونٌ - بَطْنٌ		حِصْنٌ	citadelle
صَمٌّ	sculpture	فَرَسٌ ...	le cheval de ...

Tab. 2.15 : Structures linguistiques identifiant d'autres noms propres

Les concepts, correspondant à chaque structure, ont été récupérés de l'ontologie SUMO (The Suggested Upper Merged Ontology) qu'on définira, plus en détail, dans la section 2.3.4.3. Le tableau 2.16 présente un aperçu des noms propres, les concepts auxquels ils peuvent être liés et leur total détecté dans la section *bā'*.

Nom propre de	Concept SUMO	Total
personne	Humain@	365
tribus	Group@	40
quartier - pays - village	Neighborhood@	394
	Region@	
	Nation@	
	City@	
montagne	Mountain@	49
fleuve	River@	30
source	Creek@	37
puits	MineOrWell@	6
étoile	Star@	4
château	Building@	3
citadelle	StationaryArtifact@	13
sculpture	Sculpture@	1
cheval	Hors@	25

Le symbole @ dans SUMO signifie que le lemme identifié est une instance ou un individu faisant partie du concept correspondant.

Tab. 2.16 : Concepts correspondants aux noms propres

D'autres patrons permettent, avec les même procédures, d'identifier les noms d'espèce et de genre et de spécifier la relation d'hyponymie entre le signifié et le signifiant. Par exemple, شَجَرٌ *šağarun* sous-tend "[c'est] un [genre de] arbre". Le tableau 2.17 donne un aperçu sur les patrons relevés dans la section de la lettre *b*.

Patron sémantique	Genre de	Patron sémantique	Genre de
شَجَرَةٌ - شَجَرٌ	arbre	شُجَيْرَةٌ	arbuste
نَخْلٌ	palmier	نَبْتٌ - نَبَاتٌ - عُشْبَةٌ	plante
ثَمْرٌ	dattes	طَعَامٌ	plat de nourriture
الْخَمْرُ	vin	دَاءٌ	maladie
اللَّبَنُ	lait et ses dérivés	دَوَاءٌ	médicament
حَجَرٌ	pierre	سَمَكٌ	poisson
طَائِرٌ	oiseau		

Tab. 2.17 : Structures linguistiques identifiant les relations d'espèce et de genre

Par contre, le tableau 2.18 présente les concepts correspondants aux structure relevées et leur total détecté dans la section *bā*'.

Noms d'espèce	Concept SUMO	Total
arbre - palmier	BotanicalTre+	54
arbuste - plante	Plant+	
Aliments	DateFruit+	21
	Wine+	
	Milk+	
	Food+	
animaux	Bird+	89
	canine+	
	Fish+	
pierre	Stone+	1

+ : symbole de SUMO pour signifier que le lemme identifié est subsumé par le concept correspondant.

Tab. 2.18 : Concepts correspondants aux noms d'espèce et de genre

Cependant, les structures qui permettent de détecter, automatiquement, le concept auquel peut être lié un lemme donné, concernent seulement les noms propres et les noms de genre et d'espèce. La couverture manuelle de toute la ressource se révèle une opération très couteuse et complexe. Pour gagner du temps et avoir une couverture plus large de toute la ressource, il est fondamental de bénéficier des réseaux sé-

mantiques déjà développés. Dans ce contexte d'élaboration, la WordNet de la langue anglaise (PWN), qualifiée d'ontologie légère, est un standard utilisé dans plusieurs applications, entre autres, comme pivot pour l'alignement avec des ressources lexicales de langues différentes. En plus, la PWN a été complètement aligné à l'ontologie SUMO (The Suggested Upper Merged Ontology).

2.3.2 Procédures d'alignement

Afin d'enrichir automatiquement les lemmes de *qāmūs* avec les concepts et les informations sémantiques qu'ils détiennent, il faudrait projeter les concepts et les relations sémantiques de la PWN vers la ressource arabe. Par l'intermédiaire de la PWN, la ressource *qāmūs* sera liée, automatiquement, à l'ontologie SUMO. La figure 2.30 résume les procédures mises en jeu.

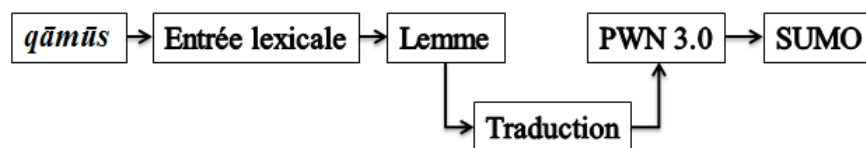


Fig. 2.30 : Procédures de mapping

Une étape primaire est celle de traduire le lemme (et ses sens, dans le cas nécessaire), ce qui permet de mettre en correspondance (mapping) le lemme de *qāmūs* avec son/ses équivalents dans PWN. Bien que la version actuelle de la PWN soit 3.1, dans ce travail, on utilise la version 3.0, parce qu'elle est entièrement alignée avec SUMO. Par le biais des identificateurs trouvés dans la PWN 3.0, le lemme arabe est mis en correspondance avec le/les concepts de SUMO.

Pour la traduction automatique, une approche simple est d'utiliser un dictionnaire bilingue et chercher, pour chaque terme arabe, la liste de ses traductions possibles. Deux dictionnaires bilingues arabe-anglais sont disponibles en format numérisé, «An Arabic-English Dictionary » (dorénavant AED)⁸ et «A Dictionary of Modern Written Arabic » (dorénavant DMWA)⁹.

8. Salmoné, 1889

9. Wehr, 1971

2.3.3 Ressources lexicales bilingues utilisées

2.3.3.1 «An Arabic-English Dictionary »

Compilé par H. Anthony Salmoné, il a été publié en 1890. L'auteur a précisé que, pour la compilation de son dictionnaire, il a utilisé, entre autres, «An Arabic English Lexicon » de Edward William Lane, lui-même basé sur *tāğ al=‘arūs min ġawāhir al=qāmūs* de az=zabīdī. La macrostructure suit l'ordre traditionnel de la langue arabe, en fonction de la première consonne des racines.

La version disponible est en format TEI, et la figure 2.31 en illustre un exemple. On note que le lemme est translittéré, *bajīHa*, accompagné de la forme arabe vocalisée, *بَجَحْ*. Cependant, la translittération utilisée n'est pas standard et nécessite une phase de mapping. En plus, la traduction du lemme verbal est à la 3^ePSM "Rejoiced at", ce qui nécessite la lemmatisation de la traduction anglaise pour en déduire l'infinif.



Fig. 2.31 : Entrée lexicale *bağīḥa* dans AED

2.3.3.2 «A Dictionary of Modern Written Arabic »

Compilée par Hans Wehr (1909-1981), il a été publiée en 1961. Bien que c'est un dictionnaire de la langue moderne, l'auteur précise dans son introduction :

[...] These results emerge from the very character of modern Arabic — a written language, powerfully influenced by traditional norms, which nevertheless is required to express a multitude of new foreign concepts, not for one country only, but for many distributed over a vast geographical area... Arabic phonology, morphology, and syntax have remained relatively unchanged from earliest times, as has much of the vocabulary... Here traditional adherence to ancient linguistic norms and to the models of classical literature, especially the Koran, has had the effect of preserving the language intact over the centuries.

Wehr ajoute que le néologisme affecte surtout les domaines de la technologie et des sciences et fait précéder les nouveaux termes techniques par le symbole O.

La version téléchargée, dont la figure 2.32 en illustre un extrait, est numérisée en format Unicode et a été générée à l'aide d'un système OCR. On note des erreurs typographiques ou de reconnaissance. Par exemple, la lettre *t̤* est translittéré, certaines fois, comme *t*. En outre, le signe de la *hamza* n'est signalé ni dans la forme translittérée, ni dans la parole arabe. Par exemple, le verbe *أثر* *'atara* se trouve sous forme écrite *اثر atara*. De conséquence, il a été nécessaire d'effectuer plusieurs phases de corrections et de normalisation surtout des diacritiques.

اثر atara u i (اثر, اثاره atāra) to transmit, pass along, report, relate (هـ s.th., عن from, or based on the authority of, s.o.) II to affect, influence (في or على s.o., s.th.), act (في or على upon), produce an effect, make an impression, have influence (في or على on); to induce (phys.) IV to prefer (على هـ s.th. to), like (هـ s.th.) more (على than); to have a predilection, a liking (هـ for), like (هـ s.th.), be fond (هـ of); to choose, deem wise or advisable (ان to do s.th.); to adore (هـ s.o., s.th.) | أثر نفسه بالخير (nafsahū bi-l-kair) to wish o.s. well, hope for the best for o.s. V to be impressed, be influenced; to let o.s. be impressed, be impressible; to be moved, be touched (ب or ل by, also من); to be excited, be stimulated; to be affected (ب by, said of materials, e.g., iron by acid); to be induced (phys.); to follow in s.o.'s (هـ) tracks, follow s.o.'s example, emulate (هـ s.o.); to pursue, follow up (هـ a question, a problem); to perceive, feel (هـ s.th.) X to claim a monopoly; to possess alone, with the exclusion of others, monopolize (ب s.th.); to appropriate (ب s.th.), take exclusive possession (ب of); to preoccupy (هـ s.th.), engross (هـ the attention) | استأثر الله به the Lord has taken him unto Himself.

Fig. 2.32 : Structure du texte téléchargé de DMWA

La macrostructure du dictionnaire suit l'ordre en fonction des racines. Par contre, la microstructure organise ses éléments en présentant le lemme arabe non vocalisé, suivi de sa translittération vocalisée et enfin des traductions possibles. Mais on note que les verbes dérivés sont listés, dans la même entrée du verbe de base, et présentés par le chiffre romain correspondant à leurs formes. Afin de rendre les informations plus explicites, il a fallu segmenter les entrées lexicales des verbes, en utilisant, comme marqueur, le chiffre romain des formes dérivées, et générer, automatiquement, une nouvelle entrée lexicale pour chaque forme dérivée. Grâce à des regex de génération, chaque verbe dérivé est accompagné de sa forme translittérée entre crochets.

La figure 2.33 montre le texte après les opérations de normalisation, de segmentation et de génération. Les procédures de correction et de normalisation ont été concentrées sur les formes translittérées pour le fait qu'elle sont entièrement vocalisées.

أثر
 اثر 'atara u i (atr, اثاره aṭāra) to transmit, pass along, report, relate (هـ s.th., عن from, or based on the authority of, s.o.)
 اثر II [ʿattara] to affect, influence (في or على s.o., s.th.), act (في or على upon), produce an effect, make an impression, have influence (في or على on); to induce (phys.)
 اثر IV [a'tara] to prefer (على هـ s.th. to), like (هـ s.th.) more (على than); to have a predilection, a liking (هـ for), like (هـ s.th.), be fond (هـ of); to choose, deem wise or advisable (ان to do s.th.); to adore (هـ, هـ s.o., s.th.) | أثر نفسه بالخير (nafsahū bi-l-kair) to wish o.s. well, hope for the best for o.s.
 اثر V [ta'attara] to be impressed, be influenced; to let o.s. be impressed, be impressible; to be moved, be touched (ب or ل by, also من); to be excited, be stimulated; to be affected (ب by, said of materials, e.g., iron by acid); to be induced (phys.); to follow in s.o.'s (هـ) tracks, follow s.o.'s example, emulate (هـ s.o.); to pursue, follow up (هـ a question, a problem); to perceive, feel (هـ s.th.)
 اثر X [ista'tara] to claim a monopoly; to possess alone, with the exclusion of others, monopolize (ب s.th.); to appropriate (ب s.th.), take exclusive possession (ب of); to preoccupy (هـ s.th.), engross (هـ the attention) | استأثر الله به the Lord has taken him unto Himself

Fig. 2.33 : Structure de DMWA après normalisation et génération

2.3.3.3 Analyse et comparaison de AED et DMWA

Comme illustré dans le tableau 2.19, en comparant certaines paroles connues pour avoir subi un glissement sémantique, on note que, au contraire du dictionnaire «A Dictionary of Modern Written Arabic » (dorénavant DMWA), «An Arabic-English Dictionary » (dorénavant AED) ne contient pas les correspondants anglais dont le sens est nouveau.

Parole	Sens dans <i>qāmūs</i>	DMWA	AED
<i>sayyārah</i>	caravane	automobile, car	caravan
<i>qāfilah</i>	caravane	caravan ; column ; convoy	caravan
<i>dabābah</i>	machine utilisée dans la guerre	tank, armored car	1- Testudo, musculus 2- Machine formerly used in sieges
<i>qiṭār</i>	file de chameaux bien rangée	train of camels ; (railroad) train ; railroad ;	file, row, string (of animals)

Tab. 2.19 : Traductions dans DMWA et AED de noms avec glissement de sens

Cependant, AED ne couvre pas tous les lemmes extraits, comme le démontre l'exemple de la famille dérivationnelle *bġḥ* dans le tableau 2.20.

Lemme arabe	AED	DMWA
<i>baġiḥa</i>	Rejoiced at	to rejoice
<i>baġġaḥa</i>	Gladdened	-
<i>tabaġġaḥa</i>	-	to vaunt, flaunt , boast, brag
<i>tabaġġuḥ</i>	-	bragging, braggery
<i>mutabaġġiḥ</i>	-	braggart

Tab. 2.20 : Traductions des lemmes dérivant de la racine *bġḥ* dans AED et DMWA

En plus, dans AED, les *maṣdar* sont inclus dans l'entrée lexicale verbale sans traduction. Par exemple, on n'y trouve pas les traductions de la parole *isti'dād*, alors qu'elle est traduite dans DMWA : "readiness, willingness, preparedness", *taṣallub* "hardness, callousness, hardening" et *iḥkām* "perfection ; accuracy, exactness, exactitude, precision ; exact performance, precise execution".

Pour avoir une couverture maximale, on a décidé d'utiliser les deux dictionnaires, mais les traductions de AED ont un poids de confiance plus grand. Néanmoins, du fait que ce sont des dictionnaires récents, la révision rigoureuse des traductions des lemmes extraits de *qāmūs* est essentielle.

2.3.4 Ressources ontologiques utilisées

2.3.4.1 Définition

Selon Guarino, l'ontologie computationnelle est :

An ontology is a logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such

a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models.¹⁰

La première phase requise pour l'élaboration d'une ontologie computationnelle est la conceptualisation qui consiste à identifier les concepts et les relations qui les organisent en hiérarchie. Selon les définitions de Gandon :¹¹

A concept : notion usually expressed by a term (or more generally by a sign) || the concept represents a group of objects or beings having shared characteristics that enable us to recognize them forming and belonging to this group.

A relation : notion of an association or a link between concepts usually expressed by a term or a graphical convention (or more generally by a sign).

On peut donc résumer que les étapes de la conceptualisation consistent à :

- identifier les groupes d'objets ou d'êtres qui partagent des caractéristiques permettant de les reconnaître et de les regrouper en classes. Les individus constituant les membres d'une classe sont appelés les instances ;
- identifier les termes qui expriment les liens ou les relations qui se trouvent entre ces classes et permettent de les classer en hiérarchie.

Une des caractéristiques de l'organisation hiérarchique est le fait qu'un objet situé plus bas dans le classement, et donc plus spécifique, hérite des caractéristiques de l'objet se situant au-dessus de lui. Au niveau conceptuel, la définition d'un terme subordonné inclut (ou implique) la définition du terme qui lui est directement supérieur hiérarchiquement. Les relations hiérarchiques peuvent être génériques ou partitives :¹²

- la relation générique (ou relation genre-espèce) est une relation d'inclusion entre deux concepts où l'ensemble des caractères,¹³ caractérisant l'un des concepts, inclut celui de l'autre et au moins un caractère distinctif supplémentaire. En linguistique, elle correspond à la relation hyponymie–hyperonymie qui lie deux termes dont le sens de l'un inclut celui de l'autre. Par exemple, la notion d'animal subsume celle de mammifère et tout mammifère est forcément un animal ;

10. [Guarino, 1998](#).

11. [Gandon, 2002](#), page 18.

12. [ISO 1087, 2000](#), p. 4.

13. l'ensemble des caractères est appelé «compréhension ». [ISO 1087, 2000](#), p. 3.

- la relation partitive (ou relation partie-tout) désigne la relation entre deux concepts dans laquelle l'un des deux constitue le tout et l'autre une partie de ce tout. Elle correspond à la relation méronymie-holonymie en linguistique.

D'autres relations, appelées associatives, peuvent lier les concepts entre eux par des liens thématiques non hiérarchiques mais fondés sur l'expérience. Elles peuvent être, par exemple, causale liant la « cause » et son « effet », ou temporelle impliquant la séquence des événements dans le temps.

La phase successive à la conceptualisation est la formalisation du modèle conceptuel obtenu afin que le système soit utilisable aussi bien par les humains que par les machines.

Basée sur un métalangage de représentation formelle, l'ontologie computationnelle est dépendante de la langue utilisée pour formaliser les concepts et les relations qui les relient. En effet, on peut avoir plusieurs ontologies même si elles partagent la même conceptualisation.¹⁴

En conclusion, dans les applications informatiques, une ontologie est un artefact d'ingénierie qui décrit une certaine réalité utilisant un vocabulaire spécifique et un ensemble d'axiomes explicites concernant le sens délibéré des mots qui décrivent cette réalité. Les axiomes ont la forme d'une théorie logique de premier ordre, où les mots peuvent représenter les concepts et les relations. Elle sert surtout pour capturer l'information sémantique existante, pour assurer leur gestion de manière automatique. Selon la nature des documents, on peut distinguer deux types d'ontologie. Les ontologies qui couvrent des domaines spécifiques et les ontologies dites supérieures qui identifient des concepts généraux.

2.3.4.2 WordNet

2.3.4.2.1 Organisation de la WordNet La WordNet de la langue anglaise, dont l'acronyme est PWN (Princeton WordNet), est une ressource lexico-sémantique, développé au laboratoire de sciences cognitives de l'Université de Princeton.¹⁵

La version actuelle 3.1 contient quatre bases de données associées aux catégories des noms, verbes, adjectifs et adverbes.

Dans l'introduction, on a déjà présenté, dans la figure 0.1, les résultats correspondants à la recherche de la parole « leaf » dans le portail de PWN. Par commodité, La figure 2.34 les illustre de nouveau. On note que la structure de WordNet ne se

14. Guarino, 1998 et Prévot et al., 2010.

15. PWN est accessible à l'adresse suivante : <http://wordnet.princeton.edu/>

base pas sur les entrées lexicales isolées accompagnées de leurs définitions. Plutôt elle est organisée sous forme de synsets (venant de synonym set) dont chacun représente l'ensemble de toutes les paroles synonymes qui identifient un même sens ou concept. Concernant la parole «leaf », on trouve six synsets, dont trois sont des noms et trois sont des verbes. Chaque synset est constitué des unités lexicales représentant le même concept, et est accompagné d'une glose qui fournit une petite définition.

The screenshot shows the WordNet Search interface. At the top, it says "WordNet Search - 3.1" with links to "WordNet home page", "Glossary", and "Help". Below this is a search bar with "leaf" entered and a "Search WordNet" button. There are also "Display Options" with a dropdown menu set to "(Select option to change)" and a "Change" button. A key explains that "S:" shows synset (semantic) relations and "W:" shows word (lexical) relations. The display options for the sense are set to "(gloss) 'an example sentence'".

Noun

- **S: (n) leaf, leafage, foliage** (the main organ of photosynthesis and transpiration in higher plants)
- **S: (n) leaf, folio** (a sheet of any written or printed material (especially in a manuscript or book))
- **S: (n) leaf** (hinged or detachable flat section (as of a table or door))

Verb

- **S: (v) flick, flip, thumb, riffle, leaf, riff** (look through a book or other written material) *"He thumbed through the report"; "She leafed through the volume"*
- **S: (v) leaf** (turn over pages) *"leaf through a book"; "leaf a manuscript"*
- **S: (v) leaf** (produce leaves, of plants)

Fig. 2.34 : Résultats de la recherche de de la parole «leaf » dans PWN

Concernant les noms, WordNet est organisée par des relations explicites et étiquetées, dont les principales sont illustrées dans l'exemple de la figure 2.35. On distingue deux types de relations :

1. des relations lexicales qui lient les unités lexicales entre elles, par exemple, la relation dérivationnelle (Derived form : leaf₃) ;
2. des relation conceptuelles, de nature sémantique-conceptuelle, qui lient les concepts entre eux :
 - Relation d'hyponymie-hyperonymie : exprimée, dans la PWN, comme «is a kind of », elle est connue, également, comme la relation «is-a » («est-un »).

Elle permet d'organiser les synsets sous forme hiérarchique, en liant entre un terme spécifique et un terme générique. Le nœud représentant le concept le plus général (par exemple plante) est considéré l'ancêtre du nœud représentant le concept plus spécifique (par exemple arbre). On dit alors que le concept plus général « subsume » celui plus spécifique : leaf is a plant ;

- Relation de méronymie-holonomie : c'est la relation qui lie le tout, appelée holonyme, avec une partie, appelée méronyme. Dans WordNet, la méronymie est exprimée par la relation « meronym » et peut être interprétée de trois manières différentes :

X is a part of Y lie le composant et l'objet complet : leaf has parts venation

X is a substance of Y lie la matière et l'objet : leaf has substance parenchyma

X is a member of Y lie le membre et la collection.

Noun leaf has 3 senses

1. leaf, leafage, foliage - the main organ of photosynthesis and transpiration in higher plants
 - ₁ is a kind of plant organ
 - ₁ has parts: venation; lobe; leaf shape, leaf form
 - ₁ has substances: parenchyma
 - ₁ has particulars:
 - cataphyll; floral leaf; dandelion green; sporophyll, sporophyl; blade, leaf blade; amplexicaul leaf; greenery, verdure; leaflet; frond; pad; scale, scale leaf; fig leaf; simple leaf; compound leaf; entire leaf; crenate leaf; serrate leaf; dentate leaf; emarginate leaf; erose leaf; runcinate leaf; lobed leaf; parted leaf; rosette
 - Derived form: verb leaf₃
2. leaf, folio - a sheet of any written or printed material (especially in a manuscript or book)
 - ₂ is a kind of sheet, piece of paper, sheet of paper
 - ₂ is a part of written communication, written language
 - ₂ has particulars: flyleaf; interleaf; page
 - Derived forms: verb leaf₂, verb leaf₁
3. leaf - hinged or detachable flat section (as of a table or door)
 - ₃ is a kind of section, segment
 - ₃ is a part of dinner table
 - ₃ has particulars: drop-leaf

Fig. 2.35 : Organisations de la base des noms dans PWN

Concernant les verbes, dont la figure 2.36 présente l'exemple « kill », on note que la PWN les organise en taxonomie grâce à :

1. la relation conceptuelle, « Troponymy » équivalente à « is one way to », lie les synsets entre eux :
le synset {kill, shoot down, defeat, vote down, vote out} is one way to {veto, blackball, negative};
2. des relations lexicales liant les unités entre elles :
 - relation de causalité : par exemple, {kill} cause to {die};
 - relation de dérivation morphologique.

Verb kill has 15 senses

1. kill - cause to die; put to death, usually intentionally or knowingly; "This man killed several people when he tried to rob a bank"; "The farmer killed a pig for the holidays"
Derived forms: noun kill₁, noun killer₂, noun killer₁, noun killing₁, noun killing₂
Sample sentence:
They want to kill the prisoners
2. kill, shoot down, defeat, vote down, vote out - thwart the passage of; "kill a motion"; "he shot down the student's proposal"
--2 is one way to veto, blackball, negative
Sample sentence:
Somebody ----s something
3. kill - cause the death of, without intention; "She was killed in the collision of three cars"
Derived forms: noun kill₁, noun killer₂, noun killing₁
Sample sentence:
Something ----s somebody

Fig. 2.36 : Organisations de la base des verbes dans PWN

Concernant, les adjectifs et les adverbes, la figure 2.37, par contre, illustre l'exemple de «bad » et «badly ». On note que les adjectifs sont organisés par l'intermédiaire de relations lexicales telles que l'antonymie. Par contre, les adverbes sont affichés et classés en fonctions des adjectifs desquels ils dérivent.

Toutefois, on ne s'empêche pas de noter que les relations dérivationnelles et sémantiques, très importantes dans le cas de la langue arabe, comme on verra plus loin, ne sont pas très développées dans la PWN. Dans le cas où elles existent, elles sont juste mentionnées sans spécifier leur nature, mais, dans plusieurs cas, elles ne sont pas mentionnées.

Adjective bad has 14 senses

1. **bad** - having undesirable or negative qualities; "a bad report card"; "his sloppy appearance made a bad impression"; "a bad little boy"; "clothes in bad shape"; "a bad cut"; "bad luck"; "the news was very bad"; "the reviews were bad"; "the pay is bad"; "it was a bad light for reading"; "the movie was a bad choice"
 Antonyms: **good**, **bang-up**, **bully**, **corking**, **cracking**, **dandy**, **great**, **groovy**, **keen**, **neat**, **nifty**, **not bad**, **peachy**, **slap-up**, **swell**, **smashing**, **good enough**, **goodish**, **redeeming**, **satisfactory**, **acceptable**, **solid**, **suitable**, **superb**, **well-behaved**, **well behaved**, **hot**
2. **bad, big** - very intense; "a bad headache"; "in a big rage"; "had a big (or bad) shock"; "a bad earthquake"; "a bad storm"
 Antonym: **mild** (indirect, via **intense**)
3. **bad, tough** - feeling physical discomfort or pain ('tough' is occasionally used colloquially for 'bad'); "my throat feels bad"; "she felt bad all over"; "he was feeling tough after a restless night"
 Antonyms: **comfortable**, **comfy** (indirect, via **uncomfortable**)

Adverbial bad has 2 senses

1. **badly**, **bad** - with great intensity ('bad' is a nonstandard variant for 'badly'); "the injury hurt badly"; "the buildings were badly shaken"; "it hurts bad"; "we need water bad"
2. **badly**, **bad** - very much; strongly; "I wanted it badly enough to work hard for it"; "the cables had sagged badly"; "they were badly in need of help"; "he wants a bicycle so bad he can taste it"

Fig. 2.37 : Organisation des adjectifs et des adverbes dans PWN

Par exemple, la parole *break* fait partie, entre autres, de deux synsets. Le synset 00334996 dont le sens est causatif, et le synset 00336260 dont le sens est résultatif mais sans aucune relation entre eux :

- le synset, Id=00334996, contenant la parole *break*; ayant la définition «*destroy the integrity of; usually by force; cause to separate into pieces or fragments; "He broke the glass plate"; "She broke the match"*»;
- le synset, Id=00336260, contenant les synonymes *break*, *check* et *crack*, et ayant la définition «*become fractured; break or crack on the surface only; "The glass cracked when it was heated"*»

Une autre perplexité surgit lorsqu'on se rend compte de la granularité très fine avec laquelle est faite la PWN. Par exemple, le nom, *water*, fait partie de huit synsets, entre autres, les deux suivants qu'on pourrait regrouper :

- le synset, Id=14845743, contenant les paroles *H2O* et *water*, ayant la définition «*binary compound that occurs at room temperature as a clear colorless odorless tasteless liquid ; freezes into ice below 0 degrees centigrade and boils above 100 degrees centigrade ; widely used as a solvent*»;
- le synset, Id= 07935504, contenant la parole *water*, ayant la définition «*a liquid necessary for the life of most animals and plants ; "he asked for a drink of water"*».

2.3.4.2.2 WordNet de la langue arabe (Arabic WordNet) La WordNet de la langue arabe "Arabic WordNet" (AWN) est une version de taille moyenne, entièrement liée à SUMO par l'intermédiaire de PWN.¹⁶ Mais, AWN a une couverture beaucoup plus limitée que la couverture de PWN (~ 11K vs ~ 118k synsets).

Elle a été enregistrée manuellement (après une détection préliminaire automatique des traductions candidats) et se limite à des ressources contemporaines.

Le projet a été interrompu pendant des années et actuellement plusieurs recherches mirent à développer des méthodes appropriées et rapides pour l'enrichir.

2.3.4.3 SUMO (The Suggested Upper Merged Ontology)

Dans le but de favoriser l'interopérabilité des données, de la recherche, la récupération de l'information et le traitement du langage naturel, l'IEEE,¹⁷ a proposé SUMO (The Suggested Upper Merge Ontology)¹⁸ comme point de départ pour la « standardisation d'ontologie supérieur ». ¹⁹ En outre, elle a l'avantage d'être distribuée sous licence libre GNU (General Public License) et on peut la télécharger gratuitement.²⁰

SUMO fournit des définitions d'usage universel et contient les concepts les plus larges et les plus abstraits. La figure 2.38 montre les concepts du plus haut niveau.

Le nœud racine de l'ontologie est «entity » 'entité' qui englobe deux concepts :

- « physical » 'physique' correspondant à tout objet ayant une position dans le temps ou dans l'espace et ;
- « abstract » 'abstrait' qui englobe les entités qui n'existent ni dans le temps ni dans l'espace.

16. Black et al., 2006.

17. The Institute of Electrical and Electronics Engineers, Inc. (<http://grouper.ieee.org/groups/>)

18. www.adampease.org/OP/

19. Niles & Pease-2001

20. Téléchargement à : <http://sigmakee.sourceforge.net/>

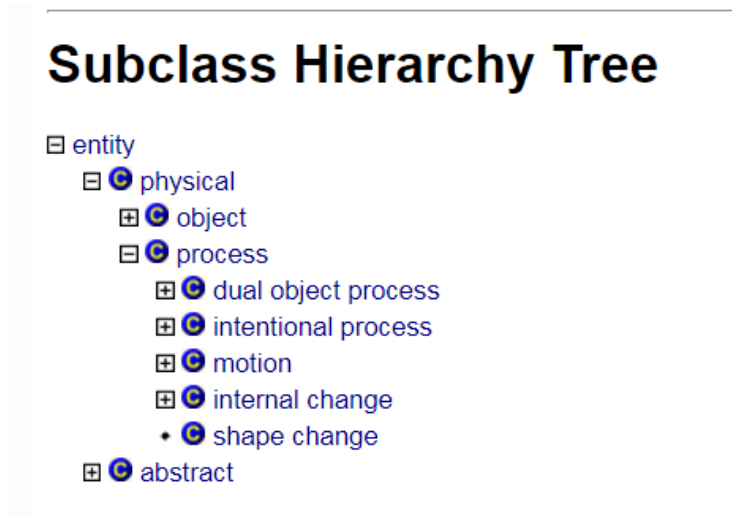


Fig. 2.38 : Partie de la hiérarchie des concepts et des relations dans SUMO

Les concepts physiques sont divisées en objet et en processus. Les objets à leurs fois, sont divisés en :

- «self connected object » concerne les objets constitués d'une seule et unique partie ;
- «region » englobe les lieux imaginaires et les zones géographiques ;
- «collections » constituées de membres qui peuvent être ajoutés ou soustraits sans modifier l'identité de la Collection. Quelques exemples sont les équipes de football et les troupeaux de moutons. La relation qui lie les parties avec la *Collection* qui leur correspond est la relation *member*;
- «agent » : quelque chose ou quelqu'un qui peut agir seul et produire des changements dans le monde.

Les processus, par contre, peuvent être :

- «dual object process » englobant tout processus nécessitant deux patients non identiques ;
- «intentional process » délibérément fait par un CognitiveAgent ;
- «motion » englobant les processus de mouvement ;
- «internal change », regroupe les processus qui impliquent la modification d'une propriété interne d'un objet, par exemple sa coloration, sa structure, etc. ;

- «shape change » englobe les processus impliquant la modification de la forme d'un objet.

Les concepts de la superclasse « abstract » sont divisibles, entre autres, en :

- «quantity » qui englobe les spécifications de quantité. En conséquence, il existe deux sous-classes de quantité : Number et PhysicalQuantity
- «attribute » qui englobe les qualités et les propriétés qui ne peuvent être classées dans les sous-classes de Objet .
- «set or class » est constitué de toutes les classes ayant des instances ou des membres.
- «proposition » sont des entités abstraites qui expriment une pensée complète ou un ensemble de pensées. Notez que les propositions ne se limitent pas au contenu exprimé par les phrases individuelles d'une langue. Ils peuvent englober le contenu exprimé par des théories, des livres et même des bibliothèques entières.

2.3.4.4 Alignement de la PWN avec SUMO

La figure 2.39 montre le résultat de la recherche du verbe «swerve » dans le portail Sigma de SUMO. ²¹ On peut identifier :

- l'identificateur (Id) du synset auquel appartient le verbe «swerve » : Id=**202033295**;
- le concept SUMO auquel correspond le synset : **DirectionChange**;
- le synset constitué de : **swerve** - sheer - curve - trend - veer - slue - slew - cut ;
- la glose correspondant à une définition du concept représenté par le synset ;
- la relation d'hyponymie : le verbe «turn » dont l'identificateur est 201907258 est l'hyperonyme du synset 202033295 auquel appartient le verbe «swerve »;
- les relations dérivationnelles qui lient le verbe aux noms pertinents ;
- la relation d'hyponymie : le synset «peel_off », dont l'identificateur est 202033703, et «yaw », dont l'identificateur est 202034004, sont les hyponymes du synset 201907258.

SUMO Search Tool

This tool relates English terms to concepts from the [SUMO](#) ontology by means of mappings to [WordNet](#) synsets.

English Word: Any

Verb Synset: 202033295

- SUMO Mappings: [DirectionChange](#) (subsuming mapping)

Words: curve, cut, sheer, slew, slue, swerve, trend, veer

Gloss: turn sharply; change direction abruptly; "The car cut to the left at the intersection"; "The motorbike veered to the right"

hypernym [201907258](#) - turn

derivationally related [101263257](#) - swerve, swerving, veering

derivationally related [108681222](#) - course, trend

derivationally related [113867641](#) - curve, curved_shape

derivationally related [102829696](#) - bend, curve

derivationally related [107411350](#) - swerve, yaw

derivationally related [101263257](#) - swerve, swerving, veering

derivationally related [101263257](#) - swerve, swerving, veering

hyponym [202033703](#) - peel_off

hyponym [202034004](#) - yaw

[Show OWL translation](#)

[Sigma web home](#) [Suggested Upper Merged Ontology \(SUMO\) web home](#)
Sigma version 2.99b (>= 2014/01/25) is [open source software](#) produced by [Articulate Software](#) and its partners

Fig. 2.39 : Résultats de la recherche du verbe «swerve » dans Sigma.

La figure 2.40, par contre, montre comment sont codifiées les informations dans la base de données de SUMO (dorénavant DB "data base"). On peut distinguer :

- l'identificateur du synset auquel appartient le verbe «swerve » : **202033295**;
- la catégorie grammaticale (POS) : *v* indique verbe ;
- le synset : **swerve** - sheer - curve - trend - veer - slue - slew - cut ;
- les relations d'hyperonymie introduites par le symbole @ ;
- les relations dérivationnelles introduites par le symbole + suivi de l'id du synset correspondant et de sa POS ; et de la lettre n qui indique que c'est un nom ;
- les relations d'hyponymie introduite par le symbole ~;
- la glose introduite par le symbole | ;

21. <http://sigma.ontologyportal.org:8080/sigma/Browse.jsp?kb=SUMO>

- le concept SUMO introduit par le symbole &% ;
- le concept est suivi d'un symbole qui peut être '=', '+' ou '@' et indique que le concept représenté par le synset de PWN est, respectivement, équivalent, subsumé ou une instance du concept général de SUMO.

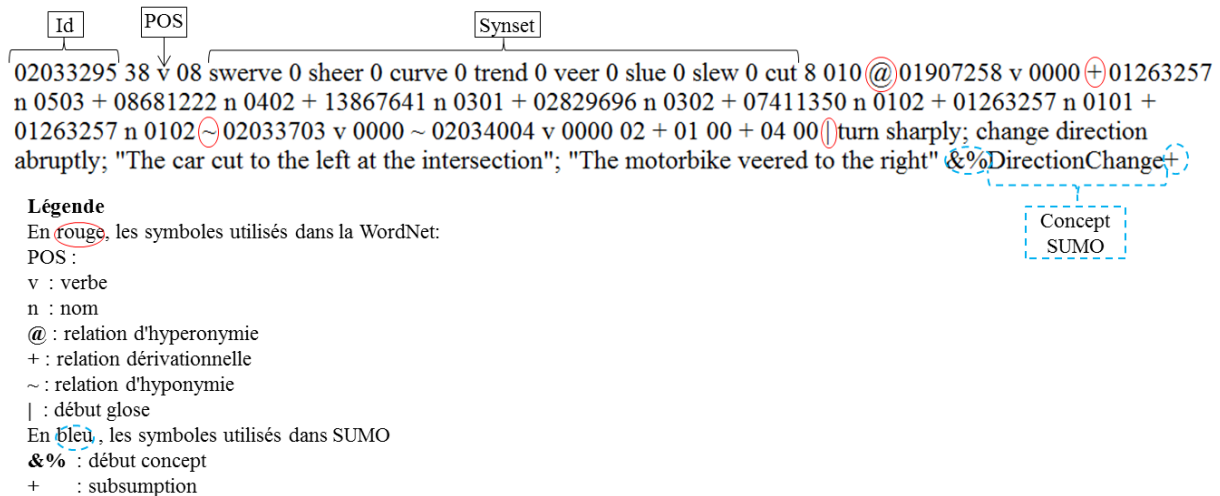


Fig. 2.40 : Exemple d'organisation de la base de donnée (DB) de SUMO

Le but de la recherche étant l'alignement entre les données arabes de *qāmūs* avec PWN et SUMO, l'objectif est chercher le correspondant anglais, d'un lemme arabe donné, dans la data-base de SUMO et, récupérer l'identificateur du synset et le concept SUMO auquel il est lié. Les champs pertinents, dans la DB, sont l'identificateur du synset et le concept.

Par souci de simplicité, on présentera, dans les exemples qui suivent, la parole anglaise accompagnée de l'identificateur, de la glose qui l'explique, introduite par le symbole |, et du concept SUMO, introduit par &%. Par exemple :

swerve 02033295 | turn sharply ; change direction abruptly ; "The car cut to the left at the intersection"; "The motorbike veered to the right" &%DirectionChange+

2.3.5 Alignement de *qāmūs* avec les ressources ontologiques

2.3.5.1 Cas de correspondance entre de *qāmūs*, PWN et SUMO

2.3.5.1.1 Levée d'ambigüité Lorsqu'on utilise des dictionnaires bilingues, un des problèmes majeurs de la traduction automatique est l'effet de la polysémie. Cependant une parole polysémique a un seul sens dans un contexte précis, d'autant plus qu'elle

constitue la lexie d'une entrée lexicale. Dans ce cas, les sens qui l'accompagne délimitent sa signification. Pour faire face au problème de la polysémie, en plus de la traduction du lemme, on a fait recours, également, à la traduction des sens qui l'accompagnent. Les traductions acceptées par le système résultent de l'intersection entre les traductions du lemme et les traductions des sens.

2.3.5.1.1.1 Cas où lemme et sens ont des traductions communes Le verbe نَكَب *nakaba* est polysémique et il a été listé six fois, dans *qāmūs*, avec des sens différents. L'exemple 41 illustre l'entrée lexicale du verbe نَكَب *nakaba* avec le sens de عدل *'adala*.

Exemple 41

نَكَبَ عَنْهُ، كَنَصَرَ وَفَرَحَ، نَكَبًا، وَنَكَبًا، وَنُكُوبًا : عَدَلَ، كَنَكَبَ، وَتَنَكَّبَ

Selon le dictionnaire DMWA, le verbe نَكَب *nakaba* a différentes traductions :²²

1. نَكَب *nakaba*: to make unhappy, miserable, afflict, distress ; to drop ; to put out of favor ;
2. نَكَب *nakaba*: to blow sideways ; to veer, shift ;
3. نَكَب *nakaba*: **to deviate, swerve.**

le verbe عدل *'adala*, qui représente le sens de نَكَب *nakaba*, dans *qāmūs*, présente aussi diverses traductions :²³

1. عدل *'adala*: to act justly, equitably, with fairness
2. عدل *'adala* : to be equal , be equivalent , be on par , be the equal , equal, match , counterbalance, outweigh ; to make equal, equalize, level , place on the same level or footing :
3. عدل *'adala*: **to deviate, swerve**, deflect, turn away ; to digress, depart, refrain,

22. Wehr, 1971, p. 997.

23. Wehr, 1971, p. 596.

desist, abstain, avert o.s., turn away, leave off, relinquish. abandon, renounce, disclaim, give up, forego, waive, drop.

En confrontant, on note que deux traductions sont communes au lemme *nakaba* et au sens 'adala : **to deviate, swerve**. Le système prendra compte, uniquement, de ces deux traductions ; et affiche leurs correspondants dans la DB de SUMO :

swerve 02033295 | turn sharply ; change direction abruptly ; "The car cut to the left at the intersection"; "The motorbike veered to the right" &%DirectionChange+

deviate 02064887 | turn aside ; turn away from &%Transfer+

deviate 02065085 | cause to turn away from a previous or expected course ; "The river was deviated to prevent flooding" &%Transfer+

On note que le verbe «deviate » appartient à deux synsets. Le verbe «deviate », du synset Id=02064885, est le causatif du verbe «deviate », du synset Id=02065087. Or, les verbes arabes trilitères de l'exemple 41, نكب *nakaba* et عدل 'adala, ne sont pas causatifs. Par conséquent, une phase de désambiguïsation est nécessaire pour sélectionner le sens 'deviate' non causatif et donc appartenant au synset Id=02064887. Les résultats obtenus sont illustrés dans la figure 2.41.

Dans le fichier XML, le lemme est accompagné d'autres informations délimitées par une balise :

- *< palain – text >* illustre la contextualisation du lemme, dans *qāmūs* ;
- *< pos >*: identifie la catégorie grammaticale comme verbe ;
- *< morphological_information >* : définit que le verbe est de base (forme I) ou dérivé (forme II, III ; IV...);
- *< root – information >* : détermine la racine de laquelle il dérive est trilitère (T) ou quadrilitère (Q);
- *< morphosyntactic – information >* : Transitif ou Intransitif ;
- *< sens >* : le/les sens qui l'accompagnent ;
- *< lemma – translate >* : la/les traductions du lemme selon le contexte et les sens ;
- *< external – link >* : détermine les liens externes à la ressource *qāmūs*, qui le mettent en correspondance avec le/les concept SUMO et le/les synsets PWN.

```

<lexical_entry>
  <plain_text>عَدَلَ</plain_text>
  <lemma>تَكَب</lemma>
  <pos>Verb</pos>
  <morphological_information>I</morphological_information>
  <masdar>تَكَب</masdar>
  <masdar>تَكَب</masdar>
  <masdar>تَكُوب</masdar>
  <sense att="عَدَلَ">
  <synonym att="عَدَلَ">
  <translate lang="eng" value="to deviate, swerve" />
  <external_links att="mapping" value="null">
    <concept ref="SUMO" value="DirectionChange+">
      <synset ref="PWN" ver="3.0" id="02033295" />
    <concept ref="SUMO" value="Transfer+">
      <synset ref="PWN" ver="3.0" id="02064887" />
    </concept>
  </external_links>
</synonym>
</sense>
  <root>ن ك ب</root>
</lexical_entry>

```

Fig. 2.41 : Traductions du lemme *nakaba* avec le sens de ‘*adala*

2.3.5.1.1.2 Cas où lemme et sens n’ont pas des traductions communes

Un exemple intéressant, qui illustre ce cas, est le verbe أَبَّ *’abba* qui, dans *qāmūs*, est listé dans six entrées lexicales avec autant de sens différents. Selon le dictionnaire DMWA, la traduction du lemme coïncide seulement avec celle du sens 2 :

Lemme أَبَّ *’abba*: to long, yearn;

sens-1 تَهَيَّأ *tahayya’a* : to be prepared, be in readiness, be ready, stand ready ; to be armed, be prepared for war, be in fighting condition (mil.); to prepare o.s., get ready ; to be possible ; to be well-dressed ;

sens-2 اِشْتَاق *ištāqa* : to long, yearn, crave, covet, desire ardently

sens-3 رَدَّ *radda* : to send back ; to bring back, take back [...];

sens-4 قصد *qaṣada* : to go or proceed straight away ; make a beeline, walk up to ; to go to see [...],

sens-5 هَزَمَ *hazama* : to put to flight, rout, vanquish, defeat [...];

sens-6 حَرَكَ *ḥarraka* : to move, set in motion, drive, propel, operate [...].

Dans ce cas, ce sont les traductions des sens qui permettent la désambiguïisation du lemme. Ce dernier assume, selon le contexte, la/les traductions du sens. Par exemple, dans la figure 2.42, selon le contexte du « plain texte », le lemme 'abba hérite, dans le premier cas, la traduction du verbe *hazama* 'vaincre' et, dans le deuxième cas, celle de *ḥarraka* 'bouger'.

```
<plain_text>أَبٌ: هَزَمَ بِحَمْلَةٍ لَا مَكْتُوبَةٍ فِيهَا</plain_text>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>T</root_information>
<morphosyntactic_information>Intransitive</morphosyntactic_information>
<sense att="هَزَمَ بِحَمْلَةٍ لَا مَكْتُوبَةٍ فِيهَا">
<troponym_of att="هَزَمَ">
  <translate lang="eng" value="to put to flight, rout, vanquish, defeat;to put out of action, neutralize" />
  <external_links att="mapping" value="null">
    <concept ref="SUMO" value="Won+">
      <synset ref="PWN" ver="3.0" id="01108148" />
    </concept>
  </external_links>
</troponym_of>
</sense>
<plain_text>أَبٌ الشَّيْءِ: حَرَكَةً</plain_text>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>T</root_information>
<morphosyntactic_information>Transitive</morphosyntactic_information>
<external_links>
</external_links>
<sense att="حَرَكَةً">
<synonym att="حَرَكَ">
  <translate lang="eng" value="to move, set in motion, drive" />
  <external_links att="mapping" value="null">
    <concept ref="SUMO" value="Transfer+">
      <synset ref="PWN" ver="3.0" id="01850315" />
    </concept>
  </external_links>
</synonym>
</sense>
```

Fig. 2.42 : Contextualisations du verbe 'abba et correspondance dans SUMO et PWN.

2.3.5.1.1.3 Cas où le lemme n'est pas traduit L'exemple 42 illustre l'entrée lexicale du lemme *ğa'nab*.

Exemple 42

الْجَانِبُ كَجَعْفَرٍ: الْقَصِيرُ، الْقَمِيءُ، [مِنَّا وَمِنَ الْخَيْلِ]، وَهِيَ بِهَاءٍ وَبِغَيْرِ هَاءٍ

le=*ğa'nab* comme *ğa'far* : le petit, le servile, [de nous et des chevaux], et elle [est] avec ou sans *hā'*

Le lemme n'est traduit dans aucun dictionnaire, par contre les sens qui l'accompagnent sont traduits. Les traductions du premier sens :

sens-1 الْقَصِيرُ *qaṣīr*: short ; small, short (of stature), low.

Selon le sens du lemme dans *qāmūs*, la traduction «short (of stature) » est, probablement, la plus adéquate. Cependant, la parole «short » fait partie de 11 synsets différents, dont les plus adéquats peuvent être :

short 02386612 | low in stature ; not tall ; "he was short and stocky"; "short in stature"; "a short smokestack"; "a little man" &%SubjectiveStrongNegativeAttribute+

short 00628687 | having a relatively short body.

Par contre, le deuxième sens, *qamī'*, a deux traductions :

sens-2 الْقَمِيءُ *qamī'*: insignificant ; despicable

Les deux traductions posent également des problèmes de polysémie :

insignificant 01288298 | not worthy of notice &%SubjectiveWeakNegativeAttribute+

insignificant 01473353 | of little importance or influence or power ; of minor status ; "a minor, insignificant bureaucrat"; "peanut politicians" &%SubjectiveWeakNegativeAttribute+

insignificant 02164402 | devoid of importance, meaning, or force &%SubjectiveWeak-NegativeAttribute+

despicable 01133017 | morally reprehensible ; "would do something as despicable as murder"; "ugly crimes"; "the vile development of slavery appalled them"; "a slimy little liar" &%SubjectiveStrongNegativeAttribute+

À cause de la granularité trop fine de la PWN, on trouve des difficultés pour distinguer le synset de la PWN et le concept SUMO adéquats. Il faut noter, aussi, que le deuxième sens de short (synset 00628687) n'est lié à aucun concept SUMO. La figure 2.43 illustre la représentation en XML de l'exemple 42, après la révision des résultats.

```

<lexical_entry>
  <plain_text>الْجَائِبُ، كَجَعْفَرٍ: الْقَصِيرُ، الْقَمِيءُ، [منا ومن الخيل.] وهي بهاءٌ وغير هاءٍ.</plain_text>
  <lemma>جَائِبٌ</lemma>
  <pos>ADJ</pos>
  <sense att="القصير">
    <synonym att="قصير">
      <translate lang="eng" value="small, short, low" />
    <external_links att="mapping" value="null">
      <concept ref="SUMO" value="SubjectiveStrongNegativeAttribute+">
        <synset ref="PWN" ver="3.0" id="02386612" />
      </concept>
      <concept ref="SUMO" value="null">
        <synset ref="PWN" ver="3.0" id="00628687" />
      </concept>
    </external_links>
  </synonym>
</sense>
<sense att="القميء">
  <synonym att="قميء">
    <translate lang="eng" value="lowly, little, small; insignificant, of little value" />
  <external_links att="mapping" value="null">
    <concept ref="SUMO" value="SubjectiveStrongNegativeAttribute+">
      <synset ref="PWN" ver="3.0" id="01133017" />
    </concept>
    <concept ref="SUMO" value="SubjectiveWeakNegativeAttribute+">
      <synset ref="PWN" ver="3.0" id="01473353" />
      <synset ref="PWN" ver="3.0" id="02164402" />
      <synset ref="PWN" ver="3.0" id="01288298" />
    </concept>
  </external_links>
</synonym>
</sense>
</root>

```

Fig. 2.43 : Correspondance synsets et concept du lemme *ǧa`nab*

2.3.5.1.2 Problèmes de désambigüisation L'exemple 43 illustre le verbe نَكَبَ *nakaba* accompagné des sens لَثَمَ *laṭama* et أَصَابَ *aṣāba*.

Exemple 43

نَكَبَتِ الْحِجَارَةُ رِجْلَهُ: لَثَمَتْهَا، أَوْ أَصَابَتْهَا، فَهُوَ مَنكُوبٌ وَنَكِيبٌ

la pierre ***nakab-at*** son pied : elle l'a touché, ou elle l'a atteint, et lui [est] *mankūb* et *nakib*

On note que le lemme n'a pas de traduction commune avec les sens. Par contre, les deux sens ont une traduction commune 'hit' :

Lemme نَكَبَ *nakaba*: to make unhappy, miserable, afflict, distress ; to drop ; to put out of favor ; to blow sideways ; to veer, shift ; to deviate, swerve ;

sens-1 لَثَمَ *laṭama*: to kiss ; to strike, **hit**, wound, injure ;

sens-2 أَصَابَ *aṣāba* : to **hit**; to attain, reach, achieve ; to reach out ; to get, obtain, win, gain, acquire ; to have, eat ; to take ; to befall, fall, happen ; to allot, bestow, make fall to s.o.'s lot ; to cause losses ; to be right, be in the right ; to do the right thing, hit the mark ; to do right, properly ; to say the right word ;

Cependant le verbe *hit* fait partie de 15 synsets correspondant chacun à un concept différent !

hit 00781903 | pay unsolicited and usually unwanted sexual attention to ; "He tries to hit on women in bars" &%Requesting+

hit 01111816 | gain points in a game ; "The home team scored many times"; "He hit a home run"; "He hit .300 in the past season" &%Game+

hit 01123887 | make a strategic, offensive, assault against an enemy, opponent, or a target ; "The Germans struck Poland on Sept. 1, 1939"; "We must strike the enemy's oil fields"; "in the fifth inning, the Giants struck, sending three runners home to win the game 5 to 2" &%ViolentContest+

- hit** 01137138 | hit with a missile from a weapon &%Shooting+
- hit** 01152396 | consume to excess ; "hit the bottle" &%Ingesting+
- hit** 01209135 | affect or afflict suddenly, usually adversely ; "We were hit by really bad weather"; "He was stricken with cancer when he was still a teenager"; "The earthquake struck at midnight" &%Process+
- hit** 01209426 | produce by manipulating keys or strings of musical instruments, also metaphorically ; "The pianist strikes a middle C"; "strike 'z' on the keyboard"; "her comments struck a sour note" &%Impacting+
- hit** **01236164** | **hit against ; come into sudden contact with ; "The car hit a tree"; "He struck the table with his elbow" &%Impacting+**
- hit** 01400044 | deal a blow to, either with the hand or with an instrument ; "He hit her hard in the face" &%Impacting+
- hit** 01405044 | cause to move by striking ; "hit a ball" &%Impelling+
- hit** 02006834 | reach a point in time, or a certain state or level ; "The thermometer hit 100 degrees"; "This car can reach a speed of 140 miles per hour" &%Attribute+
- hit** 02020590 | reach a destination, either real or abstract ; "We hit Detroit by noon"; "The water reached the doorstep"; "We barely made it to the finish line"; "I have to hit the MAC machine before the weekend starts" &%Arriving=
- hit** 02104161 | drive something violently into a location ; "he hit his fist on the table"; "she struck her head on the low ceiling" &%Transfer+
- hit** 02108791 | cause to experience suddenly ; "Panic struck me"; "An interesting idea hit her"; "A thought came to me"; "The thought struck terror in our minds"; "They were struck with fear" &%PsychologicalProcess+
- hit** 02482425 | kill intentionally and with premeditation ; "The mafia boss ordered his enemies murdered" &%Murder+

La désambiguïisation lexicale est déjà une tâche complexe, mais l'utilisation de PWN comme ressource pivot, pour faire correspondre une ressource d'une autre langue, rend le travail encore plus difficile, vu la polysémie de la langue anglaise et la granularité trop fine avec laquelle a été construite la PWN.

Dans le contexte, نَكَبَتِ الْحِجَارَةُ رِجْلَهُ "la pierre a touché son pied", le synset et le concept correspondants sont, plus probablement :

hit 01236164 | hit against; come into sudden contact with; "The car hit a tree"; "He struck the table with his elbow" &%Impacting+

La figure 2.44 affiche les résultats, concernant l'exemple 43, après la désambiguïsation.

```

<lexical_entry>
<plain_text>نَكَبَتِ الْحِجَارَةُ رِجْلَهُ: لَثَمَتْهَا، أَوْ أَصَابَتْهَا، فَهُوَ مَنَكُوبٌ وَنَكَبٌ</plain_text>
<lemma>نَكَبٌ</lemma>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<adjective>مَنَكُوبٌ</adjective>
<adjective>نَكَبٌ</adjective>
<sense att="لَثَمَتْهَا">
<synonym att="لَثَمْتُ">
<translate lang="eng" value="to hit" />
<external_links att="mapping" value="null">
<concept ref="SUMO" value="Impacting+">
<synset ref="PWN" ver="3.0" id="01236164" />
</concept>
</external_links>
</synonym>
</sense>
<sense att="أَوْ أَصَابَتْهَا">
<synonym att="أَصَابٌ">
<translate lang="eng" value="to hit" />
<external_links att="mapping" value="null">
<concept ref="SUMO" value="Impacting+">
<synset ref="PWN" ver="3.0" id="01236164" />
</concept>
</external_links>
</synonym>
</sense>
<root>ن ك ب</root>
</lexical_entry>

```

Fig. 2.44 : Traduction du lemme avec le sens de *laṭama* et *'aṣāba*

Dans d'autres cas, les définitions dans *qāmūs* sont obscures et il faudrait une étude plus approfondie. Le lemme أَتَعَبٌ *'at'aba*, dans l'exemple 44, est défini par un autre

verbe أَغْتَبَ 'a'taba qui n'est pas défini dans qāmūs.

Exemple 44

أُغْتَبَ الْعَظْمُ : أَغْتَبَهُ بَعْدَ الْخَيْرِ

'at'aba l'os : 'a'taba=hu après l'avoir éclissé.

En cherchant le verbe أَغْتَبَ 'a'taba dans tāġ,²⁴ on le trouve comme extension, avec le sens de "provoquer ou causer le 'atab". La définition de la parole 'atab est, également, inexistante dans qāmūs et se trouve comme extension dans tāġ.²⁵

Sens de 'atab dans tāġ

وَالْعَتَبُ فِي الْعَظْمِ: النَّقْصُ وَهُوَ إِذَا لَمْ يُحْسَنْ جِرَّهُ وَبَقِيَ فِيهِ وَرَمٌ أَوْ عَرَجٌ

'et=al='atab dans [le cas de] l'os : la déféctuosité, s'il n'a pas été bien éclissé et subsiste une tuméfaction ou un boitement'.

Sens de 'a'taba dans tāġ

وَيُقَالُ فِي الْعَظْمِ الْمَجْبُورِ أُغْتَبَ فَهُوَ مُعْتَبٌ، كَأَعْنَتٍ، كَالْتَعْتَابِ، وَأَصْلُ الْعَتَبِ الشَّدَّةُ كَمَا تَقَدَّمَ

'et on dit dans [le cas de] l'os éclissé 'u'tiba et donc il [est] ma'tūb comme 'u'nita, comme le it=ti'tāb, et la cause du 'atab est le resserrement, comme il a été déjà expliqué'.

N'ayant pas de traduction dans les dictionnaires bilingues, ni le lemme ni le sens, de l'exemple 44, ont un correspondant dans la PWN et dans SUMO.

En conclusion, la tâche de la désambiguïsation est complexe et longue à cause de plusieurs faits, comme, la polysémie qui caractérise aussi bien la langue arabe que la langue anglaise ; la finesse de l'organisation de PWN. En outre, dans certaines cas,

24. tāġ est considéré l'œuvre la plus complète, dans laquelle l'auteur, az=zabīdī, a surtout expliqué qāmūs, en ajoutant, comme extension, ce qu'il a considéré manquant. Voir 2.3.3 et l'exemple 1.2

25. Zabīdī, Tāġ al-'arūs, V. III, p. 307 et 312

les définitions très succinctes, dans *qāmūs*, nécessitent le recours à d'autres lexiques, comme *tāǧ*, pour les «comprendre».

2.3.5.2 Cas de non correspondance entre *qāmūs*, PWN et SUMO

Le tableau 2.21 donne le pourcentage des lemmes qui ont été révisés manuellement et n'ont pas une correspondance dans PWN et dans SUMO. On peut identifier différentes typologies de problèmes de non correspondance.

POS	Total	Sans correspondance	
		Total	%
Verbes	511	204	40
Noms	479	41	8,50
Adjectifs	200	56	29,50

Tab. 2.21 : Total et pourcentage des lemmes sans correspondance

2.3.5.2.1 Problèmes stylistiques Dans certains cas, comme l'exemple 45 et 46, le lemme n'est pas traduit dans les dictionnaires bilingues. En outre, la définition est phrastique, qu'il faudrait modifier ou interpréter pour avoir une correspondance.

Dans l'exemple 45, la définition *زَادَ شَحْمُهُ* "sa graisse a augmenté" peut bien correspondre à "s'engraisser" et en anglais à "fatten".

Exemple 45

أَثْرَبَ الْكَبْشُ: زَادَ شَحْمُهُ

"le mouton *'atraba* : sa graisse a augmenté"

Dans l'exemple 46, la définition, "ce qui affecte la terre [de la part] du missionnaire [envoyé] à pied", peut correspondre à "empreinte".

Exemple 46

الذُّوَابَةُ مِنَ النَّعْلِ: مَا أَصَابَ الْأَرْضَ مِنَ الْمُرْسَلِ عَلَى الْقَدَمِ

"la=*du'ābah* des sandales : ce qui affecte la terre [de la part] du missionnaire [envoyé] à pied".

2.3.5.2.2 Problèmes de culture et de religion Beaucoup de concepts, exprimés en langue arabe, reflètent des réalités qui n'existent pas dans le monde européen. La vie nomade dans le désert donne une importance particulière aux chameaux et au troupeau. Dans l'exemple 47, le sujet du verbe *'ahlaba*, déjà expliqué dans l'exemple 38 du paragraphe 2.2.5.2.2, est le propriétaire de chameaux qui ont enfanté des femelles.

Exemple 47

أَخْلَبَ الرَّجُلُ: وَلَدَتْ إِبِلُهُ إِنَاثًا

"un tel *'ahlaba* : ses chameaux ont mis au monde des femelles"

L'habillement donne naissance à des concepts qui se reflètent dans le lexique arabe. Dans l'exemple 48, le verbe *taḍannaba* exprime le mode pour arranger le turban en laissant une partie sous forme de *ḍanab* "queue".

Exemple 48

تَذَنَّبَ الْمَعْمُومُ: ذَنَّبَ عِمَامَتَهُ

"un tel=portant=le=turban *taḍannaba* : *ḍannaba* son turban "il a arrangé son turban avec un *ḍanab* "une queue"

L'importance de la poésie, aussi, donne naissance à des concepts qui se reflètent dans le lexique arabe. Le verbe *ḥaššaba* dérive de la racine *ḥšb*, de laquelle dérivent le substantif *ḥašab* "bois" et l'adjectif *ḥašib* "rude/gros". Le verbe de forme II, *ḥaššaba*, dans l'exemple 49, est causatif avec le sens de "rendre rude (la poésie)".

Exemple 49

حَشَّابَ الشَّعْرَ: قَالَهُ مِنْ غَيْرِ تَنَوُّقٍ [...]

"[il] *ḥaššaba* la poésie : il l'a dit sans élégance [...]"

Et enfin, la religion et le rapport des musulmans avec Allah donnent une charge cognitive à la langue arabe, des fois, difficile même à traduire. Dans l'exemple 50a, le verbe *tāba* est combiné avec la préposition *'ilā* "à". Il a toujours un sujet humain, avec le sens de "retourner à Allah/se repentir". Par contre, dans l'exemple 50b, le verbe *tāba* est combiné avec la préposition *'alā* (d'ailleurs difficile à traduire dans ce contexte). Dans ce cas, le sujet est Allah avec le sens de "guider à/accepter la *tawbah*", c'est à dire la pénitence d'une personne. Enfin, dans l'exemple 50c, le verbe *istatāba* est 'demander la *tawbah*' toujours à Allah.

Exemple 50a

تَابَ إِلَى اللَّهِ [...] : رَجَعَ عَنِ الْمَعْصِيَةِ

"[il] *tāba* à Allāh : il a renoncé à la félonie".

Exemple 50b

تَابَ اللَّهُ عَلَيْهِ [...] : وَقَفَّهُ لِالتَّوْبَةِ

"Allāh *tāba* concernant lui : il l'a guidé/aidé à la *tawbah*".

Exemple 50c

اسْتَتَابَهُ : سَأَلَهُ التَّوْبَةَ

"[il] *istatāba*=Lui : il Lui a demandé la *tawbah*"

Dans SUMO, le concept «ReligiousProcess » subsume déjà certains concepts de la religion musulmane et peut subsumer, également, les verbes *tāba* de l'exemple 50a et *istatāba* de l'exemple 50c.

Par contre, aucun concept ne considère les faits et gestes de Allah. De conséquence, lier le verbe *tāba*, de l'exemple 50b, à un concept existant lui ferait perdre une charge religieuse qui marque l'identité musulmane et la langue arabe.

2.3.5.2.3 Problèmes morphosyntaxiques Dans le lexique arabe, en général, on peut trouver des lemmes qui sont au duel. C'est le cas des exemples 51a et 51b.

Il est intéressant de noter que, dans l'exemple 51a, *ad=dā'ibāni*, est le duel de *dā'ib*, nom d'agent du verbe *da'aba* 'persévérer/ être assidu'. Littéralement, le lemme *ad-dā'ibāni* se traduirait par "les deux persévérants". En fait il s'agit du «jour et la nuit », appelés aussi, *al=ḡadīdāni* "les deux nouveaux".

Exemple 51a

الدَّائِبَانِ : الْجَدِيدَانِ

les=*dā'ibāni* : les=deux=nouveaux

Dans l'exemple 51b, *al=ḥinnābatāni*, correspondant aux 'deux narines', reste également sans correspondance dans le lexique anglais.

Exemple 51b

الْحِنَابَتَانِ [...] : ظَرْفَا الْأَنْفِ

les=*ḥinnābatāni* : les=deux=côtés du nez

2.3.5.2.4 Problèmes morphosémantiques La dérivation est un phénomène très systématique et récurrent dans la langue arabe qui bénéficie d'une richesse lexicale incomparable. En correspondance, la langue anglaise fait recours aux phénomènes dérivationnelles de manière occasionnelle. De conséquence, on note une différence de structure et d'organisation entre le lexique arabe et celui de l'anglais, ce qui provoque des lacunes lexicales. Les exemples 52 et 53 peuvent être typiques pour représenter ce genre de lacunes lexicales.

Comme déjà expliqué dans la section théorique 3.2.2.2, le verbe dérivé, de forme IV, peut avoir une valeur estimative avec le sens de «trouver/estimer quelqu'un/quelque chose + adjectif » et dérive de la même racine que l'adjectif auquel il est lié. Justement, dans l'exemple 52, le verbe *'ağdaba* est transitif avec le sens de "trouver/estimer la terre *ğadb-ah*", où *ğadb* est un adjectif dont le sens est "aride".

Exemple 52

أَجَدَبَ الْأَرْضَ: وَجَدَهَا جَدْبَةً

[un tel] *'ağdaba* la terre : il l'a trouvée *ğadbah*

Dans d'autres contextes, le verbe de forme IV peut avoir une valeur expositive "être atteint/exposé de + substantif" où le substantif, auquel il est lié, désigne une calamité ou une maladie. Dans ce cas aussi, le verbe et le substantif dérivent de la même racine. Dans l'exemple 53, le verbe *'ağdaba* est intransitif avec le sens de "être atteint de *ğadb*", où *ğadb* est un substantif avec le sens "aridité".

Exemple 53

أَجَدَبَ الْقَوْمُ: أَصَابَهُمُ الْجَدْبُ

les personnes *'ağdaba* : ils ont été atteints par le *ğadb*

Grâce aux phénomènes dérivationnelles systématiques de la langue arabe, on trouve de forts liens morphosémantiques entre le verbe *'ağdaba* et l'adjectif *ğadb*, dans l'exemple 52, et entre le verbe *'ağdaba* et le substantif *ğadb*, dans l'exemple 53.

En général, ces verbes n'ont pas de correspondance dans la PWN, et de conséquence, on se trouve en face à des lacunes lexicales.

2.3.5.2.5 Problèmes ontologiques Une ontologie est conceptualisée en fonction des concepts qui sont exprimés dans une langue donnée avec les moyens qu'elle a à disposition. Une grande partie des cas sans correspondance (presque 30%) peuvent être expliquée par une lacune très significative dans le lexique arabe, concernant le verbe *être* «to be »:

be 02604760 | have the quality of being ; (copula, used with an adjective or a predicate noun); "John is rich"; "This is not a good answer" &%Entity+

La figure 2.45 donne quelques exemples de non correspondance des verbes arabes dont la traduction commence par «to be ».

```

<lexical_entry>
  <plain_text>أَخْدَوْدَبَ الرَّمْلُ: أَحْقُوْقَفْتُ.</plain_text>
  <lemma>أَخْدَوْدَبَ</lemma>
  <synonym att="أَحْقُوْقَفْتُ" />
  <translate lang="eng" value="to be crooked, vaulted, cambered, embossed, convex" />
  <external_links value="null"/>
</lexical_entry>
<lexical_entry>
  <plain_text>تَخَدَّبَ بِهِ: تَعَلَّقَ.</plain_text>
  <lemma>تَخَدَّبَ</lemma>
  <synonym att="عَلَّقَ" />
  <translate lang="eng" value="to be attached, be devoted , be fond" />
  <external_links value="null"/>
</lexical_entry>
<lexical_entry>
  <plain_text>كَرِبَ، كَفَّرِحَ: كَلِبَ، وَاشْتَدَّ غَضَبُهُ، فَهُوَ حَرِبٌ مِنْ حَرَبٍ.</plain_text>
  <lemma>كَرِبَ</lemma>
  <synonym att="كَلِبَ" />
  <translate lang="eng" value="to be furious, enraged, angry" />
  <external_links value="null"/>
  <sense att="وَاشْتَدَّ غَضَبُهُ" />
  <translate lang="eng" value="null" />
  <external_links value="null"/>
</lexical_entry>
<lexical_entry>
  <plain_text>قَدْ حَسَبَ حَسَابَةً، كَخَطَبَ خُطَابَةً، وَحَسَبًا، مُحَرَّكَةً، فَهُوَ حَسِيبٌ مِنْ حُسَبَاءَ.</plain_text>
  <lemma>حَسَبَ</lemma>
  <sense att="null" />
  <translate lang="eng" value="to be of noble origin, be highborn; to be highly esteemed, be valued" />
  <external_links value="null"/>
</lexical_entry>

```

Fig. 2.45 : Exemples de non correspondance due au manque du verbe «to be » en langue arabe.

Ce que une langue indo-européenne formule avec «être+adjectif », la langue arabe l'exprime par un verbe d'état. Par contre, pour exprimer un état résultatif, une langue indo-européenne utilise surtout «être+participe », la langue arabe fait recours à la dérivation et, de conséquence, son lexique est très riche de verbes *muṭāwi'* (voir l'étude théorique, concernant les verbes dérivés en *t*, dans 3.2.2.5). Par conséquence, une divergence lexicale énorme se crée entre un lexique de langue indo-européenne et le

lexique arabe. D'un côté, un seul verbe «être», et de l'autre, autant de verbes d'état et de verbes résultatifs (ou *muṭāwi'*) qu'adjectifs et participes.

2.4 Représentation standard de la ressource lexicale

Convertir *qāmūs* en format standard permet de rendre l'information contenue disponible, sous forme structurée pour les recherches lexicographiques et pour les applications NLP. Cela signifie que, non seulement, il sera possible de faire des recherches des classes de mots spécifiques et de voir comment ils sont décrits dans *qāmūs*, mais aussi d'utiliser ces informations pour le traitement des textes arabes, en particulier, les textes anciens.

2.4.1 LMF : Lexical Markup Framework

Développé en collaboration par une équipe de spécialistes de la lexicographie informatique et du traitement du langage naturel, LMF (Lexical Markup Framework) a le statut de la norme ISO 24613 :2008. Il définit un standard abstrait d'architecture pour la représentation des lexiques informatiques, qui obéit à des contraintes, décrites dans une DTD. Le modèle du noyau de LMF, illustré dans la figure 2.46, regroupe les classes suivantes :

- Lexical Ressource : représente la totalité de la ressource et peut contenir un ou plusieurs Lexicon d'une ou plusieurs langues ;
- Global Information : regroupe les métadonnées concernant la ressource lexicale tels que le titre, l'auteur, etc ;
- Lexicon : représente un lexique regroupant toutes les entrées lexicales d'une langue donnée ;
- Lexical Entry : représente l'entrée lexicale ;
- Form : est une classe abstraite qui permet de gérer les différentes variantes orthographiques du mot de l'entrée lexicale, la prononciation, etc ;
- Form Representation : représente une variante orthographique si le mot possède plus d'une transcription phonologique.
- Representation : une classe abstraite qui représente une variante orthographique.


```

<lexical_entry>
<plain_text>كَبُرَ، كَبُرًا، كَعْتَبَ، وَكَبُرًا، بِالضَّمِّ، وَكَبَارَةٌ، بِالْفَتْحِ: تَقْبِضُ صَغَرَ، فَهُوَ كَبِيرٌ
</plain_text>
<lemma>كَبُرَ</lemma>
<imperfective>يَكْبُرُ</imperfective>
<pos>Verb</pos>
<morphological_information>I</morphological_information>
<root_information>I</root_information>
<morphosyntactic_information>Intransitive</morphosyntactic_information>
<masdar>كَبْرٌ</masdar>
<masdar>كَبْرٌ</masdar>
<masdar>كَبَارَةٌ</masdar>
<adjective>كَبِيرٌ</adjective>
<adjective>كَبَارٌ</adjective>
<Plural>كَبَارٌ</Plural>
<Plural>كَبَارُونَ</Plural>
<Plural>مَكْبُورَاءٌ</Plural>
<sense att="تَقْبِضُ صَغَرَ">
<antonym att="صَغَرَ"/>
<translate lang="eng" value="to grow" />
<external_links >
  <concept ref="SUMO" value="Increasing+">
  <synset ref="PWN" ver="3.0" id="00230746" />
  </concept>
</external_links>
</sense>
<root>ك ب ر</root>
</lexical_entry>

```

Fig. 2.47 : Entrée lexicale du lemme *kabura* 'être grand'

Selon les règles imposées par la macrostructure et la microstructure de *qāmūs*, le résultat des extractions dans le fichier .XML présente, dans plusieurs cas, la répétition d'un lemme avec des sens différents. Par exemple, le verbe *'abba*, listé six fois dans *qāmūs*, présente le cas typique de l'organisation lexicale dans un dictionnaire médiévale. De conséquence, dans le fichier XML, le verbe *'abba* est répété dans six entrées lexicales avec six sens différents.

Durant la conversion automatique du fichier XML en LMF, on a dû implémenter une phase d'analyse qui détecte la répétition des lemmes. Le système de conversion ne doit pas répéter un lemme donné, mais il doit créer un sens, avec un identificateur n+1, qu'il ajoute à l'entrée déjà existante. Par exemple, le système, en parcourant le fichier, doit créer une seule entrée lexicale du verbe *'abba* liée à six sens différents.

L'entrée lexicale hérite la catégorie grammaticale, *'abba* est, dans tous les cas, un verbe trilitère de forme I. Par contre, les caractéristiques syntaxiques varient en

fonction du contexte et du sens et sont explicitées dans la partie de LMF, appelée SyntacticBehaviour. La figure figure 2.48 montre une partie des résultats.

```

<Sense id="6_أَب">
  <Definition>
    <feat att="Qamus" val="أَب الشيء: حَرْكَة" />
  </Definition>
  <SenseRelation targets="حَرْكَة">
    <feat att="label" val="Synonym" />
  </SenseRelation>
  <MonolingualExternalRef>
    <feat att="type" val="near_syn" />
    <feat att="ExternalSystem" val="PWN3.0" />
    <feat att="ExternalReference" val="01850315" />
  </MonolingualExternalRef>
  <MonolingualExternalRef>
    <feat att="type" val="near_syn" />
    <feat att="ExternalSystem" val="SUMO" />
    <feat att="ExternalReference" val="Transfer+" />
  </MonolingualExternalRef>
</Sense>
<SyntacticBehaviour>
  <feat att="1_أَب" val="Intransitive" />
  <feat att="2_أَب" val="Intransitive" />
  <feat att="3_أَب" val="Transitive" />
  <feat att="4_أَب" val="Intransitive" />
  <feat att="5_أَب" val="Intransitive" />
  <feat att="6_أَب" val="Transitive" />
</SyntacticBehaviour>
</LexicalEntry>

```

Fig. 2.48 : Une partie de l'entrée lexicale 'abba

Cependant, selon le format standard, explicité dans la DTD de LMF, les racines sont, également, encodées en tant qu'entrées lexicales avec type root. Le choix de coder les racines comme *LexicalEntry* est problématique pour des raisons théoriques et pratiques. Tout d'abord, la définition d'entrée lexicale implique qu'il devrait être un lexème, à savoir une unité ayant une signification lexicale précise, ce qui n'est pas le cas des racines. Les racines arabes sont, en fait, des informations méta-lexicales qui regroupent les lexèmes dans des classes sémantiques, mais, elles-mêmes ne sont pas dotées d'une signification spécifique. Cela ne signifie pas que les racines ne devraient

pas être répertoriées. En fait, la régularité de la dérivation en arabe est telle que toute personne, utilisant un lexique arabe, s'attend à pouvoir chercher une racine et extraire toutes les entrées lexicales qui en dérivent. En outre, dans LMF, toutes les entrées lexicales, verbes, noms et adjectifs sont liés à la racine par une relation RelatedForm, sans aucune hiérarchie, comme démontré dans la figure 2.49 qui illustre une représentation partielle du résultat obtenu de la conversion du verbe *kabura*.

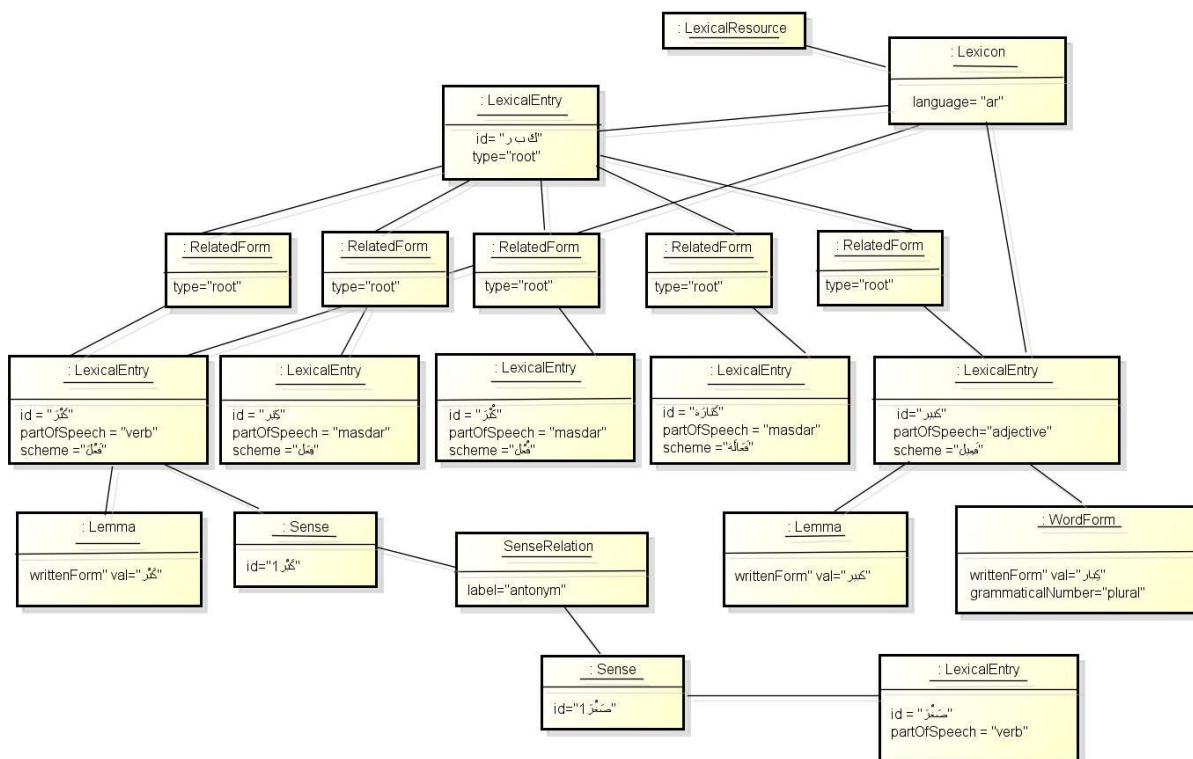


Fig. 2.49 : Schème partiel de la version LMF concernant l'entrée lexicale de *kabura*

En second lieu, la dérivation, dans la langue arabe, n'est pas seulement un phénomène morphologique, mais surtout, sémantique. Les paroles dérivent du *maṣḍar* corrélatif au verbe de base et, chacune d'elles portent une variation sémantique qui doit être formalisée.²⁸

En plus, au niveau de la représentation de stockage des données, LMF propose un format de sérialisation, uniquement, en XML. D'autres tentatives se sont focalisées pour représenter les ressources sous forme d'une ontologies accompagnées d'infrastructure robuste pour le stockage mais surtout pour l'échange et l'accès aux données. Ces efforts ont mené à la création du format Lemon (LEXicon Model for ONtologies), un

28. Pour plus de détail, voir le chapitre 2 qui traite des notions de dérivation, en particulier 2.2, .

modèle qui permet de représenter les informations lexicales relatives à une ontologie publiée sur le web sémantique.

2.4.2 LEMON : LExicon Model for ONtologies

Sur la base des technologies du web sémantique, le modèle Lemon a été développé à partir de LMF (Lexical markup framework) et contient le modèle du noyau LMF mais dans un langage d'ontologie. Le modèle noyau de LEMON, illustré dans la figure 2.50, est constitué par :

- **Ontology Entity** : l'entité ontologique qui décrit la signification du concept d'une manière indépendante de la langue ;
- **Lexical Sense** : permet de lier toutes les significations du mot ou du terme ;
- **Lexical Entry** : représente le mot ou le terme lui-même ;
- **Lexical Form** : utilisé pour décrire une entrée ou une forme (p. ex., le pluriel, duel, etc.) ;
- **Written Representation** : la réalisation précise de l'entrée lexicale.

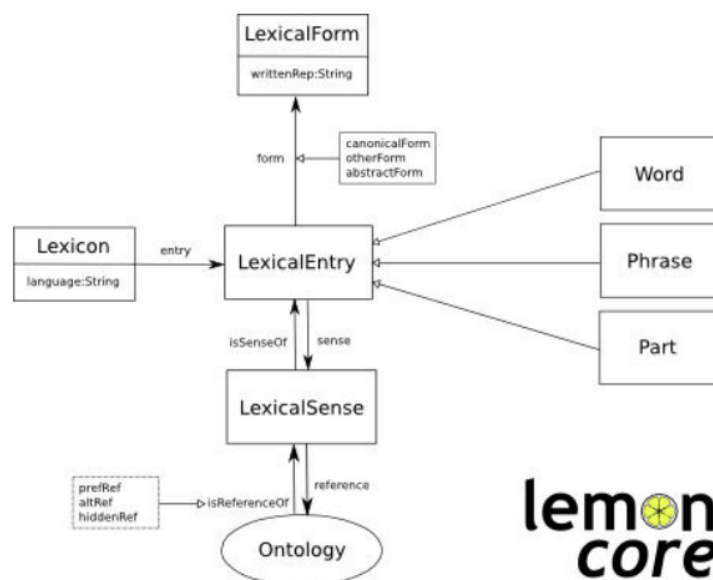


Fig. 2.50 : Modèle noyau de LEMON

Pour permettre une description linguistique complexe, LEMON utilise le vocabulaire LexInfo.²⁹ Il est écrit en OWL (Ontology Web Language) qui utilise le modèle RDF (Resource Description Framework), un langage qui permet de décrire les métadonnées et les données pour faciliter leur traitement. Les éléments les plus importants de OWL sont les classes et les propriétés :

- Les classes : une classe définit un groupe d'individus qui sont réunis parce qu'ils ont des caractéristiques similaires. L'extension de la classe est constituée de l'ensemble des individus d'une classe, appelés aussi les instances ;
- Les propriétés : permettent d'exprimer des faits au sujet des classes ; on parle dans ce cas de propriétés de type de donnée. Par contre, les propriétés d'objets permettent de relier entre les instances.

Cependant, un étude plus approfondie de LEMON fait noter que, comme dans le cas de LMF, la représentation lexicale est basée sur les entrées lexicales isolées, sans considérer les liens sémantiques et paradigmatisques qui existent entre les lemmes appartenant à la même famille dérivationnelle.

En plus, on note des problèmes de structuration et de formalisation, comme illustré dans la figure 2.51. Selon la DTD de LEMON, les formes dérivées «*derivedForm*» et la racine «*Root*» sont des sous propriété de *Lexical_Variant* qui, à son tour, est une propriété de *Lexical_Entry*. En outre, la propriété *Lexical_Variant* est définie comme «une relation non sémantique entre deux entrées lexicales »!

Plusieurs travaux récents prouvent qu'il est nécessaire de préserver la structure morphosémantique du lexique.³⁰ En exploitant les liens morphologiques et sémantiques entre les paroles de la même famille dérivationnelle, on arrive à automatiser, de manière plus efficace, la création des lexiques, la génération automatique et homogènes des définitions, la recherche de l'information, etc.

La structure du lexique arabe a une représentation séculaire, basée sur les relations morphologiques et dérivationnelles qui regroupent les paroles dérivant de la même racine. La représentation du lexique arabe en format, LMF ou LEMON, qui ne considère pas ces aspects lui fait perdre la structure morphosémantique qui le caractérise.

Une alternative viable serait de modifier le format LMF et/ou LEMON, de manière à créer une classe transcendante, qu'on peut appeler *Root* et qui représente les racines

29. <http://lexinfo.net/>

30. Ligozat & al., 2012; L'Homme & al., 2016; Fillmore & Baker, 2009.

arabes. Par contre, *Lexical_Entry*, devient une sous-classe liée à *Root* par une relation *subClassOf* et les entrées lexicales, faisant partie de la même classe *Root*, doivent être automatiquement liées entre elles par une relation *derivedForm*.

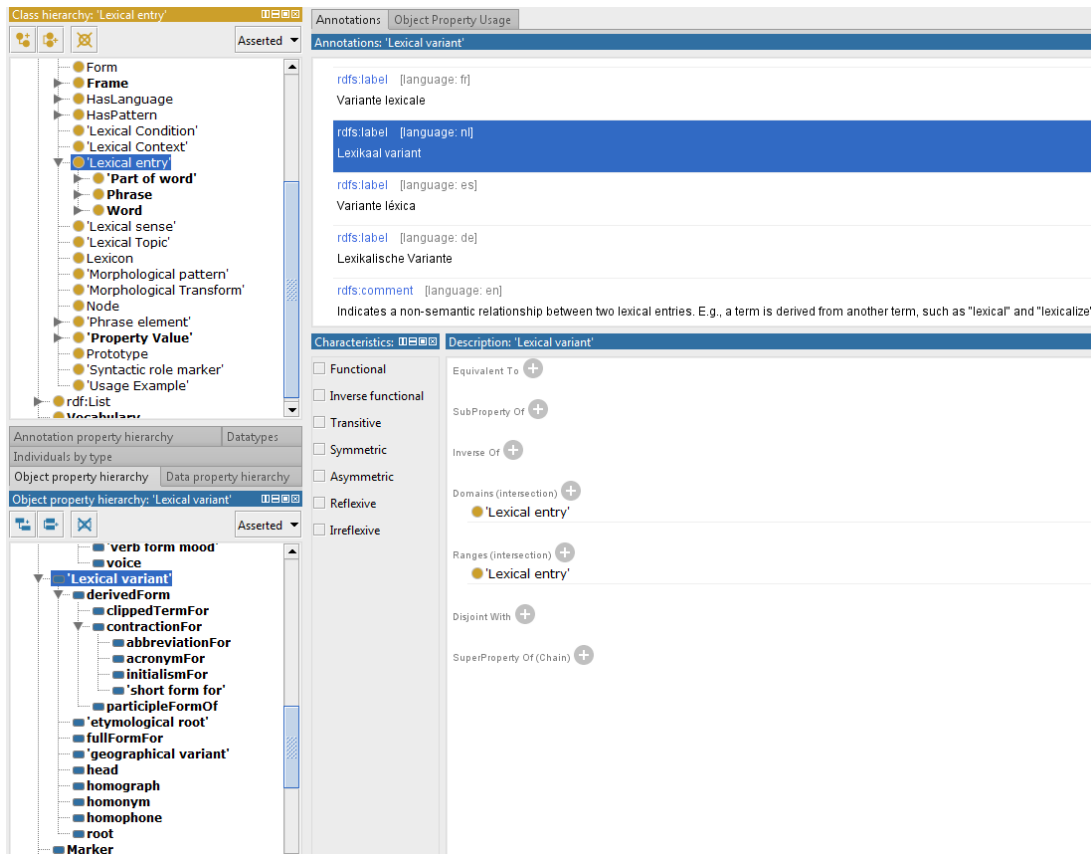


Fig. 2.51 : Organisation et caractéristiques de LEMON

2.5 Éditeur LexO

En collaboration avec le laboratoire «literary computing»³¹, on a adopté LexO, un éditeur collaboratif pour les ressources multilingues lexicales et termino-ontologiques. Basé sur le modèle Lemon, il vise à soutenir les lexicographes et les terminologues dans leur travail. Bien que le développement de LexO soit toujours en cours, l'éditeur est déjà utilisé dans d'autres projets de recherche, comme DiTMAO et Totus Mundus.³² La figure 2.52 illustre certaines fonctionnalités de LexO, qui permettent l'insertion des entrées lexicales. Pour chaque sens, on peut spécifier le type de relation sémantique

31. <http://licolab.ilc.cnr.it/index.php/it/>

32. Bellandi & al., 2017

qui le lie au lemme concerné, telle la synonymie, l'antonymie, l'hyperonymie, l'hyponymie et la troponymie.

En plus, il donne l'opportunité d'importer la ressource en format LEMON grâce à la fonctionnalité «Export », qu'on trouve en bas et à droite.

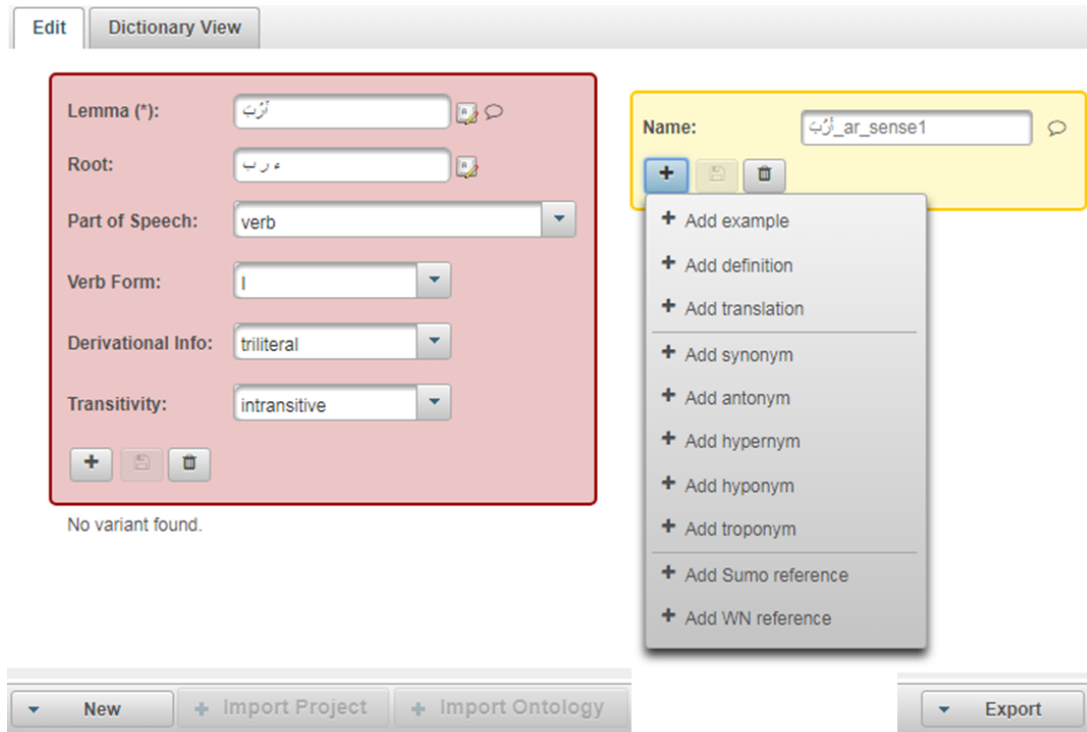


Fig. 2.52 : Interface LexO

Par contre, la figure 2.53 illustre l'exemple du verbe *'adaba* avec les deux sens extraits à partir de *qāmūs*.

En plus de la recherche d'un lemme précis, La figure 2.54 illustre une autre fonctionnalité qui a été ajoutée pour trouver tous les lemmes appartenant à la même famille dérivationnelle. Ainsi, on peut chercher tous les lemmes qui dérivent de la même racine.

CRÉATION D'UNE RESSOURCE SÉMANTIQUE
al=qāmūs al=muḥīṭ

The screenshot shows the LexO interface for the lemma 'adaba'. On the left, a sidebar lists various senses of the lemma. The main area is divided into two panels. The top panel (red border) shows the lemma details: Lemma (*): ادب, Root: ادب, Part of Speech: verb, Verb Form: ا, Derivational Info: triliteral, and Transitivity: transitive. The bottom panel (yellow border) shows two senses of the lemma. The first sense (adaba_ar_sense1) has a definition 'دبة فائده: دعاء إلى طعامه، كدته اياديا', a translation 'to invite, ask to come', and a Wordnet ref '01469770'. The second sense (adaba_ar_sense2) has a definition 'د ب ا ب ت ا د ب ا، مخرجة - حسان مخرجة وتكونه', a translation 'make', and a Wordnet ref '01664172'.

Fig. 2.53 : Exemple de l'entrée lexicale 'adaba dans LexO

The screenshot shows the LexO interface for searching by root. The 'Query filter' section is active, and a dropdown menu is open over the 'Select all the entries with root' field, showing a list of roots including 'ع ر ب', 'ه و ح', 'ن ف س', 'ق ي ح', 'ع د ب', 'د ع ي', 'و ر م', and 'ن ت س ط'. The main area shows the details for the lemma 'ع ر ب' (arab). The top panel (red border) shows the lemma details: Lemma (*): عرب, Root: ع ر ب, Part of Speech: verb, Verb Form: ا, Derivational Info: triliteral, and Transitivity: intransitive. The bottom panel (yellow border) shows two senses of the lemma. The first sense (arab_ar_sense1) has a definition 'عرب العرب: يقى اثره بعد الضوء', a translation 'cicatrize', and a Wordnet ref '01551679'. The second sense (arab_ar_sense2) has a definition 'عربك معجته: فسد', a translation 'to be disordered, swollen', and a Wordnet ref '00256507'.

Fig. 2.54 : Fonctionnalité de la recherche à travers la racine

Cinquième partie

DISSERTATION ET CONCLUSIONS

Dans cette thèse de doctorat, le travail applique les procédures de conversion d'un lexique arabe médiéval, *al=qāmūs al=muḥīṭ*, en format numérisé XML. La méthodologie utilisée, décrite dans la troisième partie de la thèse, consiste à extraire, de manière automatique, les informations contenues dans le lexique afin de les rendre explicites. Les extractions exigent une connaissance approfondie des caractéristiques de la langue arabe, décrites dans la première partie de la thèse, et des caractéristiques de la lexicographie arabe, décrites dans la deuxième partie. Telles connaissances permettent d'identifier les structures récurrentes qui sont, ensuite, formalisées et exploitées par des algorithmes, créés ad hoc, afin d'identifier les propriétés morphosyntaxiques et sémantiques des entrées lexicales. Pour vérifier la cohérence des procédures, le contrôle de la section de la lettre *bā'* a été effectué manuellement. Cependant, il faut signaler que les algorithmes développés sont généraux et peuvent être appliqués à tout le lexique.

La deuxième tâche consiste à enrichir le lexique numérisé, *qāmūs*, par les informations sémantiques et conceptuelles permettant de l'analyser et de l'utiliser avec une précision majeure. Dans ce but, les procédures adoptées consistent à aligner la version obtenue du lexique arabe avec la ressource lexico-sémantique Princeton WordNet de la langue anglaise (PWN) et l'ontologie universelle SUMO (Suggested Upper Merged Ontology). L'alignement se fait par le biais de la traduction des lemmes arabes, de *qāmūs*, en utilisant deux dictionnaires bilingues arabe-anglais disponibles en format numérisé (An Arabic-English et Dictionary of Modern Written Arabic). Les traductions obtenues permettent d'identifier les concepts et les relations sémantiques, existantes dans PWN et SUMO, qui sont ensuite ajoutés aux définitions des lemmes.

Cette partie du travail a été très intéressante, parce qu'elle a permis de mettre en évidence les aspects communs entre la langue arabe et la langue anglaise pour construire une éventuelle ontologie commune. Mais, elle a permis, surtout, de souligner les aspects différents entre les deux langues qui sont à l'origine des lacunes lexicales provoquant ainsi des problèmes d'alignement.

Les lacunes, dues aux problèmes stylistiques, de religion et de culture, peuvent être colmatées. Il suffit de faire l'effort de mieux exprimer les définitions données par *qāmūs* et/ou raffiner les concepts existants et/ou en ajouter de nouveaux pour présenter au mieux la définition du lemme ou l'aspect spécifique de la culture arabe et musulmane.

Cependant, la perplexité s'insinue lorsqu'on se penche sur les autres typologies de problèmes. Les lacunes lexicales dues aux caractéristiques morphosémantiques sont plus ou moins une conséquence des différences ontologiques sous-jacentes dans le lexique arabe et celui de l'anglais.

Plusieurs propriétés linguistiques spécifiques ont conduit la langue arabe à faire son compte avec le temps. D'un côté, le temps est un concept intrinsèque du verbe arabe et, de l'autre, évènements et états sont exprimés par une forme verbale. De conséquence, le verbe arabe fait usage d'une dimension de temps qui exprime la durée du verbe d'état et se penche sur le point final de l'action. En plus de son sens, il exprime la différence entre action vs état ; durable vs temporaire ; réel vs estimé (non réel), etc. En se basant sur un système de dérivation systématique et presque formel, toutes les différenciations se fondent avec la morphologie.

Les ontologies, PWN et SUMO, conçues selon les propriétés linguistiques de la langue anglaise, ne peuvent rendre la notion de temps, d'état et d'action sous-jacents dans des verbes comme :

عَسَرَ 'asura "être impossible/compliqué/difficile [toujours]" ;

عَسِرَ 'asira "devenir polémique/insupportable" ;

عَسَرَ 'asara "contrarier".¹

Dans la PWN, et de conséquence dans SUMO, les connaissances sont organisées en compartiments presque sans liens entre elles. La prévalence est donnée aux noms pour leur tangibilité et réalité. Les adjectifs, par contre, ont un statut isolé et ne sont presque pas envisagés du point de vue ontologique.

Par exemple, l'adjectif *red* est lexicalisé par deux synsets différents, dans PWN, subsumés par le concept Red, dans SUMO :

red 00381097 | of a color at the end of the color spectrum (next to orange) ; resembling the color of blood or cherries or tomatoes or rubies &%Red=

red 00395626 | (especially of the face) reddened or suffused with or as if with blood from emotion or exertion ; "crimson with fury" ; "turned red from exertion" ; "with puffy reddened eyes" ; "red-faced and violent" ; "flushed (or crimson) with embarrassment" &%Red=

Bien que l'anglais n'ait pas de verbes d'état qui expriment "être + adjectif", on trouve certains verbes qui expriment "devenir + adjectif", par exemple, *red*den qui dénote des verbes causatifs ou résultatifs subsumés par des concepts différents :

redden 00103317 | turn red, as if in embarrassment or shame ; "The girl blushed when a young man whistled as she walked by" &%OrganismProcess+

1. *as-saraqustī, kitābu l-'af'āl*, V. I, p. 229-230

red 00534837 | make red ; "The setting sun reddened the sky" &%Process+

red 00289031 | turn red or redder ; "The sky reddened" &%Coloring+

Les ontologies correspondantes à la langue anglaise, PWN et SUMO, classent les différents sens de l'adjectif *red* et du verbe *red* sous des nœuds différents, sans spécifier les liens évidents qui existent entre eux.

Le temps, en langue anglaise, est exprimé par la temporalité du locuteur ; l'état et sa durée sont exprimés, en utilisant le verbe *être* et d'autres paroles, telles que, «maintenant», «hier» et «toujours». Par contre, la langue arabe distingue trois verbes pour exprimer "être/devenir rouge", comme illustré dans les exemples 54.²

Dans l'exemple 54a, le verbe *إِحْمَرَّ* *iḥmarra* est un verbe d'état qui décrit la qualité permanente «être rouge».

Exemple 54a

إِحْمَرَّ الشَّيْءُ: إِذَا لَزِمَ لَوْنُهُ فَلَمْ يَتَغَيَّرْ مِنْ حَالٍ إِلَى حَالٍ.

"une chose *iḥmarra* : si sa couleur [rouge] est permanente et ne change pas d'un état à l'autre."

Dans l'exemple 54b, le verbe *إِحْمَارَّ* *iḥmārra* est également un verbe d'état, mais il décrit la qualité «être rouge» comme 'araḍ "événement casuel" non permanent.

Exemple 54b

Définition de *iḥmārra* dans *lisān al-‘arab*

وَإِحْمَارَّ يَحْمَارُّ: إِذَا كَانَ عَرَضًا حَادِثًا لَا يَثْبُتُ

"*iḥmārra yaḥmārru* : si c'est un accident qui s'est produit [qui] ne persiste pas"

Définition de *iḥmārra* dans *tāġ*

وَإِحْمَارَّ يَحْمَارُّ إِحْمَارًا: إِذَا كَانَ يَحْمَارُّ مَرَّةً وَيُصْفَرُّ أُخْرَى

"*iḥmārra yaḥmārru* : si, une fois, il devient rouge et une autre [fois, il devient] jaune".³

2. Ibn Manẓūr, *Lisān*, V. 4 ; pp 208-213

3. Ġawharī, *Ṣaḥāḥ*, V. 11 ; p. 88.

Sous la racine *ṣfr*, le verbe *إِصْفَرَّ* *iṣfarra* exprime la qualité permanente de l'être *jaune*. Par contre, *إِصْفَارًا* *iṣfāra* est un 'araḍ "évènement casuel" qui se manifeste chez l'homme.⁴

Dans l'exemple 54c, le verbe *حَمِرَ* *ḥamira* ne peut pas être classé comme verbe d'état, c'est le processus de «devenir rouge » dû à la grande colère.

Exemple 54c

يُقَالُ حَمِرَ الرَّجُلُ [...] : إِذَا تَحَرَّقَ غَضَبًا وَغَيْضًا، وَهُوَ رَجُلٌ حَمِرٌ [...]

"on dit [que] un homme *ḥamira* [...] : s'il s'est enflammé de colère et de rage ; et lui [est] un homme *ḥamir* [...]"

La capacité de la langue arabe d'utiliser les processus de la dérivation lui donne un instrument puissant qui lui permet de colmater l'absence du verbe *être* et l'organisation lexicale en découle très différente. Les connaissances sont organisées de manière que les paroles qui les expriment soient liées entre elles. Verbe, noms et adjectifs forment un tout, les uns dérivant des autres.

L'adjectif dérive du verbe dont la notion du temps est intrinsèque, et de conséquence, la catégorisation de l'état en fonction du temps est également intrinsèque. Le verbe *إِحْمَرَّ* *iḥmarra*, exprimant la qualité permanente «être rouge », est lié à l'adjectif *أَحْمَرٌ* *aḥmaru* "rouge". Par contre, le verbe *حَمِرَ* *ḥamira*, exprimant un processus, est lié à un autre adjectif, *ḥamir* "rouge [de colère]".⁵

Les verbes d'état *إِحْمَرَّ* *iḥmarra* et *إِحْمَارًا* *iḥmāra* se trouvent sans correspondance dans la langue anglaise.

Le verbe *حَمِرَ* *ḥamira*, qui exprime un processus, et l'adjectif qui lui est corrélé, *حَمِرٌ* *ḥamir* "rouge [de colère]", sont les seuls à avoir une correspondance dans PWN et dans SUMO :

red 00103317 | turn red, as if in embarrassment or shame ; "The girl blushed when a young man whistled as she walked by" &%OrganismProcess+

4. Ğawharī, Ṣaḥāḥ, Ibid. V. 12 p. 326

5. À ce propos, voir, également, les exemples 34 et 35 cités en 3.2.1.3.2

red 300395626 | (especially of the face) reddened or suffused with or as if with blood from emotion or exertion; "crimson with fury"; "turned red from exertion"; "with puffy reddened eyes"; "red-faced and violent"; "flushed (or crimson) with embarrassment"&% Red=

Un locuteur anglais trouve normal que l'adjectif *red* ne soit pas lié au verbe *redden*. Un locuteur arabe trouve l'affinité sémantique entre tous les verbes, les noms et les adjectifs qui dérivent de la racine *hmr* et ne peut concevoir de classer les paroles diversement ou de les «éparpiller» de manière qu'il n'y ait aucun lien entre elles. Toute les paroles, ayant une affinité sémantique et dérivant de la même racine, sont liées entre elles et doivent être classées sous un même nœud qui leur confère une identité sémantique partagée.

Un autre exemple qui démontre que l'organisation lexicale est différente entre les deux langues, est donné par les noms des lieux. En considérant les paroles, *cimetière* et *librairie*, par exemple, l'anglais utilise :

cemetery 115302802| A landArea which is used for burying the dead&%Cemetery=

library 103660664| a depository built to contain books and other materials for reading and study&%LibraryBuilding=

En arabe, pour exprimer un lieu caractérisé par l'abondance d'un objet/animal, la parole est dérivée de la même racine de l'objet/animal, en utilisant le schème maC₁C₂aC₃ah, qui a le sens de «lieu contenant beaucoup de objet/animal» :⁶

maqbarah 'cimetière' : lieu où on trouve beaucoup de *qabr* 'tombeau'.

maktabah 'librairie' : lieu où on trouve beaucoup de *kitāb* 'livre' ;

Mais, on trouve d'autres noms de lieu qui n'ont pas de correspondance en anglais :

ma'sadah : lieu infesté par/où vivent beaucoup de *'asad* 'lion'

maklabah : lieu infesté par/où vivent beaucoup de *kalb* 'chien'.

Faire correspondre la langue arabe à une ontologie conçue pour la langue anglaise lui ferait perdre ce qui lui confère son essence. Les exemples traités prouvent qu'une ontologie lexicale est le fruit de l'expérience d'un groupe de personne qui communiquent selon les moyens linguistiques qu'ils ont à disposition et ne peut correspondre à la perfection à une autre langue. Comme annoncé par Benjamin Lee Whorf :

6. voir également les exemples cités plus haut [4.2.2.2.4](#)

«Concepts of "time" and "matter" are not given in substantially the same form by experience to all men but depend upon the nature of the language or languages through the use of which they have been developed. They do not depend so much upon any one SYSTEM (e.g., tense, or nouns) within the grammar as upon the ways of analyzing and reporting experience which have become fixed in the language as integrated "fashions of speaking" and which cut across the typical grammatical classifications, so that such a "fashion" may include lexical, morphological, syntactic, and otherwise systematically diverse means coordinated in a certain frame of consistency ».⁷

Selon l'hypothèse Sapir-Whorf, la nature n'est pas « compartimentée » exactement de la même manière pour tout le monde et, la manière de découper et organiser les concepts est conditionnée par notre langue maternelle :

«We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face ; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds ».⁸

Beaucoup d'expériences ont été menées pour défendre ou rejeter le relativisme linguistique énoncé par l'hypothèse Sapir-Whorf.

Lucy John A. a montré que, pour mémoriser les couleurs, les locuteurs dont le lexique est plus riche pour les couleurs ont une meilleure performance que les locuteurs des langues dont le lexique en est moins riche.⁹ Cependant, d'autres expériences, menées sur les Danis de la nouvelle Guinée, ont montré que, malgré le fait que leur langue manque complètement des termes indiquant toute la gamme des couleurs basiques, ils *apprennent* rapidement les concepts de couleurs.¹⁰

L'objectif de ce travail n'est pas de défendre le relativisme linguistique. La projection sur la langue arabe des ontologies basées sur la langue anglaise a permis de mettre en évidence ce qui est commun et ce qui est différent entre ces deux langues. Mais le fait le plus important est que l'expérience constitue, en soi, la preuve que les différences entre les deux langues ne sont pas inaccessibles au langage, puisqu'on peut en parler et les décrire.

7. Whorf, *Language, Thought, and Reality*, p. 158.

8. Whorf, *Language, Thought, and Reality*, p. 213.

9. Lucy, *Language Diversity and Thought*.

10. Rosch, *Natural Categories*.

En conclusion et pour commencer le chemin vers une ontologie arabe, il est nécessaire qu'elle soit décrite en fonction de son individualité et en fonction de son propre système conceptuel qui a une structure différente par rapport à celui de l'anglais. Étant «presque » formel, le système morphologique de la langue arabe aide à formaliser les concepts exprimés. Conçu comme un système binaire fonctionnel, la création des paroles se basent sur deux concepts :

- racine qui lie et regroupe les mots appartenant sémantiquement à une même famille ;
- schème ou paradigme qui, «habillant » la racine, confère une catégorie et ajoute des traits morphosémantiques et/ou morphosyntaxiques.

Tout en continuant à analyser les correspondances entre la langue arabe et la langue anglaise, ce qui permet de retrouver les concepts communs qui les réunissent déjà, il faudrait, en parallèle, penser à une spécialisation plus poussée et plus explicite des propriétés et des relations du système ontologique qui serait capable d'intégrer les connaissances spécifiques de cultures et de langues aussi diverses :

From this fact proceeds what I have called the "linguistic relativity principle", which means, in informal terms, that users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world. From each such unformulated and naive world view, an explicit scientific world view may arise by a higher specialization of the same basic grammatical patterns that fathered the naive and implicit view. Thus the world view of modern science arises by higher specialization of the basic grammar of the Western Indo-European languages. Science of course was not CAUSED by this grammar ; it was simply colored by it. ¹¹

11. Whorf, *Language, Thought, and Reality*, p. 221..

Bibliographie

Al-'andalusī, an-Naḥwī al-Ġarnāṭī. *Tuḥfatu al-'arabi bi-mā fī l-qur'āni min al-ġaribi*. Éd. Samīr al-Mağdūb. Al-maktab al-islāmī, Beyrouth. 1983.

Al-'askarī, 'abū Hilāl al-Ḥasan Bnu 'abdallāh. *al-Furūq al-luġawiyya*. Éd. Muḥammad 'Ibrīm Salīm. Dār al-'ilm waṭ-ṭaqāfah, Caire.

Al-'astarabādī, Raḍī ad-dīn Muḥammad Bnu al-Ḥasan. *Šarḥ šāfiyyat ibn al-Ḥāġib*. Éd. Muḥammad Nūr al-Ḥasan, Muḥammad az-Zafzāf et Muḥammad Muḥyī d-dīn 'abd al-Ḥamīd. 4 Volumes. Dār al-kutub al-'ilmiyyah, Beyrouth. 1982.

Al-'astarabādī, Raḍī ad-dīn Muḥammad Bnu al-Ḥasan. *Šarḥ ar-raḍiyy li-kāfiyyati bni l-Ḥāġib*. Éd. Ḥasan Muḥammad Bnu Ibrāhīm al-Ḥifzī. Haġar liṭ-ṭibā'ah wan-našr wat-tawzī' wal-'i'lān, Ġīzah. 1993.

Al-'azharī, Abū Mansūr Muḥammad Bnu 'Aḥmad. *Tahḍīb al-luġah*. Éd. 'Abdalsalām Hārūn Et. al. 15 Volumes. Al-Mu'assasah Al-Miṣriyyah Al-'āmmah li-l-ta'līf wa-l-Anbā' wa-n-našr & Ad-dār al-miṣriyya li-t-ta'līf wa-t-tarġamah. Caire. 1964–67.

Baalbaki, Ramzi. *The Arabic Lexicographical Tradition : From the 2nd/8th to the 12th/18th Century*. Handbook of Oriental Studies. Brill, Leiden, Boston, 2014.

al-Baġdādī, 'Abd al-Qādir bnu 'Umar. *al-'Adab wa-lubb lubāb Lisān al-'Arab*. Éd. 'Abd as-Salām Muḥammad Hārūn. 13 volumes. Maktabat al-Ḥāniġī, Caire, 4^e édition, 1997.

Bandanīġī, 'Abū Bišr al-Yamān bnu al-Yamān. *at-Taqfiyah fī l-luġah*. Éd. Ḥalīl 'Ibrāhīm al-'Aṭiyyah. Maṭba'at al-'Ānī, Bagdad. 1976.

Baṭalyawsī, 'abū Muḥammad 'Abdallāh bnu Muḥammad bnu s-Sīd al-Baṭalyawsī. *al-Muṭallaṭ*. Éd. Šalāḥ Maḥdī al-Farṭūsī. 2 volumes. Dār ar-Rašīd, Bagdad. 1981.

Bellandi A., Giovannetti G., Piccini S. et Weingart A. *Developing LexO : a Collaborative Editor of Multilingual Lexica and Terminological Resources in the Humanities* ». Dans LOTKS. 2017.

Ben Braham, M. *Le pluriel brisé* (d'après les grammairiens arabes). Éd. Ernest Leroux. Paris. 1897.

Benmamoun, E. *Arabic morphology : The central role of the imperfective*. Dans *Lingua*, Elsevier B.V., Amsterdam. 1999. Pages 175-201.

Benveniste, É. *Actif et moyen dans le verbe*. Problèmes de linguistique générale. Gallimard, Paris 1966. Pages 168–175.

Berners-Lee, T. *Weaving the Web : The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. HarperBusiness, San Francisco. 1999.

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C.. Introducing the arabic wordnet project. Dans *Proceedings of the third international WordNet conference*. 2006. Pages 295–300. Citeseer

Blachère, R. *al-Jawhari et sa place dans l'évolution de la lexicographie Arabe*. ANA-LECTA. Presses de l'Ifpo, Damas. 1975.

Bohas, G. *Le PCO, la composition des racines et les conventions d'association*. Bulletin d'Études Orientales, tome XLIII. Damas. 1991. Pages 119-137.

Bohas, G. et Dat, M. *Une théorie de l'organisation du lexique des langues sémitiques : matrices et étymons*. ENS Editions, Lyon. 2007.

Bohas, G. *Ibn Khaldoun et la situation linguistique du monde arabe à son époque : description et explication*. Attarbiya wa ttakwin. 2007. Pages 9-23.

Boudelaa, S. et Marslen-Wilson, W. D. *Morphological units in the Arabic mental lexicon*. Cognition, 81(1). 2001. Pages 65-92.

Cantineau, J. *Racines et gabarits*. Mélanges offerts à William Margais par l'institut d'études islamiques de l'Université de Paris. Maisonneuve, Paris. 1950.

Carter, M. G. *Arabic lexicography*. Éd. Latham J. D. et Youg M. J. L. Learning and Science in the Abbasid Period ("The Cambridge History of Arabic Literature"). Cambridge University Press. 1990. Pages 106–117.

Charlet, J., Bachimont, B., et Troncy, R. *Ontologies pour le Web sémantique*. Dans *Revue I3*, numéro Hors Série *Web sémantique*. 2004. Pages 43-63.

CHRISTOL, A. *Introduction - La causativité en latin : problèmes théoriques*. Dans *La causativité en latin*. Éd. Bernard BORTOLUSSI et Peggy LECAUDÉ. L'Harmattan, Paris, 2014. Pages 13- 26.

Cohen, D. *'Aḍḍād et ambiguïté linguistique en arabe*. Arabica 8, 1961. Pages 1–29.

El-Dahdah, A. *A Dictionary of Arabic Grammar in Charts and Tables*. Librairie du Liban, Beirut. 1999.

ad-Dānī, Abū 'Amr 'Uṭmān bmu Sa'īd. *fī naqṭ al-maṣāḥif*. Éd. 'Izzat Ḥasan. Dār al-Fikr al-'Arabī, Damas. 1960.

Danks, W. *The Arabic verb : form and meaning in the vowel-lengthening patterns*. Thesis, PhD Doctor of Philosophy. 2010.

Dichy, J. *Sens des gabarits et sens des racines en arabe : le principe de figement lexical (pfl) et ses effets sur le vocabulaire d'une langue semitique*. Éd. L. Panier et S. Remi-Giraud. La polysemie, Presses Universitaires de Lyon, 2002.

Dichy, J. *fa'ula, fa'ila, fa'ala : dispersion et régularités sémantiques dans les trois gabarits simples du verbe arabe*. Éd. Ditters, E. et Motzki, H. Approaches to Arabic Linguistics, presented to Kees Versteegh on his sixtieth birthday. Brill, Leiden. 2007. Pages 313-365.

al-Faramāwī, 'abdu l-ḥayyi l-ḥusaynu. *Qiṣṣatu n-naqṭi wa-š-šakli fī l-muṣḥafi š-šarīfi*. Dār an-nahḍah l-'arabiyyah ; Caire. 1978.

Fawwāz, ḥikmat kašlī. *al-qāmūs al-muḥiṭ ; dirāsah wa-taḥlīl wa-naqd*. Dār al-kutub al-'ilmiyyah, Beyrouth, Liban. 1996.

Fellbaum C. *WordNet : An Electronic Lexical Database*. MIT Press, Cambridge, MA. 1998.

Fillmore, C. J., Christopher R. J., and Petruck, M.R.L. *Background to FrameNet*. International Journal of Lexicography, volume 16.3. 2003. Pages 235-250.

Fillmore C. J. et Baker C. *A frames approach to semantic analysis*. Éd. B. Heine & H. Narrog. The Oxford Handbook of Linguistic Analysis. Oxford : Oxford University Press. 2009. Pages 313-340.

al-Fīrūzābādī, Maḡd ad-dīn 'abū ṭ-Ṭāhir Muḥammad bnu Ya'qūb. *al-qāmūs al-muḥiṭ*. Éd. Muḥammad Na'im al-'Arqūsūsi. Mu'assasat ar-risālah ; Beyrouth. 8^e édition, 2005.

Fleisch, H. *La conception phonétique des Arabes d'après le Sīr Šinā'at al-'l'rāb d'Ibn Ğinnī*. Zeitschrift der Deutschen Morgenländischen Gesellschaft, volume 108. 1958. Pages 74-105.

Fournet, A. *La reconstruction de l'indo-européen et la réalité du sémitique : convergences et perspectives*. Langues et Littératures du Monde Arabe, 7, ENS éd., Lyon. 2007. Pages 3-39.

Francoy, G. et George, M. *Language Resource Management. Lexical markup framework (LMF)*. Technical report, ISO/TC 37/SC 4 N453 (N330 Rev. 16). 2008.

Frisch, S. A., Broe M. B. et Pierrehumbert J. *Similarity and phonotactics in Arabic*. Rutgers Optimality Archive 223, 1997.

Frisch, Stefan A., Pierrehumbert, J. B. et Broe, M. B. *Similarity avoidance and the OCP*. Natural Language & Linguistic Theory. 2004. Pages 179-228.

al-Ġāḥiḡ, Abū 'Uṡmān 'Amr bnu Baḥr. *al-Bayān wa-t-tabyīn*. Éd. 'Abdalsalām Muḥammad Hārūn. 7^e édition, 4 volumes. Maktabat al-ḥāniġī, Caire, 1998.

Gandon F. *Ontology engineering : a survey and a return on experience*. Report No 4396, INRIA, 2002.

Goldsmith, J. A. *Autosegmental Phonology*. Doctoral dissertation, MIT, Cambridge, Massachusetts, 1976.

Ġawālīqī, 'abū Maṡṡūr Mawḥūb bnu 'Aḥmad. *al-mu'arrab min al-kalām al-a'ġamī 'alā ḥurūf al-mu'ġam*. Éd. 'Aḥmad Muḥammad Šākir. 2^e édition. Dār al-kutub al-miṡriyya, Caire. 1969.

Ġawḥarī, 'Abū Naṡr 'Ismā'īl bnu Ḥammād. *aṡ-ṡaḥāḥ, tāġ al-luġah wa-ṡaḥāḥ al-'arabiyyah*. Éd. Aḥmad 'Abd al-Ġafūr 'Aṡṡār. Dār al-'Ilm li-l-Malāyīn, 2^e édition, 6 volumes, Beyrouth, 1979.

Grand'henry, J. *Grammaire arabe à l'usage des arabe. Traduction française et commentaires des «éléments d'arabe, morphologie et syntaxe, II da Rachid Chartouni (Beyrouth)*. PEETERS et Publications Linguistiques de Louvain, Belgique. 2000.

Guarino, N. *Formal Ontology and Information Systems*. Dans *Formal Ontology in Information Systems*. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Éd. N. Guarino. IOS Press, Amsterdam. 1998. Pages 3-15.

Guerssel, M., et Lowenstamm J. *Ablaut in Classical Arabic measure I active verbal forms*. Studies in Afroasiatic grammar. 1996. Pages 123-134.

Ġurġānī, ʿAli bnu Muḥammad. *kitāb at-taʿrīfāt*. Éd. Muḥammad Bāsil ʿUyūn s-Sūd. Dār al-kutub al-ʿilmiyyah, Beyrouth. 2000.

Habash, N. *Arabic Morphological Representations for Machine Translation*. Dans *Arabic Morphological Representations for Machine Translation*. Éd. ʿAbdelhadi Soudi, Antal Van Den Bosch et Neumann Günter. Springer Publishing Company, Incorporated. 2007. Pages 263-285.

Haiman, J. *Iconic and Economic Motivation*. Language, Volume 59, n. 4. 1983. Pages 781-819.

Ḥalīl, ʿAbū ʿAbdalraḥmān al-Ḥalīl bnu ʿAḥmad al-Farāhīdī. *Kitāb al-ʿAyn*. Éd. Maḥdī al-Maḥzūmī et ʿIbrāhīm as-Sāmarrāʿī. 8 volumes. Dār al-Rašīd, Bagdad. 1980–85.

Hamzé, H. *Les parties du discours dans la tradition grammaticale arabe*. Dans *Les classes des mots. Traditions et perspectives*. Presses Universitaires du Luon. 1994. Pages 93-115.

Hamzé, H. *La position du sujet du verbe dans la pensée des grammairiens arabes*. Dans *Language et linéarité*. Éd. Pierre Cotte. Presses Universitaires du Septentrion. 1999. Pages 127-149.

Ḥaṭṭābī, ʿAbū Sulaymān Ḥamd bnu Muḥammad al-Ḥaṭṭābī al-Bustī. *Ġarīb al-Ḥadīṯ*. Éd. ʿAbd-al-Karīm ʿIbrāhīm al-ʿIzbāwī. 3 volumes. Ġāmiʿat ʿUmm al-Qurā, Mecca. 1982.

Haywood, J. A. *Arabic Lexicography : its history and its place in the general history of lexicography*. Leiden : E. J. Brill. 1960.

Hirst, G. *Ontology and the Lexicon*. Dans *Handbook on Ontologies*. Springer Berlin Heidelberg. 2004. Pages 209 - 229.

L'Homme M. C. et San Martin A. *Définition terminologique : systématisation de règles de rédaction dans les domaines de l'informatique et de l'environnement*. Cahiers de lexicologie 109 (2). 2016. Pages 147-174.

Ibn Kamāl Bāšā, Šams al-Dīn Aḥmad bnu Sulaymān. *Risāla fī taḥqīq taʿrīb al-kalimah al-ʿaġamiyyah*. Éd. Muḥammad Sawāʿī. Institut Français de Damas, Damascus. 1991.

Ibn al-'Anbārī, Abū Bakr Muḥammad bnu al-Qāsim. *kitāb al-'Aḍḍād*. Éd. Muḥammad 'abū l-Faḍl 'Ibrāhīm. al-Maktabah al-'Aṣriyyah, Beyrouth. 1987.

Ibn al-'Aṭīr, Maḡd ad-Dīn 'Abū s-Sa'ādāt al-Mubārak bnu Muḥammad. *an-Nihāyah fī ḡarīb al-Ḥadīṯ wa-l-'aṭar*. Éd. Ṭāhir 'Aḡmad az-Zāwī et Maḡmūd Muḥammad aṭ-Ṭanāḡī. 5 volumes. Dār al-Fikr, Beyrouth. 1979.

Ibn Fāris, ṭAbū l-Ḥusayn ṭAḡmad. *aṣ-Ṣāḡibī fī fiḡh al-luḡah wa-sunan al-'arab fī kalā-mihā*. Éd. 'Aḡmad Ḥasan Basaḡ. Dār al-kutub al-'ilmiyyah. Beyrouth. 1997.

Ibn ḡinnī, 'abū l-Faḡh 'Uṭmān. *al-Ḥaṣā'iṣ*. Éd. Muḥammad 'Aliyy al-Naḡḡār. 3 Volumes. al-maktabah al-'ilmiyyah. 1956.

Ibn ḡinnī, 'abū l-Faḡh 'Uṭmān. *al-Muḡtasabu fī tabyīni wuḡūhi ṣawāḡḡi l-qirā'ti wa-l-'iḡāḡi 'anhā*. Éd. 'Aliyy an-Naḡḡī Nāṣif et 'Abd al-Ḥalīm an-Naḡḡār. 2 volumes. Dār sazākīn li-ṭ-ṭibā'ati wa-n-naṣr, 2^e édition. 1986.

Ibn ḡinnī, 'abū l-Faḡh 'Uṭmān. *sirru ṣinā'at al-'i'rāb*. Éd. Ḥasan Hindāwī. Dār al-Qalam, 2^e édition, Damas. 1993.

Ibn Manzūr, ḡamāl ad-Dīn 'Abū l-Faḍl Muḥammad bnu Mukram. *Lisān al-'Arab*. 15 vols. Dār Ṣādir, Beyrouth. 1968.

Ibn Raṣīq, 'abū 'Alī l-Ḥasan al-Qayrawānī. *al-'Umdah fī maḡāsin aṣ-ṣi'r wa-ādābihi wa-naḡḡihi*. Éd. an-Nabawī 'Abd al-Wāḡid Ṣa'lān. 2 volumes. Maktabat al-ḡāniḡī. Caire. 2000.

Ibn Sallām al-ḡumaḡī, Abū 'Abdallāḡ Muḥammad. *Ṭabaḡāt fuḡūl al-ṣu'arā'*. Éd. Maḡmūd Muḥammad Ṣākir. 2 volumes. Dār al-Madanī, ḡaddah. 1980.

Ibn Sīda, Abū l-Ḥasan 'Alī bnu 'Ismā'īl. *al-Muḡhaṣṣaṣ*. Édition Būlāq, Égypte. 1316 - 1321 A.H. (1898-1903).

Ibn Ya'īṣ, Muwaffaq ad-Dīn Ya'īṣ bnu 'Alī. *Ṣarḡ al-mufaṣṣal*. Éd. Emile Badī' Ya'qūb. 6 volumes. Dār al-kutub al-'ilmiyyah. Beyrouth. 2001.

ISO 1087-1. International standard. Première édition 15-10-2000. http://disciplinas.stoa.usp.br/pluginfile.php/312608/mod_resource/content/1/ISO_1087-1_2000_PDF_version_%28en_fr%29_CPDF.pdf.

- Jarrar M. *Building a Formal Arabic Ontology [in Arabic]*. Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. Alecso, Arab League. Tunis. 2011.
- , A., Gargouri, B., Abdelwahed, A., et Francopoulo, G. *Modélisation des paradigmes de flexion des verbes arabes selon la norme LMF-ISO 24613*. In Proceedings of TALN. 2007.
- Khemakhem, A., Gargouri, B., Haddar, K., et Ben- Hamadou, A. *Lmf for arabic*. Dans LMF, Lexical Markup Framwork. Éd. Francopoulo G. John Wiley & Sons. 2013. Pages 83–98.
- Kouloughli, D. E. *Grammaire de l'arabe d'aujourd'hui*. Éd. Pocket. Presses de Cox & Wyman Ltd. Angleterre. 1994.
- Lahrouchi, M. *Co-occurrence restrictions in the vocalic patterns of Afroasiatic plurals*. Dans *Folia Orientalia*, 47 (2). 2011. Pages 89-102.
- Larkey, L., Ballesteros, L., et Connell, M. *Light stemming for Arabic information retrieval*. Dans *Arabic computational morphology*. Éd. Soudi A., Van den Bosch A. et Neumann G. Springer, Heidelberg, Germany. 2007.
- Lancioni, G. *Sull'ordinamento dei dizionari arabi classici*. Rivista degli studi orientali, 71. 1997. Pages 113-127.
- Larcher, P. *D'une grammaire l'autre : catégorie d'adverbe et catégorie de maf'ūl muṭ- laq*. Bulletin D'études Orientales, 43. 1991. Pages 139-59.
- Larcher, P. *Syntaxe et sémantique des formes verbales dérivées de l'arabe classique : vues "nouvelles" et questions en suspens*. Quaderni di Studi Arabi, Istituto per l'Oriente C. A. Nallino, 17. 1999. Pages 3-27.
- Larcher, P. *Le système verbale de l'arabe classique*. DPSI, Aix-en-provence, deuxième édition. 2012.
- Larcher, P. *Dérivation arabisante et istiqaq arabe : histoire d'un malentendu*. ENS éditions. 2008.
- Lasersohn, P. *Plurality, Conjunction and Events*. Studies in Linguistics and Philosophy. Springer Science&Business Media B.V. 1995.

Lenci, A., Montemagni, S., et Pirrelli V. *Testo e computer, Elementi di linguistica computazionale*. Carocci editore, Roma. 2012.

Ligozat AL., Tribout D. et Grau B. *Intérêt des ressources morphologiques pour la recherche d'information précise*. Dans Actes de Coria 2012. 2012.

Loukil, N., Haddar, K. et BenHamadou, A. *Normalisation de la représentation des lexiques syntaxiques arabes pour les formalismes d'unification*. In Proceedings of the 26th conference on Lexis and Grammar, Bonifacio. 2007.

Lowenstamm, J., Prunet, J. F. *Le tigrinya et le principe du contour obligatoire*. Phonologie des langues sémitiques. Éd. Université du Québec à Montréal Revue québécoise de linguistique. 1986. Pages 181-206.

Lowenstamm, J. *A propos des gabarits*. Recherches linguistiques de Vincennes. 2003. Pages 7-30.

Lucy J. A. *Language Diversity and Thought*. Cambridge University Press. Cambridge. 1992.

Manning, C. D. et Schütze H. *Foundations of statistical natural language processing*. MIT Press, Cambridge, MA. 1999.

McCarthy, J. J. *A prosodic theory of nonconcatenative morphology*. Linguistic Inquiry. The MIT Press. 1981. Pages 373-418.

McCarthy, J. J. *OCP Effects : Gemination and Antigemination*. Linguistic Inquiry. The MIT Press. 1986. Pages 207-263.

McCarthy, J. J. *Linear Order and Phonological Representation*. Linguistic Inquiry. 1986. Pages 71-100.

McCarthy, J. J. et Alan P. *Prosodic morphology and templatic morphology*. Dans *Perspectives on Arabic linguistics II : Papers from the second annual symposium on Arabic linguistics*. Éd. Eid M. et McCarthy J. J. John Benjamins, Amsterdam. 1990. Pages 1-54.

McCarthy, J. J. *L'infixation réduplicative dans les langages secrets*. Dans *Langages*. Les javanais. 1991. Pages 11- 29.

McCarthy, J. J. *The phonetics and phonology of Semitic pharyngeals*. Linguistics Department Faculty Publication Series. 1994. Pages 191 - 233.

- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., et Miller, K. *Introduction to WordNet : an on-line lexical database*. Practical Lexicography, A Reader. 2008.
- Moscato, S. *An Introduction to the Comparative Grammar of the Semitic Languages. Phonology and Morphology*. Otto Harrassowitz. Wiesbaden. Troisième édition. 1980.
- Muğamma' al-luğah al-'arabiyyah. *al-muğam al-wasīf*. Maktabat aš-šurūq ad-dawliyyah. 2004
- al-Murādī, al-Ḥasan bnu Qāsim. *al-ğanā d-dānī fī ḥurūfi l-ma'ānī*. Éd. Faḥr ad-Dīn Qabāwah et Muḥammad Nadīm Fāḍil. Dār al-kutub al-'ilmiyyah, Beyrouth. 1992.
- Naṣṣār, Ḥusayn. *al-muğam al-'arabī : naš'atu-hu wa-taṭawwuru-hu*. 2 volumes. Dār Miṣr li-l-Ṭibā'a, Cairo, 4^e édition. 1988.
- Niles, I., et Pease, A. *Origins of the Standard Upper Merged Ontology : A Proposal for the IEEE Standard Upper Ontology*. Dans Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology, Seattle, Washington, August 6. 2001.
- Noyer, R. R. *Features, Positions and Affixes in Autonomous Morphological Structure*. Thèse Ph.D. MIT, Cambridge, Mass. 1992.
- 'Omar 'Aḥmad Muḥtār. *muğam al-luğah al-'arabiyyah al-mu'āṣirah*. Édition 'ālam al-kutub. 2008.
- Owens, J. *The foundations of grammar - An introduction to medieval arabic grammatical theory*. John Benjamins Publishing. 1988.
- Pease, A, Niles, I, et Li, J. *The suggested upper merged ontology : A large ontology for the semantic web and its applications*. Dans Working notes of the AAI-2002 workshop on Ontologies and the Semantic Web. 2002
- Pierrehumbert, J. *Dissimilarity in the Arabic Verbal Roots*. Proceedings of NELS. 1993
- Prévot L., Huang C., Calzolari N., Gangemi A., Lenci A, et Oltramari A. *Ontology and the lexicon : A multidisciplinary perspective*. Dans *Ontology and the Lexicon : A Natural Language Processing Perspective*. Cambridge University Press. 2010. Pages 3-24.
- Quṭrub, 'abū 'Alī Muḥammad bnu l-Mustanīr. *al-'aḍḍād*. Éd. Ḥannā Ḥaddād. Dār al-'ulūm li-ṭ-ṭibā'ī wa-n-našr, Riyad. 1984.
- Rastier F. *Ontologie(s)*. Revue des sciences et technologies de l'information, série : Revue d'Intelligence artificielle. 2004. Pages 15-40.

Roman, A. *La création lexicale en arabe : étude diachronique et synchronique des sons et des formes de la langue arabe*. Presses Universitaires de Lyon. 2005.

Rosch E. H. *Natural categories*. Dans *Cognitive Psychology*. 1973. Pages 328-350.

Ryding, K. C. *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press. 2005.

Salmoné, H. A. *An Advanced Learner's Arabic-English Dictionary*. Librairie du Liban, Beirut. 1889.

aş-Şālih, Şubḥī. *dirāsāt fī 'ilmi al-luġah*. Dār al-'ilm li-l-malāyīn. Beyrouth, Liban. 2009.

as-Saraqusṭī, 'Uṭmān. *kitābu l-'af'āl*. Éd. Ḥusayn Muḥammad Muḥammad Šaraf. al-Hay'ah al-'āmmah li-šu'ūn al-maṭābi'i al-'amriyyah. 1975-1978.

Sībawayhi, 'abū Bišr 'Amr ibn 'Uṭmān. *al-Kitāb*. Éd. 'Abd as-Salām Muḥammad Hārūn. Maktabat al-ḥāniġī. Caire. 1982.

as-Suyūṭī, Ġalāl ad-dīn 'abū l-Faḍl 'Abdalraḥmān bnu 'abī Bakr. *al-'iqṭirāḥ fī 'ilm 'uṣūl al-naḥw*. Éd. 'Abdu al-Ḥakīm 'Aṭīyyah. Dār al-bayrūtī, Damas, 2^e édition. 2006.

as-Suyūṭī, Ġalāl ad-dīn 'abū l-Faḍl 'Abdalraḥmān bnu 'abī Bakr. *al-muzhir fī 'ulūm al-luġa wa-'anwā'i-hā*. Éd. Muḥammad 'Aḥmad Ġād al-Mawlā, 'Alī Muḥammad al-Biġāwī et Muḥammad 'abū l-Faḍl 'Ibrāhīm. 2 volumes. Dār at-ṭurāt, Caire. 3^e édition,

aš-Šidyāq, 'Aḥmad Fāris. *al-ġāsūs 'alā l-qāmūs*. Maṭba'at al-ġawā'ib, al-quṣṭantīniyyah. 1982.

Ṭa'ālibī, 'abū Manṣūr 'Abd al-Malik bnu Muḥammad. *Fiqh al-luġah wa-'asrār al-'arabiyyah*. Éd. Yāssīn al-'Ayūbī. al-Maktabah al-'Aṣriyyah, Beyrouth. 2^e édition. 2000.

The Leipzig Glossing Rules : Conventions for interlinear morpheme-by-morpheme glosses. <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>.

Versteegh, K. *The Explanation of Linguistic Causes : Az-Zaggagi's theory of grammar - Introduction, translation, commentary*. Amsterdam 1995.

Voisin, S., et Vittrant, A. *Pluriactionnalité temporelle en birman et wolof : les cas de répétition externe*. Bulletin de la Société de linguistique de Paris. 2012. Pages 353-377.

Wehr, H. *A Dictionary of Modern Written Arabic*. Éd. Cowan J. M. Ithaca, New York, Spoken Language Services. 1971.

Whorf, B. L. *Language, Thought, and Reality*. Éd. Carroll J. B. The Technology Press of Massachusetts Institute of Technology. 4^e édition. 1959.

Wright, W. *A grammar of the Arabic language*. The University Press, Cambridge, 3^e édition. 1896.

az-Zabīdī, 'Abū l-Faḍl Muḥammad Murtaḍā bnu Muḥammad. *tāğ al-'arūs min ġawāhir al-qāmūs*. Éd. Farrāğ 'Abd as-Satār 'Aḥmad. 40 vols. Maṭba'at al-Kuwayt. 1965-2001.

az-Zağğāğī, 'abū l-Qāsim 'Abd ar-raḥmān bnu 'Ishāq. *al-'ibḍāl wa-l-mu'āqabah wa-l-naḏā'ir*. Éd. 'Izz ad-Dīn at-Tanūḥī. al-Mağma' al-'ilmī al-'arabī bi-dimašq, Damascus. 1962.

Zeroual, I., et Lakhouaja A. *Clitiques-Stemmer : nouveau stemmer pour la langue Arabe*. The First National Doctoral Symposium On Arabic Language Engineering (JDI-LA'2014). 2014.

az-Zubaydī, 'abū Bakr Muḥammad bnu l-Ḥasan al-'Andalusī. *Laḥn al-'awāmm*. Éd. Ramaḍān 'Abd at-Tawwāb. 2^e édition. 2000. al-Maktabah al-ḥāniğī, Caire. 2000.