

CONSULTATION SUR PLACE

PRET

PEB

OUI

NON

NON

enssib**Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques****Mémoire de DEA
Sciences de l'Information et de la
Communication****Option 3 :
Systèmes d'information
documentaire****Les descripteurs complexes
en indexation automatique
arabe : Etude des structures
annectives extraites d'un
corpus économique****Mabrouka EL HACHANI****Directeurs de mémoire :
Mohamed HASSOUN et Joseph DICHY**

BIBLIOTHEQUE DE L'ENSSIB



802990C

mai 1998

Université Lumière
Lyon2Ecole Nationale Supérieure
des Sciences de l'Information
et des BibliothèquesUniversité Jean Moulin
Lyon3

enssib

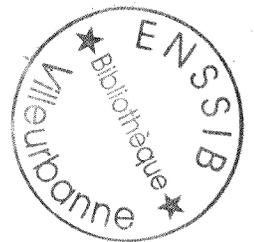
**Ecole Nationale Supérieure des Sciences
de l'Information et des Bibliothèques**

**Mémoire de DEA
Sciences de l'Information et de la
Communication**

**Option 3 :
Systèmes d'information
documentaire**

**Les descripteurs complexes
en indexation automatique
arabe : Etude des structures
annectives extraites d'un
corpus économique**

Mabrouka EL HACHANI



**Directeurs de mémoire :
Mohamed HASSOUN et Joseph DICHY**

mai 1998

Université Lumière
Lyon2

Ecole Nationale Supérieure
des Sciences de l'Information
et des Bibliothèques

Université Jean Moulin
Lyon3

1998
DEA
7

A ma famille

REMERCIEMENTS

Je remercie M. Mohamed HASSOUN pour ses précieux conseils et pour le suivi de ce travail.

J'exprime mes remerciements à M. Joseph DICHY pour la qualité pédagogique de sa direction.

Je remercie également Carole pour son aide et pour le témoignage de son amitié.

Je remercie toutes les personnes qui ont porté un intérêt sincère à ce travail et pour leurs encouragements.

**LES DESCRIPTEURS COMPLEXES EN INDEXATION
AUTOMATIQUE ARABE :
ETUDE DES STRUCTURES ANNECTIVES EXTRAITES D'UN
CORPUS ECONOMIQUE**

Mabrouka EL HACHANI

Sous la direction de Messieurs
M. HASSOUN J. DICHY
ENSSIB Université Lumière Lyon2

Résumé :

De nombreux systèmes d'information documentaire sont basés sur des logiciels d'indexation qui comportent pour beaucoup d'entre eux des modules linguistiques. Il ne s'agit pas de procéder à une analyse linguistique d'une part et à une typologie des problèmes d'indexation d'autre part.

Le but de notre étude est de repérer le fonctionnement syntactico-sémantique d'une certaine catégorie de descripteurs complexes en arabe qui sont représentés par l'annexion. L'étude de leur structuration permet d'apprécier leur portée sémantique, point essentiel en indexation.

Descripteurs en français : Documentation automatique ; TALN ; indexation automatique ; langue arabe ; descripteur complexe ; annexion

Abstract :

A majority of information retrieval systems are based on indexing software including linguistic tools. This study is not a linguistic analysis on the one hand and on the another hand a list of indexing problems. Our aim is to show the syntactic and semantic function of some complex descripteurs in arabic represented by annexation..

English key words : Automatic documentation ; natural language automatic process ; automatic indexing ; arabic language ; complex descriptor ; annexation

Table des matières

INTRODUCTION.....	1
<i>PREMIERE PARTIE.....</i>	5
I-Les méthodes d'indexation automatique.....	6
A-La méthode statistique.....	7
B-L'approche linguistique.....	9
C-La méthode mixte.....	10
D-La méthode par assignation.....	11
E-Les autres méthodes.....	12
1-Les systèmes experts.....	12
2-L'indexation automatique des titres.....	13
II-L'approche linguistique de l'indexation automatique.....	13
A- Les différentes composantes de l'analyse linguistique.....	14
1-L'analyse morphologique.....	15
2-L'analyse syntaxique.....	16
3-L'analyse sémantique ou lexicale.....	21
4-L'analyse pragmatique.....	22
B- La représentation informatique des modules linguistiques dans un logiciel documentaire.....	22
C- Le rôle de l'approche linguistique en indexation.....	23
III-L'importance du syntagme nominal dans les méthodes de l'indexation automatique.....	24
A- Le syntagme nominal et le Système d'Information Documentaire.....	24
1-Le modèle SYDO.....	25
2-L'équipe SAMIA.....	28
B-L'intérêt du syntagme nominal complexe en indexation automatique.....	32
<i>DEUXIEME PARTIE.....</i>	39
I-La définition de l'annexion dans la tradition linguistique arabe.....	40
a) L'annexion dite "pure".....	42
b) L'annexion formative ou qualificative.....	44
II-Les différentes structures annectives dans la tradition grammaticale arabe.....	45
A-L'annexion nominale.....	45
1- Les différentes catégories sémantiques.....	45
- La possession / L'appartenance "ل".....	46
- La forme / la matière "من".....	48
- La contenance "في".....	49
- L'Annexion de "reformulation spécifique".....	49
- Le spécifique.....	50
- L'élatif.....	51
B- L'annexion à un modificateur.....	54

1-L'annexion à numéral.....	54
2-Le lexique des modificateurs et quantificateurs.....	56
3-Le lexique des prépositions.....	71
C- L'annexion à un terme issu d'un déverbal.....	76
1- Le participe actif.....	76
2- Le participe passif.....	78
3- Le nom de procès.....	79
III-D'autres structures annectives.....	81
1- L'annexion à une phrase.....	81
2- L'ellipse de l'un des termes de l'annexion.....	83
3- La composition nominale en arabe.....	85
IV-Un modèle de structures annectives.....	87
1- Règles d'écriture de grammaire du syntagme nominal annexionnel.....	87
2-L'essai d'une grammaire pour le traitement morphosyntaxique du syntagme nominal annexionnel.....	89
1-L'annexion nominale.....	89
2-L'annexion à un modificateur.....	92
3-L'annexion à un terme issu d'un déverbal.....	93
3- L'essai d'une grammaire intégrant le traitement sémantique du syntagme nominal annexionnel.....	93
1-L'annexion nominale.....	94
2-L'annexion à modificateur.....	96
3-L'annexion à terme issu d'un déverbal.....	97
TROISIEME PARTIE.....	98
I-L'extraction des descripteurs complexes.....	99
A-Le choix du corpus.....	100
B-Une typologie des descripteurs extraits du corpus.....	100
1-Les descripteurs avec une structure annective à base nominale.....	101
2-Les descripteurs composé d'un quantificateur.....	104
3-Les descripteurs avec une structure annective à base déverbale.....	106
II-Le traitement du corpus à partir du modèle des structures annectives.....	110
A-Le modèle de traitement des structures annectives.....	110
B-Le traitement des descripteurs complexes selon le modèle des structures annectives.....	114
III-L'élaboration d'un lexique des méta-descripteurs.....	135
A-Les méta-descripteurs ayant une base nominale.....	136
B-Les méta-descripteurs incluant un modificateur.....	138
C-Les méta-descripteurs ayant une base verbale.....	140
CONCLUSION.....	147
BIBLIOGRAPHIE	150
ANNEXES.....	158

INTRODUCTION

INTRODUCTION

Le développement des banques de données textuelles “ (...) *est directement lié à l'essor de l'édition électronique et à la disponibilité des documents écrits sur supports électroniques lisibles par ordinateur*”¹.

Ainsi deux types d'information circulent parallèlement :

1- L'information référentielle représentée par des données bibliographiques normalisée que Dewèze appelle “référothèque”. Pour ce type de données l'indexation se fait en général manuellement et avec un vocabulaire libre ou contrôlé. L'accès à cette information se fait de façon indirect pour l'utilisateur car il doit trouver la référence de son information avant de la trouver réellement. *L'interrogation pour ce type de données se fait par des fonctions booléennes dans des langages variés et spécifiques. Cette indexation privilégie un nombre limité d'aspects du document et ne permet pas une interrogation précise et exhaustive*².

2- Le texte intégral est l'ensemble des données textuelles informatives représenté sous format numérique. Ce ne sont plus des données référentielles, l'accès à l'information est ici direct. L'indexation de ces données textuelles se fait de façon semi-automatique par diverses méthodes qui peuvent être combinées entre elles ou non, nous verrons ceci plus loin.

La différence entre une recherche dans une base textuelle et une base bibliographique vient évidemment de leur constitution. La recherche dans une base textuelle peut se faire en langage libre sur les mots du texte alors que “ (...) *les bases de données bibliographiques préconisent la recherche d'information à partir de références bibliographiques (...)*”³.

L'emploi de la langue naturelle permet à l'utilisateur de ne pas avoir à utiliser un langage formel précis pour la consultation de ces bases de données.

¹ Hanna E. NEET, *A la recherche du mot clé, analyse documentaire et indexation alphabétique*, Genève : Les Editions IES (Institut Etudes Sociales), Collection Les cours de l'IES n°2, , 1989, p. 169

² Sofiane CISSE, *Le génie linguistique : une réalité*, Processeurs, 15 mai 1992, p. 29

³ Hanna E. NEET, *op. cit.* p.169

En ce sens, le développement des bases de données textuelles a donné la priorité au développement de logiciels d'indexation automatique. *“Les programmes d'indexation utilisés dans le domaine de la documentation automatique cherchent à retrouver dans un texte les éléments de sens caractéristiques en général des termes techniques spécifiques (ils sont le plus souvent non ambigus). L'indexation de ces termes fournit une liste qui peut être considérée comme une représentation du texte bien plus petite, mais suffisamment riche pour donner une bonne idée du contenu sémantique du texte. L'index des textes est à son tour consulté par l'ordinateur lorsqu'un utilisateur recherche les textes associés à un contenu sémantique déterminé, ces contenus étant défini comme un sous-ensemble des termes techniques”*⁴.

La conception de bases de données textuelle exige la description du contenu des textes figurant dans cette base. Dans les systèmes classiques, cette description se fait à l'aide d'une liste de mots clés élaborés à partir d'une langue structurée a priori. Ces mots clés serviront lors de la recherche par une consultation de la base.

Cette consultation se fait par le biais de requêtes composées de descripteurs qui seront comparés par le système aux mots clés du lexique (fichier inverse) qui décrivent les contenus des documents.

L'indexation est la partie la plus importante de l'analyse documentaire. C'est en effet elle qui conditionne la valeur d'un système documentaire, une mauvaise indexation ou une indexation insuffisante représente 90% des causes essentielles de l'apparition de bruit ou de silence lors de la recherche [CHAUMIER, 90].

Le documentaliste traite avant tout de la sémantique. Il est clair que *l'objectif primordial de l'indexation est de faciliter la recherche de documents ou d'information contenue dans des documents*⁵, objet de recherche de l'utilisateur.

L'interrogation reste souvent booléenne, mais la difficulté liée à la langue dans sa généralité amène à résoudre certains problèmes de recherche par des palliatifs du fait du manque d'outils linguistiques. Par exemple, *l'utilisation de la troncature pour*

⁴ Max SILBERZTEIN, *Dictionnaire électronique et analyse automatique des textes, le système INTEX*, Paris : Editons Masson, 1993, p. 174

⁵ Hanna E. NEET, *A la recherche du mot clé, analyse documentaire et indexation alphabétique*, Genève : Les Editions IES (Institut Etudes Sociales), Collection Les cours de l'IES n°2, , 1989, p. 7

assimiler les différentes formes d'un même concept comme index, indexerons, ou encore l'utilisation d'opérateur de proximité ou d'adjacence etc, pour traiter la reconnaissance des mots composés⁶.

Le traitement linguistique automatique des textes et des questions, qui aboutit à une normalisation, n'est pas forcément suffisant pour permettre dans tous les cas un appariement. Il est donc souvent utile de disposer de bases de données lexicales permettant de retrouver les différentes expressions possibles pour un même concept.

Nous avons précédemment vu que pour être exploitées efficacement les bases de données textuelles nécessitaient un traitement linguistique. Ce traitement est encore plus utile pour les descripteurs complexes qui sont majoritaires dans les textes techniques⁷.

Ces descripteurs complexes peuvent être des mots composés, des locutions, des expressions lexicalisées, etc. En arabe, il existe une structure syntaxique et sémantique à partir de laquelle de nombreux descripteurs sont construits, c'est l'annexion.

On peut se demander en quoi une étude sur les descripteurs complexes incluant une structure annective peut-elle être utile à l'indexation automatique. Une grande majorité des bases de données textuelles traitent d'un domaine de spécialité, et comme nous l'avons déjà évoqué les structures annectives sont largement présentes dans ces écrits scientifiques et techniques.

Un premier inventaire des unités linguistiques a été réalisé par Wigdan ABBAS⁸. Ce travail sur la langue arabe incluait l'annexion mais était limité à l'analyse morphosyntaxique. Nous nous proposons ici de faire l'inventaire de ces structures annectives et de procéder à une double analyse morphosyntaxique et syntactico-sémantique, analyse sur laquelle nous pourrions nous appuyer pour procéder à l'indexation automatique.

Nous allons aborder dans un premier temps les méthodes d'indexation automatique et de façon détaillée celle qui nous intéresse directement : l'approche linguistique.

⁶ Sofiane CISSE, op. cit. p. 29

⁷ B. DAILLE, *Répérage et extraction de terminologie par une approche mixte statistique et linguistique*, TAL vol. 36, n° 1-2, pp. 119-132

⁸ W. ABBAS, *Définition et description des unités linguistiques intervenant dans l'indexation automatique des textes en arabe*, Thèse de doctorat, Université Lumière Lyon2, 1998

Nous étudierons ensuite les descripteurs complexes à travers les structures annectives. Ces deux premières parties nous permettront de procéder, dans une troisième partie, à l'élaboration d'un lexique des "méta-descripteurs". Ce lexique nous donnera l'occasion de voir comment s'emboîtent les descripteurs complexes à travers les structures annectives et la portée sémantique de ces structures.

PREMIERE PARTIE

Nous allons aborder dans cette partie notre contexte d'étude qui est l'indexation automatique. Nous verrons dans un premier temps les différentes méthodes utilisées dans les systèmes d'information documentaire pour l'indexation automatique. Ceci nous permettra de voir plus en détail la méthode qui nous intéresse directement : l'approche linguistique. Nous en étudierons les différentes composantes, leur représentation informatique et le rôle de cette approche linguistique dans l'indexation automatique.

L'étude de cette approche nous permettra de centrer notre travail sur le rôle du syntagme nominal (SN) dans le système d'information documentaire et d'aborder le point central de notre sujet le syntagme nominal complexe en indexation automatique.

I- Les méthodes d'indexation automatique

L'indexation automatique est l'opération qui consiste à *faire reconnaître par l'ordinateur des termes figurant dans le titre, le résumé, le texte complet (s'il est enregistré avec la notice documentaire) et parfois même l'indexation humaine, et à employer ces termes, soit tels quels soit après conversion en d'autres termes équivalents ou conceptuellement voisins, pour en faire des critères incorporés dans le fichier de recherche et utilisables pour retrouver le document*¹.

Il existe deux types d'indexation automatique ou semi-automatique : le premier consiste à enrichir automatiquement l'indexation humaine par autopostage générique² ou encore une indexation automatique non sélective (prise en compte de tous les mots non vides du document). Ce type d'indexation est utilisé de façon généralisée. Le deuxième type d'indexation automatique est l'indexation automatique sélective, c'est-à-dire une prise en compte de certains

¹ G. VAN SLYPE, *Les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires*, Paris :Les Editions d'organisation, 1987, pp.168-169

² L'autopostage générique : (...) *lors de l'initialisation d'une banque de données tous les descripteurs attribués par l'indexeur humain se verront automatiquement complétés par la machine, à l'aide de descripteurs qui leur sont attachés dans le thésaurus par une relation hiérarchique ascendante (ibidem p.170).*

termes seulement jugés par le système comme les plus représentatifs du contenu du document, soit en langage naturel soit en langage contrôlé [VAN SLYPE, 87]. Une grande majorité des systèmes bâtis sur ce type d'indexation était encore en expérimentation jusqu'à l'intégration de module linguistique il y a une dizaine d'années environ.

L'indexation automatique n'existe pas véritablement, aucun système pour le moment n'indexe de façon totalement autonome des textes numérisés, c'est pour cela que l'on parle d'indexation assistée par ordinateur. En effet, on peut dire que les systèmes actuels s'ils remplacent l'homme pour une importante part de son expertise, ils ne le remplacent pas complètement, car l'expression « indexation automatique » suppose une intervention totale du système. Ceci est loin d'être le cas et l'intervention humaine est toujours nécessaire [BLANQUET, 94]. On peut citer en exemple SINTEX et ALEXDOC comme logiciels d'indexation assistée par ordinateur.

En fait, l'indexation automatique pourrait être le complément de l'indexation humaine ou inversement. Car l'indexation automatique comporte deux phases : la première est une phase de préindexation automatique au cours de laquelle l'ordinateur analyse le texte qui lui est soumis et lui associe un certain nombre de descripteurs généralement extraits d'une liste d'autorité qu'il propose au documentaliste. La seconde phase est une phase de dialogue entre le documentaliste et l'ordinateur au cours de laquelle la liste proposée au cours de la phase précédente est affinée par voie humaine. Il semblerait que l'indexation automatique et l'indexation humaine ne soient pas si opposées [VAN SLYPE, 87].

De nombreuses méthodes (statistiques, linguistiques, systèmes experts, etc.) ont été développées pour concevoir, ou améliorer dans certains cas, les systèmes ou les logiciels d'indexation automatique. C'est ce que nous allons voir à présent.

A- la méthode statistique

L'initiateur des méthodes d'indexation automatique reste sans aucun doute H.P. Luhn avec son célèbre article *The automatic creation of literature abstracts*

paru en 1958 dans le *Journal of Research and Development* d'IBM. Il déclare : « (...) au lieu de tirer l'information au hasard comme le fait normalement le lecteur, la nouvelle méthode automatique choisit les phrases d'un article qui représentent le mieux l'information pertinente »³ [CHAUMIER, 90]. H. P. Luhn ouvrit la voie aux travaux sur l'indexation automatique par voisinage appelée aussi méthode statistique. Selon H. P. Luhn « la fréquence d'un mot dans un article fournit la mesure utile de la signification d'un mot (...), la cooccurrence relative dans une phrase de mots auxquels ont été affectés des poids de signification est une mesure utile de la signification de la phrase (...) »⁴. Plus certains mots sont souvent rencontrés en compagnies les uns des autres dans une phrase plus on peut dire que ces mots sont lourds de sens [CHAUMIER, 90].

Les méthodes statistiques, ou encore méthodes par extraction, sont basées sur deux types de traitement : le premier est basé sur le calcul de fréquence statistique (avec prise en compte des synonymes (avec dictionnaire ou sur racine)), de fréquence selon une table (table spécialisée par corpus homogène ou table selon la loi de Zipf⁵). Le deuxième type de traitement est construit sur une recherche de voisinage (que l'on appelle également méthode par co-occurrence) : avec ou sans élimination de polysémie ou avec le calcul de la distance moyenne [CHAUMIER, 90].

La méthode statistique est basée sur le mot plein (le lexique). En effet, une fois que tous les mots vides, ceux qui ne portent pas de sens en soi (mots grammaticaux, articles, etc.), sont éliminés, il ne reste que les mots pleins. On tient compte du fait que plus un mot plein est présent dans un texte et plus il est

³ J. CHAUMIER, M. DEJEAN, *L'indexation documentaire, de l'analyse conceptuelle humaine à l'analyse morphosyntaxique*, Documentaliste, vol.27, n°6, novembre-décembre 1990, p.276

⁴ ibidem, p.276

⁵ La loi de Zipf est la suivante : si l'on dresse une table de l'ensemble des mots d'un texte quelconque, classés par ordre de fréquence décroissante on constate que la fréquence d'un mot est inversement proportionnelle à son rang dans la liste. La loi est représentée de la façon suivante : $f \cdot r = c$ (f =fréquence, r =rang, et c = constante). Cette égalité, vraie en approximation, est indépendante du locuteur, du type de document et de la langue.

signifiant et servira ainsi de descripteur et pourra apparaître lors d'une interrogation [BLANQUET, 94].

B- La méthode linguistique

On constate en indexation que l'objet analysé, c'est-à-dire, le texte et les descripteurs, qui sont utilisés pour la représentation de celui-ci, font appel à la linguistique. De plus, le fait que certains systèmes d'indexation utilisent les techniques du traitement automatique des langues, démontre la pertinence d'une approche linguistique.

On a remarqué que l'absence de syntaxe lors des interrogations peut provoquer un bruit important au niveau du résultat. C'est pourquoi des méthodes basées sur une désambiguïsation syntaxique ont été élaborées. *Ces méthodes vont d'une simple marque distinctive affecté à chaque descripteur (un rôle) à une véritable structuration syntaxique de l'énoncé documentaire* ⁶. On peut citer deux systèmes qui fonctionnent sur ce modèle : PRECIS (Preserved Context Indexing System, c'est-à-dire « système d'indexation respectant le contexte ») de D. AUSTIN, adopté par la Bibliographie Nationale Britannique (BNB) et SYNTOL (Syntagmatic Organization Language), réalisé en 1960 par J.C. GARDIN et ses collègues du CNRS et ceux de la Maison des sciences de l'homme avec le soutien de l'Euratom.

Ces travaux ouvrirent la voie à divers autres études qui ont abouti à la conception de logiciels, expérimentés pour la plupart sur des banques de données de différents organismes importants. On peut citer PIAF-DOC (Programme Interactif d'Analyse du Français) pour la Documentation Française ou encore le système SINTEX (Système d'INDEXation de TEXtes), de la Société d'Information Européenne (SIE), expérimenté sur les banques de données des Communautés Européennes, ECODOC et CELEX. Les travaux de recherche réalisés à l'université ne sont pas négligeables, l'exemple du logiciel PASSAT (Programm zur Automatischen Selektion

⁶ J. MANIEZ, *Les langages documentaires : conception, construction et utilisation dans les systèmes documentaires*, Paris :Les Editions d'organisation, 1987, p.251

von Stichwartern Aus Texten) de Siemens, logiciel élaboré à partir de travaux de chercheurs de l'Université de Heidelberg à la fin des années 1970. Ce logiciel, après modification et amélioration, est toujours opérationnel et utilisé par Volkswagen pour l'élaboration de sa base de données [CHAUMIER, 90].

Ces systèmes basés sur une analyse syntaxique ne sont pas très performants lors de l'interrogation. Le taux de bruit et de silence ont été réduits par rapport aux systèmes de départ et l'analyse syntaxique a permis de nombreuses avancées. Des travaux, axés sur l'analyse de contenu et sur l'analyse de textes, ont permis de concevoir d'autres systèmes dont notamment les analyseurs syntaxiques et morphologiques qui sont apparus dans la décennie 1980. On peut citer les logiciels AlexDoc⁷ de la société GSI-Erli et SPIRIT de la société SYSTEX qui permettent une indexation sur texte intégral et sur résumé ainsi que l'interrogation des bases de données en texte intégral par le biais d'une indexation des questions. Les systèmes ALETH et DARWIN fonctionnent aussi sur le même schéma : ils cherchent la réponse par essais et erreurs successifs.

Ces systèmes, combinant différentes analyses linguistiques pour le traitement en langage naturel, sont formés de plusieurs modules de traitement linguistique.

C- La méthode mixte

Aujourd'hui de nombreux outils, logiciels ou systèmes, selon la dénomination de leurs concepteurs, ne sont pas basés sur une méthode d'indexation mais combinent plusieurs analyses en même temps. Il s'agit pour la plupart d'une combinaison linguistique (morphologie, syntaxe, lexique voire sémantique) et statistique. On peut citer quelques logiciels qui combinent ces systèmes : SPIRIT de T-GID, STAND d'IBM France. Ces systèmes permettent une indexation en langage naturel et intègrent des outils d'analyse linguistique

⁷ AlexDoc est l'une des couches d'applications du logiciels d'enregistrement et de recherche documentaire ALEXIS. AlexDoc comporte des programmes d'aide à l'indexation des documents et des questions.

multilingue [GACHOT, 95]. Ces outils sont évidemment plus performants que les autres.

D- La méthode par assignation

Le thésaurus est l'outil habituel de l'indexation, le plus souvent on y trouve trois types de relations : synonymie, hiérarchie et voisinage sémantique. Toutes ces relations sont polysémiques, elles peuvent être représentées sous forme de graphes mais elles sont différentes des réseaux sémantiques qui eux représentent les différents sens d'un même mot. Les réseaux sémantiques semblent être, sur le plan informationnel, de meilleure qualité car les relations sont définies et non polysémiques. Ils permettent une organisation multiples et multidimensionnelles, le grand avantage est que l'utilisateur final peut utiliser ces réseaux pour ses interrogations. Mais les thésaurus comme les réseaux sémantiques doivent être remis à jour fréquemment [CATTENAT, 93].

Le thésaurus est néanmoins un outil indispensable qui permet non seulement de contrôler le vocabulaire d'indexation mais aussi de gérer les problèmes de relations sémantiques entre les termes. Chaque domaine a un thésaurus spécifique, par exemple un thésaurus destiné à l'indexation et à la recherche d'image doit, en plus des relations classiques d'un thésaurus (c'est-à-dire hiérarchique), multiplier les liens d'association, d'implication et d'équivalence entre les descripteurs. On peut concevoir des liens entre les termes abstraits et les indices visuels qui les évoquent [GUILBAUD, 95]. Pour créer des relations, il faudra partir des questions des utilisateurs donc de leur vocabulaire [CATTENNAT, 93].

Depuis une dizaine d'années, on voit apparaître des modules de gestion de thésaurus associés à des logiciels de stockage et de recherche documentaire. Ces modules permettent l'enregistrement des termes et celui de leurs descripteurs (par exemple non-descripteur vers descripteur, descripteur spécifique vers descripteur générique). Ils permettent également de produire automatiquement les relations inverses (générique à spécifique). En plus de la gestion du thésaurus, ces modules permettent de faire un tri alphabétique des termes et d'en éditer une liste complète

structurée sur un niveau hiérarchique comme GOLEM ou BASIS ou sur tous les niveaux hiérarchiques MISTRAL [VAN SLYPE, 87].

On trouve également des logiciels de gestion de thésaurus multilingue comme BASIS, ou encore des logiciels qui permettent une assistance interactive pour la construction de thésaurus. C'est le cas d'ALEXIS de ERLI qui permet d'enregistrer les termes et les relations, de valider immédiatement chaque entrée et d'émettre un message quand cela est nécessaire, de mettre à jour le fichier, d'enregistrer les modifications.

Ainsi dans le cas d'une gestion automatisée d'un thésaurus, il s'agit de comparer les mots clés avec les descripteurs du thésaurus. L'indexation automatique simule ici l'indexation manuelle.

E- Les autres méthodes

1- Les systèmes experts et l'indexation automatique

Le développement de l'intelligence artificielle a permis la conception de systèmes informatiques différents des systèmes classiques, ce sont les systèmes experts. Ils sont basés sur l'exploitation dans un domaine particulier des connaissances explicites et organisées, et peuvent se substituer à un expert humain. Ces systèmes ont fait leur apparition dans les années 1970.

On peut citer COALSORT qui permet une assistance à l'indexeur tout comme JAKS qui traite en plus le texte intégral. Mais aucun de ses systèmes ne traite le langage naturel sauf RIME⁸, conçu à partir d'une méthode déterministe basée sur le traitement du langage naturel. [VANDEUR, 90]

Il est à noter que l'indexation automatique basée sur un système expert, ne peut se passer de l'intervention humaine, et on parlera plus volontiers dans ce cas d'indexation assistée par ordinateur.

⁸ RIME est spécialisé dans la gestion et l'archivage d'imagerie médicale, il est distribué par EURODIM.

2- L'indexation automatique des titres

Au départ les ordinateurs ont été utilisés en documentation pour préparer et éditer automatiquement différents types d'index, ces index servaient par la suite d'instruments manuels de recherche rétrospective ou de support de diffusion de l'information. Les index KWIC (Key Word In Context) sont une solution intéressante à la recherche et à la diffusion de l'information. Ils sont, en fait, produits directement par l'ordinateur à partir de données mémorisées et sans l'intervention d'un indexeur. Comment s'effectue dans ce cas l'indexation ? Elle se fait sur les mots du langage naturel contenus dans le texte en l'occurrence le titre, qui a été retenu et qui doit servir de base à l'édition de l'index. Le cas le plus fréquemment utilisé, pour ce type d'index, est celui des titres d'articles de périodique car plus explicite et plus précis en général que les autres types de documents, dans ce cas, on parle de Key Word In Title (KWIT), c'est-à-dire, de *mot clé du titre*. Dans ce type d'index, on exclut les mots vides de la même manière que pour la méthode statistique.

Les index KWIC et KWIT sont identiques, mais il existe une autre variante de cette méthode celle du Key Word Out of Context (KWOC) c'est-à-dire *le mot clé hors contexte*. Le principe est le même : élimination des mots vides et permutation des termes significatifs, la seule différence réside dans la présentation du mot au sein de l'index. En effet, le mot sur lequel porte la permutation n'est plus mis en évidence en colonne centrale mais placé en en-tête des titres comme des vedettes matières. La méthode KWOC est plus avantageuse dans le sens où il n'y a pas de limitation de caractères et permet une meilleure lisibilité, en revanche ces méthodes ne sont pas pratique d'utilisation.

II-L'approche linguistique de l'indexation automatique

Le documentaliste a pour activité de transcrire linguistiquement des textes qui doivent être analysé d'un point de vue documentaire. *Il transcrit linguistiquement*

*les signes d'un système sémiologique graphique précis représenté par le langage documentaire ou un thésaurus de mots clés*⁹.

*Les méthodes classiques de documentation basée sur l'indexation manuelle avec vocabulaire fermé et thésaurus font largement appel à la linguistique*¹⁰, cela vaut également pour l'indexation automatique. Car l'approche dominante concernant l'indexation postule que le sens est contenu dans le document et que pour aboutir à sa représentation, il faut analyser le plus finement possible la structure apparente du texte. Pour cela, il faut identifier et caractériser la forme que prend dans le texte, les idées de l'auteur grâce à différents outils (linguistiques et/ou statistiques) [LAINE-CRUZEL, 93].

I- Les différentes composantes de l'analyse linguistiques

L'indexation automatique basée sur des méthodes linguistiques nous semble intéressante car l'activité de l'indexeur est avant tout une activité basée sur la langue. On part, en effet, d'une information écrite ou même numérisée, on traduit ensuite cette information en concept pour pouvoir la retrouver lors d'une recherche.

Les systèmes d'indexation automatique combinent différentes analyses linguistiques. Ils se composent de plusieurs modules ayant chacun une tâche spécifique (morphologique, lexicale, syntaxique, pragmatique, ...).

En fait, ces modules se complètent les uns les autres. Il y a une interpénétration de ces analyses en reconnaissance comme c'est le cas dans les différents logiciels

⁹ R. LONG, *Linguistique et indexation*, Documentaliste-Science de l'information, vol. 17, n°3, mai-juin 1980, p.100

¹⁰ C. FLUHR, *Linguistique et documentation*, ENSSIB, Journée d'Etude : La Recherche en Sciences de l'Information (III), 23 novembre 1993, ENSSIB-Villeurbanne, p.2

d'indexation automatique¹¹. Nous avons dissocié ici ces analyses pour expliquer leur contenu respectif.

1- *L'analyse morphologique*

L'analyse morphologique est une composante de l'analyse linguistique. Celle-ci est utilisée de façon générale pour le traitement automatique d'une langue naturelle et pour l'indexation automatique en particulier. L'analyse morphologique permet d'une part de recueillir sous forme écrite l'information linguistique et d'autre part de reconnaître n'importe quelle unité lexicale ou « forme ».

L'analyse morphologique consiste à identifier des objets lexicaux (mot simple ou locution) à partir de chaîne de caractère¹² en utilisant des dictionnaires et des règles :

- *un dictionnaire comprenant les formes canoniques des unités lexicales (forme masculin singulier pour un nom ou adjectif, infinitif pour un verbe, etc.)¹³*

- *des règles morphologiques de fléchissement comme par exemple canaux ↔ canal¹⁴*

- *ces règles morphologiques peuvent être complétées par des règles dérivationnelles qui consiste à décomposer des mots inconnus du dictionnaire en éléments constituants : morphosyntaxe → morpho+syntaxe¹⁵.*

¹¹ A. BERRENDONER, *Grammaire pour un analyseur : aspect morphologique*, nouvelle édition, Grenoble Université Mendes France, cahier du CRISS n° 15, 1990

¹² A. CORET, B. MENON, D. SCHIBLER, C. TERRASSE, *Un système d'indexation structuré à l'INIST, Bilan d'une étude préalable*, Documentaliste-Science de l'information, 1994, vol. 31, n°3, p.150

¹³ *Ibidem*

Les modèles semblent être ainsi les points de liaison entre le dictionnaire et les règles de grammaire.

L'analyse morphologique ne permet pas de résoudre les ambiguïtés des syntagmes nominaux complexes, comme par exemple les mots composés. Dans ce cas, il ne s'agit pas uniquement d'étudier la structure d'un mot mais de plusieurs termes qui forment une seule unité sémantique, on a recours pour cela à l'analyse syntaxique.

2-L'analyse syntaxique

L'analyse syntaxique est également utilisée dans le traitement automatique du langage naturel. *Elle va permettre de reconnaître les relations entre les constituants d'une phrase, en retrouvant des groupes homogènes (groupes verbales, groupes prépositionnels, ...) [et] en identifiant les fonctions (circonstant, actant, ...)*¹⁶. Elle va lever les ambiguïtés de l'analyse morphologique et permettre de choisir des mots clés ou descripteurs car *l'analyse syntaxique de la phrase suppose l'identification des syntagmes et la structuration hiérarchique de ces syntagmes en liaison avec la caractérisation de leurs fonctions*¹⁷.

Dans l'analyse syntaxique, la phrase représente l'unité d'analyse elle-même formée d'unités lexicales qui sont les syntagmes. Cette structure syntaxique implique que les unités sont regroupées en constituants de différentes tailles emboîtées les unes dans les autres. Chaque phrase est analysée indépendamment des autres. On considère deux types d'analyses : une analyse des constituants et

¹⁴ Ibidem

¹⁵ Ibidem

¹⁶ ibidem, p.34

une analyse de dépendance. L'analyse des constituants est surtout basée sur la grammaire générative de Chomsky, grammaire formelle¹⁸ basée sur la notion de « catégories syntaxiques ». Celles-ci projettent des relations de dépendances syntaxiques entre des termes qui sont des constituants ou regroupement de mots à partir d'une catégorie grammaticale par exemple *groupe verbal* ou *groupe nominal*.

Les éléments, qui forment cette grammaire, sont les suivants :

- un axiome, ici ce sera P qui symbolise la phrase,
- un vocabulaire auxiliaire qui désigne les symboles des catégories grammaticales comme « groupe nominal » (GN), groupe verbal (GV), nom (N), etc.
- un vocabulaire terminal qui représente des éléments du lexique comme *voiture* ou *maison*,
- un ensemble de règles, dites « règles de réécriture ou règles de production » permettent par *une série d'étapes de « dérivation » d'engendrer des suites qui sont des phrases, en leur associant une description*¹⁹.

Prenons l'exemple d'une grammaire simple²⁰ :

P → GN+GV

GN → article+N

¹⁷ C. FUCHS et al., *Linguistique et traitements automatiques des langues*, Paris : Editions Hachette Supérieur, collection HU Linguistique, 1993, p. 110

¹⁸ Il existe plusieurs grammaires formelles (grammaire à réécriture, transformationnelle, GPSG, HPSG, grammaire d'unification, ...) qui s'appuient toutes sur l'analyse syntaxique, l'objet de notre étude n'étant pas d'en faire une synthèse nous renvoyons pour cela aux ouvrages de G. SABAH, A. ABEILLE et C. FUCHS qui en ont fait un bon résumé.

¹⁹ Magdalena PRADILLA, *Recherche de descripteurs en indexation automatique des documents*, Thèse de Doctorat mathématique appliquées aux sciences sociales, 1982, Les cahiers du CRISS n°4, p.38

²⁰ Ibidem, p. 38

GV → V+GN

V → vend, prend

N → marin, chat, jouet

article → le

La flèche signifie que la partie gauche se réécrit sous la forme de partie droite. Le symbole (+) désigne l'opération de concaténation, c'est-à-dire en prenant notre exemple que P se réécrit sous la forme d'un GN suivi d'un GV, tandis que la virgule marque un choix.

Cette grammaire simple permet d'engendrer un nombre important de phrases :

« le marin vend le chat »

« le marin vend le jouet »

« le marin vend le marin »

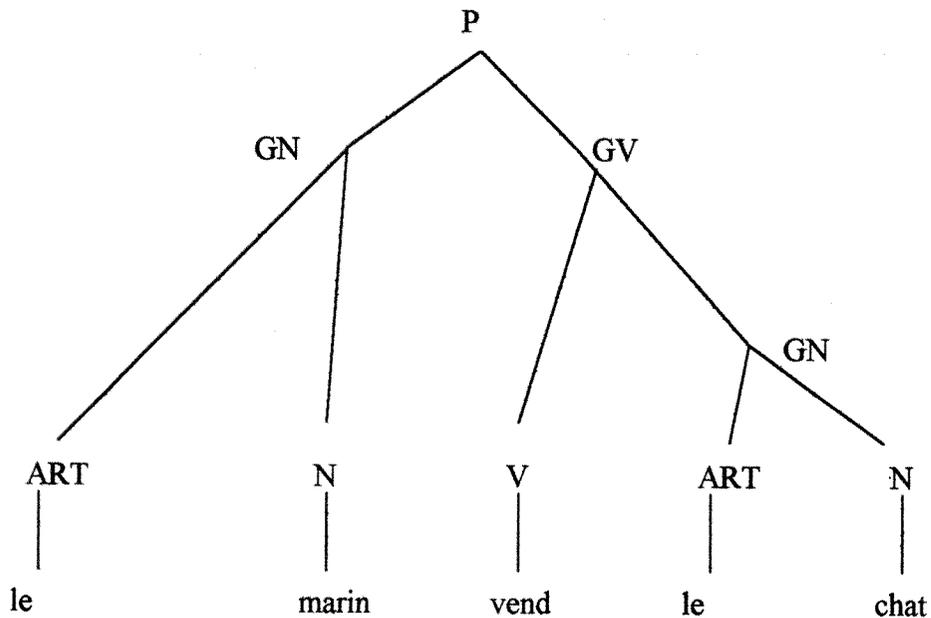
« le marin prend le chat »

etc.

Une telle grammaire engendre évidemment des phrases non-grammaticales comme par exemple *le marin court le chat**, si on introduit courir dans la liste d'éléments des verbes. On peut également introduire des règles dites contextuelles, ces règles vont spécifier si un verbe est intransitif, c'est-à-dire qui ne peut pas avoir de compléments directs (ceci va restreindre les phrases non grammaticales). Elle peut également engendrer des phrases sémantiquement déviantes comme *le jouet vend le marin*.

On peut représenter ce modèle (*le marin vend le chat*) sous forme d'arbre appelé arbre syntaxique, *l'information syntaxique est attachée aux noeuds et aux branches*²¹ :

²¹ C. FUCHS et al., *Linguistique et traitements automatiques des langues*, Paris : Editions Hachette Supérieur, collection HU Linguistique, 1993, p. 106



Un modèle linguistique qui doit être construit a priori est difficile à contrôler car on ne tient pas compte de toutes les possibilités. La mise en œuvre de ce modèle demande également un nombre de règles de grammaire limité pour bien maîtriser le processus, mais pour cela il faut :

- utiliser une grammaire qui ne compte qu'un nombre réduit de catégories ou de fonctions,

- travailler sur des textes similaires du point de vue de leur forme syntaxique, pour cela une étude statistique permettrait de dresser une typologie des règles et des structures fréquentes parmi celles qui sont cohérentes.

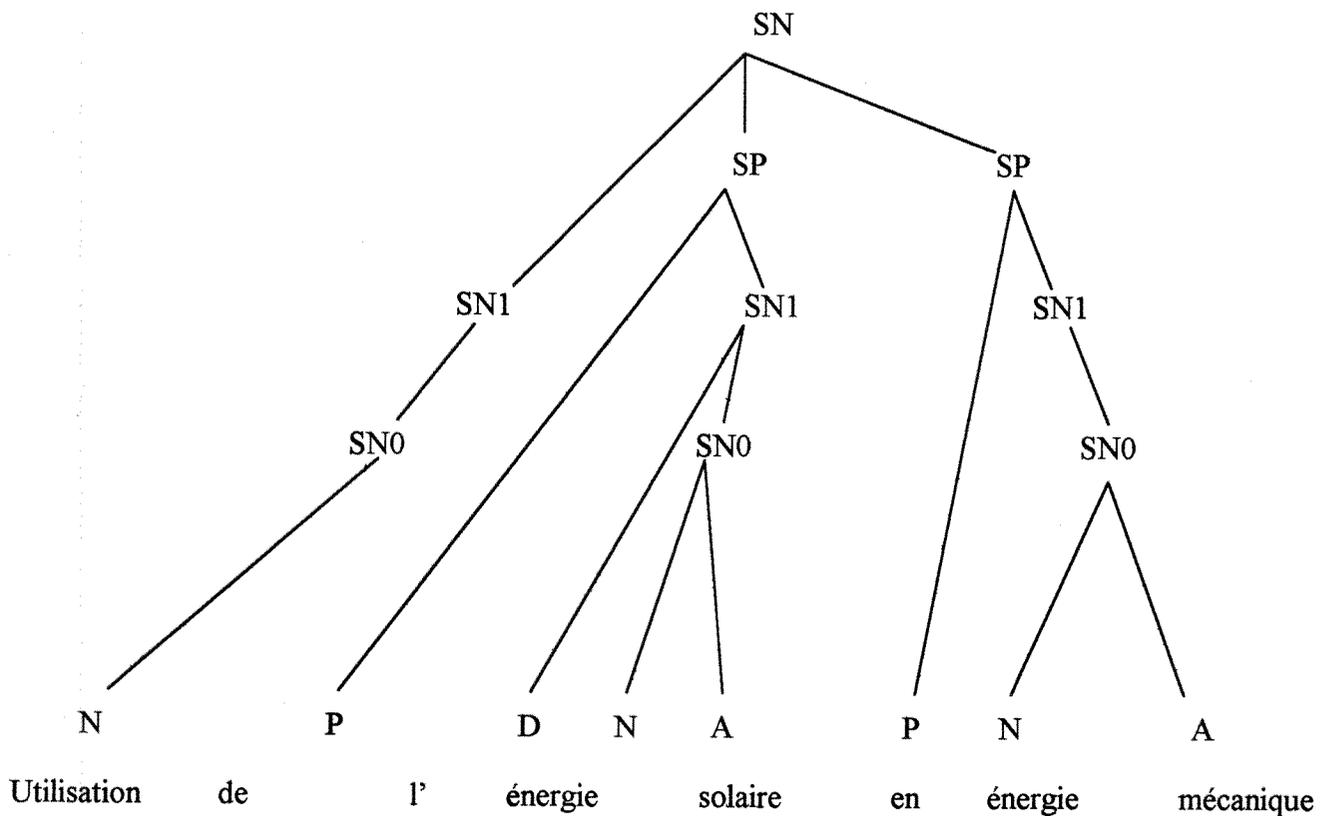
L'analyse syntaxique est ainsi le modèle qui définit la bonne cohérence des relations entre les mots d'une phrase. Elle peut permettre de sélectionner la bonne catégorie morphologique d'un mot ambigu. L'analyse syntaxique combinée à l'analyse morphologique permet de constater si deux termes successifs forment un constituant grammatical. Comme le souligne A. Dewèze *l'analyseur morphosyntaxique exploite non seulement le repérage des catégories mais aussi le genre et le nombre, etc, pour effectuer une analyse morphologique plus*

complète des mots apparaissant dans le texte²². L'analyse syntaxique permet ensuite d'effectuer la délimitation du rôle respectif des mots dans la phrase qui peut ainsi être découpée en syntagme.

L'analyse syntaxique et le choix des mots clés

Magdalena PRADILLA donne un exemple d'utilisation de l'analyse syntaxique pour extraire des syntagmes nominaux : *on part de l'hypothèse que les liens sémantiques (ou de « sens ») sont liés aux liens syntaxiques et par conséquent une analyse syntaxique non ambiguë produit des syntagmes nominaux ou groupes nominaux de niveau inférieur qui pourront être des descripteurs*²³.

L'analyse syntaxique pourrait être la suivante :



²² A. DEWEZE, *Informatique documentaire*, Edition MASSON, Collection Méthode+Programmes, Paris, 3^e édition, 1989, p.134

²³ Magdalena PRADILLA, *Recherche de descripteurs en indexation automatique des documents*, Thèse de Doctorat mathématiques appliquées aux sciences sociales, 1982, Les cahiers du CRISS n°4, p.48

Ici on dégage les syntagmes nominaux (SN₀) d'ordre inférieur, on a les candidats descripteurs suivants : *énergie solaire* et *énergie mécanique*.

Nous venons de voir les deux analyses linguistiques les plus importantes utilisées en indexation automatique. Pour perfectionner certains logiciels documentaires, il arrive qu'on intègre deux autres modules d'analyse linguistique : l'analyse lexicale et l'analyse pragmatique.

3- L'analyse sémantique ou lexicale (cf Annexe 1)

“La sémantique lexicale est très souvent assimilée à la sémantique des “mots pleins” (...) c'est-à-dire des mots qui relèvent des grandes catégories comme le substantif, le verbe ou l'adjectif (...) par opposition aux mots vides²⁴”.

En quoi consiste l'analyse sémantique ? Catherine Fuchs souligne que *les approches de la sémantique effectuées en vue de traitements automatiques ont eu jusqu'ici pour objectif essentiel la recherche de formalismes de représentation (...) issus de la logique et des réseaux sémantiques²⁵*. On va étudier tous les sens possible d'un terme, et ces sens seront représentés sous forme de réseau.

Les réseaux sémantiques trouvent leur origine en psychologie “ (...) il s'agissait de rendre compte de la façon dont les sujets catégorisent et mémorisent (l'information)²⁶”. Il s'agissait donc d'inscrire l'unité lexicale dans une structure sémantique plus vaste de telle sorte que le sens de l'unité résulte de la place qu'elle occupe dans la structure et des relations qu'elle entretient avec les autres unités de cette structure. Les réseaux sémantiques sont représentés schématiquement par des graphes formés de noeuds qui symbolisent les unités

²⁴ ²⁴ C. FUCHS et al., *Linguistique et traitements automatiques des langues*, Paris : Editions Hachette Supérieur, collection HU Linguistique, 1993, p. 140

²⁵ C. FUCHS ,op. cit. , p. 162

lexicales (porteuse d'information), ces graphes sont reliés par des arcs qui représentent les relations entre unités.

4- L'analyse pragmatique

Ce niveau d'analyse fait référence à l'aspect cognitif humain : en effet, l'être humain fait appel à ses connaissances du monde et du contexte pour résoudre les cas de polysémies. Dans le cas de l'indexation automatique, on fait appel aux réseaux sémantiques pour délimiter l'univers de connaissances à un domaine très spécialisé [CLAVEL, 93].

La différence entre l'analyse sémantique et l'analyse pragmatique c'est que cette dernière fait intervenir la psycholinguistique et notamment les compétences de l'homme à appréhender le monde qui l'entoure.

B- Représentation informatique des modules linguistiques dans un logiciel documentaire (Cf annexe 2)

Ces modules linguistiques se traduisent en informatique par :

- des dictionnaires morphologiques : incluant les flexions, dérivation des mots,
- des analyseurs syntaxiques : repérage des différentes catégories grammaticales
- des dictionnaires de mots vides : incluant les mots outils grammaticaux et fonctionnels,
- module d'analyse statistique affectant à chaque concept une fonction de poids informationnel et classant par ordre de pertinence les réponses.

Ces différents niveaux d'analyse et ces différents dictionnaires se complètent et sont imbriqués les uns dans les autres. La composante des modules linguistiques varie d'un logiciel documentaire à un autre, par exemple SPIRIT

²⁶ ibidem, p. 143

n'inclue pas de thésaurus. Il est surtout spécialisé pour l'indexation et la recherche dans de grosses bases alors que AlexDoc de GSI-ERLI se base surtout sur les réseaux sémantiques et donc la délimitation d'un domaine de spécialité pour effectuer l'indexation et la recherche.

C- Rôle de l'approche linguistique en indexation (cf Annexe 3)

L'indexation automatique suppose de travailler sur de l'information textuelle, ainsi les requêtes de l'utilisateur et l'information contenue dans le document à chercher sont de même nature, celle du langage libre.

Ceci explique que l'interrogation des bases de données textuelles, au moyen d'une requête en langage naturel, met en évidence le fait que l'on exprime mieux au moyen de ce langage, l'information que l'on veut rechercher. Ceci n'est évidemment possible que si cette information est en langage naturel, c'est-à-dire, présente dans le contenu des documents numérisés. Mais cela implique également pour l'utilisateur la capacité d'exprimer dans sa langue et de manière précise, son besoin et qu'il ait une idée du contenu de la base.

Pour pouvoir localiser l'information dans des masses importantes de textes et le plus finement possible, il est nécessaire de disposer d'outils linguistiques. Le développement des logiciels d'indexation automatique intégrant des outils linguistiques pour le traitement du langage naturel s'est considérablement étendu depuis de nombreuses années, on peut citer SPIRIT de la société SYSTEX et AlexDoc de la société GSI-ERLI, spécialisée dans l'ingénierie linguistique.

On voit bien que la linguistique intervient dans tous les niveaux de la documentation automatique. Elle joue, en particulier, un rôle essentiel dans le

problème de fonds de la recherche documentaire qui est de réduire le bruit (documents non pertinents) et le silence (documents manquants).

L'introduction d'outils utilisant abondamment la linguistique transforme le rôle de l'être humain dans la chaîne de recherche d'information. *Le documentaliste perd peu à peu son action d'intermédiaire visible entre l'utilisateur final et le système et devient de plus en plus un gestionnaire du système pour l'améliorer en fonction d'une observation du dialogue direct entre utilisateur et système*²⁷ [FLUHR, 93].

Des recherches ont été menées sur cet aspect linguistique de l'indexation automatique. On s'est aperçu que d'un point de vue documentaire le syntagme nominal, qui est la plus petite unité linguistique porteuse de sens, représentait l'information.

Nous allons voir dans la partie qui va suivre l'importance du syntagme nominal dans un système d'information documentaire et l'utilité d'une recherche en analyse du contenu l'intégrant comme élément important.

III- L'importance du syntagme nominal pour l'indexation automatique

A-Le Syntagme nominal et le Système d'Information Documentaire

C. Fuchs²⁸ définit le syntagme nominal comme *étant un groupe d'unités dominé par une tête dont la catégorie donne son nom au syntagme (syntagme*

²⁷ C. FLUHR, *Linguistique et documentation*, ENSSIB, Journée d'Etude : La Recherche en Sciences de l'Information (III), 23 novembre 1993, ENSSIB-Villeurbanne, p.3

²⁸ C. FUCHS et al., *Linguistique et traitements automatiques des langues*, Paris : Editions Hachette Supérieur, collection HU Linguistique, 1993, p. 162

nominal, syntagme verbal, syntagme adjectival, ...). Il joue un rôle fonctionnel donné²⁹. Comme nous l'avons dit précédemment, le fait que le syntagme nominal soit la plus petite unité porteuse de sens, il peut être un élément de choix pour les descripteurs. Le groupe SYDO s'est intéressé à cet aspect.

1-Le modèle SYDO

Le groupe SYDO³⁰ était formé d'une équipe pluridisciplinaire comprenant des informaticiens, des documentalistes, des linguistes, et des psychologues. L'équipe s'est attachée à élaborer des procédures d'analyse morphosyntaxique les plus complètes possibles et adaptées à des applications documentaires.

L'objectif du groupe SYDO était d'élaborer un système documentaire complet pour petites et moyennes bases de données spécialisées. Le système intégrait plusieurs modules : l'indexation, la description des documents, la création d'un thésaurus, l'aide à la consultation, etc. Afin de valider ses recherches sur l'analyse automatique de documents écrits, le groupe SYDO a tenté d'adapter son analyseur sur des corpus particuliers composés principalement de questionnaires d'enquêtes et d'études biologiques de l'entreprise Synthélabo³¹. Il s'agissait pour cette entreprise de réaliser une veille scientifique dans le domaine en question.

L'expérimentation avait été menée sur un corpus regroupant des comptes rendus médicaux des médecins d'un service de cardiologie. Le travail du groupe SYDO était d'identifier les mots isolés pouvant être pris comme descripteurs. Le

²⁹ *ibidem* p.110

³⁰ l'équipe SYDO était constitué de quatre centres de recherche français et de deux universités européennes :

- le laboratoire d'informatique documentaire de Lyon1
- le centre de recherche en sémiologie de l'université Lumière Lyon2
- le centre d'études et de recherches de l'ENSSIB
- l'équipe de recherche sur les systèmes d'information et communication des organisations Lyon3
- l'Université de Fribourg (Suisse) et l'Université de Mayence (Allemagne).

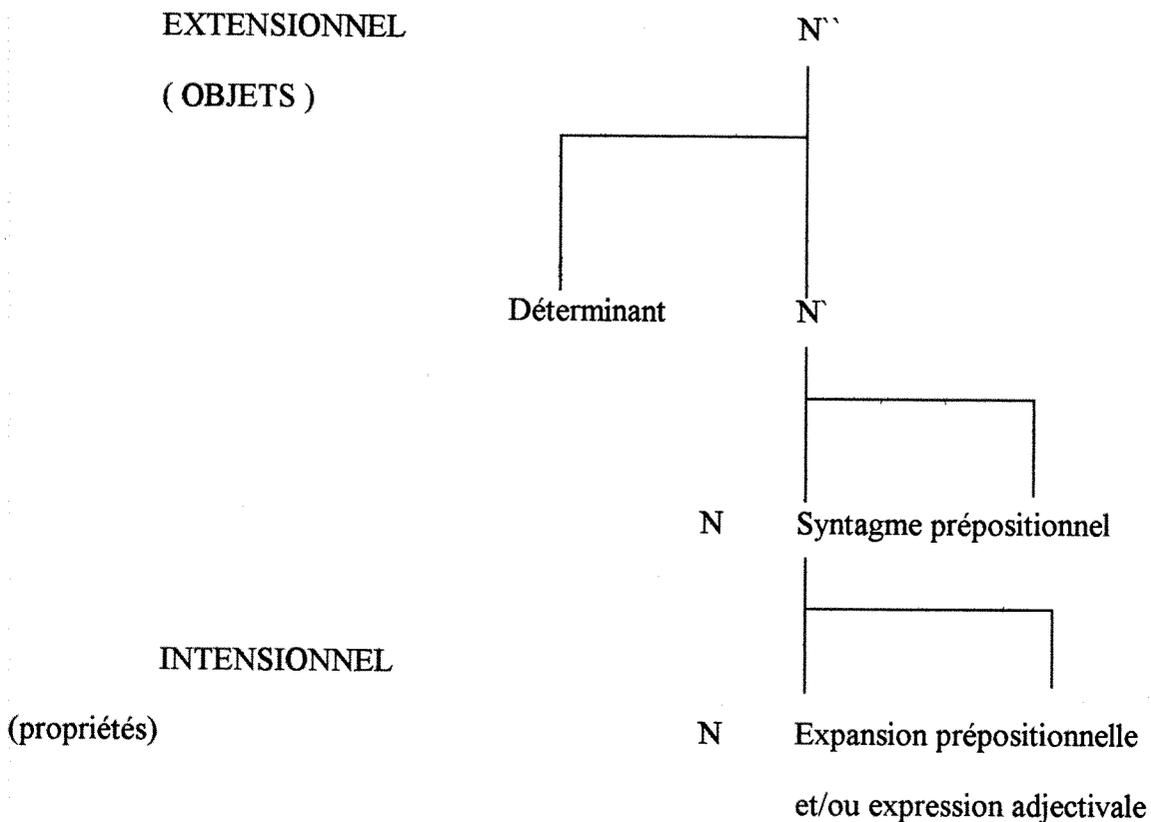
³¹ Richard BOUCHE, *Le syntagme nominal, une nouvelle approche des BDD textuelles*, META, XXXIV, 3, 1989, p.429-433

résultat n'a pas été concluant pour ces mots isolés mais plutôt pour les expressions beaucoup plus complexes construites sur les formes suivantes : nom + adjectif, nom + préposition + nom, etc. Ce travail était donc une réflexion sur le statut linguistique du mot clé. [BOUCHÉ, 89]

Quelle était donc la méthodologie développée par le groupe SYDO pour ce traitement linguistique ?

Le groupe SYDO a en effet développé une méthodologie toute particulière pour le traitement linguistique du descripteur. C'est une méthodologie de traitement automatique des textes écrits en langage naturel, qui est basée sur un modèle linguistique et utilisant une grammaire affixe. [BOUCHÉ, 91].

La grammaire de reconnaissance du syntagme nominal développé par le groupe SYDO est la suivante [BOUCHÉ, 91] :



Cette grammaire s'articule sur trois niveaux : N, N', N''

N : niveau purement intensionnel, les unités considérées sont uniquement des prédicats simples ou complexes.

N' : niveau de basculement de l'intensionnel à l'extensionnel, il définit une classe d'objet de la réalité extra-linguistique. N' est toujours un prédicat.

N'' : opération de fermeture au moyen d'un quantificateur. On a, à ce niveau, une référence qui correspond à un objet de la réalité extra-linguistique.

Les outils informatiques réalisés par l'équipe SYDO : en effet, l'équipe avait mis au point deux modules, celui de l'analyse morphosyntaxique et celui d'extraction des générateurs. Nous nous intéresserons seulement à l'analyse morphosyntaxique. Quels sont les buts de ce type d'analyse ? Tout d'abord, deux principes ont guidé la mise au point des règles d'analyse : le premier est d'éliminer les sources d'ambiguïté, c'est-à-dire, d'éviter de créer à une étape d'analyse, un « paramètre » qui serait ambigu à une étape postérieure. Le second principe consiste à réduire le nombre de règles d'analyse.

L'objectif du modèle conçu par le groupe SYDO est d'une part d'identifier les syntagmes et d'autre part de déterminer leur structure en mettant en évidence les relations entre leurs constituants. Cette détermination de la structure du syntagme nominal va faciliter la recherche d'information. Elle permettra ainsi de stocker une représentation du syntagme nominal.

Les linguistes du groupe insistent sur le fait qu'un mot, pris isolément du contexte de discours dans lequel il s'inscrit, n'a pas de sens en soi. Ce sont des prédicats libres par opposition au syntagme nominal qui lui relève du discours [LE GUERN].

Le groupe SYDO a surtout travaillé sur le français, une autre équipe lyonnaise (SAMIA) s'est aussi intéressée à l'analyse morphosyntaxique mais en l'appliquant à l'arabe.

2- L'équipe SAMIA (Synthèse et Analyse Morphosyntaxiques Informatisées de l'Arabe)

Le programme de recherche SAMIA relève de deux centres de recherches universitaires lyonnais :

- le Centre de Recherche en Traduction et Terminologie (CRTT), de l'Université Lumière Lyon 2, (J. Dichy, X. Lelubre)

- le Centre d'Etudes et de Recherche en Sciences de l'Information (CERSI) de l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB) (M. O. Hassoun)

Les recherches au sein du groupe SAMIA portent sur la conception de modèle de simulation de l'activité langagière grâce à l'avancée des recherches linguistiques en morphosyntaxe, ceci dans le but de l'élargir au niveau syntactico-sémantique.

Aujourd'hui, trois générations du modèle SAMIA ont vu le jour :

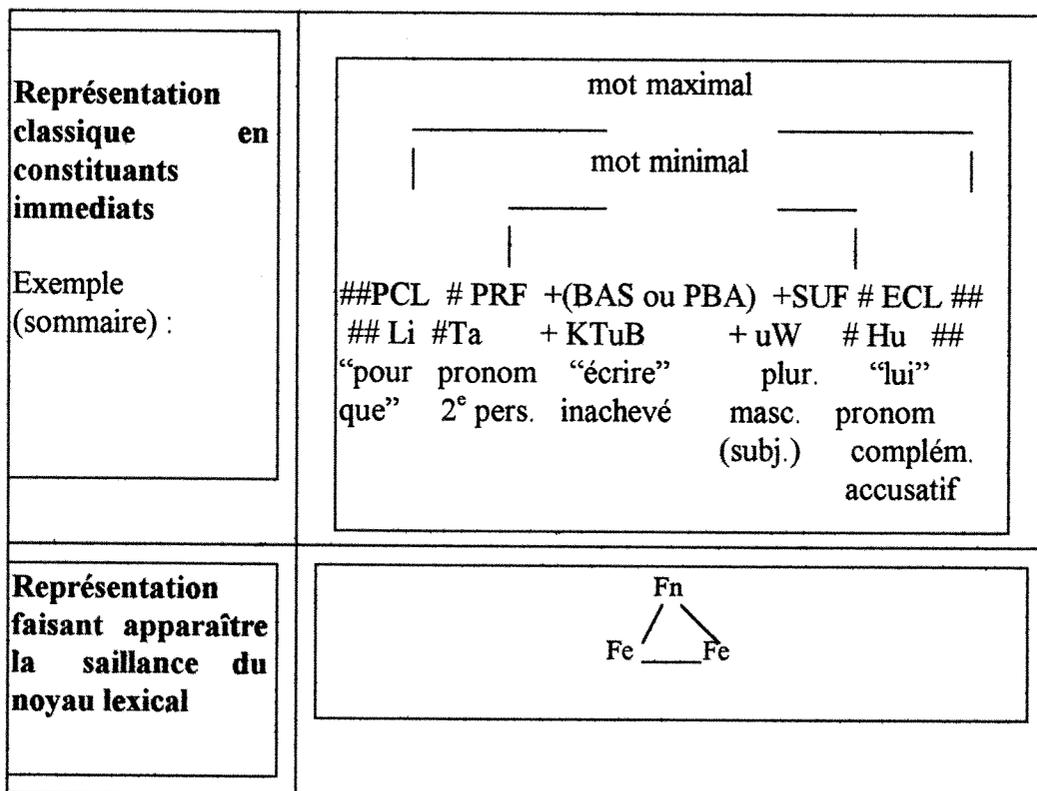
- le modèle SAMIA 1^{ère} génération

Il consistait à traiter de la génération et de l'analyse du mot graphique³² tout en prenant en compte des transformations subies par les racines anomales (ayant /w/, /y/, ou /ʔ/ comme radicales). La base de données lexicale correspondant à ce modèle n'avait pas encore été réalisée. Deux modèles de la morphologie arabe ont été conçus l'un en synthèse et l'autre en analyse, ainsi qu'un dictionnaire informatisé.

³² *Le programme porte sur la morphologie de l'arabe. Le domaine de la morphologie est le mot. Par convention et aussi parce que ce traitement automatique porte actuellement sur de la langue écrite, nous entendons ici par « mot », le mot graphique c'est à dire très classiquement ce qui est situé entre deux blancs (...), in Simulation de modèles linguistiques et EAO de l'arabe, J. DICHY et M. HASSOUN, Travaux SAMIA1, CILF, 1989, p.45*

- Le modèle en synthèse morphologique *consiste en l'occurrence à engendrer automatiquement à partir de structures initiales des mots graphiques et ce avec consultation de données enregistrées dans un dictionnaire*³³.

- Le modèle en analyse : on transcrit en écriture arabe les consonnes et les segments d'allongement (/w/, /y/, et /:/). Les voyelles brèves qui servent pour la vocalisation ne paraissent que très rarement dans les revues, les journaux et les livres, le plus important étant la marque de la gémination des consonnes (doublement des consonnes). Il faut cependant préciser que le programme de recherche SAMIA a porté sur des formes non vocalisées. Dans cette perspective, on a repéré dans le mot graphique la présence de proclitiques, préfixes, base, suffixes et enclitiques afin de mettre en évidence une base de représentation graphique non vocalisée. Le schéma suivant donne une représentation, certes simplifiée du mot graphique, c'est une représentation vectorielle du mot (J.P. Desclés) :



Le schéma du mot³⁴

³³ Ibidem, p.66

Le noyau lexical du mot graphique (ou *formant-noyau*, *F_n*) est représenté par les bases et pro-bases, les autres constituants étant des extensions (ou *formants-extensions*, *F_e*). Le schéma ci-dessus représente le mot, avec des frontières fortes (+) et faibles (#) de morphème.

Le dictionnaire quant à lui représente la base de données de l'analyseur et synthétiseur.

Le modèle SAMIA 2^{ème} génération : DIINAR.1

Le logiciel de synthèse et d'analyse des formes verbales de l'arabe, fait appel à la base de données lexicale de l'équipe SAMIA. Cette base de données regroupe les travaux de l'équipe ainsi que des données linguistiques modélisées concernant la conjugaison de l'arabe réalisé par S. Ammar et J. Dichy. On peut également citer les travaux de N. ABU ALCHAY³⁵ sur la conception d'un modèle adapté au traitement automatique des verbes arabes dans le cadre de l'EAO au sein du programme SAMIA pour l'apprentissage de la conjugaison.

La première version de la base de données lexicale, DIINAR.1³⁶, a été conçue et réalisée dans le prolongement du programme de recherche SAMIA. Limitée au traitement au niveau du mot, elle inclue les spécificateurs morphosyntaxiques pour gérer des relations contextuelles entre la base du mot (nom, verbe, mot outil) et les morphèmes grammaticaux entourant la base (prépositions, les conjonctions, l'article, les marques casuelles, etc). Cette base de données est le fruit d'une coopération entre les deux partenaires lyonnais (CRTT, ENSSIB) et le partenaire

³⁴ Joseph DICHY, *Pour une lexicomatique de l'arabe : l'unité lexicale simple et l'inventaire fini des spécificateurs du domaine du mot*, in *Lexicomatique et Dictionnaire*, Actes des Journées de Lyon, collection "Actualités scientifiques" (AUPELF-UREF, réseau "Lexicologie, Terminologie et Traduction"), Québec : DPLU et Paris : Hachette, juin 1996, p.6

³⁵ N. ABU ALCHAY, *Un système expert pour l'analyse et la synthèse des verbes arabes dans un cadre d'enseignement assisté par ordinateur*, Thèse de doctorat Université Claude Bernard Lyon1, 1988

³⁶ Dictionnaire INformatisé de l'ARabe

tunisien, l'Institut de Recherche en Sciences de l'Informatique et des Télécommunications (IRSIT)³⁷.

Le modèle SAMIA 3^{ème} génération : DIINAR.2 (en cours)

La base de données DIINAR.1 est une base au niveau du mot graphique et l'extension de celle-ci à des niveaux supérieures est en cours. La méthodologie du modèle linguistique SAMIA a ainsi permis la réalisation de cette base.

Une version de DIINAR étendue aux spécificateurs du niveau de la phrase (DIINAR.2) doit être réalisée en collaboration avec l'Université de Nimègue (Institut de Langue et Culture du Moyen-Orient, TCMO, Pays-Bas)³⁸.

Les applications des travaux de recherches SAMIA :

- l'analyseur et le synthétiseur permettent des développements en Enseignement Assisté par Ordinateur (EAO) et en Traduction Assistée par Ordinateur (TAO) de l'arabe.

Antérieurement aux travaux SAMIA, M. Hassoun a essayé dans sa thèse de doctorat³⁹ d'adapter le modèle d'analyse morphologique du français (AMEDE) du groupe SYDO à la langue arabe. Cette adaptation a démontré que le système conçu pour les spécificités de la langue française pouvait être appliqué à la langue arabe après traitements car les traits morphologiques de l'arabe sont différents de ceux du français. Il a, en effet, fallu déterminer des traits morphologiques propres à la langue arabe. C'est à ce niveau que M. Hassoun a déterminé en 1982 un modèle linguistique pour l'analyse morphologique.

³⁷ Ibidem, p.1

³⁸ Joseph DICHY, op. cit. p.1

³⁹ M. HASSOUN *Système d'analyse morphologique automatique de la langue arabe*, Thèse de doctorat en mathématique-informatique, 3^e cycle ; 1982, 153p.

Ces travaux ont permis de déterminer les outils nécessaires au développement d'un ensemble de traits morphologiques de la langue arabe : modèle linguistique, dictionnaire, processus d'analyse et de synthèse permettant le développement en EAO et TAO.

En plus des développements en EAO et TAO, les travaux du programme SAMIA prévoit dans le cadre de DIINAR.2 la réalisation d'un premier système d'indexation automatique. Celui-ci s'inspirera des travaux sur la langue française du groupe SYDO ainsi que de l'analyseur morphologique de l'arabe élaboré par l'équipe SAMIA. Après une étude complète de la morphologie arabe et le développement d'un analyseur morphosyntaxique, il serait intéressant de dépasser le cadre du mot simple et de faire l'inventaire des autres problèmes du syntagme nominal arabe pour résoudre les ambiguïtés des descripteurs complexes dans le cadre d'une indexation automatique.

B- L'intérêt du syntagme nominal complexe en indexation automatique

Il est très rare de désigner une notion ou un concept d'un domaine spécialisé par des termes simples. La majorité de ces termes sont en fait des termes complexes c'est-à-dire formés de plus d'une unité lexicale.

Comment espérer traiter automatiquement des termes si leur construction, leur association, leur formation n'est pas identifiée ?

Une étude statistique sur corpus pour l'extraction des descripteurs est certes une méthode en indexation automatique mais les résultats ne sont pas toujours concluants, en témoigne le taux de bruit ou de silence lors de la recherche. Cela d'autant plus que les textes indexés sont en langage naturel de même que les requêtes de l'utilisateur. Depuis l'introduction d'outils linguistiques dans les logiciels d'indexation automatique, le pourcentage de bruit et de silence a énormément diminué. On ne va plus calculer le nombre d'occurrence des termes présent dans la

requête et dans les documents qui comportent ces termes mais on va procéder à l'analyse linguistique des termes (morphologique, syntaxique, sémantique, etc...) présent dans la requête et éviter ainsi les synonymies ou polysémies possible lors de la recherche. Cette analyse linguistique est complétée par un calcul statistique des réponses qui va permettre un classement par ordre de pertinence.

Le problème qui se pose en documentation est que l'on traite toujours d'un domaine de spécialité. Comme nous l'avons précédemment évoqué, la majorité des notions ou concepts sont décrits par des termes complexes : mot composé, expression lexicalisée, etc.

Benoît HABERT rappelle que *reconnaître les séquences appelées « noms composés » dans les textes généraux, les termes dans les textes spécialisés, constituent un objectif fondamental pour tous les travaux documentaires. (...) En dehors de l'informatique documentaire, un traitement adéquat de ces séquences s'avère tout aussi crucial pour l'analyse automatique dans son ensemble.*⁴⁰

L'inventaire de ces structures est donc important, en effet, *une partie essentielle de la néologie lexicale en particulier dans le domaine technique et scientifique s'opère par le biais de séquences complexes*⁴¹. Face à l'accroissement de la masse de textes à traiter, les professionnels de l'information se dotent de plus en plus d'outils permettant d'extraire les syntagmes nominaux qui peuvent servir de descripteurs pour le domaine en question.

Qu'en est-il pour la langue arabe ? Pour rendre compte de la construction du SN arabe, il faudrait rappeler sa place dans l'organisation de la langue.

La langue arabe possède un système dérivationnel très riche, les éléments et les opérations qui interviennent dans la formation des mots sont :

⁴⁰ Benoît HABERT, Christian JACQUEMIN, *Noms composés, termes, dénominations complexes : problématiques linguistiques et traitements automatiques*, TAL, volume 34, n°2, 1993, p.6

⁴¹ Ibidem, p.6

- les racines : le système phonologique de l'arabe établit une disjonction entre les deux ensembles formés chacun par les consonnes et les voyelles. Il en découle que les racines du système de nomination sont consonantiques. La racine arabe ne peut être constituée que de deux ou trois consonnes. Une racine réfère en principe à une même notion. Lorsqu'on ajoute à une racine des voyelles, elle forme un modèle que l'on appelle schème et qui peut être nominal ou verbal.

- les schèmes désignent donc des combinaisons de consonnes et de voyelles auxquelles peut s'ajouter la génération d'une consonne radicale ou par extension quatre. Celle-ci appliquée à la racine permet d'obtenir différentes dérivations. Ces schèmes correspondent à des modèles à partir desquels on peut former des mots avec des racines. Les grammairiens arabes pour représenter l'ensemble des schèmes ont choisi la racine /f^c-l-/.

Il y a deux types de dérivation au niveau du mot :

- une dérivation interne qui part du formant noyau (RAC+SCH) composé de racine et schème,

- une dérivation externe pour les termes qui sont des unités lexicales, ces unités lexicales sont composées d'un formant noyau et d'un formant extension lexicalisé, on a ainsi le schéma suivant : UL= <Fn+Fel >

par exemple : pour la racine /s-y-r/ sur le schème /fa^{cc}âl/

le noyau ou Fn= <sayyar > et UL= <sayyart+a&> soit sayyara&

Il faut noter que tous les verbes et une partie des noms relèvent de ce système (RAC, SCH), une partie des noms n'en relève pas, ce sont en fait les emprunts qui fonctionnent comme un calque linguistique comme *elektronî* pour électronique.

Selon M. LE Guern *le descripteur n'est pas un mot de la langue en tant qu'élément du lexique mais ce mot actualisé. Cette actualisation se fait par la construction d'un syntagme. Le descripteur n'est donc pas mot de la langue mais*

*syntagme du discours ou plus exactement syntagme type par rapport aux occurrences des énoncés*⁴². En arabe, comme dans toute langue, les mots sont en rapport syntagmatique et parce que le syntagme nominal se fait dans le discours, il ne peut se concevoir sans syntaxe. M. Le Guern signale à ce sujet que la *conception linguistique du descripteur entraîne une conséquence nécessaire : le traitement linguistique inhérent à un système documentaire ne peut se concevoir sans une syntaxe, syntaxe dont la face signifiante rend compte de l'agencement des morphèmes et dont la face signifiée soit l'organisation*⁴³.

André Roman, définit la syntaxe d'une langue comme étant *l'organisation des unités de nomination de cette langue en système de relations qui structurent en fait une certaine unité de communication : la phrase*⁴⁴. L'arabe se sert des deux systèmes : le système de communication et le système de nomination pour former les syntagmes nominaux ainsi que pour distribuer des étiquettes à des classes d'objets.

En fait l'opération de nomination est la mise en relation de deux entités, l'une linguistique, l'autre extra-linguistique : *l'homme a eu recours à une combinatoire linguistique. Cette combinatoire devait produire ses langues comme autant de « système de systèmes », chaque système est composé chacun de quatre systèmes interdépendants : un système de phonème, un système de syllabe, un système de nomination, un système de communication*⁴⁵.

La recherche des descripteurs d'un texte ou d'un ensemble de textes constitue la première étape de l'indexation automatique. *La fonction des descripteurs ne*

⁴² M. LE GUERN, *Un analyseur morphosyntaxique pour l'indexation automatique*, in revue *Le Français moderne*, n° 1 juin 1991, p. 26

⁴³ M. Le Guern, *Les descripteurs d'un système d'information documentaire : essai de définition*, in acte du Colloque *Traitement automatique des langues naturelles de systèmes documentaires*, Condenser, suppl 1 ; Université de Clermont Ferrand, 1984, p.163

⁴⁴ A. Roman, *les relations syntaxiques*, in *cercle linguistique d'Aix en Provence, travaux Aix en Provence*, 1985, p.95

⁴⁵ *Ibidem*, p.95

*fait pas problème, ce sont des éléments du corpus appelés à être intégrés au thésaurus de telle manière qu'ils permettent de retrouver les documents pertinents, (...). (...) La finalité du descripteur exclut qu'on puisse l'envisager en faisant abstraction de la valeur référentielle de ses occurrences dans le corpus*⁴⁶.

Ainsi le descripteur est toujours en relation avec un objet extra-linguistique ce qui le différencie du simple mot de la langue.

Cette définition du descripteur vaut aussi bien pour le descripteur simple que pour le descripteur complexe. Les descripteurs constitués de plusieurs unités lexicales ont très souvent une sémantique propre reflétant celle du texte d'origine. *Analyser le fonctionnement sémantique du descripteur implique qu'on tienne compte de la relation qui le lie à l'objet extra-linguistique désigné par les occurrences du terme descripteur dans le corpus*⁴⁷. Ainsi pour définir un descripteur complexe, nous pouvons nous appuyer sur le modèle de l'opérateur étoile ou la base serait "l'opérande" et les expansions seraient des "opérateurs étoiles". Prenons comme exemple le prédicat complexe "réseau numérique" "réseau" est un prédicat ou opérande (et) "numérique" étant un prédicat opérateur ou opérateur étoile, permettant la construction du prédicat complexe⁴⁸. M. Le Guern définit un prédicat complexe comme suit : *on peut considérer un prédicat complexe comme le résultat de l'application d'un prédicat opérateur sur le prédicat opérande. Si le prédicat opérateur et le prédicat opérande sont tous deux des prédicats libres, le résultat est un prédicat lié ou en revanche il suffit qu'un des deux prédicats, l'opérateur ou l'opérande, soit lié, pour que le résultat soit nécessairement un prédicat lié*⁴⁹.

⁴⁶ M. Le Guern, op. cit., p.163-164

⁴⁷ Ibidem, p. 166

⁴⁸ W. MUSTAFA ELHADI, *La terminologie arabe des télécommunications : Faits de variation*, Thèse de doctorat en Sciences du Langage, Université Lumière, 1989, tome 1 p.279

⁴⁹ M. LE GUERN, *Sur les relations entre terminologie et lexique*, in actes du colloque : Les terminologies spécialisées. Approches quantitative et logico-sémantique, Méta, vol. 34, n°3, 1989, p.

L'information contenue dans un descripteur complexe est rarement rendue par les descripteurs qui le constituent. En effet, très souvent les descripteurs complexes perdent de leur signification si leurs constituants sont pris séparément.

G. ANTONIADIS⁵⁰, donne comme exemple *chemin de fer en France* si on prend « chemin » et « fer » cela n'aura pas la même signification que « chemin de fer ». D'où l'importance de tenir compte des descripteurs complexes dans les systèmes d'indexation automatique. Mais le problème qui se pose est le suivant : comment tenir compte de ces descripteurs dans les systèmes d'indexation automatique ? Combien de termes doit former le descripteur complexe ? On ne peut pas compter les occurrences de descripteurs contenant 2 à n mots consécutifs. Cela est coûteux et inutile, il y a un nombre limité d'ensemble de mots consécutifs pouvant être considérés comme descripteur complexe. *On constate qu'un descripteur ne peut pas être constitué de n'importe quelle suite de catégories grammaticales, il est rare voire impossible qu'un descripteur complexe contiennent des verbes, des adverbes, des signes de ponctuations*⁵¹.

Ceci est certes vrai pour le français mais qu'en est-il pour l'arabe ? Nous avons précédemment cité le travail de W. ABBAS sur la définition des unités linguistiques qui intervenaient dans l'indexation automatique arabe. Après délimitation des catégories (catégories des nominaux et leurs fonctions dans la construction des SN, les prédéterminants et leurs fonctions, etc.) elle a procédé à une étude des relations morphosyntaxiques entre elles ainsi que l'attribution de traits grammaticaux correspondant à chacune d'entre elles. Elle a examiné les différentes structures linguistiques des constituants de la phrase arabe et les éléments constituants de la catégorie du SN à savoir la tête, les modificateurs, les compléments, les prédéterminants, etc. L'objectif de son travail était l'essai d'une grammaire de reconnaissance pour l'extraction du SN. Notre étude vient en complément de ce travail, nous nous intéressons non pas à l'approfondissement de

⁵⁰ G. ANTONIADIS, *Recherche des descripteurs complexes pour l'indexation automatique des documents*, Mémoire de DEA, juin 1980, p.3

ces unités linguistiques mais à l'étude de l'une d'entre elle l'annexion, structure à la fois syntaxique et sémantique et sur laquelle nombre de descripteurs complexes sont bâtis. Ce sont ces descripteurs complexes qui posent des difficultés de reconnaissance pour les systèmes et donc posent également des difficultés d'extraction.

Nous allons voir dans la partie qui suit ce qu'il en est pour les descripteurs complexes en langue arabe.

⁵¹ Ibidem, p.4

DEUXIEME PARTIE

Nous allons voir dans cette partie ce que représente l'annexion dans la tradition grammaticale arabe. Nous verrons ensuite les différentes catégories syntaxiques et sémantiques de cette structure annective. Ces deux points nous permettront d'élaborer en troisième lieu des modèles de structures de l'annexion pour l'essai d'une grammaire du syntagme annexionnel à travers un traitement morphosyntaxique et syntactico-sémantique.

I-La définition de l'annexion dans la tradition linguistique arabe

Comme nous l'avons vu précédemment, la langue arabe a recours à deux systèmes pour la constitution des SN, le système de nomination et le système de communication. Nous avons également remarqué que les structures les plus fréquentes en documentation spécialisée sont des structures complexes.

Ces termes complexes sont représentés en arabe par ce que l'on appelle l'annexion. Qu'est-ce-que cela désigne exactement ?

L'annexion met "*en relation deux termes nominaux essentiels à sa formation. Le premier terme, la tête de syntagme est le mudâf ou l'annexant. Le deuxième terme, le complément est le mudâf ilayhi, l'annexé*"¹. L'annexion a un rôle essentiellement syntactico-sémantique.

Les règles grammaticales qui régissent l'annexion sont les suivantes :

1- le premier terme (l'annexant) ne prend jamais l'article et se met au cas voulu par sa fonction dans la phrase. Il ne porte jamais la nounation (le tanwin).

2- le second terme (l'annexé) porte l'article ou non selon le sens et il se met au cas indirect : exemple : *le drapeau du pays* علم البلاد

¹ Abdelkader FASSI-FEHRI, *La linguistique arabe, formes et interpretationm* Rabat, 1982, p.189-190

² Gérard LECOMTE, Ameer GHEIDIRA, *Méthode d'arabe littéral*, premier et deuxième livre, deuxième édition, Paris : Editions Klincksieck, 1970, p. 54

le drapeau d'un pays

علم بلاد³

ou *un drapeau de pays*

Comme nous venons de voir l'annexion comporte deux termes au minimum, elle peut également en comporter trois, rarement plus :

exemple :

la porte de la chambre à coucher

باب غرفة النوم⁴

Dans le cas d'une annexion à trois termes, seul le dernier terme peut avoir pour la détermination l'article (si c'est un nom) ou être un pronom affixe. Tous les termes sauf le premier se mettent au cas indirect. Rien ne peut séparer les termes d'une annexion sauf le démonstratif, ce qui signifie que l'épithète de

exemple .

la porte de cette chambre

باب هذه الغرفة⁵

L'annexion peut être de deux types, l'une dite "pure" ou "logique" (ou encore de dépendance) الاضافة المعنوية, de type groupe nom+complément de nom en français, et l'autre est l'annexion "formative" ou "qualificative" الاضافة اللفظية, comme l'indique Ibn Ya'îš :

واضافة الاسم الى اسم على ضربين معنوية ولفظية (...)⁶

L'annexion d'un nom à un nom est de deux types l'annexion logique et l'annexion qualificative (...).

³ Ibidem

⁴ Ibidem, p. 138

⁵ M. NEYRENEUF, op. cit. p. 138

⁶ *šarḥ al-mufaṣal li-ibn Ya'îš bin Ya'îš al-naḥwî*, 3^e partie Librairie al-mutanabî Le Caire, p. 118,

W. Wright précise également que :

*The Arab grammarians speak of two kinds of an annexation. The one is called "the proper" or "real" annexation, the pure annexation or the logical annexation ; the other "the improper annexation, the impure annexation, or the (merely) verbal or grammatical annexation."*⁷

Les grammairiens arabes distinguent deux types d'annexion. L'une appelée "la vraie" annexion ou l'annexion proprement dite "l'annexion pure" ou l'annexion logique. L'autre "l'annexion impropre, l'annexion impure, l'annexion verbale ou l'annexion grammaticale.

a) L'annexion dite "pure" (l'annexion connotative)

Le premier terme de cette annexion est toujours un nom défini par le second terme. Sachant qu'un nom ne peut être défini (التعريف) ou déterminé (التخصيص) que par un article ou une annexion comme le souligne Ibn Yaʿiš :

⁸ فالمعنوية ما افاد تعريفها كقولك دار عمر او تخصيصا كقولك غلام رجل (...)

L'annexion qualificative est définie quand tu dis "la maison de Amr" et elle est déterminée quand tu dis le serviteur d'un homme (...).

Wright donne également la définition d'un terme défini et d'un terme déterminé :

⁷ W. WRIGHT, *A grammar of the Arabic language*, Cambridge University Press 1859 3rd Edition, Librairie du Liban, 1981, 3rd part, p. 198

⁸ *ṣarḥ al-mufaṣṣal li-ibn Yaʿiš bin Yaʿiš al-naḥwī*, 3^e partie Librairie al-mutanabī Le Caire, p.

The *التخصيص* consists in qualifying an indefinite noun by an adjective or an expression equivalent to an adjective, as a proposition with a genitive or the genitive of an undefined noun, *التعريف* is the defining of the noun by the genitive of a defined noun(...) ⁹

Le *التخصيص* consiste à déterminer un nom indéfini par un adjectif ou une expression équivalente à un adjectif comme dans le cas d'une préposition suivi d'une annexion ou l'annexion d'un nom indéfini. *التعريف* consiste à définir un nom par une annexion d'un nom défini (...).

Le premier terme n'aura donc ni article ni la nounation (tanwin) ni le "ن" final pour les pluriels externes masculins et les duels. Il se met au cas voulu par sa fonction dans la phrase .

Exemples ¹⁰ :

Il a ouvert la porte de la maison فتح باب الدار

Les enseignants de l'école sont ici مدرسو المدرسة هنا

Le second terme peut être un nom avec ou sans article suivant le sens au cas indirect :

Exemples ¹¹ :

Le château du roi قصر الملك

Un château de roi ou le château d'un roi قصر ملك

⁹ Ibidem p.198

¹⁰ exemples pris dans Michel NEYRENEUF, Ghalib AL HAKKAK, *Grammaire active de l'arabe*, Librairie générale française, collection Les langues modernes, 1996, p. 137

¹¹ Ibidem, p.138

Dans ce dernier exemple (un château de roi) le premier terme est défini grammaticalement par le second terme et ne prend donc pas de tanwin. On tolère dans le cas d'une annexion pure que le premier terme prenne l'article " ال " si celui-ci est également joint au second terme comme dans l'exemple suivant :

celui dont la chevelure est crépue الجعد الشعر¹²

Le premier terme peut également être pris comme complément du second terme. Dans ce cas uniquement il faut que cette préfixation à l'adjectif soit sous la forme d'un duel ou d'un pluriel :

Zayd est celui qui frappe la tête au coupable زيد الضارب رأس الجاني¹³

b) L'annexion formative ou qualificative (الاضافة اللفظية)

A la différence de l'annexion logique le premier terme d'une annexion qualificative est un adjectif, il prend l'article s'il est épithète d'un nom défini. Le second terme est toujours défini au cas indirect :

un homme de grande taille ("un homme grand de taille") رجل طويل القامة¹⁴

le philosophe aux idées étranges الفيلسوف الغريب الافكار¹⁵

La détermination du premier terme se fait au moyen du second, si l'antécédent comme l'adjectif se trouve être analogue à " يفعل ", il ne sort point de son indétermination :

رب راحينا عظيم الامل مرؤع القلب قليل الحيل

*Combien comptent sur nous, dont l'espoir est immense, le coeur épouvanté, les ressources faibles !*¹⁶

¹² La *Alfiyya d'Ibnu- Malik*, traduit par G. GOGUYER, Librairie du Liban Publishers, p. 94

¹³ Ibidem, p.94

¹⁴ M. NEYRENEUF, op. cit. p. 142

¹⁵ Ibidem, p. 142

¹⁶ La *Alfiyya d'Ibnu- Malik*, op. cit. p. 94

Cette annexion qualificative est propre aux adjectifs qu'ils soient épithètes ou attributs. Elle rappelle l'annexion "pure" mais l'adjectif, en tant que premier terme de l'annexion qualificative, prendra l'article s'il est épithète d'un nom défini. C'est la seule exception possible à la règle qui précise qu'un nom ne peut être défini que d'une seule façon (par l'article ou l'annexion). Il sera toujours défini par l'annexion même si le nom auquel il se rapporte est un indéfini. Quant au second terme, il sera toujours défini au cas indirect. Les autres règles de l'annexion de dépendance s'appliquent :

Exemples¹⁷:

Un homme de grande taille رجل طويل القامة

mot à mot "un homme grand de taille"

Le philosophe aux idées étranges الفيلسوف الغريب الافكار

II-Les différentes structures annectives dans la tradition grammaticale arabe :

A-l'annexion nominale

1- Les différentes catégories sémantiques

A cette fonction grammaticale qu'est l'annexion correspondent différents rôles sémantiques représentant la relation syntactico-sémantique entre l'annexant et l'annexé. Ils sont au nombre de six :

- possession/appartenance
- forme/matière
- contenance
- annexion de "reformulation spécifique"
- spécifique

¹⁷ M. NEYRENEUF, op. cit. p.142-145

- élatif

- L'appartenance ou la possession (prépositon ل)

Cette catégorie désigne deux choses : la première est celle de l'appartenance et la deuxième est l'objet possédé par une personne, c'est notamment le cas de :

l'argent de Zayd مال زيد¹⁸

l'esclave de Zayd غلام زيد¹⁹

W. Wright désigne cette catégorie de la façon suivante :

By the genitive is indicated the thing possessed by a possessor and the possessor of a thing possessed as : خزينة السلطان *the treasure of the sultan :*

سلطان البر والبحر *the lord of the land and sea (...)*²⁰

Par le génitif on indique la chose possédée par un possesseur et le possesseur d'une chose possédée telle que : خزينة السلطان *la trésorerie du sultan ou* سلطان البر والبحر *le maître des terres et des mers.*

On peut également prendre l'exemple suivant :

la fortune et la terre de Zayd مال زيد و ارضه²¹

C'est ce que Ibn Ya'îš et les grammairiens arabes appellent "الاضافة بالمعنى اللام",

c'est-à-dire l'annexion ayant pour sens "pour, appartient à" :

إذا كانت الاضافة بمعنى اللام كان معناها الملك و الاختصاص و ذلك قولك مال زيد و ارضه أي مال له

و ارض له أي يملكه²²

¹⁸ Šarḥ al-mufaṣṣal li-ibn Ya'îš bin Ya'îš al-naḥwî, 2^e partie Librairie al-mutanabî Le Caire, p. 118

¹⁹ Ibidem

²⁰ W. WRIGHT, op. cit., p. 199

²¹ Šarḥ al-mufaṣṣal li-ibn Ya'îš bin Ya'îš al-naḥwî, 2^e partie Librairie al-mutanabî Le Caire, p. 119

Dans le cas d'une annexion dont le sens serait celui de la particule "ل" cela désigne la possession ou une particularité comme lorsque l'on dit "la fortune et la terre de Zayd" c'est-à-dire "une fortune à lui et une terre à lui" cela signifie qu'il les possède.

W. Wright précise également que les grammairiens arabes déclaraient que la vraie annexion était celle qui avait pour sens celui de la préposition "ل" c'est-à-dire "appartient à, pour, etc."

The Arab grammarians say that in the real annexation is implied the force of a preposition with is either ل (...) Zeid's slave غلام زيد or the slave who belongs to Zeid أي الغلام الذي لزيد²³ :

Mais dans l'exemple d'Ibn Ya'îš, on a une seule et même structure pour désigner sémantiquement deux notions différentes de la possession :

²⁴مال زيد و ارضه وابنه و سيده و عبده

La fortune de Zayd, sa terre, son fils, son esclave et son maître.

La nuance entre la possession et l'appartenance est ici implicite "سيده و عبده" en effet "سيده" c'est-à-dire son maître n'est nullement la possession de Zayd mais Zayd travaille pour lui, il fait partie de son monde, ils appartiennent tous deux au même monde. Cette structure démontre le caractère dual de la possession : le premier celui de posséder un objet ou une chose ou un esclave comme dans l'exemple cité, et le second sens "appartenir à quelque chose, à un groupe etc"²⁵. Il s'agit donc de faire la différence entre ce qui m'appartient réellement comme "mon stylo" par exemple et ce qui ne m'appartient pas réellement mais fait partie de mon univers comme "mon école, mon agence, mon train ou mon bus".

²²Šarḥ al-mufaṣṣal li-ibn Ya'îš bin Ya'îš al-naḥwî, 2^e partie Librairie al-mutanabî Le Caire, p. 119

²³ W. WRIGHT, Op. cit., p.200

²⁴ Šarḥ al-mufaṣṣal li-ibn Ya'îš bin Ya'îš al-naḥwî, 2^e partie Librairie al-mutanabî Le Caire, p. 119

- La forme et la matière (préposition من)

L'annexion a le sens de "من" lorsque le terme annexé complète la nature du terme annexant comme dans les exemples suivants :

c'est un vêtement soyeux (en soie) هذا ثوب خز²⁶

une bague en fer خاتم حديد²⁷

Ainsi ces deux exemples d'annexion désignent la matière ou la forme, nous pourrions ainsi dire :

هذا ثوب خز اي هذا ثوب من خز

خاتم حديد أي خاتم من حديد

*The annexation is resolved by "من" when the مضاف اليه is the genitive or material of the مضاف as ثوب خز a silk dress (...)*²⁸

L'annexion est résolue par "من" quand l'annexé désigne une forme ou une matière de l'annexant comme dans *ثوب خز un vêtement soyeux (...)*.

Ibn 'Aqîl donne la même règle :

(...) فيتعين تقدير "من" ان كان المضاف اليه جنسا للمضاف نحو "هذا ثوب خز" و "خاتم

حديد" و "التقدير هذا ثوب من خز و خاتم من حديد (...)"²⁹.

Pour signifier "من" dans le cas ou l'annexé (المضاف اليه) est une catégorie ou forme pour l'annexant (المضاف) comme dans "هذا ثوب خز" un

²⁵ J. DICHY, communication verbale

²⁶ Šarḥ ibn 'aql, 3^e partie, Dar al-fikr p. 43

²⁷ Ibidem, p.43

²⁸ W. WRIGHT, Op. cit., p.200

²⁹ Šarḥ ibn 'aql, 3^e partie, Dar al-fikr p. 44

vêtement en soie et "خاتم حديد" une bague en fer et la structure sous jacente serait " هذا ثوب من حر " un vêtement en soie et حاتم من حديد une bague en fer.

-L'annexion exprimant la contenance (préposition في)

Nous avons vu précédemment différentes catégories sémantiques de l'annexion, nous allons voir à présent une autre catégorie représentant la contenance.

Dans ce type de structure l'annexant où le premier terme désigne le contenant et l'annexé où le deuxième terme désigne le contenu comme dans l'exemple suivant :

une tasse de café فنجان قهوة³⁰

اي فنجان فيه قهوة

un verre d'eau كأس ماء

اي كأس فيه ماء

Nous remarquerons que le français a également la même structure pour exprimer une idée de contenance qui ne concerne que les liquides ou assimilés.

- L'annexion de "reformulation spécifique"

C'est une annexion qui spécifie la nature du terme annexant comme dans *نهر النيل le fleuve Nil, شهر رمضان le mois de ramadan* ou encore *مدينة بغداد la ville de Bagdad*³¹. C'est une double spécification car on sait que le Nil est un fleuve et le Ramadan est un mois du calendrier lunaire musulman, la fonction de ce type d'annexion est d'insister sur la nature sémantique du terme.

³⁰ A. FASSI-FEHRI, *Linguistique arabe, forme et interprétation*, Rabat, p. 132

³¹ R. BLACHERE, M. GAUDEFROY-DEMOMBYNES, *Grammaire de l'arabe classique*, Maisonneuve et Larose, 3^e Edition, 1978, p.324

- le spécifique

Le spécifique ou تمييز a une structure syntaxique différente de l'annexion mais a en commun avec celle-ci un complément second. Dans ce cas, le groupe nominal est composé du nom et de l'épithète suivi d'un nom au cas direct indéfini qui spécifie la nature de la qualification comme par exemple :

un homme laid de visage رجل قبيح وجهاً pour reprendre la structure de l'annexion qualificative رجل قبيح الوجه.³²

Le spécifique comprend également le sens de "من", il est explicatif. Lorsqu'il est indéterminé, il est régi à l'accusatif comme spécifique c'est-à-dire pour clarifier ou préciser le terme qui précède, comme dans les exemples suivants :

un empan de terre شبر ارضاً³³

une mesure de froment قفيز برأ³⁴

deux livres de miel et de dattes منوان عسلاً و تمرأ³⁵

Peuvent être mis au génitif les mots que l'on emploie comme antécédents d'annexion comme dans l'exemple suivant :

مد حنطة غداً *une mesure de blé c'est un aliment*³⁶.

L'accusatif est nécessaire après l'antécédent d'annexion si celui-ci est comme ملء الارض ذهباً (..) *pleine la terre d'or*³⁷. On peut faire régir le nom logiquement sujet verbal, à l'accusatif par افعل qui agirait comme un superlatif, comme par exemple انت اعلى منزلاً *tu es très élevé en position*³⁸.

³² Michel NEYRENEUF, op. cit., 1996, p.142

³³ La Alfyyah d'Ibnu Malik, op. cit., p.86

³⁴ Ibidem, p.86

³⁵ Ibidem, p.86

³⁶ Ibidem, p. 85

³⁷ Ibidem, p.85

³⁸ Ibidem, p. 86-87

Le spécificatif peut être également employé pour exprimer l'admiration comme par exemple : *يا أيُّ بكرٍ أباً! Abu Bakr, quel noble père!*³⁹

-L'élatif

L'élatif est construit sur le même schème que celui des couleurs *افعل* (mais avec un féminin en *فعلى* et un pluriel en *فعل*) à partir d'adjectifs de racines trilitères.

Il exprime le comparatif ou le superlatif : *كبير أي أكبر*.

L'élatif sert à exprimer l'admiration il est implicitement ou non suivi de "من".

Il reste au masculin singulier lorsqu'il est annexé à un complément indéterminé ou lorsqu'il est isolé. Préfixe de l'article, il s'accorde avec le terme annexé à un complément déterminé lorsqu'on sous-entend le sens de "من", dans les autres cas, il s'accorde avec son antécédent.

L'élatif régit le substantif au nominatif uniquement quant il supplée un verbe comme par exemple :

*لن ترى في الناس من رفيق اولى به الفعل من الصديق*⁴⁰

Vous ne verrez pas dans le monde un compagnon chez qui la prééminence sied mieux que chez l'ami

Ibn Ya'îš donne également un exemple de cette utilisation :

اضافة الفعل الى ما بعده اضافة البعض الى الكل و الواحد الى الجنس مثل حمارك افرح الحمير و عبدك خير

العبيد⁴¹

L'annexion de افعل à ce qui suit est l'annexion de certain à tout, de l'unique à l'ensemble comme l'exemple "ton âne est le plus gai des ânes, et ton esclave est le meilleur des esclaves".

³⁹ Ibidem, p. 86

⁴⁰ Ibidem, p.118-121

⁴¹ *Šarḥ al-mufaṣṣal li-ibn Ya'îš bin Ya'îš al-naḥwî*, 3^e partie Librairie al-mutanabî Le Caire, p.5

Ceci pour signifier que :

عبدك احسن الاحرار و حمارك افراح البغال أي عبدك احسن من الاحرار و حمارك افراح من البغال⁴²

Ton âne est le plus gai des ânes, et ton esclave est le meilleur des esclaves

Pour Ibn Yaʿīš, lorsque l'on dit :

زيد افضل الرجال او هو افضل الرجال يعني انه واحد منهم و اصله افضل الرجال لانك خفت

الواحد المنكر يدل على الجنس⁴³

Zayd est le meilleur des hommes ou il est le meilleur des hommes c'est-à-dire qu'il est l'un d'eux et sa nature est d'être le meilleur des hommes car tu as indiqué que le masculin singulier désigne le genre.

W. Wright précise un autre point sur cet usage :

Nouns of the forms فعل افعل etc used as superlatives are construed as substantives in the singular masculine with the genitive of the word denoting the objects among which the one spoken of is preeminent. The generative is at times indefinite and explicative at times definite and partitive.

for example of the indefinite genitive :

هو افضل رجل *he is the most excellent man*

example of the definite genitive :

هي افضل النساء ⁴⁴ *she is the best of the women*

Les noms de la forme فعل افعل utilisés aussi pour le superlatif sont construits comme des substantifs au masculin singulier avec le génitif du mot dénotant les objets parmi lesquels celui dont on parle est prééminent.

⁴² Ibidem, p.5

⁴³ Šarḥ al-mufaṣṣal li-ibn Yaʿīš bin Yaʿīš al-naḥwî, 3^e partie Librairie al-mutanabî Le Caire, p.5

⁴⁴ W. WRIGHT, op. cit. , p.226-227

Le génitif est parfois indéfini et explicatif d'une part et défini et partitif d'autre part.

exemple d'un génitif indéfini :

C' est un homme remarquable هو افضل رجل

exemple d'un génitif défini:

elle est la plus remarquable des femmes هي افضل النساء

Si le génitif est défini, l'adjectif peut s'accorder en genre et en nombre avec l'objet ou les objets dont on parle comme par exemple :

elle est la plus remarquable des femmes هي فضلى النساء

ce sont les femmes les plus remarquables هن فضليات النساء

ou encore هن فضل النساء

On doit également mentionner l'annexion indéfinie après اول premier et اخر dernier, ils peuvent tous les deux être utilisés pour la structure de l'élatif comme par exemple :

ان اول بيت وضع للناس للذي ببكة

The first house (temple) which was founded for mankind, was that in Bekka (Mekka)⁴⁵

La première maison qui a été fondée pour l'humanité c'est celle de la Mecque

On peut trouver pour certaines constructions temporelles comme celle de اول يوم une structure plus fréquente اليوم الاول pour dire le premier jour.

⁴⁵ Ibidem, p.226-227

B-L' annexion à un modificateur

Nous allons voir d'une part l'annexion à un modificateur à proprement parler et d'autre part l'annexion numérale, qui est une partie des annexions à un modificateur.

1.L'annexion numérale

C'est l'annexion d'une catégorie particulièrement bien codée et ne représentant pas de différences de règle par rapport aux autres annexions. Il existe néanmoins plusieurs variations d'usage allant de la langue très classique inutilisée aujourd'hui à l'arabe moderne

Comment doit-on considérer l'annexion à un numéral ? Est-ce une tête de syntagme comme le nom ou des termes nominalisés ? Ou est-ce un spécifieur de la construction nominale et en l'occurrence un quantificateur ? Le fait de le traiter ici dans la partie des modificateurs nous amène à le considérer comme un quantificateur.

La position du numéral à l'intérieur d'un SN simple : nous allons voir les trois ordres possibles entre l'article et le numéral en arabe classique :

- l'ordre NUM+ART+N

Cet ordre se trouve être dans toute construction numérale simple, c'est-à-dire, où le numéral n'est pas composé :

exemple :

les cinq vêtements

خمسة الاثواب⁴⁶

- l'ordre NUM+N

Ici le nombre est annexé et ne porte pas l'article car il est défini par le terme annexé (règle classique de l'annexion) :

exemple :

⁴⁶ Šarḥ al-mufaṣṣal li-ibn Yaʿīš bin Yaʿīš al-naḥwī, 2^e partie Librairie al-mutanabî Le Caire, p.121

les trois dirhams

ثلاثة دراهم⁴⁷

- l'ordre ART+NUM+N

C'est l'ordre des nombres composés comme عشرون vingt qui ne peuvent pas prendre le و de la nounation ou tanwin. Cette désinence disparaît normalement avec l'annexion, on a ainsi :

*les treize hommes*الثلاثة عشر رجلاً⁴⁸*les vingt dirhams*العشرون درهماً⁴⁹

Nous remarquons ici que رجل et درهم ne portent pas l'article et fonctionnent comme tamyiz⁵⁰. On dira plutôt en arabe moderne, الاثواب الخمسة (ceci ne fonctionne pas pour les nombres).

Nous avons en fait trois types d'ordre : NUM+ART+N, NUM+N, ART+NUM+N

En arabe classique, le nombre peut être construit sur le modèle annectif, l'article défini quand il apparaît est porté par le complément du nom :

exemple :*les trois hommes d'affaires*ثلاثة رجال الاعمال⁵¹

De la même manière, lorsque le numéral est construit avec un nom de mesure (auquel il est annexé) et non avec le cardinal :

exemple :*les trois mètres en longueur*ثلاثة الامتار طولاً⁵²

⁴⁷ A. FASSI-FEHRI, op. cit. p.209

⁴⁸ Ibidem

⁴⁹ Ibidem, p.208

⁵⁰ Il y a une grande discussion à ce sujet entre les Basriens et les Koufiens, se reporter à la thèse de Fassi-Fehri, *Linguistique arabe, forme et interprétation*, Rabat, p. 208

⁵¹ A. FASSI-FEHRI, op. cit. p. 209

⁵² A. FASSI-FEHRI, op. cit. p.210

En arabe moderne on dira plutôt :

الثلاثة امتار طولاً⁵³

On ne pourra pas avoir :

الامتار الثلاثة طولاً

Pour considérer le numéral comme tête de syntagme, il faut qu'il ait la même valeur qu'un nom. Pour Fassi-Fehri, le numéral est à la fois partiellement nominal et adjectival, ce qui signifie qu'il est pleinement nominal. Il a donc les propriétés d'un nom, il reçoit le même cas que le terme auquel il est annexé en fonction du rôle que celui-ci remplit dans la phrase. Comme tête nominale ordinaire, le numéral est au nominatif et le dénombré au génitif ou bien ils sont tous les deux au nominatif. Le numéral a la possibilité d'avoir le duel, le pluriel, la non-variabilité en fonction du nom dénombré. Il peut avoir un déterminé qui s'accorde avec lui et non avec le dénombré.

2. Lexique des modificateurs et quantificateurs

Les modificateurs désignent certains mots outils qui sont des noms combinés avec d'autres noms ou pronoms suffixes.

Les modificateurs ont ici le rôle de prédéterminants comme dans *tous كل الناس* *les gens*, où *كل* est modificateur de *ناس* c'est-à-dire prédéterminant. Car si *كل* est le premier terme d'une annexion, alors dans *كل الناس, الناس* est complément de *كل* ce qui est contraire au résultat d'une analyse sémantique, d'où son statut de modificateur.

Le quantificateur exprime la totalité, la dualité ou la partie. Il est également exprimé par l'annexion du numéral. Ces quantificateurs sont souvent annexés à des noms ou pronoms et ont une portée sémantique précise. C'est dans ce sens que nous avons pensé à l'utilité d'un lexique qui pourrait servir d'index pour un analyseur dans le cadre d'un module d'indexation automatique.

⁵³ Ibidem

	REGLES	EXEMPLES
QUANTIFICATEURS		
كل	<p>Ce mot désigne la <i>totalité</i>. il est couramment utilisé et peut se construire de nombreuses façons.</p> <p>- lorsqu'il est en rapport d'annexion avec un nom indéfini au singulier il a pour signification <i>chaque, chacune, chacun, tous les</i>.</p> <p>Lorsqu'il complète un nom indéfini quantifieur au pluriel : <i>tous les, chaque</i></p>	<p style="text-align: right;">استقبل كل ضيف⁵⁴</p> <p><i>Il a accueilli chaque invité</i></p> <p style="text-align: right;">كل شهر⁵⁵</p> <p><i>chaque mois (tous les mois)</i></p> <p style="text-align: right;">كل واحد⁵⁶</p> <p><i>chacun</i></p> <p style="text-align: right;">كل واحدة⁵⁷</p> <p><i>chacune</i></p> <p style="text-align: right;">كل ستة شهر⁵⁸</p> <p><i>tous les six mois, chaque semestre</i></p>
	- complété par un nom indéfini ou un pronom suffixe : <i>tout (e), tous, toutes</i>	<p style="text-align: right;">استقبل كل الضيوف⁵⁹</p> <p><i>il accueillit tous les invités</i></p>
	<p>En apposition après un nom indéfini : <i>tout, tous, toutes, en entier</i>. Dans ce cas كل est repris par un pronom suffixe qui reprend le nom.</p>	<p style="text-align: right;">جاء الضيوف كلهم⁶⁰</p> <p><i>Tous les invités sont venus</i></p>

⁵⁴ M. NEYRENEUF et HAKKAK, op. cit. p.156

⁵⁵ Ibidem, p. 156

⁵⁶ Ibidem, p. 156

⁵⁷ Ibidem, p. 156

⁵⁸ Ibidem, p. 156

⁵⁹ Ibidem, p. 155

⁶⁰ Ibidem, p. 156

		فرح الشعب كله ⁶¹ <i>Le peuple tout entier s'est réjoui</i>
	Lorsqu'il est tout seul indéfini et suivi de la préposition من il a le sens de <i>chacun de</i>	كل منهم = كل من الضيوف ⁶² <i>Chacun des invités, chacun d'eux</i>
	seul, défini par l'article il signifie le <i>tout, la totalité, tout le monde</i>	لكل يعرف الخير ⁶³ <i>Tout le monde connaît la nouvelle</i>
جميع	il exprime <i>la totalité</i> ou <i>l'ensemble</i> il peut se substituer à lui dans les constructions suivantes :	
	-seul défini par l'article il signifie <i>tous, tout le monde</i>	الجميع ⁶⁴
	-sous forme adverbiale, il signifie en <i>totalité, entièrement</i>	جميعا ⁶⁵
	-comme premier terme d'une annexion complété par un nom défini ou un pronom suffixe, il signifie <i>tout (e), tous, toutes</i>	جميع الناس ⁶⁶ <i>Tous les gens</i> جميعكم ⁶⁷ <i>Vous tous</i>
	-en apposition après un nom défini (tout, tous, toutes, en entier) complété par un pronom suffixé	فرح الشعب جميعه ⁶⁸ <i>Le peuple tout entier s'est réjoui</i> احتفلت المدينة جميعها بالعيد ⁶⁹ <i>La ville toute entière s'est</i>

⁶¹ Ibidem, p. 156

⁶² Ibidem, p. 157

⁶³ Ibidem, p. 157

⁶⁴ Ibidem, p. 157

⁶⁵ Ibidem, p. 157

⁶⁶ Ibidem, p. 157

⁶⁷ Ibidem, p. 157

⁶⁸ Ibidem, p. 157

⁶⁹ Ibidem, p. 157

		<i>réjouï</i>
مجموع / كامل	Ces mots sont précédés de la préposition ب avec un sens identique, on a également une reprise du terme qui précédé par le pronom affixé.	<p>70 جاء الناس بكاملهم</p> <p><i>Tous les gens (sans exception) sont venus</i></p> <p>71 جاء الناس بجمعهم</p> <p><i>idem</i></p>
اجمع	C'est l'élatif de جميع , il s'emploie surtout en apposition, complété par le pronom suffixé (<i>tout (e), tous, en entier</i>)	<p>حتفلت المدينة اجمعها (باجمعها)</p> <p>72 بالعيد</p> <p><i>La ville entière a célébré la fête</i></p>
كافة	Ce nom s'emploie surtout comme premier terme d'une annexion ou en apposition (<i>l'ensemble sans exception</i>)	<p>73 كافة انحاء الارض</p> <p><i>Toute la région de la terre</i></p>
كلا/كلتا	Exprime la dualité, il signifie <i>les deux</i> ou <i>tous les deux</i> : Comme premier terme d'une annexion, il s'accorde en genre mais reste figé au cas sujet s'il est complété par un nom.	<p>74 بكلا يديه</p> <p><i>Avec les deux mains</i></p> <p>75 تكلمت مع كليهم</p> <p><i>J'ai parlé avec tous les deux</i></p>
بعض	Expression de la partie : - employé seul avec l'article il signifie <i>certain</i> en arabe moderne et <i>l'un d'entre eux</i> en arabe classique. - employé comme premier terme d'une annexion, si son complément est un singulier on le traduira par <i>quelque, un peu de ...</i>	<p>76 بعض الراحة .</p> <p><i>Quelque tranquillité, un peu de tranquillité</i></p> <p>77 قال بعضهم</p> <p><i>Certains d'entre eux disent (classique : l'un d'entre eux a dit)</i></p>

70 Ibidem, p. 157

71 Ibidem, p. 157

72 Ibidem, p. 157

73 Ibidem, p. 158

74 Ibidem, p. 158

75 Ibidem, p. 158

76 Ibidem, p. 158

77 Ibidem, p. 158

معظم	Expression de la <i>majorité, la plus grande partie</i> (sens moderne) <i>l'ensemble de, le reste de,</i> - utilisé comme premier terme d'une annexion, il signifie <i>la plupart de</i>	في معظم الاوقات ⁷⁸ <i>La plupart du temps</i>
سائر	Idem	سائر دول العالم ⁷⁹ <i>L'ensemble des pays du monde</i>
احد/احدى	Exprime la partie : - comme premier terme d'une annexion <i>un des.</i> A la différence des autres termes celui-ci s'accordera en genre et en nombre avec le nom ou le pronom qui le complète.	احدى البنات ⁸⁰ <i>Une des filles</i> احد هم ⁸¹ <i>L'un d'eux</i>
واحد/ واحدة	Ce nom désigne la partie et se construit comme premier terme d'une annexion. Dans ce cas, il est défini par l'article et est suivi de la préposition <i>من chaque, chacun d'eux.</i> Il s'accorde en genre avec le nom ou le pronom qui suit.	كان واحد منهم يعمل كل النهار ⁸² <i>Chacun d'entre eux travaillait toute la journée</i>
نفس	Pronom réfléchi il signifie <i>le même.</i> Construction possible : - en apposition à un nom (avec un pronom de renvoi) ou comme premier terme d'une annexion : <i>le même,</i> - en apposition à un pronom isolé (avec un pronom de renvoi) ou précédé de la préposition <i>ب</i> , elle-même suivie d'un pronom de renvoi (<i>moi/toi, même</i>). Dans le cas de ces deux constructions,	نفس الوقت ⁸³ <i>Le même moment</i> الشيء نفسه ⁸⁴ <i>La même chose</i> ou bien نفس الشيء Cette construction de plus en plus fréquente aujourd'hui était autrefois jugée incorrecte

⁷⁸ Ibidem, p. 159

⁷⁹ Ibidem, p. 159

⁸⁰ Ibidem, p. 159

⁸¹ Ibidem, p. 159

⁸² Ibidem, p. 159

⁸³ Ibidem, p. 160

⁸⁴ Ibidem, p. 160

	il peut être remplacé par بعينه ou par ذاته.	<p>جاء الملك نفسه⁸⁵</p> <p><i>Le roi est venu lui-même</i></p> <p>رايت الرجل ذاته⁸⁶</p> <p><i>J'ai vu l'homme lui-même</i></p>
ذو	<p>Nom proche du démonstratif ذا . Il s'emploie toujours comme premier terme d'une annexion. Il signifie <i>possesseur de</i>. Le groupe formé par l'ensemble de l'annexion a une valeur d'adjectif qualificatif. Qualifiant un nom défini par l'article ce groupe sera aussi défini par la présence de l'article entre les deux termes de l'annexion. ذو varie en cas :</p> <ul style="list-style-type: none"> - cas sujet : جاء رجل ذو مال - cas indirect : اتكلم عن رجل ذي مال - cas direct : رأيت رجلا ذا مال <p>Il s'accorde en genre et en nombre avec le nom auquel il se rapporte qu'il soit en apposition ou en prédicat. Il est nécessairement défini par un nom complément.</p>	<p>رجل ذو مال⁸⁷</p> <p><i>Un homme riche</i> (<i>possesseur d'argent</i>)</p> <p>الرجل ذو المال⁸⁸</p> <p><i>L'homme riche (possesseur d'argent)</i></p> <p>الرجال ذو القبعات السوداء⁸⁹</p> <p><i>Les hommes au chapeau noir</i></p> <p>البنات ذات الوجه الجميل⁹⁰</p> <p><i>La fille au beau visage</i></p>

Les modificateurs peuvent être assimilés à des prépositions (حروف الجر) on les désigne par le terme de quasi-prépositions (ظروف). Ce sont en fait des noms employés généralement au cas direct. Ils peuvent jouer le rôle de préposition lorsqu'ils sont complétés par des substantifs, ils se mettent alors au cas indirect. Elles peuvent toutes être précédées d'une préposition (essentiellement من ل من) et se mettent alors au cas

⁸⁵ Ibidem, p. 160

⁸⁶ Ibidem, p. 160

⁸⁷ Ibidem, p. 160

⁸⁸ Ibidem, p. 160

⁸⁹ Ibidem, p. 161

⁹⁰ Ibidem, p. 161

indirect, alors que les prépositions proprement dites ne peuvent être précédées par une autre préposition.

QUASI- PREPO- SITION	REGLES	EXEMPLES
MODIFICATEURS		
مع	<p>- <i>en compagnie de</i></p> <p>- <i>au moment de</i> (équivalent de عند)</p> <p>- sert à exprimer le fait <i>d'avoir quelque chose sur soi, avec soi,</i></p> <p>- <i>malgré, en dépit de</i></p>	<p>رجع معي⁹¹</p> <p><i>Il est revenu avec moi</i></p> <p>مع الصباح⁹²</p> <p><i>Au matin</i></p> <p>معها قلم⁹³</p> <p><i>Elle a (avec elle) un stylo</i></p> <p>مع ذلك⁹⁴</p> <p><i>Malgré cela</i></p>
عند	<p>1. <i>chez, auprès de, à côté,</i></p> <p>2. <i>au moment de</i></p> <p>3. comme مع elle sert à exprimer la possession ou la comparaison</p> <p>4. <i>de l'avis de, pour quelqu'un, aux yeux de ...</i></p>	<p>عنده الطبيب⁹⁵</p> <p>1. <i>chez le médecin</i></p> <p>عنده الصباح⁹⁶</p> <p>2. <i>au matin</i></p> <p>عنده خبرة⁹⁷</p> <p>3. <i>il a de l'expérience</i></p> <p>ملوك الارض عند الله تراب⁹⁸</p> <p>4. <i>The kings of the earth are dust compared with (lit. at the side of) God (Les rois de la terre sont</i></p>

⁹¹ M. NEYRENEUF, op. cit. p.117

⁹² Ibidem, p.117

⁹³ Ibidem, p. 117

⁹⁴ Ibidem, p.117

⁹⁵ Ibidem, p.117

⁹⁶ Ibidem, p.117

⁹⁷ Ibidem, p.117

⁹⁸ W. WRIGHT, op. cit., p.179

	5.utilisée pour le temps, cette préposition indique qu'une chose est très liée par rapport à un événement qui a eu lieu à un moment particulier soit simultanément soit tout de suite après.	seulement de la poussière comparés à (lit. à côté de Dieu)) فقال عند موته ⁹⁹ 5.He said as he was dying (il dit en mourrant)
لدى	- devient لديه devant un pronom suffixé, - équivalent de عند mais beaucoup moins fréquent - sert à exprimer la possession : NB : on peut également trouver لدن en arabe classique à la place de لدى	لدى وصوله ¹⁰⁰ Dès son arrivée ليس لدينا ثقة فيه ¹⁰¹ Nous n'avons pas confiance en lui
اثناء	au cours, pendant	اثناء زيارته ¹⁰² Pendant sa visite
رغم	malgré, en dépit de	رغمه ¹⁰³ Malgré lui
ازاء	face à, en présence de	ازاء الوزير ¹⁰⁴ En présence du ministre
ضد	contre, opposé à, anti-	ضد الحكومة ¹⁰⁵ Anti-gouvernemental
امام	devant, face à, exprime le temps et le lieu	امام البلدية ¹⁰⁶ Devant la mairie الصلاة امامك The (time of) prayer is before thee ¹⁰⁷ ("le temps de la prière est devant toi", c'est l'heure de ta prière)

⁹⁹ Ibidem, p.178

¹⁰⁰ M. NEYRENEUF, op. cit. p.118

¹⁰¹ Ibidem, p.118

¹⁰² Exemple créé par nous.

¹⁰³ Ibidem

¹⁰⁴ Ibidem

¹⁰⁵ Ibidem

¹⁰⁶ Ibidem

¹⁰⁷ W. WRIGHT, op. cit. p.187

طبق	<i>d'après, selon, conforme à</i>	نسخة طبق الاصل ¹⁰⁸ <i>Une copie conforme</i>
بدل	<i>au lieu de, en échange de</i>	بدلاً من الاسبوع القادم ¹⁰⁹ <i>Au lieu de la semaine prochaine</i>
عبر	<i>à travers, pendant</i>	عبر الاجيال ¹¹⁰ <i>A travers les générations</i>
بعد	<i>après,</i>	بعد ثلاثة ايام ¹¹¹ <i>After three days (après trois jours).</i>
عوض	<i>au lieu de, en contrepartie, en compensation</i>	اعطاه عوض عن الاخر ¹¹² <i>Il le lui donna en compensation de l'autre</i>
بين	Signifie "intervening space", c'est-à-dire, "séparant deux espaces" : <i>entre</i> Si deux termes en annexion sont suivis de بين celui-ci est répété si les deux termes sont différents, en revanche si ce sont deux substantifs, il ne sera répété qu'une fois. On peut également rencontrer ما بين ou فيما بين	ألف الله بين قلوبكم ¹¹³ <i>God has united your hearts (Dieu a réuni vos coeurs)</i> بينني و بينه ¹¹⁴ <i>Between me and him (entre lui et moi)</i> اقتتلوا فيما بينهم ¹¹⁵ <i>They fought with each other (ils se sont entretués)</i>
تجاه	<i>en face de, en direction de</i>	اتجاه المطار ¹¹⁶

¹⁰⁸ Exemple créée par nous

¹⁰⁹ Ibidem

¹¹⁰ Ibidem

¹¹¹ W. WRIGHT, op. cit. p.186

¹¹² al-Munjid fî al-luga wal-i'lâm Dâr al-Mašraq, Beyrouth Liban, 1992, p. 634

¹¹³ Ibidem, p. 180

¹¹⁴ Ibidem, p. 180

¹¹⁵ Ibidem, p. 180

¹¹⁶ Exemple créé

		<i>En direction de l'aéroport</i>
فوق	<i>au-dessus de</i>	فوق المائدة ¹¹⁷ <i>Au-dessus de la table</i>
تحت	<i>sous, en dessous, bas</i>	كانت تحت فلان <i>She was under the authority of someone--for a married woman¹¹⁸--(elle était sous l'autorité de quelqu'un (pour une femme mariée))</i>
قبالة	<i>En face de, vis-a-vis de</i>	جلس قبالته ¹¹⁹ <i>Il s'assit en face de lui</i>
قبل	<i>avant</i>	قبل موته ¹²⁰ <i>Before his death (avant sa mort)</i>
حول	<i>autour de</i>	طاف حوله ¹²¹ <i>He went around it (the Temple) (Il fit le tour du Temple)</i>
حين	<i>lors, au moment de</i>	حين وصوله ¹²² <i>Au moment de son arrivée</i>
مثل	<i>comme</i> On trouve parfois المثل avec l'article	انت مثل اخي ¹²³ <i>Tu es comme mon frère (sur certains points).</i> ثمن المثل <i>The price paid for a similar thing¹²⁴ (le prix payé pour une chose similaire)</i>

¹¹⁷ Idem

¹¹⁸ W. WRIGHT, op. cit. p. 182

¹¹⁹ al-Munjid fî al-luġa wal-i'lâm op. cit., p. 608

¹²⁰ Ibidem, p. 186

¹²¹ Ibidem, p. 188

¹²² Exemple crée

¹²³ M. NEYRENEUF, op. cit. p. 119

¹²⁴ W. WRIGHT, op. cit. p. 210

جنب	à côté de	125 جنبه A côté de lui
نحو	Désigne 1. une direction utilisée à l'accusatif comme préposition signifie <i>vers, en direction de</i> . 2. au nominatif elle est employée comme un substantif ou adjectif pour signifier <i>comme, tel que</i> . 3. également une quantité approximative <i>environ</i> comme مقدار, elle est suivi de la préposition من	1. اتجه نحو الحديقة ¹²⁶ <i>Il s'est dirigé vers le jardin</i> 2. رجل نحو زيد <i>A man like Zeid</i> ¹²⁷ (un homme tel que Zeid) 3. والشمع نحو من مائة شمعة <i>The wax candles were about a hundred</i> ¹²⁸ . (il y avait environ cent bougies)
خارج	à l'extérieur	129 خارج الوطن A l'extérieur du pays
وراء	derrière	دخل عنتر وراءها <i>Antar came in behind her</i> ¹³⁰ (Antar entra derrière elle)
خلال	pendant, à travers	خلال الايام القادمة ¹³¹ Pendant les prochains jours
وسط	au milieu de, parmi	وسط روضة <i>In the midst of a garden</i> ¹³² (au milieu du jardin)
خلف	derrière, synonyme de وراء	والشعر يسحب خلف ظهرها

¹²⁵ Exemple créé

¹²⁶ Exemple créé

¹²⁷ W. WRIGHT, op. cit. p. 212

¹²⁸ Ibidem, p. 212

¹²⁹ Exemple créé

¹³⁰ Ibidem, p. 187

¹³¹ Exemple créé

¹³² Ibidem, p. 188

		<i>And her hair hung down to the ground behind her back</i> ¹³³ (Et ses cheveux tombaient par terre derrière son dos)
وفق	<i>conformément à</i>	¹³⁴ وفق اتفاقنا Conformément à notre accord
داخل	<i>à l'intérieur de</i>	¹³⁵ داخل السيارة A l'intérieur de la voiture
دون	Exprime plusieurs sens : 1. Indique qu'une personne ou une chose est <i>située à côté de</i> ou <i>en dessous de</i> . Cela est aussi valable pour une action qui peut avoir eu lieu <i>en dessous</i> ou <i>à côté</i> . 2. Indique qu'une chose ou une personne arrive <i>à côté</i> ou <i>approximativement</i> de l'autre qui est plus élevée. 3. Indique que pour un espace ou une distance, un nombre n'est pas égal ou ne complète pas un autre. 4. Indique qu'une chose est plus petite qu'une autre.	1. مصر فما دونها من البلاد <i>Egypt and the adjacent countries</i> ¹³⁶ (l'Egypte et les pays voisins) 2. هذا دون ذاك <i>This nearer than that</i> ¹³⁷ (ceci est plus près que l'autre) 3. و بينهما دون رمية حجر <i>And between them there was less than a stone's throw</i> ¹³⁸ (entre eux il y a moins d'écart qu'un jet de pierre) 4. النحاس دون الفضة و الحديد دون النحاس

¹³³ Ibidem, p. 187¹³⁴ Exemple créé¹³⁵ Exemple créé¹³⁶ Ibidem, p. 183¹³⁷ Ibidem, p. 183¹³⁸ Ibidem, p. 183

	<p>5. Indique qu'une qualité qui appartient à une personne ou une chose n'est pas possédée par une autre qui y en est l'opposée.</p> <p>6. Indique l'exclusion : <i>excepté, négligé</i>. Indique qu'une action est <i>renvoyée</i> à plus tard, ou que l'action affecte une personne ou une chose à un degré plus faible que l'autre.</p> <p>7. Indique qu'un objet est placé devant ou avant une autre</p>	<p><i>Copper is of less value than silver and iron of less value than copper</i> ¹³⁹ (le cuivre a moins de valeur que l'argent et le fer a moins de valeur que le cuivre).</p> <p>5. منهم الصالحون و دون ذلك <i>Some of them are good some of them are not so</i> ¹⁴⁰ (Certains d'entre eux sont bon d'autres ne le sont pas)</p> <p>6. يدعون من دون الله بما لا يضره و ما ينفعه <i>He invokes to the exclusion or neglect of God what can neither injure him nor benefit him</i> ¹⁴¹. (Il invoque excepté Dieu ce qui peut ni l'injurier ni l'avantager)</p> <p>7. ليس دونهم حجاب <i>There is no curtain or barrier in front of them</i> ¹⁴² (il n'y a ni rideau ni barrière devant eux)</p>
يمين	à droite de	يمين الشارع A droite de la rue ¹⁴³

¹³⁹ Ibidem, p. 183

¹⁴⁰ Ibidem, p. 184

¹⁴¹ Ibidem, p. 184

¹⁴² Ibidem, p. 185

¹⁴³ Exemple créé

شبه	Signifie <i>comme, quasi semblable à elle</i> peut être suivi de ب	<p>نحت الاولين في الجبل شبه درج يصعد عليه</p> <p><i>And the Ancients have cut out in the mountain like step (or a stair) by which one can ascend</i>¹⁴⁴. (Et les Ancients avaient sculpté dans la montagne des semblants de marches grâce auxquelles on pouvait descendre)</p>
غير	<p>Signifie <i>autre, autre que</i> ou encore à l'apposé de mais dans ce cas il correspond au préfixe <i>in, im, un</i>, comme impossible, improbable, etc.</p> <p>Quand il est à la forme accusative, il est souvent traduit par la préposition ou conjonction <i>excepté, sauf</i></p> <p>Précédé des prépositions ب ou من elle a le sens de <i>sans</i>.</p> <p>Pour éviter de répéter deux fois la préposition لا on la fait suivre de la particule négative</p>	<p>الملوك وغيرهم</p> <p><i>Kings and others</i>¹⁴⁵, Les rois et les autres</p> <p>قام القوم غير ابي بكر</p> <p><i>The poeple stood up except Abu Bakr</i>¹⁴⁶ (Tout le monde se leva excepté Abu Bakr)</p> <p>من غير خلاف</p> <p><i>Without controversy</i>¹⁴⁷ (Sans controverse)</p> <p>من غير اب و لا ام</p> <p><i>Without a father and a mother</i>¹⁴⁸ (sans père ni mère)</p>
سوى	Exprime l'exception : <i>excepté, sauf</i>	<p>وسواك مانع فضة المحتاج</p> <p><i>Whilst others than thou</i></p>

¹⁴⁴ Ibidem, p. 211

¹⁴⁵ Ibidem, p. 208

¹⁴⁶ Ibidem, p. 208

¹⁴⁷ Ibidem, p. 208

¹⁴⁸ Ibidem, p. 209

		<i>with hold their benefits from the needy</i> ¹⁴⁹ (Excepté toi, les autres qui privent les démunis de leurs avantages)
خلال	<i>au cours de, pendant, durant</i>	خلال ايام الدراسة ¹⁵⁰ <i>Durant les années d'études</i>
رب	<p>1. Elle signifie <i>beaucoup de, maint.</i> Elle est suivie d'un substantif quand il est annexé, d'un adjectif indéfini ou au même cas, ou d'un nom voire une phrase verbale.</p> <p>2. Elle peut signifie <i>plusieurs, beaucoup</i> avec une construction syntaxique en tamyiz</p> <p>3. elle peut avoir le sens de <i>peut-être</i></p>	<p>1. رب رجل كريم لقيت <i>Many a noble man have I met</i>¹⁵¹ (Que beaucoup d'hommes nobles ai-je rencontrés)</p> <p>2. ربهن نساء <i>Many women</i>¹⁵² (beaucoup de femmes)</p> <p>3. ربما زيد في الدار¹⁵³ Zeid est peut-être à la maison</p>

Il est à noter que la liste des quasi-prépositions est loin d'être exhaustive, nous nous sommes attachées à en étudier les plus usités.

¹⁴⁹ Ibidem, p. 209

¹⁵⁰ Exemple crée

¹⁵¹ W. WRIGHT, op. cit. 214

¹⁵² Ibidem, p. 215

¹⁵³ Exemple crée

Ces quasi-prépositions peuvent être précédées d'une préposition proprement dite (alors que celle-ci ne peut être précédée d'une autre préposition). Dans ce cas, elles se mettent au cas indirect :

Il sortit de chez le coiffeur : نخرج من عند الحلاق¹⁵⁴

Elle regarda derrière le rideau : نظرت من وراء الستار¹⁵⁵

Certaines peuvent aussi être utilisées comme des adverbes construits au cas sujet défini, sans complément, précédées ou non de la préposition من . C'est le cas notamment de : بعد/قبل/تحت/فوق/امام/وراء :

Il est venu auparavant, il est déjà venu : قد جاء من قبل¹⁵⁶

3. Lexique des prépositions incluses dans le dictionnaire des mots vides

Nous avons remarqué que ces prépositions et quasi-prépositions étaient en rapport d'annexion avec les termes qui suivent. D'autant plus que les quasi-prépositions sont généralement des noms. Elles sont construites sur une même structure : Préposition ou quasi-préposition suivi d'un nom.

Dans le cas des annexions nominales, nous avons vu une structure composée de deux termes nominaux. Nous avons remarqué qu'en réalité la structure sous-jacente contenait implicitement une préposition. Une construction avec l'une de ces prépositions a un sens particulier, d'où l'utilité d'un dictionnaire de mots vides.

*“Les mots vides sont des mots qui ne sont pas utilisés pour constituer le fichier inverse d'une référothèque. Ce sont essentiellement les articles et prépositions, les mots outils et locutions outils (en particuliers les locutions comprenant des articles, des prépositions, des conjonctions, des pronoms : par rapport à, de l', à partir de, etc.). L'ensemble des mots vides constitue l'antidictionnaire.”*¹⁵⁷ C'est pourquoi nous avons élaboré le lexique suivant :

¹⁵⁴ M. NEYRENEUF, op. cit. p. 119

¹⁵⁵ Ibidem. p.119

¹⁵⁶ Ibidem

¹⁵⁷ A. DEWEZE, op. cit. p.119

PREPOSITION	REGLES	EXEMPLES
من	<p>- Partitif et distinctif</p> <p>- Elle marque l'origine dans l'espace</p> <p>- Elle marque aussi l'origine dans le temps</p> <p>- Elle peut régir un nom indéterminé dans le cas où elle est employée comme explétive avec une négation</p>	<p>¹⁵⁸ و من الناس من يقول امنا با الله</p> <p><i>Et parmi les gens, il y a ceux qui crurent en Dieu</i></p> <p>سبحان الذي اسرى بعبده ليلا من</p> <p>¹⁵⁹ المسجد الحرام الى مسجد الاقصى</p> <p><i>Gloire à celui qui a fait voyager son serviteur la nuit depuis la mosquée de Médine jusqu'à celle de Jérusalem.</i></p> <p>لمسجد اسس على التقوى من أول يوم</p> <p>احق أن تقوم فيه ¹⁶⁰</p> <p><i>Car une mosquée a été fondée sur la piété dès le premier jour, (depuis) il est plus digne d'y prier.</i></p> <p>ما لباغ من مفر</p> <p><i>Point de refuge pour un méchant</i>¹⁶¹</p>

¹⁵⁸ Šarḥ ibn 'aḳīl, 3^e partie, Dar al-fikr p. 15

¹⁵⁹ Šarḥ ibn 'aḳīl, 3^e partie, Dar al-fikr p. 15

¹⁶⁰ Šarḥ ibn 'aḳīl, 3^e partie, Dar al-fikr p. 15 - Coran, Sourate *At-tawbat*, verset 108

إلى	Exprime la <i>finalité</i> , le <i>but</i> (lieu) <i>pour</i> , à	¹⁶² إليك <i>Pour toi</i>
حتى	Exprime la <i>finalité jusqu'à</i> , est suivie d'un substantif	¹⁶³ حتى اليوم Jusqu'à aujourd'hui
في	Exprime l'état limitatif <i>dans</i> , <i>au moyen</i> <i>de</i> et aussi la cause	نحن أقارب في أقارب <i>Our relations are those of</i> <i>kinsfolk</i> ¹⁶⁴ (Nous sommes entre gens de même famille)
عن	Elle est très souvent utilisée dans le sens de بعد pour exprimer le dépassement	رمى السهم عن القوس (معنى "بعد") ¹⁶⁵ <i>J'ai jeté la flèche de l'arc</i>
على	Exprime la supériorité - sens de في et de عن - sens de dépassement avec عن qui se trouve pour بعد - très souvent utilisé dans le sens de في	زيد على السطح ¹⁶⁶ <i>Zayd est sur le toit</i> و دخل المدينة على حين غفلة من أهلها (أي في حين غفلة) ¹⁶⁷ <i>Il est entré dans la ville à</i> <i>l'insu de ses habitants</i>

¹⁶¹ *La Alfiyya*, p. 90¹⁶² Exemple créé¹⁶³ Exemple créé¹⁶⁴ Hans Wehr, *A dictionary of modern written arabic, Arabic-English*, edited by J. Milton Conan, Otto Harrassowitz, Wiesbaden Librairie du Liban 1980, p. 734¹⁶⁵ *Šarḥ ibn ʿaql*, 3^e partie, Dar al-fikr p. 23¹⁶⁶ *Ibidem*

ل	<p>1- Sert à marquer la propriété.</p> <p>2- Elle exprime :</p> <ul style="list-style-type: none"> - la finalité, le but - la possession - l'appartenance <p>3- Elle sert de préposition à un verbe intransitif</p>	<p>1-</p> <p style="text-align: right;">المال لزيد¹⁶⁸</p> <p><i>L'argent est à Zayd</i></p> <p style="text-align: right;">الباب للدار¹⁶⁹</p> <p>2-</p> <p><i>"La porte est à la maison"</i></p> <p>c'est-à-dire la porte de la maison</p> <p>3-</p> <p style="text-align: right;">جئتك لآكرامك¹⁷⁰</p> <p><i>Je suis venu te remercier</i></p>
ك	<p>-elle sert pour la comparaison <i>comme</i></p> <p>- <i>en tant que, en qualité de</i></p> <p>- elle sert également pour la corroboration</p>	<p style="text-align: right;">زيد كالأسد¹⁷¹</p> <p><i>Zayd est comme un lion</i></p>

¹⁶⁷ Ibidem, p.20

¹⁶⁸ Ibidem

¹⁶⁹ Ibidem

¹⁷⁰ Ibidem

¹⁷¹ Ibidem, p.25

ب	<p>Elle exprime le moyen la transitivité, la compensation, la contiguïté, (elle peut être employée comme synonyme de من/عن/مع</p> <p>- prépositionnel, elle a le sens de في <i>dans</i></p> <p>-l'échange</p> <p>- elle a le sens de "من" (source de quelque chose, point de départ)</p> <p>-elle signifie avec, au moyen de (instrument)</p>	<p>كان بالمدينة تاجر</p> <p><i>Il y avait dans la ville un marchand</i>¹⁷²</p> <p>اشترت الفرس بالف درهم¹⁷³</p> <p><i>- J'ai acheté un cheval pour mille dinars</i></p> <p>اشترت بماء البحر (من ماء البحر)¹⁷⁴</p> <p><i>- J'ai bu de l'eau de mer</i></p> <p>قتله بالسيف</p> <p><i>Il le tua avec le sabre</i>¹⁷⁵</p>
---	---	--

¹⁷² R. BLACHERE, op. cit. p. 330

¹⁷³ *Šarḥ ibn ʿaql*, 3^e partie, Dar al-fikr p. 22

¹⁷⁴ *Šarḥ ibn ʿaql*, 3^e partie, Dar al-fikr p. 22

¹⁷⁵ R. BLACHERE, op. cit. p. 333

La liste des prépositions est loin d'être exhaustive, nous nous sommes attachées à en représenter les plus importantes.

C- L'annexion à un terme issu d'un déverbal

1- le participe actif (اسم الفاعل)

L'annexion à un participe actif est possible dans deux cas :

1- lorsque la valeur nominale du participe actif prévaut, on a une annexion, et cela dans deux possibilités :

- premièrement, quand le complexe formé par le participe actif et son complément est le sujet de la phrase comme dans l'exemple suivant :

عائد المريض يحوض في الرحمة

*Le visiteur du malade baignera dans la miséricorde*¹⁷⁶ (hadit)

- deuxièmement, quand le participe actif est au masculin pluriel avec l'apocopé du ن :

ما نحن بتاركي أهتنا

Et nous ne serons pas des "négligeurs" de nos divinités (Coran XI, 53)¹⁷⁷

2- Quand le participe actif a un complément d'objet celui-ci peut avoir comme fonction syntaxique soit l'accusatif soit le génitif (marque de l'annexion) :

(état d'annexion)

انك جامع الناس

¹⁷⁸ *Tu es le "collecteur" des hommes* (Coran III, 9)

(tournure semi-verbale)

انك جامع الناس

¹⁷⁹ *Tu (es) celui qui réunira les hommes*

¹⁷⁶ Ibidem p. 325

¹⁷⁷ Ibidem

¹⁷⁸ Ibidem

¹⁷⁹ Ibidem

Le participe actif d'un verbe transitif peut être construit soit à l'accusatif soit au génitif en ayant le sens de l'inaccompli, du présent de narration ou du futur.

Car dans le cas d'une annexion "non proprement dite" (غير الحقيقية), le premier terme de l'annexion (le nom de procès) pourra prendre l'article et l'annexion aura le sens de l'inachevé :

exemple :

¹⁸⁰ *One who kills people* (tournure annective) قاتلُ الناسِ

¹⁸¹ *One who kills people* (tournure semi-verbale) قاتلُ الناسِ

(Quelqu'un qui tue les gens)

ou bien

¹⁸² *he who kills people* (tournure annective) القاتلُ الناسِ

¹⁸³ *he who kills people* (tournure semi-verbale) القاتلُ الناسِ

(Celui qui tue les gens)

Ce dernier exemple (القاتلُ الناسِ) équivaut en fait à la tournure الذي يقتل الناسِ

Celui qui tue les gens.

En revanche, si le participe actif d'un verbe transitif a le sens de l'accompli, il est plus proche de la nature d'un nom, dans ce cas sa construction ne peut être qu'une annexion. Ainsi si cette construction du génitif est vraie, le participe actif ne pourra pas avoir l'article on dira : قاتلُ الناسِ et non القاتلُ الناسِ ou القاتلُ الناسِ .

Lorsque le participe actif est indéfini, il se construit à l'accusatif avec le nom qui suit seulement dans le cas suivant :

- quand il est attribut ou prédicat d'un sujet (qui le précède) pour exprimer un état ou une condition :

¹⁸⁰ W. WRIGHT, op. cit. p. 64

¹⁸¹ Ibidem.

¹⁸² Ibidem

¹⁸³ Ibidem

¹⁸⁴*Zeid is beating (or will beat) Amr* زيد ضارب عمرا

(Zayd est en train de battre (ou va battre) °Amr)

- après une particule négative ou interrogative quand il est attribut d'un sujet qui le précède ou le suit :

¹⁸⁵*Thou dost not follow their kibla* ما انت بتابع قبايتهم

Tu ne suis pas leur kibla (orientation du corps lors de la prière)

- après une interjection comme le prédicat d'un sujet supprimé :

¹⁸⁶*O (tho who art) climbing a hill* يا طالعا جبلا (أي يا رجلا طالعا جبلا)

(*O toi qui monte la colline*)

Le participe actif au singulier quand il est suivi d'un substantif au génitif peut avoir l'article dans le cas seulement ou celui-ci est lui-même défini :

¹⁸⁷*He who beats the slave* الضارب العبد
(Celui qui frappe l'esclave)

2- le participe passif (اسم المفعول)

Le participe passif de tous les verbes transitifs de la forme simple est construit sur un même schème مفعول. Ce participe désigne soit celui qui subit l'action soit parfois le résultat de l'action : جمع كمجموع *rassembler réunir* (action) , *l'ensemble de* (résultat de l'action).

Employé comme épithète le participe passif d'un verbe transitif indirect s'accorde en cas et en définition avec le nom auquel il se rapporte, mais reste toujours au masculin singulier. Il est suivi de préposition (propre à la construction du verbe) à laquelle on accroche un pronom suffixe, il s'accorde en genre et en nombre avec le nom :

¹⁸⁴ Ibidem, p. 65

¹⁸⁵ Ibidem, p. 65

¹⁸⁶ Ibidem, p. 66

¹⁸⁷ Ibidem, p. 66

¹⁸⁸*La femme condamnée est entrée*

دخلت المرأة المحكوم عليها

¹⁸⁹*La marchandise désirée est arrivée*

وصلت البضاعة المرغوب فيها

Le participe passif semble être employé comme un adjectif, d'un point de vue annectif, il ne pourra être utilisé que dans le cas d'une annexion qualificative.

3.-le nom de procès (masdar)

Chaque verbe a un ou plusieurs nom d'action, de qualité ou d'état, formé sur la même racine et signifiant "le fait de..." Pour les verbes simples, il existe de nombreux schèmes de formation du nom de procès. Il faudra donc les mémoriser en même temps que les voyelles de la radicale médiane du verbe. En revanche, chaque forme dérivée a un ou deux schèmes précis sur lesquels se construisent le nom d'action.

Il sera tantôt utilisé comme substantif et tantôt avec un sens verbal clair (il reste alors au singulier). Il pourra avoir des compléments ou remplacer une construction verbale. Voici quelques exemples :

le fait de sortir, sortie خروج (sortir) خرج

Le nom de procès ou masdar est très souvent utilisé comme substantif dans une annexion comme dans l'exemple suivant :

¹⁹⁰*Le renouvellement de la carte d'identité*

تجديد بطاقة التعريف

Valeur nominale d'un nom de procès quand il est le premier d'une annexion :

Tu as multiplié (l'action de) frapper la porte

أكثرت دق الباب

Tu as fréquemment frappé à la porte ¹⁹¹

Cependant le nom de procès a une valeur verbale et un complément au cas direct, quand ce complément vient après une annexion ou quand ce complément ne suit pas immédiatement le nom de procès :

¹⁸⁸ M. NEYRENEUF, op. cit. p. 143

¹⁸⁹ M. NEYRENEUF, op. cit. p. 143

¹⁹⁰ Exemple créé

¹⁹¹ BLACHERE, op. cit. p. 326

انكرت لبسك هذا الكساء

Je blâme ton port de ce vêtement (que tu portes ce vêtement) ¹⁹²

Dans l'annexion à un déverbal, le second terme à une valeur de "complément d'objet" du procès par le déverbal :

W. Wright indique en ce sens la règle suivante :

If the nomen agentis be derived from a verb which governs two or three objective complements it takes the first either in the accusative (which is by far the more usual) or in the genitive and the others in the accusative :

¹⁹³ *I will dress Zeid in a splendid robe* انا كاس زيدياً ثوباً فاخراً

Si le nom de procès est dérivé d'un verbe qui comporte deux ou trois compléments d'objet, le premier est soit à l'accusatif (ce qui est de loin le plus) ou au génitif et les autres à l'accusatif.

J'habillerai Zeid d'un splendide vêtement انا كاس زيدياً ثوباً فاخراً

Quand le nom de procès d'un verbe est doublement transitif, c'est-à-dire, lorsque celui-ci comporte deux compléments d'objet, ces derniers peuvent être suffixés au nom de procès :

¹⁹⁴ *He who gives it to you* معطيكه

Il te l'a donné

¹⁹⁵ *He who gives me it to eat* مطعمنيها

Il me l'a donné à manger

W. Wright précise également concernant le double complément d'un nom de procès :

¹⁹² Ibidem, p. 326

¹⁹³ W. WRIGHT, op. cit. 67

¹⁹⁴ Ibidem, p. 67

¹⁹⁵ Ibidem, p. 67

The second of the two complements of nomen agentis or that which is in the accusative is very rarely inserted between the nomen agentis and the first complement or that in the genitive :

وسواك مانع فضله المحتاج

*whilst others than thou withhold their benefits from the needy*¹⁹⁶

Autre remarque : le second des deux compléments d'un nom de procès, ou celui qui est à la forme accusatif, est très rarement inséré entre le nom de procès et le premier complément qui est au génitif :

وسواك مانع فضله المحتاج

Mis à part toi, ceux qui mettent leurs richesses loin des nécessiteux.

Nous venons de voir l'annexion de termes nominaux et déverbaux. Il existe également d'autres structures annectives d'un degré supérieur au mot. Nous avons vu l'annexion d'un terme déverbal et nous avons constaté que malgré sa nominalisation, il garde sa valeur et les règles syntaxiques qui régissent sa construction avec les compléments et notamment avec le nom de procès.

Nous allons à présent étudier d'autres structures annectives qui se situent à un niveau supérieur au mot, c'est l'annexion de terme à une phrase.

III- D'autres structures annectives

L'annexion concerne certes les substantifs, mais elle peut également relier un terme à une phrase comme nous allons le voir dans cette partie.

1- L'annexion à une phrase

Certaines particules peuvent s'annexer à un terme d'une phrase comme : منذ / منذ

Elle exprime le temps et est suivie d'un substantif

- elle peut avoir le sens de "في"

¹⁹⁶ Ibidem, p. 67

ما رأيته منذ يوم الجمعة¹⁹⁷

Je ne l'ai pas vu depuis vendredi

ما رأيته منذ يومنا (اي في يومنا)¹⁹⁸

Je ne l'ai pas vu depuis notre journée

- صمت يوم الجمعة و صليت يوم الخميس¹⁹⁹

J'ai jeuné vendredi et j'ai prié jeudi

- D'autres particules peuvent s'annexer à une phrase :

- une phrase nominale avec حيث

- اجلس حيث زيد جالس²⁰⁰

Je m'assieds où Zayd s'assied

- à une phrase verbale

- اجاس حيث جلس زيد (يجاس زيد)²⁰¹

Je m'assieds où s'assied Zayd

- اذا à une phrase nominale

- جئتك اذا زيد قائم²⁰²

Je suis venu à toi quand Zayd s'est levé

- à une phrase verbale

- جئتك اذا قام زيد²⁰³

¹⁹⁷ Šarḥ ibn ʿaql, 3^e partie, Dar al-fikr p. 54

¹⁹⁸ Ibidem, p. 57

¹⁹⁹ Ibidem, p.12

²⁰⁰ Šarḥ ibn ʿaql, 2^e partie, Dar al-fikr p. 44

²⁰¹ Ibidem, p. 44

²⁰² Ibidem, p. 44

²⁰³ Ibidem, p. 44

– "ما" et "إذا" ne peuvent être annexés qu'à une phrase verbale :

آتيتك اذا قائم²⁰⁴

Je viens chez toi s'il est levé

"ما" et "إذا" ne peuvent être annexés à une phrase nominale, ainsi on ne peut pas dire :

* آتيتك اذا زيد قائم²⁰⁵

Nous venons de voir l'annexion à une phrase par le biais de certaines particules (اذا حيث). Il existe d'autres structures annectives complexes, nous nous proposons ici de faire l'étude de l'ellipse de l'un des termes de l'annexion nominale.

2- L'ellipse de l'un des termes de l'annexion

2.1. L'ellipse du premier terme :

- Le premier terme de l'annexion, l'annexant ou le mudāf, peut être ellipsé, sa fonction syntaxique est reprise par le deuxième terme de l'annexion, l'annexée ou le mudāf ilayhi :

exemple :

اشربوا في قلوبهم العجل بكفرهم (اي حب العجل)²⁰⁶

Par impiété, ils remplirent leur coeur d'amour pour le Veau

و جاء ربك (أي امر ربك)²⁰⁷

Ton Seigneur est venu (=l'Acte de ton Seigneur)

²⁰⁴ Šarḥ ibn 'aql, 2^e partie, Dar al-fikr p. 44

²⁰⁵ Ibidem, p46

²⁰⁶ Ibidem, p.47 (extrait du Coran, Sourate *La Vache*, Verset 93)

²⁰⁷ Ibidem, p.48

On voit dans ces deux exemples l'ellipse de (حب) et de (امر) et chacun de ces deux termes a pris la fonction syntaxique de l'annexant.

- L'annexant peut être ellipsé mais l'annexé garde sa fonction syntaxique, c'est-à-dire, la marque de l'annexé la voyelle "i" pour rappeler la présence de l'annexant à condition que le terme éllipsé représente ce terme auquel il est attaché.

exemple :

أكل أمرئ تحسين أمرا

و نار توقد بالليل نارا²⁰⁸

Crois-tu que chaque être humain est un donneur d'ordre

Et chaque flamme qui s'embrase dans la nuit (est) un feu ?

Dans cet exemple, le terme ellipsé est dans le deuxième vers pour "كل نار" chaque flamme ou feu, le "كل" est ellipsé et il ne reste que le terme "نار" au génitif. La condition est ici présente puisque le terme ellipsé est présent dans le début du premier vers puisqu'on le trouve dans "كل أمرئ".

- L'annexant peut être ellipsé mais le terme ellipsé n'est pas le terme présent au début mais un terme équivalent :

exemple :

تريدون عرض الدنيا و الله يريد الآخرة²⁰⁹

Vous voulez les richesses d'ici bas mais Dieu veut ceux de l'au-delà

On peut lire "الله يريد باقي الآخرة" ce qui équivaut à "الله يريد الآخرة" soit "Dieu souhaite

les richesses de l'au-delà" Ibn 'Aqîl propose une autre lecture :

"الله يريد عرض الآخرة". Dans ce cas, le terme ellipsé serait le l'annexant de la

première partie de l'exemple, cette version est celle de Ibn Abi Rabi'a, et repris par Ibn 'Aqîl pour exemple.

²⁰⁸ *Šarḥ ibn 'aql*, 2^e partie, Dar al-fikr p.77

²⁰⁹ *Ibidem*, p.78

2.2.L'ellipse du second terme de l'annexion

Lorsque le deuxième terme de l'annexion est ellipsé, il ne reste que la tête de syntagme, c'est-à-dire, le premier terme. Sa fonction syntaxique est la même, il ne garde pas son tanwîn. Le terme ellipsé est toujours l'équivalent d'un terme déjà évoqué dans une première annexion, il n'est pas repris une deuxième fois pour éviter une répétition.

exemple :

قطع الله يد ورجل من قالها (اي قطع الله يد من قالها ورجل من قالها)²¹⁰

Dieu a coupé le pied et la main de celui qui l'a prononcé

On peut ellipser un terme et coordonner deux termes avec un terme ellipsé seulement dans le cas où il est équivalent au terme ellipsé.

3- La composition nominale en arabe

La composition nominale en arabe est l'un des procédés morphosyntaxique de la création lexicale .

J. Ghazi distingue la composition conjointe de la composition disjointe par "(...) *la présence et le rapprochement de deux unités autonomes, après suppression de radicales finales du premier composant et initiale du second composant : afro-asiatique* افروآسيائوي *et indo-européen* هندو-اوروبسي

"²¹¹اوروبسي". L'exemple donné par J. Ghazi ressemble plus à un calque linguistique français. En arabe classique ou moderne nous dirions par exemple pour *indo-européen* هندي اوروبسي soit les deux adjectifs l'un derrière l'autre.

C'est le modèle savant de la composition française. On peut également trouver des termes composés avec la particule de négation لا comme dans لا شيء l'infini ou encore لا نهاية le néant.

²¹⁰ Šarḥ ibn 'aql, 2^e partie, Dar al-fikr p. 79

²¹¹ J. GHAZI, *Propositions pour une typologie nouvelle de la création lexicale*, ARABICA, Revue d'Etudes Arabes tome XXXIV, mars 1987, p.156

Quant à la composition disjointe, elle consiste à mettre côte à côte deux termes nominaux autonomes et à lexicaliser la formation ainsi obtenue. Ces termes sont essentiellement des adjectifs et des substantifs par exemple des noms de régions ou de ville comme *حضرموت* Hadramaout ou *بيت لحم* Betléem. Il peut également s'agir de nom de personne comme *عبدالله* ou encore *نورالدين*. La composition nominale concerne aussi les nombres ordinaux et cardinaux de 11 à 19 : *ثلاثة عشر* (13) ou *تسعة عشر* (19).

Il est à remarquer que la langue arabe pour exprimer le terme *capital* associe deux termes *رأس* et *المال*. Il en résulte deux écritures, on peut trouver *رأس المال* ou *الرأسمال*. La fusion sera de règle pour l'adjectif ainsi que pour le substantif *الرأسمالية* *le capitalisme*.

Il existe également en arabe des expressions lexicalisées bâties sur une structure complexe. En effet, si chaque terme d'une expression lexicalisée est pris isolément, il n'aura pas un sens identique à celui de l'ensemble des termes de l'expression.

Lorsqu'une locution est prise dans une acceptation spécifique, on l'appelle phrasème²¹². Ces expressions lexicalisées ou phrasèmes ont un signifié spécifique :

Il existe trois types de phrasème, on peut représenter le phrasème mathématiquement par la formule suivante $s(a,b) \neq s(a) + s(b)$ ²¹³ :

- Phrasème complet : AB au signifié C tel que ces signifiés n'incluent ni le signifié de A ni celui de B,

Exemple : le signifié de *رجال السند*²¹⁴ est *source d'information*

²¹² S. MINKO-MINSERVE, *Les phrasèmes dans le sens de Melcu 'k en arabe*, Mémoire de DEA LTMT, Université Lumière Lyon2, 1996

²¹³ s=signifié ; a=terme ; b=terme

²¹⁴ Ibidem, p.47

Ce sens n'inclut ni le sens de رجال *hommes*, ni celui de السند, *appui*

- Quasi-phasème : phrasème AB signifié ABC incluant les signifiés des deux constituants et un sens imprévisible,

Exemple : le signifié de رئيس الوزراء²¹⁵ est *premier ministre*, inclut les signifiés de رئيس *chef, président* et de الوزراء *ministres* avec en plus un sens imprévisible celui de *premier ministre*, c'est-à-dire, le chef et le premier de tous les ministres.

- Semi-phasème : phrasème AB au signifié AC ou BC incluant signifié de l'un des constituants.

Exemple : le signifié de رأس المال²¹⁶ est *capital*, il n'inclut que le sens de l'un de ses éléments المال *la richesse* mais pas celui de رأس qui signifie *tête*.

Nous avons élargi dans cette partie notre étude des structures annectives à un degré supérieur des annexions de termes, celle de l'annexion de phrase par le biais de particule, l'ellipse de l'un des termes d'une annexion et enfin la composition nominale. Nous projetons de réaliser dans la partie qui va suivre une grammaire en vue d'un double traitement des annexions morphosyntaxique d'une part et syntactico-sémantique d'autre part.

IV-Modèle de structures annectives

1-Règles d'écriture de grammaire du syntagme nominal

Le documentaliste cherche des concepts clés qu'il traduit en descripteur générique ou spécifique. La particularité de l'indexation automatique est de repérer les SN. Pour savoir si un élément du discours est porteur de sens, le documentaliste fait référence à ses connaissances antérieures, la machine ne peut

²¹⁵ Ibidem, p.41

avoir le vécu d'un être humain. Aussi est-il nécessaire dans le cadre du traitement automatique des langues appliquées à la documentation de connaître la structuration syntaxique de ces descripteurs complexes et leur portée sémantique.

Nous allons déterminer pour chaque mot son appartenance à une catégorie, c'est-à-dire, son trait morphologique. Pour cela nous allons construire un lexique dans lequel devra se trouver la catégorie et la sous catégorie de chaque terme utilisé. Ces catégories syntaxiques²¹⁷ sont empruntées aux travaux du groupe SYDO, nous y avons apporté quelques modifications pour les besoins d'adaptation à la langue arabe :

Catégorie F (substantif et adjectif)

NOM: mot utilisable en place de nom (N, N')

N-COM : nom commun

N-PRP : nom propre

N-PRO : pronom (elle, moi, lui, eux, ...)

ADJ : mot utilisable en place d'adjectif (A, A', A'')

NAN : mot utilisable en place d'adjectif ou de nom (grand, noir, joli,)

DEV : déverbal

Catégorie D (prédéterminant ou quantificateur)

Det : Article (le, les,...)

Ø : article ellidé

Pré-Det : Démonstratif (ce, celle, celui-ci, ...)

Num : numéraux cardinaux (deux, trois, ...), numéraux ordinaux (premier, second, ...)

QUA : Quantificateur (assez, beaucoup, pas, peu, ...)

Catégorie Prep (prépositions) (à, de, devant, malgré, avec, par, ...)

Catégorie W (adverbe) :

Catégorie C (conjonction de coordination) : et, ou, ni, mais, ...

Catégorie V (verbe)

²¹⁶ Ibidem, p.40

²¹⁷ O. LAROUK, *Extraction de connaissances à partir de documents textuels, traitement automatique de la coordination (connecteurs et ponctuation)* thèse de doctorat en Informatique, Université Claude Bernard Lyon1, 1993, chap, 3, p.22

Catégorie Y (particule préverbale ppv)

Catégorie Q (conjonction de subordination) : que, si, quoique, parce que, ...

Catégorie T (ponctuation)

Ces catégories sont principalement utilisées lors de l'analyse morphosyntaxique.

2- L'essai d'une grammaire pour le traitement morphosyntaxique du syntagme annexional

Dans la phrase arabe, on a trois cas de déclinaison correspondant à des structures syntaxiques :

nominatif (ـ) cas sujet, apposition

accusatif (ـ) cas complément direct

génitif (ـ) cas complément indirect, complément de nom en cas d'annexion, complément prépositionnel.

Pour notre étude, le cas grammaticale qui nous intéresse, est le génitif.

Nous envisageons dans un premier temps un traitement morphosyntaxique de ces structures pour faire resurgir les éléments morphosyntaxiques appartenant à la structure profonde de ces annexions. Ce traitement constitue un prétraitement au traitement sémantique. De plus, le traitement morphosyntaxique permettra la préparation d'une interface de saisie. Nous envisageons ensuite un traitement syntactico-sémantique de ces structures annexives dans l'optique d'une indexation automatique. Ceci permettrait une meilleure indexation de ces termes et donc une plus grande chance d'apparition de ces descripteurs lors de l'interrogation.

2.1.L'annexion nominale

2.1.1- La détermination de l'annexion nominale

a-L'annexion "pure"

Dans le cas d'une annexion déterminée comme pour علم البلد , on a la règle suivante :

$N^2 = N + N'$ علم البلد

$N^2 = \text{Det} + N$ البلد

Dans le cas d'une annexion indéterminée comme pour علم بلد ou l'article est ellidé du second terme, la règle est la suivante :

$$N''=N+N' \text{ علم بلد}$$

$$N'=Ø+N \text{ بلد}$$

Lorsqu'on a une annexion de trois termes comme dans باب غرفة النوم la règle est :

$$N''=N+N' \text{ باب غرفة النوم}$$

$$N'=N+N' \text{ غرفة النوم}$$

$$N=\text{Det}+N \text{ النوم}$$

b- L'annexion formative ou qualificative

La règle de réécriture du SN dans le cas d'une annexion formative est la suivante :

$$N''=N+N'' \text{ رجل طويل القامة}$$

$$N''=A+N' \text{ طويل القامة}$$

$$N'=\text{Det}+N \text{ القامة}$$

L'exemple suivant est une variante de la règle de réécriture d'une annexion formative. Il diffère uniquement au niveau de la détermination du premier terme qui porte l'article :

$$N''=N+N'' \text{ الفيلسوف الغريب الافكار}$$

$$N''=A'+N' \text{ الغريب الافكار}$$

$$A=\text{Det}+A \text{ الغريب}$$

$$N'=\text{Det}+N \text{ الافكار}$$

2.1.2. L'appartenance / la possession

Nous avons ici l'annexion de deux nominaux dont l'un est nom propre, la règle s'écrit de la façon suivante :

$$N''=N+SP \text{ مال لزيد}$$

$$SP=P+N-PRP \text{ لزيد}$$

$$N-PRP= \text{زيد}$$

Ici nous avons également l'annexion de deux termes nominaux, ce sont souvent des noms communs :

$$N''=N+SP \text{ غلام للرجل}$$

$$SP=P+N' \text{ للرجل}$$

$$N^1 = \text{Det} + N \text{ الرجل}$$

Nous avons ici une annexion à trois termes dont une ellipse pour éviter une répétition, nous avons سلطان البر و البحر au lieu de سلطان البحر : la coordination permet de retrouver le sens originel :

$$N^{2'} = N^{2''} + C + N^2 \text{ سلطان البر و البحر}$$

$$N^{2''} = N + N^2 \text{ سلطان البر}$$

$$N^2 = \text{Det} + N \text{ البر}$$

$$C = \text{و}$$

$$N^1 = \text{Det} + N \text{ البحر}$$

2.1.3. La forme et la matière

La règle de réécriture pour cette catégorie est la suivante :

$$N^{2'} = N + N^2 \text{ خاتم من فضة}$$

$$N^2 = \text{SP} + N \text{ من فضة}$$

$$N = \emptyset + N \text{ فضة}$$

2.1.4. La contenance

La règle est ici différente des précédentes du fait de la portée sémantique de la contenance. Celle-ci exige en arabe une répétition par le pronom complément du contenant. La règle est donc la suivante :

$$N^{2'} = N + N^2 \text{ فنجان قهوة}$$

$$N^2 = \text{SP} + N \text{ فيه قهوة}$$

$$\text{SP} = \text{P} + \text{N} - \text{PRO} \text{ فيه}$$

$$N = \emptyset + N \text{ قهوة}$$

2.1.5. La reformulation spécifique

Ici la règle d'écriture du SN est très simple (deux nominaux) :

$$N^{2'} = N + N^2 \text{ شهر رمضان}$$

$$N^2 = \emptyset + N \text{ رمضان}$$

2.1.6. Le tamyiz

La règle de réécriture est ici la suivante :

$$N^{2'} = N + N_{(\text{cas direct})} \text{ شير ارضنا}$$

$N = \emptyset + N$ ارض

2.1.7.L'élatif

La règle de réécriture dans le cas de l'élatif est identique à l'annexion qualificative :

$N'' = N + N''$ هو افضل رجل من الرجال

$N'' = N' + SP$ افضل رجل من الرجال

$N' = A + N'$ افضل رجل

$N' = \emptyset + N$ رجل

$SP = P + N'$ من الرجال

$N' = Det + N$ الرجال

2.2.L'annexion à un modificateur

2.2.1.L'annexion à un numéral

La règle de réécriture du SN dans le cas d'une annexion à un numéral est la suivante :

$N'' = Num + N''$ خمسة الاثواب

$N' = Det + N$ الاثواب

Nous avons ici une autre variante mais la règle est la même :

$N'' = Num + N$ الثلاثة عشر رجلا

$N' = \emptyset + N$ رجل

2.2.2. L'annexion à un modificateur ou à un quantificateur

La règle est ici identique à l'annexion précédente, cependant nous avons un quantificateur à la place du numéral:

$N'' = QUA + N''$ كل الشعب

$N' = Det + N$ الشعب

Nous avons, dans l'exemple suivant, une règle identique à la précédente avec pour variante une inversion du quantificateur par rapport au nom quantifié. C'est en quelque sorte un effet de style :

$N'' = N' + N''$ كله الشعب

$N' = Det + N$ الشعب

$N' = QUA + NPRO$ كله

2.2.3. Les prépositions

Un SN annexionnel peut inclure un SN prépositionnel qui en arabe est en rapport d'annexion avec le nom qui suit la préposition :

$N'' = N' + SP$ الولد مع الرجل

$SP = Prep + N'$ مع الرجل

$N' = Det + N$ الرجل

2.3.L'annexion à un terme issu d'un déverbal

Ici le déverbal nominalisé agit comme un nom :

2.3.1.Le participe actif

$N'' = Dev + N'$ قاتل الناس

$N' = Det + N$ الناس

2.3.2.Le participe passif

$N'' = N'' + SP$ المرأة المحكوم عليها

$N'' = N' + N'$ المرأة المحكوم

$N' = Det + N$ المرأة

$N' = Det + Dev$ المحكوم

$SP = P + NPRO$ عليها

2.3.3.Le nom de procès

$N'' = Dev + N''$ تحديد بطاقة التعريف

$N'' = N + N'$ بطاقة التعريف

$N' = Det + N$ التعريف

Nous venons de voir le traitement morphosyntaxique du SN annexionnel pour les annexions nominales, déverbale et celles incluant les modificateurs. Ceci nous a permis de procéder aux règles de réécriture du SN pour le traitement informatique des descripteurs bâtis sur ce modèle. Nous allons procéder dans la partie qui va suivre au traitement syntaxique et sémantique de ces mêmes SN annexionnels .

3-L'essai d'une grammaire intégrant le traitement sémantique du syntagme nominal annexionnel

Nous allons représenter sous une forme schématique le traitement syntaxique et sémantique de ces annexions. Le traitement morphosyntaxique nous a permis

de repérer les constituants des SN construits sur ce modèle. Nous allons maintenant voir la nature sémantique des constituants du SN annexionnel.

Nous nous sommes inspirées pour ce traitement des grammaires d'unification, qui font le lien entre la grammaire et le lexique. Elles (...) *proposent des inventaires de traits morphosyntaxiques et sémantiques validant les relations à divers niveaux entre les éléments de la phrase*²¹⁸. Elles ont "(...) *pour principe de base l'utilisation d'un même formalisme pour représenter les éléments du dictionnaire, les règles de grammaire et les structures internes des phrases*"²¹⁹.

Nous avons représenté les relations de dépendances syntaxiques et sémantiques des constituants d'un SN à l'intérieur d'un cadre, ce cadre est délimité par deux grands crochets symbole des frontières du SN traité.

A l'intérieur de ce cadre les éléments morphosyntaxiques symbolisent les constituants syntaxiquement profonds du SN. Chaque ligne représente un couple de constituants, c'est-à-dire, une information morphosyntaxique à laquelle correspond une valeur sémantique :

Elément morphosyntaxique=valeur sémantique
--

C'est à travers l'empilement de chaque ligne qu'un lien syntactico-sémantique apparaît donnant ainsi l'information nécessaire à la compréhension du SN.

3.1.L'annexion nominale

3.1.1-L'appartenance / la possession

Exemple : مال لزيد

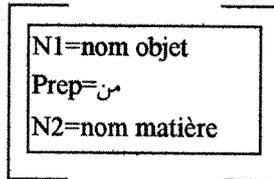
N1=objet possédé Prep=ل N2=possesseur (personne)

²¹⁸ J. DICHY, op. cit. p.3

²¹⁹ G. SABAH, op. cit., p. 127

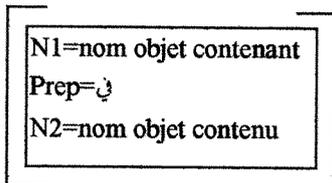
3.1.2-La forme et la matière

Exemple : خاتم من فضة



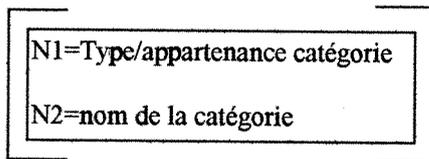
3.1.3-La contenance

Exemple : فنجان فيه قهوة



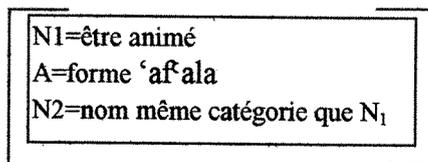
3.1.4.L'annexion de reformulation spécifique

Exemple شهر رمضان



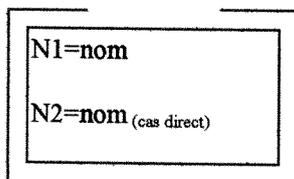
3.1.5. L'élatif

Exemple هو افضل رجل



3.1.6.Le tamyiz

Exemple : شير ارضاً



Le tamyiz ou spécifique a trois rôles sémantiques :

1- il spécifie la nature de la qualification d'un terme comme dans

رجل قبيح وجهاً *un homme laid de visage*

2- il est explicatif et a donc le sens de la préposition

شبر ارضاً *un empan de terre*

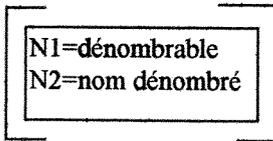
3- il exprime l'admiration comme dans :

اكرم يا ابي بكر اباً *Abu Bakr, quel noble père*

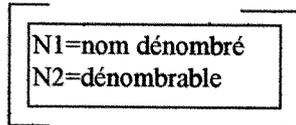
3.2.Modificateur

3.2.1. L'annexion à un numéral

Exemple : خمسة الاتواب

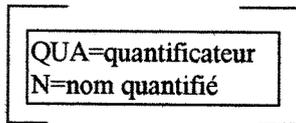


Exemple : الاتواب خمسة

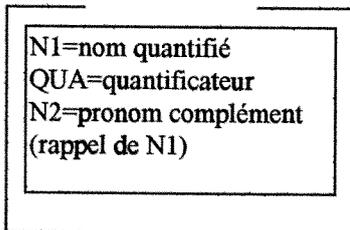


3.2.2.L'annexion à un modificateur ou à un quantificateur

Exemple : كل الناس

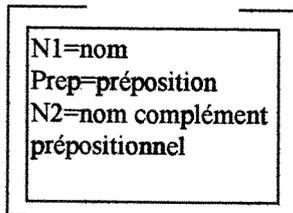


Exemple : الشعب كله



3.2.3.Les prépositions

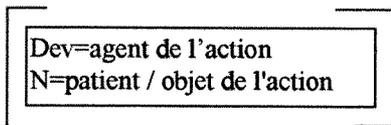
Exemple : الولد مع الرجل



3.3. L'annexion à un terme issu d'un déverbal

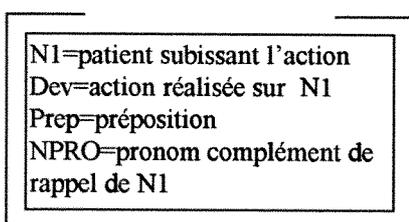
3.3.1. Le participe actif

Exemple : قاتل الناس



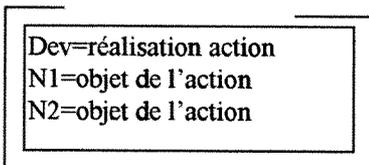
3.3.2. Le participe passif

Exemple : المرأة المحكوم عليها



3.3.3. Le nom de procès

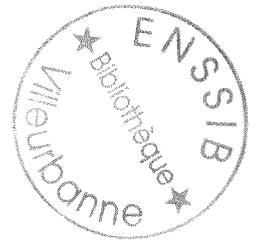
Exemple : تجديد بطاقة التعريف



Un double traitement tout d'abord morphosyntaxique et sémantique ensuite est utile dans le sens où le premier traitement permet de voir la constitution syntaxique et morphologique du SN annexionnel. Cela permet également d'en limiter la structure apparente et de faire resurgir la structure profonde notamment pour les annexions nominales. Nous avons ensuite procédé à une représentation schématique des catégories sémantiques des constituants du SN annexionnel.

Les analyses morphosyntaxiques actuelles arrivent à donner des résultats probants mais sur des unités lexicales simples. Pour les unités lexicales complexes, il faut identifier les structures les plus répandues en faire un lexique ou un dictionnaire. Lors de l'analyse cela nous permettra de savoir si tel terme complexe candidat à l'indexation est élément du lexique ou non.

TROISIEME PARTIE



Nous avons vu au début de cette étude l'intégration d'outils linguistiques dans certains logiciels documentaires en vue d'une indexation automatique.

Nous avons également vu que d'un point de vue documentaire les requêtes incluant des descripteurs complexes avaient un temps de réponse long et un résultat peu concluant. Aussi nous-a-t-il paru pertinent de procéder à une étude d'une catégorie des descripteurs complexes, ceux bâtis sur des structures de type annectif.

Nous avons répertorié ces structures (nominale et déverbale), nous en donnons ci-dessous le classement en trois grandes classes de modèles, dans lesquelles nous trouvons des sous-classes.

Nous avons opté pour ce classement car il nous paraît plus pertinent et plus clair de les classer par catégories grammaticales en reprenant le classement des grammairiens (classiques et modernes).

Nous pourrons ainsi voir si ces modèles peuvent être appliqués à un corpus de texte économique. Cette application va nous permettre de procéder à la constitution d'un "lexique de méta-descripteur", c'est-à-dire un lexique comprenant les éléments syntaxiques et sémantique pour des descripteurs n-terme incluant des déverbaux ou des prépositons ou particules.

I- Extraction des descripteurs et traitement du corpus

Notre étude se fait dans un cadre précis : nous extrayons les SN complexes en vue de repérer leur structuration et de voir lesquels peuvent être des candidats à l'indexation. Sachant que cette étude s'inscrit dans le cadre de l'indexation automatique, nous nous proposons ici d'adopter la méthode d'extraction des SN du groupe SYDO.

Nous allons déterminer pour chaque mot son appartenance à une catégorie, c'est-à-dire, son trait morphologique. Pour cela nous reprenons le lexique des

catégories et sous-catégories de chaque terme utilisé, lexique que nous avons déjà vu dans la deuxième partie de cette étude.

A-Le choix du corpus

Pour savoir si notre analyse peut être appliquée à l'indexation automatique nous devons l'expérimenter sur un corpus spécialisé.

Le corpus choisi traite des concepts économiques, c'est un manuel¹ d'enseignement de l'économie en Irak. Les bases des sciences économiques et les concepts, qui s'y rattachent, y sont évoquées.

Le traitement par corpus nous permettra de voir si ces modèles peuvent être appliqués à un domaine de spécialité.

B- Une typologie des descripteurs extraits

Les descripteurs appelés parfois encore mots clés, sont les termes qui sont autorisés [à être employé] à l'exception de tout autre pour l'indexation des documents et des questions. Ils servent à représenter les concepts ou notions des documents et des questions. Un descripteur peut être formé d'un mot ou d'une expression. (...). Une notion peut être représentée par un descripteur (cas le plus général) ou par plusieurs descripteurs² (dans ce cas, nous avons une description morphologique et sémantique).

Les descripteurs que nous avons extraits de notre corpus appartiennent à une catégorie de descripteurs composés de deux éléments ou plus. Ils sont classés par ordre de complexité syntaxique, c'est-à-dire, formé de deux termes ou plus et pouvant inclure des quantifieurs, des prépositions, etc. C'est pour cela que nous

¹ *Al-mafâhîm al-iqtisâdiyât aljamhuriyat al-irâqiyat- Wizarat at-t aqâfat wal-i'lâm da'irat al-i'lâm al-dâhilî al-âmat -as-salsilat al-iqtisâdiyât raqm 7, 88p.*

² J. CHAUMIER, *Analyse et langage documentaire, le traitement linguistique de l'information documentaire*, Entreprise Moderne d'Édition, 1982, p. 90

nous sommes attachées à représenter ces SN complexes de façon détaillée. Nous sommes au niveau le plus bas de l'analyse syntaxique. Les parenthèses permettent de repérer la détermination des SN et notamment les éléments auxquels se rapporte l'adjectif lorsqu'il y a une coordination.

1-Les descripteurs avec une structure annective à base nominale

1.1. Une catégorie de descripteurs de type : N+ (Det/Ø+N)

le processus de production	عملية الانتاج
Les conditions de crédit	شروط القرض
le prix de l'intérêt	سعر الفائدة
le marché du travail	سوق العمل
l'organisation OPEC	منظمة اوبك
les capitaux	روؤس الاموال

1.2. Une catégorie de descripteurs de type : N+((Det+N)+(Det+A))

le domaine de la production manufacturière	ميدان الانتاج المادي
le problème du chômage technique	مشكلة البطالة الفنية
le fonds monétaire international	صندوق النقد الدولي
la baisse de l'activité économique	هبوط النشاط الاقتصادي
l'organisation de coopération économique	منظمة التعاوض الاقتصادي

Nous remarquons ici que l'adjectif est déterminé et se rapporte au second terme de l'annexion. S'il ne portait pas l'article, il se serait rapporté au premier terme de l'annexion.

1.3. Une catégorie de descripteurs de type : $N+((\text{Det}+N)+C+(\text{Det}+N))$

les domaines de l'industrie et de l'agriculture	ميدان الصناعة و الزراعة
les niveaux de vie et niveaux culturels	مستويات المعيشة و الثقافة
la production de marchandises et d'articles	انتاج السلع و البضائع

Nous remarquons ici la présence du coordonnant qui relie deux termes de même nature pour éviter la répétition du premier élément de l'annexion qui est ellipsé dans la seconde partie du descripteur.

Nous rappelons ici la règle de l'ellipse du second terme que nous avons vu dans la deuxième partie de cette étude. Lorsque le deuxième terme de l'annexion est ellipsé, il ne reste que la tête de syntagme c'est-à-dire le premier terme. Sa fonction syntaxique est la même, il ne garde pas son tanwin. L'ellipse est toujours l'équivalent d'un terme déjà évoqué dans la première partie annexion, il n'est pas repris une deuxième fois pour éviter une répétition. Un terme ne peut donc être ellipsé et être coordonné à un autre terme uniquement s'il est équivalent au terme ellipsé.

1.4. Une catégorie de descripteurs de type :

- $N+((\text{Det}+\text{Dev})+C+(\text{Det}+\text{Dev}))$

l'appareil de production et d'exécution	اجهزة الانتاج و التنفيذ
---	-------------------------

- $N+((\text{Det}+\text{Dev})+C+(\text{Dev}+\text{NPRO}))$

le processus de production et sa réalisation	عملية الانتاج و تحقيقه
--	------------------------

Nous remarquons, dans ces deux exemples, que la tête de syntagme est un nom, celui-ci est ellipsé dans la deuxième partie du descripteur. Nous avons une coordination de deux déverbaux en l'occurrence un nom de procès (ou masdar). Cette catégorie est une variante de la précédente.

1.5. Une catégorie de descripteurs de type :

- $N+(\text{Dev}+(\text{N}+(\text{Det}+\text{N})))$

une méthode d'exploitation de la force de travail

طريق استغلال قوة العمل

- $N+(\text{Dev}+(\text{Det}+\text{N}))$

les crises de surproduction

ازمات فيض الانتاج

Nous remarquons dans le premier exemple, un premier terme annexé à un déverbal, celui-ci est suivi de plusieurs termes déterminés. C'est un exemple d'une annexion de plus de trois termes.

1.6. Une catégorie de descripteurs de type 6 :

- $((\text{Det}+\text{N})+(\text{Det}+\text{A}))+\text{Prep}+(\text{N}+(\text{Def}+\text{N}))$

les infrastructures de base de l'économie nationale

الهياكل الارتكازية للاقتصاد الوطني

l'organisation économique de l'ouest asiatique (ASEAN)

اللجنة الاقتصادية لغرب آسيا

la force productrice de travail

القوة الانتاجية للعمل

Nous remarquons que les exemples cités dans cette catégorie incluent comme préposition, la préposition "ل". Comme nous l'avons précédemment cité dans le lexique des prépositions, elle exprime entre autres *la finalité*, *le but* ou *la possession*. Elle sert également de préposition à un verbe intransitif. Ici elle exprime *le but*, *la finalité*.

1.7. Une catégorie de descripteur de type :

- $N+(\text{Dev}+\text{Def}+\text{N})+\text{Prep}+\text{Dev}+(\text{Def}+\text{N}))$

la rareté des capitaux nécessaire à l'investissement

ندرة رأس المال الازم للاستثمار

La préposition contenue dans ce descripteur fait partie de la construction syntaxique du déverbal. Celui-ci fonctionne comme un verbe, il garde donc la préposition qui l'accompagne quand il est déverbal.

1.8. Une catégorie de descripteur de type : $N+(Dev+A)$

Un problème d'organisation économique

مسألة تنظيم اقتصادي

Nous avons ici un exemple d'annexion indéterminée.

1.9. Une catégorie de descripteur de type :

- $((Det+N)+(Det+A)+Prep+(Det+N)+Prep+((Det+N)+A))$

le budget général de l'Etat avec des recettes suffisantes

الميزانية العامة للدولة بالارادات الكافية

La préposition “ل” exprime ici l'appartenance. La présence de la préposition “ب” exprime ici le moyen.

2-Les descripteurs composés d'un quantificateur

2.1. Une catégorie de descripteur de type : $(N+A)+(QUA +(N+A))$

des matières premières d'entreprises industrielles مواد اولية ذات منشاء صناعي

Nous avons ici un exemple d'annexion indéterminée. Nous avons vu dans le lexique des modificateurs un certain nombre de quantificateurs dont celui qui est présent dans ce descripteur : “ذات”. C'est un nom proche du démonstratif “ذا”, il s'emploie toujours comme premier terme d'une annexion. Il signifie *possesseur de*, le groupe qu'il forme à une valeur d'adjectif qualificatif, il est nécessairement défini par un nom complément.

2.2 Une catégorie de descripteur de type : $Num+N+N+(Det+N)+(Det+N)$

le triple du nombre des autres ingénieurs

ثلاثة اضعاف عدد المهندسين الاخرين

Nous avons ici une annexion avec un numéral comme tête de syntagme.

2.3. Une catégorie de descripteur de type :

- (Prep+QUA+N)+(N+A)+Prep+(Det+N)

pour chaque entreprise de productivité une capacité précise de rendement

لكل منشأة انتاجية كفاءة معينة للانتاج

Nous avons vu dans le lexique des prépositions que certaines prépositions pouvaient s'annexer à d'autres. Dans le cas de cet exemple, la préposition "ل", qui exprime *la possession* ou *l'appartenance*, est suivi du quantificateur "كل". Celui-ci a pour signification *chaque, tous les*, lorsqu'il est complété par un nom indéfini au singulier.

2.4. Catégorie de descripteurs de type :

- Dev+QUA+((Det+N)+(Det+A))

l'investissement de l'ensemble des ressources économiques

استثمار كافة الموارد الاقتصادية

l'utilisation de tout le potentiel technique

استخدام كل الامكانيات التقنية

- QUA+(Det+N)

l'ensemble du matériel

مجموع الادارات

Nous avons ici trois types de quantificateurs dont un en tête de syntagme "مجموع".

Nous avons vu dans la catégorie précédente le quantificateur "كل"; en ce qui concerne "كافة", il exprime *la totalité*, il s'emploie surtout comme premier terme d'une annexion ou en apposition (il désigne la *totalité* sans exception).

3-Les descripteurs avec une structure à base déverbale

La langue arabe est bâtie sur des schèmes. Nous avons procédé ici à un classement par type de schèmes. Ces schèmes désignent des combinaisons de trois consonnes auxquelles peut s'ajouter la génération d'une consonne radicale et par extension un schème de quatre consonnes. Ces schèmes correspondent à des moules dans lesquels viennent se placer les racines pour former les radicaux de mots.

Certains descripteurs extraits peuvent appartenir à plusieurs catégories morpho-syntaxiques, il n'est donc pas étonnant de les voir dans la catégorie des quantificateurs et dans celle des déverbaux.

a-Le nom de procès

1. Les descripteurs dont la tête de syntagme est bâtie sur le schème

taf'îl "تفعيل"

- Dev+(Det+N)

le renouvellement de la production تجديد الانتاج

- Dev+((Det+N)+(Det+A))

la distribution du revenu national توزيع الدخل القومي

- Dev+C+Dev+C+Dev+(Det+N)

تخطيط و توجيه و ادارة الانتاج

la planification, l'orientation et la gestion de la production

- (Dev+(Det+N))+C+(Dev+NPRO)

la description du travail et son organisation توصيف العمل و تنظيمه

Nous avons dans cet exemple une ellipse du second terme qui est présent dans la première partie de l'annexion, il est repris par le pronom de rappel.

- Dev+((Det+N)+(Det+A)+C+(Det+A))

تحديد الاهداف الاقتصادية والاجتماعية

la définition des objectifs économiques et sociaux

- Dev+((Det+N)+(Det+A))

le financement du commerce international تمويل التجارة الدولية

2. Les descripteurs dont la tête de syntagme est bâtie sur le schème

istif^{al} “ استفعال ”

- Dev+((Det+N)+(Det+A))

l'exploitation des machines abondantes استغلال الآلات المتوفرة

- Dev+(QUANT+(Det+N)+(Det+A))

l'investissement de l'ensemble des ressources économiques

استثمار كافة الموارد الاقتصادية

l'utilisation de tout le potentiel technique

استخدام كل الامكانيات التقنية

3. Les descripteurs dont la tête de syntagme est bâtie sur le schème

tafa^{ul} “ تفعل ”

- Dev+((Ø +N)+(Det+N))

la possession des moyens de production

تملك وسائل الانتاج

- (Det+Dev)+Prep+((Det+Dev)+C+(Det+Dev))

l'équilibre entre la production et la consommation

التوازن بين الانتاج و الاستهلاك

4. Les descripteurs dont la tête de syntagme est bâtie sur le schème

’ifti’âl “فتعال”

- Dev+(Det+N)+C+(Dev+NPRO)

l’augmentation et la baisse de la productivité ارتفاع الانتاجية او انخفاضها

- Dev+(Ø+N)+(Det+N)+(Det+A)

la hausse des prix des matières premières ارتفاع اسعار المواد اولية

- Dev+((Det+N)+(Det+A))

l’élargissement des marchés étrangers اتساع السوق الخارجية

5. Les descripteurs dont la tête de syntagme est bâtie sur le schème

’inf’âl “انفعال”

- Dev+(Det+N)+(Det+A)

la réduction de l’énergie productrice انخفاض الطاقة الانتاجية

6. Les descripteurs dont la tête de syntagme est bâtie sur le schème ’if’âl

“فعال”

- Dev+Dev+((Det+N)+((Det+N)+(Det+A)))

le renouvellement de la distribution du revenu du PNB

اعادة توزيع الدخل الناتج الاجمالي

- Dev+((Ø+N)+(Ø+A))

l’établissement des industries de transformations اقامة صناعات تحويلية

7. Les descripteurs dont la tête de syntagme est bâtie sur le schème

”فعال“ *fi‘âl*

- Dev+(Prédét+N)

l'évaluation de cette productivité

قياس هذه الانتاجية

b- Les déverbaux bâtis sur des schèmes de formes non dérivées :

Les descripteurs dont la tête de syntagme est bâtie sur le schème

”فعل“ *fu‘ûl*

- Dev+Prep+((Det+N)+(Det+A))

le redressement du secteur agricole

النهوض بالقطاع الزراعي

- Dev +((Det+N)+(Det+A))

la baisse de la valeur la monnaie fiduciaire

هبوط قيمة النقود الورقية

Ce descripteur a comme tête de syntagme un déverbal, il faut noter que la majorité des structures annectives construites sur les schèmes déverbaux désignent un processus ou le cours d'une action de transformation.

c- Le participe passif :

Les descripteurs dont la tête de syntagme est bâtie sur le schème *maf'ul*

”مفعول“

le produit industriel

منتوج الصناعات

II- Le traitement du corpus à partir du modèle des structures annectives

Nous avons repris, pour cette partie, le classement des structures annectives que nous avons étudié dans la deuxième partie de ce mémoire.

A-Le modèle de traitement des structures annectives :

Nous avons réparti ces structures en trois types de modèles comme suit :

Modèle 1 : Annexion nominale

M1.1 - Possession / appartenance :

Exemple : *mal li zaydin* مال لزويد

Det1=∅ N1=objet possédé <i>mal</i> Prep= <i>li</i> Det2=∅ N2=possesseur <i>zaydin</i>

M1.2 - Forme / matière :

Exemple : *hatim min fuḍa* حاتم من فضة

Det1=∅ N1=nom objet <i>hatim</i> Prep= <i>min</i> Det2=∅ N2=nom matière <i>fuḍa</i>

M1.3 - Contenance :

Exemple : *finjân fihi qahwa* فنجان فيه قهوة

Det1=∅
 N1=nom objet contenant *finjân*
 Prep=*fi*
 NPRO=Pronom (reprise de N1) *hi*
 Det2=∅
 N2=nom objet contenu *qahwa*

M1.4 - Reformulation spécifique :

Exemple : *šahr ramadân* شهر رمضان

Det1=∅
 N1=Type/appartenance catégorie *šahr*
 Det2=∅
 N2=nom de la catégorie *ramadân*

M1.5 - Elatif :

Exemple : *hûwa 'afaḍalu rajulin* هو افضل رجل

Det1=∅
 N1=être animé *hûwa*
 A=forme 'af'ala '*afaḍalu*
 Det2=∅
 N2=nom même catégorie que N₁ *rajulin*

M1.6 - Tamyiz :Exemple : *Šabr arđan* شبر ارضا

Det1=∅ N1=nom <i>Šabr</i> Det2=∅ N2=nom (cas direct) <i>arđan</i>
--

Modèle 2 : Annexion à un modificateur**M2.1 - Annexion à un numéral :**Exemple : *ħamsa at-twabin*: خمسة الاثواب

Det1=∅ N1=dénombrable <i>ħamsa</i> Det2=∅ N2=nom dénombré <i>at-twabin</i>

Exemple : *at-twabu al-ħamsa* الاثواب الخمسة

Det1=∅ N1=nom dénombré <i>at-twabu</i> Det2=∅ N2=dénombrable <i>al-ħamsa</i>

M2.3 - Quantificateur :Exemple : *kulu al-nâsi* كل الناس

QUA=quantificateur <i>kulu</i> Det= <i>al</i> N=nom quantifié <i>nâsi</i>

Exemple : *aš-ša^cbu kuluhu* الشعب كله

<p>Det1= <i>al</i> N1=nom quantifié <i>ša^cbu</i> QUA=quantificateur <i>kulu</i> Det2=∅ N2=pronom complément (rappel N1) <i>hu</i></p>

M2.2 - Préposition / Quasi-Préposition :

Exemple: *al-waladu ma^ca ar-rajuli* الولد مع الرجل

<p>Det1= <i>al</i> N1=nom <i>waladu</i> Prep=préposition <i>ma^ca</i> Det2= <i>al</i> N2=nom complément prépositionnel <i>rajuli</i></p>

Modèle 3 : Annexion déverbale

M3.1 - Participe actif :

Exemple : *qâtilu al-nâsi* قاتل الناس

<p>Dev= agent de l'action <i>qâtilu</i> Det= <i>al</i> N= patient /objet de l'action <i>nâsi</i></p>
--

M3.2 - Participe passif :

Exemple : *al-mar'a al-mahkum^calayha* المرأة المحكوم عليها

<p>Det1= <i>al</i> N1= patient subissant l'action <i>mar'a</i> Dev (participe passif)= action réalisée sur N1 <i>al mahkum</i> Prep=préposition <i>alâ</i> NPRO=pronom complément de rappel N1 <i>ha</i></p>

M3.3 - nom de procès :

Exemple : *tajdīd biṭâqat al ta'rif* تجديد بطاقة التعريف

Dev= réalisation action *tajdīd*
 Det1=∅
 N1= objet de l'action *biṭâqat*
 Det2= *al*
 N2= complément de N1 *ta'rif*

B-Le traitement des descripteurs complexes selon le modèle des structures annectives

1-Le modèle de l'annexion nominale

a) La possession / l'appartenance :

Modèle : *mal li zaydin*

le budget général de l'Etat avec des recettes suffisantes

الميزانية العامة للدولة بالارادات الكافية

al-miyzanyiat al-ʿamat li al-dawlat bi al-ʿidaʿât al-kâfiyat

Det1= **al**
 N1=objet possédé -**miyzanyiat**
 A1=adjectif relatif à l'objet possédé **al-ʿamat**
 Prep= **li**
 Det2= **al**
 N2=possesseur -**dawlat**
 Prep= **bi**
 Det3= **al**
 N3=*spécification objet possédé* -**ʿidaʿât**
 A3= *adjectif qualificatif de N3* **al-kâfiyat**

l'organisation économique de l'ouest asiatique (ASEAN) اللجنة الاقتصادية لغرب آسيا

al-lajnat al-iqtisadiyat li garb 'âsiya

<p>Det1= al N1=Objet possédé-lajnat A1=adjectif relatif à l'objet possédé al-iqtisadiyat Prep= li Det=∅ N2=Possesseur garb 'âsiya</p>

la force productrice de travail

القوة الانتاجية للعمل

al-qûwat al-intâjiyat li al-'amal

<p>Det= al N1=Objet possédé -qûwat A1=Adjectif relatif à l'objet possédé al-intâjiyat Prep= li Det2= al N2=Nom possesseur -'amal</p>

Pour le traitement des descripteurs sur le modèle de la possession, on note une nuance par rapport au modèle initiale, c'est l'ajout des adjectifs relatifs qui se rapporte à l'objet possédé. L'adjectif suit directement le nom auquel il se rapporte. Ici le possesseur n'est pas nécessairement une personne c'est une entité qui reflète un possesseur abstrait.

b) La contenance

marché du travail

سوق العمل

sûq fihi al-‘amal

Det1=∅
N1=nom contenant sûq
Prep= fi
NPRO= hi
Det2= al
N2=Nom contenu -‘amal

le fonds monétaire internationale

صندوق النقد الدولي

şandûq fihi al-naqd al-dawl

Det1=∅
N1=Contenu şandûq
Prep= fi
NPRO= hi
Det2= al
N2=Contenant -naqd
Det= al
A2=adjectif qualificatif de N2 -dawli

le domaine de la production manufacturière

ميدان الانتاج المادي

maydân fihi al-intâj al-mâdi

Det1=∅
N1=Contenu -maydân
Prep= fi
NPRO= hi
Det2= al
N2=Contenant -intâj
Det2= al
A2=adjectif qualificatif de N2 -mâdi

les domaines de l'industrie et de l'agriculture

ميدان الصناعة و الزراعة

maydân fihi al-sinâ'at wa al-zirâ'at

Det1=∅
N1=Contenant maydân
Prep= fi
NPRO= hi
Det2= al
N2=Contenu -sinâ'at
C= wa
Det3= al
N3=contenu -zirâ'at

Comme dans le modèle précédent, la tête de syntagme exprime ici l'idée du contenant. Cependant dans le cas des descripteurs que nous traitons ce n'est pas un contenant réel ni un contenu réel mais le concept de contenu et de contenant, c'est-à-dire, une représentation abstraite.

c) La reformulation spécifique

I' OPEC

منظمة اوبك

munaḍamat obek

Det1=∅
N1=Type / appartenance catégorie munaḍamat
Det2=∅
N2=nom de la catégorie obek

Le traitement du descripteur de cette catégorie est identique au modèle initiale de traitement.

2-Le modèle de l'annexion à un modificateur

a) L'annexion à un numéral

le triple du nombre des autres ingénieurs ثلاثة اضعاف عدد المهندسين الاخرين

talâṭat adʿâf ʿadad al-muhandisin al-aḥarîn

Det1=∅
 N1=dénombrable **talâṭat**
 QUA1=**adʿâf**
 QUA2=**ʿadad**
 Det2= **al**
 N2=nom dénombré **-muhandisin**
 Det2= **al**
 A2=adjectif qualifiant le nom dénombré **-aḥarîn**

Le traitement de ce descripteur est à peu de chose près identique au modèle initiale du traitement du numéral, mise à part le fait que nous avons deux quantificateurs qui suivent le numéral. Ceci est dû au fait que pour signifier le triple ou le quadruple, c'est-à-dire, la multiplicité d'une quantité, l'arabe utilise le terme "adʿâf" de la racine "dʿf" qui signifie "multiplier, accroître en nombre".

Pourquoi avoir un deuxième quantificateur si nous en avons déjà un ? Le terme adʿaf doit être suivi d'un autre terme sémantiquement proche, c'est-à-dire, qui désigne soit la quantité soit le nombre de la chose dénombrée.

Ces deux points mises à part, le traitement de ce descripteur correspond au modèle initiale de traitement de l'annexion à un numéral.

b) La préposition et la quasi-préposition

la force productrice de travail

القوة الانتاجية للعمل

al-qûwat al-intâjiyat li al-‘amal

Det1= al
 N1=nom -qûwat
 Det1= al
 A1=Adjectif -intâjiyat
 Prep=préposition li
 Det2= al
 N2=nom complément prépositionnel -‘amal

l'organisation économique de l'ouest asiatique (ASEAN) اللجنة الاقتصادية لغرب آسيا

al-lajna al-iqtisâdiyât li ġarb asiyat

Det1= al
 N1=nom -lajna
 Det1= al-
 A1=Adjectif iqtisâdiyât
 Prep=préposition li
 Det2=Ø
 N2=nom complément prépositionnel ġarb 'asiyâ

Nous avons repris ici deux exemples de descripteurs traités dans le cas du modèle possession : le traitement est identique à celui des prépositons.

c) Le quantificateur

les matières premières des entreprises industrielles

مواد اولية ذات منشأ صناعي

mawâd ulawiyat dât manša' şan'î

<p>Det1=∅ N1=nom mawâd Det1=∅ A1=adjectif qualificatif de N1 ulawiyat QUA=dât Det2=∅ N2=nom quantifié manša' Det=∅ A2=adjectif qualificatif de N2 şan'î</p>
--

Nous avons vu dans la partie "typologie des descripteurs" la signification du quantificateur "ذات" qui précise une caractéristique appartenant au terme qui le précède.

Pour le traitement des annexions nominales, nous avons trouvé à travers l'étude de notre corpus d'autres types ou catégories sémantiques non présentes dans la liste précédente. Ce sont les catégories suivantes :

- processus,
- organisation / structure,
- méthode / mode / technique,
- résultat d'une action.

d) Le processus

Modèle :

Det1=Ø/al N1=type de processus Det2=Ø/al N2=nature du processus
--

le processus de production et sa réalisation

عملية الانتاج وتحقيقه

‘amiliyat al-intâj wa taḥqîq hi

Det1=Ø N1=type de processus ‘amiliyat Det2= al N2=nature du processus -intâj C= wa Dev=déverbal taḥqîq NPRO= pronom de rappel de N2 hi
--

La tête de syntagme désigne le type de processus.

e) L'organisation

Modèle :

Det1=Ø N1=Nom organisation ou structure Det2=Ø N2=nom nature organisation
--

اللجنة الاقتصادية لغرب آسيا (ASEAN) l'organisation économique de l'ouest asiatique

al-lajna al-iqtisâdiya li ġarb 'asiyâ

<p>Det1= al N1= Organisation -lajna Det1= al A1= Nature organisation -iqtisâdiya Prep= li Det2=Ø N2=ġarb 'asiyâ</p>

Dans ce modèle le sème *organisation* est présente dans la tête de syntagme.

f) La méthode

Modèle :

<p>Det1=Ø/al N1=nom méthode Dev= type méthode Det2=Ø/ al N2=ce sur quoi porte la méthode</p>
--

une méthode d'exploitation de la force de travail

طريقة استغلال قوة العمل

ṭarîqat istiġlâl qûwat al-ʿamal

<p>Det1=Ø N1= ṭarîqat Dev= istiġlâl Det2=Ø N2= qûwat Det3=al- N3= ʿamal</p>

Le traitement des descripteurs inclut comme dans le cas du modèle précédente le sème de la catégorie en question : nous avons ici le sème de la méthode.

g) Le résultat d'une action

Modèle :

Det1=∅
N1=Résultat action
Dev=nature action
Det2=∅
N2=objet l'action

les crises de surproduction

ازمات فيض الانتاج

'azmât fayḍ al-intâj

Det1=∅
N1=Résultat action **'azmât**
Dev=nature action **fayḍ**
Det2= **al**
N2=objet de l'action **-intâj**

la force productrice de travail

القوة الانتاجية للعمل

al-quwat al-intâjyat li al-ʿamal

Det1= **al**
N1=Résultat action **-quwat**
Det1= **al**
A1=nature action **-intâjyat**
Prep= **li**
Det2= **al**
N2=objet de l'action **-ʿamal**

Nous venons de voir le traitement syntactico-sémantique des descripteurs ayant comme tête de syntagme un nom. Nous avons remarqué qu'ils correspondent au modèle initiale de traitement syntactico-sémantique mais ces descripteurs contiennent plus d'éléments dans leur structuration.

3-Le modèle de l'annexion à un déverbal

a) Le masdar

Nous avons opté pour un classement par schème car en arabe les schèmes ont une valeur sémantique précise.

1- modèle sur le schème *taf'il* “ تفعيل ”

Dev=déverbal forme II taf'il : intensité action , “faire faire” Det=Ø/al N1=nom complément déverbal (objet de l'action)
--

le renouvellement de la production

تجديد الانتاج

tajdîd al-intâj

Dev= tajdîd Det1= al N1= -intâj
--

la distribution du revenu national

توزيع الدخل القومي

tawzîc al-daḥl al-qawmî

Dev=tawzîf
 Det1= al
 N1= -dahî
 Det1= al
 A1= -qawmî

la planification, l'orientation et la gestion de la production

تخطيط و توجيه و ادارة الانتاج

taḥṭîf wa tawjîh wa 'idârat al-intâj

Dev=taḥṭîf
 C=wa
 Dev=tawjîh
 C=wa
 Dev='idârat
 Det1= al
 N1= -intâj

la description du travail et son organisation

توصيف العمل و تنظيمه

tawsîf al-'amal wa tanzîmi hi

Dev=tawsîf
 Det1= al
 N1= -'amal
 C=wa
 Dev=tanzîmi
 NPRO=hi

تحديد الاهداف الاقتصادية والاجتماعية

la définition des objectifs économiques et sociaux

taḥṭīṭ al-ahdâf al-iqtisâdiyat wa al-ijtimâ'iyat

Dev=taḥṭīṭ
Det1= al
N1= -ahdâf
Det1= al
A1= -iqtisâdiyat

le financement du commerce international

تمويل التجارة الدولية

tamwîl al-tijârat al-dawlyat

Dev= tamwîl
Det1= al
N1= -tijârat
Det1= al-
A1= dawlyat

Nous observons que les déverbaux tête de syntagme du descripteur pluriterme sont construits sur le schème *tafîl*. Ce schème correspond à la forme II des verbes dérivés qui est la forme *fa^{cc}ala*. Le redoublement de la deuxième radicale implique sémantiquement l'intensité de l'action ou signifie "faire faire" comme *°alama* savoir et *°allama* enseigner, ou la répétition d'une action.

2- Le modèle sur le schème *istaf'ala* “ استفعال ”

Dev=déverbal forme IX *istif'âl*: sens factitif causatif “faire faire” c’est le réfléchi passif de la forme IV *'af'ala*
 N1=nom complément déverbal (objet de l’action)

l’exploitation des machines abondantes

استغلال الآلات المتوفرة

istiglâl al-'alât al-mutawafarat

Dev=*istiglâl*
 Det1= *al*
 N1= *'alât*
 Det1= *al*
 A1= *-mutawafarat*

l’investissement suffisant des ressources économiques

استثمار كافة الموارد الاقتصادية

istitmâr kâfat al-mawârd al-iqtisâdiyât

Dev=*istitmâr*
 QUA=*kâfat*
 Det1= *al*
 N1= *- mawârd*
 Det1= *al*
 A1= *-iqtisâdiyât*

l'utilisation de tout le potentiel technique

استخدام كل امكانيات التقنية

istiḥdām kulu 'imkāniyat al-tiqniyat

Dev=istiḥdām
 QUA=kulu
 Det1=∅
 N1='imkāniyat
 Det1= al
 A1= -tiqniyat

C'est la forme réfléchie passive de la forme IV ('aḥ'ala) qui a un sens factitif ou causatif. Cette forme a également le sens de "faire faire", elle comprend aussi le sens de "demander", de "rechercher", comme dans les verbes suivant 'aḥ'bara informer, istaḥ'bara s'informer ou enquêter .

3- Le modèle sur le schème *tafa'ul* "تفعل"

Dev=déverbal formeV *tafa'ul* sens réfléchi passif de la forme
 II fa'ala
 Det1=∅
 N1=nom complément déverbal (objet de l'action)

la possession des moyens de production

تملك وسائل الانتاج

tamaluk wasa'il al-intâj

Dev=**tamaluk**
 Det1=∅
 N1=**wasa'il**
 Det2=∅
 N2=**al-intâj**

l'équilibre entre la production et la consommation

التوازن بين الانتاج و الاستهلاك

al-tawazun bayna al-intâj wa al-istihlâk

Dev=**al-tawazun**
 Prep=**bayna**
 Det1=∅
 N1=**al-intâj**
 C=**wa**
 Det=**al**
 Dev=**-istihlâk**

C'est le schème de la forme V *tafa^cala* c'est également la forme II mais avec le "t" préfixé. Elle a le sens réfléchi passif de la forme II, par exemple *allama* enseigner *ta^callama* apprendre.

4- Modèle sur le schème "افعال" 'fti'âl

Dev=déverbal formeVIII **tafa^cul** sens réfléchi passif de la
 forme I **fa^cala** action du sujet a son profit
 N1=nom complément déverbal (objet de l'action)

l'augmentation et la baisse de la productivité

ارتفاع الانتاجية او انخفاضها

al-irtifâc al-intâjijyat aw inḥifâḍ ha

Dev=**al-irtifâc**
 Det1= **al**
 N1= **-intâjijyat**
 C=**aw**
 Dev=**inḥifâḍ**
 NPRO=**ha**

la hausse des prix des matières premières

ارتفاع اسعار المواد اولية

irtifâc as'âr al-mawâd al-ulawiyat

Dev=**irtifâc**
 Det1=**Ø**
 N1=**as'âr**
 Det2= **al**
 N2= **-mawâd**
 Det2= **al**
 A2= **-ulawiyat**

l'élargissement des marchés étrangers

اتساع السوق الخارجية

itisâc al-suq al-ḥarijyat

Dev=**itisâc**
 Det1= **al**
 N1= **-suq**
 Det1= **al**
 A1= **-ḥariyyat**

La forme VIII dénote le réfléchi passif de la forme I (fa'ala) ou désigne une action que le sujet fait à son profit comme *jamā'a* rassembler, réunir et *ijtama'a* se réunir.

5- Modèle sur le schème *infi'âl* "انفعال"

Dev=déverbal formeVII **infi'âl** sens réfléchi passif de la forme I **fa'ala** insistance sur le passif correspond aux verbes pronominaux français.
 N1=nom complément déverbal (ce sur quoi porte l'action)

la réduction de l'énergie productrice

انخفاض الطاقة الانتاجية

inhifâḍ al-ṭaqat al-intâjīyat

Dev=**inhifâḍ**
 Det1= **al**
 N1= **-ṭaqat**
 Det1= **al**
 A1= **-intâjīyat**

6- Modèle sur le schème if'âl " افعال "

Dev=déverbal forme IV **if'âl** sens factitif /causatif "faire faire"
 N1=nom complément déverbal (ce sur quoi porte l'action)

le renouvellement de la distribution du revenu du PNB

اعادة توزيع الدخل الناتج الاجمالي

i'âdat tawzîf al-daħl al-nâtîj al-ijmâlî

Dev=i'âdat.
 Dev=tawzîf
 Det1= al
 N1= -daħl
 Det2= al
 N2= -nâtîj
 Det2= al
 A2= -ijmâlî

l'établissement des industries de transformation اقامة صناعات تحويلية

iqâmat şinâ'ât taħwîliyat

Dev=iqâmat.
 Det1=Ø
 N1=şinâ'ât
 Det1=Ø
 A1=taħwîliyat

8- Modèle sur le schème fi'âl "فعال"

Dev=déverbal forme III fâ'il, sens faire un effort en direction de quelque chose.
N1=nom complément déverbal (objet de l'action)

l'évaluation de cette productivité

قياس هذه الانتاجية

qiyâs haḍihi al-intâjijyat

Dev=**qiyâs**
Det1=**haḍihi**
Det1= **al**
N1= **-intâjijyat**

C'est le schème de la forme III qui dénote l'idée de "faire un effort en direction de quelqu'un ou de quelque chose". Les verbes construits sur ce schème, sont la plupart du temps transitif direct comme *qâṭala* combattre ou *kâṭaba* écrire à, correspondre avec.

-Participe passif (le participe actif est moins présent dans le corpus)

9- Modèle sur le schème maf'ûl "مفعول"

Dev=déverbal participe passif **maf'ûl**
N1=nom complément déverbal (ce sur quoi porte l'action)

le produit industriel

منتوج الصناعات

mantûj al-şinâ'ât

Dev= mantûj

Det1= al

N1= -şinâ'ât

10- Autres exemples de masdar de formes non dérivées :

modèle forme verbale :

Dev=déverbal réalisation de l'action

N1=nom complément déverbal (ce sur quoi porte l'action)

le redressement du secteur agricole

النهوض بالقطاع الزراعي

al-nuhûḍ bi al-qitâ' al-zirâ'i

Dev=al-nuhûḍ

Prep=bi

Det1= al

N1= -qitâ'

Det1= al

A1= -zirâ'i

la baisse de la valeur de la monnaie fiduciaire

هبوط قيمة النقود الورقية

hubuṭ qimat al-nuqud al-warqiyat

Dev=hubuṭ
Det1=∅
N1=qimat al-nuqud
Det1= al
A1= -warqiyat

Les descripteurs ayant comme tête de syntagme un déverbal sont majoritairement construits sur les schèmes de formes dérivées. Ils désignent un processus ou le cours d'une action.

Les descripteurs complexes ayant comme tête de syntagme un déverbal sont syntaxiquement construits sur un seul modèle pour tous les schèmes. Cependant ils varient sémantiquement avec le changement de schème.

III-Elaboration d'un lexique de métadescripteurs :

Les structures annectives en tant que descripteur complexe candidat à l'indexation

Nous avons recensé toutes les catégories sémantiques de l'annexion sous forme synthétique de tableau. Dans ce tableau nous avons la nature sémantique du modèle annectif à laquelle correspond la structure syntactico-sémantique deux types d'exemples : le premier est celui du modèle initiale de traitement et le second celui des descripteurs extraits du corpus. Il arrive qu'il y ait pour un même modèle deux variantes. Nous avons représenté de façon la plus exhaustive

possible la structure syntactico-sémantique de chaque modèle afin de voir les différents constituants des descripteurs complexes qui ont été extraits du corpus.

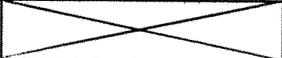
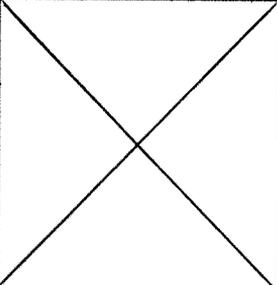
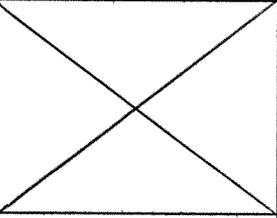
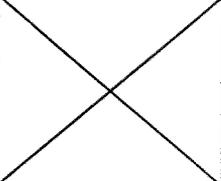
Les cases barrées d'une croix représentent les modèles pour lesquels nous n'avons pas d'exemples. Car ce sont soit les nouveaux modèles trouvés à la suite de l'étude des descripteurs, soit les exemples correspondant au modèle initiale et pour lesquels nous n'avons pas d'exemples correspondant dans notre corpus.

Nous avons réparti ce lexique en trois parties :

- la première concerne les méta-descripteurs ayant une base nominale
- la seconde concerne les méta-descripteurs incluant des modificateurs
- la troisième concerne les méta-descripteurs ayant une base verbale.

A- Méta-descripteur ayant une base nominale

n° modèle	Nature modèle	Structure syntactico-sémantique	Exemples du modèle initiale	Exemples du Corpus
1	Possession / Appartenance	Det1=Ø/al N1=objet possédé Det1=Ø/al A1=adjectif qualifiant N1 Prep=li Det2=Ø/al N2=possesseur (Animé/non-animé) Det2=Ø/al A2=adjectif qualifiant N2	<i>mâlu li Zaydîn</i>	al-miyzânyiat al-âmat li al- dawlat bi al- 'idafât al-kâfiyat
2	Forme /matière	Det1=Ø/al N1=nom objet Prep=min Det2=Ø/al	<i>hâtim min</i> <i>fudat</i>	X

		N2=nom matière		
3	Contenance	Det1=Ø/al N1=nom objet contenant Prep=fi NPRO=rappel de N1 Det2=Ø/al N2=nom objet contenu Det2=Ø/al A2=adjectif qualificatif de N2	<i>fjân fihi</i> <i>qahwat</i>	<i>şanduq fihi naqd</i> <i>al-dawli</i>
4	reformulation spécifique	Det1=Ø/al N1=Type/ appartenance catégorie Det2=Ø/al N2=nom de la catégorie	<i>şahr ramadân</i>	<i>munadamat obek</i>
5	Elatif	Det1=Ø/al N1=être animé A=forme 'âf'ala Det2=Ø/al N2=nom même catégorie que N1	<i>'afaḍalu</i> <i>rajulin</i>	
6	Tamyiz	Det1=Ø/al N1=nom Det2=Ø/al N2=nom _(cas direct)	<i>Şabr arḍan</i>	
7	Processus	Det1=Ø/al N1=type de processus Det2=Ø/al N2=nature du		<i>'amiliyat al-intâj</i> <i>wa taḥqîq hi</i>

		processus C=wa Dev=déverbal NPRO= pronom complément de N2		
8	Organisation / Structure	Det1=Ø/al N1=type de processus Det2=Ø/al N2=nature du processus		al-lajna al- iqtisadiya li ġarb 'asiya
9	Méthode	Det1=Ø/al N1=Nom méthode Dev=type de méthode Det2=Ø/al N2=objet de la méthode		ṭariq istiglâl quwat al-ʿamal
10	Résultat d'une action	Det1=Ø/al N1=Résultat action Dev=nature action Det2=Ø/al N2=objet de l'action		'azmat fayḍ al- intâj

B-Méta-descripteur incluant un modificateur

n° modèle	Nature modèle	Structure syntactico- sémantique	Exemples du modèle initiale	Exemples du Corpus
1	Annexion avec un numéral	Det1=Ø/al 1=dénombrable Det2=Ø/al	<i>ḥamsa aṭ- ṭwabin</i>	ṭalâṭat adʿaf ʿadad al-

		N2=nom dénombré Det2=Ø/al A2=adjectif qualifiant le dénombré		muhandisin al-aḥarîn
		Det1=Ø/al N1=nom dénombré Det2=Ø/al 2=dénombrable	<i>at-twabû al- ḥamsa</i>	X
2	Annexion avec une préposition et quasi préposition	Det1=Ø/al N1=nom Prep=préposition -tion Det2=Ø/al N2=nom complément prépositionnel	<i>al-waladu ma^ca ar- rajuli</i>	al-quwat al- intajiyat li al- ʿamal
3	Annexion avec quantificateur	QVA =quantifi- -cateur Det2=Ø/al N=nom quantifié	<i>kul aš-ša^cb</i>	mawâd ʿulawiyat dât manša' šan'f
		Det1=Ø/al N1=nom quantifié Det1=Ø/al A1=adjectif qualifiant N1 QVA =quantifi- -cateur NPRO=pronom complément	<i>aš-ša^cb kuluhu</i>	X

		(rappel)		
--	--	----------	--	--

C-Méta-descripteur ayant une base verbale

n° modèle	Nature modèle	Structure syntactico-sémantique	Exemples du modèle initiale	Exemples du corpus
1	Participe actif	Dev=déverbal Det=Ø/al N=nom complément du déverbal	<i>qâtilu al-nassi</i>	
2	Participe passif	Dev=déverbal participe passif maf'ul Det1=Ø/al N1=nom complément déverbal (objet de l'action)	<i>al-mar'a al-mahkum 'alayha</i>	matûj al- šinâ'ât
3	Masdar/ Nom de procès	Dev=déverbal Det=Ø/al N=nom	<i>tajdîd bitaqat al ta'rif</i>	taḥdîd al- intâj
	Schème II taf'il: intensité de l'action ; "faire faire"	Dev= taf'il Det1=Ø/al N1=nom complément déverbal (objet de l'action)		taḥdîd al- intâj
	Schème III fi'âl sens faire un	Dev= fi'âl Det1=Ø/al N1=nom complément déverbal		qiyâs haḍihyat al-intâjijyat

	effort en direction de quelque chose	(objet de l'action)		
	Schème IV if'âl: sens factitif /causatif "faire faire"	Dev= if'âl Det1=Ø/al N1=nom complément déverbal (objet de l'action)		i'âdat tawzî al- dahî al- nâtij al- ijmâlî
	Schème V tafa'ul : sens réfléchi passif de la forme II	Dev= tafa'ul Det1=Ø/al N1=Nom complément déverbal (objet de l'action)		tamaluk wasa'il al- intâj
	Schème VII infi'âl : sens réfléchi passif de la forme I fa'ala insistance sur le passif correspond aux verbes pronominaux français	Dev= infi'âl. Det1=Ø/al N1=nom complément déverbal (objet de l'action)		inhi'âd al-ṭaqat al- intâjijyat
	Schème VIII tafa'ul : sens	Dev= tafa'ul Det1=Ø/al N1=nom		al-irtifâ'c al-intâjijyat

	réfléchi passif de la forme I fa'ala action du sujet a son profit	complément déverbal (objet de l'action)		aw inhifâḍ hi
	Schème IX : istifâl : sens factitif causatif "faire faire" c'est le réfléchi passif de la forme IV ' af'ala	Dev= istifâl Det1=Ø/al N1=nom complément déverbal (objet de l'action) Det1=Ø/al A1=adjectif qualificatif de N1		taḥdîd al- intâj
	Masdar forme non dérivée	Dev=déverbal Det1=Ø/al N1=Nom complément déverbal (objet de l'action)		al-nuḥḍ bi al-qitâc al-zirâ'i

Nous avons vu dans la deuxième partie de cette étude les différentes formes syntaxiques des structures annectives. Nous les avons répertoriées et nous avons remarqué que chaque structure annective, suivant la nature de ses constituants, avait une portée sémantique précise.

Nous avons constaté que par rapport aux structures annectives classiques, les structures annectives extraites d'un corpus économique sont assez complexes car elles se limitent très rarement à deux ou trois termes. De plus, la détermination par l'article semble également jouer un rôle important, en effet son absence ou sa présence dans l'une des structures influe sur le sens du descripteur. Cette détermination est importante surtout dans les structures impliquant comme

constituant un adjectif, celui-ci peut se rapporter à un déverbal ou à un nom comme dans l' exemple suivant :

la baisse de l'activité économique هبوط النشاط الاقتصادي

dont la structure est (N+((Det+N)+(Det+A))). Le parenthésage nous permet de constater ici que l'adjectif se rapporte au second terme de l'annexion.

Nous avons élaborés dans la deuxième partie de cette étude un lexique des modificateurs, quantificateurs et prépositions. Ce lexique nous a servi lors de l'extraction de descripteurs de notre corpus, car nous avons trouvé des structures qui font intervenir les prépositions "ل" et "من" :

La coordination de deux termes est très présente dans les descripteurs. Elle a pour but d'éviter la répétition de la tête de syntagme.

En ce qui concerne l'annexion à un déverbal, la langue de spécialité semble en faire grand usage. La majorité des descripteurs ont comme premier terme des noms de procès (masdar), ainsi les participes actif et passif sont moins présents. Le participe passif est surtout utilisé comme un adjectif. Le schème taf'îl "تفعيل" est le plus présent dans cette majorité de descripteurs. Il désigne le processus d'une action, la transformation de quelque chose :

la distribution du PNB توزيع الدخل القومي

dont la structure est (Dev+((Det+N)+(Det+A))).

Ces déverbaux sont très souvent des têtes de SN.

Nous remarquons, à travers notre tableau de modèles, la présence des catégories sémantiques. Ce sont en fait des représentations conceptuelles. En se référant à la relation prototypique, notion empruntée à la psychologie cognitive³,

³ on peut dire que le marteau est la représentation la plus typique de l'outil, c'est celui qui vient en premier à l'esprit pour cette catégorie.

nous avons trouvé les relations sémantiques les plus typiques et les plus représentatifs de la i'dâfat. Ce sont :

- la possession
- la forme et la matière
- la réformulation spécifique
- l'élatif
- le spécifique ou *tamiyz*
- l'annexion à un modificateur
- l'annexion à un déverbal

Nous remarquons également qu'à travers le traitement des descripteurs extraits d'un domaine de spécialité, nous avons trouvé d'autres types de modèles. Ces modèles sont propres à l'économie mais ils sont assez généraux pour être repris dans un autre domaine de spécialité :

- l'organisation,
- le processus,
- la méthode
- le résultat d'une action

Suite au traitement syntactico-sémantique nous avons remarqué les points suivants :

- les descripteurs complexes contiennent plus d'éléments que les modèles initiales de traitements
- le traitement syntactico-sémantique a permis la clarification quant à la relation sémantique entre les différents composants de ces descripteurs
- il s'avère que l'on peut limiter le nombre de composants d'un descripteur car dépassant cinq éléments ce n'est plus un descripteur complexe mais une phraséologie
- le traitement de ces descripteurs complexes et leur étude à travers un corpus a permis de voir l'émergence d'autres catégories sémantiques non présentes

chez les différents grammairiens classiques et modernes. Ces catégories sémantiques sont en réalité une représentation abstraite des concepts circonscrits au domaine étudié. Il est probable qu'en étudiant un corpus juridique ou médicale nous trouverions d'autres types de structures annectives et donc d'autres modèles qui viendraient s'ajouter aux modèles initiaux définis en fin de seconde partie de cette étude.

Nous avons vu dans la première partie de cette étude l'approche du groupe SYDO et celle de l'équipe SAMIA et le choix de leur analyse donnant la priorité à l'approche morphosyntaxique. Cette approche entre dans le cadre des techniques documentaires afin de déterminer les structures porteuses d'éléments informatifs.

Dans le cadre d'un système de recherche d'information (SRI) le problème linguistique touche aussi bien les textes écrits (documents textuels) que la formulation des requêtes, c'est-à-dire, les questions des usagers, demandeurs d'information.

L'objectif du modèle linguistique est de permettre l'identification des syntagmes nominaux, tout en mettant en évidence, à travers la structure syntaxique reconnue, les mots appartenant au lexique.

Les structures ainsi répertoriées ont permis d'identifier des SN et de déterminer la structure de ces syntagmes tout en mettant en évidence les relations entre leurs constituants. Ceci permet le stockage d'une représentation du syntagme nominal qui facilitera la recherche d'information.

Ces modèles de structures montrent le mécanisme de passage d'une logique intensionnelle (mots de la langue) vers une logique extensionnelle (mot à valeur référentielle). Les SN fonctionnent dans une logique extensionnelle car ils pointent sur des objets de la réalité extra-linguistique.

Le résultat de cette étude n'est pas une simple liste de syntagmes nominaux mais un lexique. Celui-ci contient des modèles de structures syntactico-sémantiques d'un type de descripteurs complexes (SN annexionnel). Les relations syntaxiques ainsi observées pourraient être mémorisées dans une base et servir au

processus de recherche d'information. Ces structures ont permis de définir la nature des relations syntaxiques des constituants d'un SN et ses emboîtements qui peuvent traduire des relations d'hypéronomie ou d'hyponomie.

CONCLUSION

CONCLUSION

Nous avons vu, dans la première partie de ce mémoire, l'apport de la linguistique dans l'analyse automatique en général et l'indexation automatique en particulier.

Nous avons constaté que le traitement linguistique automatique des textes numériques avait permis de concevoir des logiciels documentaires permettant le traitement du langage naturel et par ce biais, l'amélioration de l'indexation automatique et de la recherche sur texte intégral. Nous avons également remarqué que le traitement linguistique était la combinaison de différentes analyses (morphologique, syntaxique, sémantique, etc...).

Cependant ces modules d'analyse linguistique ne peuvent traiter de façon automatique certains termes ambigus (même si la combinaison de ces analyses permet de désambigüiser certaines expressions). Ces termes sont des descripteurs pluritermes qui peuvent être des mots composés, des expressions, etc. Nous nous intéressons aux structures annectives car elles sont les plus nombreuses dans les domaines scientifiques et techniques.

Ceci explique l'utilité d'une analyse linguistique préalable (deuxième partie de notre mémoire). Ce traitement linguistique nous a permis de cerner les différents types de catégories ou modèles de ces structures annectives dans les grammaires classiques .

Les différents lexiques (quantificateurs, modifieurs, prépositions) pourront servir à au traitement des mots vides. Ils permettront aussi de repérer plus facilement la portée syntactico-sémantique des structures annectives incluant un quantificateur ou une préposition par exemple.

Cette étude nous a permis d'élaborer l'essai d'une grammaire du SN pour un traitement à la fois morphosyntaxique et syntactico-sémantique des descripteurs extraits de notre corpus..

A la suite de ce traitement, nous avons trouvé d'autres catégories ou modèles de structure annective. Ces modèles sont issus d'un corpus économique, ils réfèrent à un domaine précis. Nous en avons élaboré à partir des modèles généraux et ceux extraits du corpus, un lexique des "méta-descripteurs". Ce lexique représente la structure, à la fois syntaxique et sémantique, d'un certain type de descripteur complexe. C'est en quelque sorte le prototype des catégories sémantiques de l'annexion.

Ce lexique des méta-descripteurs peut s'appliquer à d'autres domaines que l'économie. Il va de soi que l'on pourra trouver d'autres modèles en étudiant un corpus d'une autre spécialité.

La présence dans ce lexique des métadescripteurs des déverbaux démontrent qu'on peut également rechercher l'information sur des unités déverbales et pas uniquement nominales.

Ainsi devant le développement croissant d'internet, d'intranet et du travail partagé en réseau, le traitement automatique de l'information devient crucial. La recherche de l'information pertinente passe par des outils performants. Cela ne peut se faire sans un traitement à la base : il faut passer par des modèles linguistiques qui traite cette information de manière à pouvoir la retrouver. La sémantique est un domaine encore mal délimitée pour la langue arabe. Cependant, grâce aux grammaires lexiques, les modèles, qui constuisent des dépendances sémantiques entre les termes à partir des relations syntaxiques, peuvent contribuer à l'avancement des travaux sur le traitement automatique de l'arabe.

BIBLIOGRAPHIE

BIBLIOGRAPHIE

I-MONOGRAPHIE

a) Monographie arabe

- *Al-mafâhîm al-iqtisâdiyât aljahuriyat al-‘irâqiyat*- Wizarat at-t aţ-ţaqafat wal-
i‘lâm da‘irat al-i‘lâm al-dâhilî al-‘âmat -as-sasilat al-iqtisâdiyât raqm 7, 88p
- *Šarḥ al-mufaṣal li-ibn Ya‘îš bin Ya‘îš al-naḥwî*, 3^e partie Librairie al-mutanabî
Le Caire, p. 118,
- *Šarḥ ibn ‘aqîl*, 3^e partie, Dar al-fikr p. 43
- *Al-Munjid fî al-luġa wal-i‘lâm* Dâr al-Mašraq, Beyrouth Liban, 1992, p. 634

b) Monographie occidentale

- [ABEILLE, 93] A. ABEILLE, *Les nouvelles syntaxes, grammaires d'unification
et analyse du français*, Paris : Editions Armand Collin, collection
Linguistique, 1993, 327p.
- [ALBERTINI, 91] A. ALBERTINI, A. SILEM, *Encyclopedie de l'économie et de
gestion*, Paris : Edition Hachette, 1991, 592p.
- [BLACHERE, 52] R. BLACHERE, *Grammaire de l'arabe classique*, Paris :
Editions GP Maisonneuve, 1952, 508p.
- [BLANQUET, 94] M.F. BLANQUET, *Intelligence artificielle et système
d'information*, Collection Systèmes d'information et nouvelles
technologies, Paris : ESF Editeurs, 1994, 269p.
- [BRETON, 82] J. BRETON, et al. , *Guide pratique d'indexation en vocabulaire
libre*, Quebec : Edition Microfor, 1982, 219p.
- [CHAUMIER, 82] J. CHAUMIER, *Analyse et langages documentaires : le
traitement linguistique de l'information documentaire*, Paris :
Entreprise moderne d'Édition, 1982, 186p.
- [COHEN, 70] D. COHEN , *Etude de linguistique sémitique et arabe : Essai d'une
analyse automatique de l'arabe*, Paris : Mouton , 1970, 178p (pp 49-
79)

- [FASSI-FEHRI, 82] A. FASSI-FEHRI, *La linguistique arabe, formes et interprétation*, Rabat, 1982, 341p.
- [FUCHS, 93] C. FUCHS et al. , *Linguistique et traitements automatiques des langues*, Paris : Editions Hachette Supérieur, collection HU linguistique, 1993, p.110
- [HODGE, 92] G. M. HODGE, *Automated support to indexing*, 1992 NFAIS Report Series tome 3, The National Federation of Abstract and Information Services, Philadelphia PA USA, 1992, 176p.
- [IBN MALIK] Ibn Malik, *Alfiya*, traduction par G. GOGUYER, Librairie du Liban Publishers, 199, p.
- [LECONTE ,70] G. LECONTE, A. GHEIDIRA, *Méthode d'arabe littéral*, premier et deuxième livre, deuxième édition, Paris : Editions Klincksieck, 1970
- [MANIEZ, 87] J. MANIEZ, *Les langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires*, Paris : Les Editions d'Organisation, 1987, 277p.
- [NEET, 89] H. E. NEET, *A la recherche du mot-clé, analyse documentaire et indexation alphabétique*, les Editions IES (Instituts d'Etudes Sociales), Collection "Les cours de l'IES", n°2, Genève, 1989, 187p.
- [NEYRENEUF, 96] M. NEYRENEUF, G. AL HAKKAK, *Grammaire active de l'arabe*, Librairie générale française, collection Les langues modernes, 1996, p.
- [ROMAN, 91] A. ROMAN, *Grammaire de l'arabe*, Paris : Edition PUF, collection Que sais-je, 1991, 125p.
- [RICHTER, 81] N. RICHTER, *Pratique de l'indexation*, Bibliotheque de l'Université du Maine, 1981, 124p.
- [SILBERZTEIN,93] M. SILBERZTEIN, *Dictionnaires électroniques et analyse automatique de textes*, Le système INTEX, Edition MASSON, 1993, 233p.
- [SILEM, 92] A. SILEM, *Lexique d'économie*, Paris : Edition Dalloz, 1992, 576p.

[VAN SLYPE, 87] G. VAN SLYPE, *Les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires*, Paris, les Editions d'Organisation, 1987,p.

[WRIGHT, 81] W. WRIGHT, *A grammar of the Arabic language*, 3rd Edition, Librairie du Liban, 1981, 767p.

II-PUBLICATIONS EN SERIE

[ANDREEWSKI, 96] A. ANDREEWSKI, *Les systèmes documentaires SPIRIT et MICRO-MIND*, Bulletin du Centre de Hautes Etudes Internationales d'Informatique Documentaire (BCHEIID) n° 61, 1996, pp, 29-41

[BERRENDONNER, 95] A. BERRENDONNER, *Quelques notions utiles à la sémantique des descripteurs nominaux*, TRANEL (Travaux neuchâtelois de linguistique) 1995, n° 23, pp. 9-39

[BOUCHÉ ,89] R. BOUCHÉ, *Le syntagme nominal, une nouvelle approche des BDD textuelles*, META XXXIV,3, 1989, p. 429-433

[BOUCHÉ, 90], R. BOUCHÉ, S. LAINE, J.P. METZGER, *Extraction de connaissances à partir d'une collection de documents*, In Tools of Knowledge Organization and the human interface, Congrès organisé par l'ISKO (International Society of Knowledge) 14-17 aout 1990

[BOUCHÉ, 91] R. BOUCHÉ, N. GERMAIN, *Bibliométrie, infométrie, et analyse automatique de documents écrits*, Bulletin de la Société Française de Bibliométrie Appliquée (SFBA) et de l'Association pour le Développement de la Scientométrie et de la Technométrie (ADEST) 1991, n 9, p. 352-366

[CATTENAT, 93] A. CATTENAT, PAUL G. *Thésaurus ou réseaux sémantiques pour l'aide à l'interrogation de base de données textuelles par l'utilisateur final*. IDT 93. 6^e congrès, 1993, pp.138-142

[CHAUMIER, 90] J. CHAUMIER, M. DEJEAN, *L'indexation documentaire, de l'analyse conceptuelle à l'analyse morphosyntaxique*, Documentaliste, vol. 27, n°6, 1990, pp.275-279

- [CHAUMIER, 92] J. CHAUMIER, M. DEJEAN, *L'indexation assistée par ordinateur : principes et méthodes*, Documentaliste, vol. 29, n°1, 1992, pp.3-6
- [CORET, 94] A. CORET, B. MENON, D. SCHIBLER, C. TERRASSE, *Un système d'indexation structuré à l'INIST, Bilan d'une étude préalable*, Documentaliste, vol. 31, n°3, 1994, pp.3-6
- [DICHY, 96] J. DICHY, *Pour une lexicomatique de l'arabe : l'unité simple et l'inventaire des spécificateurs du domaine du mot*, in *Lexicomatique et Dictionnaire*, Actes des Journées de Lyon, collection "Actualités scientifiques" (AUPELF-UREF, réseau "Lexicologie, Terminologie et Traduction"), Québec : DPLU et Paris : Hachette, juin 1996,
- [DICHY, 89] J. DICHY, M. HASSOUN, *Simulation de modèles linguistiques et EAO de l'arabe*, Travaux SAMIA1, CILF, 1989, p.45
- [FLUHR, 93] C. FLUHR, *Linguistique et documentation*, ENSSIB, journée d'étude : la Recherche en Sciences de l'information (III), 23 novembre 1993, Villeurbanne
- [GACHOT, 95] [GACHOT, 95] GACHOT, Isabelle, *Linguistique + statistique + informatique = indexation automatique ?*, Archimag, n°84, mai 1995, pp.34-37
- [GUILBAUD, 95] GUILBAUD Elisabeth, *Comment indexer l'image ?*, Archimag, n°86, juillet-août 1995, pp.41-43
- [GSI-ERLI, 89] GSI-ERLI, *Recueil d'articles*, ENSSIB, novembre 1989
- [HABERT, 93] B. HABERT, C. JACQUEMIN, *Noms composés, termes, dénominations complexes : problématiques linguistiques et traitements automatiques*, TAL, vol. 34, n°2, 1993
- [LAINE-CRUZEL, 92] S. LAINE-CRUZEL, *Vers un nouveau modèle prenant en compte le profil des utilisateurs*, ENSSIB, journée d'étude : la Recherche en Sciences de l'information (III), 23 novembre 1993, Villeurbanne
- [LE GUERN, 84] M. LE GUERN, *Les descripteurs d'un système d'information documentaire : essai de définition*, in acte du Colloque Traitement

automatique des langues naturelles de systèmes documentaires,
Condenser, suppl, Université de Clermont Ferrant, 1984.

- [LE GUERN, 91] M. LE GUERN, *Un analyseur morpho-syntaxique pour l'indexation automatique*, Le Français moderne, 1991, tome LIX, n°1, pp.22-35
- [LONG, 80] B. LONG, *Linguistique et indexation*, Documentaliste-Science de l'information, vol.17, n°3, mai-juin 1980, p.100
- [ROMAN, 85] A. ROMAN, *Les relations syntaxiques*, in cercle linguistique d'Aix en Provence, travaux Aix en Provence, 1985
- [POEHL, 92] K. POEHL, *Le syntagme nominal comme outil de l'indexation automatique, le cas de la langue allemande*, Edition les Cahiers des Sciences de l'information et Communication, DEA Rhone-Alpes, n°2, 1992, pp. 3-21
- [PRADILLA, 82] M. PRADILLA, *Recherche de descripteurs en indexation automatique de documents*, Thèse de Doctorat mathématique appliquées aux sciences sociales, 1982, Les cahiers du CRISS n°4, p.29
- [VANDEUR, 90] VANDEUR, Marc, *Approche de l'élaboration d'un système expert d'aide à l'indexation*, Cahier de la Documentation-Bladen Voor de Documentatie, n°4, 1990, pp.75-90

III-MEMOIRES, NOTES DE SYNTHÈSE ET THESEES

a) Mémoires et notes de synthèse

- [ABBAS, 92] W. ABBAS, *Définition des unités linguistiques intervenant dans l'indexation automatique en arabe*, Mémoire de DEA en Sciences de l'Information et Communication, ENSSIB, 1992, 70p
- [ANTONIADIS, 80] G. ANTONIADIS, *Recherche des descripteurs complexes pour l'indexation automatique des documents*, Mémoire de DEA en Sciences de l'Information et Communication, juin 1980, 40p.
- [EL HACHANI, 97] M. EL HACHANI, *L'indexation automatique*, note de synthèse, ENSSIB, 1997, 31p.

[KURAMOTO, 95] H. KURAMOTO, *Maquette d'un système de recherche d'information en utilisant des syntagmes nominaux*, Mémoire de DEA en Sciences de l'Information et Communication, ENSSIB, 1995,

[MINKO-MI-NSEME, 96] S.A. MINKO-MI-NSEME, *Les phrasèmes dans le sens de Mel'cuk, en arabe*, Mémoire de DEA Langues et civilisations étrangères mention LTMT, Université Lumière Lyon2, 1996, 86p.

b) Thèses

[ABBAS, 98] W. ABBAS, *Définition et description des unités linguistiques intervenant dans l'indexation automatique des textes en arabe*, Thèse de doctorat Sciences de l'information et communication, Université Lumière Lyon2, 1998

[ABU ALCHAY, 88] N. ABU ALCHAY, *Un système expert pour l'analyse et la synthèse des verbes arabes dans un cadre d'enseignement assisté par ordinateur*, Thèse de doctorat Université Claude Bernard Lyon1, 1988

[EZZAHID, 96] S.S. EZZAHID, *Les entrées d'une BDD arabe*, Thèse doctorat 3^e cycle Sciences du langage, Université Lumière Lyon2, 1996

[HASSOUN, 82] M. HASSOUN, *Système d'analyse morphologique automatique de la langue arabe*, Thèse de doctorat en mathématique-informatique, 3^e cycle, 1982

[HASSOUN, 87] M. HASSOUN, *Conception d'un dictionnaire pour le traitement automatique de l'arabe dans différents contextes d'application*, Thèse pour le doctorat d'Etat, Université Lyon1, 1987

[LAINE, 82] S. LAINE, *Extraction et sélection de descripteurs complexes dans un ensemble de textes pour leur indexation automatique*, Thèse de doctorat d'ingénieur en Sciences et Techniques communes : sciences de l'information et documentation, Lyon1, 1982

[LAROUK, 93] O. LAROUK, *Extraction des connaissances à partir de documents textuels : traitement automatique de la coordination (connecteurs et ponctuation)*, Thèse de doctorat spécialité informatique, Lyon1, 1993.

[METZGER, 88] J.P. METZGER, *Syntagmes nominaux et information textuelle*,

Thèse d'Etat es sciences, Université Lyon1, 1988

[MUSTAFA ELHADI, 89] W. MUSTAFA ELHADI, *La terminologie arabe des*

télécommunications : Faits de variations, Thèse de doctorat,

Université Lyon2, 1989

ANNEXES

TRANSCRIPTION ARABICA

Alphabet arabe	transcription
ء	'
أ	â
ب	b
ت	t
ة	at
ث	<u>t</u>
ج	j
ح	ḥ
خ	ḫ
د	d
ذ	<u>d</u>
ر	r
س	s
ش	š
ص	ṣ
ض	ḍ
ط	ṭ
ظ	ẓ
ع	c
غ	g
ف	f
ق	q
ك	k
ل	l
م	m
ن	n
ه	h
و	w, u
ي	y, î
ى	â

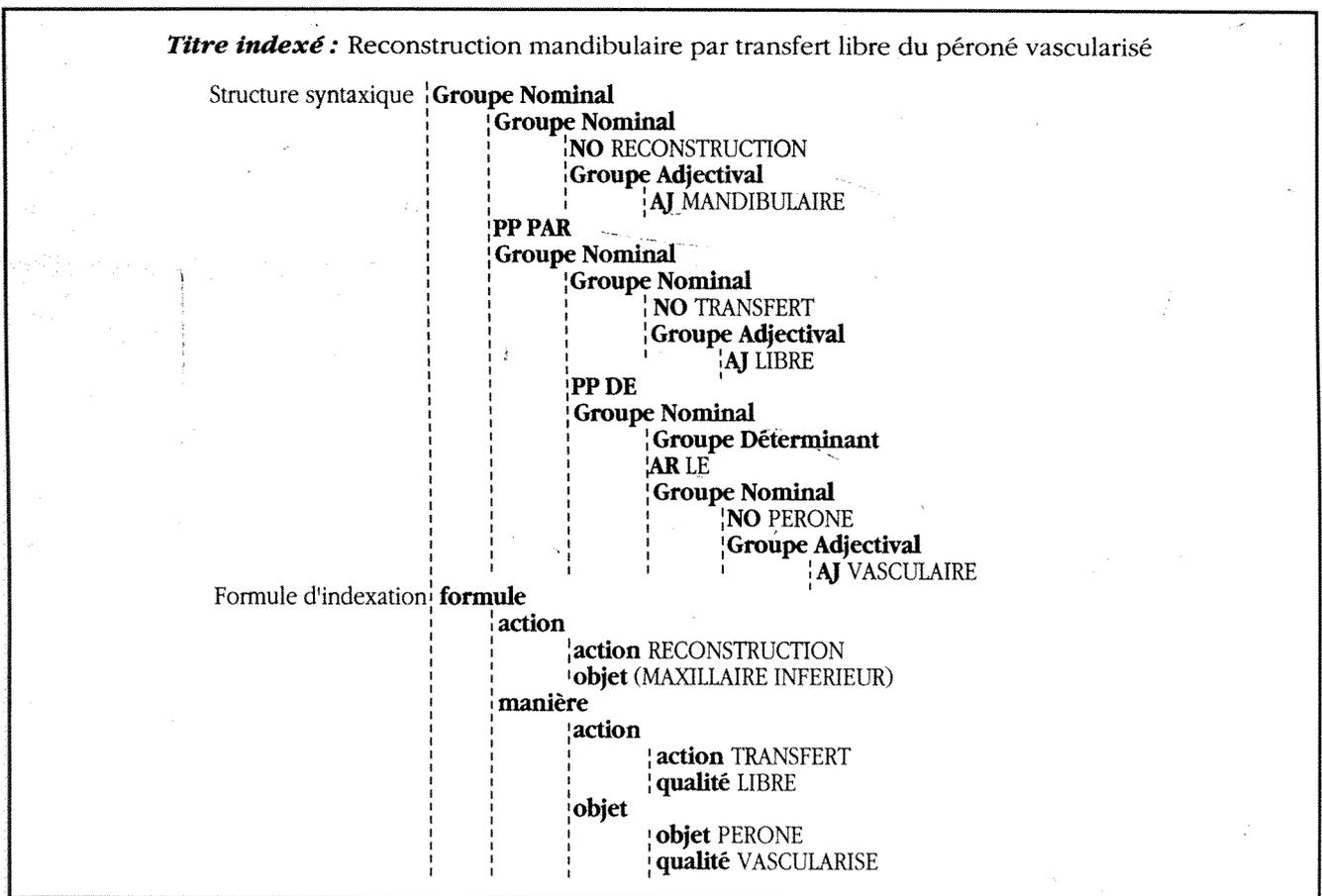


Figure 2 : Exemples de structures obtenues par l'analyse linguistique et sémantique

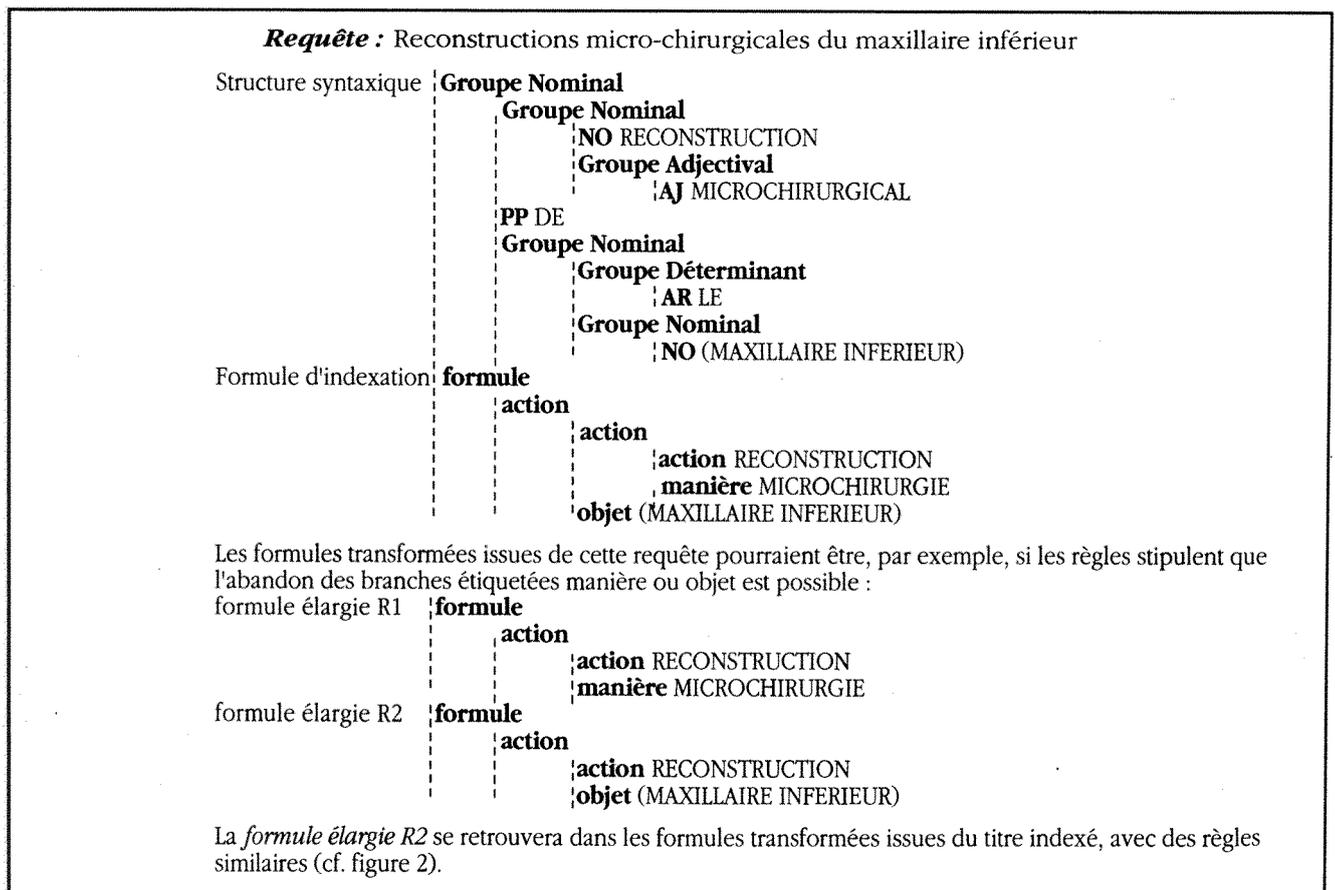


Figure 2 bis : Exemple de traitement d'une requête

Economique

◆ SPIRIT l'est à plusieurs titres :

- . Par sa simplicité d'utilisation et son automatisme, il ne nécessite pas la présence d'un spécialiste et permet un gain de temps important.
- . Par sa capacité à rendre l'information accessible, il améliore la productivité de ceux qui en font usage.
- . Par sa mise en oeuvre immédiate sans travail préparatoire ni thesaurus.

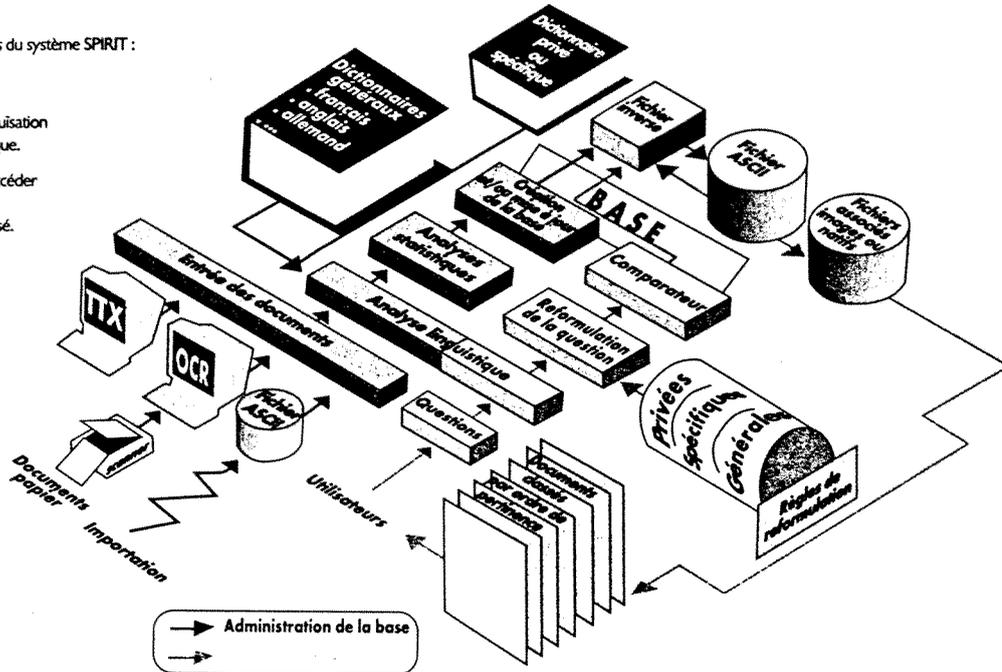
Quatre modules constituent les fondements du système SPIRIT :

■ Le module linguistique réalise la segmentation du texte, la désambiguïsation entre élément fonctionnel et sémantique.

Le module fichier inverse permet d'accéder aux documents à partir des concepts pondérés, et ce dans un temps optimisé.

■ Le module analyse statistique permet d'affecter à chaque concept une fonction de poids informationnel et d'ordonner les réponses en fonction de leur pertinence.

■ Le module comparateur fait les intersections ensemblistes entre les concepts qu'il a sélectionnés dans la question et le fichier inverse.



Universel

◆ SPIRIT fonctionne sous les systèmes d'exploitation MVS/TSO, VM/CMS et VAX/VMS. Les versions UNIX, OS/2, MS-DOS/Windows 3.1 également disponibles sont réalisées selon l'architecture du modèle client/serveur. Une API (Application Programm Interface) permet de constituer toute application spécifique à l'environnement de l'utilisateur.

Un formidable potentiel

◆ Initialement conçu comme outil de recherche d'information, SPIRIT vise maintenant une multitude d'utilisations nouvelles :

- . Aide à la décision et au diagnostic
- . Bureautique intelligente
- . Hypertexte dynamique
- . Construction automatique de graphes de concepts
- . Diffusion d'informations par le contenu
- . Intégration dans des systèmes de GED et de Workflow

◆ SPIRIT est un puissant générateur d'applications pour administrer et retrouver l'information; il fonctionne sur poste de travail personnel comme sur de grands centres de traitement de l'information.

L'originalité du traitement linguistique de SPIRIT offre des capacités d'adaptation immédiate aux exigences propres à chaque domaine d'activité.

UN SYSTEME D'INDEXATION STRUCTUREE A L'INIST

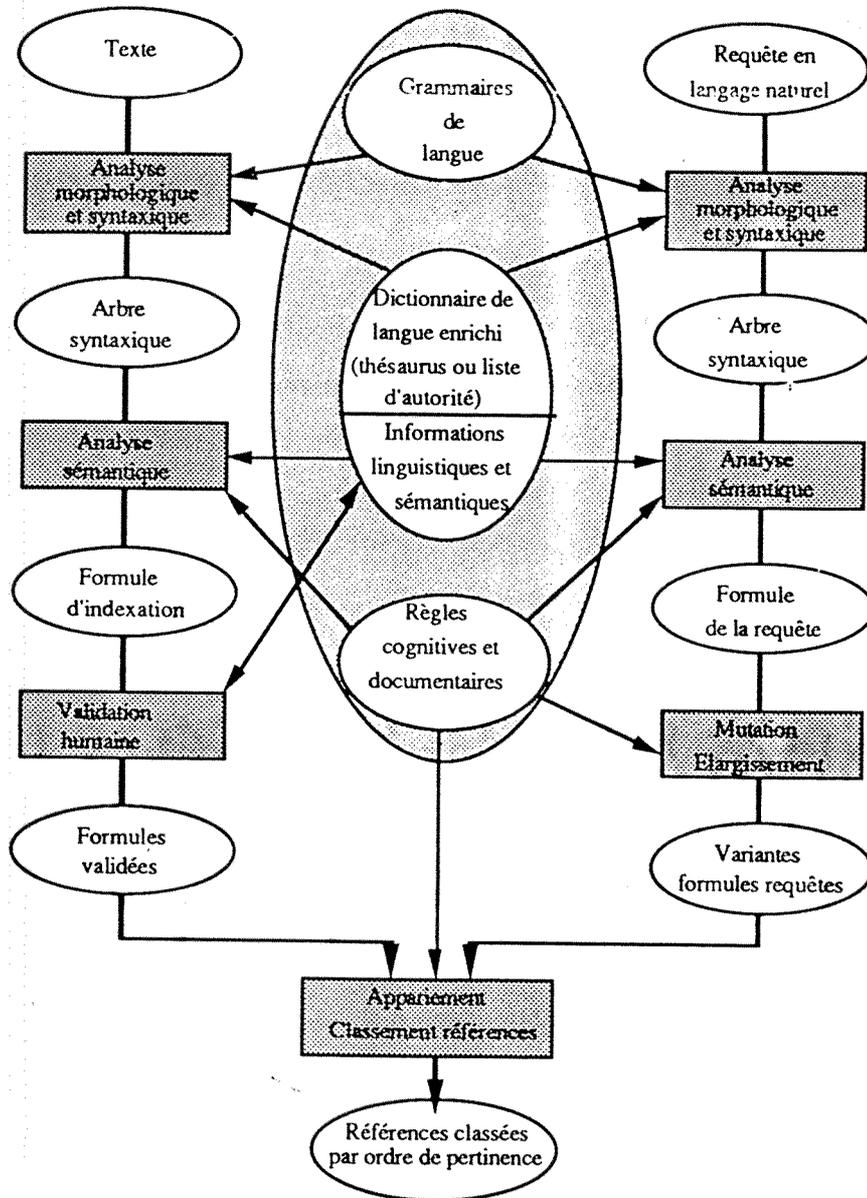
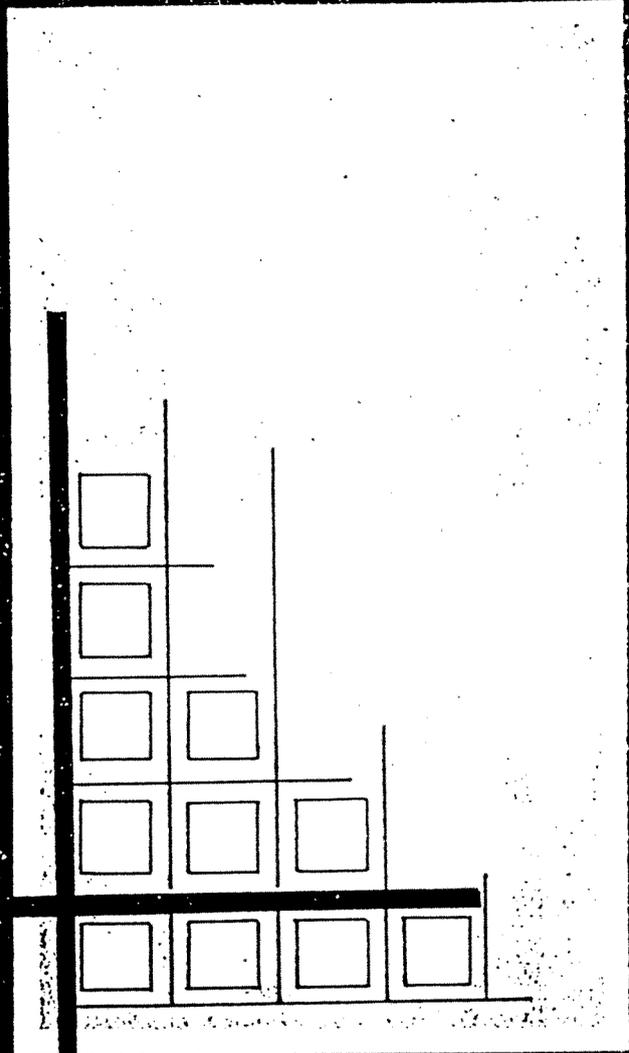


Figure 1 : Chaîne de traitement très schématisée de l'indexation structurée

مفاهيم
اقتصادية



Textes extraits du Corpus
(les descripteurs extraits de
ce corpus sont soulignés)

الجمهورية العراقية
وزارة الثقافة والاعلام - دائرة الاعلام الدائري العامة
السلسلة الاقتصادية رقم / ٦

مفاهيم إقتصادية

القومي عن طريق الضرائب والرسوم واسعار مبيعات العديد من سلع الخدمات .

وبالطبع فإن جميع الدخول الاولية والثانوية ينطوي على حساب مكرر مبالغ فيه وهو ما درج عليه الاقتصاديون البورجوازيون الذين يدخلون في الدخل القومي كل دخل سواء جاء من الانتاج المادي او من الميدان غير الانتاجي .

ويطلق لفظ القومي على هذا الدخل للإشارة الى الدخول الصافية المتولدة في داخل البلد المعين والعائدة الى رعاياه فقط وكذلك ما يحصل عليه رعاياه في البلدان الاجنبية ، بينما يقال الدخل المحلي للدلالة على مجموع صافي قيمة الانتاج في البلد المعين العائد لرعاياه ولغيرهم .

[٢ - الناتج الاجمالي]

الانتاج كما هو معروف ، عملية يؤثر فيها العمال بواسطة ادوات العمل في مواضع العمل وهي مواد الخام والارض وغيرها بهدف تحويلها الى خيرات مادية متنوعة . ومجموع ما ينتجه المجتمع من هذه الخيرات المادية خلال العام يطلق عليه المنتوج او الناتج الاجتماعي الاجمالي او احياناً الناتج الاجمالي .

ويتكون الناتج الاجتماعي الاجمالي في ميدان الانتاج المادي وحده ، اي في ميدان الصناعة والزراعة والبناء . وبالطبع فإن عملية الانتاج وتحقيقه تحتاج الى عمليات مكتملة من اجل تحقيق القيمة الاستعمالية للمنتج كالنقل والمواصلات والتخزين والتجارة . ان هذه الفروع لا تخلق بحد ذاتها قياً استعمالية جديدة ، لكنها ضرورية لتحقيق هذه القيمة ، وبالتالي

فأنها تزيد من قيمة المنتجات . وبالطبع فإنه كلما نما المنتوج الاجتماعي الاجمالي تطور المجتمع . لهذا فحين تتركز ثورة السابع عشر من تموز على إيجاد المستلزمات المادية للتهوض بالانتاج ، فإنها توفر الاساس لرفاه المجتمع .

ويضاف الى مصطلح الناتج الاجمالي عادة اما لفظة المحلي او القومي فحين يقال الناتج المحلي الاجمالي يراد به مجموع المنتجات التي انتجت من قبل المنتجين المقيمين في البلد المعين (بما فيهم أولئك الذين هم ليسوا من رعاياه . اما الناتج القومي الاجمالي ، فإنه يؤلف بمجموع المنتجات التي أنتجها رعايا البلد المعين في داخله وخارجه اي انه قد استبعد من الناتج الاجمالي المحلي ما انتجه المقيمون فيه من غير رعاياه ومضافاً اليه ما انتجه مواطنوه في الخارج) .

وبالطبع فإنه اكبر من الدخل الصافي الذي يتألف في الحالتين مما يتبقى بعد طرح قيمة وسائل الانتاج المستهلكة خلال العام . ان الشرط الذي لا غنى عنه بالنسبة الى تجديد الانتاج في المجتمع الاشتراكي يتضمن ثلاثة جوانب :

١ - التعويض عن وسائل الانتاج التي استهلكت في مجرى النشاط الاقتصادي .

٢ - توفير كميات اضافية من وسائل الانتاج الضروري تسمح بالتوسيع اللاحق للانتاج وتوفير المخزين الاحتياطي الضروري .

٣ - زيادة كمية سلع الاستهلاك باستمرار لمواجهة الحاجات المتنامية لآبناء الشعب .

تعني كلمة «الانتاج» كل ما ينتجه النشاط العملي للانسان من اجل صنع الخيرات المادية ، ضمن التشكيلة الاجتماعية التي يعيش فيها . كالمأكولات والالبسة والمساكن ، والآلات والمكائن وكل الحاجيات التي ينتجها العمل البشري .

ويختلف هدف الانتاج بين مجتمع وآخر ، اي بين تشكيلة اجتماعية واخرى ، فالهدف المباشر بالنسبة للأعمالية ، هو الحصول على اعلی الارباح عن طريق استغلال قوة العمل . اما في المجتمع الاشتراكي حيث الملكية الجماعية لوسائل الانتاج الاساسية ، فإن هدف الانتاج هو تلبية حاجات المجتمع ، اي حاجات جميع اعضاء المجتمع ، ولهذا فإن المجتمع الاشتراكي يعمل على توسيع الانتاج باستمرار ويزيد من اتقانه من اجل رفع الرفاهية المادية والمستوى الثقافي للجماهير الشعب باستمرار .

أما الانتاجية فهي معدل ما تنتجه وحدة انتاجية معينة ، في وحدة زمنية محددة ، ساعة مثلاً ، فارتفاع الانتاجية او انخفاضها تعني في النهاية ، زيادة او نقص الانتاج النهائي الذي تتوقف عليه زيادة الرفاه المادي للمجتمع .

ولهذا تهتم المؤسسة المعنية بمراقبة مؤشر الانتاجية ، لكي تراقب تنفيذ خططها الانتاجية ومعالجة اي معوق يؤثر على الانتاج النهائي للوحدة الانتاجية المعنية .

تعني انتاجية العمل نسبة كمية الانتاج الى عدد العمال المشغولين (او عدد ساعات العمل) . وهذه الانتاجية تعتبر من حيث الجوهر كمقياس لكفاية المشروع الانتاجي ، التي يعبر عنها عادة بمقدار الناتج للعامل الواحد او للساعة الواحدة ، او اية وحدة زمنية اخرى ، ولكنها لا تعد مقياساً للمجهود البشري المبذول ، لان حجم القوة الانتاجية في ظل ظروف معينة يتوقف على كيفية تجاوز العمل البشري والانتفاع به وعلى الجهود التي يقدمها العمال انفسهم . ولكن قوة الانتاجية في الزراعة تناس غالباً بالمشارت من الارض ، ولو ان ناتج المشارة الواحدة من الارض لا يتوقف على خصوبة التربة فحسب ، بل يتوقف ايضاً على كيفية قيام المزارع باستغلال التربة .

ومن العوامل التي تزيد من قوة الانتاجية استخدام الآلات الحديثة - المكننة - والاكتشافات الجديدة والتخصص - وادخال التحسينات المستمرة على الآلات وطرق الانتاج والكفاية الادارية وكفاية العامل نفسه

ولمعرفة التغيرات التي تطرأ على القوة الانتاجية للعمل ، فلا بد من مقارنة الانتاج لصناعة معينة وفي فترة معينة بانتاجها في فترة اخرى ، ويمكن قياس هذه الانتاجية ، بمقدار ما ينتج في وحدة من العمل ، كأن يقال مثلاً انها تساوي كذا وحدة من سلعة ما في وحدة زمنية معينة . وعلى أية حال فقد جرت العادة على مقارنة القوة الانتاجية من سنة لآخرى .