

M 1992 DEA 23

**ECOLE NATIONALE SUPERIEURE
DES SCIENCES DE L'INFORMATION
ET DES BIBLIOTHEQUES**

DEA en sciences de l'information
et de la communication

OPTION 3: Systèmes d'information documentaire

MEMOIRE DE DEA

**DEFINITION DES UNITES
LINGUISTIQUES INTERVENANT DANS
L'INDEXATION AUTOMATIQUE EN
ARABE**

WIGDAN ABBAS

Sous la direction de MOHAMED HASSOUN,
(CERSI-ENSSIB)

1992

A MA MERE ET A MON PERE

A MES SOEURS ET A MES FRERES

REMERCIEMENTS :

J' exprime ma reconnaissance à Monsieur Mohamed HASSOUN qui m'a constamment suivie pour l'élaboration de ce travail.

Je tiens à remercier également Monsieur Joseph DICHY qui a accepté de relire mon texte et m'a conseillée utilement.

Je remercie Messieurs Richard BOUCHE et Michel LE GUERN pour leurs précieux conseils .

Je remercie également, Mademoiselle Widad MUSTAPHA ELHADI et Monsieur Omar LAROUK, pour leurs disponibilités et leurs différents conseils.

Ma sympathie s'adresse à , Josiane PEYRON, Vincent PEYRON, Nadia, Margarita, Ion, Malik et Nabil pour leurs soutiens respectifs.

RESUME :

Cette étude porte sur l'analyse morpho-syntaxique des unités linguistiques pouvant servir de descripteurs à des fins d'indexation automatique en arabe ; ainsi, elle détermine le statut du syntagme nominal, tente d'élaborer une grammaire de sa réécriture et examine les méthodes de repérage des meilleurs descripteurs possibles.

MOTS CLES :

analyse morpho-syntaxique, unités linguistiques, arabe, indexation automatique, syntagme nominal, descripteur, grammaire.

ABSTRACT :

This study deals with the morphological syntactical analysis of linguistic units capable of being used as descriptors with a goal towards automatic indexing in arabic; it thus determines the status of the nominal syntagm, endeavours to construct a grammar system of its revised version and examines the ways pointing out possible descriptors.

KEY WORDS :

morphological syntactical analysis, automatic indexing, linguistic units, nominal syntagm, descriptor, grammar, Arabic.

TABLE DES MATIERES

INTRODUCTION	6
1. L'INDEXATION AUTOMATIQUE	9
2. SYNTAGME NOMINAL ET DESCRIPTEUR	15
2.1 DEFINITION DU DESCRIPTEUR	16
2.2 DESCRIPTION DU SYNTAGME NOMINAL	18
3. QUELQUES TRAVAUX SUR LE TRAITEMENT AUTOMATIQUE DE L'ARABE REALISES EN FRANCE	21
4. LA CONSTRUCTION DU SYNTAGME NOMINAL DANS L'ORGANISATION DE LA LANGUE	27
5. LES PROBLEMES DE L'ANALYSE LINGUISTIQUE DE LA LANGUE NATURELLE	35
6. LES CATEGORIES GRAMMATICALES POUVANT SERVIR POUR LA RECONNAISSANCE DU SYNTAGME NOMINAL.....	40
6.1. NOMS COMMUNS	40
6.2. NOMS PROPRES	41
6.3. MASDARS	42
6.4. ADJECTIFS	43
6.5. LES PREDETERMINANTS	44
6.6. PRONOMS	47
6.7. COORDONNANTS	48
6.8. PREPOSITIONS	49
6.9. PSEUDO-CATEGORIE : ADVERBES	49
7. LES REGLES DE REECRITURE DU SYNTAGME NOMINAL EN ARABE	50
CONCLUSION	60
BIBLIOGRAPHIE	63

I. INTRODUCTION

Nous tenterons dans le cadre de ce travail une première définition des unités linguistiques de la langue arabe qu'il convient d'isoler pour l'indexation automatique. De ce fait, il est donc nécessaire de procéder à l'analyse morpho-syntaxique de ces unités. L'analyse morpho-syntaxique dont il sera question ici devrait permettre de repérer les syntagmes nominaux dans un texte.

On examinera tout d'abord la notion de l'indexation automatique au chapitre premier, on analysera ensuite la signification du descripteur et du syntagme nominal (qu'on désignera désormais par SN) au deuxième chapitre.

L'objet du chapitre trois sera l'analyse de quelques travaux sur le traitement automatique de la langue arabe.

Etant donné que la construction du SN est importante pour notre étude, on examinera les phénomènes morpho-syntaxiques de la langue au chapitre quatrième. On s'intéressera à la relation entre l'énoncé et le système de la langue lui-même, c'est-à-dire aux fonctions des termes du point de vue syntaxique, et à leurs marques morphologiques, d'où la nécessité de l'analyse morpho-syntaxique. Nous étudierons la notion de schème et de racine et on examinera les structures linguistiques des constituants de la phrase arabe ainsi que le contenu logico-sémantique du discours écrit en arabe.

Les thèmes et les rhèmes (topics and comments), d'un grand nombre de discours écrits en français sont

essentiellement véhiculés par des SN. Ces SN font ainsi référence à la réalité extra-linguistique. Nous verrons ce qu'il en est pour l'arabe.

L'objet de chapitre cinq sera l'analyse de problèmes linguistiques de la langue naturelle. On examinera surtout les exigences que demandent l'analyse linguistique.

A partir d'une recherche d'éléments pertinents de la langue, chapitre sixième, on essaiera de définir les unités linguistiques qui seront susceptibles d'être classées comme descripteurs.

Nous terminerons par un essai d'élaboration de règle de réécriture permettant d'isoler le SN au chapitre septième.

Dans une première étape nous analysons, selon les méthodes linguistiques, les phrases d'un texte donné en essayant d'appliquer les règles d'écriture que nous élaborons. Dans une étape ultérieure de recherche interviendrait l'application de ces règles sur un modèle informatique au moyen d'un analyseur morpho-syntaxique.

Il ne s'agit pas d'écrire une grammaire qui rende compte de tous les phénomènes de la langue, mais d'une grammaire adaptée aux objectifs visés, en l'occurrence l'indexation automatique.

Il n'existe pas, à notre connaissance, de travaux dans ce domaine se rapportant à l'arabe. Cela s'explique peut-être, d'une part parce que peu de linguistes s'intéressent à

la reconnaissance automatique de la langue arabe écrite, et d'autre part parce que les informaticiens, sans l'aide de linguistes, ont tendance à formaliser la langue naturelle pour pouvoir la traiter avec des outils développés pour les langages de programmation, plutôt qu'à considérer cette dernière comme étant un domaine d'application différent de celui des langages de programmation.

Si notre travail se limite aux SN, c'est parce que nous voulons examiner les unités linguistiques qui servent le plus à l'indexation des documents. Néanmoins, nous allons passer par un essai d'analyse d'ensemble. En effet, l'analyse syntaxique des SN passe par la prise en considération de tous les phénomènes linguistiques qui affectent l'ensemble de la langue considérée.

"l'indexation automatique, c'est l'opération consistant 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 et conceptuellement voisins pour en faire des critères incorporés dans le fichier de recherche et utilisables pour retrouver le document."¹

L'opération de l'indexation peut se décomposer en trois étapes :

- appréhender le contenu du document.
- extraire du contenu du document les notions jugées les plus importantes.
- traduire les notions en terme de descripteurs.

Indexer un document c'est donc lui donner une représentation facilitant l'accès à l'information par son contenu.

Lors de l'indexation automatique des documents écrits en français, deux approches sont souvent envisagées :

- une approche statistique qui consiste à définir un modèle probabiliste des occurrences des mots dans un document

1. Van Slype, Georges, les langages d'indexation : conception, construction et utilisation dans les systèmes documentaires, les éditions d'organisation, Paris, 1987.

; le but est de savoir si le comportement statistique de ces mots permet leur participation à décrire le contenu;

- une autre approche qui tient compte de l'importance des problèmes linguistiques surtout le phénomène de référence à la réalité extra-linguistique, cette approche a été suivie par le Groupe SYDO déterminant la réflexion sur le statut du descripteur.

STRATEGIE DU GROUPE SYDO-LYON :

L'équipe SYDO-LYON² s'est beaucoup intéressée à l'analyse morpho-syntaxique de la langue française et à la représentation des textes écrits en français en vue de la construction d'une base de données textuelle. Les chercheurs de l'équipe ont conçu une grammaire de reconnaissance des SN. Cette grammaire s'inspire des grammaires génératives mais elle s'en différencie dans la mesure où c'est une grammaire de reconnaissance et non pas de génération de phrases.

La solution adoptée pour l'indexation automatique par le Groupe SYDO est la constitution d'un analyseur morpho-syntaxique permettant d'extraire les syntagmes nominaux en construisant une grille de catégories morphologiques.

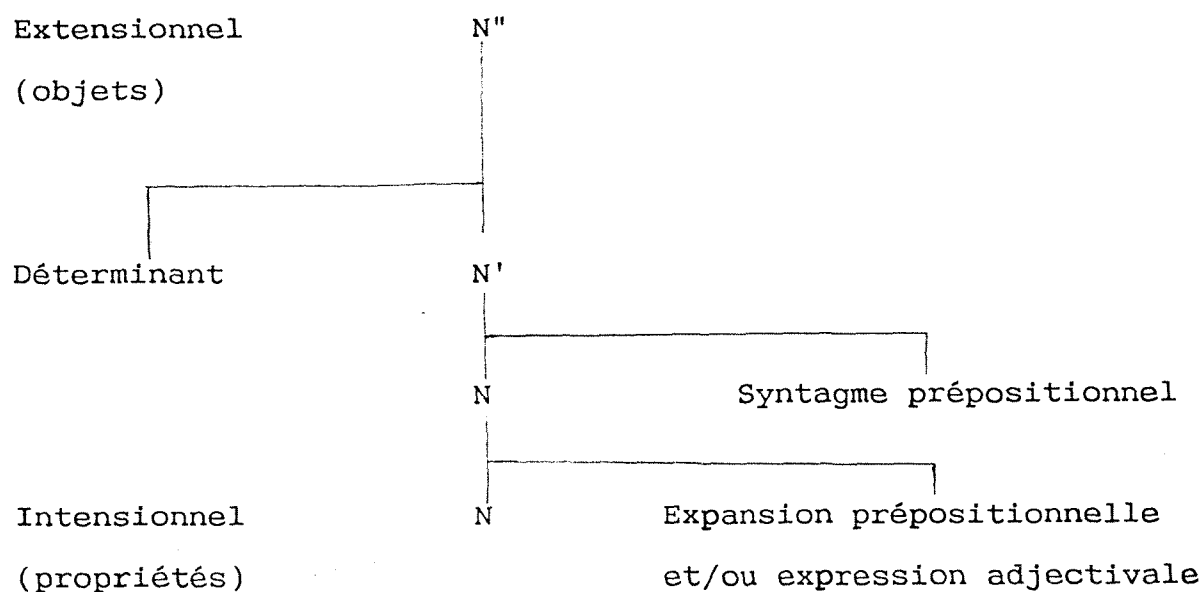
Le lexique de l'analyseur fournit les éléments d'informations nécessaires dans l'ordre de la morphologie et

2. Le Groupe SYDO réunit le laboratoire d'informatique documentaire de LYON 1, le Centre de Recherche en Linguistique et Sémiologie de LYON 2, le Centre d'Etude et de Recherche en Science de l'Information de l'ENSIB et l'Equipe de Recherche sur les systèmes d'Information et de Communication des organisations de LYON 3.

de la syntaxe. Il indique la catégorie et les valeurs prises par les variables flexionnelles, syntaxiques ou lexicales.

L'analyse syntaxique est présentée sous forme de règles de réécriture.

La grammaire de reconnaissance des syntagmes nominaux s'articule autour de trois niveaux : N, N', N'', comme le montre la figure que nous avons empruntée à M. Bouché:³



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

N' : c'est le basculement de l'intensionnel à l'extensionnel.

N'' : c'est l'opération de fermeture, au moyen d'un quantificateur : référence à un objet de la réalité extralinguistique.

3. BOUCHE, R. Le syntagme nominal, une nouvelle approche des bases de données textuelles, in : Actes du Colloque : terminologie et industries de la langue, META, journal des traducteurs, Montréal, vol 34, N 3, septembre 1989, p. 429-434.

Donc, l'analyse syntaxique automatique de la langue naturelle est un problème à la fois informatique et linguistique. L'aspect linguistique de l'indexation automatique nous oblige à réfléchir sur le statut linguistique du descripteur. En effet, quand on pense à l'approche linguistique de l'indexation automatique, on a une autre conception des systèmes d'information documentaire et on se rend compte du phénomène de référence à la réalité extralinguistique. C'est ainsi qu'on arrive à distinguer entre les mots de la langue et ceux du discours. L'importance du fait référentiel a été clairement indiquée dans l'approche de R. BOUCHE : "les deux approches, langage d'indexation et prise en compte d'une valeur référentielle, apparaissent donc comme complémentaires."⁴

Certaines études linguistiques se limitent aux syntagmes nominaux pour aborder le problème de l'indexation automatique. En effet, il semble convaincant d'utiliser les SN comme descripteurs, ils sont les thèmes explicites des composantes du texte à indexer.

L'indexation automatique peut être conçue comme une traduction lexicale des unités de la langue ou encore une traduction syntaxique quand il s'agit d'exprimer les liens syntaxiques entre les différentes parties de discours qui décrivent son contenu.

4. BOUCHE, R., valeurs référentielle et langage d'indexation, communication faite, le 28 novembre 1988, au colloque "Archives et temps réel" organisé à Lille par le CEDRO (Université Lille III), L'ABDS Nord, et les archives du Nord.

Nous allons nous inspirer des travaux de l'équipe SYDO pour l'élaboration des règles de réécriture de grammaire du SN en arabe.

2. SN ET DESCRIPTEUR

Pour arriver à désigner ce que nous entendons par descripteur, nous sommes partis de notion du terme ; en effet le terme est d'une part, l'unité linguistique qui sert à désigner un concept appartenant à une discipline particulière et, d'autre part, l'unité qui sert à l'indexation des documents dans les systèmes d'information documentaire. Il est appelé, dans le domaine des systèmes documentaires, descripteur : "le terme ne peut-être autre chose que le terme préférentiel que le documentaliste choisira parmi tous les termes qui se trouvent être autant de candidats-descripteurs"⁵.

Si on veut décrire un texte à des fins d'indexation automatique, nous pensons que le moyen le plus fiable est de se servir du SN comme descripteur. Dans le traitement automatique de la langue naturelle, c'est l'analyse syntaxique qui devrait montrer quels descripteurs linguistiques prendre pour le traitement du texte.

5. MUSTAFA ELHADI, W. , La terminologie arabe des télécommunications : Faits de variations, Thèse pour l'obtention du titre de docteur en sciences du langage, Université Lyon 2, 1989.

2.1 DEFINITION DU DESCRIPTEUR :

Selon M. LE GUERN, le descripteur n'est pas le mot de la langue mais le mot actualisé : "le descripteur n'est pas le mot de la langue en tant qu'éléments 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, plus exactement syntagme type par rapport aux occurrences des énoncés. Et cela n'est pas vrai seulement pour les descripteurs complexes. Le descripteur simple est lui aussi un syntagme ; s'il semble identique à un mot de la langue, ce n'est qu'une apparence. L'analyse sémantique de descripteur entraîne à y voir une ellipse de l'article, alors qu'il n'y a rien de tel pour les mots de la langue."⁶

Le terme descripteur peut être, dans une certaine perspective, le synonyme de terme :

"*descripteur* et *terme* renvoient à la même réalité et sont donc une relation de synonymie référentielle. Cette synonymie est due à une variation de facette comme le dit M. LE GUERN, c'est-à-dire selon le point de vue sous lequel on considère le concept désigné par les deux termes ; en traduction, ce concept figure sous l'étiquette *terme*, en documentation, en revanche, il figure sous l'étiquette *descripteur* et son rôle est la représentation du monde."⁷

6. M. LE GUERN, M., Les descripteurs d'un système documentaire : Essai de définition, in acte du colloque Traitement Automatique des langues naturelles de systèmes documentaires, BES.(G.6), FAUCHERE (P.M), LAGUEUNIERE (F.) Condenser supp. 1, Université de Clermont Ferrand, 1984.

7. MUSTAFA ELHADI, W. . La terminologie arabe des télécommunications : Faits de variations, Thèse de Doctorat, Université Lyon 2, 1989.

Il faut signaler aussi que parmi les éléments de distinction entre le mot de la langue et le mot du discours figure la différence du statut sémiotique. Cette différence trouve son origine dans les catégories de signes de PEIRCE : "entre le mot de la langue et le descripteur d'un système documentaire, on constate d'abord une différence de statut sémiotique, en prenant cette fois le terme "sémiotique" au sens de PEIRCE. Le recours à la typologie de PEIRCE permet de caractériser le descripteur en soulignant ce qui le différencie du mot de la langue, tout en montrant les traits de ressemblance : ainsi se trouvent expliqués à la fois le fait qu'on ait eu tendance à assimiler les descripteurs aux mots du lexique, et la nécessité à la fois théorique et pratique de les distinguer."⁸

Il convient d'indiquer que la sémantique du descripteur ne relève pas de la sémantique lexicale, il faut considérer qu'elle a pour objet les mots de la langue et non la réalisation effective de ces mots dans le discours. Dans son article sur "les descripteurs", M.LE GUERN établit une comparaison entre le mot de la langue et le descripteur et met en évidence la différence entre la synonymie lexicale et la synonymie référentielle : "la prise en compte de la fonction référentielle des descripteurs permet de poser en d'autres termes la question de la synonymie : deux descripteurs sont synonymes s'ils ont la même référence; il ne s'agit donc pas,

8. M. LE GUERN, M., Les descripteurs d'un système documentaire : Essai de définition, in acte du colloque Traitement Automatique des langues naturelles de systèmes documentaires, BES, (G.6), FAUCHERE (P.M), LAGUEUNIERE (F.) Condenser supp. 1, Université de Clermont Ferrand, 1984.

9. M. LE GUERN, M., Les descripteurs d'un système documentaire : Essai de définition, in acte du colloque Traitement Automatique des langues naturelles de systèmes documentaires, BES, (G.6), FAUCHERE (P.M), LAGUEUNIERE (F.) Condenser supp. 1, Université de Clermont Ferrand, 1984.

dans une perspective documentaire, de synonymie référentielle, alors que la seule prise des signifiés linguistiques conduirait plutôt à y voir une certaine antonymie."⁹

2.2. DESCRIPTION DU SN :

Un SN, selon M. LE GUERN, est "l'unité minimale de discours qui a la possibilité de signifier un objet [...] Maison, le mot du lexique, ne signifie aucune maison que ce soit, alors qu'il suffit que le discours construise le syntagme "une maison" pour que soit désigné un objet concret. La fermeture du prédicat par le quantificateur le transforme en terme."¹⁰

De ce qui précède, on peut comprendre que le terme est un élément référentiel du discours, c'est-à-dire qu'il fait référence à un objet situé dans un univers donné.

pour comprendre ce que nous entendons par élément référentiel, nous reprenons ce que M. LE GUERN a expliqué à propos de ce sujet : nous partons de l'exemple "la maison de Paul", cet expression est analysable : ce syntagme est obtenu par une opération de fermeture de quantificateur "la" sur le prédicat complexe "maison de Paul". Ce prédicat complexe ne fonctionne pas comme le prédicat "maison", mot du lexique, et cette différence de fonctionnement n'est pas due du seul fait que l'un des prédicats est complexe et l'autre ne l'est pas.

10. M. LE GUERN, M. Sur les relations entre terminologie et lexique, in actes du colloque : Les terminologies spécialisées. Approches quantitative et logico-sémantique, Meta, vol.34 NO 3, sept.89.

Le prédicat "maison de Paul" a ceci de particulier qu'il suppose un univers donné, univers dans lequel existe une personne du nom Paul. Appelons le prédicat-lié. En revanche, "maison", mot du lexique, ne suppose aucun univers déterminé(...). Maison, en tant que mot du lexique est considéré par l'auteur comme prédicat libre qui signifie en dehors de prise en compte de quelque univers que ce soit et s'oppose ainsi au prédicat lié "maison de Paul".¹¹

Ici M. LE GUERN distingue entre les mots du lexique et les mots de la langue. En effet, les mots du lexique d'un système documentaire n'ont pas le statut des mots de la langue, il nous l'explique : "dans un lexique comme dans la terminologie, nous retrouvons des "mots" mais d'un côté comme de l'autre ce ne sont pas les mêmes "mots",. L'objet "mot" du lexique est une réalité qui appartient à la terminologie. Le lexique concerne les mots indépendamment des choses, alors que dans la terminologie les "mots" sont liés aux choses ; les mots que les lexicographes désignent comme substantifs sont en réalité des prédicats, ils disent des qualités et non des substances."¹²

M. LE GUERN poursuit la description du SN en disant "je prends dans le lexique le prédicat libre "maison" et je le place dans mon univers de discours ; de ce fait, il n'est plus libre, il devient un prédicat lié, et je suis passé dans une logique extensionnelle. Transformer un prédicat libre en

11. M. LE GUERN, M. Sur les relations entre terminologie et lexique, in actes du colloque : Les terminologies spécialisées. Approches quantitative et logico-sémantique, Meta, vol.34 NO 3, sept.89.

12. M. LE GUERN, M. Sur les relations entre terminologie et lexique, in actes du colloque : Les terminologies spécialisées. Approches quantitative et logico-sémantique, Meta, vol.34 NO 3, sept.89.

prédicat lié, c'est lui associer une classe d'objets pris dans un univers déterminé."¹³

Un SN a donc une référence, qui est le segment de réalité qui lui est associé, il serait donc doté d'une autonomie grammaticale, il est l'expression fermée au sens de la logique et il peut constituer de bons points d'accès à l'information, il se différencie des autres unités syntagmatiques par son statut référentiel.

13. M. LE GUERN, M. Sur les relations entre terminologie et lexique, in actes du colloque : Les terminologies spécialisées. Approches quantitative et logico-sémantique, Meta, vol.34 NO 3, sept.89.

3. PANORAMA SUR LES TRAVAUX DE TRAITEMENT AUTOMATIQUE DE LA LANGUE ARABE :

Ce chapitre a pour objet d'examiner quelques travaux réalisés en France sur le traitement automatique de l'arabe.

Le traitement automatique de la langue naturelle utilise des méthodes linguistiques pour l'analyse des constituants de la phrase, des méthodes statistiques ainsi que des arguments logiques pour déterminer le statut logico-sémantique d'une langue donnée.

Sur le traitement automatique des langues naturelles A. ROMAN souligne l'importance du recours aux deux systèmes de la langue à savoir, le système de communication et le système de la nomination : "le traitement automatique d'une langue est, tout comme la création terminologique dans cette langue, une exploitation de ses deux systèmes de nomination et de communication. Il devra donc reprendre toutes les spécifications requises par la création lexicale. Les autres spécifications requises seront celles nécessaires à l'identification des constituants de chaque phrase et à l'identification des phrases comme des constituants logiques du texte. Cette dernière opération fera recours essentiellement aux coordonnants et, le cas échéant, s'ils ne sont pas suffisamment spécifiés, à des codes de connexion qui les suppléeront."¹

1. ROMAN. A, La création lexicale en arabe : ressources et limites d'une langue humaine naturelle, Paris, GRLM (ed) 1990.

La plupart des travaux réalisés en traitement automatique des langues naturelles et en particulier en arabe ont adopté une démarche d'analyse consistant à reconnaître l'unité de traitement et une démarche de synthèse permettant la génération de l'unité de traitement. Ces travaux définissent souvent un modèle linguistique et un modèle informatique.

Pour réaliser un modèle linguistique, une grammaire doit être définie. Ceci implique une interprétation des formes linguistiques en s'appuyant sur l'analyse morpho-syntaxique qui consiste à établir des liens entre les constituants de la phrase.

Le problème le plus sérieux est posé par le traitement automatique de la composante logico-sémantique du discours comme l'a bien indiqué M. LE GUERN "les informations de nature logico-sémantique fournies par la syntaxe sont, le plus souvent, redondantes par rapport à celles que l'être humain emprunte à la sémantique lexicale et à la connaissance de l'univers. Ce qui fait son attention aux faits syntaxiques peut se relâcher sans que soit nécessairement altérée son interprétation des énoncés. Les applications informatiques, au contraire, maîtrisent mal la sémantique lexicale et ne peuvent simuler la connaissance de l'univers que pour des domaines extrêmement limités et encore à un coût élevé : il est donc essentiel que le traitement automatique exploite, de la manière la plus complète possible, les informations fournies par la syntaxe."²

2. LE GUERN, Michel. un analyseur morpho-syntaxique pour l'indexation automatique, in : le français moderne, Tome LIX, N 1, 1991, p. 24.

Plusieurs travaux de traitement automatique de l'arabe ont été réalisés dans divers objectifs et applications : certains d'entre eux traitent de la langue arabe en vue de la traduction assistée par ordinateur (TAO) : nous signalons le travail de Moghrabi³ sur la traduction conceptuelle des recettes de cuisines.

En ce qui concerne les travaux d'analyse morpho-syntaxique, on peut signaler le travail de HLAL ainsi que le travail d'ABBAB qui était réservé à la morphologie.

Parmi les travaux qui examinent les problèmes du traitement automatique de l'arabe, il convient de signaler les travaux du programme SAMIA (Synthèse et Analyse Morphologiques Informatisées de l'Arabe) dont l'objectif général consiste à engendrer automatiquement à partir de structures initiales des mots graphiques non vocalisés, mais aussi au besoin, des mots vocalisés avec consultation de données enregistrées dans un dictionnaire selon les besoins définis en E.A.O et suivant le mode utilisé (synthèse ou analyse).

Il est à noter que le programme SAMIA se caractérise par l'élaboration de modèles de simulation informatique de certains aspects de l'activité langagière. Il porte sur les connaissances linguistiques relatives à la morphologie.

Les trois éléments principaux sont :

Le synthétiseur : son déroulement s'effectue en trois étapes :

3. Voir bibliographie, nous signalons la référence de ce travail ainsi que pour les travaux qui suivent dans la bibliographie.

- la grammaire des formants ou(G.1) permet de passer d'une donnée, le "vecteur mot" à une représentation phonologique (RP) du mot. Le vecteur mot contient les informations concernant les formats constitutifs du mot. La représentation phonologique rend compte de façon purement formelle, de la succession correcte des morphèmes (fonnants) constituant le mot.

- la grammaire 2 (G2) permet de passer de forme phonologique à la forme phonétique. Cette grammaire est conçue comme un système ordonné de règle à partir de l'analyse phonologique de l'Arabe. Elle doit distinguer les formes phonétiques attestables de celle qui ne le sont pas, tout en tenant compte des phénomènes d'alomorphie (un même morphème peut avoir des représentations phonétiques différentes en fonction du contexte). et d'homophonie (des morphèmes, ou même des séquence de morphèmes, ont quoique phonologiquement distincts, une représentation phonétique différente).

- la grammaire G3, permet de passer de la forme phonétique à la forme graphique du mot arabe. Elle doit aussi, prendre en compte les phénomènes d'allographie et d'homographie.

Ce type de modèle permet de EAA, un certain nombre d'exercices portant sur la formation des mots en arabe.

L'analyseur : le groupe SAMIA a voulu tenir compte du phénomène de non vocalisation de certains textes, c'est pourquoi, le démarche d'analyse consiste à associer à chaque mot graphique non vocalisé, en entrée du système, l'ensemble des analyses autorisées par la morphologie de l'arabe.

L'analyseur doit alors repérer dans ce mot graphique, l'éventuelle présence de proclitiques, préfixes, suffixe et enclitiques.

L'appel au dictionnaire permettra alors de retrouver l'ensemble des bases réalisés qui peuvent lui correspondre.

Le dictionnaire :

Analyseur et synthétiseur font appel à ce dictionnaire. En synthèse, pour engendrer un mot graphique à partir d'une racine, d'un schème et d'un ensemble de spécifications d'ordre morphologique, il faut vérifier :

- l'existence de cette racine;
- après avoir engendré la base de ce mot, vérifier des bases de la racine en question;
- attester l'insertion de la base dans la mot maximal envisagé.

En analyse, la consultation de dictionnaire permet :

- de vérifier l'existence de chacun des constituants du mot, y compris la racine et la base.
- d'assurer la comptabilité de la base trouvée avec les différents formats du mots.

Il convient de mentionner aussi le travail d'ABU AL CHAY,N qui a conçu un système basé sur une structuration des constituants de verbes en préfixe-base-suffixe ; ce système permet l'apprentissage de la conjugaison qui est une

composante importante du programme E. A. O de la morphologie arabe.

Les réalisations informatiques portant sur le modèle d'analyse conçu par ABU AL-CHAY, 84, le modèle de synthèse de MOURAD, 85 et MOURAD, 87 et notamment la modélisation des verbes arabes ABU AL-CHAY, 86 ont permis de dégager le maximum d'informations qui ont été prises en considération par le dictionnaire réalisé par M. HASSOUN, 87 dans le cadre du programme SAMIA.

Il faut signaler enfin le travail de Zouari, 89 qui consistait à construire un dictionnaire permettant l'analyse des textes aussi bien voyellés que non voyellés et dont la structure permet de prendre en compte d'autres types d'informations utiles pour d'autres applications ; exemples : la traduction automatique, l'analyse morphologique et la documentation automatique.

D'après les travaux qui ont été présentés ci-dessus, on remarque que quelques-uns seulement se sont intéressés à certains des problèmes que pose la langue arabe. On constate aussi que la langue arabe traitée a toujours été voyellée, sauf dans un cas ou deux, exemples : le cas de (PROGRAMME SAMIA,), (HASSOUN, 87) et (ZOUARI, 89).

4. LA CONSTRUCTION DU SN DANS L'ORGANISATION DE LA LANGUE ARABE :

L'objectif de ce chapitre est d'examiner l'organisation de la langue arabe du point de vue morphologique et syntaxique afin d'arriver à comprendre la construction du SN.

Pour rendre compte de la construction du SN arabe, il faudrait rappeler sa place dans l'organisation de la langue.

On appelle **morphologie** l'étude des règles gouvernant la structure interne des mots. Un mot est constitué d'une unité de base et éventuellement d'un certain nombre d'affixes.

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

Les racines : le système phonologique de l'arabe établit une disjonction entre les deux ensembles que sont les consonnes et les voyelles, dont il découle que les racines du système de nomination des langues sémitiques sont consonantiques. Une racine arabe ne peut être constituée que de trois consonnes. La racine "f@al" (faire) est retenue par les grammairiens arabes pour représenter l'ensemble des schèmes.

Les schèmes: ce sont des combinaisons de voyelles auxquelles peut s'ajouter la génération d'une consonne radicale, appliquée à la racine, permettant d'obtenir différents dérivations de celle-ci. Ils correspondent à des moules dans lesquels viennent se couler les racines pour former des radicaux des mots.

Les préfixes : ils n'interviennent que dans la conjugaison des verbes.

Les suffixes : ce sont les morphèmes situés immédiatement après la base.

Les variables en arabe existent en genre et en nombre ; celles qui concernent le nombre sont : le singulier, le duel et le pluriel, le genre se limite au masculin et au féminin, il n'existe pas de mots neutres.

A. ROMAN souligne à ce propos que : "les modalités qui déterminent l'appartenance d'une res à une classe sont ou étaient: l'animé qui incluait une res dans le sous-ensemble des "vivants" ; le genre masculin qui incluait une res dans le sous-ensemble des "mâles" ; le genre féminin qui incluait une res dans le sous-ensemble des "femelles" ; l'inanimé qui incluait une res dans le sous-ensemble des "non-vivants". Quant au modus, ils n'existent que par les res, ils ne peuvent donc être classés en tant que tels, indépendamment des res sans lesquelles ils n'existent point ; autrement dit, alors que les res préexistent à leur catégories, les modus ne préexistent pas à une catégorie éventuelle".¹⁷

17. ROMAN. A.. Grammaire de l'arabe. Que sais-je ?, presse universitaire de France, 1990.

Etant donné que la construction du syntagme nominal se fait dans le discours, il ne peut se concevoir sans une 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" 18

La syntaxe est l'étude de l'organisation des mots dans le discours. On peut étudier la syntaxe soit au niveau de la phrase toute entière soit au niveau de segment de phrase :

"la syntaxe d'une langue est l'organisation des unités de nomination de cette langue en système de relations qui structurent en fait une certaine unité de communication : la phrase." 19

L'arabe se sert de la combinaison de son système de communication et de celui de la nomination pour former des syntagmes 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 : une entité linguistique et une entité extra-linguistique : "l'Homme a eu recours à une combinatoire linguistique. Cette combinatoire devait produire ses langues comme autant de "système de système", des systèmes composés chacun de 4 systèmes interdépendants : un système de

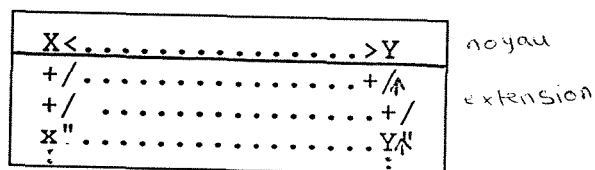
18. M. LEGUERN, M., Les descripteurs d'un système documentaire : Essai de définition, in acte du colloque Traitement Automatique des langues naturelles de systèmes documentaires, BES. (G.6), FAUCHERE (P.M), LAGUEUNIÈRE (F.) Condenser supp. 1, Université de Clermont Ferrand, 1984.

19. ROMAN, A., les relations syntaxiques, in Cercle linguistique d'Aix en Provence, Travaux 3, Aix-en-Provence, 1985.

phonème, un système de syllabe, un système de nomination, un système de communication."²⁰

Le terme étant une unité de discours, sa place est prévue dans la construction de la phrase mise en évidence par André ROMAN comme suit :

"Le système de communication, c'est-à-dire le système constituant les unités de nomination en phrase s'est établi, pour toutes les langues du monde semble-t-il, selon le plan suivant :



Dans ce plan, les unités de nomination "X" et "Y" sont les deux éléments fondamentaux, structurellement nécessaires et inomissibles, du noyau de la phrase ; ces éléments "nucléaires", reliés par une relation biunivoque, sont de co-occurrence comme dans un duo, la première voix et la deuxième voix".²¹

Ces éléments sont au nombre de deux parce que un système ne peut comprendre moins de deux éléments; et ils ne sont que deux parce qu'un système comprenant plus de deux éléments est

20. ROMAN. A., La création lexicale en arabe : ressources et limites d'une langue naturelle, Paris, GRML (ed) 1990.

21. ROMAN. A., La création lexicale en arabe : ressources et limites d'une langue humaine naturelle, Paris, GRML (ed) 1990.

un système complexe et qu'au demeurant sa complexité serait ici inutile.

Dans ce plan, "x", "y" ... sont les unités de nomination éventuellement appelées par le locuteur, selon son besoin, ce sont des unités sans aucune nécessité structurelle. Leur sous-ensemble peut-être vide; "x", "y" sont "extension" reliée à leurs bases soit par la "relation de coordination", symbolisée par (+), soit par "relation de subordination" symbolisée par (↑).

Cependant, dans le cadre de ce travail, nous nous inspirons de règle de réécriture de groupe SYDO, nous nous contentons d'emprunter aux analyses de A. ROMAN, les catégories grammaticales objets du chapitre VI.

En arabe, le sujet et le prédicat sont distingués par la relation d'implication sémantique réciproque qui les identifie comme deux éléments de la phrase.

André ROMAN distingue dans le plan de la phrase deux types de relation entre le sujet et le prédicat : "deux relations syntaxiques générales, R, ont donc été reconnues. Il semble qu'elle soient les seules relations binaires générales possibles. La première est la relation biunivoque entre deux constituants singuliers de la phrase, sa " première voix" et " sa seconde voix" ; cette relation égale, qui sera dite "réciproque" ou de " prédication", est unique. La deuxième est une relation univoque entre deux constituants quelconques de

la phrase, une "base" B, et son "extension" E ; cette relation peut être multiple ; elle peut être réalisée :

- soit comme une relation égale, non hiérarchisée, la relation de "coordination" qui jumelle à sa base une extension de même statut syntaxique que celle-ci ; les coordonnants en sont les différents moyens ;

- soit comme une relation inégale, hiérarchisée, la relation de subordination qui rattache une extension à sa base ; l'extension recevant le statut syntaxique que lui octroie le fonctionnement."²²

Pour expliquer la formation de certains SN sur le modèle de l'opérateur étoile, sur lequel nous reviendrons, on tiendra compte de l'extension par subordination. M. ROMAN la définit ainsi : "l'extension par subordination : la subordination est le moyen de relier, hiérarchiquement, dans le cadre d'une même phrase, à une unité fléchie, la base, une autre unité fléchie "expansion" qui sera avec elle dans une relation d'implication à sens unique. La situation syntaxique de l'extension subordonnée lui est propre : elle est déterminée par le fonctionnel qui l'intègre dans la phrase. Elle est sa fonction."²³

Nous avons pu constater que les expansions que nous avons mentionnées ci-dessus pourraient être envisagées dans le modèle de l'opérateur étoile. Selon M. BERRENDONNER, l'opérateur étoile "s'applique à un couple de prédicats pour former à partir d'eux un prédicat complexe. Il y a toutefois

22. ROMAN. A. La création lexicale en arabe : ressources et limites d'une langue humaine naturelle. Paris, GRLM (ed) 1990.

23. ROMAN. A., La création lexicale en arabe : ressources et limites d'une langue naturelle. Paris, GRLM.(ed) 1990.

des restrictions importantes sur la nature de ses arguments. Les prédicats sur lesquels il opère doivent être des "noms", des "adjectifs" ou des "prédictions", mais en aucun cas des "verbes". Il faudra donc qu'ils soient marqués [-v] dans le lexique [...]."24

En effet, l'arabe peut se servir de ses expansions (d'identification, d'annexion et modale) pour formuler des SN sur le modèle de l'opérateur étoile où la base serait "l'opérande" et les expansions seraient des "opérateurs étoiles".

Voici quelques exemples formulés sur ce modèle :

révolution * syntaxique :	ثورة نحوية
les langues * humaines :	اللغات إنسانية
la théorie * linguistique :	النظرية اللغوية

La base est donc l'opérande (révolution, langues, théorie), et les expansions d'identification qui servent à identifier ces bases (syntaxique, humaine et linguistique) sont des opérateurs étoiles.

On remarque que ces exemples sont des prédicats complexes; donc on peut dire qu'un prédicat complexe est un ensemble de prédicats simples muni de classes d'opérateurs étoile. Ce fait a été souligné par M. LE GUERN : "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

24. BERRENDONNER, A., les référents nominaux du français et la structure de l'énoncé, Thèse de Doctorat d'Etat, Lyon 2, 1978.

sont tous deux des prédicats libres, le résultat est un prédicat libre ; 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. Le choix d'un modèle commun pour le lexique et la terminologie a le double avantage de bien rendre compte à la fois de ce qui les rapproche et de ce qui les distingue."²⁵

Une grammaire d'opérateur ne fonctionne que sur une langue donnée, c'est la raison pour laquelle nous formulons certaines réserves. Même si l'ordre de construction des prédicats est le même pour l'arabe et pour le français, il faut préciser cependant que la grammaire d'opérateurs du français et celle de l'arabe sont deux grammaires d'opérateurs différentes mais homologues et non pas une même grammaire qui s'appliquerait à l'une et à l'autre.

Comme le SN en français, le SN en arabe est donc doté d'une autonomie référentielle : "un SN peut être interprété hors contexte : il a un contenu informationnel au même titre qu'une proposition (au sens logique). Il semble que lorsqu'un locuteur emploie un SN (défini ou indéfini) il pose ou suppose l'existence d'un ou plusieurs objets."²⁶

25. M. LEGUERN, M. Sur les relations entre terminologie et lexique, in actes du colloque : Les terminologies spécialisées. Approches quantitative et logico-sémantique, Meta, vol.34 NO 3, sept.89.

26. METZGER. J.P., Syntagmes nominaux et formation textuelle : reconnaissance automatique et représentation, Thèse d'Etat, Université Claude Bernard Lyon 1, octobre 1988.

5. LES PROBLEMES DE L'ANALYSE LINGUISTIQUE DE LA LANGUE NATURELLE

Nous essayons d'examiner dans ce chapitre les exigences que demandent l'analyse linguistique d'une langue donnée.

Pour analyser linguistiquement une langue donnée, il faut donc concevoir une grammaire qui permettrait de catégoriser ses unités linguistiques. Catégoriser veut dire classer conventionnellement des unités minimales significatives en fonction des propriétés morpho-syntaxiques de la langue.

L'analyse d'un énoncé en langage naturel ne peut s'opérer sans faire appel à des fondements linguistiques. En effet, la linguistique propose des modèles pour traiter des données de nature textuelle et peut apporter des modèles et des outils généraux pour l'analyse morpho-syntaxique des énoncés de types différents (texte, dialogue écrit, etc...).

Toute analyse de la langue naturelle s'appuie sur une grammaire sous-jacente aux analyseurs. Elle est constituée d'un ensemble d'hypothèses portant sur la langue et définissant la nature de ses unités grammaticales.

L'analyse linguistique comprend l'analyse morphologique et l'analyse syntaxique. Pour bien mener l'analyse des unités linguistiques il est nécessaire que les catégories soient définies sur la base de véritables propriétés syntaxiques.

L'analyse morphologique à deux but fondamentaux :

36

- le premier est de reconnaître les "mot" dont se compose le texte que l'on veut analyser. Cette formation fait appel à une définition intuitive du mot.

- le deuxième but est d'associer à chaque mot ainsi reconnu un ensemble d'informations linguistiques :

- l'origine lexicale du mot.
- les catégories grammaticales potentielles.
- le genre et le nombre.
- le temps, le mode, la personne.
- les connaissances de natures sémantiques.

J.P. Metzger signale que : "l'analyse morphologique combine deux fonctions :

- fonction classificatoire : de nature lexicale, qui consiste à attribuer à chaque forme du texte une catégorie syntaxique.
- fonction calculatrice ou flexionnelle, qui consiste, à partir d'une forme donnée, à calculer sa base pour accéder au lexique."²⁷

Il est à noter que pour résoudre une ambiguïté morphologique, il est souvent nécessaire d'avoir préalablement effectué une analyse syntaxique partielle du secteur de texte qui la contient.

Sur cette question A. BERRENDONNER souligne que "Analyse morphologique et analyse syntaxique s'entredéterminent donc, et devront être constamment entremêlées. Cela signifie, notamment, que même si nous présentons ici l'analyse morphologique, du fait de sa spécificité, comme un programme isolé, ce programme n'est pas à concevoir comme une composante

27. METZGER. J.P., Syntagmes nominaux et formation textuelle : reconnaissance automatique et représentation, Thèse d'Etat, Université Claude Bernard Lyon 1, octobre 1988.

autonome et monolithique du traitement automatique "du texte analysé". Il faut, bien au contraire, l'envisager comme un corps de règles articulé en sous-ensembles discontinu, entre lesquels peuvent être intercalées des phases d'analyse syntaxique."²⁸

Dans l'état actuel de nos recherches, nous ne pourrions pas aborder la manière dont l'analyse syntaxique et l'analyse morphologique "s'entredéterminent". Cette question est cependant très importante dans le contexte qui est le nôtre car elle permettrait d'établir la jonction entre les travaux du groupe SYDO (analyse morphologique et détermination du SN) et du programme de recherche SAMIA sur l'analyse morphologique de l'arabe.

En effet, l'analyse syntaxique est un passage obligé pour réaliser des systèmes élaborés de traitement automatique des langues naturelles. Ayant fait l'objet de recherches soutenues durant les trente dernières années, les techniques d'analyse syntaxique sont suffisamment avancées pour pouvoir donner lieu à de nombreuses réalisations. Cependant, aucun module d'analyse syntaxique actuel n'est en mesure de traiter la totalité d'une langue naturelle. En conséquence, une grande variété de modules d'analyse syntaxique ont été développés en vue d'applications particulières ou pour explorer les possibilités d'un formalisme grammatical ou d'une technique d'analyse.

28. BERRENDONNER, Alain, Grammaire pour un analyseur : aspects morphologiques, nouvelle édition, Grenoble, Université Mendès France, cahier du CRISS, N 15, 1990.

Signalons enfin que la langue naturelle est ambiguë, et qu'une sémantique de la langue naturelle a besoin d'une opération de spécification qui permettrait de sélectionner les notions mises en discours par leur compréhension, c'est-à-dire en enrichissant la description des notions visées. De cela on peut comprendre qu'interpréter une forme linguistique, c'est être capable, à partir d'un certain état de représentation de l'univers et sur la base de la structure morpho-syntaxique de cette forme linguistique, d'établir certains types de liens:

- lien de constituants de cette forme entre eux : c'est l'analyse morpho-syntaxique qui va décider si deux termes successifs forment ou non un constituant d'un niveau donné, et si oui, quelles opérations entrent en jeu dans la formation des constituants.

- lien de cette forme à l'environnement : c'est le niveau de l'analyse interprétative, où lexicale, contexte et situation entrent en interaction pour la bonne interprétation de la forme considérée et pour l'unifier avec les représentations préexistantes.

A. BERRENDONNER souligne la difficulté de l'analyse de toute langue naturelle étant donnée leur ambiguïté intrinsèque :

"Comme la langue naturelle est intrinsèquement ambiguë, toute grammaire de reconnaissance d'une telle langue est ambiguë. A une séquence de symboles terminaux, peuvent donc correspondre plusieurs structures syntaxiques. Parmi celles-ci, certaines ne sont pas attestées par la langue. Or un simple analyseur de type combinatoire ne peut distinguer les

solutions parasites des solutions attestées. Il est donc nécessaire de le guider tout au long de l'analyse afin de prévenir la construction de solutions parasites."

6. LES CATEGORIES GRAMMATICALES POUVANT SERVIR POUR LA RECONNAISSANCE DU SYNTAGME NOMINAL :

Nous essayons de définir dans ce chapitre les catégories grammaticales pouvant servir de descripteurs pour l'indexation automatique de textes écrits en arabe.

Dans son travail sur la reconnaissance et l'extraction des syntagmes nominaux, A. AJJOUR²⁹ a classé les catégories grammaticales de la langue arabe exactement de la même façon que les chercheurs pour la reconnaissance automatique du français. Nous pensons que l'arabe ne fonctionne pas tout à fait comme le français ; quelques catégories ne trouvent pas leurs homologues (catégorie des adverbes). Nous essayons d'examiner les points qui différencient les deux langues.

Nous pensons que pour l'extraction des SN en arabe nous aurons besoin de classer les catégories comme suit :

les noms communs, les Masdars (que nous examinerons plus loin), les noms propres, les adjectifs, les pronoms (personnels : toniques et enclitiques, démonstratifs, possessifs, relatifs) les particules (équivalentes de prépositions en français), les coordonnants. Nous prenons certaines réserves à propos de la catégorie des adverbes.

6.1 LA CATEGORIE DES NOMS :

41

A l'origine le nom est indéfini en arabe mais il existe des cas où il est défini : le nom propre, le pronom, le démonstratif, le relatif, le vocatif.

D'après les travaux qui ont été faits par M. BERRENDONNER, sur les référents nominaux, on constate qu'en français les "noms" appelés parfois "substantifs" sont considérés par lui comme des prédicats : "Quant aux noms de la langue, c'est à dire aux items comme (chiens, table, artiste...) dont il est parfois soutenu qu'ils servent à désigner des objets, on a vu que de très nombreux arguments syntaxiques me conduisent à les considérer comme des prédicats unaires, entités sémantiques dépourvues de référence "en soi" au même titre que tout autre prédicat."³⁰

En arabe, selon M. ROMAN : "le nom propre semble différent du nom "commun" en ceci qu'il est un "nom généalogique" ; sa première fonction est, à l'évidence, de déclarer l'appartenance d'un être humain à un groupe social. Aussi le nom propre arabe était-il un nom sans tanwin /n/ parce qu'il était, en tant que nom propre construit, toujours comme base d'une expansion d'annexion, ellipse le plus souvent, le nom de sa famille, de son clan, de sa tribu, de sa communauté."³¹

6.2. LE NOM PROPRE :

30. BERRENDONNER, A., les référents nominaux du français et la structure de l'énoncé, Thèse de Doctorat d'Etat, Lyon 2, 1978.

19. ROMAN, A., La création lexicale en arabe : ressources et limites d'une langue naturelle, Paris, GRIM (ed) 1990.

31. ROMAN, A., Grammaire de l'arabe, Que sais-je ?, presse universitaire de France, 1990.
abe, que sais je ? ,

Le nom propre en arabe peut être employé avec prédéterminant ;

Exemples : *L'Algérie* الجزائر *Le Soudan* السودان

Il peut être aussi employé sans prédéterminant ;

Exemples: *L'Egypte* مصر *La Syrie* سوريا

Cependant, il existe des noms propres qui ont conservé la trace étymologique du déterminant d'origine, laquelle est dans le lexique d'aujourd'hui est considérée comme faisant partie intégrante du nom propre,

Exemples : *La Jordanie* الأردن *La Chine* الصين

Cette question sera examinée plus profondément dans une recherche ultérieure.

6.3. LES MASDARS :

Il faut signaler aussi qu'en arabe il existe des noms qui sont en relation avec les verbes. C'est à dire ils sont en connexions avec les verbes sous l'aspect de dérivation. Ce qui va nous amener à parler des "Masdars" qui représentent un abstrait ; ils peuvent désigner une profession comme par exemple : (commerce, تجارة), (tourisme, سياحة), (couture, حياكة) etc...

Ils représentent des noms issus de dérivation du verbe exemples : (bibliothèque, مكتبة), (école, مدرسة) etc...

Le "Masdar" représente un nom abstrait analogue à l'infinitif français employé substantivement (le boire, le manger, l'existence).

Lors de l'indexation, nous pensons qu'ils seront extraits comme SN, étant donné qu'ils ont un statut référentiel nominal. Il va de soi que cette abstraction n'ira pas sans engendrer des ambiguïtés, que nous nous réservons de traiter ultérieurement.

6.4. L'ADJECTIF :

On constate que la catégorie des noms et celle de l'adjectif, en français, se caractérisent par la non distinction entre elles ; en effet, la frontière est très faible puisque une même forme peut donner des noms concrets, des noms abstraits ou des adjectifs. En arabe, pour l'extraction des SN, on envisagerait cependant deux catégories distinctes : la catégorie des noms et celle des adjectifs. . Nous signalons par contre que certains prénoms en arabe ont le statut d'un adjectif, comme par exemples :

(beau = جميل), (bon, حسن), (généreux, كريم),
(vainqueur, منتصر)

Il existe également en arabe, des noms qui peuvent être parfaitement des adjectifs.

Exemple : le savant,

Il est savant : إنه عالم
 c'est un savant : هو العالم

Cela laisse entendre que l'on trouve une sous catégorie non affectée. Nous traitons cette question ultérieurement.

6.5. LES PREDETERMINANTS :

C'est cette catégorie qui sert à fermer les expressions nominales et à les rendre précises. C'est le moyen par lequel un mot peut être défini, ce mot change dès lors de statut, il n'est plus le mot de la langue mais il devient mot du discours à caractère référentiel.

L'article en français est considéré comme un des facteurs dont le rôle essentiel consiste à relier les classes virtuelles aux objets qui leur correspondent dans la réalité.

En arabe on a un seul article défini invariable /al/. L'article indéfini est présent sous forme de "Tanwin" à la fin du mot et il est confirmé par l'absence de l'article /al/ :

»la nunation est un signe ajouté au nom singulier et au pluriel brisé déclinable, pour souligner leur appartenance pleine et entière à la classe nominale."³²

Deux rôles sont assignés à /al/ ; c'est un morphème de quantification, et une modalité de détermination : "/al/ devant une res : si la res est "nom commun", /al/ est une "unité numérale", l'article défini ; le nombre de res est

32. NASIF, A. Nâr al-qirâ fi 'sarh gawf al-firâ, Beyrouth, 1863.

arrêté par /al/ ; le syntagme constitué par / al/ et par la res est l'étiquette d'un ensemble /al + kalb/, = "le chien" ou /al + kila:b/, = "les chiens" ; si la res est un nom propre /al/ est une modalité de notoriété."33

Lorsqu'on a une fermeture au moyen de quantificateur, on a un passage obligé du niveau de la logique intensionnelle à celle dite extensionnelle ; donc la distinction entre ces deux niveaux de logique est nécessaire pour distinguer une propriété de l'objet dont on parle. En fait, on peut dire que les déterminants sont des éléments du passage du domaine intensionnel au domaine extensionnel.

M. LE GUERN a bien expliqué la différence entre les deux niveaux de la logique de la manière suivante :

"si l'on prend un prédicat ou un mot de la langue, par exemple "cheval", on est dans une logique intensionnelle : aucun univers n'est pris en compte. Si on considère le prédicat cheval, mais en tenant compte cette fois-ci d'un univers on construit un concept "un cheval", il y a une classe virtuelle. Le prédicat cheval ne peut être considéré comme concept que lorsqu'il bascule d'une logique intensionnelle à une logique extensionnelle."34

Du point de vue de la logique, on peut envisager en arabe le niveau extensionnel au moyen des déterminants. Cette détermination se fait à l'aide du quantificateur universel

33. ROMAN. A., La création lexicale en arabe : ressources et limites d'une langue naturelle, Paris, GRLM. (ed) 1990.

34. LE GUERN, M., "sur les relations entre terminologie et lexique", in actes du colloque Les terminologies spécialisés. Approche quantitatives et logico sémantique, Meta. vol 34 no 3, 1989.

/al/, les substantifs qui sont précédés de ce quantificateur universel sont au niveau extensionnel et ils ont un référentiel, ils sont donc des syntagme nominaux car ils représentent des expressions fermées.

Le niveau intensionnel peut être envisageable par l'intermédiaire du "Tanwin" : les substantifs qui ont la marque de nunation demeurent donc au niveau intensionnel, ils sont indéfinis. Le tanwin sert à les actualiser mais il ne les identifie pas, c'est à dire que ces substantifs sont des propriétés, ils ne renvoient pas à des objets référentiels.

On peut en conclure qu'il existe deux classes d'actualisateurs, une classe qui actualise le substantif sans l'identifier, c'est le cas du "Tanwin", et une classe qui l'identifie en même temps qu'il l'actualise, c'est le cas de /al/.

En ce qui concerne les prédicats complexes, comme par exemple l'expression "la maison blanche", on constate une différence entre le français et l'arabe au niveau de la détermination. Car la détermination porte sur toute l'expression "maison blanche" en français, alors qu'en arabe on a "la maison la blanche" ; c'est-à-dire double détermination.

Donc en français, logiquement parlant, c'est une proposition qui signifie la propriété "d'être maison blanche" à un individu, alors qu'en arabe, c'est une proposition qui signifie la propriété "d'être blanche" à un individu.

On peut conclure de cet exemple que la linguistique et la logique convergent ; comme le confirme MC. CALWAY "la prise en considération du seul point de vue logique d'une part, ou du seul point de vue linguistique d'autre part, ne peut que nous amener à pratiquer un grand nombre de choix concurrents qui se présentent pour exprimer la structure sémantique des phrases ; par contre, si le point de vue logique et le point de vue linguistique sont considérés ensemble, l'éventail des possibilités relatives à la structure sémantique se rétrécit de manière rigoureuse, ce qui ne va pas toutefois jusqu'à imposer des choix absolus, c'est à dire l'entreprise du linguiste et du logicien vont de pair."³⁵

6.6. LES PRONOMS :

Les pronoms sont tous, à leur origine morphologique, en dehors du système de la racine.

La première personne ne distingue pas les genres, seulement le nombre, singulier **أنا** , pluriel **نحن** .

On peut les diviser en deux catégories :

- les pronoms personnels.
- les pronoms démonstratifs.

Les pronoms personnels se divisent à leur tour en deux catégories : les autonomes qui sont appelés les pronoms isolés, et les enclitiques qui sont dépendants soit d'un nom, soit d'un verbe ou d'une préposition, ils sont appelés les

35. GALMICHE (M), LAWLER (J), MC CALWAY. (J.D), MULLER (u), SUEUR (J-P). Quantificateurs et référence, in Language, NO, 48, p. 30, decembre 1977.

pronoms affixes. Les pronoms personnels en arabe peut avoir une valeur sémantique comparable à certain point, avec les adjectifs possessifs du français, mais syntaxiquement, ils sont construits selon la structure de "l'annexion". La différence en ce qui concerne l'accord en genre est que le possessif s'accorde en français avec le nom de l'objet, alors qu'en arabe il s'accorde avec la personne à qui appartient l'objet.

EX : il est venu avec sa soeur =	جاء مع أختها
elle est arrivée avec son frère =	جاءت مع أخيها
j'ai lu son livre (le livre de Myriam) =	قرأت كتابها
j'ai lu son livre (le livre de Mohamed) =	قرأت كتابه

L'arabe a organisé ses pronoms démonstratifs de façon à désigner l'objet proche et l'objet éloigné. La différence entre la construction des syntagmes nominaux arabes formés à l'aide d'un démonstratif et celle d'un SN français est qu'en arabe intervient l'article défini, entre le démonstratif et le nom, alors qu'il est absent en français

6.7. LES COORDONNANTS :

Les coordonnants, selon M. ROMAN, sont des unités amorphes qui intègrent dans un cadre syntaxique donné des syntagmes auxquels ils donnent la même situation syntaxique que les syntagmes auxquels ils les associent. Les coordonnants sont donc par là différents des fonctionnels qui, eux,

introduisent des syntagmes auxquels ils donnent une situation syntaxique propre.

6.8. PREPOSITIONS : Les prépositions ainsi que les locutions prépositionnelles sont importants pour la reconnaissance et l'extraction des syntagmes prépositionnels.

6.9. Pseudo-catégorie : L' ADVERBE :

Il existe beaucoup de polémique, surtout dans les textes anciens, en ce qui concerne la définition de cette catégorie. Pour ce qu'on appelle communément en français, des adverbes, l'arabe a des formes qui sont construites de cette façon :
 Nom + marque casuelle
 (accusatif + a pour les noms déterminés ou + an - avec nunation- pour les noms indéterminés).

Ce que nous pouvons appeler adverbes comme le permet les textes modernes, ce sont les adverbes qui concernent le temps (al-'a`na : maintenant), (al-yawma : aujourd'hui) etc, d'autres qui concernent le lieu (l'action de sortir : dehors), (l'action d'entrer : dakilan), certains pour désigner la manière, (ruwaydan : lentement). Mais cela ne constitue pas à proprement parler une catégorie grammaticale comme c'est le cas en français.

7. REGLES D'ECRITURE DE GRAMMAIRE DU SN

Après avoir examiné dans le chapitre précédent les catégories grammaticales pouvant servir de descripteurs, nous tentons d'élaborer une première grammaire de règles de réécriture du SN. Cependant, cette première étape ne représente qu'un début vers une longue recherche à travers laquelle nous espérons pouvoir compléter l'élaboration de cette grammaire.

"pour mener à bien l'analyse syntaxique d'un langage L, il faut lui associer une grammaire G, telle que toute phrase du langage puisse être engendrée par la grammaire et que toute phrase engendrée par la grammaire soit élément du langage. Chomsky a proposé une classification des langages en fonction de la grammaire qui leur est associée".²³

Une grammaire G est toujours désignée comme suit :

- un alphabet sur lequel doivent être construites les chaînes (symboles terminaux).
- ensemble de symboles non terminaux.
- un symbole de départ.
- ensemble des règles.

Dans notre essai de grammaire, le vocabulaire terminal est l'ensemble des catégories définies plus haut.

23. LALLICHE-BOIDIN, Geneviève, analyse syntaxique automatique du français : application à l'indexation automatique, thèse de Doctorat, Université des Sciences Sociales de Grenoble, octobre 1986.

Le vocabulaire non terminal est : { N", N', N, A", A', A, D', D, P, SP, SPⁿ }.

N" est le symbole du SN

Un SN s'écrit simplement :

N" = D+N' : déterminant plus nom, on remarque que la fermeture de SN au moyen de prédéterminant rend l'expression bien précis, exemple : le jardin : البستان

Sur les SN simples, il est intéressant de rappeler la remarque de Basier selon laquelle : "ces nominaux peuvent surgir dans un texte (ou discours) sans avoir été introduits au préalable (l'article défini qui les détermine n'est donc pas anaphorique), ni être accompagnés d'aucun élément à valeur spécifiante (adjectif, relative restrictive, complément de nom...) susceptible d'expliquer leur caractère défini. Ces SN, peuvent fonctionner tout seuls, et jouissent d'un statut qui les apparente aux noms propres."²⁴

L'expression "le jardin" peut avoir en arabe une expansion et on envisagera une règle que se réécrit :

N" = D + N + D + Ø + A

le beau jardin :

البستان الجميل

On remarque que l'adjectif épithète fonctionne comme le nom : quand le nom est précédé de l'article /al/, lui aussi doit l'être, quand le nom est suivi d'un suffixe casuel comportant une marque de nunation lui aussi doit l'avoir.

Lors de l'indexation d'un document, extrait dans ce cas ~~on~~ le SN composé d'une base plus son expansion, c'est à dire un prédicat complexe. Cette expression, on le constate bien, est formulée par l'opérateur étoile où la base "albustan" est l'opérande, et son expansion "aj-jamîl" est l'opérateur étoile.

Pour le démonstratif, on a la règle suivante :

$$N'' = \text{Dém} + \text{Art} + N$$

Dans les expressions qui portent un adjectif cardinal, la règle est la suivante :

$$N'' = \text{cardinal} + N$$

أكثر رجال

On extrait comme syntagme toute l'expression parce que c'est l'adjectif cardinal qui quantifie le nom qui le suit, donc il l'identifie en quelque sorte et la fermeture de l'expression se fait par l'intermédiaire de cet adjectif cardinal.

En ce qui concerne les quantificateurs :

Ex : quelques femmes

بعض النساء

most of the members of "NP"

la plupart des hommes

أكثر الرجال

la règle est :

$$N'' = Q + \text{Art} + N$$

Pour les quantificateurs suivis de DE

peu d'étudiants :

قليل من الطلاب

beaucoup d'enfants :

كثير من الأطفال

la règle est :

$$N'' = Q + \text{pré(min)} \text{ Art} + N$$

Pour le possessif :

$N'' = N + \text{PRO(poss)}$

exemple : notre chien =	كلبنا
sa recherche =	بحثه، بحثها
son livre =	كتابه، كتابها

Il existe trois cas en arabe : le nominatif, l'accusatif et le génitif. Le génitif existe sous deux formes :

- en présence d'une préposition
- en cas d'annexion.

A partir de ces deux cas on peut envisager deux règles :

Le syntagme annexionnel:

$N'' = N + N''$

Le syntagme prépositionnel :

$SP = p + N''$

Les Règles correspondant au texte analysé :

A titre d'exemple, nous présentons l'analyse d'un texte arabe²⁵ extrait d'une étude faite par Dr. ALI. Nabil, intitulée "la langue arabe et l'ordinateur". Cette analyse a pour objet d'isoler les SN, et de proposer un premier essai de grammaire.

$N'' = N'$

$N' = N + SP$

$SP = P + N''$

$N'' = D+N$

النظريات الرسمية للغة :

لِ اللغة

لغة ال

$$N'' = N'$$

$$N' = N + SP$$

$$SP = P + N''$$

$$N'' = N + N''$$

$$N'' = D + N$$

التوزيع النسبي لأنماط الجمل

أنماط الجمل

أنماط الجمل

الجمل

$$N'' = N'$$

$$N' = N + SP$$

$$SP = P + N''$$

$$N'' = D+N + D + \emptyset + A$$

السلوك النحوي للجماعة اللغوية.

للجماعة اللغوية

الجماعة اللغوية

GRAMMAIRE DU SN

SYNTAMES NOMINAUX :

$$N'' = D' + N'$$

$$N'' = NPR$$

$$N'' = PRO$$

$$N'' = N'' + N''$$

$$N'' = N'' + A''$$

$$N'' = N'$$

SYNTAGME ANNEXIONNEL :

$$N'' = N + N''$$

SYNTAGMES ADJECTIVAUX :

$$A'' = A' + SP$$

$$A'' = A'$$

$$A'' = D + \emptyset + A$$

$$A' = A$$

EXPRESSIONS NOMINALES :

$$N' = N + SP$$

$$N' = N$$

$$N' = N + A'$$

EXPRESSIONS PREDETERMINATIVES:

$$D' = D$$

$$D' = \text{Numéral}$$

$$D' = \text{cardinal (num)}$$

$$D' = D' + D$$

$$D' = Q$$

$$D' = Q + \text{Prép}(\text{min})$$

$$D' = \text{Dém}$$

SYNTAGME PREPOSITIONNEL:

$$SP = P + N''$$

$$SP = sp + sp^n$$

LES NOMINAUX :

$$N = \text{F-NOM}$$

$$N = \text{F-ADJ}$$

$$N = \text{HASDAR}$$

Remarques :

L'ambiguïté des phrases écrites en arabe résulte en grande partie de l'absence de voyellation, car un mot non voyellé peut se lire selon différentes manières, et pour un analyseur morpho-syntaxique, c'est souvent très difficile de distinguer entre les mots qui s'écrivent de la même manière et qui manquent de voyellation.

Ex : écrire =

كُتِبَ

livres =

كُتِبَ

il est parti =

ذَهَبَ

or =

ذَهَبَ / ذَهَبَ

Nous avons hésité sur la conception d'une règle pour un certain nombre d'expressions qui représentent des syntagmes nominaux complexes (des prédicats complexes formés sur le modèle de l'opérateur étoile). Nous sommes arrivés à cette forme :

$N'' = D + N + D + \emptyset + A$

ce qui nous a amené à la règle suivante :

$N'' = N'' + A''$

où le symbole vide renvoie à base nominale de l'expression.

Ex : l'école*syntaxique=

المدرسة
النحوية

la distribution*relative=

التوزيع النسبي

Nous réservons également cette question - qui mérite une étude sur les phénomènes de l'anaphore - pour un travail ultérieur.

CONCLUSION

Nous avons essayé au cours de cette étude d'examiner les unités linguistiques de la langue arabe intervenant dans l'indexation automatique des textes écrits dans cette langue.

N'ayant pas trouvé de travaux proprement consacrés à ce problème en arabe, nous avons jugé utile de nous intéresser aux travaux relatifs à cette question en français pour en adopter un modèle concernant la langue examinée.

A ce stade de notre étude, nous pouvons dire que l'analyse de quelques unités linguistiques de la langue arabe nous a permis de constater que chaque langue adopte un modèle linguistique qui lui est propre pour la description de son système grammatical.

Chaque langue est un système qui opère une sélection au travers et aux dépens de la réalité objective, chaque langue structure la réalité à sa manière, choisit les éléments fondateurs de son propre fonctionnement linguistique.

Les éléments de la réalité du langage dans une langue donnée ne représentent jamais tout à fait la même forme dans une autre langue. L'arabe, par exemple, établit ses phénomènes linguistiques par lesquels on le distingue des autres langues. Une langue ne peut jamais être la copie directe d'une autre langue.

Examiner une langue donnée de point de vue linguistique amène à l'analyser morphologiquement et syntaxiquement. Toutefois, la logique est un phénomène important pour la compréhension des lacunes linguistiques, d'où la combinaison de ces deux domaines pour la plupart des travaux de traitement automatique de la langue naturelle.

Les linguistes qui travaillent dans le domaine des sciences de l'information sont avantagés par rapport aux linguistes théoriciens dans la mesure où les premiers peuvent mettre à leur disposition un outil informatique pour la réalisation d'une application souhaitée.

Dans les systèmes d'information documentaire, les procédures de la recherche de l'information demandent que l'on se place dans un univers référentiel, d'où l'appel à la référence du phénomène de la réalité extra-linguistique.

Parmi les unités du discours à capacité référentielle en arabe, nous ne nous sommes intéressés qu'aux référents nominaux, et parmi eux les syntagmes nominaux que nous considérons comme syntagmes documentaires. En nous limitant au SN, nous avons développé les concepts théoriques permettant de distinguer les éléments d'un texte capables de servir de descripteur de ceux qui ne le sont pas. Or, nous avons remarqué que les verbes jouent un rôle important dans la construction de la langue arabe c'est-à-dire que la langue arabe est véhiculée en grande partie par les verbes ; on peut donc se demander si le verbe ne peut être tenu pour une unité référentielle du discours, s'il ne peut jouer le rôle d'un

meilleur descripteur et si ne retenir que les SN comme candidats descripteurs ne revient pas à se priver d'une importante clé d'accès à l'information, le problème reste posé, une étude plus approfondie sur les syntagmes verbaux apporterat peut être une solution.

BIBLIOGRAPHIE

ABAAB, H. Contribution au traitement automatique de la langue arabe. Conception d'un synthétiseur morphologique utilisable en E.A.O., Thèse pour le Doctorat du 3^e cycle, Université de Paris 7, 1984.

ABU AL - CHAY, N. Un système expert pour l'analyse et la synthèse des verbes arabes, Thèse pour le Doctorat, Université Claude Bernard-Lyon 1, juillet 1988.

AJJOUR, A. reconnaissance et extraction des syntagmes nominaux à partir d'un corpus en arabe et sa traduction en français, mémoire de DEA, l'ENSB, 1991.

ALI, N., La langue arabe et l'ordinateur, Ta'reep, 1988.

ARSAC, J. , FLUHR, C. , Development of documentray, softword. INFODIAL. 2nd International Congress and Exhibition on Data Based and Data Banks, 1983, p. 276- 281.

BERRENDONNER, A. Grammaires formelles et analyse morpho-syntaxique automatique, in : Ecole d'Eté des sciences de l'information / sous la direction de Y. Le Coadic, J. Roualt (ed), Vignieu, DBMIST, septembre 1983.

BERRENDONNER, A. Grammaire pour un analyseur : aspects morphologiques, Nouvelle édition, Grenoble, Université Mendès France, Cahier du CRISS N 15, 1990.

BERRUT, C. Une méthode d'indexation fondée sur l'analyse sémantique de documents spécialisés, thèse pour le Doctorat, Université Joseph Fourier, Grenoble1, décembre 1988.

BOUCHE, R. Valeur référentielle et langage d'indexation dans les systèmes d'informations documentaires, in : Colloque sur Archives et Temps réel, Lille, CREDO (Université Lille3) ADBS Archives du Nord, 28 novembre 1988.

BOUCHE, R. Le syntagme nominal, une nouvelle approche des bases de données textuelles, in : Actes du Colloque terminologie et industries de la langue, META, Journal des traducteurs, Montréal, vol 34, N 3, septembre 1989, P. 429-434.

BOUCHE, R., LAINE, S., METZGER, J. P. Extraction de connaissances à partir d'une collection de documents: knowledge retrieval from a documentray set, in: IST international ISKO-conference on tools for knowledge organization and the human interface. Darmstadt (RFA) 14-17 august 1990.

BOUCHE, R. , GERMAIN, N. Bibliométrie, infométrie et analyse automatique de documents écrits, Communication pour les journées d'études organisées par la Société Française de Bibliométrie Appliquée de 5 au 7 juin 1990, Ile Rousse.

- CHALON, J. M. Les applications du traitement automatique du langage naturel, Note de synthèse, Ecole Nationale supérieure de bibliothécaires, 1991.
- CHARTON, G. Le traitement de l'information, le logiciel LEXI-NET dans une chaîne de contrôle de flux. Société française de bibliométrie appliquée. Congrès, 1987.
- CONTINSOUZAS, L. Indexation automatique et recherche documentaire, mémoire de DEA : Science de l'information et de la communication, Université Lumière Lyon 2, 1991.
- COULON, D. Informatique et langage naturel: présentation générale des méthodes d'interprétation des textes écrits, Technique et Science Informatique, 1986, vol. 5, N 2, p.103-128.
- DEBRITO, M. Réalisation d'un analyseur morpho-syntaxique pour la reconnaissance du syntagme nominal. Utilisation des grammaires d'affixes, thèse pour le Doctorat, Université Claude Bernard Lyon 1, septembre 1991.
- DESCLES, J. P. Langages applicatifs, langues naturelles et cognition, Hermès, 1990.

DESCLES, J. P. Quelques systèmes de représentations linguistiques et méta linguistiques, in : la notion de système dans les sciences contemporaines, Colloque CNRS, Lyon, 1980, presse universitaire Aix-en-Provence.

DICHY, J.,

DICHY, J., HASSOUN, M. Simulation de modèles linguistiques et enseignement assisté par ordinateur de l'arabe, Travaux SAMIA, Paris : Fondation postuniversitaire interculturelle, 1989..

DUPONT, P. Eléments logico-sémantiques pour une analyse du français, Thèse d'Etat : Université Lumière LYON 2, 1983.

JAYEZ, J. H. Un survol des recherches sur le traitement automatique du langage naturel, Linguisticae Investigationes, revue internationale de linguistique française et de linguistique générale, 1980, vol. 4, N 1, p; 39-109.1

HASSOUN, M. Système d'analyse morphologique automatique de la langue arabe, Thèse pour le Doctorat du 3ème cycle de mathématiques (informatique), Université de Lyon 1, 1982.

HASSOUN, M. Conception d'un dictionnaire pour le traitement automatique de l'arabe dans différents contextes d'application, Thèse pour le Doctorat d'Etat, Université Lyon 1, juillet 1987.

HLAL, Y. Méthodes d'apprentissage pour l'analyse morpho-syntaxique (expérimentation dans le cas de l'Arabe et du Français), Thèse pour le Doctorat d'Etat, Université de Paris 9, 1979.

KOULOUGHY, D. Vers un traitement automatique de la prédication verbale en arabe standard moderne, T.A. Informations Vol, 28, N 1, 1987, p. 23-45.

LAINE, S. Extraction et sélection de descripteurs complexes dans un ensemble de textes pour leur indexation automatique, Thèse pour de Docteur-Ingénieur en mathématiques (informatique), Université Claude Bernard-Lyon 1, juin 1982.

LAINE, S. , LAROUK, O. , VIDALENC, I. Système d'informations textuelles: l'apport des logiques extensionnelles et intensionnelles. (French Textual Information Systèmes: the contribution of extensional and intensional logics), in: 11th international conference on research & development in information retrieval, ACM/SIGIR/sous la responsabilité de Y. CHAIRAMELLA: presses universitaires de Grenoble, 13-15 juin 1988, p. 507-518

LAINE, S. Systèmes d'informations textuelles, in : colloque sur la recherche en sciences de l'information (I). Des enjeux professionnels . ENSB/ADBS 23 NOVEMBRE 1989 AU 1989.

LALLICH-BOIDIN, G. Analyse syntaxique automatique du français: Applications à l'indexation automatique, Thèse de Doctorat: Université des sciences sociales de Grenoble, octobre 1986, 167 P.

LALLICH-BOIDIN, G., HENNERON, G., PALERMITIR, R. Analyse du français: Achèvement et implémentation de l'analyseur morpho-syntaxique, Grenoble, Université Mendès France, cahiers du CRISS N 16, 1990.

LALLICH-BOIDIN, G., ANTONIADIS, G., POLITY, Y. ROUALT, J. A french texte recognition model for information retrieval système, SIGIR, 1988.

LEENS, D. About machine translation systems and machine-aided translation systems, or how translators and computers can interact, The candien journal of information science, vol. 15, N 4 décembre 1990, p. 23-38.

LE GUERN, M. Les descripteurs d'un système documentaire : essai de définition, in: Colloque sur le traitement automatique des langues naturelles et systèmes documentaires, Clermont-Ferrand, 6 et 7 mai, 1982.

LE GUERN, M. Sur les relations entre terminologie et lexique, in : Actes du colloque terminologie et industrie de la langue,

META, journal des traducteurs, Montréal, Vol. 34, N 3 ,
septembre 1989, p. 340-343.

LE GUERN, M. Un analyseur morpho syntaxique pour l'indexation automatique, le Français Moderne, juin 1991, tome LIX, N 1, p. 22-35.

METZGER, J. P. Systèmes nominaux et information textuelle, Thèse d'Etat ès sciences, Université Lyon 1, 1988.

MOURRAD, F. Réalisation d'un synthétiseur morphologique de l'arabe, Mémoire de DEA, Université Lyon 1 et Lyon 2, 1985.

MUSTAFA ELHADI, W. , La terminologie arabe des télécommunications : Faits de variations, Thèse de Doctorat, Université Lyon 2, 1989.

ROMAN, A. Analyse morpho-syntaxique générale des constituants de la phrase arabe en vue d'une traduction en arabe assistée par ordinateur, T. A. information, vol. 27, N 1, 1986, p.17-26.

ROMAN, A. La création lexicale en arabe : ressources et limites d'une langue humaine naturelle, Paris, GRLM (ed) 1990.

SALKOFFF, M. Analyse automatique du français. Traitement des langues naturelles, congrès 4 juillet 1988, p. 217-237.

SIOUFI, A. La détermination dans le syntagme nominal arabe,
Thèse de Doctorat, Université Lyon 2, 1981.

VANDEUR, M. Approche de l'élaboration d'un système expert
d'aide à l'indexation, cahiers de la documentation, Bladen
Voor de Documentatie, N 4, 1990, p. 75-90.

VIDALENC-SABOURIN, I. Traitement automatique des anaphores en
français, étude linguistique préalable, Thèse de Doctorat,
Université Lyon 2, 1989.

WOODS, W. Transition network grammars for language
anaylsis, journal of the ACM, vol. 13, N 10, 1970.

ZOUARI, L. Construction automatique d'un dictionnaire orienté
vers l'analyse morpho-syntaxique de la langue arabe écrite
voyellée ou non voyellée, Thèse pour le Doctorat, Université
Paris 11, 1989.

BIBLIOTHEQUE DE L'ENSSIB



8023896