

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

**DEA en sciences de l'information  
et de la communication**

**Option : Systèmes d'information documentaire**

**MEMOIRE DE DEA**

**Conception et réalisation d'un dictionnaire  
informatisé de l'arabe, compatible avec les besoins  
de l'analyse morphologique et  
de la correction orthographique**

**présenté par**

**Malek GHENIMA**

**sous la direction de Mr Mohamed HASSOUN,  
Ecole Nationale Supérieure des Sciences de  
l'Information et des Bibliothèques (E.N.S.S.I.B)**

**1992**



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

**DEA en sciences de l'information  
et de la communication**

**Option : Systèmes d'information documentaire**

**MEMOIRE DE DEA**

**Conception et réalisation d'un dictionnaire  
informatisé de l'arabe, compatible avec les besoins  
de l'analyse morphologique et  
de la correction orthographique**

**présenté par**

**Malek GHENIMA**

**sous la direction de Mr Mohamed HASSOUN,  
Ecole Nationale Supérieure des Sciences de  
l'Information et des Bibliothèques(E.N.S.S.I.B)**

**1992**

**Conception et réalisation d'un dictionnaire  
informatisé de l'arabe, compatible avec les besoins  
de l'analyse morphologique et  
de la correction orthographique**

**présenté par Malek GHENIMA**

**RESUME** : Le travail que nous présentons a pour objet la conception et la réalisation d'un dictionnaire informatisé pour le traitement automatique de l'arabe dans le cadre du programme S.A.M.I.A (Synthèse et Analyse Morphologique Informatisées de l'Arabe) en vue de l'analyse morphologique et de la correction orthographique.

**DESCRIPTEURS** : Dictionnaire Automatisé, Descripteur Linguistique, Traitement Automatisé, Arabe, Synthèse Morphologique, Analyse Morphologique, Correction Orthographique, SAMIA.

**ABSTRACT** : This work aims to conceive and realize an electronic dictionary for automatic processing of arabic language as part of S.A.M.I.A research program (Synthèse et Analyse Morphologique Informatisées de l'Arabe). This dictionary will be used to perform morphological analysis and spelling correction.

**KEYWORDS** : Automatic Dictionary, Linguistic Keyword, Automated Processing, Arabic, Morphological Synthesis, Morphological Analysis, Spelling Correction, SAMIA.

A ma Mère

A mon Père

A mon Frère

A mes deux Soeurs

A mes Amis Proches

et A Monica

## REMERCIEMENTS

Je tiens à exprimer mes remerciements à Monsieur M. HASSOUN qui a assuré la direction de mes travaux. Ses conseils, ses critiques et son perpétuel soutien ont permis de mener à bien ce travail.

Je lui adresse l'expression de ma profonde et sincère reconnaissance.

Monsieur J. DICHY, m'a aidé efficacement durant la période de mes recherches par son expérience acquise sur le traitement automatique de l'arabe, ses remarques et ses précieux conseils.

Qu'il trouve ici l'expression de mes sincères remerciements.

Je tiens à remercier monsieur X. LELUBRE pour sa précieuse collaboration.

Je porte aussi mes remerciements aux enseignants de l'E.N.S.S.I.B pour les connaissances acquises et le soutien qu'il m'ont apporté tout au long de ce DEA.

Que toutes les personnes qui m'ont apporté leur collaboration scientifique, matérielle, morale ou amicale trouvent ici l'expression de mon amitié et de ma reconnaissance.

## TABLE DES MATIERES

<b>I. Introduction</b> .....	<b>4</b>
- Présentation .....	4
- Applications Envisagées .....	5
<b>II. Les Travaux de Recherches de l'Equipe S.A.M.I.A</b> .....	<b>7</b>
II.1 Introduction .....	7
II.2 Les constituants du mot graphique .....	8
II.2.1 Le mot maximal .....	8
II.2.2 Proclitiques (PCL) et Enclitiques (ECL) .....	9
II.2.3 Préfixes (PRF) et Suffixes (SUF) .....	10
II.2.4 La notion de Base (BAS) .....	10
II.2.5 La notion de Racine (RAC) .....	10
II.2.6 La notion de Schème (SCH) .....	11
II.3 Réalisations Informatiques déjà effectuées .....	12
<b>III. Les Dictionnaires Informatisés</b> .....	<b>15</b>
III.1 Introduction .....	15
III.2 Dictionnaire et Traitement Automatique des Langues .....	19
III.2.1 Le système de traduction SYSTRAN .....	19
III.2.2 Base de données et dictionnaire intégrés pour la traduction automatique .....	22
III.2.3 Le système de traduction ARIANE-78 .....	23
III.2.4 Lexique interactif pour l'analyse automatique du français .....	24
III.3 Dictionnaire et Traitement Automatique de l'Arabe .	26
<b>IV. Les Descripteurs Linguistiques de la Base</b> .....	<b>31</b>
IV.1 Formants-noyau (Fn) et Formants-extension (Fe) .....	31
IV.2 Les descripteurs linguistiques de la base lexicale .	34
IV.2.1 Les descripteurs linguistiques associés aux bases verbales .....	35
IV.2.2 Les descripteurs linguistiques associés aux bases nominales .....	38
<b>V. Entrées et Organisation du Dictionnaire</b> .....	<b>41</b>
V.1 Introduction .....	41
V.2 Les Entrées du Dictionnaire .....	42
V.2.1 Les morphèmes .....	42

V.2.1.1	Les bases .....	42
V.2.1.2	Les autres morphèmes .....	43
V.2.1.3	Les combinaisons de morphèmes .....	43
V.2.2	Les Méta-Entrées .....	45
V.2.2.1	Les Racines .....	45
V.2.2.2	Les Descripteurs Linguistiques .....	45
V.3	Organisation du Dictionnaire .....	46
V.3.1	Les Exceptions .....	47
V.3.2	Les informations relatives aux bases .....	47
V.3.2.1	Les informations relatives aux bases nominales .....	47
V.3.2.2	Les informations relatives aux bases verbales .....	48
VI.	Réalisation Informatique .....	49
VI.1	Choix du Langage Informatique .....	49
VI.2	Matériel Utilisé .....	49
VI.3	Structure et Contenu des Fichiers du Dictionnaire ..	50
VI.4	Principaux Algorithmes .....	56
VI.4.1	Principe de l'Algorithme de Synthèse de Bases Verbales .....	56
VI.4.2	Algorithme de Synthèse de Bases Nominales .....	56
VI.5	Utilitaires d'Ajout des Bases .....	59
VI.5.1	Ajout d'une Base Verbale .....	59
VI.5.2	Ajout d'une Base Nominale .....	61
VI.6	Conclusion .....	61
VII.	Conclusion .....	62
ANNEXES	.....	64
1 .	Liste des Proclitiques .....	67
	Liste des Préfixes .....	67
	Liste des Suffixes .....	67
	Liste des Enclitiques .....	67
2 .	Table de Compatibilité Proclitiques-Préfixes .....	69
3 .	Table de Compatibilité Suffixes-Enclitiques .....	71
4 .	Liste des Prébases .....	73
5 .	Liste des Postbases .....	75
6 .	Table de Compatibilité Prébases-Postbases .....	77
7 .	Liste des Vecteurs Suffixes .....	78
7.1	Liste des vecteurs suffixes de l'accompli .....	79
7.2	Liste des vecteurs suffixes de l'inaccompli réel ..	79
7.3	Liste des vecteurs suffixes de l'inaccompli potentiel subjonctif .....	80

7.4 Liste des vecteurs suffixes de l'inaccompli potentiel apocopé .....	80
7.5 Liste des vecteurs suffixes de l'impératif .....	81
7.6 Liste des vecteurs suffixes de l'accompli passif ..	81
7.7 Liste des vecteurs suffixes de l'inaccompli réel passif .....	82
8 . Liste des Vecteurs Préfixes .....	84
9 . L'article du Journal "AL SABAH" .....	86
10. Lexique de Définitions .....	88
<b>REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>92</b>

## I. INTRODUCTION

### Présentation

Le présent travail s'inscrit dans le prolongement des recherches effectuées à Lyon 1, à Lyon 2, et à l'E.N.S.S.I.B, sur le traitement automatique de la morphologie de l'arabe (thèses de MM. M. HASSOUN en informatique [HAS87], J. DICHY en linguistique [DIC90], vol. des Travaux S.A.M.I.A I (Synthèse et Analyse Morphologiques Informatisées de l'Arabe) [SAM89], etc.).

Il prolonge par ailleurs la recherche - donnant lieu à la réalisation d'un prototype et à la rédaction d'un mémoire - que j'ai effectué dans le cadre d'un stage de Maîtrise à Lyon 2 (Département d'Etudes Arabes) au 2ème semestre de 1989/90. Cette recherche, sous la direction de M. J. DICHY, portait sur : "Un système d'enseignement assisté par ordinateur traitant de la conjugaison de l'arabe, avec simulation de modèle linguistique".

Il s'agit de mettre au point un prototype de dictionnaire des bases nominales et verbales de l'arabe, comportant (très schématiquement) :

- un ensemble de relations entre les bases relevant d'une même racine;

- un ensemble de relations entre bases corrélées d'un point de vue morphologique (telles que relations singulier-pluriel "brisé", masculin-féminin, achevé-inachevé, etc.);

- un ensemble de traits morphosyntaxiques associés aux bases et validant ou invalidant des relations entre les bases nominales ou verbales, et les autres morphèmes du mot (préfixes ou suffixes, proclitiques ou enclitiques).

Nous distinguerons soigneusement, en cours de développement, les aspects informatiques des aspects linguistiques.

### Applications Envisagées

Un tel dictionnaire, compatible avec l'analyse morphologique trouve une application immédiate dans la réalisation d'un correcteur orthographique (Sujet de DEA en Sciences de l'Information et de la Communication pour l'année 1991/92 de M. N. GADER), mais il est susceptible d'applications ultérieures, moyennant les extensions nécessaires, en recherche documentaire (extraction du syntagme nominal, peu envisageable en arabe en l'absence d'un tel dictionnaire), ou dans d'autres domaines de traitement automatique des données linguistiques.

Apport de ce travail d'un point de vue d'informatique documentaire:

- Conception d'une base de données lexicale
  - °) conception en intension d'une base de donnée très étendue;
  - °) conception d'un système (ouvert) de gestion de cette base de donnée.

- Contribution à résoudre le difficile problème de l'indexation ou de l'analyse automatique de textes en arabe non-vocalisés.

## II. LES TRAVAUX S.A.M.I.A

### II.1 Introduction.

Le programme de recherche S.A.M.I.A. (Synthèse et Analyse Morphologiques Informatisées de l'Arabe en vue d'une application en Enseignement Assisté par Ordinateur (E.A.O)) trouve son origine dans une étude de "faisabilité" réalisée par une équipe réunie autour de J.P. Desclés, alors à l'Université Paris 7, à la demande de la sous-direction de la politique linguistique du Ministère français des Affaires Etrangères, et réalisée au cours de l'année 1982. Les travaux sont actuellement menés au Centre de Recherche en Terminologie et Traduction (C.R.T.T) de l'université Lumière Lyon 2.

Le programme SAMIA se caractérise par l'élaboration de modèles de simulation informatique de certains aspects de l'activité langagière. Ces aspects sont analysés et représentés dans des modèles linguistiques opérant en synthèse et/ou en analyse.

En l'état actuel des recherches, ce programme porte sur les connaissances linguistiques relatives à la morphologie, mais les modèles sont construits avec le souci d'extensions ultérieures vers la morpho-syntaxe.

Le modèle linguistique sur lequel s'est fondé le projet, tel qu'il est défini, comprend trois éléments principaux :

- l'analyseur
- le synthétiseur

- le dictionnaire

Grâce à une équipe pluridisciplinaire, ce modèle a subi quelques modifications, celles-ci pour faire face aux applications orientées vers l'indexation automatique (la construction d'une base de données textuelles en arabe).

Des informations détaillées concernant ces trois éléments se trouvent dans la thèse de M. Hassoun [HAS87] et dans celle de J. Dichy [DIC90], ainsi que dans un ouvrage collectif J. Dichy et M. Hassoun [SAM89].

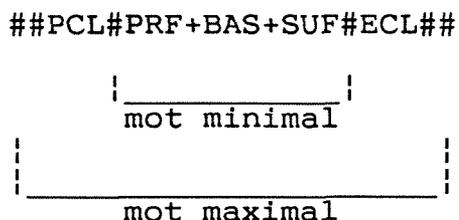
## II.2 Les constituants du mot graphique.[DIC90]

Le mot graphique en arabe pouvant comprendre plusieurs morphèmes, l'analyse aura pour tâche d'en identifier les constituants. D.COHEN , 1961/70, appelle mot maximal l'unité décomposable, selon les termes du rapport Desclés [DES83], en : proclitique(s), préfixe, base, suffixe(s), enclitique(s) (ci-après respectivement PCL, PRF, BAS, SUF, ECL).

Pour la clarté de l'exposé, nous présentons ci-dessous un ensemble de ces termes ou de ces conventions, partiellement basé sur des critères empiriques. (Ces définitions renvoyant les unes aux autres, nous nous sommes trouvé contraint d'employer par anticipation des termes dont le sens apparaît un peu plus loin; le lecteur excusera également d'inévitables répétitions.)

### II.2.1 Le mot maximal.

On peut représenter le mot selon le schéma suivant :



Lire ## comme frontière de mot graphique, # et + respectivement comme frontière faible et forte de morphème(1).

### II.2.2 Proclitiques (PCL) et enclitiques (ECL).

Les pro- ou enclitiques (classiquement appelés enclinomèmes) peuvent être désignés par le terme générique clitique. Dépouillé de ses clitiques, le mot, c'est-à-dire, la base accompagnée éventuellement de ses préfixe et suffixe, demeure une unité du discours susceptible de déplacement dans la phrase ou d'apparaître à l'initiale ou à la pause.

Les bases doivent nécessairement, en revanche, être suivies d'un suffixe, qui peut être un suffixe zéro (les bases verbales à l'inachevé doivent être précédées d'un préfixe). Les morphèmes clitiques appartiennent, comme les préfixes et suffixes, à des inventaires limités (cf Annexe 1).

Les **proclitiques** (PCL) sont des morphèmes comprenant une seule consonne, et rattachés de ce fait, dans l'écriture, au mot qui les suit. Ce sont : des coordonnants, conjonctions, prépositions, etc.

(1) Pour les frontières +, # et ##, voir N. Chomsky et M. Halle, traduction française 1973, notamment pp. 219-231.

L'article / ل / (/al/, "le") est également un proclitique, bien qu'il comporte deux segments-lettres.

Les **enclitiques** (ECL) sont des morphèmes attachés à une catégorie lexicale (noms ou verbes), qui les précède séquentiellement, l'ensemble ne formant qu'un seul mot. En arabe, les enclitiques sont des pronoms personnels compléments.

### II.2.3 Préfixe (PRF) et Suffixes (SUF).

Par convention, les **préfixes** (PRF) ne comprennent que les morphèmes verbaux (préfixés) de l'inaccompli. Le préfixe est connecté de manière nécessaire à la structure syllabique de l'inaccompli.

Les **suffixes** (SUF) sont des morphèmes situés immédiatement après la base, qui est un nom ou un verbe nécessairement suivi d'un suffixe (y compris du suffixe zéro).

### II.2.4 La notion de Base (BAS).

La **base** (BAS), pour la partie du lexique qui relève du système dérivationnel de l'arabe, s'analyse en racine et schème qu'on définira par la suite.

### II.2.5 La notion de Racine (RAC).

Le système phonologique, proto-sémitique dont la langue arabe a hérité, établit une disjonction entre les deux ensembles

constitués par les consonnes et les voyelles; il en découle que les racines du système de nomination de l'arabe, comme des autres langues sémitiques, sont consonantiques [ROM82].

Il y a deux types de racines :

(a) les racines monoconsonantiques, qui dénotent les "unités de nomination générales"; par exemple : /  $\text{ك}$  / (/mä/(\*), "quoi"), ou le pronom clitique complément /  $\text{ك}$  / (/Ka/, "toi, masculin");

(b) les racines triconsonantiques, correspondant aux "unités de nomination particulières", qui se réalisent, dans le discours, et en combinaison avec les "unités du système de communication", qui sont des voyelles ou des allongements vocaliques ou consonantiques ([ROM89], paragraphe II.1).

Les unités du système de nomination comportent également :

[1] pour les "unités de nomination particulières", des combinaisons de racines tri- et monoconsonantiques.

[2] pour les "unités de nomination générales", des combinaisons de racines monoconsonantiques.

#### II.2.6 La notion de Schème (SCH).

(\*) ä : le tréma sur la voyelle "a" signifie qu'on doit prolonger la prononciation de cette voyelle. Cette remarque est valable aussi pour les autres voyelles dans les transcriptions qui suivent.

Le schème (SCH) est la représentation de la base obtenue en substituant aux trois consonnes qui constituent habituellement la racine, respectivement R1, R2 et R3, ou comme dans la morpho-phonologie arabe médiévale, les phonèmes / ف /, / ع / et / ل /.

Le statut linguistique de la notion de schème est strictement limité à la partie du lexique de l'arabe qui est rapportable à des racines tri- ou quadri-consonantiques. Il est clair qu'une base telle que / إ ل ك ت ر و ن / (/iliktroun/, "électron") n'est pas représentable en schème et racine.

### II.3 Réalisations Informatiques déjà effectuées.

Quatres mémoires de maîtrise dus respectivement à J.AKAICHI, N.GADER, M.GHENIMA et R.NOUIRA présentés à l'Institut Supérieur de Gestion de Tunis (I.S.G) en 1990 ont été réalisés dans le cadre du programme S.A.M.I.A.

L'objet de ces quatres mémoires est la réalisation de deux prototypes de didacticiels d'enseignement des formes verbales et déverbales de l'arabe, en synthèse et en analyse. Ces deux didacticiels portent respectivement

- sur l'enseignement de la construction et de l'analyse de formes verbales, c'est-à-dire, sur la conjugaison de l'arabe, et
- sur les formes déverbales, c'est-à-dire, sur le nom verbal ( مصدر ), le participe actif ( اسم الفاعل ), le participe passif ( اسم المفعول ), l'adjectif ressemblant aux participes actif et passif ( الصفة المشبهة ) le nom de temps et de lieu ( اسم الزمان ).

Le principe général caractérisant l'approche de l'E.A.O développé dans le programme S.A.M.I.A est celui de la simulation de systèmes de connaissance. Les deux systèmes de connaissance ci-dessus (formes verbales et déverbales) sont mis en oeuvre en synthèse et en analyse : l'apprenant doit acquérir ces deux démarches. La notion de simulation implique par ailleurs l'élaboration de moteurs d'inférence susceptibles de "calculer" la réponse à la question traitée par l'apprenant, les modules d'E.A.O ayant en charge la comparaison entre les résultats de la simulation d'une part, et les réponses de l'apprenant d'autre part . En outre, la structure même de la simulation est elle même conçue de manière à aider l'apprenant à saisir la structure d'ensemble du système de connaissance qu'il cherche à acquérir (L'apprenant n'est pas seulement aidé par les exercices qui lui sont proposés, mais aussi par la représentation sous-jacente à la structure d'ensemble du système d'E.A.O avec simulation).

Concrètement, cela implique un double travail de conception:

- module de simulation
- module d'E.A.O

Pour des raisons de portabilité, le langage adopté est TURBO-PASCAL. Ce critère excluait par ailleurs l'utilisation du MS-DOS arabisé, dont le coût aurait exclu la possibilité d'utilisation dans un contexte pédagogique. Il a donc fallu élaborer une interface d'entrées/sorties en caractères arabes, compatible avec la plupart des cartes vidéo disponibles (Hercules, EGA, VGA, etc...), sans pour autant perdre la possibilité d'afficher en caractères latins.

La répartition des tâches a été la suivante :

- Jalel AKAICHI : module de simulation de la conjugaison, en synthèse et en analyse [AKA90].

- Malek GHENIMA : module d'E.A.O correspondant et programme d'arabisation des entrées/sorties, en collaboration avec N.GADER [GHE90].

- Rym NOUIRA : module de simulation de la synthèse et de l'analyse des formes déverbales [NOU90].

- Nabil GADER : module d'E.A.O correspondant et programme d'arabisation des entrées/sorties, en collaboration avec M.GHENIMA [GAD90].

### III. LES DICTIONNAIRES INFORMATISES

#### III.1 Introduction.

Les dictionnaires électroniques (informatisés) entrent dans la plupart des programmes qui mettent en jeu une procédure de reconnaissance d'unités linguistiques significantes.

Dans de tels programmes, un texte est, dans une phase initiale, soumis à une procédure de segmentation en unités graphiques (les mots). Un dictionnaire est alors consulté afin de déterminer la nature de chacune de ces unités.

Si un mot n'est pas trouvé dans le dictionnaire, l'analyse plus approfondie du texte qui devrait suivre la première phase sera bloquée, ou gênée dans le meilleur des cas. En conséquence, les mots du texte doivent être conformés aux entrées du dictionnaire avec une précision entièrement nouvelle.

Cette activité nécessite la mise en oeuvre :

- de procédures de correction et de normalisation orthographiques,
- de dictionnaires extrêmement complets,
- et, le plus souvent, des procédures de traitements de mots inconnus (noms propres, nouvelles productions).

Outre les applications à base d'analyse de textes, c'est-à-dire de reconnaissance de configurations significantes, il se développe des programmes de génération de textes qui font appel

aux mêmes informations linguistiques, il est donc important que les dictionnaires aient une forme neutre vis-à-vis de ces deux classes d'applications.

### Dictionnaires électroniques et dictionnaire usuels

Il est important de bien préciser les différences fondamentales qui existent entre les dictionnaires électroniques et les dictionnaires usuels aujourd'hui disponibles sur support informatique. Ces derniers peuvent être monolingues, généraux ou spécialisés.

Les dictionnaires électroniques sont destinés à des utilisations informatiques. Ils ont la forme de tables de données formalisées auxquelles des programmes divers ont accès. Par exemple, un vérificateur d'orthographe d'usage consultera un dictionnaire électronique, qui peut être une simple liste de mots, pour vérifier si les mots d'un texte donné figurent dans le dictionnaire. Un dictionnaire électronique consacré à l'analyse de textes devra comporter des indications grammaticales codées et formatées utilisables pour un analyseur syntaxique.

Les informations entrant dans les programmes sont par essence codifiées. Ainsi, en français, pour obtenir toutes les formes conjuguées d'un verbe, entré dans un dictionnaire sous sa forme infinitive, il est nécessaire de disposer d'un algorithme de conjugaison qui opère sur une centaine de classes de verbes : des codes alphanumériques représentant ces classes doivent donc affecter chaque entrée de verbe. Il en va de même pour la mise au

féminin ou du pluriel des adjectifs et des noms, puisqu'en général, ils sont entrés au masculin singulier pour les adjectifs, au singulier avec leur genre pour les noms. Un dictionnaire électronique syntaxique doit comporter des informations systématiques sur les possibilités de sujet et de compléments qui s'offre pour chaque verbe, ainsi que des indications formelles sur les positions et les formes diverses que peuvent prendre les sujets et compléments dans les phrases quelconques (indications sur les possibilités de passif, de pronominalisation, etc.). De telles informations ne sont pas données dans les dictionnaires usuels (e.g. Larousse de la langue française (LEXIS), Petit Larousse, Petit Robert).

Il importe de réaliser que la modernisation de l'imprimerie qui exige que les textes des dictionnaires soient composés sur support informatique n'a rien changé au contenu des dictionnaires courants. La formalisation des caractères et des règles typographiques qui est intervenue est sans rapport avec le contenu. Cette observation vaut pour les dictionnaires ou encyclopédies usuels consultables sur un terminal minitel par exemple. L'informatisation ne fait guère qu'introduire des supports de textes nouveaux, une typographie nouvelle et des procédés de consultation nouveaux du point de vue du dictionnaire en papier.

Le terme ambigu d'informatisation a parfois introduit une confusion entre les deux catégories de dictionnaires. S'il est certain que dans les deux cas des listes de mots se trouvent sous forme électronique, les informations que contient chaque type de

dictionnaire n'ont rien de commun : ce sont des codes opaques (et non publiables) dans un cas, des textes à destination d'un public étendu dans l'autre. Il n'est même pas clair que les seules listes des mots soient identiques dans les deux types de dictionnaires pour lesquels il existe des différences fondamentales de conception. Les dictionnaires usuels ont tendance à ne pas donner d'informations trop évidentes pour l'utilisateur, donc encombrantes et désagréables à lire. Par contre les dictionnaires électroniques doivent être bien complets et explicites jusqu'au moindre détail. Un dictionnaire électronique du français devra par exemple comporter une liste complète des adjectifs en -able dérivés de verbes. Comme la dérivation (accrocher-accrochable) est intuitivement régulière, les dictionnaires courants laissent au lecteur le soin d'appliquer la règle. Néanmoins, dans ces dictionnaires, on trouve parfois et de manière imprévisible des paires évidentes à constituer comme (abaisser-abaissable), à côté de cas difficiles que l'on trouve dans des verbes dits du "2<sup>ème</sup>" ou "3<sup>ème</sup>" groupe (lire-lisible, élire-éligible) qui eux sont assez systématiquement énumérés. Cette abréviation du lexique est générale, elle affecte de nombreuses formes (e.g. adjectifs en -ant, adverbes en -ment) qui doivent toutes figurer explicitement avec leurs variations dans un dictionnaire électronique, mais surtout il n'existe pas de règles pour ce type de condensation, qui ne dépend que du goût et des instructions données aux lexicographes sur la taille, donc le prix du dictionnaire.

De même, les dictionnaires du commerce diffèrent sur de nombreuses variantes orthographiques, ou encore, ils omettent des informations qui sont nécessaires à la conjugaison détaillée des

verbes, ou à la formation de féminins ou de pluriels pour les noms et les adjectifs dans les cas évidents. Par contre, comme nous l'avons déjà signalé, de telles informations détaillées et formalisées doivent toutes figurer dans les dictionnaires électroniques. Ici encore, les règles adoptées par les lexicographes pour inclure ou rejeter ces informations ne sont jamais explicitées.

On voit d'après ces quelques remarques que dictionnaires électroniques et dictionnaires usuels sur support informatique n'ont en commun que des listes partielles de mots. Il en va de même pour les lexiques bilingues : on a pu penser que les dictionnaires multilingues et les données terminologiques informatisées pourraient être utilisés par des programmes de traduction automatique. Il n'en est rien, et le principal obstacle à la réalisation de programme de TA (Traduction automatique) ou de TAO (Traduction Assistée par Ordinateur) est à la fois scientifique et économique : il consiste en une absence totale de dictionnaires électroniques multilingues dont on sait qu'ils seront très coûteux. Nous insistons sur le fait que la distance entre les lexiques spécialisés des traducteurs et les dictionnaires électroniques nécessaires à la traduction est plus grande encore que dans le cas discuté ci-dessus des dictionnaires monolingues usuels.

## III.2 Dictionnaire et Traitement Automatique des Langues [HAS87]

### III.2.1 Le système de traduction SYSTRAN.

Le système SYSTRAN (SYSTEM TRANslation) a été mis au point à partir de 1970 par P. TOMA de l'Université de GEORGETOWN [VAN79]. SYSTRAN fonctionne depuis cette date à l'U.S. Air Force pour traduire du russe en anglais. Au début de l'année 1982, la société GACHOT acquiert les droits de commercialisation du système SYSTRAN et propose plusieurs couples de langues et notamment (Anglais ---> Arabe).

La société GACHOT dans sa présentation du système SYSTRAN indique qu' "il est important de noter que les dictionnaires utilisés par le système SYSTRAN ont un caractère très particulier : non seulement ils doivent indiquer les traductions possibles de chaque terme, mais ils doivent comporter en outre un ensemble très riche d'informations afin de réaliser une construction syntaxique et grammaticale correcte en langue cible".

Le système nécessite un vaste dictionnaire dont la taille a été estimée par la Communauté Européenne à près de 50.000 entrées (mots ou expressions). En fait, le traitement de domaines technologiques particulier peut augmenter ce nombre.

#### **Les dictionnaires utilisés dans le système Anglais-Français :**

1) *Dictionnaire des unitermes (Stems)* : Pour faciliter l'accès aux données, ce dictionnaire est divisé en trois sections : termes courts, moyens et longs. Une procédure de génération permet d'obtenir automatiquement les formes régulières et les informations qui s'y rapportent.

2) *Dictionnaire des unitermes de haute fréquence* : il contient quelques éléments de la langue source qui constituent 30% des "mots" des textes à traduire. Il s'agit en particulier des signes de ponctuation, des lettres (A à Z), des chiffres, et certains mots très fréquents sans ambiguïté, syntaxique ou polysémique (AND, AN, OF, etc...).

3) *Table des idiotismes (Idioms)* : elle est divisée en deux parties, la première servant à réduire les substitutions idiomatiques unitermes avant le processus de traduction, la deuxième servant à repérer les idiotismes et à leur fournir la traduction correspondante dès le début du processus de traduction, sans empêcher chaque élément de recevoir toutes les informations grammaticales et sémantiques nécessaires pour le fonctionnement des différents programmes d'analyse et de synthèse.

4) *Dictionnaire des expressions L.S.* : Il est actionné par un système d'étiquettes au niveau des unitermes (renvoi automatique de chaque mot susceptible de constituer une expression, vers l'entrée correspondante dans le dictionnaire des expressions L.S.).

5°) *Dictionnaire des expressions C.L.S.* : Il est actionné à partir de règles contextuelles ou grammaticales, ce dictionnaire permet de traiter les ambiguïtés sémantiques dans la mesure où les conditions contenues dans une règle sont satisfaites par l'analyse du texte source.

### III.2.2 Base de données et dictionnaire intégrés pour la traduction automatique [DUR87].

La notion de dictionnaire intégré en fourche [BOI86] du GETA (Groupement pour l'Etude de la Traduction Automatique à Grenoble) est issue d'un souci de réunir dans un dictionnaire les informations suivantes :

- informations générales et terminologiques,
- informations monolingues et multilingues,
- informations naturelles et codées.

Le terme "fourche" indique que les dictionnaires sont monolingues, les traductions données pour un article étant des entrées dans les dictionnaires intégrés des langues correspondantes.

On désire pouvoir générer des vues à partir d'une même source : par exemple un dictionnaire de traduction pour un couple de langues donné, ou des dictionnaires d'analyse ou de transfert codés, ou encore un recueil des verbes propres à un domaine technologique donné, etc. Cette unification permet le maintien de la cohérence des "vues" obtenues.

Un article de dictionnaire intégré a la structure d'un article de dictionnaire naturel. Une entrée est donc formée du nom de cet article, de sa classe syntaxique, des définitions de ses sens, d'exemples et de ses traductions.

Pour les applications informatiques (telles que la traduction automatique ou la synthèse de la parole) , on greffe les informations codées sur cette structure.

Une maquette d'un dictionnaire intégré a été réalisée au moyen du langage "Prolog". Des fonctions d'accès et de vues sur ce dictionnaire ont été définies.

### III.2.3 Le système de traduction ARIANE-78 .

"Le système "ARIANE-78", conçu et réalisé au GETA, a pour but de fournir aux linguistes un outil complet et d'utilisation simple, pour la mise au point d'une chaîne de traduction automatique multilingue".

Nous allons nous limiter ici à la présentation des différents dictionnaires employés par le système.

Il y a trois sortes de dictionnaires : de bases, d'affixes et de tournures. Chaque dictionnaire est une liste d'article dont voici la syntaxe simplifiée :

<art. de D. de bases> ::= <morphe> == <format M> (<format S ou G> , <UL>).

<art. de D. d'affixes> ::= <morphe> == <format M> (<format S ou G>).

<art. de D. de tournures> ::= <tournure> == <format M> (<format S ou G>, <UL>).

<morphe> ::= <suite de symboles non blancs de moins de 34 car.>

<tourneure> ::= <suite de symboles, sans sous-suite de deux blancs, de 34 car. au plus>

<format i> ::= <identificateur>

La syntaxe est positionnelle, le signe "==" étant en colonnes 35 et 36.

On accède aux dictionnaires par les morphes ou les tournures. Ces dictionnaires contiennent toutes les chaînes élémentaires intervenant lors du découpage d'un texte par l'automate. On peut distinguer : les formes complètes, les mots invariables, les morphes libres et les morphes liés. Le système permet l'utilisation de un à sept dictionnaires, dictionnaire des tournures inclus. On a donc la possibilité de créer plusieurs dictionnaires pour la répartition des bases et des désinences. Cette possibilité jointe à l'utilisation des différentes fonctions dans les règles de grammaire permet de guider la consultation des dictionnaires, en vue d'accélérer le découpage et l'interprétation des textes et d'éviter les résultats parasites.

#### III.2.4 Lexique interactif pour l'analyse automatique du français.

Le travail de A. EYANGO-MOUEEN est une composante d'un projet plus général mené par le Laboratoire d'Informatique Documentaire de l'Université Lyon 1 au sein du groupe SYDO(\*) visant à l'élaboration d'un système automatisé d'informations textuelles.

(\*) Le groupe SYDO regroupait les équipes de recherche provenant des institutions suivantes :

- Laboratoire d'Informatique Documentaire de l'Université Lyon I
- Centre de Recherche en Linguistique et Sémiologie de l'Université Lyon II
- Centre de Recherche en Informatique pour les Sciences Sociales (CRISS) de l'Université Grenoble II

Le groupe SYDO travaillait depuis plusieurs années sur les problèmes d'analyse morphosyntaxique, avec deux voies parallèles et complémentaires :

- Une recherche à long terme et de nature fondamentale relative à la définition du modèle linguistique le mieux adapté au traitement automatique d'une langue dans une optique de reconnaissance.

- Une recherche à court terme et de nature plus appliquée, qui consiste à développer des outils de traitement automatique des langues en vue de la conception de nouveaux systèmes d'informations spécialisées.

Le lexique, partie intégrante de l'analyseur morphosyntaxique, est l'objet du travail d'A. EYANGO-MOUCEN. Il est conçu sous la forme d'une base de données relationnelle. Les informations fournies par le modèle linguistique sont structurées pour constituer un ensemble de traits linguistiques susceptibles d'être attribués à toute forme rencontrée dans un texte en français.

### Structuration des données linguistiques :

#### (1) La notion de "PROFIL" : aspect grammatical du lexique

"Le profil d'une base est une suite d'informations linguistiques hors-contexte qui lui est attribuée en se référant à son comportement dans une phrase et éventuellement à sa morphologie". On distingue trois types de "profils" :

- Le "profil syntaxique" : Il est constitué par une suite d'informations essentiellement de type syntaxique.
- Le "profil lexical" : Il est constitué d'un ensemble d'informations du type lexical exclusivement.
- Le "profil flexionnel" : Il est constitué d'informations de type flexionnel.

## (2) La notion de BASE : aspect morphologique

On se limite à la morphologie flexionnelle ce qui suppose que les mots dérivés font partie du lexique aussi bien que les bases.

## (3) La notion de BASE VIRTUELLE

Cette notion a été introduite pour qu'on puisse prendre en compte dans le lexique un maximum de cas d'irrégularités de surface, à savoir : les amalgames orthographiques et morphographiques, les formes élidées et les variantes orthographiques.

### III.3 Dictionnaire et Traitement Automatique de l'Arabe. [HAS87]

Le travail qu'on va présenter faisait l'objet de la thèse de M. HASSOUN [HAS87].

Ce travail avait pour objet l'élaboration d'un dictionnaire pour le traitement automatique de l'arabe dans le cadre du programme de recherche SAMIA en vue d'une application d'E.A.O.

Partant du modèle linguistique conçu par le programme de recherche SAMIA (traits morphologiques, règles de contextualisation), il a été question au cours de ce travail de définir une structuration de données permettant la construction d'un dictionnaire utilisable dans les deux contextes de la synthèse et l'analyse de la morphologie, ainsi que de l'E.A.O avec simulation d'un modèle linguistique.

Le dictionnaire a été conçu sous forme d'une base de données relationnelle. En voici une synthèse des relations les plus importantes :

### 1. Relation schème

Cette relation contient tous les schèmes attestés dans la morphologie de la langue arabe.

### 2. Relation classe-syntaxique

Une classe syntaxique regroupe un ensemble de catégories morphologiques de bases et se définit par l'ensemble de morphèmes non-lexicaux qui sont compatibles avec celles-ci.

Cette relation permet d'établir le lien entre un numéro de classe syntaxique et une catégorie morphologique de cette classe.

### 3. Relation morphèmes autres que la base

Cette relation regroupe tous les morphèmes de l'arabe qui sont susceptibles d'accompagner les bases dans les mots écrits de la langue (proclitiques, préfixes, suffixes et enclitiques).

#### 4. Relation vecteur-descripteur

Cette relation définit les descripteurs linguistiques relatifs aux bases verbales et nominales.

#### 5. Relation vecteur-suffixe-casuel

Ce vecteur regroupe les suffixes casuels acceptés par une base donnée dans les deux cas possibles : déterminé et indéterminé.

#### 6. Relation classe de suffixes lexicalisés

Cette relation regroupe les informations relatives à la classe de suffixes lexicalisés (une classe permettant d'identifier tous les suffixes lexicalisés acceptés par une base donnée).

#### 7. Relation vecteur-affixe

Cette relation contient les différents vecteurs suffixes ou préfixes. Chaque vecteur suffixe ou préfixe représentant les suffixes ou les préfixes acceptés par une base ou par un ensemble de bases.

#### 8. Relations prébase et postbase

Ce sont les relations qui contiennent respectivement la liste des prébases et la liste des postbases.

#### 9. Relation racine

Cette relation contient la liste des racines.

#### 10. Relation verbes

Cette relation contient la liste des verbes arabes.

#### 11. Relation module de conjugaison

Cette relation contient la liste des modules de conjugaison. Un module de conjugaison se caractérise par l'ensemble des informations suivantes :

- le verbe module conjugué à la 3PMS(\*) de l'accompli et à l'inaccompli réel actif,
- le type des racines des verbes concernés ,
- le numéro de la forme verbale dérivée.

#### 12. Relation base schématique

Cette relation contient la liste des bases schématiques. Une base schématique étant obtenue par une simple substitution de chaque lettre radicale des bases du verbe par son rang dans la racine.

(\*) signifie 3<sup>ème</sup> personne du masculin singulier /هو/ (/Houa/, "il").

### 13. Relation Base

Cette relation contient la liste des bases.

### 14. Relation base-descripteur

Cette relation permet de réaliser une association entre les bases et leurs descripteurs linguistiques (qui se trouvent dans la relation vecteur-descripteur).

Une implémentation de ce dictionnaire a été faite sur le S.G.B.D Informix (Système de Gestion de Bases de Données sous le système d'exploitation UNIX).

#### IV. LES DESCRIPTEURS LINGUISTIQUES DE LA BASE

Jusqu'à présent, nous avons vu de quoi était constitué un mot graphique (de proclitiques, préfixes, suffixes, etc...) et comment sont situés les dictionnaires informatisés au niveau du traitement automatique des langues ainsi que dans le cas particulier de l'arabe. Il nous reste maintenant à présenter les descripteurs linguistiques de la base qui ont une part importante dans la conception du dictionnaire.

Afin de présenter ces descripteurs, il nous faut tout d'abord définir les termes formants-noyau (Fn) et formants-extensions (Fe).

##### IV.1 Formants-noyau (Fn) et Formants-extensions (Fe)[DIC90]

Cette structure peut se résumer ainsi :

[1°] le mot est une unité linguistique dont la manifestation la plus aisément observable est le mot graphique (ce qui implique une conception de la langue comme système compatible avec ses deux manifestations orale et écrite).

[2°] Les morphèmes constitutifs de l'unité-mot sont appelés des formants de mot, c'est-à-dire, des signes linguistiques minimaux dont les relations de contextualisation sont limitées aux autres morphèmes inclus dans l'unité composée que constitue le mot dans sa manifestation graphique.

La convention de définition des formants du mot (une autre convention serait, par exemple, celle du formant de phrase) donne un fondement formel à ce que le rapport Desclés [DES83] appelait la grammaire des formants du mot.

[3°] Ces formants peuvent être répartis en formant-noyau (Fn) et formants-extensions (Fe). Ils sont caractérisés par les deux types de relations :

(a) la relation d'ordre, représentée par les positions occupées par le formant sur le vecteur de représentation du mot ;

(b) un ensemble de relations de collocation, qui ont pour objet :

- de valider ou d'invalider la compatibilité des formants entre eux (par exemple, le proclitique /*شـ*/, qui est la marque du futur n'est compatible qu'avec des préfixes, bases et suffixes du paradigme verbal de l'inachevé);
- de permettre les modifications affectant un nombre important des formants, en fonction de leur contexte (par exemple : modifications de bases verbales relevant de racines anomales selon la personne , le genre et le nombre).

La grammaire des formants du mot doit être construite, en ce qui concerne la manifestation graphique de cette unité, de manière à être compatible avec les deux fonctionnements de la synthèse et de l'analyse, en tenant compte de la dissymétrie qui doit amener la première à produire des formes "vocalisées" et la seconde à analyser des mots "non-vocalisés" [DIC87].

[4°] L'unité-mot ainsi définie est donc caractérisée structurellement par :

(a) une relation d'ordre au sens strict entre les formants qui la constituent.

(b) la présence du formant-noyau (Fn) correspondant à une catégorie lexicale (la base), i.e. à des unités d'inventaires non limités, et celle d'un ensemble de formants-extensions (Fe), correspondant aux autres champs du vecteur de représentation du mot (VRM), instanciés par des fonctionnels, des modalités ou des pronoms, et dont les unités sont en nombre limité.

Ce dernier point implique que l'on distingue au sein de l'ensemble des mots graphiques entre :

- les mots-formes, qui incluent un noyau lexical, et
- des mots tels que / <sup>◦</sup>بِكُوم / (/Bikom/, "par-vous") et / <sup>◦</sup>لِيَا نَكُوم / (/LiaNNaKom/, "parce-que-vous") qui en sont dépourvus.

[5°] Le Fn ne coïncide pas nécessairement avec une unité lexicale (UL). Certains Fe sont en effet susceptibles, lorsqu'ils sont associés à une base nominale, de se trouver, pour ainsi dire, pris avec elle dans un processus de lexicalisation. Un formant-extension sera dit lexicalisé (appelé formant-extension lexicalisé (Fel)) lorsque l'unité <Fn, Fel> résultant de son association avec un formant-noyau constitue une unité du lexique (UL) indépendante.

Une UL est donc constituée :

- soit d'un Fn (on écrira : UL = <Fn>); c'est le cas des bases verbales, ou des bases nominales dépourvues de Fe lexicalisé, par exemple : / نَادِر / (/Nädir/, "rare");

- soit d'un ensemble UL = <Fn, Fel>, où Fel peut inclure plus d'un formant, et dont l'ordre séquentiel est assigné par la grammaire des formants du mot;

par exemple, / نَادِرَةٌ / (NädiRa) (Plur. / نَوَادِرُ / (Naouädir)), "chose rare", "bon mot", constitue une UL distincte de l'exemple précédent, analysable ainsi :

UL = <Fn = / نَادِرُ /, Fel = / ة />.

#### IV.2 Les descripteurs linguistiques de la base lexicale [DIC90]

Associés aux bases dans le dictionnaire propre au système, les descripteurs linguistiques gèrent les rapports entre ces dernières et les formants appartenant aux autres champs du vecteur, c'est à dire, les relations entre le formant-noyau (Fn) et les formants extensions (Fe). Il s'agit d'indications codées ou formulées en langage naturel représentant les relations contextuelles de nature morpho-syntaxique ou sémantique entre la base lexicale et les autres formants du mot-forme (cf. M.O. Hassoun, [HAS87] et, en ce qui concerne les bases verbales, N. Abu Al-Chay, [ABU88]).

Les descripteurs linguistiques sont exclusivement associés aux bases lexicales :

- Par convention, les informations relatives aux règles de collocation gérant les relations des Fe entre eux, sont en effet incluses dans la grammaire des formants du mot.

- Les descripteurs portant sur les relations entre Fe et Fn pouvaient être associés soit aux premiers, soit aux seconds; mais

c'est naturellement aux  $F_n$ , dont l'inventaire, à la différence des  $F_e$ , n'est pas limité, qu'il convient de les rattacher, pour des raisons de faisabilité évidentes. (si l'on veut que les descripteurs linguistiques soient en nombre fini, il convient de les associer aux bases lexicales i.e. aux  $F_n$  : les relations  $F_n \rightarrow F_e$  sont en effet limitées par le nombre relativement restreint des  $F_e$ , ce qui ne serait pas le cas des relations  $F_e \rightarrow F_n$ , l'inventaire des  $F_n$  n'étant pas fini.)

Voici la liste de ces descripteurs :

#### IV.2.1 Les descripteurs linguistiques associés aux bases verbales

Trois catégories de descripteurs, dues à la modélisation du traitement automatique des verbes arabes en synthèse et en analyse établie par N. Abu Al-Chay [ABU88], permettent de gérer les relations entre bases, préfixes et suffixes verbaux, quelle que soit la racine de la base. (On sait que les racines anormales entraînent dans la conjugaison de l'arabe des modifications tant de la base que des suffixes, voire dans quelques cas, des préfixes.)

Ce sont, en désignant les catégories de descripteurs de bases verbales par les initiales DBV assorties d'un numéro :

DBV N°1 : Le paradigme de conjugaison de la base verbale concernée.

DBV N°2 : Le "numéro de module" identifié par N. Abu Al-Chay [ABU88], qui désigne le type du verbe en fonction de la composition de sa racine, et pour les verbes de forme non-augmentée, de la voyelle de la deuxième radicale à l'achevé et l'inachevé.

DBV N°3 : Le "numéro de vecteurs-suffixes", correspondant à la liste des suffixes verbaux compatibles avec la base verbale concernée. Les paradigmes de conjugaison des verbes de racines anormales comportent en effet plusieurs réalisations de la base verbale, chaque réalisation étant compatible avec un sous-ensemble déterminé de la liste des suffixes.

A quoi s'ajoute trois DBV propres aux relations entre bases verbales et enclitiques :

DBV N°4 : Le caractère transitif (T) ou non (nT) du verbe, c'est-à-dire, sa compatibilité, ou son incompatibilité avec un pronom enclitique (ECL). Dans le premier cas, il y a lieu de considérer si le verbe admet des compléments d'objet humains ou non : dans l'affirmative, il est compatible avec tous les pronoms enclitiques, ce qui ne suppose pas de code particulier; dans la négative, le verbe exclut les pronoms ECL du masculin pluriel, ainsi que la première et deuxième personne, ce qui nécessite un descripteur "transitif à des non-humains" (Tnh). On a donc, par réduction, trois descripteurs de transitivité : DBV N°4 = { T, nT, Tnh} .

DBV N°5 : Une règle générale interdit en arabe un complément d'objet de la première ou de la deuxième personne si le sujet du verbe (i.e., dans cette langue, le préfixe ou le suffixe inclus dans la forme verbale) est de la même personne. La grammaire des formants du mot comporte donc une incompatibilité entre les préfixes (PRF) ou suffixes (SUF) de la 1<sup>ère</sup> ou de la 2<sup>ème</sup> personne et les pronoms ECL correspondants. Cette règle doit cependant être levée avec un petit nombre de verbes, par exemple :

(أَظُنُّنِي, "je me crois") ou / تَرَكَكَ / ("tu te vois", au sens "mental" du verbe voir). d'où la nécessité d'un descripteur X correspondant, qui ne doit cependant apparaître qu'avec les verbes concernés; on écrira donc :

$$\text{DBV N}^\circ 5 = \{ X, 0 \}$$

DBV N°6 : Le double enclitique. Certains verbes admettent un double complément d'objet, le premier étant toujours humain (DBV N°4 = T), le second pouvant être humain ou non-humain selon le sens du verbe. d'où un descripteur Y (si les ECL n'admettent pas de restriction) et Y' (si le deuxième ECL est incompatible avec les pronoms dont le référent est exclusivement humain), qui ne doit être associé qu'aux verbes concernés. On a donc DBV N°6 = {Y, Y', 0}.

Une dernière catégorie de descripteurs concerne la structuration du lexique propre aux langues sémitiques :

DBV N°7 : Le fléchage Base --> Base. On dit qu'il y a fléchage entre deux bases lorsqu'une d'entre elles est supplétive de l'autre, c'est à dire, lorsque l'une d'entre elles remplace

l'autre dans les paradigmes morpho-syntaxiques, d'une manière qui ne peut être prédite par des règles (opérant au sens strict). En ce qui concerne les formes verbales, les fléchages concernent surtout :

- la relation entre l'achevé et l'inachevé de la forme simple (non-augmentée) du verbe (problème de la voyelle de la deuxième radicale à l'achevé et à l'inachevé, dont la prédictibilité est loin d'être assurée par les indications de "tendances générales" que l'on trouve dans les ouvrages pédagogiques, où elles ont leur utilité);

- la relation entre le verbe et ce que l'on appelle couramment ses dérivés, notamment le nomen verbi ( مصدر ), qui n'est pas prédictible dans un nombre important de cas.

#### IV.2.2 Les descripteurs linguistiques associés aux bases nominales

Les catégories de descripteurs linguistiques des bases nominales (DBN) sont au nombre de six :

DBN N°1 : La catégorie de la base nominale, en termes de :

- {Sing. {Masc., Fém.}, Pluriel {Masc., Fém.}, Collectif}
- {Nom propre, Nom commun}
- l'unité lexicale (UL) à laquelle elle appartient ou qu'elle constitue, en fonction des Fe lexicalisées (Fel) associés ou non à la base nominale.

On a donc  $UL = \{Fn, \langle Fn, Fel \rangle\}$

Rappelons que le Fn coïncide avec la base, que Fel peut inclure plus d'un formant, et que l'ordre séquentiel d'occurrence

des éléments de l'ensemble <Fn, Fel> est assigné par la grammaire des formants du mot.

La catégorie DBN N°4 comportera l'indication du ou des Fel associés à la base nominale.

DBN N°2 : La catégorie du fléchage Base --> Base. Il s'agit, pour les bases nominales, de relations :

- Singulier <---> Pluriel,
- Masculin <---> Féminin,
- Base <---> forme de la Base associée au Fe / ~~o~~ / du nom-  
adjectif de relation.

On tiendra compte de ce que le fléchage intervient dans l'organisation du dictionnaire informatisé que nous proposons, entre deux bases, uniquement en tant qu'elles constituent une partie d'une unité lexicale ou qu'elles coïncident avec elle.

DBN N°3 : Le vecteur-suffixes casuels correspondant à la liste des SUF de cas compatibles avec la base considérée. Cette catégorie de descripteurs prend en charge les déclinaisons dites "diptotes", ainsi que les modifications affectant la base et les suffixes casuels lorsque la base comporte une racine anormale.

DBN N°4 : Le numéro de Fe lexicalisé (Fel), pris dans la liste suivante :

- [0] absence de Fel.
- [1] SUF / ~~o~~ / du nom-adjectif de relation.
- [2] SUF / ~~ö~~ / le plus fréquemment : marque du féminin.

- [3] SUF / **ات** / pluriel du précédent.
- [4] SUF / **ون** / du masculin pluriel, cas sujet.
- [5] SUF / **ين** / cas direct et indirect de [6].
- [6] SUF / **ون** / ou / **ين** /, alternant selon le cas.
- [7] SUF / **ان** / du duel, cas sujet.
- [8] SUF / **ين** / du duel, cas sujet.
- [9] SUF / **ان** / ou / **ين** /, alternant selon le cas.
- [10] SUF / **ه** / du cas direct non-déterminé.
- [11] PCL / **ا** / article.

DBN N°5 : Le(s) numéro(s) de Fe incompatible avec la Base.

La liste est la même que pour les DBN N°4.

DBN N°6 : Le double enclitique. Certains noms, correspondant à un nomen verbi ( **مصدر** ), admettent, tout comme le verbe auxquels ils sont associés par la dérivation (voir la catégorie DBV N°6), un double complément d'objet, le premier étant toujours humain , le second pouvant être humain ou non-humain selon le cas du verbe associé. D'où un descripteur Y (si les ECL n'admettent pas de restriction) et Y' (si le deuxième ECL est incompatible avec les pronoms dont le référent est exclusivement humain), qui ne doit être associé qu'aux "noms verbaux" concernés. On a donc :

$$\text{DBN N°6} = \{ Y, Y', 0 \}.$$

Les données présentées dans ce chapitre sont reprises de la thèse de J. Dichy [DIC90]. Les deux chapitres suivants présenteront, en revanche, la synthèse que nous en avons extraite, en vue de la réalisation d'un prototype de dictionnaire.

## V. ENTREES ET ORGANISATION DU DICTIONNAIRE

### V.1 Introduction.

Un examen approfondi des mots présents dans un texte arabe écrit, vocalisé ou non, nous montre qu'on peut les répartir en trois grandes classes :

#### Classe 1

Elle regroupe les mots, dont la base est représentable par un schème et une racine.

##### Exemple :

" استَرَدَّ " (/isTaRaDDa/) qui a comme schème " استَفْعَلَ " (/isTaFaLa/) et comme racine " رَدَدَ " .

Il s'agit des mots qui sont assimilés par le système morphologique dérivationnel de l'arabe.

#### Classe 2

Elle regroupe les mots non assimilés par le système morphologique dérivationnel de l'arabe. Ici on trouve plusieurs catégories de mots :

- des emprunts;

##### Exemples :

" كُمْبِيُوْتَر " ("Computer"), " تِكْنُوْلُوْجِيَا " ("Technologie"), etc.

- des noms propres

Exemples :

" تونس " (/TouNiSSa/, "Tunisie"), " فرانسسا " (/FiRaNSSa/, "France"), etc.

Classe 3

Elle regroupe des mots outils dont certains sont composés de clitiques, de préposition et des pronoms affixes.

Exemple : " له " (/LaHou/, "pour lui"), " به " (/BiHi/, "avec lui"), " بهم " (/BiHim/, "avec eux"), " في " (/Fi/, "dans"), " كي " (/Kay/, "pour"), " منذ " (/MonDou/, "depuis"), etc.

V.2 Les Entrées du Dictionnaire.

On trouve deux grands groupes d'informations :

- les morphèmes
- les méta-entrées (il s'agit de constituants abstraits tel que la racine).

V.2.1 Les morphèmes.

On peut répartir les informations concernant les morphèmes en trois groupes :

- les bases;
- les autres morphèmes;
- les combinaisons de morphèmes.

V.2.1.1 Les bases.

Les bases nominales et verbales seront représentées sous deux formes :

- La forme graphique non vocalisée;
- la forme graphique vocalisée.

Exemples :

\*) Le mot " قَالَ " (/qäla/, "dire") a comme :

- base graphique non vocalisée " قال "
- base graphique vocalisée " قَال "

\*) Le mot " يَكْسِرُونَ " (/Yaksirün/, "ils cassent") a comme :

- base graphique non vocalisée " كسر "
- base graphique vocalisée " كَسِر "

V.2.1.2 Les autres morphèmes.

On peut préciser, en se référant aux constituants du mot graphique (§ II.2), quatre types de morphèmes autres que la base:

- des proclitiques,
- des préfixes,
- des suffixes,
- des enclitiques.

V.2.1.3 Les combinaisons de morphèmes.

Le modèle linguistique, développé au sein de l'équipe SAMIA, segmente le mot à analyser en trois sous-chaînes : prébase, base et postbase. La consultation du dictionnaire en analyse permettra d'attester leur existence. Cela nécessite d'enregistrer dans le dictionnaire les éléments correspondants.

(1) la liste des prébases (cf annexe 4) :

On obtient cette liste par la combinaison des éléments de la liste des proclitiques et de la liste des préfixes, en respectant la grammaire des formants du mot. Un élément de cette liste est composé :

Au *minimum* :

- d'un préfixe;
- ou d'un proclitique;

et au *maximum* :

- de trois proclitiques avec un préfixe;
- ou de quatre proclitiques.

Exemples :

\*)  $\text{وَيَكْتُوبُ}$  ("et il écrit") la prébase est "  $\text{وَيَ}$  ", un proclitique "  $\text{و}$  " et un préfixe "  $\text{يَ}$  " ;

\*)  $\text{أَنْتَ إِفْرَاقٌ}$  la prébase est "  $\text{أَنْتَ إِفْرَاقٌ}$  ", deux proclitiques "  $\text{أَنْتَ}$  " et "  $\text{إِفْرَاقٌ}$  ".

(2) La liste des post-bases (cf annexe 5) :

Cette liste comme la précédente, est obtenue par la combinaison d'éléments de la liste des suffixes avec des éléments de la liste des enclitiques, en respectant la grammaire des formants du mot. Un élément de cette liste est composé :

- d'au moins un morphème;
- d'au plus cinq morphèmes (trois suffixes et deux enclitiques).

Exemple :

\*)  $\text{يَكْتُبُونَهُ}$  ("ils l'écrivent") la postbase est "ونكه", un suffixe "ون" et un enclitique "ه";

V.2.2 Les Méta-Entrées.

V.2.2.1 Les Racines

Ce sont les racines correspondantes aux bases (nominales ou verbales) des mots de la classe 1. Plusieurs bases (nominales ou verbales) peuvent avoir une racine commune; c'est pour cela (on le verra avec plus de détail par la suite) que ces bases sont rattachées à cette même racine commune.

V.2.2.2 Les Descripteurs Linguistiques

Les descripteurs linguistiques des bases verbales ou bases nominales sont représentés de façon normalisée afin de pallier aux différentes éventualités.

Par exemple pour la relation Singulier <---> Pluriel un mot peut avoir plus d'un pluriel ou ne pas en avoir du tout.

Exemples :

\*) Le mot "أَدْنَى" (/aDNä/) a trois pluriels "أَدْنَى", "أَدْنَى", et "أَدْنُونَ";

\*) par contre le mot "حَيْرَةٌ" (/HayRaToun/, "inquiétude") n'a pas de pluriel.

V.3 Organisation du Dictionnaire.

Le dictionnaire, représenté sous forme de base de données, peut être divisé en deux groupes principaux de données :

### V.3.1 Les Exceptions.

Il y a deux types de listes d'exceptions :

1) **Liste de Type 1** : Cette liste contient essentiellement les mots appartenant à la classe 3 (cf § V.1). Pour être le plus exhaustif possible, cette liste contient aussi toutes les combinaisons possibles des mots outils et des prépositions avec les clitiques.

L'intérêt d'enregistrer toutes les combinaisons possibles est de faciliter la procédure d'analyse et ainsi alléger le traitement donc le temps de réponse.

Cette liste d'exceptions étant finie, on pourra ultérieurement au niveau de la réalisation informatique la maintenir constamment en mémoire centrale afin d'éviter des accès fréquents au fichier qui la contiendra. Au début de la réalisation informatique, un fichier est préférable car il pourra être mis à jour chaque fois qu'il y aura une exception à rajouter. Une fois qu'on est sûr que la liste est complète, on pourra la mettre dans une table en mémoire centrale.

2) **Liste de Type 2** : Cette liste contient essentiellement les mots appartenant à la classe 2 (cf § V.1). Il y a d'autres mots, n'appartenant pas à la classe 2, qui entrent dans cette liste et dont la base n'est pas représentable par un schème et une racine, mais par contre ont des descripteurs linguistiques.

Exemple : " سَفَرَجَل " (/safarjal/, "coing") qui a un pluriel (assimilé par le système) : " سَفَارِيْج " (/safärij/).

La différence entre ces deux listes d'exceptions est que la première (liste de type 1) contient des éléments qui n'ont pas de descripteurs linguistiques tandis que ceux de la deuxième (liste de type 2) ont des descripteurs linguistiques.

### V.3.2 Les informations relatives aux bases.

L'information racine est précisée aussi bien pour les bases nominales que pour les bases verbales. Dans beaucoup de cas la racine est commune à des bases nominales et verbales en même temps.

#### V.3.2.1 Les informations relatives aux bases nominales.

\*) *La catégorie* : il s'agit des combinaisons possibles de "singulier", "pluriel" et "collectif" avec "masculin" et "féminin". Voici les valeurs possibles :

"SM" : Singulier-Masculin.

"SF" : Singulier-Féminin.

"PM" : Pluriel-Masculin.

"PF" : Pluriel-Féminin.

"C" : Collectif.

\*) *L'Unité Lexicale* : elle peut prendre soit la valeur <Fn> ou <Fn, Fel> (cf DBN N°1 § IV.2.2).

La précision de l'unité lexicale entraîne la précision d'une liste de formants extensions (Fe) (cf DBN N°4 et DBN N°5 § IV.2.2).

\*) *Le Pluriel* : une base nominale peut avoir ou ne pas avoir de pluriel. Elle peut aussi en avoir plusieurs en même temps.

\*) *Le Féminin* : c'est la forme du féminin de la base nominale. Elle peut ne pas en avoir.

#### V.3.2.2 Les informations relatives aux bases verbales.

\*) *Le Verbe* : c'est le verbe correspondant à la base verbale.

\*) *La Transitivité du verbe* : il faut préciser si le verbe est transitif ou non (cf DBV N°4 IV.2.1).

\*) *Le Nom Verbal* ( مصدر ) : une base verbale peut avoir ou ne pas avoir de nom verbal.

Une base verbale peut aussi avoir plus d'un nom verbal en même temps.

## VI. REALISATION INFORMATIQUE.

### VI.1 Choix du Langage Informatique

Nous avons opté pour une programmation procédurale, en utilisant le langage PASCAL pour des raisons d'efficacité, de portabilité, et techniques.

Nous avons programmé à l'aide de la version 4.0 de TURBO-PASCAL qui est un langage fait pour satisfaire les besoins des utilisateurs des IBM PCs et compatibles. C'est un langage structuré, de haut niveau qui permet d'écrire les programmes de différents types et de différentes tailles.

Le langage PASCAL nous a permis d'effectuer un traitement graphique sur les caractères arabes, ainsi qu'un traitement efficace sur les chaînes de caractères utilisées fréquemment dans notre programmation.

### VI.2 Matériel Utilisé.

Notre prototype de dictionnaire a été développé sur un micro-ordinateur, compatible IBM PCs, MICROSPOT possédant les caractéristiques techniques suivantes :

- Micro-Processeur : 386 SX à 16 Méga Hertz (MHZ),
- Mémoire Vive (RAM) : 2 Méga Octets (MO),
- Capacité Disque Dur : 40 Méga Octets (MO),
- Ecran Couleur,

- Clavier étendu 102 touches,
- Système d'exploitation : MS-DOS Version 5.0,
- Carte Graphique : VGA .

### VI.3 Structure et Contenu des Fichiers du Dictionnaire

#### FRacine : Fichier des Racines

CODE	INFORMATION	TYPE	TAILLE
NRac	Numéro de la Racine	Entier	
Racine	La Racine des Bases Nominales et Verbales	C(\$)	4

#### FVerbe : Fichier des Verbes

CODE	INFORMATION	TYPE	TAILLE
NRac	Numéro de la Racine de la Base Verbale	Entier	
NClas	Numéro de la Classe de conjugaison du Verbe	Entier	
Verbe	Verbe à la 3 <sup>ème</sup> personne de l'accompli	C	10
DBV4	Descripteur de la Base Verbale 4 (§ IV.2.1)	Entier	
DBV5	Descripteur de la Base Verbale 5 (§ IV.2.1)	Entier	
DBV6	Descripteur de la Base Verbale 6 (§ IV.2.1)	Entier	
DBV7	Descripteur de la Base Verbale 7 (§ IV.2.1)	Entier	

(§) C veut dire chaîne de caractère

**FBasVerb : Fichier des Bases verbales**

CODE	INFORMATION	TYPE	TAILLE
NClas	Numéro de la Classe de conjugaison du Verbe	Entier	
BasVerb	La Base Verbale	C	20
Nvs	Numéro du Vecteur Suffixe (cf annexe 7)	Entier	

**FBasNom : Le Fichier des Bases Nominales**

CODE	INFORMATION	TYPE	TAILLE
NRac	Numéro de la Racine de la Base Nominale	Entier	
NBasN	Le Numéro de la Base Nominale	Entier	
BasNom	La Base Nominale	C	20
Categ	La Catégorie de la Base Nom. ( § V.3.2.1)	C	2
UL	L'Unité Lexicale ( § V.3.2.1 ) (*)	Entier	
CPlur	Le Code d'accès à la forme du Pluriel	Entier	
CFem	Le Code d'accès à la forme du Féminin	Entier	
DBN2	Descripteur de la Base Nom. 2 ( § IV.2.2)	Entier	
DBN3	Descripteur de la Base Nom. 3 ( § IV.2.2)	Entier	

(\*) L'unité lexicale prend deux valeurs possibles :

- 0 pour <Fn>.
- 1 pour <Fn, Fel>.

FPluriel : Le Fichier des Formes du Pluriel

CODE	INFORMATION	TYPE	TAILLE
CPlur	Le Code de la forme du Pluriel	Entier	
Pluriel	La Forme du Pluriel	C	20

FFéminin : Le Fichier des Formes du Féminin

CODE	INFORMATION	TYPE	TAILLE
CFem	Le Code de la forme du Féminin	Entier	
Feminin	La Forme du Féminin	C	20

Ffe : Le Fichier des Formants Extensions

CODE	INFORMATION	TYPE	TAILLE
NBasN	Le Numéro de la Base Nominale	Entier	
ListFe	La liste des Fe sous forme de Vecteur	Vect1 (1)	

(1) Le type Vect1 est composé de onze éléments de valeur binaire (0 ou 1) (cf DBN N°4 § IV.2.2) et a la structure suivante :

01	02	03	04	05	06	07	08	09	10	11
0	1	1	0	1	0	1	0	1	0	1

0 : veut dire que le Fe correspondant n'est pas possible.  
 1 : veut dire que le Fe correspondant est possible.

FDBN2 : Le Fichier du Descripteur de la Base Nominale N°2

CODE	INFORMATION	TYPE	TAILLE
DBN2	Descripteur de la Base Nominale N°2	Entier	
BasAm	La Base Amputée	C	20
SufAm	Le Suffixe à Rajouter à la Base Amputée	C	8

FDBN3 : Le Fichier du Descripteur de la Base Nominale N°3

CODE	INFORMATION	TYPE	TAILLE
DBN3	Descripteur de la Base Nominale N°2	Entier	
TDec	Table de Déclinaison. Chaque case de cette table contient une chaîne de 3 caractères permettant de décliner une forme dérivée dans un type de déclinaison spécifié	TAB	6

FEXCEP1 : Le Fichier des Exceptions de Type 1 (cf § V.3.1).

CODE	INFORMATION	TYPE	TAILLE
Excep	L'élément Exception en question	C	10

FEXCEP2: Le Fichier des Exceptions de Type 2 (cf § V.3.1).

CODE	INFORMATION	TYPE	TAILLE
NBasN	Le Numéro de la Base Nominale	Entier	
Excep	L'élément Exception en question	C	10

FSuffixe : Le Fichier des Suffixes relatifs à la Conjugaison.

CODE	INFORMATION	TYPE	TAILLE
Nvs	Numéro du Vecteur Suffixe	Entier	
Png	Personne + Nombre + Genre	C	3
Suf	Le Suffixe correspondant	C	10

**FPréfixe : Le Fichier des Vecteurs Préfixes correspondant aux différentes Classes.**

CODE	INFORMATION	TYPE	TAILLE
NClas	Le Numéro de la Classe	Entier	
Nvp	Le Numéro du Vecteur Préfixe (*)	Entier	

(\*) Le numéro du Vecteur Préfixe varie entre 1 et 2 car il existe seulement deux vecteurs préfixes qui sont les suivants :

LES PERSONNES													
DUEL			PLURIEL					SINGULIER					
3DF	3DM	2D	3PF	3PM	2PF	2PM	1P	3SF	3SM	2SF	2SM	1S	NVP
ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	1
ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	ف	2

**FClas : Le fichier contenant pour chaque classe un verbe modèle qui est en fait le premier verbe saisi de la classe.**

CODE	INFORMATION	TYPE	TAILLE
NClas	Le Numéro de la Classe	Entier	
Verbe	Le premier Verbe saisi de la Classe	C	20

## VI.4 Principaux Algorithmes.

### VI.4.1 Principe de l'Algorithme de Synthèse de Bases Verbales.

Si le verbe existe

- déterminer sa classe;
- passer en revue le(s) vecteur(s) suffixe(s) de la classe jusqu'à trouver le vecteur suffixe correspondant à la bonne base verbale qui est elle aussi déterminée au préalable;
- ensuite, on parcourt le vecteur suffixe jusqu'à déterminer le suffixe correspondant à la personne, au nombre et au genre données;
- on concatène ensuite base verbale et suffixe pour obtenir la forme verbale (le verbe conjugué). Si un préfixe est nécessaire alors il faudrait l'ajouter au début de la forme verbale ainsi obtenue;
- enfin, on rajoute à cette forme verbale toutes les combinaisons possibles entre clitiques (enclitiques et proclitiques) et affixes (préfixes et suffixes).

### VI.4.2 Algorithme de Synthèse de Bases Nominales.

#### Principe

Etant donnée une base nominale, on va essayer d'avoir tous les mots possibles à partir de cette base en utilisant les différents descripteurs.

En premier lieu, on regarde si la base nominale a une forme plurielle auquel cas calculer le (les) forme(s) plurielles. Idem pour la forme féminin mais il y a une seule forme à calculer.

Ensuite, il faut calculer les mots à partir de la base nominale et les formants extensions. On les obtient en parcourant la liste de bits correspondant aux formants extensions (ListFe) du fichier Fe. S'il y a '1' on ajoute le suffixe correspondant à la base nominale pour obtenir un mot. Le onzième bit correspond à un préfixe. S'il contient '1' alors il faut le rajouter à chaque mot obtenu par les formants extensions.

### Algorithmes et Modules

Algorithme SyntBaseNom (BaseNom, NBasNom, UL) ---> (ListeMots)

#### Début

Si CPlur # 0 alors DetermPluriel (CPLUR, ListePluriel)

Si CFem # 0 alors DetermFeminin (CFem, Feminin)

ParcoursFe (NbasN, BaseNom, UL, ListeBaseFe)

ListeMots <--- ListePluriel + ListeFeminin + ListeBaseFe

#### Fin

Module DetermFeminin (CFem) ---> (Feminin)

#### Début

Liref<sup>(+)</sup> (FFeminin, EnregFFeminin, CFem)

Feminin <--- EnregFFeminin.feminin

#### Fin

(+) Liref signifie une lecture de fichier.

Module DetermPluriel (CPLUR) ---> (ListePluriel)

Début

ListePluriel <--- Vide

Liref (FPluriel, enregFlur, CPLUR)

Tantque (non(fdf(FPLur)) et (CPLUR = enregFPLUR.CPLUR)

faire

ListePluriel <--- ListePluriel + EnregPluriel.Pluriel

Liref (FPluriel, EnregFlur)

Fintantque

Fin

Module ParcoursFe (NBasN, BaseNom, UL) ---> (ListeBaseFe)

Début

ListeBaseFe <--- Vide

Liref (FFe, EenregFFe, NBasN)

Si UL = 0

alors Parcours1 (EnregFFe.ListeFe, ListeBaseFe)

sinon Parcours2 (EnregFFe.ListeFe, ListeBaseFe)

Fin

Module Parcours1 (ListeFe) ----> (ListeBaseFe)

Début

j <--- 0

Pour i = 1 jusqu'à 10

Faire

Si listeFe[i] # 0

alors

Début

ListeBaseFe <--- ListeBaseFe + BaseNom + TabFe[i]

j <--- j + 1

Fin

FinPour

Si ListeFe[11] # 0

alors

Pour i = 1 jusqu'à j Faire

ListeBaseFe <--- ListeBaseFe + liste[11] + ListeBaseFe[i]

Fin

## VI.5 Utilitaires d'Ajout des Bases.

### VI.5.1 Ajout d'une Base Verbale.

Pour saisir une base verbale, il faut avant tout saisir le verbe correspondant. Saisir un nouveau verbe étant donné le numéro de sa classe se passe de la façon suivante :

Si le numéro de la classe du nouveau verbe existe déjà, c'est à dire il existe un verbe dans la base de même classe, ainsi le nouveau verbe hérite des caractéristiques du verbe de même classe déjà existant.

Si le numéro de la classe n'existe pas alors il faut saisir pour chaque aspect-mode la(les) bases(s) du nouveau verbe correspondante(s).

Un verbe peut avoir dans un aspect-mode au minimum 1 base et au maximum 4 bases (Voir Thèse N.Abu Al-Chay [ABU88]).

Dans chaque aspect-mode une fois une base saisie, il y a un choix de vecteur suffixe (liste des suffixes associés à une base du verbe dans un aspect-mode) à faire à partir des vecteurs possibles présentés pour cet aspect-mode.

Une fois les bases, des différents aspects-mode, saisies il faut choisir le vecteur préfixe correspondant à ce verbe parmi les deux vecteurs possibles.

Le vecteur préfixe correspond à la liste des préfixes distribués par rapport au vecteur des pronoms (1SN, 2SM, 2SF, 3SM, 3SF, 1PN, 2PM, 3PM, 2DN, 3DM, 3DF) à l'aspect-mode inaccompli réel. Il est présenté comme suit :

LES PERSONNES													
DUEL			PLURIEL					SINGULIER					
3DF	3DM	2D	3PF	3PM	2PF	2PM	1P	3SF	3SM	2SF	2SM	1S	NUP
ت	ي	ت	ي	ي	ت	ت	ت	ت	ي	ت	ت	أ	1
ت	ي	ت	ي	ي	ت	ت	ت	ت	ي	ت	ت	أ	2

### VI.5.2 Ajout d'une Base Nominale.

Pour saisir une base nominale, on commence à saisir la catégorie de la base nominale c'est-à-dire qu'il faut préciser :

- si la base nominale est au singulier (masculin, féminin) ou au pluriel (masculin, féminin) sinon au collectif;
- si la base nominale est un nom commun ou un nom propre.

Ensuite il faut saisir la valeur de l'unité lexicale de la base nominale qui est soit <Fn> soit <Fn,Fel>.

Si la base nominale à un (des) pluriel(s) alors il faut le(s) saisir.

Si elle admet aussi une forme du féminin, il faut également la saisir.

Il faut préciser aussi les formants extensions possibles pour la base nominale.

Enfin, on précise la table de déclinaison de la base.

### VI.6 Conclusion.

Pour réaliser notre prototype de dictionnaire, on a utilisé comme corpus un échantillon d'une trentaine de mots tiré d'un article de la presse écrite quotidienne tunisienne en langue arabe et plus particulièrement du journal " الصباح " ("AL SABAH") du 7 août 1992 (l'article dans son intégralité est porté en annexe 9).

## VII. CONCLUSION

Notre objectif tout au long de ce travail a consisté avant tout à concevoir un dictionnaire structuré et facilement modifiable. L'utilisation directe qui est faite de ce dictionnaire est la réalisation d'un correcteur orthographique de l'arabe au niveau du mot ; elle est en cours de réalisation par N. GADER dans le cadre de son mémoire de DEA en Sciences de l'Information et de la Communication pour l'année 1992.

Comme le lecteur a pu le constater tout au long de ce travail, la conception de ce dictionnaire est basée sur un modèle linguistique définissant l'ensemble des traits morphologiques des mots arabes.

L'organisation des données dans le dictionnaire et la structure même de ce dernier en tant qu'ensemble de relations interconnectées, sont totalement indépendantes des traitements envisagés.

Plusieurs perspectives d'utilisation de ce dictionnaire sont possibles. On peut en citer :

- la correction orthographique par phonémisation;
- le problème de la vocalisation de textes arabes qui sont généralement non voyellés.

Enfin, une utilisation plus importante et plus complète de ce dictionnaire permettant d'autres types d'exploitations (syntaxiques, sémantiques, etc...), nécessitera, sans aucun doute, d'importants développements. Les plus importants sont essentiellement d'ordre linguistique.

En effet, le schéma général du dictionnaire montre que les descripteurs linguistiques de la base se trouvent au centre de cette structure. Ces descripteurs sont, pour l'instant, relatifs aux bases nominales et verbales, et sont limités aux aspects morphologiques. Il est nécessaire de les compléter afin de répondre aux problèmes de la syntaxe (analyse et synthèse syntaxique).

La définition de l'ensemble des catégories morphologiques de la langue et des sous-catégorisations possibles est un travail linguistique à poursuivre. Cette définition doit tenir compte des applications possibles du modèle linguistique et du dictionnaire (en traduction assistée par ordinateur et en documentation automatique).

**A N N E X E 1**

**Liste des Proclitiques, Préfixes,  
Suffixes et Enclitiques.**



**A N N E X E 2**

**Table de Compatibilité Proclitiques-Préfixes.**

## Table de Compatibilité Proclitiques-Préfixes

ت	ي	ن	أ	Uid	Préf.	Procl.
					Procl.	Préf.
+	+	+	+	+	أو	
+	+	+	+	+	أف	
				+	أب	
				+	أك	
+	+	+	+		أس	
				+	وب	
				+	وك	
+	+	+	+	+	ول	
+	+	+	+		وس	
				+	فب	

ت	ي	ن	أ	Uid	Préf.	Procl.
					Procl.	Préf.
+	+	+	+	+	Vide	
+	+	+	+	+	أ	
+	+	+	+	+	و	
+	+	+	+	+	فا	
				+	ب	
				+	ك	
+	+	+	+	+	ل	
+	+	+	+		ع	
				+	ال	
+	+	+	+	+	أل	

ت	ي	ن	أ	Uid	Préf.	Procl.
					Procl.	Préf.
				+	فلل	
				+	وال	
				+	فال	
				+	بال	
				+	كال	
				+	بكال	
				+	وكال	
				+	فبال	

ت	ي	ن	أ	Uid	Préf.	Procl.
					Procl.	Préf.
				+	فك	
+	+	+	+	+	فل	
+	+	+	+		فس	
				+	لل	
				+	أوب	
+	+	+	+		أفس	
				+	أفب	
+	+	+	+	+	أول	
+	+	+	+	+	أفل	
				+	ولل	

**A N N E X E 3**

**Table de Compatibilité Suffixes-Enclitiques.**

**Table de Compatibilité Suffixes-Enclitiques**

نا	نوی	ك	ی	كن	كم	هن	هم	ه	هما	ها	Uid	Encl / Suf
+	+	+	+	+	+	+	+	+	+	+	+	Uide
+	+	+	+	+	+	+	+	+	+	+	+	ت
+	+	+	+	+	+	+	+	+	+	+	+	ون
+	+	+	+	+	+	+	+	+	+	+	+	تا
+	+	+	+	+	+	+	+	+	+	+	+	و
+		+	+	+	+		+	+	+	+	+	ات
+		+	+	+	+		+	+	+	+	+	ن
+	+		+			+	+	+	+	+	+	ین
+	+	+		+	+		+	+	+	+	+	ا
		+		+	+		+	+	+	+	+	نا

نا	نوی	ك	ی	كن	كم	هن	هم	ه	هما	ها	Uid	Encl / Suf
+	+					+	+	+	+	+	+	تمو
+	+					+	+	+	+	+	+	ی
	+					+	+	+	+	+	+	نما
						+	+	+	+	+	+	تن
	+			+	+	+	+	+	+	+	+	ان
						+	+	+	+	+	+	تین
											+	ة
											+	یہ
											+	تان
											+	وا
											+	تم

**A N N E X E 4**

**Liste des Prébases.**

## Liste A : Liste des Prébases

(1) [ ]

(2) [ أو، فـ ]

(3) [ و، فـ ]

(4) [ أ، أو، أفـ، ألـ، أسـ، أفـ، أفسـ، ولـ، وسـ، فلـ، فسـ، أ، لـ، سـ ]

(5) [ أ، أو، أفـ، ألـ، أسـ، أفـ، أفسـ، و، ولـ، وسـ، فـ، فلـ، فسـ، نـ، لنـ، سنـ ]

(6) [ أ، أو، أفـ، ألـ، أسـ، أفـ، أفسـ، و، ولـ، وسـ، فـ، فلـ، فسـ، يـ، ليـ، سيـ ]

(7) [ أ، أو، أفـ، ألـ، أسـ، أفـ، أفسـ، و، ولـ، وسـ، فـ، فلـ، فسـ، تـ، لتـ، ستـ ]

(8) [ لـ، ولـ، فلـ ]

(9) [ لـ، ولـ، فلـ ]

(10) [ الـ، والـ، وكـ، فـ، فـ، كـ ]

(11) [ بـ، فـ ]

(12) [ كـ، و، فـ ]

(13) [ بـ، و، فـ ]

(14) [ و، بـ، كـ، أ، بـ، كـ، أو، أولـ، أولـ، أولـ، أولـ ]

**A N N E X E 5**

**Liste des Postbases.**

**Liste B : Liste des Postbases**

- (1) [ ]
- (2) [ ك ، كن ، كم ، ه ، هم ، ها ، هن ، هما ]
- (3) [ ته ، نكن ، نكم ، نحم ، نحا ، ن ]
- (4) [ اني ، أنا ، ني ، نني ، ونى ، ونه ، ونني ، وننا ، ونحم ، ونحا ، ونهن ، ونا ، ننا ]
- (5) [ ون ، ان ، انحم ، انحا ، انهن ، انحما ، وا ، وه ، وهم ، وها ، وهن ، وهما ]
- (6) [ ا ، اه ، اهم ، اها ، احما ]
- (7) [ ي ]
- (8) [ ين ]
- (9) [ نك ، اك ، اكم ]
- (10) [ وك ، وكم ]
- (11) [ ونك ، ونكن ، ونكم ]
- (12) [ ينه ، ينني ، يننا ، ينهم ، ينحا ، ينهن ، ينحما ، ينني ، يننا ، يه ، يهم ، يها ، يهن ، يهما ]
- (13) [ تني ، ة ، يه ، ات ]
- (14) [ تان ]
- (15) [ تا ، تاك ، تاه ، تاني ، تانا ، تاكن ، تاكم ، تاهم ، تاها ، تاهن ، تاهما ]
- (16) [ تنه ، تنهم ، تنحا ، تنهن ، تنحما ، ناه ، ناهم ، ناك ، ناكن ، ناكم ، ناهما ، ناهما ، ت ، تي ، تك ، ته ، تنني ، تننا ]
- (17) [ تكن ، تكم ، تهم ، تهن ، تحما ، تما ، تماه ، تمانني ، تماهم ، تماهن ، اتحا ، تم ، اته ، اتنا ، اتكم ، اتهم ، تماهما ]
- (18) [ تحا ، ونحما ، ونه ، ونني ، ونني ، وننا ، تاي ، و ، وكن ، وي ، اتحما ، اتكن ، اتني ، اتك ، اتني ، اكن ، تمو ، تموها ]
- (19) [ تموه ، تموهما ، تموهم ، تموهن ، تمونني ، تمونا ، انه ، انكم ، اكن ، انني ، تينحا ، تينحما ، تينه ، تينهم ، تينهن ]

**A N N E X E 6**

**Table de Compatibilité Prébases-Postbases.**

**Table de Compatibilité Prébases-Postbases**

12	11	10	9	8	7	6	5	4	3	2	1	A / B
+	+	+	+	+	+	+	+	+	+	+	+	1
+	+				+	+	+	+	+	+	+	2
						+	+	+		+	+	3
						+	+			+	+	4
	+		+			+	+			+	+	5
+	+					+	+			+	+	6
+	+	+	+		+	+				+	+	7
+	+	+	+	+	+	+					+	8
							+			+	+	9
+	+						+				+	10
							+				+	11
						+				+	+	12
+	+	+	+	+	+						+	13
	+		+								+	14
+	+					+					+	15
											+	16

**A N N E X E 7**

**Liste des Vecteurs Suffixes.**

### 7.1 Liste des vecteurs suffixes de l'accompli

Suffixe	NUS								
تَنْ	107	نَ	106	نَ	103	ا	102	تْ	101
نَ	107	ا	106	تَمَّا	103	تَا	102	تْ	101
تَمَّا	107	تَمَّ	106	ا	103	تْ	103	تْ	101
ا	107	تَمَّا	106	تَا	103	تْ	103	نَا	101
تَا	107	تْ	107		104	تْ	103	تَمَّ	101
وَا	108	تْ	107	ا	105		103	تَنْ	101
تَا	109	تْ	107	تْ	106	تْ	103	نَ	101
وَا	109		107	تْ	106	نَا	103	تَمَّا	101
تْ	109	تْ	107	تْ	106	تَمَّ	103		102
		نَا	107	نَا	106	تَنْ	103	تْ	102
		تَمَّ	107	تَنْ	106	وَا	103	وَا	102

### 7.2 Liste des vecteurs suffixes de l'inaccompli réel

Suf.	NUS	Suf.	NUS	Suf.	NUS	Suf.	NUS	Suf.	NUS	Suf.	NUS
	209	اَنْ	208	نَ	207		204	نَ	202	وْ	201
نَ	209	وْ	208	نَ	207	اَنْ	204	رِينَ	203	رِينَ	201
نَ	209	وْ	208	اَنْ	207	اَنْ	204	وْنَ	203	وْ	201
اَنْ	210	وْ	208	اَنْ	207		205	وْنَ	203	وْ	201
اَنْ	210	نَ	208	اَنْ	207		205		204	وْ	201
اَنْ	210	نَ	208	وْ	208		205		204	وْنَ	201
		اَنْ	208	رِينَ	208		205		204	وْنَ	201
			209	وْ	208		205	نَ	204	اَنْ	201
			209	وْنَ	208	رِينَ	206	نَ	204	اَنْ	201
			209	وْنَ	208	وْنَ	206	اَنْ	204	اَنْ	201
			209	اَنْ	208	وْنَ	206		204	نَ	202

### 7.3 Liste des vecteurs suffixes de l'inaccompli potentiel subjonctif

Suf.	NUS										
ـ	309	نَ	307	ـ	307	ـِ	304	نَ	302	ـ	301
اَ	309	اَ	307	ـِ	307	وَا	304	ـ	303	ـِ	301
اَ	309	نَ	308	ـ	307	وَا	304	ـ	303	ـ	301
اَ	309	نَ	308	وَا	307	ـِ	305	ـ	303	ـ	301
		اَ	308	وَا	307	وا	305	نَ	303	ـ	301
		اَ	308	اَ	307	وا	305	نَ	303	وَا	301
		اَ	308	اَ	307		306	اَ	303	وَا	301
		ـ	309	ـ	307		306	ـ	303	اَ	301
		ـ	309	ـ	307		306	ـ	303	اَ	301
		ـ	309	ـ	307		306	اَ	303	اَ	301
		ـ	309	نَ	307		306	اَ	303	نَ	302

### 7.4 Liste des vecteurs suffixes de l'inaccompli potentiel apocopé

Suf.	NUS												
	411	نَ	409	نَ	407	ـَ	407		405	نَ	402	ـ	401
	411	ـِ	410	اَ	407	ـِ	407		405	ـِ	403	ـِ	401
نَ	411	وَا	410	اَ	408	ـَ	407		405	وَا	403	ـ	401
نَ	411	وَا	410	اَ	408	وَا	407		406	وَا	403	ـ	401
وَا	412	اَ	410	اَ	408	وَا	407		406	نَ	404	ـ	401
وَا	412	اَ	410		409	اَ	407	ـِ	406	نَ	404	وَا	401
اَ	412	اَ	410		409	اَ	407		406	اَ	404	وَا	401
اَ	412		411		409	ـَ	407		406	اَ	404	اَ	401
اَ	412		411		409	ـَ	407		406	اَ	404	اَ	401
		ـِ	411		409	ـَ	407	وا	406		405	اَ	401
			411	نَ	409	نَ	407	وا	406		405	نَ	402

### 7.5 Liste des vecteurs suffixes de l'impératif

Suffixe	NUS	Suffixe	NUS	Suffixe	NUS
نَ	509	رِي	505	رِي	501
اَ	510	وَا	505	وَا	501
		نَ	506	اَ	501
		اَ	506		502
			507	نَ	502
		ي	507	°	503
		وَا	507	وَا	503
		اَ	507	اَ	503
		اَ	508	رِي	503
		رِي	508	نَ	503
		وَا	508		504

### 7.6 Liste des vecteurs suffixes de l'accompli passif

Suffixe	NUS	Suffixe	NUS	Suffixe	NUS	Suffixe	NUS
نَا	604	تَمَمَا	603	نَا	602	تَ	601
تَمَمَ	604	نَا	603	نَ	603	تَ	601
نَ	604	اَ	603	تَ	603	نَا	601
اَ	604	تَ	604	نَا	603	تَمَمَ	601
وَا	605	تَ	604	تَمَمَ	603	تَمَمَ	601
		نَا	604	وَا	603	نَ	601
		تَمَمَ	604	نَا	603	تَمَمَا	601
		تَمَمَا	604	اَ	603	اَ	602
		نَا	604	نَا	603	نَا	602
		تَ	604	تَمَمَ	603	وَا	602
		اَ	604	نَ	603	اَ	602

**7.7 Liste des vecteurs suffixes de l'inaccompli réel passif**

Suffixe	NUS	Suffixe	NUS	Suffixe	NUS	Suffixe	NUS
'	706		705	نَ	702	'	701
'	706		705	نَ	703	رِينَ	701
نَ	706		705	نَ	703	'	701
نَ	706	'	706	اَنَ	703	'	701
اَنَ	706	رِينَ	706	اَنَ	703	'	701
اَنَ	709	'	706	اَنَ	703	وَنَ	701
اَنَ	709	وَنَ	706	وَنَ	704	وَنَ	701
اَنَ	709	وَنَ	706	وَنَ	704	اَنَ	701
		اَنَ	706	يَنَ	704	اَنَ	701
		اَنَ	706		705	اَنَ	701
		'	706		705	نَ	702

**A N N E X E 8**

**Liste des Vecteurs Préfixes.**

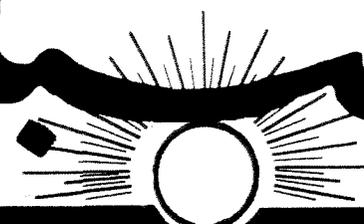
## Liste des Vecteurs Préfixes

Vect Pref 2	Vect Pref 1	Pronoms
أُ	أَ	أَنَا
تُ	تَ	نَحْنُ
تُ	تَ	أَنْتَ
يُ	يَ	أَنْتِ
تُ	تَ	أَنْتُمَا
نُ	نَ	أَنْتُمْ
تُ	تَ	أَنْتُنَّ
تُ	تَ	هِيَ
يُ	يَ	هِيَ
يُ	يَ	هُمَا
تُ	تَ	هُمَا
يُ	يَ	هُمْ
نُ	نَ	هُنَّ

A N N E X E 9

L'article du Journal "AL SABAH".

# الصباح



الصحيفة اليومية الأولى بتونس

## الانتخابات الأمريكية

في ظلّ احتمالات الانتخابات الأمريكية ، أصابت الحيرة محيط الرئيس الأمريكي « بوش » ، و أخذ يثمرّ ف من موقع الخاسر في الانتخابات الرئاسية المقبلة. فقد دلّ آخر سبر آراء على أنّ شعبية الرئيس الأمريكي بلغت أدنى مستوى وصلته شعبية رئيس في الولايات المتحدة الأمريكية.

**A N N E X E 10**

**Lexique de Définitions.**

## LEXIQUE DE DEFINITION [SAM90]

### **Anomale (Racine -)**

Ce terme, dû à A.Roman, désigne les racines dont les bases sont susceptibles d'être affectées de transformations. Il s'agit des racines dans lesquelles :

- on trouve une radicale /ʃ/ ou /ʒ/;
- la deuxième et la troisième radicale correspondent à une même consonne (R2 = R3);
- l'une des radicales est /i/;

### **Ductus du Mot**

Dans la graphie, le corps même du mot, dépourvu des signes diacritiques dits de vocalisation.

### **Formants du Mot**

Morphèmes constitutifs de cette unité. Nous posons la convention selon laquelle les opérations de contextualisation, qu'elles relèvent du signifiant, du signifié ou du syntactique d'un morphème inclus dans un unité observable donnée, ici, le mot, sont limitées aux relations associées aux morphèmes susceptibles d'être compris dans cette unité. Ces morphèmes seront dits alors formant de cette dernière, c'est à dire, en ce qui concerne le programme SAMIA, formants du mot.

### **Grammaire des Formants du Mot**

La convention de définition des formants du mot donne un fondement formel à ce que le rapport Desclés [DES83] appelait la grammaire des formants du mot. Ces derniers sont représentés par projection sur le vecteur de représentation du mot. Des relations de collocation sont en outre associées aux morphèmes, afin de :

- (1°) valider ou invalider leur compatibilité avec d'autres formants, et de
- (2°) permettre les transformations affectants un nombre important de formants, en fonction de leur contexte. Cette grammaire doit être construite de manière à être compatible avec les deux fonctionnements, de la synthèse et de l'analyse.

## Graphèmes

Par convention, sont appelés graphèmes, non sans quelques abus de langage, les unités du mot graphique non-vocalisé, c'est-à-dire les seuls graphèmes notés dans le ductus du mot, que l'on peut aussi appelé lettres ou - faute d'un meilleur terme - segments ( حروف ).

## Morphème

En anglais, morph, signe segmental élémentaire (I. Mel'cuk, 1982, p.63 [MEL82]). Cet auteur introduit dans la définition formelle du signe linguistique, classiquement représenté en termes de <Signifiant, Signifié>, un troisième élément, le syntactique.

Le signe est représenté par le triplé ordonné : <Signifiant, Signifié, Syntactique> (p.40). Le syntactique est défini, p.26, comme l'ensemble des spécifications d'ordre contextuel associées à un couple <Signifiant, Signifié>, non déductibles de l'un ou de l'autre : combinaisons avec d'autres couples du même type, modifications du signifiant ou du signifié en contexte.

La définition du morphème de Mel'cuk est à la base de la convention adoptée pour celle du formant du mot.

## Mot (en Arabe)

(1°) Définition empirique : unité graphique (quand on parle des systèmes d'écriture, mot a le sens de "portion de texte délimitée par deux blancs ou par des marqueurs comparables", Gleason, 1961-69 [GLE69]) et d'unité accentuelle (Brame, 1970 [BRA70], Dichy, 1978 [DIC78] - Ces deux unités ne coïncident pas entièrement). Le mot en arabe constitue donc une unité syntagmatique empiriquement observable.

(2°) Définition axiomatisée : A partir de ces observations, on décrit un vecteur de représentation du mot. Les éléments de cette représentation vectorielle, associés à un rang de vecteur et à des règles qui en gèrent les autres relations contextuelles au sein de la même unité définissent axiomatiquement le mot arabe.

## Mot Graphique Non-Vocalisé (Mg)

Le Mg est le mot, dans sa réalisation en graphie non-vocalisée, qui constitue l'entrée du système de traitement en analyse. Quelquefois abrégé en mot graphique, par référence à l'écriture courante.

## Paradigme

Les différents paradigmes de conjugaison possibles sont répartis suivant les aspects et les modes. Pour le verbe à la voix "active", nous avons cinq paradigmes :

1. Accompli ( ما فعل ) ;
2. Inaccompli réel ( مفاعله مرفوع ) ;
3. Inaccompli potentiel - "subjunctif" ( مفاعله منصوب ) ;
4. Inaccompli potentiel - "apocopé" ( مفاعله مجزوم ) ;
5. Impératif ( أمر ) ;

et quatre paradigmes pour le verbe à agent non exprimé, couramment, bien qu'improprement appelé "passif"(\*) ( مبنو للمجهول ) :

6. Accompli ;
7. Inaccompli réel ;
8. Inaccompli potentiel - "subjunctif" ;
9. Inaccompli potentiel - "apocopé".

## Radical

Le radical comprend, à l'intérieur de la base, les phonèmes (ou les graphèmes) compris entre la première et la dernière consonne radicale apparente.

Dans < سبعة > correspondant à < سمات > (/Simät/, "traits"), < س > est la première radicale apparente, mais la racine est en fait / وسم / , et / و / la première radicale phonologique).

## Radicale

Consonne appartenant à la racine.

## Vecteur de Représentation du Mot

Voir, ci-dessus, la définition axiomatisée du mot. En voici le schéma, commenté dans [DIC90], paragr. X.3.6.2 :

(\*) Par convention, nous n'indiquons que la voix "passive". En l'absence d'une spécification de la voix, le verbe est donc à la voix "active".

PCL												PRF		<----- BAS ----->					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

----- BAS -----																			
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
R1						R2						R3				R4			

--- BAS --->				SUF								ECL-1						ECL-2	
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60

ECL-2			
61	62	63	64

**Vecteur-Mot**

Formulation abrégée pour vecteur de représentation du mot.

## REFERENCES BIBLIOGRAPHIQUES

- [ABA83] ABAAB H., J.-P. DESCLES, J. DICHY, D.E. KOULOUGHLI, M. ZIADAH , 1983, Conception d'un synthétiseur et d'un analyseur morphologique de l'arabe, en vue d'une utilisation en Enseignement assisté par Ordinateur. Rapport rédigé sous la direction J.-P. DESCLES, à la demande du Ministère Français des Relations Extérieures, souvent cité comme le rapport Desclés.
- [ABA84] ABAAB H., 1984, Contribution au traitement automatique de la langue arabe - Conception d'un synthétiseur morphologique utilisable en E.A.O., Thèse pour le doctorat du 3ème cycle.
- [ABU84] ABU AL-CHAY N., 1984, La constitution automatique du vecteur-mot d'un mot graphique non-vocalisé en arabe, mémoire de D.E.A., Univ. Lyon-1 et Lyon-2.
- [ABU86] ABU AL-CHAY N., 1986, Modélisation des verbes arabes - Automatisation de la conjugaison, document interne, Programme SAMIA.
- [ABU88] ABU AL-CHAY N., 1988, Un système expert pour l'analyse et la synthèse des verbes arabes, Thèse de doctorat Univ. Lyon-1.

- [AKA90] AKAICHI J, 1990, Conception d'un Analyseur et d'un Synthétiseur des Verbes Arabes dans une perspective d'Enseignement Assisté par Ordinateur, Memoire de Maitrise (MIAGE), Institut Supérieur de Gestion (I.S.G.) de Tunis (Tunisie)/ C.R.T.T - Univ. Lyon-2, .
- [BRA70] BRAME, M, 1970, Arabic Phonology, Implication for Phonological Theory and Historical Semitic, Ph.D, Diss., Massachisetts, Institue of Technology.
- [BOI86] BOITET C., 1986, Toward integrated dictionaries for M(a)T: motivations and linguistic organisation, Proceedings of COLING'86, Bonn.
- [BOU86] BOUZIDI L., 1986, Conception d'un système d'information pour l'E.A.O de la morphologie de l'arabe, document interne, Programme SAMIA.
- [COU87] COURTOIS B., 1986-1987, Dictionnaire électronique du français. Alphabétisation directe, alphabétisation inverse. Rapports de recherche du Laboratoire d'Automatique Documentaire et Linguistique (LADL). Univ. Paris VII, n°17, 19, 20.
- [COU88] COURTOIS B., 1988, Dictionnaire DELAR des racines du système DELA. Rapport de recherche du LADL. Univ. Paris VII, n°23.

- [DAN80] DANLOS L., 1985, Génération automatique de textes en langues naturelles. Masson, Paris.
- [DES83] DESCLES J.-P. , 1983, Rapport Desclés, voir [ABA83] ci-dessus.
- [DES89] DESCLES J.-P., 1989, «Informatique et sciences humaines», L'informatique en perspective, K. Fall and G. Vignaux, eds., Presses de l'Université du Québec, p.141-160.
- [DIC78] DICHY, J, 1978, Le Rythme dans l'Apprentissage Oral de l'Arabe Littéraire Moderne, in Revue de Phonétique appliquée, n°45, Université de Mons (Belgique), pp. 21-42.
- [DIC86] DICHY J., 1986, Traitement automatique de la conjugaison arabe : conception linguistique, document interne, Programme SAMIA.
- [DIC87] DICHY J., 1987, «The SAMIA Research Program, Year Four: Progress and Prospects», in Processing Arabic Report, n°2, Katholieke Universiteit Nijmegen, p.1-26.
- [DIC88] DICHY J., 1988, «Compétences de base et compétences complexes - Vers un modèle de transfert intra-linguistique: l'exemple de l'arabe», L'enseignement des langues en Europe de l'Ouest - Actes de la rencontre de Nantes (25-30 août 1988), Cahiers de l'E.R.E.L. n°3

spécial, 1990, A.P.L.V. - Université de Nantes, p.78-87.

[DIC89] DICHY J., 1989, «Propositions pour un dictionnaire informatisé et un système de correction orthographique de l'arabe», Conférence sur l'informatique et la langue arabe, Institut du Monde arabe, Paris, 3-4 octobre 1989.

[DIC90] DICHY J., 1990, L'écriture dans la représentation de la langue : la lettre et le mot en arabe, Thèse pour le doctorat d'Etat, Univ. Lyon-2.

[DUR87] DURAND J. C. et LEPAGE Y., 1987, Base de données et dictionnaire intégrés pour la traduction automatique, Congrès IDT-87, Strasbourg, p. 177-181.

[GAD90] GADER N., 1990, Un Système d'Enseignement Assisté par Ordinateur des Formes Nominales Dérivées en Arabe avec Simulation, Memoire de Maitrise (MIAGE), Institut Supérieur de Gestion (I.S.G.) de Tunis (Tunisie)/ C.R.T.T - Univ. Lyon-2,.

[GHE90] GHENIMA M., 1990, Conception d'un Sytème d'Enseignement Assisté par Ordinateur pour la Conjugaison des Verbes Arabes avec Simulation de Modèles Linguistiques, Memoire de Maitrise (MIAGE), Institut Supérieur de Gestion (I.S.G.) de Tunis (Tunisie)/ C.R.T.T - Univ. Lyon-2,.

- [GLE69] GLEASON, H. A., An Introduction to Descriptive Linguistics, Holt, Rinehart & Winston, New-York, 1961; trad. Franç. : Introduction à la linguistique, Larousse, Paris, 1969.
- [GRI90] GRIZE J.-B., 1990, Logique et langage, Paris, Orphys.
- [GRO86] GROSS M., NIVAT M., PERRIN D., 1986, Présentation du Centre d'Etudes et de Recherches en Informatique Linguistique, CERIL.
- [GRO89] GROSS M., 1989, La construction de dictionnaires électroniques, Annales des Télécommunications, Tome 44, n° 1-2 Janvier-Février.
- [HAS85] HASSOUN M. O., DICHY J., 1985, La base de mot en arabe et ses descripteurs, document interne, Programme SAMIA.
- [HAS86] HASSOUN M.O., 1986, «Dira:sa wa-tahqi:q li-bank macluma:t min ?azli l-muca:laza l-?a:lijja li-l-luga l-carabijja», 9th National Computer Conference, Riadh, vol. 2, p.61-80.
- [HAS87] HASSOUN M.O., 1987, Conception d'un dictionnaire pour le traitement automatique de l'arabe dans différents contextes d'application, Thèse pour le doctorat d'Etat, Univ. Lyon-1.

- [IGH87] IGHILAZA C., 1987, Un didacticiel avec simulation d'un modèle linguistique: "SAMIA VERBE Na:Qis", mémoire de D.E.A., Univ. Lyon-1 et Lyon-2.
- [LEF91] LEFEVRE P., CAILLAUD N., 1991, Logiciel d'accès par voisinage à un dictionnaire informatique du français courant, Electricité de France : Bulletin de la direction des études et recherches - Serie C-Mathématiques- Informatique, N°1.
- [LEL85] LELUBRE X., 1985, Projet pour un didacticiel de conjugaison de verbes arabes, Ministère de l'Education nationale.
- [LEL87] LELUBRE X., 1987, Scénario de DIALOGAISON, logiciel d'arabe présenté au Concours national de scénarios de logiciels à usage éducatif, organisé par la Ministère de l'Education Nationale, Paris.
- [MEL82] MEL'CUK, I. A., 1982, Towards a Language of Linguistics, a System of Formal Notions for Theoretical Morphology, Wilhem Fink Verlag, Munchen.
- [MOU85] MOURAD F., 1985, Réalisation d'un synthétiseur morphologique de l'arabe, mémoire de D.E.A, Univ. Lyon-1 et Lyon-2.

- [NOU90] NOUIRA R, 1990, Conception d'un Analyseur et d'un Synthétiseur des Dérivées Verbales de l'Arabe dans une perspective d'Enseignement Assisté par Ordinateur, Memoire de Maitrise (MIAGE), Institut Supérieur de Gestion (I.S.G.) de Tunis (Tunisie)/ C.R.T.T - Univ. Lyon-2, .
- [REI83] REIG D., 1983, La conjugaison arabe, Paris, G.P. Maisonneuve et Larose.
- [ROM84] ROMAN A., 1984, «Les modalités de voix du verbe arabe», Cercle linguistique d'Aix-en-Provence, Travaux 2, Le Passif, Université de Provence, p.145-158.
- [ROM90] ROMAN A., 1990, Grammaire arabe, Paris, Presses Universitaires de France (Coll. «Que sais-je?»).
- [SAM84] SAMIA (signature collective), 1984, «Enseignement assisté par ordinateur de l'arabe - Simulation à l'aide d'un modèle linguistique : la morphologie», Colloque international EAO 84 organisé par l'Agence de l'Informatique, Lyon, les 4-5 septembre 1984, pp.81-96.
- [SAM89] DICHY J. and HASSOUN M.O., eds. 1989, Simulation de modèles linguistiques et Enseignement assisté par ordinateur de l'arabe - Travaux SAMIA I, Paris, Fondation Postuniversitaire Interculturelle (Conseil international à la langue française - CILF).

- [SCH78] SCHEIDLIN R., 1978, 201 Arabic Verbs, New-York, Barron's Educational Series, Inc.
- [TRE86] TREMBLAY D., 1986, Représentation syntaxique et sémantique de termes dans les dictionnaires électroniques. Thèse de doctorat Univ. Paris VII, LADL.
- [VAN79] VAN SLYPE G., 1979, Description de traduction automatique SYSTRAN, in Documentaliste, vol. 16, n°4, Juillet-Août.



\* 9 5 5 4 4 2 8 \*