

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE  
APPLIQUÉES

PAR  
**SAMMOUD MOHAMED SADOK**

**APPLICATIONS LINGUISTIQUES MULTILINGUES : APPORT DES  
GRAMMAIRES CATÉGORIELLES ET DE LA LOGIQUE COMBINATOIRE**

Janvier 2010

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

CE PROJET A ÉTÉ ÉVALUÉ  
PAR UN JURY COMPOSÉ DE :

M. Ismail Biskri, directeur de projet  
Département de mathématiques et informatique à l'Université du  
Québec à Trois-Rivières

M. Mhamed Mesfioui, professeur  
Département de mathématiques et informatique à l'Université du  
Québec à Trois-Rivières

M. François Meunier, professeur  
Département de mathématiques et informatique à l'Université du  
Québec à Trois-Rivières

## **REMERCIEMENTS**

En préambule à ce mémoire, j'aimerais exprimer mes profonds remerciements au Professeur Ismail Biskri, mon directeur de recherche, qui m'a constamment orientée et soutenue depuis mon arrivé au Canada.

Mes remerciements s'adressent également à tous les professeurs qui m'ont assisté tout le long de mes études à l'université du Québec à Trois-Rivières.

Je n'oublierais pas mes parents, toujours présents par leur soutien et leur encouragement, pour les sacrifices et les efforts qu'ils ont fait.

Je tiens aussi à remercier et à exprimer ma gratitude à mon frère Dr Sammoud Hamed et ma femme Sahar qui étaient toujours à mes côtés dans les moments les plus difficiles.

J'adresse également mes remerciements à mes sœurs et à toute ma famille.

Enfin, je tiens à remercier tout mes proches et amis qui m'ont toujours soutenue et encouragée au cours de la réalisation de ce mémoire.

Un grand merci à tous et à toutes.

## TABLE DES MATIERES

REMERCIEMENTS.....	i
LISTE DES TABLEAUX.....	v
LISTE DES FIGURES .....	vi
CHAPITRE 1 .....	1
INTRODUCTION .....	1
CHAPITRE 2 .....	5
GRAMMAIRES CATEGORIELLES .....	5
2.1 Introduction.....	5
2.2 Historique des Grammaires Catégorielles .....	5
2.2.1 Fondements philosophiques des grammaires catégorielles : Husserl..	5
2.2.2 Lesniewski (1922) .....	6
2.2.3 Ajdukiewicz (1935) .....	7
2.2.4 Les Grammaires Bar-Hillel .....	9
2.2.5 Calcul de Lambek (1958).....	11
2.2.6 Grammaire Catégorielle combinatoire de Steedman .....	13
2.2.7 Grammaire Applicative Universelle de Shaumyan .....	15
2.2.8 Grammaire Applicative et Cognitive (1990) .....	17
2.3 Conclusion .....	17
CHAPITRE 3 .....	18
LOGIQUE COMBINATOIRE.....	18
3.1 Introduction.....	18
3.2 Cadre historique .....	18
3.3 Combinateurs logiques .....	19
3.3.1 Combinateur d'effacement K .....	19
3.3.2 Le combinateur de distribution S .....	19
3.3.3 Le combinateur identité I .....	20
3.3.4 Le combinateur de composition B .....	20
3.3.5 Le combinateur de duplication W .....	21
3.3.6 Le combinateur de coordination $\phi$ .....	22
3.3.7 Le combinateur de distribution $\psi$ .....	22
3.3.8 Le combinateur de changement de type $C^*$ .....	22
3.3.9 Le combinateur de permutation C .....	23
3.4 Combinateurs complexes.....	24
3.4.1 Utilisation des puissances .....	24
3.5 Conclusion .....	24
Chapitre 4.....	26
GRAMMAIRE CATEGORIELLE COMBINATOIRE .....	26

APPLICATIVE .....	26
4.2 Principe de la GCCA .....	26
4.3 Les règles de la GCCA .....	28
4.4 Typification d'unités d'une langue .....	31
4.4.1 Quantificateur nominal .....	31
4.4.2 Verbes intransitifs.....	31
4.4.3 Verbes transitifs.....	32
4.4.4 Les adjectifs.....	33
4.4.5 Les modificateurs de verbes.....	34
4.5 Traitement d'une structure concaténée.....	35
4.6 Les métarègles:.....	37
4.6.1 Métarègle 1 :.....	37
4.6.2 Métarègle 2 :.....	38
4.6.3 Métarègle 3 :.....	39
4.6.4 Métarègle 4 :.....	39
4.6.5 Métarègle 5 :.....	39
4.6.6 Métarègle 6 :.....	39
4.6.7 Métarègle 7 :.....	40
4.6.8 Métarègle 8 :.....	40
4.6.9 Métarègle 9 :.....	41
4.6.10 Métarègle 10 :.....	41
4.7 Conclusion : .....	41
Chapitre 5 : .....	43
PRINCIPES DE BASE DE LA LANGUE ARABE .....	43
5.1 Introduction.....	43
5.2 Caractéristiques de la langue arabe .....	43
5.2.1 Quelques généralités.....	43
5.2.2 Les voyelles.....	44
5.2.3 Les consonnes.....	44
5.3 Formes de base de la langue arabe .....	46
5.3.1 Les verbes.....	46
5.3.1.1 Morphologie des verbes en arabe .....	46
5.3.1.2 Conjugaison d'un verbe en arabe .....	47
5.3.2 Les Noms .....	48
5.3.2.1 Les noms primitifs :.....	48
5.3.2.2 Les noms dérivés : .....	49
5.4 Règles d'accord sujet-verbe en arabe : .....	50
5.4.1 Accord Verbe-Sujet pour les phrases SVO : .....	50
5.4.2 Accord Verbe-Sujet pour les phrases VSO : .....	51
5.5 Pronoms dans la langue arabe : .....	51
5.5.1 Les pronoms libres : .....	51
5.5.2 Les pronoms liés.....	53
5.6 Travaux réalisés dans le traitement automatique de la langue arabe .....	54
5.6.1 Traitement automatique de la langue arabe parlée .....	55
5.6.2 Traitement automatique de la langue arabe écrite.....	56

5.6.2.1	La reconnaissance des caractères arabes imprimés .....	56
5.6.2.2	La traduction automatique des textes arabes.....	57
5.6.2.3	L'analyse syntaxique des textes arabes .....	57
5.6.3	La Grammaire générative.....	57
5.6.4	Les Grammaires d'unification.....	58
5.7	Application des GC pour l'analyse de textes en arabe .....	59
5.8	Conclusion .....	62
Chapitre 6 : .....		63
ÉTUDE THEORIQUE : ANALYSE DU CORPUS .....		63
6.1	Introduction.....	63
Chapitre 7 : CONCLUSION .....		77
BIBLIOGRAPHIE.....		79

## LISTE DES TABLEAUX

TABLEAU I : AFFECTATION DE TYPES POUR LA PHRASE DE TYPE SUJET-VERBE INTRANSITIF.....	32
TABLEAU II : AFFECTATION DE TYPES POUR UNE PHRASE COMPORTANT UN VERBE TRANSITIF DIRECT. ....	33
TABLEAU III : AFFECTATION DE TYPES POUR UNE PHRASE CONTENANT UN VERBE INTRANSITIF, UN ADJECTIF ET UN QUANTIFICATEUR NOMINAL. ....	33
TABLEAU IV : AFFECTATION DE TYPES POUR UNE PHRASE COMPORTANT UN ADVERBE ET UN VERBE INTRANSITIF. ....	34
TABLEAU V : AFFECTATION DE TYPES POUR LA PHRASE COMPORTANT UN VERBE TRANSITIF DIRECT. ....	37
TABLEAU VI : AFFECTATION DE TYPES POUR UNE PHRASE COMPORTANT UN ADVERBE.....	38
TABLEAU VII : LES NOMS PRIMITIFS DANS LA LANGUE ARABE.....	49
TABLEAU VIII : LES NOMS DERIVES DANS LA LANGUE ARABE. ....	50
TABLEAU IX : PRONOMS LIBRES DU SINGULIER ET DU DUEL.....	52
TABLEAU X : PRONOMS LIBRES POUR LE PLURIEL. ....	52
TABLEAU XI : PRONOMS LIES DU SINGULIER. ....	53
TABLEAU XIII : PRONOMS LIES DU DUEL.....	54
TABLEAU XIII : PRONOMS LIES POUR LE PLURIEL.....	54



## LISTE DES FIGURES

FIGURE 1 : LE MODELE DE LA GRAMMAIRE CATEGORIELLE UNIVERSELLE DE SHAUMYANN. ....	17
FIGURE 2 : PASSAGE DE LA STRUCTURE CONCATENEE A UNE STRUCTURE APPLICATIVE FONCTIONNELLE. ....	36

# CHAPITRE 1

## INTRODUCTION

L'évolution démesurée des technologies de l'information a fait apparaître de nouveaux besoins de la part de « nouvelles catégories d'utilisateurs ». Ces utilisateurs ont commencé à exprimer leur volonté de trouver des applications informatiques permettant de prendre en considération leurs langues maternelles surtout dans le Web.

En plus de faciliter l'interaction Homme-Machine entre cette « nouvelle catégorie d'utilisateurs » et leurs ordinateurs, ces applications offriront la possibilité de mieux profiter des révolutions technologiques dans le domaine de l'informatique.

Une nouvelle branche de l'informatique a donc vu le jour, elle porte le nom de « informatique linguistique ». L'informatique linguistique a pour tâche de faciliter l'interaction homme-machine pour les utilisateurs de par le monde. Elle s'intègre dans le domaine de l'Intelligence artificielle. Ce « nouveau domaine » a pour but de faciliter le développement des applications informatiques qui offrent la possibilité d'analyser des textes issus de différentes langues.

Si nous interprétons les statistiques [31] se rapportant au nombre d'utilisateurs d'internet, on remarque que l'Anglais est la langue la plus utilisée sur le web suivie des Chinois avec respectivement 495.8 millions et 407.7 millions d'utilisateurs. Autrement dit, 27.7 % des utilisateurs d'internet utilisent la langue anglaise et 22.6 % utilisent le Chinois. Cette place qu'occupe l'Anglais n'est pas due à un choix de la part des internautes. En effet, la langue anglaise a été imposée : presque la majorité des sites internet ainsi que les applications informatiques sont développés aux États-Unis et en Anglais.

Si l'on prend l'exemple de la langue arabe. Il existe plus de 320 millions locuteurs de cette langue repartis sur 22 pays dans le monde. Il s'agit de la quatrième langue la

plus parlée dans le monde après le Mandarin, l'Espagnol, l'Anglais et le Hindi. Les statistiques [31] montrent qu'il existe plus de 60.252.100 utilisateurs dans le monde qui utilise l'Arabe pour surfer sur le net. Ce nombre correspond à 3.3 % du nombre global d'utilisateurs d'Internet.

Ce qui est marquant dans ces statistiques c'est que : entre 2000 et 2009, l'utilisation de la langue arabe sur internet a enregistré la plus grande croissance par rapport à toutes les autres langues : 2,297.7 % de croissance. Cette croissance ne cesse d'augmenter d'où l'intérêt de notre étude.

Cependant, l'analyse de la langue arabe fait défaut dans la littérature. Cela est dû au caractère très spécial de cette langue sémitique et les difficultés rencontrées par les chercheurs. En effet, l'adaptation de modèles déjà existants et ayant montré leurs efficacités dans le traitement des langues latines à la langue arabe pose problème. Il faudra prendre en considération tous les aspects de la langue arabe. L'adaptation du modèle HPSG (head driven structure Grammar) par exemple a montré que cette grammaire n'offre pas un schéma permettant de représenter les phrases nominales arabes.

Dans notre cas, nous adapterons le modèle de Grammaire catégorielle combinatoire applicative (GCCA) à l'analyse de l'Arabe.

Nous montrerons l'efficacité de ce modèle en l'appliquant sur un Corpus de 50 phrases différentes. Le choix de ces phrases est fait de façon à prendre en considération plusieurs caractéristiques de la langue arabe. Nous nous intéresserons à la coordination, nous traiterons le cas d'absence de sujet. D'autre part, nous prouverons l'efficacité des GCCA dans le traitement des versions verbales et nominales de quelques phrases.

Le présent mémoire est dévissé en 6 chapitres. Le premier chapitre représente une introduction dans laquelle nous présenterons la problématique et le déficit qu'on va relever tout en décrivant le contenu de chacun des chapitres..

Dans le chapitre 2, on donnera un état de l'art sur les grammaires catégorielles. On essaiera de montrer son évolution depuis sa version la plus basique. On commencera par la présentation des fondements philosophiques des grammaires catégorielles en donnant un aperçu sur les travaux de Husserl [3]. On passera par la suite à l'analyse des différentes versions ayant vu le jour dans l'ordre chronologique de leurs apparitions. Nous partirons du modèle développé par Lesniewski (1922) en mettant le point sur théorie des catégories sémantiques qu'il a mis en place. On passera par la suite à l'analyse des principes du modèle d'Ajdukiewicz (1935). Nous y introduirons le principe de calcul de fraction et nous donnerons un exemple illustratif montrant l'application de ce modèle dans l'analyse d'une phrase en langue française. Nous nous intéresserons par la suite aux modèles de la grammaire AB avec son aspect bidirectionnel. Pour finir, nous nous pencherons sur le Calcul de Lambek (1958), la grammaire catégorielle combinatoire de Steedman, la Grammaire applicative universelle de Shaumyan, et la Grammaire applicative et cognitive. Ce chapitre a une importance capitale dans notre mémoire. En effet, il nous permettra d'introduire le chapitre qui constitue le cœur de notre travail et qui porte sur la grammaire catégorielle combinatoire applicative.

Le chapitre 3 sera dédié à l'étude des fondements de la logique combinatoire. Nous commencerons par une présentation du cadre historique dans lequel elle est apparue. On passera par la suite à la définition et au développement de chacun de ces combinateurs logiques. Pour finir avec l'introduction de la notion de combinateur complexe et de la puissance.

Au chapitre 4 on s'intéressera à l'étude de la grammaire catégorielle combinatoire applicative. Ce chapitre est la base de notre mémoire, nous y décrirons le principe de ce modèle ainsi que ses règles de réduction. Une partie de ce chapitre sera consacrée à l'explication de la méthode d'analyse d'une structure concaténée et à la typification des unités linguistiques.

Des exemples traitant des phrases de la langue Française seront développés dans le but de comprendre cette méthodologie d'analyse. Nous finirons ce chapitre par la

présentation d'un ensemble de métarègles qui organisent le choix des règles de réduction dans le processus d'analyse d'une phrase.

Le chapitre 5 est un panorama de concepts généraux se reliant à la langue arabe. Nous consacrerons cette section à la présentation de quelques caractéristiques de base de la langue arabe. Nous mettrons le point sur ses formes de base et sur les règles d'accord verbe sujet dans le cas des phrases de type VOS et VSO. Nous présenterons aussi un récapitulatif des travaux ayant été réalisés dans le domaine de traitement automatique de la langue arabe.

Les chapitres 6 et 7 seront consacrés à l'analyse d'un corpus contenant un ensemble de phrases de la langue arabe et à l'interprétation des résultats obtenues.

## **CHAPITRE 2**

### **GRAMMAIRES CATEGORIELLES**

#### **2.1 Introduction**

Ce chapitre sera dédié à la présentation de quelques notions générales se rapportant au modèle des grammaires catégorielles. Il s'agit en d'autres termes d'une base théorique essentielle pour comprendre la partie la plus essentielle de notre desserte, à savoir l'étude de la Grammaire Catégorielle Combinatoire Applicative appliquée à la langue arabe.

On commencera ce chapitre par un aperçu sur les origines historiques des grammaires catégorielles en partant des fondements philosophiques de Husserl.

On passera par la suite à la présentation des différentes versions de grammaires catégorielles dans l'ordre chronologique de leurs apparitions. On analysera aussi les modèles de Lesniewski (1922), Ajdukiewicz (1935), Lambek (1958), Steedman (1962) et Shaumyann (1998).

#### **2.2 Historique des Grammaires Catégorielles**

Pour retrouver les origines des grammaires catégorielles, il faut remonter aux travaux du philosophe Husserl, ainsi que des logiciens Lesniewski, Ajdukiewicz et Bar-Hillel.

##### **2.2.1 Fondements philosophiques des grammaires catégorielles : Husserl**

Ayant commencé sa carrière comme mathématicien, Husserl (1913) s'est concentré sur un domaine encore inconnu à cette époque, à savoir, la phénoménologie. L'exemple le plus connu de sa théorie est la question « Qu'est-ce qu'une fleur ».

Plusieurs réponses à cette interrogation (pourtant si banale) sont possibles en variant son aspect dans notre pensée.

Ca peut être une rose, une fleur sauvage, de couleur rouge, blanc, etc. Il est donc impossible de définir un objet sans les éléments qui l'entourent. De même dans la linguistique, on ne peut trouver de signification indépendante à tout symbole grammatical sans les éléments qui s'y attachent [3].

La théorie des catégories de signification instaurée par Edmund Husserl a constitué une source d'inspiration pour Lesniewski ainsi que plusieurs autres chercheurs qui ont essayé de développer ces idées dans le domaine de la logique et de la linguistique.

### **2.2.2 Lesniewski (1922)**

Contrairement à ce qui est connu, et d'après Godart-Wendling [16] [17], Lesniewski est le premier à s'intéresser à la linguistique et a avoir explicitement montré son intérêt pour l'analyse des langages naturels. La théorie des catégories sémantiques qu'il a inventée a permis d'éviter considérablement les Antonine. Elle se base sur deux catégories sémantiques fondamentales : (N) pour les noms et (P) pour les propositions.

Partant des fondements philosophiques instaurés par Husserl, Lesniewski a voulu retrouver une solution au paradoxe Russelien à savoir, le paradoxe du Barbier<sup>1</sup>. C'est grâce à la théorie des catégories sémantiques définies par ce logicien que ce paradoxe a pu être résolu.

En même temps, ce chercheur a proposé dans une de ses publications l'analyse d'une phrase en langage naturel en se basant sur la théorie qu'il a mise en œuvre [17].

---

<sup>1</sup> On peut lire sur l'enseigne d'un barbier " Je rase tous les hommes du village qui ne se rasent pas eux-mêmes, et seulement ceux-là. " Le problème qui se pose est : qui rase le barbier ? S'il se rase tout seul, il ne respectera pas son enseigne (il raserait quelqu'un qui se rase lui-même). Et s'il y a quelqu'un qui le rase, il aura démenti ce qui est écrit dans son enseigne qui dit qu'il raserait tous les hommes du village.

On ne peut cependant affirmer si Lesniewski a continué ou non dans le domaine de traitement des langages naturels. La majorité de ses publications et manuscrits ont été perdus et même détruits après la démolition de Varsovie en 1944.

### 2.2.3 Ajdukiewicz (1935)

Les premières versions des grammaires catégorielles sont attribuées à Kazimierz Ajdukiewicz grâce au principe de calcul de fraction qu'il a développé.

En effet, Ajdukiewicz est le premier à avoir présenté un système formalisé ayant permis de vérifier la bonne connexion syntaxique d'une expression [31]. Ses travaux sont en réalité un prolongement des recherches de Lesniewski ayant pour base la théorie des catégories sémantiques.

Pour vérifier la bonne connexion syntaxique des phrases, Ajdukiewicz a mis en œuvre une grammaire catégorielle unidirectionnelle. Cette grammaire présente plusieurs avantages, comparativement à la théorie des catégories sémantiques de Lesniewski. En effet, une notation algébrique a été mise en place pour faciliter le calcul sémantico - syntaxique dans l'analyse des phrases [16].

Dans ses travaux, Ajdukiewicz a proposé la division des catégories en deux sous-ensembles : des catégories de base et des catégories foncteurs [28]. Les catégories de base regroupent les catégories syntagmes nominaux notées par « N » et les catégories phrases notées par « S ».

De leur part, les catégories foncteurs sont obtenues grâce à une combinaison entre les catégories de base avec le symbole d'application : « — ».

Toutes les catégories du modèle d'Ajdukiewicz obéissent à ces deux règles [8] :

- (i) Les catégories de base du modèle sont des catégories de ce modèle.
- (ii) Supposons que X et Y sont deux catégories du modèle alors  $\frac{X}{Y}$  est une catégorie du modèle qu'on appelle communément « Catégorie Foncteur ».

La barre de fraction « — » symbolise l'application. Le type de résultat



est celui du numérateur « X ». Le dénominateur « Y » est le type de l'argument.

En plus de ces règles, un calcul formel a été mis en place. Ce Calcul se fait d'une manière unique de gauche à droite comme suit.

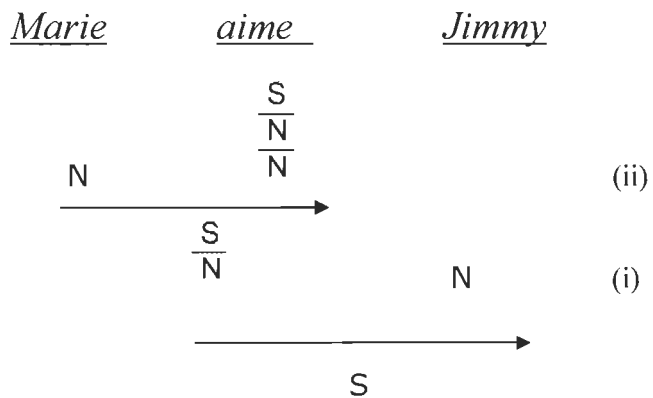
$$(i) \quad \frac{X}{Y} \quad Y \quad \longrightarrow \quad X$$

$$(ii) \quad Y \quad \frac{X}{Y} \quad \longrightarrow \quad X$$

Le but du calcul est le suivant : Ayant assigné des types catégoriels aux différents composants d'une phrase donnée, on se basera sur les différentes règles de réductions données ci-dessus dans le but d'arriver à un résultat égal à l'une des catégories de base [29].

Dans ce cas, la phrase est dite syntaxiquement correcte.

Partons maintenant du triplet (S, N, —) et appliquons-le sur un cas réel.



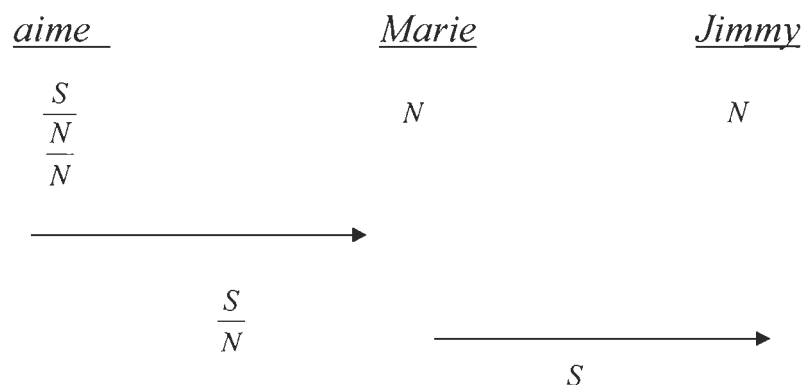
Le point de départ dans l'analyse de la phrase « *Marie Aime Jimmy* » est l'affectation des types aux différentes entités lexicales. Pour ce faire, nous avons choisis le type « N » pour les entités « Marie » et « Jimmy » et le type  $\frac{S}{\frac{N}{N}}$  pour

l'entité « aime ».

Cette affectation de type n'est pas faite au hasard. En effet, le verbe « aimer » est transitif direct. Il nécessite au moins un sujet et un complément d'objet direct. Dans notre cas le sujet et le complément d'objet direct sont tout les deux des noms propres d'où l'affectation du type « N » à ces deux entités, et le type  $\frac{S}{\frac{N}{N}}$  à l'entité « aime ».

Essayons maintenant d'expliquer le processus d'analyse. Une première réduction a été faite en se basant sur la règle (ii) elle nous a donné le résultat  $\frac{S}{N}$ . Par la suite, une deuxième réduction a eu lieu en se basant sur la règle de réduction (i) pour trouver enfin le résultat (S). Ce qui prouve que la phrase est syntaxiquement correcte.

Dans ce type d'analyse, Ajdukiewicz a défini la notion d'opérateurs principaux. Au cours de l'analyse d'une phrase, cet opérateur doit être mis en premier, suivi de ses arguments.



## 2.2.4 Les Grammaires Bar-Hillel

Ayant concentré ses travaux sur le calcul de fractions introduit par Ajdukiewicz, Bar-Hillel a réussi à développer une version encore plus évoluée de Grammaire Catégorielle appelée : Grammaire catégorielle bidirectionnelle. Cette nouvelle version a été largement utilisée dans le domaine de la linguistique.

Plus connue sous le nom de la Grammaire AB, cette génération de Grammaire Catégorielle possède plusieurs particularités par rapport à la grammaire unidirectionnelle.

En effet, les grammaires AB sont totalement lexicalisées et possèdent des règles de productions universelles [29] [16].

Les catégories de base assignées aux éléments d'une phrase sont au nombre de deux :

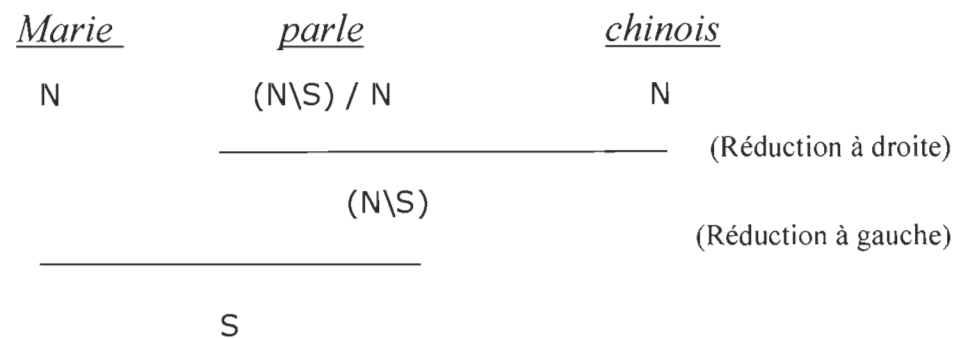
- Les types atomiques : Ces types représentent des expressions complètes.
- Les types composés : Désigne l'ensemble de types atomiques formés grâce aux opérateurs bidirectionnels ( $/$ ,  $\backslash$ ) qui offrent la possibilité de réaliser des réductions dans les deux sens (gauche et droite).

#### Règles de réduction

(i)  $X/Y \quad Y \longrightarrow X$  Réduction à droite

(ii)  $Y \quad Y \backslash X \longrightarrow X$  Réduction à gauche

Voyons maintenant comment peut-on appliquer ce modèle dans le cadre de l'analyse de la phrase « *Marie parle chinois* »



Dans la phrase « *Marie parle chinois* » on a affecté le type « N » à l'entité « Marie » qui représente le groupe nominal dans cet exemple. Le verbe « parler » est transitif direct, il attend un sujet à sa gauche et un complément d'objet à sa droite d'où l'affectation du type  $(N \backslash S) / N$  à cette entité.

Cette analyse nous a donné comme résultat la catégorie de base « S » ce qui prouve que la phrase ci-dessus est syntaxiquement correcte.

La Grammaire AB a permis de standardiser le typage de quelques mots de la langue française. Prenons par exemple le cas des déterminants : un, le, une, la, les, etc. Ces déterminants doivent s'associer à des noms dans le but de former des Groupes nominaux. On leur associe alors le type : N/N

De même pour les verbes transitifs, ces derniers sont précédés par des sujets et précédents des compléments d'objet directs ou indirects. D'où l'assignation du type  $(N \backslash S)/N$  à ces verbes [8].

#### **2.2.5 Calcul de Lambek (1958)**

Lambek propose dans ses recherches publiées en 1958 et 1961, une formulation logique des grammaires catégorielles et des types syntaxiques [8].

Le modèle proposé par ce chercheur traite les types syntaxiques comme des formules.

Lambek a introduit un ensemble de règles et de notations permettant de faciliter l'analyse des phrases d'un langage naturel [25] [8].

Le système décrit par Lambek comprend :

- (i) Les types primitifs S et N ;
- (ii) Deux opérateurs constructeurs / et \ ;
- (iii) Un opérateur de concaténation représenté par le symbole (+) : Pour deux expressions  $e_1$  et  $e_2$  de types respectifs X et Y, la concaténation de ces deux expressions sera notée :  $e_1 + e_2$  et sera de type  $X+Y$  ;
- (iv) Une Règle de réduction «  $\longrightarrow$  ». Selon cette règle, la notation  $X \longrightarrow Y$  signifie que le type X se réduit au type Y. De même, la notation  $X \leftrightarrow Y$  signifie à la fois que le type de X se réduit au type Y et que le type Y se réduit au type X.

Le système de calcul de types de Lambek se base sur les règles suivantes [8] [22] :

$$(1) \quad \text{La réflexivité : } X \longrightarrow X$$

$$(2) \quad \text{L'associativité : } (X+Y)+Z \longrightarrow X+(Y+Z)$$

$$X+(Y+Z) \longrightarrow (X+Y)+Z$$

$$(3) \quad \text{La transitivité :}$$

$$\text{Si } X \longrightarrow Y \text{ et } Y \longrightarrow Z \text{ Alors : } X \longrightarrow Z$$

$$(4) \quad \text{Si } X + Y \longrightarrow Z \text{ Alors } X \longrightarrow Z / Y$$

$$(5) \quad \text{Si } X + Y \longrightarrow Z \text{ Alors } Y \longrightarrow Z \setminus X$$

$$(6) \quad \text{Si } X \longrightarrow Z / Y \text{ Alors } X + Y \longrightarrow Z$$

$$(7) \quad \text{Si } Y \longrightarrow Z \setminus X \text{ Alors } X + Y \longrightarrow Z$$

À partir des règles présentées ci-dessus, on peut déduire les théorèmes suivants :

$$(T1) \quad X \longrightarrow (X + Y)/Y$$

$$(T2) \quad (Z/Y) + Y \longrightarrow Z$$

$$(T3) \quad Y \longrightarrow Z \setminus (Z / Y)$$

$$(T4) \quad (Z/Y) + (Y/X) \longrightarrow (Z / X)$$

$$(T5) \quad Z/Y \longrightarrow (Z/X) / (Y/X)$$

$$(T6) \quad (Y \setminus X) / Z \longrightarrow (Y / X) \setminus Z$$

$$(T7) \quad (X / Y) / Z \longrightarrow X / (Z + Y)$$

$$(T8) \quad \text{Si } X \longrightarrow X' \text{ et } Y \longrightarrow Y' \text{ Alors } X + Y \longrightarrow X' + Y'$$

(T9) Si  $X \longrightarrow X'$  et  $Y \longrightarrow Y'$  Alors  $X / Y' \longrightarrow X' / Y$

## 2.2.6 Grammaire Catégorielle combinatoire de Steedman

La grammaire catégorielle de Steedman est une généralisation de tous les modèles de grammaires catégorielles qui l'ont précédé. Toutes les propriétés des grammaires catégorielles classiques sont adoptées.

Cependant, plusieurs autres opérateurs entrent en scène.

L'un des apports de la grammaire de Steedman est l'introduction de nouvelles catégories autres que les catégories de bases (S, N) :

- La catégorie PP désignant un « Groupe Prépositionnel »
- La catégorie VP désignant un « Syntagme Verbal »
- La catégorie NP désignant un « Syntagme Nominal »

Nous verrons dans ce qui suit un exemple donné par Marck Steedman qui traite le cas d'attribution de catégorie pour le verbe : aimer [9].

Cette attribution est faite comme suit :

**(S :aimer` np2 np1\NP : np1)/ NP : np2**

Essayons maintenant d'expliquer et de clarifier la notation citée ci-dessus. Les lettres minuscules représentent les variables sémantiques, les constantes sont identifiées par les ( ` ).

Les types syntaxiques sont en lettres majuscules.

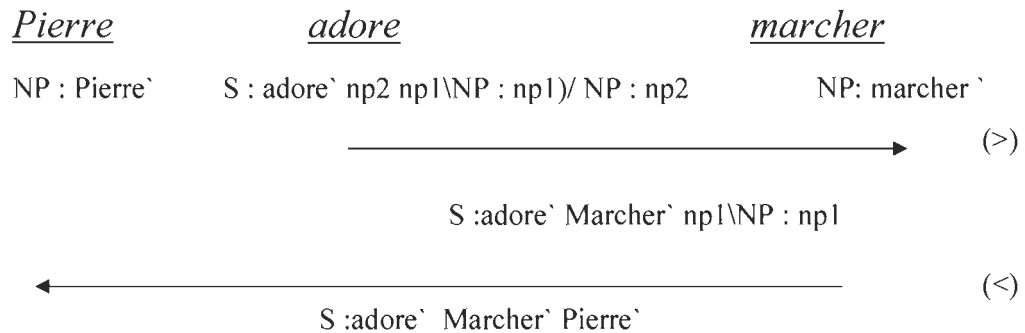
Comme exemple, appliquons cela à une phrase de la langue française

On utilisera les règles de réduction de base (avant et arrière):

$X/Y \quad Y \quad \longrightarrow \quad X \quad \text{Réduction à droite ( > )}$

$Y \quad Y \backslash X \quad \longrightarrow \quad X \quad \text{Réduction à gauche ( < )}$

**Exemple :**



Cette théorie se base sur deux catégories sémantiques fondamentales : (N) pour les noms et (P) pour les propositions. En plus des règles de réduction classiques, Steedman a introduit de nouvelles règles dans le but de remédier à quelques problèmes que les grammaires AB trouvent des difficultés à résoudre.

Parmi ces règles, on peut citer :

- Les règles de composition fonctionnelles.
- Les règles de changement de type.
- Les règles de substitution fonctionnelle.

On se limitera dans cette section à énumérer les règles sans s'y attarder. Mais on mettra l'emphasis sur la manière avec laquelle on peut les utiliser [7] [8].

(i) Les règles de composition fonctionnelles.

Ces règles sont au nombre de 4 :

$\frac{[X/Y : F] + [Y/Z : G]}{[X/Z : \lambda x (F(Gx))]} \rightarrow > (B) ;$	$\frac{[X/Y : F] + [Y\backslash Z : G]}{[X\backslash Z : \lambda x (F(Gx))]} \rightarrow > (Bx)$
$\frac{[Y\backslash Z : G] + [X\backslash Z : F]}{[X\backslash Z : \lambda x (F(Gx))]} \rightarrow < (B) ;$	$\frac{[Y/Z : G] + [X\backslash Y : F]}{[X/Z : \lambda x (F(Gx))]} \rightarrow < (Bx)$

(ii) Les règles de changement de types

Steedman présente 4 règles de changement de type. Ces règles offrent la possibilité de changer des fonctions en des arguments [7] [8].

$$\begin{array}{c}
 \frac{[X : x]}{[Y/(Y \setminus X) : \lambda F(Fx)]} \rightarrow > (T) ; \quad \frac{[X : x]}{[Y/(Y/X) : \lambda F(Fx)]} \rightarrow > (Tx) \\
 \\
 \frac{[X : x]}{[Y \setminus (Y/X) : \lambda F(Fx)]} \rightarrow < (T) ; \quad \frac{[X : x]}{[Y \setminus (Y \setminus X) : \lambda F(Fx)]} \rightarrow < (Tx)
 \end{array}$$

(iii) Les règles de substitution fonctionnelle

Les règles de substitution fonctionnelle proposée par Steedman permettent d'apporter des solutions aux blocages que les règles de changement de types et les règles de composition fonctionnelles n'ont pu résoudre [7].

Il existe aussi 4 règles de substitution fonctionnelles de base :

$$\begin{array}{c}
 \frac{((X/Y)/Z : F) - (Y/Z : G)}{(X/Z) : \lambda_x (Fx (Gx))} \rightarrow (>S) ; \quad \frac{((X/Y) \setminus Z : F) - (Y \setminus Z : G)}{(X \setminus Z) : \lambda_x (Fx (Gx))} \rightarrow (>Sx) \\
 \\
 \frac{(Y \setminus Z : G) - ((X \setminus Y) \setminus Z : F)}{(X \setminus Z) : \lambda_x (Fx (Gx))} \rightarrow (<S) ; \quad \frac{(Y/Z : G) - ((X \setminus Y)/Z : F)}{(X/Z) : \lambda_x (Fx (Gx))} \rightarrow (<Sx)
 \end{array}$$

### 2.2.7 Grammaire Applicative Universelle de Shaumyan

Shaumyan propose dans ses travaux (1967, 1977, 1987, 1998) une distinction entre deux niveaux de représentation et d'analyse d'une langue. Le niveau phénotype et le niveau génotype. Ses travaux sont fondés sur une opposition avec les hypothèses



instaurées par les grammaires de réécriture. Le principal défi relevé par Shaumyan est de mettre en œuvre une grammaire applicative universelle [8] [29].

Nous allons commencer par présenter les deux niveaux définis par Shaumyan pour conclure avec une présentation générale du modèle de la grammaire Applicative.

Shaumyan nous a permis de distinguer entre deux niveaux d'analyse et de représentation d'une langue, le niveau phénotype et le niveau génotype [9] [11].

**Le niveau phénotypique :** Communément connu sous le nom de « niveau de surface » ou encore 'niveau observable', ce niveau regroupe l'ensemble des représentations concaténées. Il permet d'évoquer tous les traits spécifiques d'une langue donnée à travers la description de l'ordre de ses mots, sa syntaxe ou encore tous ses traits morphologiques [14].

**Le niveau génotypique :** C'est un niveau abstrait tel que décrit par Shaumyan. Ce niveau permet de retracer les invariants sémiotiques, principaux constituants des langages naturels. La description de ces invariants se fait sous forme d'opérations et de relations [9][14].

Contrairement au niveau phénotypique présenté sous forme d'expressions concaténées telles qu'observées dans la langue naturelle d'origine, le niveau génotypique est présenté sous forme d'un langage abstrait appelé langage génotypique.

Supposé totalement indépendant de toute langue naturelle, ce langage formel possède ses propres règles se basant sur la logique combinatoire de Curry et sur le principe d'opérateur/opérande [8].

Selon Shaumyan, le langage génotypique est « chargé de mettre en évidence les opérateurs universels des langues [15] » d'où la nécessité d'une grammaire universelle GU pour une telle représentation.

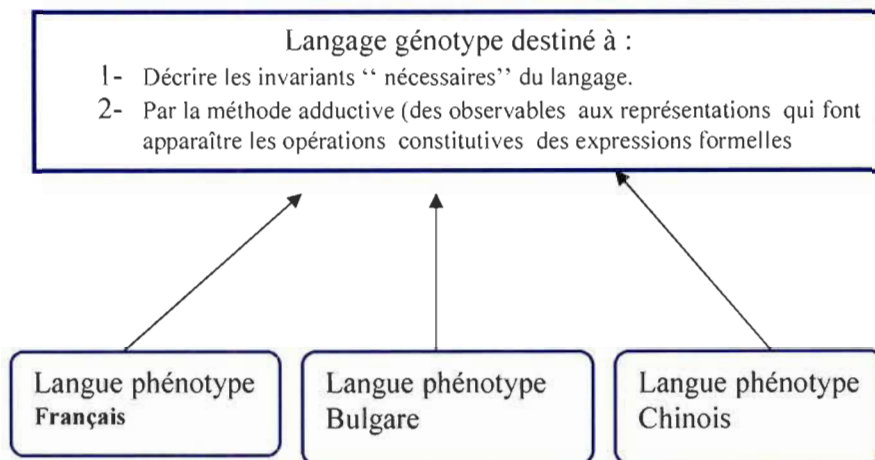


Figure 1 : Le modèle de la Grammaire Catégorielle Universelle de Shaumyann [15].

### 2.2.8 Grammaire Applicative et Cognitive (1990)

La grammaire Applicative et cognitive est une version évoluée de la grammaire applicative universelle de Shaumyann. Créée par Desclés, cette nouvelle version de grammaire applicative universelle a permis d'introduire un troisième niveau de représentation d'une langue naturelle.

En effet, la grammaire applicative et cognitive reprend le niveau génotype et phénotype tout en y rajoutant un autre niveau appelé niveau cognitif. Ce niveau aura pour tâche de décrire et représenter formellement les significations des prédicats du niveau génotype ainsi que le sens des configurations linguistiques [8].

Nous éviterons de nous approfondir dans l'étude de la grammaire applicative et cognitive et nous nous limiterons à une description générale de ce nouveau formalisme.

## 2.3 Conclusion

Nous nous sommes consacrés dans ce chapitre à l'étude de différentes versions des grammaires catégorielles tout en prenant soin de conserver l'ordre chronologique de leurs apparitions. Cette analyse nous a permis de comprendre le passage du concept instauré par Husserl aux versions récentes de grammaires catégorielles qui sont utilisés dans l'analyse de textes.

## **CHAPITRE 3**

### **LOGIQUE COMBINATOIRE**

#### **3.1 Introduction**

Dans ce chapitre, nous introduirons la notion de la logique combinatoire fondamentale. Cette partie est nécessaire pour pouvoir présenter notre prochain chapitre à savoir, les grammaires catégorielles combinatoires applicatives.

On commencera par introduire le cadre historique dans lequel est apparu le concept de la logique combinatoire. On passera ensuite à l'explication et à la présentation de quelques notions rattachées à ce concept. Et on finira par la description de quelques combinateurs qui nous seront très utiles dans le reste de notre travail.

#### **3.2 Cadre historique**

Introduite par M. Schonfinkel en 1924, la logique combinatoire n'a cessé d'évoluer tout en montrant son efficacité dans plusieurs domaines.

Cette logique avait pour finalité de supprimer les variables mathématiques et de mieux formaliser les notions de fonctions. Reprise par Curry et Feys en 1958, la logique combinatoire a été adaptée dans le but de pouvoir l'utiliser dans le domaine de calcul de prédicats [8].

C'est ainsi que la logique combinatoire a permis d'étudier et de trouver des solutions logiques à des paradoxes très connus à cette époque tel que celui de Russel.

Dans sa forme actuelle, elle est rendue « un système applicatif sans variable, utilisant des opérateurs abstraits » appelés combinateurs. Pour chaque combinateur est associée une règle de simplification (règles  $\beta$ réduction) et une règle d'introduction.

### 3.3 Combinateurs logiques

Un combinateur est un opérateur « abstrait qui nous permet de construire à partir d'opérations, des opérations de plus en plus complexes » [8].

En définissant le principe de combinateurs logiques, Shoenfinkel et H.B Curry ont voulu rendre la logique comme un calcul ordinaire. Cette volonté de mécaniser le raisonnement n'a pas donné ses fruits. De nos jours, plusieurs travaux ont montré que la mécanisation du raisonnement induit le paradoxe de Curry. Selon ce principe, on peut arriver à n'importe quelle conclusion en partant d'un ensemble de règles logiques et d'une phrase autoréférentielle.

L'application d'un quelconque combinateur sur un argument obéit à une règle nommée  $\beta$ réduction.

On s'intéressera dans ce qui suit à la description de quelques combinateurs logique [8] [14] [26].

#### 3.3.1 Combinateur d'effacement K

Le combinateur d'effacement K est traduit par la règle  $\beta$ réduction suivante :

$$K \text{ } f \text{ } x \longrightarrow f$$

Il s'agit d'une fonction-processus prenant deux arguments en entrées et rend le premier comme résultat.

#### 3.3.2 Le combinateur de distribution S

Étant donné deux opérateurs X et Y respectivement unaire et binaire. L'association de ces opérateurs à un combinateur logique S forme un opérateur complexe SXY. Appliqué à un opérande X, ce combinateur agit comme suit :

$$S \text{ } X \text{ } Y \text{ } Z \longrightarrow X \text{ } Z \text{ } (Y \text{ } Z)$$

Cet opérateur a déjà été utilisé avec Steedman.

### 3.3.3 Le combinateur identité I

Associé à un quelconque opérateur P, le combinateur d'identité I n'a aucun effet. La règle  $\beta$ réduction qui le caractérise est la suivante :

$$IP \longrightarrow P$$

Voyons maintenant comment on peut représenter le Combinateur identité I à l'aide d'un combinateur complexe composé de S et K.

Par définition, le combinateur d'identité est défini par le combinateur complexe SKK.

$$\begin{aligned} SKKx &= \\ &= Kx(Kx) \\ &= Kx \\ &= x \end{aligned}$$

### 3.3.4 Le combinateur de composition B

Le combinateur B s'applique à deux opérateurs X et Y pour former un opérateur complexe BXY. En associant à BXY un argument Z, on produit la règle  $\beta$ réduction suivante :

$$BXYZ \longrightarrow X(YZ)$$

En d'autres termes, le combinateur logique B se comporte comme une composition de deux fonctions X et Y.

$$BXYZ \Leftrightarrow X \circ Y(Z)$$

Le combinateur B peut à son tour être représenté à l'aide des combinateurs K et S. En effet, B est par définition donné par S(KS)K.

Essayons de le démontrer :

Il faut noter que l'application des combinateurs complexes se fait en appliquant les combinateurs élémentaires par ordre de gauche à droite

$$\begin{aligned}
 BXYZ &= \\
 &= S(KS)KXYZ \\
 &= KSX(KX)YZ \\
 &= S(KX)YZ \\
 &= KXZ(YZ) \\
 &= X(YZ)
 \end{aligned}$$

### 3.3.5 Le combinateur de duplication W

Étant donné un opérateur  $f$ , l'application du combinateur  $W$  à  $f$  produit un opérateur complexe  $Wf$ . Cet opérateur agit sur un opérande  $x$  comme suit :

$$Wf x \longrightarrow f x x$$

Le combinateur duplicateur  $W$  est surtout utilisé dans le cas des prédicats réfléchis (*se-laver, se-blesser, se-raser,...*). Prenons l'exemple du prédicat *se-blesser* et appliquons le à Pierre. L'expression résultante sera *se-blesse Pierre*.

Posons *se-blesse* =  $W$  *blesse*, le remplacement de cette expression combinatoire dans l'expression *se-blesse Pierre* donne  $W$  *blesse Pierre*.

La réduction de cette nouvelle expression combinatoire donne comme résultat l'expression : *blesse Pierre Pierre*.

Cela montre que les deux propositions *se-blesse Pierre* et *blesse Pierre Pierre* sont sémantiquement équivalentes [8].

Voyons maintenant comment  $W$  est exprimé à l'aide des combinateurs  $K$  et  $S$ .

$W$  est par définition donné comme suit :

$$W = SS(KI)$$

Nous avons aussi :

$$SS(KI)fx \longrightarrow fxx$$

$$SS(KI)fx = SSI(KI)fx$$

$$\begin{aligned}
&= \text{SI}f(KIf)x && //\text{passage après application de S} \\
&= f_x(KIfx) \\
&= f_x(Ix) \\
&= \mathbf{fxx} && \mathbf{(vrai)}
\end{aligned}$$

### 3.3.6 Le combinateur de coordination $\phi$

Supposons le cas où on dispose de trois opérateurs F, G, H et d'un opérande x. L'opérateur complexe  $\phi$  FGH appliqué à l'opérande x donne le résultat suivant :

$$\phi \text{ FGH}x \longrightarrow \mathbf{F(Gx)(Hx)}$$

Le combinateur  $\phi$  a une grande importance dans l'étude du cas de coordination dans le domaine de la linguistique.

### 3.3.7 Le combinateur de distribution $\psi$

Appliqué sous les mêmes conditions citées ci-dessus, le combinateur  $\psi$  donne le résultat qui suit la règle  $\beta$ réduction suivante :

$$\psi \text{ FGH}x \longrightarrow \mathbf{F(Gx)(GH)}$$

### 3.3.8 Le combinateur de changement de type $C^*$

Le combinateur de changement de type  $C^*$  permet de changer le statut d'un opérateur en opérande et inversement.

Ce combinateur est caractérisé par la règle  $\beta$ réduction suivante :

$$\mathbf{C^* XY} \longrightarrow \mathbf{YX}$$

Autrement dit, si X est un opérateur auquel est associé l'opérande Y alors, X devient opérande de l'opérateur  $(C^* Y)$  [8].

### 3.3.9 Le combinateur de permutation C

Le combinateur de permutation C permet d'exprimer la commutativité d'une opération.

Ce combinateur se traduit par la règle  $\beta$ réduction suivante :

$$\mathbf{CFXY} \longrightarrow \mathbf{FYX}$$

Le combinateur « C » est utilisé pour la description des verbes symétriques (épouser, rencontrer, croiser, etc.) [8].

Prenons le cas de l'expression : *Pierre rencontre Marie*. Et montrons que cette proposition est équivalente à *Marie rencontre Pierre*.

L'expression combinatoire relative à *Pierre rencontre Marie* est *rencontre Pierre Marie*.

Posons : *rencontre* = *C rencontre*.

*rencontre Pierre Marie*  $\leftrightarrow$  *C rencontre Pierre Marie*

L'application de la règle de réduction associée au combinateur C nous donne : *rencontre Marie Pierre*.

*Pierre rencontre Marie*  $\leftrightarrow$  *Marie rencontre Pierre*.

Voyons maintenant comment ce combinateur peut être représenté à l'aide des combinateurs K, B et S.

Par définition, le combinateur de permutation est équivalent au combinateur complexe S(BBS)(KK) en voici la démonstration :

$$\begin{aligned} \mathbf{CFXY} &= \mathbf{S(BBS)(KK)FXY} \\ &= \mathbf{BBSF(KKF)XY} \\ &= \mathbf{B(SF)(KKF)XY} \\ &= \mathbf{SF(KKFX)Y} \\ &= \mathbf{FY(KKFX)} \\ &= \mathbf{FY(KXY)} \\ &= \mathbf{FYX} \end{aligned}$$



### 3.4 Combinateurs complexes

Les combineurs complexes sont le résultat de l'association de plusieurs combineurs élémentaires à la fois.

Exemple : BCC, BC\*. etc

L'action d'un combineur complexe est déterminée par l'application successive de ses combineurs élémentaires de gauche à droite [8].

L'exemple suivant [8] montre l'application du combineur complexe BCC aux opérandes x, y, z. La réduction de BCC se fait selon l'ordre suivant : Réduction de B puis de C puis de C.

- (1) BCCxyz
- (2) C(Cx)yz
- (3) (Cx)zy
- (4) Xy

#### 3.4.1 Utilisation des puissances

L'attribution d'une puissance « n » à un combineur nous permet de l'itérer n fois.

Ce genre de combineurs complexes obéit à la règle suivante [8]:

« Si  $\chi$  est un combineur alors  $\chi^n$  itère n fois l'action du combineur  $\chi$  tel que :

$\chi^1 = \chi$  et  $\chi^n = B \chi^n \chi^{n-1}$  » [8].

Exemple :

- (1)  $B^2abcd$
- (2) BBBabcd
- (3) B(Ba)bcd
- (4) Ba(bc)d
- (5) a(bcd)

### 3.5 Conclusion

On assiste de nos jours à un grand intérêt prêté à la logique combinatoire spécialement dans le domaine de l'informatique. Ces combineurs sont largement

utilisés dans la conception de langages de programmation fonctionnelle où ils sont considérés comme modèle théorique de calcul. Dans le chapitre qui va suivre, nous montrerons comment les combinateurs logiques sont ingérés à la grammaire catégorielle combinatoire pour former une nouvelle version de grammaire catégorielle appelée Grammaire catégorielle combinatoire applicative.

## **CHAPITRE 4**

### **GRAMMAIRE CATEGORIELLE COMBINATOIRE**

### **APPLICATIVE**

#### **4.1 Introduction**

Nous consacrerons ce chapitre à la partie la plus importante de notre mémoire : La grammaire catégorielle combinatoire applicative notée : GCCA.

Le chapitre précédent a été consacré à la présentation de l'évolution de la grammaire catégorielle en partant des premières versions ayant vu le jour avec Lesniewski qui s'est basé sur des fondements philosophiques instaurés par Husserl. Nous avons présenté les travaux d'Ajdukiewicz qui, pour sa part, a contribué par une distinction entre les catégories opérateurs et les catégories de base et a permis de présenter le principe de calcul de fractions.

Par la suite, on a essayé d'expliquer les fondements de la grammaire AB de Bar-Hillel et al. Le calcul de Lambek, la grammaire catégorielle combinatoire de Steedman pour finir avec la grammaire applicative universelle de Shaumyan et avec la grammaire applicative et cognitive.

On s'intéressera dans ce chapitre à une nouvelle version des grammaires catégorielles à savoir la grammaire catégorielle combinatoire applicative [8].

#### **4.2 Principe de la GCCA**

La Grammaire Catégorielle Combinatoire Applicative n'a cessé de montrer son efficacité en apportant des solutions à des problèmes face auxquels les anciennes versions des grammaires catégorielles s'avèrent inefficaces (cas de la pseudo ambiguïté par exemple) [13].

Ce nouveau modèle de grammaire catégorielle perçoit toute langue comme un « système d'agencement d'unités linguistiques ». Tout y est traité comme étant un

opérateur ou un opérande. L'application d'un opérateur à un opérande génère une nouvelle unité qui sera à son tour un nouvel opérateur ou opérande [13].

Comme toute grammaire catégorielle, la GCCA part du principe d'assignation de types aux entités linguistiques formant la structure concaténée qu'on désire analyser. Ces types sont le résultat d'une combinaison entre les types de bases (S, et N) avec les opérateurs ( /, et \).

La question qui se pose est : Quels sont les principaux apports de la Grammaire catégorielle applicative et cognitive ?

Étant donné une généralisation de plusieurs modèles catégoriels, cette nouvelle version de grammaire catégorielle hérite de tous les avantages de ces anciennes versions tout en y ajoutant de nouveaux concepts.

Le principal apport étant l'introduction d'un système formel permettant de relier le niveau phénotype au niveau génotype [14].

De plus, ce modèle permet « une complète abstraction de la structure grammaticale de sa représentation linéaire, ainsi qu'une complète abstraction de la grammaire de son lexique » [13].

Cette grammaire est donc indépendante de la langue qu'on veut analyser.

Essayons maintenant de voir la logique sur laquelle se base ce nouveau modèle de Grammaire catégorielle.

Comme on l'a déjà cité ci-dessus, la GCCA part du principe fondamental des grammaires catégorielles qui consiste à assigner des types aux différentes entités linguistiques formant la structure concaténée qu'on désire analyser.

Cette assignation est accompagnée d'une « association canonique entre les règles catégorielles combinatoires de Steedman et les combinateurs de la logique combinatoire de Curry » [14].

### 4.3 Les règles de la GCCA

Dans le but de réaliser les réductions des différents types qu'on a assignés initialement aux entités de la structure qu'on veut analyser, on aura recours aux règles citées ci-dessous. Il s'agit des mêmes règles utilisées par Steedman pour vérifier la bonne connexion syntaxique d'une phrase. La différence réside dans le fait qu'on ne s'arrêtera pas à une simple vérification de structures.

En effet, l'application des combinateurs logiques nous permettra par la suite d'avoir de nouvelles structures appelées : expressions applicatives [7].

(i) Les règles d'application avant :

$$\frac{[X/Y : U_1] - [Y : U_2]}{[X : (U_1 U_2)]} \quad ( > )$$

L'exemple ci-dessous montre comment cette règle est appliquée. Dans le cas de la Phrase : *La cigarette tue*, on commence par assigner des types aux différents unités lexicales. Une fois cette affectation de types réalisée on regarde s'il y a des possibilités de réduction.

*La cigarette tue*

$[N/N : La] - [N : cigarette] - [S \backslash N : tue]$

$[N : (La cigarette)] - [S \backslash N : tue] \quad ( > )$

La règle d'application avant qu'on a utilisé est la suivante :

$$\frac{[N/N : La] - [N : cigarette]}{[N : (La cigarette)]} \quad ( > )$$

(ii) Les règles d'application arrière :

$$\frac{[Y : U_1] - [X \backslash Y : U_2]}{[X : (U_2 U_1)]} \quad (<)$$

Si l'on continue avec le même exemple, on aura à appliquer la règle d'application arrière (<) :

$$\begin{array}{l} La \qquad \qquad cigarette \qquad \qquad tue \\ [N/N : La] - [N : cigarette] - [S \backslash N : tue] \\ [N : (La cigarette)] - [S \backslash N : tue] \qquad \qquad \qquad (>) \\ [S : (tue (La cigarette))] \qquad \qquad \qquad (<) \end{array}$$

$$\frac{[N : (La cigarette)] - [S \backslash N : tue]}{[S : (tue (La cigarette))]} \quad (<)$$

Remarque : Pour toutes les règles qu'on va citer, on utilisera les standards suivants :

- (1) X, Y, X/Y, X\Y sont des types orientés.
- (2) La notation :  $[X : U_1]$  signifie :  $U_1$  est une unité linguistique de Type X
- (3) La notation :  $[X/Y : U_1] - [Y : U_2]$  représente deux unités linguistiques  $U_1$  et  $U_2$  de types respectifs X/Y et Y.

En plus des règles d'application avant et arrière, la GCCA utilise les règles B, C\* issues de la grammaire catégorielle combinatoire de Steedman. Ces règles sont introduites dans les séquences syntagmatiques, ce qui permet de passer d'une structure concaténée à une structure applicative [9].

On s'intéressera dans ce qui suit aux règles de changement de types et de composition fonctionnelle [9][13][10][7] :

Pour chaque règle, on donnera un exemple illustratif montrant comment on peut l'appliquer.

Règles de changement de type :

$$\frac{[X : u]}{\text{-----}>\mathbf{T}} \\ [Y/(Y \backslash X) : (\mathbf{C}_* u)]$$

$$\frac{[X : u]}{\text{-----}(<\mathbf{T})} \\ [Y \backslash (Y/X) : (\mathbf{C}_* u)]$$

$$\frac{[X : u]}{\text{-----}>\mathbf{T}_x} \\ [Y/(Y/X) : (\mathbf{C}^* u)]$$

$$\frac{[X : u]}{\text{-----}<\mathbf{T}_x} \\ [Y \backslash (Y \backslash X) : (\mathbf{C}^* u)]$$

Exemple :

*Le sport préserve votre jeunesse*

$[N/N : Le] - [N : sport] - [(S \backslash N)/N : préserve] - [N/N : votre] - [N : jeunesse]$

$[N : (Le \ sport)] - [(S \backslash N)/N : préserve] - [N/N : votre] - [N : jeunesse] \quad (>)$

$[S/(S \backslash N) : (\mathbf{C}^*(Le \ sport))] - [(S \backslash N)/N : préserve] - [N/N : votre] - [N : jeunesse] \quad (>\mathbf{T})$

Règles de composition fonctionnelle :

$$\frac{[X/Y : u_1] - [Y/Z : u_2]}{\text{-----}>\mathbf{B}} \quad ; \quad \frac{[Y \backslash Z : u_1] - [X \backslash Y : u_2]}{\text{-----}<\mathbf{B}} \\ [X/Z : (\mathbf{B} u_1 u_2)] \quad [X \backslash Z : (\mathbf{B} u_2 u_1)]$$

$$\frac{[X/Y : u_1] - [Y \backslash Z : u_2]}{\text{-----}>\mathbf{B}_x} \quad ; \quad \frac{[Y/Z : u_1] - [X \backslash Y : u_2]}{\text{-----}<\mathbf{B}_x} \\ [X \backslash Z : (\mathbf{B} u_1 u_2)] \quad [X/Z : (\mathbf{B} u_2 u_1)]$$

Voyons maintenant comment se fait réellement l'analyse d'une structure concaténée en suivant ce modèle.

#### 4.4 Typification d'unités d'une langue

Le point de départ dans toute analyse catégorielle est la typification de ses éléments. Cette typification obéit à un certain nombre de règles selon la nature de l'élément en question.

En effet, le type qu'on associera à un verbe diffère selon sa nature (verbe transitif ou intransitif). De même, le type qu'on associera à un adjectif sera différent de celui associé à un déterminant, etc.

Les trois types de base qu'on utilisera sont : les syntagmes nominaux qu'on désignera par N et les propositions notées S et N\*. La combinaison de ces trois types de base avec les opérateurs (/,\) donne une infinité de possibilités de types [15].

On essayera dans ce qui va suivre de répertorier et d'affecter les types associés à chaque élément d'une phrase suivant sa fonction grammaticale.

Cette affectation concernera des mots de la langue française. Une affectation de types similaires sera réalisée dans le chapitre suivant et visera la langue arabe.

##### 4.4.1 Quantificateur nominal

Le type associé aux quantificateurs nominaux (le, la, des, des,...) est : N\*/N. Cet opérateur s'appliquera à un syntagme nominal de type N dans le but de former un syntagme nominal complet de type N\*.

Exemple :

*La conscience*

[N\*/N : *La*] - [N : *conscience*]

[ N\* : ( *La conscience* ) ] ( > )

##### 4.4.2 Verbes intransitifs

Le modèle de la GCCA associe le type S\N\* aux verbes intransitifs.



La présence d'un complément d'objet direct ou indirect n'est pas obligatoire dans une phrase contenant un verbe intransitif.

Ce verbe s'applique sur un syntagme nominal complet N\* dans le but de former une phrase syntaxiquement et sémantiquement correcte.

Exemple : le verbe mangé, sourire, sauté, etc..

Exemple d'application : Dans le cas de la phrase *Isabelle dort*, nous utiliserons l'affectation de types donnée dans le tableau I :

Unité lexicale	Fonction grammaticale	Type affecté
<i>Isabelle</i>	Sujet	N*
<i>Dort</i>	Verbe intransitif	S\N*

tableau I : Affectation de types pour la phrase de type Sujet-Verbe intransitif.

*Isabelle*

*dort*

$[N^* : Isabelle] \quad - \quad [S\backslash N^* : dort]$   
 $[S/(S\backslash N^*): (C^* Isabelle)] - [S\backslash N^* : dort] \quad (>T)$   
 $[S : ((C^* Isabelle) dort)] \quad (>)$   
 $[S: (dort (Isabelle))] \quad (C^*)$

#### 4.4.3 Verbes transitifs

Un verbe transitif est, par définition, une unité linguistique qui attend un syntagme nominal complet noté N\* à sa droite (un complément d'objet) et un sujet ou un nom propre à sa gauche.

Le type associé aux verbes transitifs en suivant le modèle de la grammaire catégorielle combinatoire applicative est : (S\N\*)/N\* [13].

Exemple : Prenons l'exemple de la phrase « *La jalousie détruit l'amour* » et essayons de l'analyser en suivant la typification donnée dans le tableau II ci-dessous.

Unité lexicale	Fonction grammaticale	Type affecté
<i>La</i>	Quantificateur nominal	N*/N
<i>Jalousie</i>	Sujet	N
<i>Détruit</i>	Verbe transitif	(S\N*)/N*
<i>L'</i>	Quantificateur nominal	N*/N
<i>Amour</i>	Complement d'objet direct	N

tableau II : Affectation de types pour une phrase comportant un verbe transitif direct.

La                      jalousie                      détruit                      l'                      amour

$[N^*/N : La] - [N : jalousie] - [(S\N^*)/N^* : détruit] - [N^*/N : l'] - [N : amour]$   
 $[N^* : (La jalousie)] - [(S\N^*)/N^* : détruit] - [N^*/N : l'] - [N : amour] \quad (>)$   
 $[S/(S\N^*) : (C^*(La jalousie))] - [(S\N^*)/N^* : détruit] - [N^*/N : l'] - [N : amour] \quad (>T)$   
 $[S/N^* : (B(C^*(La jalousie))détruit)] - [N^*/N : l'] - [N : amour] \quad (>B)$   
 $[S/N : ((B(C^*(La jalousie)) détruit) l')] - [N : amour] \quad (>)$   
 $[S : (((B(C^*(La jalousie)) détruit) l') amour)] \quad (>)$

#### 4.4.4 Les adjectifs

Le type associé aux adjectifs suivant le modèle de la grammaire catégorielle combinatoire applicative est N/N.

Prenons un exemple d'application montrant l'analyse de la phrase « *Le grand aigle vole* » contenant un adjectif tout en utilisant l'affectation de types présentée dans le tableau III.

Unité lexicale	Fonction grammaticale	Type affecté
Le	Quantificateur nominal	N*/N
Grand	Adjectif	N/N
Aigle	Sujet	N
Vole	Verbe intransitif	S\N*

tableau III : Affectation de types pour une phrase contenant un verbe intransitif, un adjectif et un quantificateur nominal.

Le                      grand                      aigle                      vole.

$[N^*/N : Le] - [N/N : grand] - [N : aigle] - [S \setminus N^* : vole]$

$[N^*/N : B(Le\ grand)] - [N : aigle] - [S \setminus N^* : vole]$  (B)

$[N^* : (B(Le\ grand)\ aigle)] - [S \setminus N^* : vole]$  (>)

$[S : (vole\ B(Le\ grand)\ aigle)]$  (<)

Cette analyse montre que la phrase est syntaxiquement correcte.

#### 4.4.5 Les modificateurs de verbes

Plus connues sous les noms d'adverbes, ces unités lexicales s'adjoignent au verbe pour préciser son sens (*rapidement, lentement...*)

Le type associé à ce type d'adverbe est  $(S \setminus N^*) \setminus (S \setminus N^*)$

Essayons de faire l'analyse syntaxique de la phrase « *Le garçon marche péniblement* » en suivant la typification contenue dans le tableau IV.

Les affectations de types se font comme suit :

Unité lexicale	Fonction grammaticale	Type affecté
<i>Le</i>	Quantificateur nominal	$N^*/N$
<i>Garçon</i>	Sujet	$N$
<i>Marche</i>	Verbe intransitif	$S \setminus N^*$
<i>Péniblement</i>	Adverbe	$(S \setminus N^*) \setminus (S \setminus N^*)$

tableau IV : Affectation de types pour une phrase comportant un adverbe et un verbe intransitif.

*Le                      garçon                      marche                      péniblement.*

$[N^*/N : Le] - [N : garçon] - [S \setminus N^* : marche] - [(S \setminus N^*) \setminus (S \setminus N^*) : péniblement]$

$[N^* : Le\ garçon] - [S \setminus N^* : marche] - [(S \setminus N^*) \setminus (S \setminus N^*) : péniblement]$  (<)

$[N^* : Le\ garçon] - [S \setminus N^* : (péniblement\ marche)]$  (<)

$[S : ((péniblement\ marche)\ (Le\ garçon))]$

On voit que la phrase est syntaxiquement correcte ce qui prouve que notre affectation de type est bien correcte.

#### **4.5 Traitement d'une structure concaténée**

Le modèle de la grammaire catégorielle combinatoire applicative offre un nouveau processus de traitement d'une structure concaténée. Ce processus obéit à des règles basées sur une association canonique entre les règles catégorielles combinatoires présentées par Steedman et les combinateurs de Curry.

Une telle association permet une bonne vérification de la connexion syntaxique de structures analysées grâce aux règles combinatoires citées ci-dessus, et la formalisation d'expressions applicatives à l'aide des combinateurs logiques de Curry. Le traitement d'une structure concaténée suivant le modèle de la grammaire catégorielle combinatoire applicative se fait en deux étapes :

Une première étape consiste en une analyse syntaxique de la structure à traiter en utilisant les règles catégorielles combinatoires. Cette vérification de bonne connexion syntaxique est accompagnée par une introduction de combinateurs logiques au niveau de la structure syntagmatique. On aura comme résultat, une expression appartenant au langage du génotype. Le déroulement de cette étape se fait dans le niveau phénotypique [14].

Une fois l'analyse syntaxique terminée, on passe ensuite à une analyse génotypique. Cette analyse utilise les combinateurs de Curry.

La réduction successive de ces combinateurs logiques nous permet de passer à une nouvelle structure appelée « Forme normale » [14] comme le montre la figure 2.

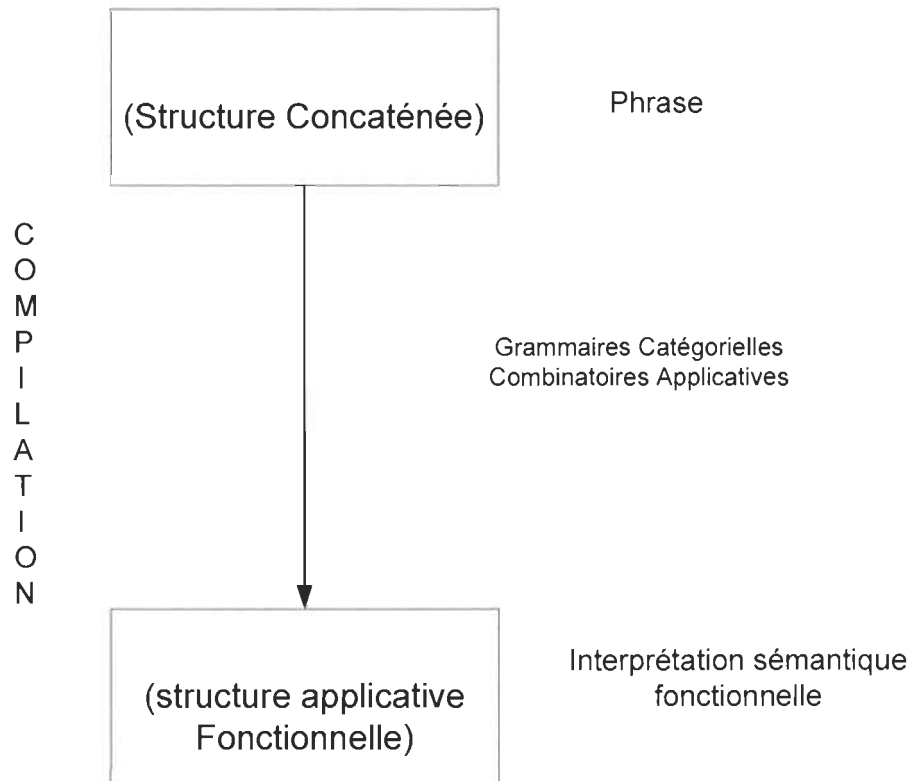


Figure 2 : Passage de la structure concaténée à une structure applicative fonctionnelle [14].

On présentera dans ce qui suit un exemple d'application du modèle de la grammaire catégorielle combinatoire applicative et mettant en évidence l'analyse d'une phrase en langue française :

*La fin justifie les moyens*

$[N^*/N:La]-[N:fin]-[(S\backslash N^*)/N^*:justifie]-[N^*/N:les]-[N:moyens]$

Analyse phénotypique :

$[N^*:(La\ fin)]-[(S\backslash N^*)/N^*:justifie]-[N^*/N:les]-[N:moyens]$  ( $>$ )

$[S/(S\backslash N^*):(C^*(La\ fin))]-[(S\backslash N^*)/N^*:justifie]-[N^*/N:les]-[N:moyens]$  ( $>T$ )

$[S/N:(B(C^*(La\ fin)\ justifie))]-[N^*/N:les]-[N:moyens]$  ( $>B$ )

$[S/N:(B(B(C^*(La\ fin)\ justifie)\ les))]-[N:moyens]$  ( $>B$ )

$[S:(B(B(B(C^*(La\ fin)\ justifie)\ les)\ moyens))]$  ( $>$ )

Génotype :

(B( B (C\*(*La fin*) justifie) les) moyens)

( B (C\*(*La fin*)( justifie (les moyens)) (B)

(C\* (*La fin*) ( justifie (les moyens)) ) (B)

(( justifie (les moyens)) (*La fin*) ) (C\*)

#### 4.6 Les métarègles:

Le traitement d'une structure concaténée suivant le modèle de la grammaire catégorielle combinatoire applicative n'est pas si évident que cela en a l'air.

L'utilisation des règles de changement de type obéit à un certain nombre de métarègles qu'on essayera d'expliquer dans ce qui suit. Cela nous permettra d'éviter l'explosion combinatoire due à l'utilisation excessive des règles de changement de type [14]. L'ensemble des métarègles qu'on va citer est expliqué en détail dans [8] [14], elles traitent du cas de la langue française.

##### 4.6.1 Métarègle 1 :

En présence de deux unités lexicales concaténées  $u_1$  et  $u_2$  de types respectifs :  $[u_1 : N]$   $[u_2 : (Y \setminus N)/Z]$ , nous serons amenés à appliquer le changement de type ( $>T$ ) à  $u_1$ .

Ainsi :  $[N : u_1 \longrightarrow Y / ( Y \setminus N) : (C^* u_1)]$

Par la suite, on aura à appliquer le combinateur logique ( $>B$ ).

Cette métarègle est généralement utilisée au niveau des phrases contenant un verbe transitif. « *La folie frôle la sagesse* ». Le tableau V nous permet de connaître les différents types catégoriels affectés aux unités de la phrase.

Unité lexicale	Fonction grammaticale	Type affecté
<i>La</i>	Quantificateur nominal	$N^*/N$
<i>Folie</i>	Sujet	$N$
<i>Frôle</i>	Verbe transitif	$(S \setminus N^*)/N^*$
<i>La</i>	Quantificateur nominal	$N^*/N$
<i>Sagesse</i>	Complément d'objet direct	$N$

tableau V : Affectation de types pour la phrase comportant un verbe transitif direct.

*La folie frôle la sagesse*

$[N^*/N:La] - [N:folie] - [(S\backslash N^*)/N^*:frôle] - [N^*/N:la] - [N:sagesse]$

**Analyse phénotypique :**

$[N^*: (La folie)] - [(S\backslash N^*)/N^*: frôle] - [N^*/N: la] - [N: sagesse]$  ( $>$ )

$[S/(S\backslash N^*): (C^*(La folie))] - [(S\backslash N^*)/N^*: frôle] - [N^*/N: la] - [N: sagesse]$  ( $>T$ ), M1

$[S/N^* (B(C^*(La folie) frôle))] - [N^*/N: la] - [N: sagesse]$  ( $>B$ )

$[S/N (B( B(C^*(La folie) frôle) la))] - [N: sagesse]$  ( $>B$ )

$[S/N (B( B(C^*(La folie) frôle) la) sagesse)]$  ( $>$ )

#### 4.6.2 Métarègle 2 :

Avec les mêmes conditions citées ci-dessus, si l'on suppose que  $u_1$  est de type X et  $u_2$  de type  $(Y\backslash X)/Z$ .

Avec X, Y et Z sont les résultats d'une association entre les types de base (S, N,  $N^*$ ) avec les opérateurs ( $/, \backslash$ ).

On aura à appliquer la règle de changement de Type ( $>T$ ) à  $u_1$ .

$[N: u_1 \rightarrow Y \backslash (Y / N): (C^* u_1)]$

Essayons d'appliquer cela sur la phrase suivante :

« Les symptômes disparaîtront trop vite »

L'affectation de types est donnée dans le tableau VI ci-dessous.

Unité lexicale	Fonction grammaticale	Type affecté
Les	Quantificateurs nominaux	$N^*/N$
Symptômes	Sujet	N
Disparaîtront	Verbe intransitif	$S\backslash N^*$
Trop	Adverbe	$((S\backslash N^*) \backslash (S\backslash N^*)) / ((S\backslash N^*) \backslash (S\backslash N^*))$
Vite	Adverbe	$(S\backslash N^*) \backslash (S\backslash N^*)$

tableau VI : Affectation de types pour une phrase comportant un adverbe.

*Les symptômes disparaîtront trop vite*

$[N^*/N:Les] - [N:symptomes] - [S\backslash N^*:disparaîtront] - [((S\backslash N^*) \backslash (S\backslash N^*)) / ((S\backslash N^*) \backslash (S\backslash N^*)) : trop] - [(S\backslash N^*) \backslash (S\backslash N^*): vite]$  ( $>$ )

$[N^*Les symptômes] - [S\backslash N^*:disparaîtront] - [((S\backslash N^*) \backslash (S\backslash N^*)) / ((S\backslash N^*) \backslash (S\backslash N^*)) : trop] - [(S\backslash N^*) \backslash (S\backslash N^*): vite]$

[N\*les symptômes]- [S\N\*/((S\N\*)\ (S\N\*)):(C\*disparaîtront)]-[(S\N\*)\ (S\N\*)]/ (S\N\*)\ (S\N\*))  
: trop]-[(S\N\*)\ (S\N\*): vite] ( $>T$ )

Dans ce cas, X est S\N\*, Y est S\N\* et Z est (S\N\*)\ (S\N\*)

#### 4.6.3 Métarègle 3 :

Considérons le cas d'une structure concaténée u1-u2 telle que u1 est de type (Y/N)/X et u2 est de type N.

La métarègle 3 précise qu'un changement de type arrière ( $<T$ ) doit être appliqué à u2.

Le paracétamol traite efficacement les migraines et rarement les inflammations.

#### 4.6.4 Métarègle 4 :

Pour la même structure u1-u2, si u1 est de type Y/S et u1 est de type N, on appliquera la règle de changement de type avant ( $>T$ ) à u2.

#### 4.6.5 Métarègle 5 :

La métarègle 5 stipule que pour une structure concaténée 'u-u1-u2', tel que u est une conjonction (et). Si les deux unités lexicales u1 et u2 sont de type N\*, on aura à appliquer la règle de changement de type avant ( $>T$ ) à u1.

Exemple

« Annie admire Van Gogh et Marie Picasso. »

[CONJD : et]-[N\* :Marie]-[N\* : Picasso]

[CONJD : et]-[S/(S\N\*) : (C\*(Marie))]-[N\* : Picasso] ( $>T$ )

#### 4.6.6 Métarègle 6 :

Dans le cas d'une structure concaténée u-u1-u2, tel que u est la conjonction (et), et u1, u2 sont de types respectifs : Y/X et N\*.

Dans ce cas, la règle de changement de type arrière ( $<T$ ) sera appliquée à u2.

On dira que la métarègle 6 s'applique directement après la métarègle 5 comme le montre l'exemple suivant.



Supposons que  $Y = S$  et  $X = S \setminus N^*$

Exemple

« *Annie admire Van Gogh et Marie Picasso.* »

[CONJD : *et*]-[ $N^*$  : *Marie*]-[ $N^*$  : *Picasso*]

[CONJD : *et*]-[ $S/(S \setminus N^*)$  : ( $C^*(Marie)$ )]-[ $N^*$  : *Picasso*] (<T)

[CONJD : *et*]-[ $S/(S \setminus N^*)$  : ( $C^*(Marie)$ )]-[ $(S \setminus N^*) \setminus ((S \setminus N^*) / N^*)$  : ( $C^* Picasso$ )] (<T)

#### 4.6.7 Métarègle 7 :

Pour une structure concaténée  $u-u_1-u_2$  tel que  $u$  est une conjonction de coordination (*et*) et  $u_1, u_2$  sont de types respectifs :  $N^*$ , et  $Y \setminus X$ , un changement de type arrière (<T) devra s'appliquer à  $u_1$ .

Exemple :

« *Pierre parle français couramment et italien moyennement.* »

Dans notre cas  $X = Y = (S \setminus N)$

[CONJD : *et*]-[ $N^*$  : *Italien*]-[ $(S \setminus N^*) \setminus (S \setminus N^*)$  : *moyennement*]

[CONJD : *et*]-[ $(S \setminus N^*) \setminus ((S \setminus N^*) / N^*)$  : ( $C^*(Italien)$ )]-[ $(S \setminus N^*) \setminus (S \setminus N^*)$  : *moyennement*] (<T)

#### 4.6.8 Métarègle 8 :

Supposons maintenant que pour la même structure  $u-u_1-u_2$ , telle que  $u$  est une conjonction,  $u_1$  de type  $N$  et  $u_2$  est de type  $(X \setminus N) \setminus Y$ .

On appliquera la règle de changement de type avant (>T) à  $u_1$ .

Exemple :

« *Pierre parle français couramment et italien moyennement.* »

Dans notre cas  $X = S$  et  $Y = (S \setminus N)$

[CONJD : *et*]-[ $N^*$  : *Italien*]-[ $(S \setminus N^*) \setminus (S \setminus N^*)$  : *moyennement*]

[CONJD : *et*]-[ $(S \setminus N^*) / X$  (*Italien*)]-[ $(S \setminus N^*) \setminus (S \setminus N^*)$  : *moyennement*]

L'ensemble de métarègles citées ci-dessus (1..8) sont utilisées dans le but de pouvoir appliquer les règles d'application, de composition, et de composition fonctionnelle ou distributive.

Les métarègles 9 et 10 qu'on expliquera ci-dessous traitent le cas de "coordination distributive de deux opérandes" dans ce cas le combinateur  $\phi$  sera appliqué [8].

#### 4.6.9 Métarègle 9 :

La métarègle 9 traite le cas des structures concaténées de type : u1-et-u2-u3. Si u1 et u2 sont de type N\* et u3 est de type (S\N\*)/X, on sera amené à appliquer la règle de changement de type avant (>T) aux structures u1 et u2.

Pour X=N\*

Exemple

« Annie et Julie sont amies »

[N\* : Annie]-[CONJD :et]-[N\* : Julie]-[ (S\N\*)/N\* : sont]-[N\* : amies]

[(S\N\*)/N\* : (C\*(ANNIE))]-[CONJD :et]- [(S\N\*)/N\* : ( C\*(Julie))]-[ (S\N\*)/N\* : sont]-[N\* : amies] (>T)

#### 4.6.10 Métarègle 10 :

Cette métarègle s'applique lorsqu'on est en présence d'une structure de type u1-u2-et-u3.

Si u1 est de type Y/N\*, et u1, u2 sont de type N\*.

Nous aurons à appliquer le changement de type arrière (>T) à chacune des unités u2 et u3.

Exemple :

Supposons que Y = (S\N\*)

Soit la phrase :

« Stéphane préfère manger et dormir. »

[N\* :Stéphane]-[ (S\N\*)/N\* : préfère]-[N\* :manger]-[CONJD :et]-[N\* :dormir]

[N\* :Stéphane]-[ (S\N\*)/N\* : préfère]-[ (S\N\*)\((S\N\*)/N\*):(C\*manger)]-[CONJD :et]-

[(S\N\*)\((S\N\*)/N\*):(C\*dormir)]

(<T)

#### 4.7 Conclusion :

L'application du modèle de la GCCA dans l'analyse de phrases de la langue française a montré son efficacité. Nous essayerons maintenant de voir s'il est possible de réaliser une analyse catégorielle de l'arabe. Pour ce faire, nous aurons

besoin d'explorer cette langue dans le but de prendre en considération quelques unes de ses particularités. Nous consacrerons le chapitre suivant à la présentation de quelques principes de base de la langue arabe. Nous nous limiterons aux points que nous jugerons pertinents.

## **CHAPITRE 5 :**

### **PRINCIPES DE BASE DE LA LANGUE ARABE**

#### **5.1 Introduction**

Nous présenterons dans cette section un panorama de concepts généraux associés à la langue arabe. Ces notions auront une importance capitale puisqu'elles vont nous aider à comprendre le contexte général de notre travail et des méthodes qu'on utilisera.

On commencera ce chapitre par une présentation générale de la langue arabe, de ses spécificités ainsi que des différents constituants de base d'une phrase arabe (verbes, noms,...).

On finira par donner un aperçu sur les différents travaux ayant été réalisés dans le domaine du traitement automatique de la langue arabe.

#### **5.2 Caractéristiques de la langue arabe**

##### **5.2.1 Quelques généralités**

La langue arabe est la langue sémitique la plus parlée dans le monde avec plus de 320 millions de locuteurs repartis sur plus de 22 pays.

On désigne par sémitique l'ensemble des langues parlées dans l'antiquité au moyen orient. Parmi les facteurs ayant permis à une telle expansion de cette langue on trouve la religion. L'arabe est en effet la langue du Coran et de l'islam de façon générale.

La langue arabe est composée de 28 consonnes dont trois lettres (ا, و, ي) pouvant être considérés à la fois comme consonne et voyelle.

Une autre particularité de la langue arabe c'est le fait qu'elle se lit de droite à gauche, et ne contient aucune majuscule.

### 5.2.2 Les voyelles

En arabe, les voyelles sont de deux types : brèves et longues. Les voyelles brèves sont au nombre de trois : ( ' , ◊ , ´ ). Ces voyelles sont rajoutées aux consonnes pour donner 3 façons différentes de prononcer la même lettre. On trouve aussi d'autres types de voyelles rajoutées à la fin d'un mot, lesquelles sont représentées sous forme de doublement de voyelles normales.

Le deuxième type de voyelle est la voyelle longue, cette voyelle est la combinaison d'une voyelle brève avec l'une des lettres ( و , ي , ا ).

Les voyelles jouent un rôle très important dans le domaine de traitement automatique de la langue arabe. En plus de leur importance pour enlever les ambiguïtés, ces voyelles sont de parfaits marqueurs qui peuvent nous aider à donner la fonction grammaticale d'un mot dans une phrase donnée [21].

Malgré la grande importance des voyelles dans le domaine de traitement automatique de la langue arabe, on remarque que la majorité des travaux réalisés traitent l'arabe non voyelle.

### 5.2.3 Les consonnes

L'écriture arabe composée principalement de consonnes. Ces consonnes sont au nombre de 28 et sont accompagnées par les voyelles.

La représentation graphique des lettres arabes change suivant leurs positions dans un mot donné. On sera donc en présence de 100 graphies différentes [21]. (Voir le Tableau ci-dessous)

Lettres de l'alphabet arabe	Graphies correspondantes
ء	ء - ئ - ا - أ
ب	ب - ب - ب
ت	ة - ت - ت - ت
ث	ث - ث - ث
ج	ج - ج - ج
ح	ح - ح - ح
خ	خ - ح - ح
د	د - د - د
ذ	ذ - ذ - ذ
ر	ر - ر - ر
ز	ز - ز - ز
س	س - س - س
ش	ش - ش - ش
ص	ص - ص - ص
ض	ض - ض - ض
ط	ط - ط - ط
ظ	ظ - ظ - ظ
ع	ع - ع - ع
غ	غ - غ - غ
ف	ف - ف - ف
ق	ق - ق - ق
ك	ك - ك - ك
ل	ل - ل - ل
م	م - م - م
ن	ن - ن - ن
هـ	هـ - هـ - هـ
و	و - و - و
ي	ي - ي - ي - ي - ي

### 5.3 Formes de base de la langue arabe

Un mot en arabe peut être de 3 types : un nom, une lettre ou un verbe. On donnera dans ce qui suit un panorama de ces trois sous-ensembles en se basant sur la référence la plus connue et la plus respectée dans le domaine de la linguistique et de la grammaire arabe à savoir : “ ALFIAT IBNOU MALEK”. Cette référence est un poème constitué de 1000 vers résumant toutes les règles de grammaire et de conjugaison dans la langue arabe.

#### 5.3.1 Les verbes

##### 5.3.1.1 Morphologie des verbes en arabe

En arabe, les verbes sont classés selon le nombre de consonnes qui les composent. Ou selon leurs natures.

- Si l'on se base sur le nombre de ces consonnes, on distingue les verbes ternaires (ثلاثي) : composé de trois consonnes. La majorité des verbes arabes sont extraits à partir de cette catégorie de verbes. On retrouve aussi les verbes constitués de 4 consonnes appelés verbes quadrilatères (رباعي).
- En se basant sur la nature des consonnes, on retrouve : Les verbes sains (صحيح) : les consonnes qui composent ce type de verbes ne contiennent pas les lettres ( ي، ا، و ). Les verbes non sains (تلمع) : correspondant aux verbes contenant les lettres : ( ي، ا، و ).

Si on revient à la première catégorisation, on remarque que plusieurs changements peuvent affecter la forme initiale d'un verbe.

Cette transformation peut être un dédoublement de consonne, un allongement de voyelle,...

Suite à un changement pareil, on assistera à un changement du schéma du verbe initial.

Un verbe sous forme initiale est appelé verbe nu (فعل مجرد). Ces verbes sont constitués de 3 ou 4 consonnes en plus des voyelles brèves. Le reste des verbes sont appelés Dérivés (مزيد) [21].

### 5.3.1.2 Conjugaison d'un verbe en arabe

En arabe, un verbe apparaît sous trois formes possibles selon son accomplissement. Il peut être conjugué à l'accompli (le passé, الماضي) à l'inaccompli (futur, المضارع) ou à l'impératif (الجزم) [24].

- L'accompli (le passé):

Un verbe conjugué à l'accompli (الماضي) désigne une action dont le déroulement s'est produit et s'est achevé.

Un suffixe est rajouté à la racine de tout verbe conjugué à l'accompli pour se raccorder en genre et en nombre avec le sujet [24].

Pour la racine AKALA (كَلَّ) (en Français (manger)). Si on conjugue ce verbe au passé avec le pluriel féminin (elles, هنّ) on aura à écrire :

هَنَّ أَكَلْنَ نَ

HONNA AKAL - NA

- L'inaccompli (المضارع) :

Cette forme désigne en même temps le présent et le futur. Un verbe conjugué à l'inaccompli désigne une action non achevée.

La conjugaison d'un verbe à l'inaccompli se caractérise par le rajout d'un préfixe ou de plusieurs infixations (substitution ou duplication de lettres) [24].

Pour le verbe (KATABA, écrire), en le conjuguant avec la première personne du singulier (هو) à l'inaccompli on aura à rajouter le préfixe (ي, YA) au début du verbe:

YA KTOUBOU

يَكْتُبُ

Si c'est la deuxième personne du singulier, le préfixe (ت, TA) sera rajouté :

TA KTOUBOU

تَكْتُبُ

Si l'on conjugue le même verbe au pluriel féminin, le suffixe (ن) sera rajouté à la fin du verbe.

KATAB NA

نَكْتُبُ



- L'impératif :

Dans la langue arabe, la conjugaison d'un verbe à l'impératif ne peut se faire qu'avec la deuxième personne du singulier et du pluriel féminin (singulier et duel) [24].

### **5.3.2 Les Noms**

En Arabe, les noms peuvent être classés en 3 sous catégories : les noms primitifs, les noms dérivés et les nombres [24].

Cette classification se base sur le rattachement du nom à une racine verbale ou non.

On présentera dans ce qui suit un panorama de ces différents sous ensembles. Pour chaque sous-catégorie on donnera un tableau montrant les fonctions grammaticales que peuvent occuper ces noms ainsi que le type catégoriel qui leur correspond.

#### **5.3.2.1 Les noms primitifs :**

Cette sous-catégorie regroupe l'ensemble de noms arabes qui ne sont pas dérivés d'un verbe. Ces noms gardent la même forme indépendamment du contexte dans lequel ils sont énoncés (voir le Tableau VII).

Ces noms peuvent être: des noms propres, des pronoms personnels, des prénoms démonstratifs,...

Noms Primitifs	Fonctions Grammaticale	Exemples
	Noms propres	سَلْمَى ، تَمِيمٌ ، مَهْدِي
	Pronoms personnels (ضمير منفصل)	أنا، نحن، هم، ت
	Prenoms relatifs (إسم موصول)	الذي، التي
	Pronoms démonstratifs	هَذَا، هَذِهِ، ذَلِكَ، هُؤْلَاءِ
	Prépositions	في، على، فوق، تحت
	Les conjonctions de coordination	ثم، و
	Adverbes	يَث، أينَ ح
	Quantificateurs	كل، بعض
	Noms Communs	سيارة، شجرة
	Noms du verbe	أه، أف
	Noms de voix	نَح، نَحْ
	Pronoms interrogatifs	متى، كيف، أين
	Pronoms allusifs	كَمْ، كي

tableau VII : Les noms primitifs dans la langue arabe.

### 5.3.1.2 Les noms dérivés :

Comme nous l'avons énoncé ci-dessus, cette sous-catégorie contient l'ensemble de noms qui sont dérivés de la racine d'un verbe donné.

Ces noms peuvent avoir différentes fonctions grammaticales comme le montre le tableau VIII.

Noms dérivés	Fonctions Grammaticales	Exemples
	Le nom verbal (مصدر)	وَدَّ / كَتَبَ مَوْدَّةً / كِتَابَةً
	Participe Actif	سَامِعٌ، عَارِفٌ، قَاتِلٌ
	Participe passif	مَرْفُوضٌ، مَنْقُوصٌ
	Adjectif	مَسْرُورٌ، جَمِيلٌ
	Nom de lieu	مَدْرَسَةٌ، مَرْفَأٌ
	Nom de temps	مَغْرَبٌ، عِشَاءٌ
	Nom d'une fois	هَمْسَةٌ
	Nom de maniere	كَلَامٌ "الملوك" جَلِيسَةٌ "القرفصاء"
	Nom d'instrument	مَطْرَقَةٌ، مَسْمَارٌ

tableau VIII : Les noms dérivés dans la langue arabe.

#### 5.4 Règles d'accord sujet-verbe en arabe :

Contrairement à la langue française, les phrases arabes peuvent être de deux types :

- Des phrases nominales (SVO): Commencent par un sujet qui peut être un nom primitif ou dérivé suivi d'un verbe et d'un ensemble de mots.
- Des phrases verbales (VSO): Représentent la majorité des phrases utilisées en arabe. Ces phrases commencent par un verbe suivi d'un ensemble de mots.

On verra dans ce qui suit la manière avec laquelle s'accorde le verbe avec le sujet qui lui correspond.

##### 5.4.1 Accord Verbe-Sujet pour les phrases SVO :

Ce type d'accord est appelé accord riche [20] [19]. En effet, dans les phrases de type SVO, le verbe s'accorde en personne, en genre et en nombre avec son sujet comme le montre l'exemple suivant.

ALBANATOU	ARAF	NA	ALHALLA
Les-filles	ont-résolu		la-solution
البنات	عرفن		البنات

#### 5.4.2 Accord Verbe-Sujet pour les phrases VSO :

Dans les phrases de types VOS l'accord Verbe-Sujet est appelé : accord pauvre [20][19]. Le verbe ne s'accorde qu'en genre et en personne avec son sujet comme le montre l'exemple suivant :

ARAF	AT	ALBANATOU	ALHALLA
résolu-ont	les-filles		la-solution
عرفت	البنات		الحل

Dans tous les cas, cet accord est traduit par l'ajout de marqueurs morphosyntaxiques qui varient selon la nature du sujet.

Nous allons voir dans cette section un aperçu sur ces différents marqueurs ainsi que sur les différents pronoms en langue arabe [19][20].

#### 5.5 Pronoms dans la langue arabe :

Dans la langue arabe, les pronoms peuvent se présenter sous deux formes possibles [19]. Cette classification se base sur l'attachement de ce pronom à l'unité lexicale à laquelle il s'associe. On peut distinguer les pronoms libres et les pronoms liés.

##### 5.5.1 Les pronoms libres :

Il s'agit de pronoms indépendants morphologiquement des entités lexicales auxquelles ils sont rattachés [19] [20]. Les tableaux IX et X ci-dessous regroupent l'ensemble des pronoms libres dans la langue arabe [19]. On associera à chacun de ces pronoms un type catégoriel qu'on utilisera par la suite dans notre analyse catégorielle du Corpus. Chaque type catégoriel faisant référence à l'un de ces pronoms libres sera de la forme  $N_{xyz}$ . Il comportera deux marqueurs (x, y,z) tels que :

- $x \in \{1,2,3\}$  : les chiffres 1,2 et 3 font référence respectivement la première la deuxième et la troisième personne.
- $y \in \{m, f, \phi\}$  : les lettres m et f désignent respectivement les genres masculin et féminin. Dans quelques cas, y peut avoir la valeur nulle ( $\phi$ ).
- $z \in \{s, d, p\}$  : Les lettres s, p, et d désignent respectivement le singulier et le pluriel et le dual.

Personne	Genre	Pronom	Type catégoriel
1 ère	Masculin	ANA	N <sub>1S</sub>
	Féminin	أَنَا	
2 ème	Masculin	ANTA أَنْتَ	N <sub>2MS</sub>
	Féminin	ANTI أَنْتِ	N <sub>2FS</sub>
3 ème	Masculin	HOUA هُوَ	N <sub>3MS</sub>
	Féminin	HIYA هِيَ	N <sub>3FS</sub>

Personne	Genre	Pronom	Type catégoriel
1 ère	Masculin	NAHNOU	N <sub>1D</sub>
	Féminin	نَحْنُ	
2 ème	Masculin	ANTOUMA	N <sub>2D</sub>
	Féminin	مَآنَت	
3 ème	Masculin	HOUMA	N <sub>3D</sub>
	Féminin	هُمَا	

tableau IX : Pronoms libres du singulier et du Dual.

Personne	Genre	Pluriel	Type catégoriel
1 ère	Masculin	NAHNOU	N <sub>1P</sub>
	Féminin	نَحْنُ	
2 ème	Masculin	ANTOM أَنْتُمْ	N <sub>2MP</sub>
	Féminin	ANTONNA أَنْتُنَّ	N <sub>2FP</sub>
3 ème	Masculin	HOM هُمْ	N <sub>3MP</sub>
	Féminin	HONNA هُنَّ	N <sub>3FP</sub>

tableau X : Pronoms libres pour le pluriel.

### 5.5.2 Les pronoms liés

Ces pronoms sont présents sous forme de suffixe qui sont rajoutés à différents types d'entités lexicales (Verbe, Adjectifs, Noms, Préposition).

On se limitera dans cette partie à la présentation d'un panorama de pronoms liés associés au verbe (*ARAF*, عَرَفَ, « *Savoir* ») conjugués à l'accompli [19].

Pour chacun des tableaux XI et XII donnés ci-dessous on associera un type catégoriel pour chacun de ces pronoms. Ces tableaux traiteront respectivement le cas des pronoms liés du singulier, du duel et du pluriel.

Les types catégoriels affectés seront sous la même forme donnée aux pronoms libres :  $N_{xyz}$ .

L'association de ces types catégoriels sera d'une grande importance dans la suite de notre travail au niveau de l'analyse de notre corpus. En effet, contrairement à la langue française, on peut trouver dans la langue arabe des phrases qui ne contiennent pas de sujet. C'est les pronoms liés qui joueront ce rôle.

Personne	Genre	Pronom	Exemple	Type catégoriel
1ère	Masculin	-TOU	ARAF-	N <sub>1S</sub>
	Féminin	تُ	TOU عَرَفْتُ	
2ème	Masculin	-TA	ARAF-TA	N <sub>2MS</sub>
	Féminin	-TI	ARAF-TI	N <sub>2FS</sub>
3ème	Masculin	-A	ARAF-A	N <sub>3MS</sub>
	Féminin	-AT	ARAF-AT	N <sub>3FS</sub>

tableau XI : Pronoms liés du singulier.

Personne	Genre	Pronom	Exemple	Type catégoriel
1 ère	Masculin	-NAA	ARAF-NAA	N <sub>1D</sub>
	Féminin	نا	عرفنا	
2 ème	Masculin	-	ARAF-TOUMAA	N <sub>2D</sub>
	Féminin	TOUMAA نُما	عرفنُما	
3 ème	Masculin	-AA ل	ARAF-AA عرفا	N <sub>3MD</sub>
	Féminin	-ATAA نا - ة	ARAF-ATAA عرفنا	N <sub>3FD</sub>

tableau XIII : Pronoms liés du Duel

Persone	Genre	Pronom	Exemple	Type catégoriel
1 ère	Masculin	- NAA	ARAF-NAA	N1S
	Féminin	خَا	عَرَفْنَا	
2 ème	Masculin	- TOM تُم	ARAF-TOM عَرَفْتُم	N2MS
	Féminin	-TONNA تُنْ	ARAF-TONNA عَرَفْتُنْ	N2FS
3 ème	Masculin	- UU وَا	ARAF-UU عَرَفُوا	N3MS
	Féminin	- NA نْ	ARAF- NA عَرَفْنَ	N3FS

tableau XIII : Pronoms liés pour le pluriel.

## 5.6 Travaux réalisés dans le traitement automatique de la langue arabe

De nos jours, on assiste à une forte demande d'un nouveau type d'applications informatiques à savoir les applications multilingues. L'arabe est la 7e langue la plus parlée au monde avec plus de 230 millions de locuteurs. Ce classement a imposé le

besoin de trouver des techniques permettant d'automatiser son traitement surtout avec l'apparition du Web. Cependant, la langue arabe est considérée comme étant une langue présentant plusieurs complexités. En effet, ses aspects morphologiques et syntaxiques sont très difficiles à adapter dans le cadre d'un système de traitement automatisé.

Plusieurs travaux sont recensés. Ils ont mené à des résultats encourageants. Les chercheurs ayant contribué à ces travaux ont pris des voies d'explorations multiples.

- Le traitement de la langue arabe écrite.
- Le traitement de la langue arabe parlée.

Chacune de ces voies d'exploration se ramifie à son tour en d'autres voies d'exploration selon la méthode utilisée.

Nous essayerons dans ce qui suit de présenter brièvement ces différentes voies d'exploration. Nous mettrons le point sur l'une d'entre elles à savoir l'utilisation des grammaires HPSG dans le cadre du traitement de la langue arabe.

#### **5.6.1 Traitement automatique de la langue arabe parlée**

Les progrès technologiques dans le domaine de traitement de signal constituent un atout majeur ayant permis de réaliser des progrès considérables dans le domaine de traitement de la langue arabe parlée.

La reconnaissance automatique de la parole dans la langue arabe est sans doute le principal axe de recherche dans ce domaine.

Le but étant d'interpréter le signal analogique généré par la parole dans le but de comprendre le message transmis. Cette interprétation se base sur des outils et des méthodes informatiques issus du domaine de traitement de signal ainsi que de l'intelligence artificielle [27].

Dans le cas de la langue arabe, cette reconnaissance s'avère plus complexe à cause de la présence de l'effet de coarticulation entre les éléments de la chaîne à analyser. Il est donc primordial de spécifier des méthodes efficaces permettant la détection des fins de mots [12].



En plus de la reconnaissance automatique de l'Arabe parlé, plusieurs autres aspects de cette langue ont été étudiés. L'aspect phonétique par exemple a intéressé beaucoup de chercheurs.

On peut citer par exemple le travail fait dans [23] ayant pour but d'affirmer l'hypothèse qui dit que la prosodie représente "un moyen puissant dans l'organisation linéaire et hiérarchique de nos expressions verbales" [23].

Il s'agit en d'autres termes de mettre en relief l'importance du rôle que joue la prosodie dans la construction et la structuration des phrases dans la langue arabe.

Pour démontrer son hypothèse, Mourad Mawhoub [23] s'est basé dans ses travaux sur une analyse phonétique d'une phrase arabe verbale non voyelle qu'il a présentée à un échantillon de 100 personnes. Cet échantillon est formé de plusieurs sujets dont les connaissances en arabe varient d'une personne à une autre. À chaque fois, une question portant sur l'un des éléments de cette phrase est posée à une personne et sa réponse est sauvegardée. Finalement, 5 phrases ont été recueillies, ces réponses ont été enregistrées dans les studios de la radio nationale de la ville de Marrakech. Une analyse phonétique de ces réponses a été réalisée avec le programme « winpitch ».

Cet analyseur en temps réel a permis de visualiser un ensemble de phrases traitées sous forme de graphes.

Sur la base des graphes obtenus l'auteur a tenté de localiser à chaque fois l'emplacement de l'élément accentué dans chaque phrase. Par la suite il a été montré que la prosodie joue un rôle très important dans l'organisation des mots dans la langue arabe standard [23].

## **5.6.2 Traitement automatique de la langue arabe écrite**

Nous nous intéresserons dans cette section à la partie la plus importante dans notre travail. On essayera de donner un panorama des différentes voies d'exploration en relation avec le domaine de traitement automatique de la langue arabe écrite.

### **5.6.2.1 La reconnaissance des caractères arabes imprimés**

À partir d'un texte imprimé, il est demandé dans la mesure du possible de reconnaître les différents caractères formant le texte. Cette reconnaissance se base sur différentes techniques d'imagerie. Une interprétation sera effectuée par la suite

soit par l'écriture des chaîne de caractères données en entrée ou en les prononçant. Malheureusement, la non-voyellation des textes arabes rend cette interprétation très ambiguë et complexe [18].

#### **5.6.2.2 La traduction automatique des textes arabes**

La compréhension et l'interprétation de la langue arabe constituent un défi majeur dans le domaine de traitement automatique de la langue arabe. Plusieurs travaux se sont intéressés à la traduction des textes arabes et ont permis la réalisation d'un grand nombre d'outils de traduction. Le traducteur le plus performant et le plus connu à ce jour est le logiciel Sakhr. Ce logiciel permet la traduction de textes de l'anglais à l'arabe et inversement. De plus, cet outil comprend une fonction de reconnaissance de la parole permettant de traduire un enregistrement vocal. Cette traduction peut être sous forme de texte écrit ou encore sous forme de message vocal. Sakhr offre, selon l'entreprise égyptienne qui l'a développé, des résultats corrects à 85%. Récemment, une version Sakhr mobile a vu le jour. Ces applications ont été développées pour iPhone et BlackBerry. Malheureusement, on voit que l'outil Sakhr est dédié principalement aux militaires, au département de justice Américaine,... et n'est pas accessible au public [5].

#### **5.6.2.3 L'analyse syntaxique des textes arabes**

Il s'agit d'étudier la syntaxe des textes en arabe dans le but de vérifier la bonne connexion syntaxique des énoncés. Cette analyse se base sur une multitude de modèles. Allant des grammaires génératives en passant par les grammaires d'unifications en arrivant jusqu'aux Grammaires catégorielles. On présentera dans ce qui suit un survol des principes des grammaires génératives et d'unification. Cet aperçu nous permettra de comprendre les divergences existant entre ces différents modèles et les grammaires catégorielles.

### **5.6.3 La Grammaire générative**

Cette théorie syntaxique est le fruit de recherches effectuées par Noam Chomsky en 1958 [8]. Elle se base sur un principe fondamental selon lequel : la génération des

phrases d'une langue donnée obéit à un ensemble de règles prédéfinies. Chomsky va jusqu'à affirmer l'existence d'un ensemble de règles communes entre toutes les règles [30]. La grammaire générative a été appliquée essentiellement à la langue anglaise.

#### **5.6.4 Les Grammaires d'unification**

Il s'agit d'une version évoluée de la grammaire générative, très critiquée à la fin des années 70. Ces critiques visaient plusieurs points des grammaires génératives, essentiellement « la place de la sémantique dans ce modèle,... » [1]. On se concentrera dans cette partie à la présentation d'un exemple de grammaires d'unifications ayant été utilisées dans le cadre du traitement automatique de la langue arabe écrite [4] : La Grammaire HPSG (Head Driven structure Grammaire).

Le modèle HPSG est inspiré de GPSG (Generalised phrase structure grammar), des travaux réalisés par C. Pollard sur les Grammaires guidées par les têtes, de tous les modèles antérieurs des grammaires d'unification (LFG...) et même des grammaires catégorielles. Ce modèle a été développé au début des années 80. Le principal but de sa mise en œuvre est de « permettre une intégration plus explicite des différents niveaux d'analyse linguistique : phonétique, syntaxe et sémantique » [1].

HPSG peut être considéré comme l'un des modèles les plus révolutionnaires puisqu'il est basé sur des travaux en logique et en informatique (héritage et typage). Plusieurs versions de HPSG se sont succédées pour améliorer ce modèle en le rendant plus simple comparé à son ancêtre GPSG. En effet, ce modèle réduit le nombre de règles syntaxiques grâce au grand nombre d'informations données dans la représentation des unités lexicales. Pour chaque entité lexicale, on donne grâce à HPSG, un ensemble d'informations (que ce soit d'ordre syntaxique, phonétique ou encore d'ordre sémantique). Cette représentation du lexique est faite à l'aide d'une seule structure de données : « les structures de traits » qui sont aussi utilisées pour la représentation de la sémantique [1].

Parmi les avantages que présente ce modèle, il y a la possibilité de son extension. En effet, HPSG est extensible et peut traiter plusieurs langues comme le cas de la langue arabe [4].

Parmi les exemples d'application du modèle HPSG sur la langue arabe on retrouve l'analyseur syntaxique SYNTAXE (analyseur syntaxique de la langue arabe non-voyellée).

L'évaluation de SYNTAXE a donné des résultats encourageants (81% des phrases traitées ont été correctement analysées). Le seul problème avec ce système c'est qu'il ne traite pas quelques types de phrases.

En effet, le modèle HPSG n'offre pas un schéma permettant de représenter les phrases nominales arabes. C'est la raison qui a poussé le groupe de chercheurs qui l'a développé à proposer un nouveau schéma qu'ils ont appelé Head Attribut Schéma.

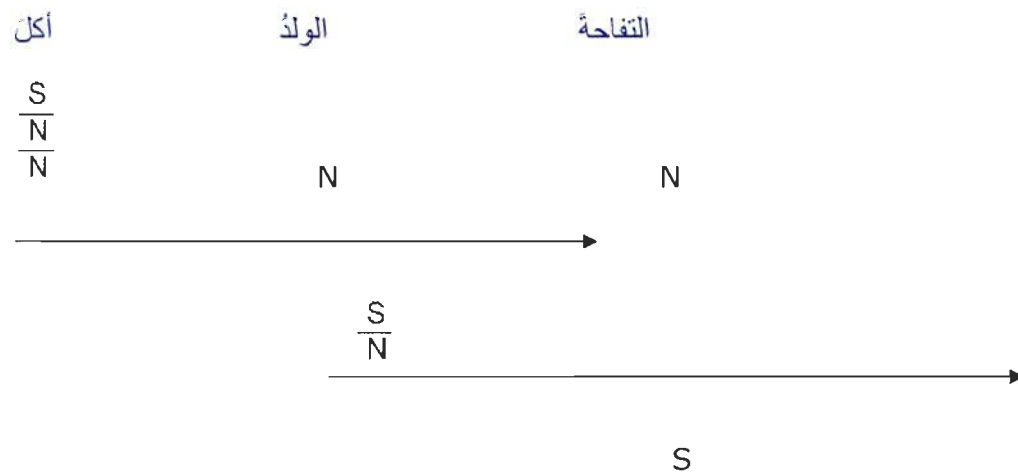
### 5.7 Application des GC pour l'analyse de textes en arabe

On essayera dans cette partie d'analyser un ensemble de phrases en arabe. Cette analyse se basera sur des modèles de grammaires catégorielles présentés dans le chapitre 1.

(i) Application du modèle de Ajdukiewicz :

Voyons comment on peut appliquer la grammaire catégorielle de Adjukiewicz pour l'analyse de la phrase : أكل الولد التفاحة : « l'enfant a mangé la pomme ».

L'opérateur principal désigné dans cette phrase est le verbe. Cette analyse se fait comme suit :



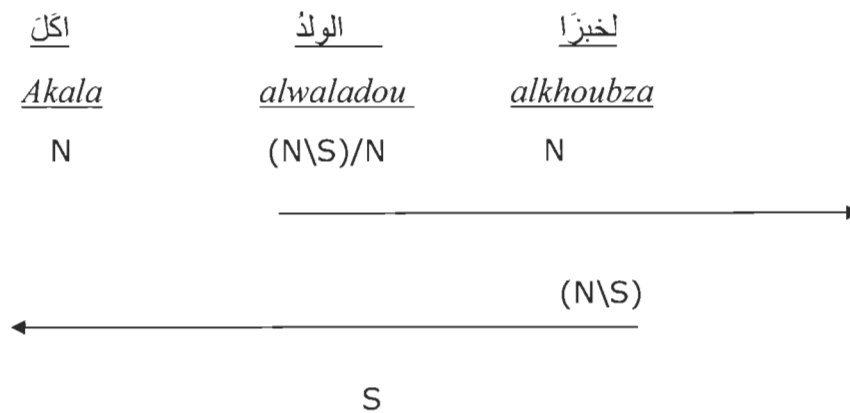
(ii) Application du modèle de la grammaire AB pour la langue arabe :

Voyons maintenant comment on peut appliquer cela sur un exemple de phrase arabe :

أكل الولد الخبز Akala alwaladou alkhoubza

Traduction : « *L'enfant a mangé le pain.* »

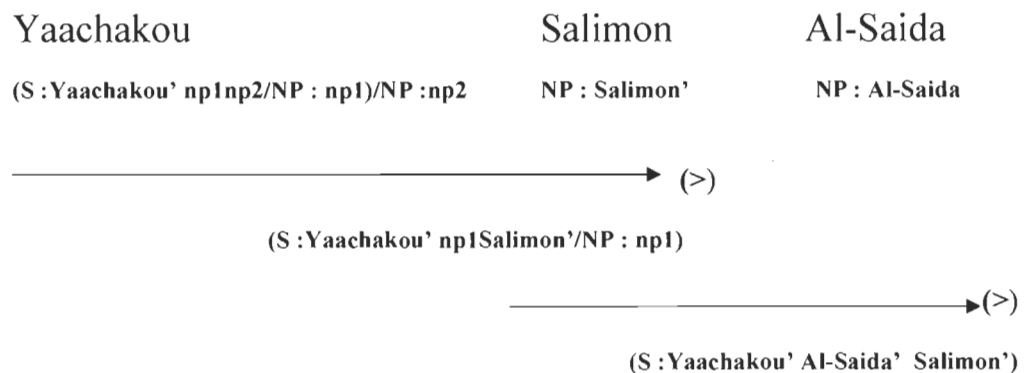
On va assigner les types suivants aux différents éléments de la phrase :



(iii) Application du modèle de la grammaire catégorielle combinatoire de Steedman :



On verra dans ce qui suit comment se fait l'analyse d'une phrase arabe suivant la méthode donnée par Steedman.



(iv) Application du modèle de la GCCA

Essayons maintenant d'appliquer le modèle de la GCCA sur un exemple en arabe.  
Cette analyse portera sur les deux versions verbale et nominale d'une même phrase.

Version nominale :

الغاية تُبرّر الوسيلة

الغاية	الوسيلة	تُبرّر
(ALGHAYATOU	TOUBARRIROU	ALWASILTA)
La-fin	justifie	les-moyens

**Phénotype :**

[N : ALGHAYATOU]-[(S/N)/N : TOUBARRIROU]-[N : ALWASILATA]  
 [S/(S/N) : (C\*ALGHAYATOU)]-[(S/N)/N : TOUBARRIROU]-[N : ALWASILATA] (>Tx)  
 [S/N : (B(C\*ALGHAYATOU) TOUBARRIROU)]-[N : ALWASILATA] (>B)  
 [S : (B(C\*ALGHAYATOU) TOUBARRIROU) ALWASILATA] ( > )

**Génotype :**

(B(C\*ALGHAYATOU) TOUBARRIROU) ALWASILATA  
 ( (C\*ALGHAYATOU) (TOUBARRIROU ALWASILATA) ) (B)  
 ( (TOUBARRIROU ALWASILATA) ALGHAYATOU ) (C\*)

Version Verbale :

Avec les mêmes affectations de types que ci-dessus, montrons que la phrase :  
 « تُبرّر الغاية الوسيلة » est syntaxiquement correcte.

تُبرّر	الغاية	الوسيلة
( TOUBARRIROU	ALGHAYATOU	ALWASILTA)
Justifie	La-fin	les-moyens

**Phénotype:**

[(S/N)/N : TOUBARRIROU]-[N : ALGHAYATOU] -[N : ALWASILATA]  
 [(S/N)/N : TOUBARRIROU]-[S/(S/N) : (C\*ALGHAYATOU)]-[N : ALWASILATA] (<T)  
 [S/N : (B (C\*ALGHAYATOU)TOUBARRIROU)]-[N : ALGHAYATA] (<Bx)  
 [S : ((B (C\*ALGHAYATOU)TOUBARRIROU) ALGHAYATA)] ( < )

**Génotype :**

S : ((B (C\*ALGHAYATOU)TOUBARRIROU) ALGHAYATA)  
 (C\*ALGHAYATOU)(TOUBARRIROU ALGHAYATA) ( B )  
 (TOUBARRIROU ALGHAYATA) ALGHAYATOU ( C\* )

On remarque que l'analyse des deux versions nominale et verbale de la phrase :  
تُبَرَّرُ الغاية الوسيلة nous ramène à un même résultat. On a montré que chacune de ces  
deux phrases est syntaxiquement correcte. De plus, on arrive à la même forme  
normale.

### **5.8 Conclusion**

L'étude de quelques principes de base de l'arabe montre la complexité de cette  
langue. Nous consacrerons le chapitre suivant à l'analyser d'un ensemble de phrases  
suivant le modèle de la GCCA ce qui représente un grand défi à relever. Nous  
allons diversifier le choix de nos phrases dans le but de vérifier l'efficacité du  
modèle GCCA.

## CHAPITRE 6 :

### ÉTUDE THEORIQUE : ANALYSE DU CORPUS

#### 6.1 Introduction

Nous tenterons dans ce chapitre d'appliquer le modèle de la grammaire catégorielle à un ensemble de phrases de la langue arabe. Pour ce fait, nous nous baserons sur un corpus comportant un ensemble de phrases arabe non voyelle. L'analyse de ce corpus sera accompagnée par une explication d'un ensemble de métarègle.

Nous analyserons les versions VSO et VOS de quelques-unes de ces phrases dans le but de démontrer l'efficacité du modèle de la GCCA dans l'analyse de phrases de l'arabe sous leurs deux versions. Nous traiterons aussi le cas de la coordination en donnant la solution permettant d'y remédier.

Nous reprenons dans cette partie l'exemple donné dans le chapitre précédent dans le but de montrer l'application de notre première métarègle :

« M1 : Pour une séquence concaténée u1-u2, si u1 est de type N et u2 de type (Y/N)/Z, alors nous appliquons le changement de type (>Tx) à u1 :

$[N : u1 \implies Y/(Y/N) : (C^* u1)] [7].$  »

Dans notre cas u1 et u2 sont représentés respectivement par « ALGHAYATOU » et « TOUBARRIROU ».

الغاية تُبرِّرُ الوسيلة

لغاية

تُبرِّرُ

الوسيلة

(ALGHAYATOU  
La-fin

TOUBARRIROU  
justifie

ALWASILTA)  
les moyens

Phénotype :

$[N : ALGHAYATOU] - [(S/N)/N : TOUBARRIROU] - [N : ALWASILATA]$

$[S/(S/N) : (C^*ALGHAYATOU)] - [(S/N)/N : TOUBARRIROU] - [N : ALWASILATA] \quad (>Tx) M1$

$[S/N : (B(C^*ALGHAYATOU) TOUBARRIROU)] - [N : ALWASILATA] \quad (>B)$



[S : (B(C\*ALGHAYATOU) TOUBARRIROU) ALWASILATA] ( > )

Génotype:

(B(C\*ALGHAYATOU) TOUBARRIROU) ALWASILATA  
 ( (C\*ALGHAYATOU) (TOUBARRIROU ALWASILATA) ) (B)  
 ( (TOUBARRIROU ALWASILATA) ALGHAYATOU ) (C\*)  
 TOUBARRIROU ALWASILATA ALGHAYATOU

Voyons maintenant ce qui se passe si on passe à la version verbale de cette phrase.  
 En d'autres termes, essayons de montrer qu'avec les mêmes affectations de types on arrive à démontrer que la phrase verbale « تُبرَّرُ الوسيلة لغايةً » « TOUBARRIROU ALGHAYATOU ALWASILTA » est syntaxiquement correcte.  
 L'analyse de cette phrase nous permettra d'introduire une nouvelle métarègle permettant de traiter le cas où on est en présence d'une structures concaténée u1-u2 de types respectifs (S/N)/N et N.

« M2 : Pour une séquence concaténée u1-u2, si u1 est de type (Y/N)/Z et u2 de type N, alors nous appliquons le changement de type (<T) à u2 :

[N : u2 ==> Y\ (Y/N) : (C\* u2)] » [7]

تُبرَّرُ

الوسيلة

لغايةً

( TOUBARRIROU ALGHAYATOU ALWASILTA )  
 justifie La fin les moyens

Phénotype :

[(S/N)/N : TOUBARRIROU]-[N : ALGHAYATOU] -[N : ALWASILATA]  
 [(S/N)/N : TOUBARRIROU]-[S\ (S/N): (C\*ALGHAYATOU)]-[N : ALWASILATA] (<T) M1  
 [S/N: (B (C\*ALGHAYATOU)TOUBARRIROU)]-[N: ALGHAYATA] (<Bx)  
 [S : ((B (C\*ALGHAYATOU)TOUBARRIROU) ALGHAYATA)] (< )

Génotype

S : ((B (C\*ALGHAYATOU)TOUBARRIROU) ALGHAYATA)  
 (C\*ALGHAYATOU)(TOUBARRIROU ALGHAYATA) (B)  
 (TOUBARRIROU ALGHAYATA) ALGHAYATOU (C\*)

TOUBARRIROU ALGHAYATA ALGHAYATOU

On voit qu'avec les mêmes affectations on a eu la même structure applicative et le même résultat S.

Voyons maintenant comment se fait l'analyse de la phrase française correspondante :

*La fin justifie les-moyens*

Phénotype :

*La fin justifie les moyens*  
 [N/N:La]-[N : fin]-[(S\N)/N : justifie]-[N/N : les] – [N : Moyens]  
 [N: (La fin)]-[(S\N)/N : justifie]-[N/N : les] – [N : Moyens] ( > )  
 [S/(S\N): (C\*(La fin))]-[(S\N)/N : justifie]-[N/N : les] – [N : Moyens] (>T)  
 [S/N: ( B (C\*(La fin) )justifie)]-[N/N : les] – [N : Moyens] (>B)  
 [S/N: (B( B (C\*(La fin) justifie) les)] – [N : Moyens]  
 (>B)  
 [S: (B( B (C\*(La fin) justifie) les) Moyens)] ( > )

Génotype :

(B( B (C\*(La fin) justifie) les) Moyens)  
 ( B (C\*(La fin)( justifie (les Moyens)) (B)  
 (C\* (La fin) ( justifie (les Moyens)) ) (B)  
 ((justifie (les Moyens)) (La fin) ) (C\*)  
*justifie les Moyens La fin*

Appliquons le même procédé à la phrase verbale « مَالِكٌ أَكَلَ الْخُبْزَ ».

أَكَلَ	الْخُبْزَ	مَالِكٌ
(AKALA	MALIKON	ALKHOBZA)
Mange	Malék	le-pain

Phénotype:

[(S/N)/N : AKALA]-[N : MALIKON] -[N : ALKHOBZA]

$[(S/N)/N : AKALA] - [S/(S/N) : (C^*Malikon)] - [N : ALKHOBZA]$  ( $<T$ )  
 $[S/N : (B (C^* Malikon) AKALA)] - [N : ALKHOBZA]$  ( $<Bx$ )  
 $[S : ((B (C^* Malikon) AKALA) ALKHOBZA)]$  ( $<$ )

## Génotype

$((B (C^* Malikon) AKALA) ALKHOBZA)$   
 $((C^* Malikon) (AKALA ALKHOBZA))$  (B)  
 $(AKALA ALKHOBZA) Malikon$  ( $C^*$ )  
 $AKALA ALKHOBZA Malikon$

L'analyse de la phrase ci-dessous nous permettra d'introduire la troisième métarègle selon laquelle [7]

« M3 : Pour une séquence concaténée  $u_1-u_2$ , si  $u_1$  est de type N et  $u_2$  de type  $(Y/N)/Z$ , alors nous appliquons le changement de type ( $>Tx$ ) à  $u_1 : [N : u_1]$   
 $\implies Y/(Y/N) : (C^* u_1)$  »

مالك	أكل	الخبز
(MALIKON	AKALA	ALKHOBZA)
Malék	Mange	le-pain

$[N : MALIKON] - [(S/N)/N : AKALA] - [N : ALKHOBZA]$   
 $[S/(S/N) : (C^* MALIKON)] - [(S/N)/N : AKALA] - [N : ALKHOBZA]$  ( $>Tx$ )  
 $[S/N : (B(C^* MALIKON) AKALA)] - [N : ALKHOBZA]$  ( $>B$ )  
 $[S : (B(C^* MALIKON) AKALA) ALKHOBZA]$  ( $>$ )

## Génotype :

$(B(C^* MALIKON) AKALA) ALKHOBZA$   
 $((C^* MALIKON) (AKALA ALKHOBZA))$  (B)  
 $((AKALA ALKHOBZA) MALIKON)$  ( $C^*$ )  
 $AKALA ALKHOBZA MALIKON$

Analyse de la version en langue française:

Malek mange le pain

## Phénotype :

$[N : Malek] - [(S \backslash N) / N : mange] - [N / N : le] - [N : pain]$   
 $[S / (S \backslash N) : (C^*(Malek))] - [(S \backslash N) / N : mange] - [N / N : le] - [N : pain] \quad (>T)$   
 $[S / N : (B (C^*(Malek)mange))] - [N / N : le] - [N : pain] \quad (>B)$   
 $[S / N : (B (B (C^*(Malek) mange) le))] - [N : pain] \quad (>B)$   
 $[S : (B (B (C^*(Malek) mange) le) pain)] \quad (>)$

Génotype :

$(B (B (C^*(Malek) mange) le) pain)$   
 $(B (C^*(Malek)(mange (le pain))) \quad (B)$   
 $(C^*(Malek) (mange (le pain))) \quad (B)$   
 $((mange (le pain)) (Malek)) \quad (C^*)$   
 mange le pain Malek

Voyons maintenant un autre exemple de phrases arabe, où le sujet est 'absent'. Ce dernier est référencé à la fin du verbe par un pronom lié.

Essayons d'analyser la phrase verbale : الباب فتحتُ (J'ai ouvert la porte).

Dans cette phrase, le sujet est référé par un pronom relatif (تُ), (TOU) lié au verbe (فتح) (FATAHA, ouvre).

L'analyse de cette phrase se fait comme suit :

الباب فتحتُ  
 فتح تُ الباب  
 (FATAH TOU ALBÉBA)  
 Ouvert je la-porte

FATAH TOU ALBÉBA

N<sub>IMS</sub> fait référence à la première personne masculin singulier.

N° fait référence à N Objets.

Ns fait référence au sujet : N<sub>IMS</sub> dans notre cas.

Phénotype:

$[(S / N_s) / N^\circ : FATAH] - [N_{IMS} : TOU] - [N^\circ : ALBÉBA]$   
 $[(S / N_s) / N^\circ : FATAH] - [S \backslash (S / N_{IMS}) : (C^* TOU)] - [N^\circ : ALBÉBA] \quad (<T) M1$

[S/N°: (B (C\* *TOU*) *FATAH*)]-[N°: *ALBÉBA*] (<Bx)

[S : ((B (C\* *TOU*) *FATAH*) *ALBÉBA*)] ( < )

### Génotype

((B (C\* *TOU*) *FATAH*) *ALBÉBA*)

(( (C\* *TOU*) (*FATAH* *ALBÉBA*) ) ( B )

(*FATAH* *ALBÉBA*) *TOU* ( C\* )

*FATAH* *ALBÉBA* *TOU*

La version nominale dans ce cas pose problème. En effet, on ne peut pas passer à la version SOV en gardant les mêmes constituants de la phrase puisque le sujet est absent. Par contre, on peut avancer le complément d'objet direct (الباب, *la porte*) et le mettre au début ce qui donne la phrase (فتحتُ الباب, *la porte j'ai ouvert*) de type OVS. Ce genre de phrase est rarement utilisé en arabe, il exprime un procédé de style permettant d'insister sur le complément d'objet.

Nous essayerons dans ce qui suit d'analyser un exemple de phrase OVS.

فتحتُ الباب.

الباب

فتح

ت

(*ALBÉBA*

*FATAH*

*TOU*)

*la-porte*

*Ouvert*

*je*

*ALBÉBA*

*FATAH*

*TOU*

Phénotype :

[N° : *ALBÉBA*]-[(S/Ns)/N° : *FATAH*] - [N1MS : *TOU*]

[S/ (S/N°):(C\*( *ALBÉBA*))]-[ (S/Ns)/N°: *FATAH*]-[N1MS:*TOU*] (>Tx)

[S/Ns: ( B (C\*( *ALBÉBA*) *FATAH*)]- [N1MS : *TOU*] (>B)

[S: ( ( B (C\*( *ALBÉBA*) *FATAH*)) *TOU*)] (>)

Génotype :

(( B (C\*( *ALBÉBA*) *FATAH*)) *TOU*)

( B (C\*( *ALBÉBA*)( *FATAH* (*TOU*))) (B)

((FATAH (TOU)) (ALBEBBA) ) (C\*)  
 FATAH TOU ALBEBBA

On remarque dans cet exemple qu'on n'arrive pas à la même forme normale que précédemment. Sauf qu'on arrive à montrer que la phrase est syntaxiquement correcte.

On s'intéressera dans notre mémoire à l'étude des versions VOS et VSO des phrases dans la langue arabe. On évitera donc de s'attarder sur les phrases OVS.

Voyons le cas de la phrase (نظرن في الأمر) (elles étudient le sujet)

نَظَرْنَ	نَ	فِي	الأمر
(NADHAR	na	FI	ALAMRI)
Étudient	elles	le	sujet

Phénotype:

[(S/Ns)/N° : NADHAR]-[N<sub>2PF</sub> : NA] -[N°/N : FI]- [N : ALAAMRI]  
 [(S/Ns)/N° : NADHAR]-[S\ (S/ N<sub>2PF</sub>): (C\* NA)]- [N°/N : FI]- [N : ALAAMRI] (<T)  
 [S/N°: (B (C\* NA) (NADHAR))]- [N°/N : FI]- [N: ALAAMRI] (<Bx)  
 [S/N°s: (B(B (C\* NA) (NADHAR)) FI)]- [N: ALAAMRI] (<B)  
 [S : (B(B (C\* NA) (NADHAR) FI ALAAMRI))]] (<)

Génotype

B B (C\* NA) (NADHAR) FI ALAAMRI  
 B (C\* NA) (NADHAR) (FI ALAAMRI) (B)  
 (C\* NA) (NADHAR FI ALAAMRI) (B)  
 (NADHAR FI ALAAMRI) (NA) (C\*)  
 NADHAR FI ALAAMRI NA

Pour l'analyse de la version nominale, on assiste ici au même problème présenté ci-dessus à savoir l'analyse d'une phrase OVS.

فِي	نَ	نظرن	الأمر
( FI	ALAMRI	NADHAR	NA )
le	sujet	Étudient	elles

Phénotype :

[N°/N : FI] - [N : ALAAMRI] - [(S/N)/N : NADHAR] - [N2PF : NA]  
 [N° : FI ALAAMRI] - [(S/Ns)/N° : NADHAR] - [N2PF : NA] ( > )  
 [S/(S/N°): C\* (FI ALAAMRI) ] [(S/Ns)/No : NADHAR] - [N2PF : NA] (Tx )  
 [S/Ns : B(C\* (FI ALAAMRI) NADHAR) ] - [N2PF : NA] ( >B )  
 [S : B(C\* (FI ALAAMRI) NADHAR NA)] ( > )

Genotype:

B(C\* (FI ALAAMRI) NADHAR NA)  
 (C\* (FI ALAAMRI) NADHAR NA) (B)  
 NADHAR NA (FI ALAAMRI) (C\*)

### حَبَّتِ السُّحُبُ السَّمَاءَ

حَبَّتِ	السُّحُبُ	السَّمَاءَ
(HAJABAT	ALSOUHOUBOU	ALSAMAAA)
Couvrent	les-nuages	le-ciel

L'analyse de la version verbale de cette phrase est donnée comme suit :

Phénotype:

[(S/N)/N : HAJABAT]-[N : ALSOUHOUBOU ]-[N : ALSAMAAA]  
 [(S/N)/N : HAJABAT]-[S(S/N): (C\* ALSOUHOUBOU )]-[N : ALSAMAAA] (<T)  
 [S/N: (B (C\* ALSOUHOUBOU ) HAJABAT)]-[N: ALSAMAAA] (<Bx)  
 [S : ((B (C\* ALSOUHOUBOU ) HAJABAT) ALSAMAAA)] ( < )

Génotype

((B (C\* ALSOUHOUBOU ) HAJABAT) ALSAMAAA)  
 ( (C\* ALSOUHOUBOU ) (HAJABAT ALSAMAAA) ) ( B )  
 (HAJABAT ALSAMAAA) ALSOUHOUBOU ( C\* )

HAJABAT ALSAMAAA ALSOUHOUBOU

Voyons maintenant le cas de l'analyse de la version nominale de cette même phrase :

السُّحُبُ	حَجَبَتْ	الْأَسْمَاءُ
( ALSOUHOUBOU	HAJABAT	ALSAMAAA)
les-nuages	Couvrent	le-ciel

Phénotype :

[N : ALSOUHOUBOU ] - [(S/N)/N : HAJABAT] - [N : ALSAMAAA]  
 [S/(S/N) : (C\* ALSOUHOUBOU )]-[(S/N)/N : HAJABAT]-[N : ALSAMAAA] (>Tx)  
 [S/N : (B(C\* ALSOUHOUBOU ) HAJABAT)]-[N : ALSAMAAA] (>B)  
 [S : (B(C\* ALSOUHOUBOU ) HAJABAT) ALSAMAAA] (>)

Génotype :

(B(C\* ALSOUHOUBOU ) HAJABAT) ALSAMAAA  
 ((C\* ALSOUHOUBOU ) (HAJABAT ALSAMAAA) ) (B)  
 ((HAJABAT ALSAMAAA) ALSOUHOUBOU ) (C\*)  
 HAJABAT ALSAMAAA ALSOUHOUBOU

Analyse de la version traduite de cette phrase en langue française:

حَجَبَتْ	السُّحُبُ	الْأَسْمَاءُ
( ALSOUHOUBOU	HAJABAT	ALSAMAAA)
les-nuages	Couvrent	le-ciel

[N/N:Les]-[N : nuages]-[(S\N)/N : couvrent]-[N/N : le] – [N : ciel]  
 [N : (Les nuages)]-[(S\N)/N : couvrent]-[N/N : le] – [N : ciel] (> )  
 [S/(S\N): (C\*( Les nuages))]-[(S\N)/N : couvrent]-[N/N : le] – [N : ciel] (>T)  
 [S/N: ( B (C\*( Les nuages) couvrent)]-[N/N : le] – [N : ciel] (>B)  
 [S/N: (B ( B (C\*( Les nuages) couvrent) le)] – [N : ciel] (>B)  
 [S: (B( B (C\*( Les nuages) couvrent) le) ciel)] (>)

Génotype :

(B( B (C\*( Les nuages) couvrent) le) ciel)



( B (C\*( *Les nuages*)( *couvrent (le ciel)*) (B)  
 (C\* ( *Les nuages*) ( *couvrent (le ciel)*) ) (B)  
 (( *couvrent (le ciel)*) ( *Les nuages*) ) (C\*)  
*couvrent le ciel Les nuages*

جلستُ أمامَ المدفأةِ

جَلَسْتُ

المدفأةِ

أمامَ

ثُ

(JALAS TOU AMAMA AL-MEDFAAATI)

M'assois je devant la-cheminée

Phenotype:

[(S/N)/N : JALAS]-[N1S : TOU] -[N°/N : AMAMA]- [N : AL-MEDFAAATI]

[(S/N)/N : JALAS]-[S(S/ N1S): (C\* TOU)]- [N°/N : AMAMA]- [N : AL-MEDFAAATI] (<T)

[S/N: (B (C\* TOU) (JALAS)]- [N°/N : AMAMA]- [N: AL-MEDFAAATI] (<Bx)

[S/N°s: (B(B (C\* TOU) (JALAS)) AMAMA)]- [N: AL-MEDFAAATI] (<B)

[S : (B(B (C\* TOU) (JALAS) AMAMA AL-MEDFAAATI))]] (<)

Génotype

B B (C\* TOU) (JALAS) AMAMA AL-MEDFAAATI

B (C\* TOU) (JALAS) (AMAMA AL-MEDFAAATI) (B)

(C\* TOU) (JALAS AMAMA AL-MEDFAAATI) (B)

(JALAS AMAMA AL-MEDFAAATI) (TOU) (C\*)

JALAS AMAMA AL-MEDFAAATI TOU

Analyse de la version Nominale:

تتساقط أوراق الأشجار بسرعة

تَتَساقَطُ

أوراق

الأشجار

بسرعة

tombent

les-feuilles

des-arbres

vite

Phénotype :

[(S/N)/N : TATASAKATOU]-[N : AWRAKOU]-[N : ALACHJAR]-[(S/N)(S/N) : BISO RATIN]

[(S/N)/N : TATASAKATOU]-[S(S/N):(C\* AWRAKOU)]-[N : ALACHJAR]-[(S/N)(S/N) : BISO RATIN] (<T)

[S/N : (B (C\* *AWRAKOU*) *TATASAKATOU*)]-[N : *ALACHJARI*]- [(S/N)\(S/N) : *BISORATIN*] (<Bx)

[S : ((B (C\* *AWRAKOU*) *TATASAKATOU*) *ALACHJARI*)]-[(S/N)\(S/N) : *BISORATIN*] (>)

[S/(S/N) : (C\**AWRAKOU*)]-[S/N : (*TATASAKATOU* *ALACHJARI*)]-[(S/N)\(S/N) : *BISORATIN*]

[S/(S/N) : (C\**AWRAKOU*)]-[S/N : (*BISORATIN* (*TATASAKATOU* *ALACHJARI*)))] (<)

[S : ((C\**AWRAKOU*) (*BISORATIN* (*TATASAKATOU* *ALACHJARI*)))]

Genotype:

((C\**AWRAKOU*) (*BISORATIN* (*TATASAKATOU* *ALACHJARI*)))

((*BISORATIN* (*TATASAKATOU* *ALACHJARI*)) (*AWRAKOU*)) (C\*)

أمام	المِدْفَاءُ	جَلَسْتُ	تُ
أمام	المِدْفَاءُ	جَلَسْتُ	تُ
(AMAMA	AL-MEDFAAATI	JALAS	TOU
Devant	la-cheminée	M'assois	je

Phénotype

[N°/N : *AMAMA*]- [N : *AL-MEDFAAATI*] -[(S/N)/N : *JALAS*]-[N|S : *TOU*]

[N°: *AMAMA* *AL-MEDFAAATI*] -[(S/N)/N : *JALAS*]-[N|S : *TOU*] (>)

[S/(S/N°) : (C\* *AMAMA* *AL-MEDFAAATI*)]-[(S/N)/N : *JALAS*]-[N : *TOU*] (>Tx)

[S/N : (B(C\* *AMAMA* *AL-MEDFAAATI*) *JALAS*)]-[N : *TOU*] (>B)

[S : (B(C\* *AMAMA* *AL-MEDFAAATI* ) *JALAS*) *TOU*] (>)

Genotype :

(B(C\* *AMAMA* *AL-MEDFAAATI* ) *JALAS*) *TOU*

(C\* *AMAMA* *AL-MEDFAAATI* ) (*JALAS* *TOU*) ) (B)

(*JALAS* *TOU*) (*AMAMA* *AL-MEDFAAATI*) (C\*)

*JALAS* *TOU* *AMAMA* *AL-MEDFAAATI*

Phénotype:

B B (C\* *TOU*) (*JALAS*) *AMAMA* *AL-MEDFAAATI*

B (C\* *TOU*) (*JALAS*) (*AMAMA* *AL-MEDFAAATI*) (B)

(C\* *TOU*) (*JALAS* *AMAMA* *AL-MEDFAAATI*) (B)

(JALAS AMAMA AL-MEDFAAATI) (TOU) (C\*)

JALAS AMAMA AL-MEDFAAATI TOU

On voit que le résultat concorde avec l'analyse de la version verbale de cette phrase.

وصل المسافرين إلى المرفأ

وصل	المسافرين	إلى	المرفأ
WASALA	ALMOUSAFIROU	ILA	ALMARFAII
Arrive	le-voyageur	au	port

Phenotype:

[(S/N)/N : WASALA]-[N: ALMOUSAFIROU] -[N/N : ILA]- [N : ALMARFAII]

[(S/N)/N : WASALA]-[S(S/ N): (C\* ALMOUSAFIROU)]- [N/N : ILA]-

[N : ALMARFAII] (<T)

[S/N: (B (C\* ALMOUSAFIROU) (WASALA))- [N/N : ILA]- [N: ALMARFAII] (<Bx)

[S/N: (B(B (C\* ALMOUSAFIROU) (WASALA)) ILA))- [N: ALMARFAII] (<B)

[S : (B(B (C\* ALMOUSAFIROU) (WASALA) ILA ALMARFAII)))] (< )

Génotype

B B (C\* ALMOUSAFIROU) (WASALA) ILA ALMARFAII

B (C\* ALMOUSAFIROU) (WASALA) (ILA ALMARFAII) ( B )

(C\* ALMOUSAFIROU) (WASALA ILA ALMARFAII) ( B )

(WASALA ILA AL- ALMARFAII) (ALMOUSAFIROU) ( C\* )

اشتقت إلى وطني

اشتق	ت	إلى	وطني
(ÉCHTAK	TOU	ILA	WATANNI)
manque	me		mon-pays

Phénotype:

[(S/N)/N : ÉCHTAK]-[NIP : TOU] -[N/N : ILA]- [N : WATANNI]

[(S/N)/N : ÉCHTAK]-[S(S/ NIP): (C\* TOU)]- [N/N : ILA]- [N : WATANNI] (<T)

[S/N: (B (C\* TOU) (ÉCHTAK))- [N/N : ILA]- [N: WATANNI] (<Bx)

[S/N°s: (B(B (C\* TOU) (ÉCHTAK)) ILA))- [N: WATANNI] (<B)

[S : (B(B (C\* TOU) (ÉCHTAK) ILA WATANNI)))] (< )

Génotype

B B (C\* TOU) (ÉCHTAK) ILA WATANNI  
 B (C\* TOU) (ÉCHTAK) (ILA WATANNI) (B)  
 (C\* TOU) (ÉCHTAK ILA WATANNI) (B)  
 (ÉCHTAK ILA WATANNI) (TOU) (C\*)  
 ÉCHTAK ILA WATANNI TOU

عَادَ الصَّقْرُ النَّبِيلُ إِلَى عَشِّهِ

عَادَ	الصَّقْرُ	النَّبِيلُ	إِلَى	عَشِّهِ
AADA	ALSAKRO	ALNABILOU	ILA	OCHIH
Retourne	le-faucon	prestigieux	à	son-nid

Analyse:

Phénotype:

[(S/N)/N : AADA]-[N : ALSAKROU]-[N : ALNABILOU]-[(S/N)S : ILA]-[N : OCHIH]  
 [(S/N)/N : AADA]-[S(S/N): (C\*ALSAKROU)]-[N : ALNABILOU]-[(S/N)S : ILA]-[N : OCHIH] (>T)  
 [S/N : (B (C\*ALSAKROU)AADA)]-[N : ALNABILOU]-[(S/N)S : ILA]-[N : OCHIH] (<Bx)  
 [S : (B (C\*ALSAKROU)AADA) ALNABILOU]-[(S/N)S : ILA]-[N : OCHIH] (>)  
 [S/N : ILA (B (C\*ALSAKROU)AADA) ALNABILOU]-[N : OCHIH] (<)  
 [S : ILA ((B (C\*ALSAKROU)AADA) ALNABILOU) OCHIH] (>)

Génotype :

ILA ((B (C\*ALSAKROU)AADA) ALNABILOU) OCHIH  
 ILA ((C\*ALSAKROU)(AADA ALNABILOU) OCHIH) (B)  
 ILA ((AADA ALNABILOU) OCHIH) (ALSAKROU) (C\*)

عَرَفَ الطِّفْلُ أُمَّهُ

عَرَفَ	الطِّفْلُ	أُمَّهُ
(ARAF	ALTÉFLOU	OMMAHOU)
a-Reconnu	le-garçon	sa - mère

Phénotype :

$[(S/N)/N : ARAFA] - [N : ALTÉFLOU] - [N : OMMAHOU]$   
 $[(S/N)/N : ARAFA] - [S(S/N) : (C^* ALTÉFLOU)] - [N : OMMAHOU]$  ( $<T$ )  
 $[S/N : (B(C^* ALTÉFLOU) ARAFA)] - [N : OMMAHOU]$  ( $<Bx$ )  
 $[S : ((B(C^* ALTÉFLOU) ARAFA) OMMAHOU)]$  ( $<$ )  
 Génotype

$((B(C^* ALTÉFLOU) ARAFA) OMMAHOU)$   
 $(C^* ALTÉFLOU) (ARAFA OMMAHOU)$  ( $B$ )  
 $(ARAFA OMMAHOU) ALTÉFLOU$  ( $C^*$ )  
 $ARAFA OMMAHOU ALTÉFLOU$

أُمُّهُ	عَرَفَ	الطِّفْلُ
(ALTÉFLOU	ARAFA	OMMAHOU)
le-garçon	Reconnais	sa - mère

Phénotype :

$[N : ALTÉFLOU] - [(S/N)/N : ARAFA] - [N : OMMAHOU]$   
 $[S/(S/N) : (C^* ALTÉFLOU)] - [(S/N)/N : ARAFA] - [N : OMMAHOU]$  ( $>Tx$ )  
 $[S/N : (B(C^* ALTÉFLOU) ARAFA)] - [N : OMMAHOU]$  ( $>B$ )  
 $[S : (B(C^* ALTÉFLOU) ARAFA) OMMAHOU]$  ( $>$ )

Génotype:

$(B(C^* ALTÉFLOU) ARAFA) OMMAHOU$   
 $(C^* ALTÉFLOU) (ARAFA OMMAHOU)$  ( $B$ )  
 $((ARAFA OMMAHOU) ALTÉFLOU)$  ( $C^*$ )  
 $ARAFA OMMAHOU ALTÉFLOU$

## Chapitre 7 : CONCLUSION

Le but de notre mémoire est d'appliquer la Grammaire catégorielle combinatoire applicative à la langue arabe. Il s'agit en d'autres termes de démontrer à travers des résultats concrets l'efficacité de cette approche dans l'analyse catégorielle de l'arabe. Pour ce fait, nous nous sommes basés sur un corpus constitué d'un ensemble de phrases variées et bien choisies.

Le choix de ces phrases a été fait en prenant soin de cibler plusieurs aspects de la langue arabe. Nous avons consacré le chapitre 5 pour présenter quelques un de ces aspects.

Au cours de notre analyse du corpus, nous nous sommes intéressés par exemple au cas des phrases de types VSO (verbe sujet Objet). Ce type de phrase est une des particularités de la langue arabe, elles ne sont pas tolérées ni dans la langue française ni dans la langue anglaise.

Notre défi consistait à démontrer que ce type de phrase est syntaxiquement correct en suivant une analyse suivant la grammaire catégorielle combinatoire applicative.

Autrement dit, avec une même assignation de type pour les différentes unités d'une phrase donnée on arrive à prouver que ses deux versions (VSO et SVO) sont syntaxiquement correctes.

Les résultats obtenus suite à l'analyse de notre corpus prouvent qu'il n'y a aucun obstacle théorique dans l'analyse catégorielle de l'arabe.

Cette analyse nous a aussi permis de standardiser quelques affectations de types catégoriels. Par exemple, les types catégoriels affectés à un verbe, un déterminent et un adverbe sont respectivement  $(S/N)/N$ ,  $N/N$  et  $(S/N)/(S/N)$ .

L'application de cette méthodologie à l'arabe ouvre plusieurs perspectives de recherche. En effet, l'analyse menée dans ce mémoire concerne uniquement le traitement de l'arabe standard non voyellé. Cependant, la voyelation joue un rôle capital dans l'arabe. Elle permet parfois de localiser et de deviner la fonction d'une unité lexicale dans une phrase donnée. Par conséquent, elle pourra être un indicateur efficace permettant l'attribution d'unités lexicales.

Nous ne présentons pas ici une application novatrice, mais plutôt, l'adaptation d'une méthodologie d'analyse des langues à la langue arabe. Ce qui en soit est un résultat intéressant.

Les applications potentielles, même si ce n'est pas l'objet de ce mémoire, sont quand même nombreuses. Elles couvrent les domaines du web, des bases de données textuelles, etc. De nos jours, ces applications sont devenues indispensables vu l'augmentation spectaculaire du nombre d'utilisateurs qui ne cessent de montrer leur besoin d'avoir des outils permettant de traiter leur langue maternelle surtout dans le cas des applications web.

## BIBLIOGRAPHIE

- [1] ABÉILLÉ A. 1993, "Les nouvelles Syntaxes, Grammaires d'unification et analyse du Français." Paris, Armand Colin.
- [2] ABEILLÉ A., 1998, "Grammaires génératives et Grammaires d'unification " Langages, 32e année, n° 129. . Diversité de la science du langage aujourd'hui. pp. 24-36.
- [3] ANOUN H., 2009 "Approche logique des grammaires pour les langues naturelles" Thèse de doctorat, Université de Bordeaux I, pages 11-19, Septembre 2009.
- [4] BAHOU Y.,HADRICH BELGUITH L., ALOULOU C., BEN HAMADOU A., 2006 " Adaptation et implémentation des grammaires HPSG pour l'analyse de textes arabes non-voyelles ". 15<sup>e</sup> congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle RFIA'2006, 25/26/27 Janvier 2006, Tours/France.
- [5] BALOUL S.," Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyelle". Thèse de doctorat, université du MAINE. Académie de Nantes,
- [6] BAR-HILLEL Y., "A quasi arithmetical notation for syntactic Description" pages 47-58, 1953.
- [7] Biskri I., 2006, « Applications linguistiques multilingues destinées au WEB : Apport des Grammaires Catégorielles ». In Colloque international « annotation automatique de relations sémantiques et recherche d'informations : vers de nouveaux accès aux savoirs », Université de Paris-Sorbonne, octobre 2006, Paris.
- [8] BISKRI I., "La Grammaire Catégorielle Combinatoire Applicative dans le cadre des Grammaires Applicatives et cognitives " Thèse do doctorat, École des Hautes Études en Sciences Sociales, Juillet 1995
- [9] Biskri I., DESCLÉS J.-P., 1996, "Du phénotype au génotype : la Grammaire Catégorielle Combinatoire Applicative", Conférence internationale Traitement Automatique des Langues Naturelles (TALN96), Marseille, France.
- [10] BISKRI I., MEUNIER, J.G., JOUIS, C., 1997, "Un modèle hybride pour l'extraction des connaissances : le numérique et le linguistique", La mémoire des mots, Cinquièmes Journées scientifiques du réseau thématique : Lexicologie, Terminologie, Traduction:, Tunis, Tunisie, pages 57-71.



- [11] BISKRI I., DESCLES, J.P., 1995, "Applicative and Combinatory Categorical Grammar (from syntax to functional semantics)", Acte du Colloque RANLP, Bulgarie.
- [12] DAHMANI H., BEDDA M., "Conception d'un Système pour La Reconnaissance de Mots Enchaînés Arabes", 2004 JEP-TALN, Traitement Automatique de l'Arabe, Fès, 20 avril 2004
- [13] DESCLÉS J.-P., KANG J., DJIOUA B., CHEONG K., H. SON "Anti-Anti Relativisme et Grammaire catégorielle appliquée au Coréen" Institut coréen de l'éducation des portes françaises, portes françaises éducation, l'article 28 de la Chambre 2008,6, pp. 475 - 530
- [14] DESCLÉS J.-P., BISKRI I., 1995 " Logique combinatoire et linguistique : Grammaire catégorielle combinatoire applicative " Mathématique et sciences humaines, tome 132, pp 39-68, 1995
- [15] DESCLÉS J.-P., 2003 " La Grammaire Applicative et Cognitive construit-elle des représentations universelles ?" Danielle Leeman, editor, Approches syntaxiques contemporaines, volume 48, pages 139–160. Nanterre : 2003.
- [16] GODART-WENDLING B., 2002 " Les trois premières grammaires catégorielles ", Langages Volume 36 Numéro 148 pp 51-66, 2002.
- [17] GODART-WENDLING B., JORAY P., " De la théorie des catégories sémantiques de Lesniewski à l'analyse de la quantification dans la syntaxe d'Ajdukiewicz" Langages, Volume 36, Numéro 148, 2002
- [18] HACHOUR O., 2004 "Reconnaissance hybride des caractères Arabes imprimés" JEP-TALN, Traitement automatique de l'Arabe, Fès 20 Avril 2004.
- [19] JEBALI A., 2005 « Les pronoms liés en arabe classique sont-ils des clitiques ».. Revue des Étudiants en Linguistique du Québec (RELQ), vol. 1, no 1, p22.
- [20] JEBALI A., " Les pronoms clitiques nominatifs du perfectif en arabe standard : vers une analyse lexicaliste." Proceedings of the Niagara Linguistic Society Symposium, Toronto Working Papers in Linguistics, vol. 25 p, 10.
- [21] KHEMAKHEM A., "ArabicLDP : une base lexicale normalisée pour la langue arabe ". Mémoire, Université de Sfax, faculté de Science économique et de Gestion, Tunisie.
- [22] LECOMTE A., "Introduction au calcul de Lambek et à sa version symétrique" cours en ligne source Web : [http://lecomte.al.free.fr/ressources/PARIS8\\_LSL/LamSym.pdf](http://lecomte.al.free.fr/ressources/PARIS8_LSL/LamSym.pdf)

- [23] MAWHOUB M., 2004. “Prosodie et ordre de constituants dans l’énoncé en Arabe Standard Moderne” JEP-TALN, Traitement automatique de l’Arabe, Fés 20 Avril 2004.
- [24] MESFAR S., 2008 Thèse de doctorat “Analyse Morphosyntaxique automatique et reconnaissance des entités nommées en arabe standard”. Université de Franche-Comté, France Novembre 2008.
- [25] RETORÉ C., 2000 “Systèmes déductifs et traitement des langues : un panorama des grammaires catégorielles ” N°3917,Tème 1. Avril 2000.
- [26] ROGER HINDLEY J., LERCHER B., SELDIN J.-P., “Introduction to combinatory logic” Livre pages 14-25.
- [27] SATORI H., HARTI M., CHENFOUR N., 2010’Système de Reconnaissance Automatique de l’arabe basé sur CMUSphinx” : Annals. Computer Science Serie. 8th Tome I Fasc. 2010
- [28] SEGOND F., “Approche des grammaires catégorielles”, mathématique et sciences humaines, tome 1, pages 47- 60, 1990.
- [29] SEGOND F., CHANOD J.-P, “Grammaire applicative : traitement informatique de la composante morpho-syntaxique”, Mathématiques et sciences humaines, tome 103 (1988), page 23-43
- [30] Source Web <http://membres.multimania.fr/rouahcompere/universelle.htm>
- [31] Source Web: <http://internetworldstats.com/stats7.htm>

