

MÉMOIRE DE FIN D'ÉTUDES

Pour l'obtention du diplôme

MASTER EN INFORMATIQUE

Spécialité : Informatique Fondamentale

Thème

**Etude et implémentation d'un analyseur
syntaxique a base de grammaire hors contexte
proposé avec prise en charge les phrases
fondamentaliste et non fondamentaliste**

Présenté par :

✍ Bourouis Azeddine

Encadrés par :

✍ Dr. Yahiaoui Yasser

Année Universitaire : 2018-2019

Remerciements

Ce mémoire n'aurait pas pu être ce qu'il est, sans l'aide d'ALLAH qui m'a donné la force afin de l'accomplir. On tient à exprimer nos profonds remerciements et nos vives reconnaissances à notre promoteur monsieur Yahiaoui Yasser, Maître de conférence au niveau du centre universitaire Salhi Ahmed de Naama, d'avoir proposé et dirigé ce travail, on le remercie infiniment pour ses conseils judicieux et son attention qu'il a apporté et pour l'ambiance sympathique qu'il a créé pour la réalisation de ce mémoire.

C'est avec un grand plaisir que nous adressons aussi nos vifs remerciements à nos enseignants qui ont accepté d'examiner ce travail, on les exprime nos sincères remerciements et notre profonde gratitude.

Enfin à tous ceux qui ont contribué de près ou de loin à l'élaboration de ce modeste travail.

Dédicaces

Dédicaces Ce travail modeste est dédié :

À ma chère mère et mon père ;

À tous mes proches de la famille Bourouis

*s, et plus particulièrement, mes frères et mes sœurs tout à son nom sans oublier la famille
Benhikel;*

À tous mes chers amis et mes camarades de

Centre Universitaire de Naama ; Et à tous ceux qui

M'ont enseigné tout le long de mon parcours

Scolaire et universitaire ;

Azeddine



Sommaire

Liste des figures	
Index des Tables	
Introduction Générale.....	Erreur ! Signet non défini.
Principale travaux du TALARABE.....	5

Chapitre 01: Traitement automatique du langage arabe (TALA)

1. Définition traitement automatique du langage arabe (TALA)	8
2. Les niveaux de traitement en TAL A	8
2.1 Niveau morpho lexical (morphologique)	9
2.2 Niveau syntaxique	9
2.3 Niveau sémantique	11
2.4 Niveau pragmatique	11
3. Les approches de TALN	11
3.1. Approche symbolique	11
3.2. Statistique	12
3.3. Approche connexionniste	12
4. Les applications de TAL	12
4.1. Le traduction automatique	13
4.2. La correction orthographique	13
4.3. Les interfaces naturelles	13
5. Outils facilitateur et réducteur de la complexité du TALA	14
5.1. Absence de voyelles	14
5.2. Segmentation de textes	15
5.3. Analyse morphologique	16
5.4. Étiquetage grammatical	17
5.5. Analyse syntaxique	17

Chapitre 02: Le mot et la phrase Arabe

1. Le mot Arabe.....	21
1.1. Définition du mot arabe.....	21
1.1.1. Le mot (الكلمة)	21
1.1.2. Le mot arabe (الكلمة العربية)	21
1.2. Particularité du mot arabe	21
2. Classification par rapport au rôle syntaxique	21
2.1. Le verbe (الفعل).....	21
2.2. Le nom (الاسم)	21
2.3. Les particules (الحروف)	21
3. La phrase arabe.....	22

3.1 Définition de la phrase arabe.....	22
4. classification de la phrase par rapport à ces éléments.....	26
4.1 Les phrase fondamentaliste (الجملة الاصولية).....	27
4.2 Les phrase non fondamentaliste (الجملة غير الاصولية).....	28

Chapitre 03: Conception et Architecture du système de reconnaissance de phrase

1. Aperçu sur le niveau lexical du mot arabe.....	30
1.1. Le formalisme utilisé.....	30
1.2. Schémas global du mot arabe.....	30
2. Architecture globale du système morphosyntaxique.....	34
2.1 Les démarche du système.....	34
2.1.1 Nettoyage du texte.....	34
2.1.2 Analyse lexicale.....	34
2.1.3 Analyse syntaxique.....	34
3. Diagramme de classe.....	35
4. Grammaire.....	36
4.1 Grammaire formelle.....	Erreur ! Signet non défini.
4.1.1 Définition.....	Erreur ! Signet non défini.
4.1.2 Classification des grammaires.....	Erreur ! Signet non défini.
4.1.3 Construction de la grammaire initiale.....	Erreur ! Signet non défini.
4.1.4 Codage et nettoyage de la grammaire.....	45
4.1.5 Construction de la grammaire propre.....	37
4.1.6 Vérification de la régularité de la grammaire propre.....	39
4.2 Transformation de la grammaire a la forme normale de Chomsky.....	Erreur ! Signet non défini.
4.2.1 Définition de la forme normale de Chomsky.....	Erreur ! Signet non défini.

Chapitre 04: Implémentation

1. Environnement matériel.....	53
2. Environnement logiciel.....	53
3. Implémentation.....	54
3.1 Fenêtre principale.....	54
3.2 Interface expert.....	56
4. Résultat.....	56
4.1 Analyse syntaxique et étiquetage.....	56
4.2. Segmentation du texte.....	57
5. Discussion et commentaire sur les résultats.....	57
Conclusion Et Perspectives.....	59
Bibliographie.....	62
Annexe.....	64

Figure -1- Les niveaux d'analyse d'un texte.....	7
Figure -2- Arbre syntaxique 1	8
Figure -3- Arbre syntaxique 2.	9
Figure -4- : Ambiguïté du mot " بين " causée par l'absence des diacritiques... ..	13
Figure -5- : ambiguïté morphologique du mot " فصل "	15
Figure 6. .Analyse-Génération.....	16
Figure 7.les alphabet arabe.....	18
Figure 8. Classification du mot arab.....	19
Figure.10 les types de phrase arabe.....	22
Figure 11 : Répartition formelle de la phrase fondamentaliste	23
Figure 12 : Répartition formelle de la phrase non fondamentaliste.....	24
Figure 13 : L'ensemble des critères auquel régit la phrase nominale non fondamentaliste.	24
Figure 14 : L'ensemble des critères auquel régit la phrase verbale non fondamentaliste ...	25
Figure.15 le principe de l'hierarchisation des classes de mots arabe.....	26
Figure16 : classification du mot arabe	26
Figure.17 : Classification du verbe	27
Figure.17 : Classification du nom	28
Figure.19 : Classification de la particule (الحروف او الأدوات) arabe	29
Figure 20 : Architecture globale du système morphosyntaxique	30
Figure 21 : Diagramme de class modélisant la phrase Arabe	32
Figure 22 : l'interface principale	52
Figure 1 : l'interface experte pour classer les nouveaux mots apparus	53
Figure 24 : Segmentation après un prétraitement réussi	54

Tableau 1 : codage de la catégorie syntaxique.....	42
Tableau 2 : codage du vocabulaire.....	44
Tableau 3 : prétraitement et étiquetage d'un petit texte.....	53
Tableau 4 : prétraitement d'un petit texte, contient des mots non reconnus par le corpus.....	54

INTRODUCTION

1. Introduction générale

Le traitement automatique de l'arabe est un domaine de recherche stimulant. Les problématiques de ce domaine se présentent comme des problématiques délicates. Ils sont divers et ils présentent des degrés de complexité élevés à cause de la basse maîtrise des phénomènes linguistiques par les techniques calculables.

Le TAL « ...combine en effet plusieurs défis intéressants, parmi lesquels on peut citer la complexité morphologique de la langue, son haut degré d'ambiguïté et l'existence de nombreux dialectes présentant des variantes significatives. L'arabe représente en outre un enjeu géopolitique considérable puisqu'il est parlé par plus de 400 millions de locuteurs, dans des pays très variés en termes de stabilité et de prospérité. C'est la principale langue des réfugiés les plus récents, que ce soit au Moyen Orient ou en Europe. Travailler sur l'arabe et ses dialectes, c'est ainsi prendre en compte un vaste territoire, allant du monde arabe à la région méditerranéenne, et plus généralement l'ensemble du reste du monde ». [1]

1. Des techniques et méthodes fondamentales proposées s'intéressent aux sujets suivants : analyse morphologique, désambiguïsation, segmentation, étiquetage morpho-syntaxique (POS), détection des entités nommées, chunking, analyse syntaxique, annotation en rôles sémantiques, reconnaissance d'opinion et de sentiments, modélisation des dialectes de l'arabe, etc ... ,
2. Les applications a savoir : traduction automatique, reconnaissance et synthèse de la parole, reconnaissance de caractères (OCR), systèmes question/réponse, pédagogie, technologies d'assistance, médias sociaux, textométrie, etc... ,
3. Les ressources : dictionnaires, données annotées, corpus, etc... ,
4. Les articles théoriques ayant une pertinence particulière pour le traitement de l'arabe.

Ce travail est la suite d'un travail déjà réalisé par des étudiant du centre Universitaire de Naâma en 2018 qui porte sur la segmentation des textes en phrase basé sur la classification syntaxique des mots mais sans prise en compte des phrases dites non-fondamentalistes donc notre apport sera sur la prise en compte de ce type de phrases.

La classification utilisée propose une représentation des connaissances dite *SHA* pour *Subsumption Hierarchical Attribute* qui permet de remplacé chaque classe de mots par un vecteur significatif permettant de symboliser la sens de la classe et ses relations directs et indirect avec les autres classes.

Ce mémoire est organisé de la façon suivante :

Une introduction générale exposant l'essentiel du travail réalisé.

Un premier chapitre permettant de décrire un background sur le traitement automatique de la langue Arabe.

Un deuxième chapitre qui parle des mots et de la phrase Arabe ou on essaye de montrer les différentes caractéristiques particulières, autrement dit la phase d'analyse de l'existant.

Le troisième chapitre qui propose une grammaire génératrice des phrases arabes quel que soit sa nature (fondamentaliste ou non) et montre la conception l'architecture globale.

Le quatrième chapitre représente les résultats de l'implémentation et les commentaires sur les résultats de l'évaluation.

Etat de l'art

1. Principale travaux du TALARABE :

travail	fonctionnement	Approche
STAr (Belguith Hadrach; L. , Baccour L.; Mourad G., 2005)	« Ce système accepte en entrée un texte en arabe non voyellé de type TXT et produit en sortie un texte segmenté en phrases et propositions sous forme d'un fichier de type XML. »[4]	« le système STAr, se base sur la technique de l'exploration contextuelle c'est à dire des signes de ponctuation, des conjonctions de coordination et une liste de particules qui sont considérées comme des critères de segmentation. L'exploration contextuelle repose sur une étude des indices linguistiques déclencheurs appelés indicateurs et des indices complémentaires associés à ces indicateurs et sur un ensemble de règles ». [4]
(Khalifa; I., Feki; Z., Farawila; A. , 2011)	« ont proposé une méthode d'apprentissage pour la segmentation des textes arabes en clauses ». [4]	« Basé sur l'exploitant unique des fonctions rhétoriques du connecteur " و /et". Les auteurs ont défini six sens pour ce connecteur " و " : (waw: (1)"alquassam/ (2)alrob/ (3)alistiinaf/ (4)alhal/ (5)almaeyat/ (6)eleatf"). Parmi ces six sens, deux classes ont été définies : "Fasl" (contenant 1, 2 et 3), qui est considérée comme un bon indicateur de segmentation, et la classe contenant (4),(5) et (6) qui n'a pas d'effet sur la segmentation. Un ensemble de 22 traits syntaxiques ont ensuite été utilisés afin de classer automatiquement chaque instance du connecteur " و " dans ces deux classes. »[4]
(Habash et Rambow, 2004)	« La TAG utilisée pour la représentation de la syntaxe arabe. ont construit une grammaire d'arbres adjoints par extraction d'arbres élémentaires à partir d'un corpus arboré. »[19]	« Pour ce faire, ils ont utilisé la partie 1 de la version 2.0 de PATB (Penn Arabic TreeBank) (Maamouri et al., 2004). Cette partie contient 160,000 mots annotés qui composent des textes du style journalistique. L'extraction passe par une réinterprétation du corpus en des structures de dépendances. Les chercheurs ont pu extraire des structures avec variation des positions de leurs composants vu que la langue arabe est à ordre semi-fixe. Ils ont, alors, obtenu des phrases à composition VSO (Verbe Sujet Objet) et d'autres à SVO, alors qu'ils ont échoué à obtenir la composition VOS. Cet échec est dû à l'absence de ce genre de structures dans le corpus utilisé pour l'extraction (Habash et Rambow, 2004). Le problème de couverture est posé vu que cette construction est liée aux styles de textes contenus dans le corpus »[19]

<p>(Said A. Salloum, Mostafa Al-Emran and Khaled Shaalan,2016)</p>	<p>Une étude de la grammaire fonctionnelle lexicale dans le contexte arabe.</p>	<p><i>« Le Lexical Functional Grammar (LFG) est une hypothèse linguistique de grammaire qui concerne la nature de la structure de l'énoncé et génère un cadre réaliste pour le traitement du langage naturel .LFG a été inventé par Joan Bresnan en 1970 et il a donné le formalisme de structure de phrase destiné à typologiquement dans diverses langues</i></p> <p><i>LFG joue un rôle crucial dans le traitement du langage naturel (TALN) car il donne un formalisme pour représenter les connaissances</i></p>
<p>(BENDEBICHE Rafiq - BOUDJENANE Salah,2018)</p>	<p>définir une grammaire formelle de type 3(classification de N. Chomsky) qui permet une analyse syntaxique des phrases fondamentalistes arabes.</p>	<p>Fait un classification pour les éléments du phrase par exemple le verbe est défini par un vecteur 1.1.0.0.0 Le sujet par 1.2.0.0.0 Et les conjonctions par 1.3.0.0.0</p>

Tableau 1 : état de l'art [10]

Chapitre 01

Traitement Automatique du Langage Arabe T.A.L.A

Introduction

L'arabe est parlée dans plus de 22 pays, du Maroc jusqu'à l'Iraq et dans toute la péninsule arabe. C'est la première langue pour plus de 250 millions de personnes et la deuxième pour 40 millions. L'arabe, langue du Coran, est devenue la langue d'une civilisation et ne sert plus seulement à désigner les seuls habitants de la péninsule arabe qui la parlaient. [3]

Les outils du traitement automatique de la langue en général, et de la langue Arabe en particulier, se caractérisent par leur diversité au niveau des langages de développement utilisés, des entrées/sorties manipulées, des représentations internes et externes des résultats, etc. Cette diversité ne favorise ni l'interopérabilité entre ces différents outils, ni leur réutilisabilité dans de nouveaux contextes. [4]

1. Définition traitement automatique du langage arabe (TALA) :

Est une discipline à la frontière de la linguistique, de l'informatique et de l'intelligence artificielle, qui concerne l'application de programmes et techniques informatiques à tous les aspects du langage humain. En tant que tel, le TAL ou TALN appartient à l'ingénierie linguistique. [2]

Aussi TALN est une discipline s'appliquant au domaine de l'informatique et du langage. Il est utilisé par exemple pour les traductions, la reconnaissance vocale ou encore les réponses automatiques aux questions. Ces domaines représentent des défis majeurs, car les mots du langage sont souvent traités un à un par l'ordinateur. [5]

Grâce au traitement du langage naturel, une cohérence tente d'être apportée aux textes en s'attachant au sens des phrases et formules. Ces avancées ne sont pas uniquement utiles pour les traducteurs ou chat bots mais aussi lorsque les ordinateurs exécutent des ordres oraux ou communiquent de manière vocale afin de faciliter par exemple la communication pour les personnes aveugles. Pour pouvoir résumer des textes longs, ou extraire des informations précises, les ordinateurs ont besoin également de comprendre la cohérence linguistique des textes. [5]

2. Les niveaux de traitement en TALA :

On distingue plusieurs niveaux d'analyse d'un texte :

Le niveau morphologique ou lexical qui s'intéresse à la décomposition d'un texte en différents mots (ex : en Français découpage selon les espaces).

Le niveau syntaxique qui définit comment les mots sont agencés dans une phrase, et l'analyse de la forme d'un texte.

Le niveau sémantique qui correspond à la signification des mots et des groupes de mots

(Analyse du sens d'un texte (ex : typage)).

Le niveau pragmatique grâce auquel, le contexte est pris en compte. (Voir la figure -1-).
[6]

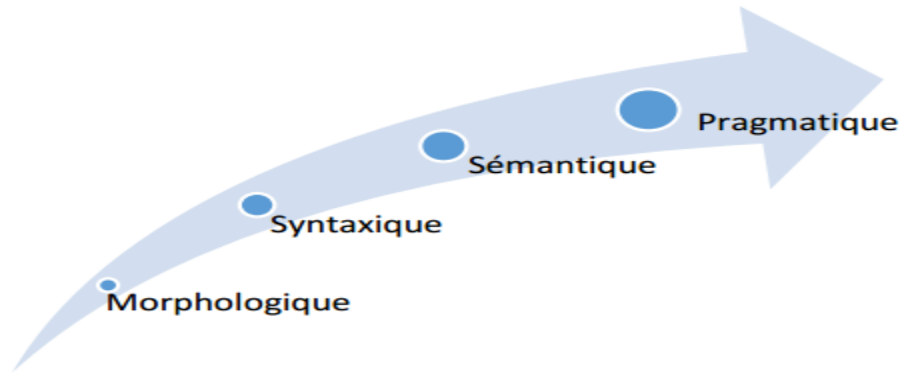


Figure -1- Les niveaux d'analyse d'un texte

2.1 Niveau morfo lexical (morphologique) :

La morphologie est l'étude de la forme des mots (de leur flexion – indications de cas, genre, nombre, mode, temps, etc. de leur dérivation préfixes, suffixes, infixes et de leur composition mots composés). Sous l'appellation de morphosyntaxe, elle représente également l'étude des règles de combinaison des morphèmes (unités minimales de sens) selon la configuration syntaxique de l'énoncé. En pratique, dans le cadre du traitement automatique de la langue, l'analyse morphologique consiste à segmenter le texte en unités élémentaires (tokenisation) et à déterminer les différentes caractéristiques de ces unités [7]

2.2 Niveau syntaxique :

L'analyse syntaxique est l'étude de la structure de la phrase, dans le but de définir comment les lexèmes sont organisés et quelles fonctions ont les mots qui servent à les mettre en relation. Les lexèmes sont réunis en syntagmes, des groupes de mots dont la fonction est connue. Il par constituants divise tout d'abord la phrase en plusieurs groupes de mots appelés syntagmes. Un syntagme est un intermédiaire entre l'ensemble global qu'est la phrase et la division unitaire que sont les mots. Il permet de représenter, sous forme symbolique ou graphique, la ou les structures syntaxiques d'un texte. En d'autres termes, il s'agit de la mise en évidence des structures d'agencement des catégories grammaticales (nom, verbe, adjectif, etc.), afin d'en découvrir les relations formelles ou fonctionnelles (par exemple, sujet, verbe et complément) [8].

Notion de Syntagme

Un **syntagme** est un ensemble de mots formant une seule unité catégorielle et fonctionnelle, mais dont chaque constituant, parce que dissociable (contrairement au mot composé), conserve sa signification et sa syntaxe propres. Un syntagme constitue donc une association occasionnelle, libre, alors que le mot composé est une association permanente (lorsqu'un syntagme se fige, il devient bien sûr un composé détaché, soit une locution). Il s'agit donc d'un groupe de mots formant une unité à l'intérieur de la phrase. Dans le cadre de l'analyse syntaxique d'une phrase, il s'agit d'une segmentation en unités fonctionnelles appelées syntagmes. Par exemple, on peut citer les types de syntagme suivants : syntagme nominal, syntagme verbal, syntagme adjectival, etc[11] . Les arbres de dérivation/syntaxique sont un moyen de représentation des syntagmes. Exemple 1 : pour la phrase « Le facteur apporte une lettre », l'arbre de dérivation est donné dans la figure 2. Soit la grammaire qui donne d'abord le sens des abréviations (SV, SN , S , V ,Det)[8].

$P \rightarrow SN \mid SV$

$SN \rightarrow Det \mid N$

$SV \rightarrow V \mid SN \text{ pp}$

$PP \rightarrow DT \mid N$

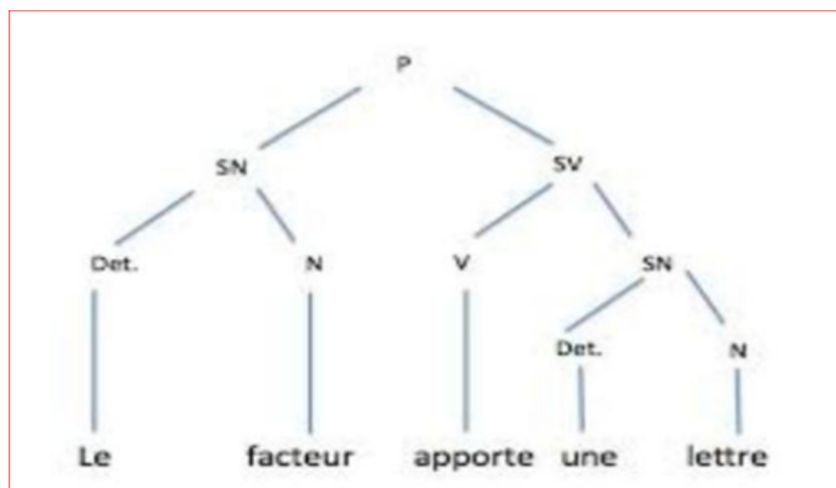


Figure -2- Arbre syntaxique 1

Exemple 02 :pour la phrase « son chien mange un os dans la cuisine. »

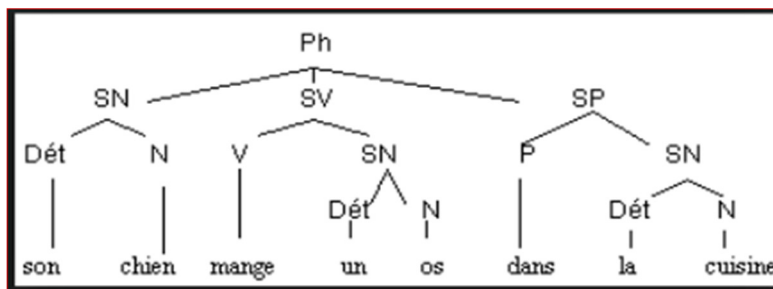


Figure -3- Arbre syntaxique 2

2.3 Niveau sémantique :

La sémantique vise à l'étude de sens hors contexte. Le traitement sémantique prend comme unité d'analyse de la phrase, et conduit à représenter sa partie significative. L'analyseur sémantique d'écrit le sens des mots de la phrase ; ces mots sont identifiées par l'analyse morphologique, et regroupes en structures par l'analyse syntaxique [9] .

2.4 Niveau pragmatique :

L'analyse pragmatique permet d'utiliser les connaissances pragmatiques afin d'interpréter des situations du monde réel. Cette étape est importante pour le processus de compréhension d'un texte ou d'une phrase, elle représente le lien entre l'analyse linguistique et le monde réel [8].

3. Les approches de TALN

3.1. Approche symbolique

Elle est caractérisée par une analyse profonde des phénomènes linguistiques, basée sur les niveaux cités auparavant.

La principale source dans les systèmes symboliques est le développement des règles et des lexiques. Un bon exemple de l'approche symbolique est vu dans des systèmes logiques à base de règles. La structure symbolique est généralement sous forme de propositions logiques. Les manipulations de ces structures sont définies par des procédures d'inférence ; le moteur d'inférence sélectionne à plusieurs reprises la règle dont la condition est remplie. Un autre exemple d'approches symbolique est les réseaux sémantiques. L'approche symbolique a été employée dans plusieurs domaines de recherche comme l'extraction d'information, la catégorisation de textes, la résolution d'ambiguïté, etc.

Les techniques typiques sont : l'apprentissage à base de règles, les arbres de décision, l'algorithme de K plus proches voisins, etc. [6].

3.2. Statistique

Elle emploie des techniques mathématiques, ainsi elle utilise des corpus de textes pour le développement des modèles des phénomènes linguistiques. Un modèle fréquemment utilisé est le modèle caché de Markov (HMM). Le HMM est un automate d'états finis avec un ensemble d'états et des probabilités attachés aux transitions (automate flou). L'approche statistique a été typiquement utilisée dans des tâches telles que la reconnaissance vocale, l'acquisition de lexique, le POS tagging, la traduction statistique, etc [10].

3.3. Approche connexionniste

Comme l'approche statistique, l'approche connexionniste développe des modèles à partir des exemples de phénomènes linguistiques. Ce qui diffère l'approche connexionniste de l'approche statistique est que les modèles connexionnistes combinent l'apprentissage statistique avec différentes théories de représentation. Ainsi dans les systèmes connexionnistes, les modèles linguistiques sont difficiles à observer en raison des architectures qui sont moins limités qu'aux statistiques.

Le modèle connexionniste est un réseau interconnecté de simples unités de traitement avec des connaissances stockées dans le poids des connexions entre les unités.

Les interactions entre les unités peuvent entraîner un comportement dynamique global qui à son tour, conduit à un calcul.

NB : Les recherches actuelles tendent vers le développement par des techniques hybrides qui utilisent les points forts de chaque approche pour tenter de résoudre les problèmes [10].

4. Les applications de TAL :

Concernant les applications, la demande de TAL provient, pour dire vite, de deux tendances « lourdes » : d'une part la nécessité de concevoir des interfaces de plus en plus ergonomiques, d'autre part la nécessité de pouvoir traiter (produire, lire, rechercher, classer, analyser, traduire) de manière de plus en plus « intelligente » les informations disponibles sous forme textuelle, de manière à pouvoir résister à leur prolifération exponentielle. Les applications des techniques de TAL sont donc nombreuses et variées. Nous avons regroupé ces applications en trois grandes familles, qui correspondent aux aides à la lecture de documents, aux aides à la production de documents, et enfin aux interfaces homme-machines [11].

4.1. Traduction automatique (TA) :

Il est probable que la TA fasse l'objet d'améliorations importantes dans les années à venir.[6]

4.2. Correction orthographique :

Intégrée à toute application informatique impliquant la rédaction, correction basée sur des lexiques. Exemple : traitement de texte, courrier électronique, navigateur Internet (zone de saisie).[6]

4.3. Atelier d'aide à la rédaction :

Orthographe, grammaire, style..., Exemple : Antidote pour le français.[6]

4.4 Correction grammaticale :

Les meilleures applications fonctionnent bien mais sont payantes, actuellement aucune application libre pour le français.[6]

4.5 La reconnaissance de la parole :

Discipline ayant fait des progrès considérables , Grandes étapes : {Segmentation du flux continu de paroles en unités discrètes ,identification du phonème correspondant à chaque unité ,regroupement des unités pour constituer des mots, prise en compte de la syntaxe pour finaliser le texte écrit } , logiciels de dictée vocale (Via Voice, Dragon Dictate...) , reconnaissance de la parole ou commande vocale (Reconnaissance vocale de Windows, Systèmes de navigation routière GPS, Smartphone...) , prototype Google de sous-titrage automatique de You tube.[6]

4.6 Synthèse de la parole :

Créer de la parole artificielle à partir d'un texte quelconque. Ces systèmes ont largement franchi le seuil de l'intelligibilité permettant leur utilisation, difficultés : désambiguïsation des homographes hétérophonies, gestion de la prosodie (intonation, rythme et intensité)...[6]

4.7. Le traitement documentaire

Les applications les plus immédiates du TAL sont celles qui visent à faciliter le traitement par l'humain des immenses ressources disponibles en langage naturel.[6]

4.8. La production de documents

Si autant de documents électroniques sont aujourd'hui disponibles, c'est bien que quelqu'un les a écrits. Dans le domaine de l'aide à la production de texte (la génération de textes), les applications du TALN sont également nombreuses

- les claviers « auto-correcteurs » (par exemple pour les handicapés) ;
- la reconnaissance optique de caractères. De nombreux systèmes commerciaux sont aujourd'hui disponibles, avec des performances très satisfaisantes : Recognita, Omnipage, ScanWorX... .

- la génération automatique de documents à partir de spécifications formelles. En fait, de nombreux secteurs d'activité impliquent la production massive de textes très stéréotypés à partir de spécifications plus ou moins formelles (textes juridiques, compte-rendu d'exploration d'une base de données, rapports d'analyses statistiques, documentations techniques, etc). Pour cette classe de documents, il est parfaitement possible de générer automatiquement, sinon des textes complètement définitifs, du moins des versions préliminaires qui seront ensuite finalisés par des rédacteurs humains. On retrouve dans ces applications la même dialectique que dans les applications destinées à faciliter la gestion de documents. D'un côté des applications à large couverture, qui utilisent essentiellement des ressources lexicales, avec des fonctions d'accès tolérant (permettant la correction d'erreurs) au lexique : c'est le cas des applications qui tournent autour de la correction orthographique. De l'autre, des applications qui intègrent des mécanismes de traitement de plus haut niveau (typiquement la génération), mais qui ne fonctionnent que pour des domaines beaucoup plus restreints [11].

4.9. Les interfaces naturelles

Dernier domaine d'application, qui est sans doute celui dans lequel la demande de traitements linguistiques est la plus forte, le domaine des interfaces naturelles (i.e. en langage naturel) telles que :

- l'interrogation en langage naturel de bases de données (traduction langage naturel ↔ SQL) ou de moteurs de recherche sur la toile. De multiples applications de ce type commencent à se mettre en place sur la toile.

- les interfaces vocales, qui mettent en œuvre de manière variable suivant les applications des modules de reconnaissance de parole, synthèse de parole, génération et gestion de dialogue, accès aux bases de connaissance,..., chacun de ces modules demandant des traitements spécifiques (désambiguïsation morpho syntaxique et identification de syntagmes pour la synthèse, grammaires stochastiques pour la reconnaissance de la parole...).

Les premiers systèmes « grand-public » de dictée vocale commencent à véritablement arriver sur le marché (Via Voice, le produit d'IBM, Dragon Dictate, et bien d'autres encore), et l'intégration dans Windows d'une API de traitement de la parole devrait, dans les années qui viennent, faire littéralement exploser le marché des technologies vocales. De nombreux services commerciaux existent déjà, ou sont proches de la commercialisation qui font appel à l'ensemble de ces techniques (reconnaissance-dialogue-synthèse), pour des applications diverses telles que les ordinateurs « main-libre », la lecture téléphonique de courriers 21 électroniques, la réservation de billets (train, avion)... la liste est pratiquement sans limite. Le marché des technologies vocales représentait en 1995 quelques centaines de millions de dollars pour les États-Unis, et est estimé, pour l'an 2 000, à près de 3 Milliards de dollars [11].

5. Outils facilitateur et réducteur de la complexité du TALA :

Les principaux problèmes de l'analyse automatique de l'arabe écrit présentes dans les textes à savoir : l'absence de voyelles, la segmentation de textes, l'analyse morphologique, la lemmatisation du pluriel brisé, l'étiquetage grammatical et l'analyse syntaxique.

5.1. Absence de voyelles :

Les voyelles courtes sont généralement absentes dans la plupart des textes écrits en arabe, ce qui constitue un grand facteur d'ambiguïté dans les tâches d'annotation. Généralement, c'est au lecteur d'inférer les voyelles au moment de la lecture. Les textes non diacritiques sont très ambigus et la proportion des mots ambigus dépasse les 80% [10].

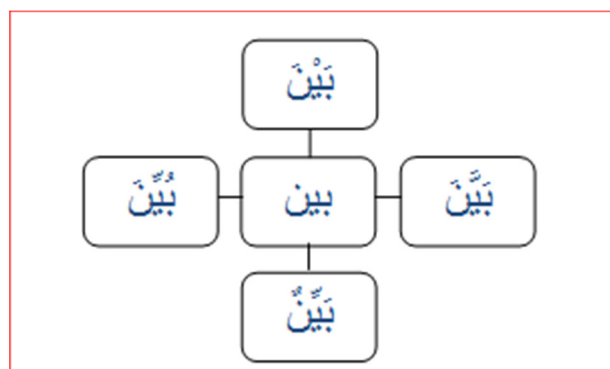


Figure -4- : Ambiguïté du mot " بين " causée par l'absence des diacritiques

5.2. Segmentation de textes

La segmentation d'un texte est une étape fondamentale pour le traitement automatique; son rôle est de découper le texte en unités d'un certain type qu'on aura défini et repéré préalablement. En effet, l'opération de segmentation d'un texte consiste à délimiter les

segments de ses éléments de base en éléments constituants différents niveaux structurels tels que : le paragraphe, la phrase, le syntagme, le mot graphique, etc. La langue arabe n'est pas appuyée principalement sur les signes de ponctuations et les marqueurs typographiques ; il est à noter que ces derniers ne sont pas utilisés de façon régulière dans les textes arabes actuels, et même dans le cas où ils y figurent, ils ne sont pas gérés par des règles précises d'utilisation. Par ailleurs, nous pouvons trouver tout un paragraphe ne contenant aucun signe de ponctuation à part un point à la fin de ce dernier. Ainsi, il convient de noter que la présence des signes de ponctuation ne peut guider la segmentation comme c'est le cas du français ou de l'anglais. D'autres particularités de cette langue s'ajoutent et posent un problème lors de la segmentation du texte en phrases nous en citons: (1) l'absence de majuscules qui marquent le début d'une nouvelle phrase, et (2) l'ambiguïté des signes de ponctuation et de certains particules comme "et" " و " et "donc", "ف" qui peuvent dans certains cas, ne plus jouer le

rôle de déclencheurs d'une nouvelle phrase. L'une des solutions serait de prendre en compte leur contexte droit et gauche. Et c'est cette solution que nous avons adoptée dans notre travail [6].

5.3. Analyse morphologique

L'opération de l'analyse morphologique tient à étudier la forme d'un mot en faisant une analyse interne de sa structure. Le but étant de le décomposer en des éléments plus petits (préfixes, suffixes, etc.) selon des règles de combinaison relatives à ces derniers. L'analyse morphologique fait la séparation et l'identification des morphèmes semblables aux mots préfixés (comme les conjonctions "et" " و " et "puis" "ف" etc.), aux prépositions préfixées (comme "avec" "ب" et "pour" "ل" l'article défini "ال" etc.), aux suffixes de pronom possessif, puis détermine un schéma possible. Les préfixes et suffixes sont trouvés en enlevant progressivement des enclitiques et des proclitiques et en essayant de faire correspondre toutes les racines produites par un schème afin de retrouver la racine. Le problème principal de cette analyse réside dans l'agglutination et l'absence de voyellation. Cette caractéristique engendre une ambiguïté morphologique au cours de l'analyse. Ainsi, la reconnaissance des unités lexicales qui composent une unité morphologique n'est pas toujours facile à détecter. Le problème est de reconnaître que la bonne segmentation réside ainsi, dans la difficulté de distinction entre une proclitique ou enclitique et un caractère original du mot.

L'absence de voyellation pose un autre problème important. En effet, les mots non voyelles engendrent beaucoup de cas ambigus au cours de l'analyse morphologique. Prenons l'exemple illustré par la figure 1.6, le mot non voyelle "لصف" pris hors contexte peut être un verbe au passé conjugué à la troisième personne du singulier "il a licencié/ كُصف" ou un nom masculin singulier "chapitre/ saison | لصف", ou encore une

concaténation de la conjonction de coordination "puis / ف" avec le verbe " لئ " : impératif du verbe "lier" conjugué à la deuxième personne du singulier masculin [10].

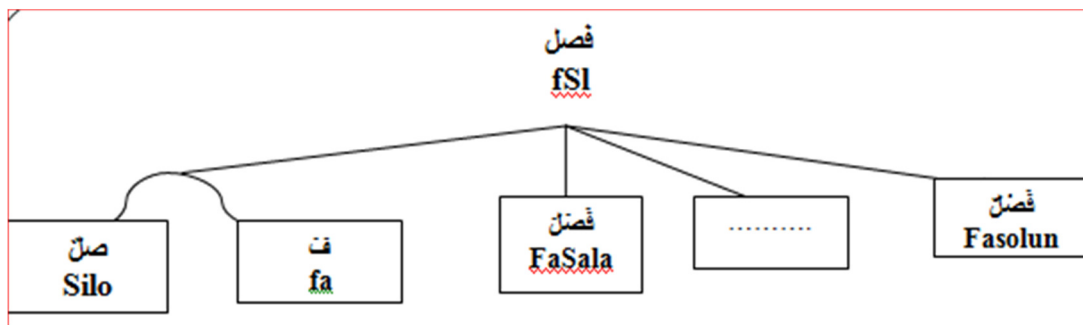


Figure -5- : ambiguïté morphologique du mot " فصل "

5.4. Étiquetage grammatical

L'étiquetage grammatical est le fait d'assigner une étiquette à chacun des mots du texte (nom, verbe, adjectif, article défini, etc.) dans le contexte où il apparaît. L'étiquette contient des informations morphosyntaxiques sur le mot, c'est à dire des informations concernant sa forme et sa fonction. Elles comprennent notamment la catégorie grammaticale, le genre, le nombre, le temps et le mode. La difficulté de l'étiquetage grammatical s'amplifie lorsque les textes visés se présentent sous leur forme partiellement vocalisée ou encore totalement non vocalisée, ce qui correspond au cas le plus courant. Dans ces conditions, le but général de l'étiquetage grammatical consiste à répondre à la question suivante : Comment associer aux différents mots qui composent un texte l'étiquette qui leur convient, compte tenu du contexte où ils occurrent ? [10].

Le problème de la vocalisation d'un mot est ainsi posé du fait que le choix de l'accentuation qui convient au mot est difficile et dépend essentiellement du contexte [10].

5.5. Analyse syntaxique

La syntaxe est l'étude des contraintes portant sur les successions licites de formes qui doivent être prises en compte lorsque l'on cherche à décrire les séquences des phrases grammaticalement correctes. L'analyse syntaxique permet d'associer à un énoncé sa ou ses structures syntaxiques possibles, en identifiant ses différents constituants et les rôles que ces derniers entretiennent entre eux. Toutefois, l'analyse syntaxique prend en entrée le résultat de l'analyse lexicale (éventuellement de l'étiquetage morphosyntaxique) et fournit en sortie une structure hiérarchisée des groupements structurels et des relations fonctionnelles qui unissent les groupements. Enfin, il est à signaler que les ambiguïtés lexicales et grammaticales, relatives à la non vocalisation des mots, pose des difficultés au niveau de l'analyse syntaxique.

Ainsi, une phrase, en absence de la vocalisation, peut être interprétée et traduite selon plusieurs interprétations qui sont toutes syntaxiquement correctes ce qui pose des ambiguïtés d'ordre sémantique [10].

6. Objectif du TALN :

L'objectif du TALN est la conception de logiciels, capables de traiter de façon automatique des données linguistiques, c'est-à-dire des données exprimées dans une langue (dite "naturelle"). Ces données linguistiques peuvent être des textes écrits, ou bien des dialogues écrits ou oraux, ou encore des unités linguistiques de taille inférieure à ce que l'on appelle habituellement des textes (par exemple : des phrases, des énoncés, des groupes de mots ou simplement des mots isolés). En fonction des objectifs attendus d'un système de TALN, on peut le classer dans l'un des domaines suivants : analyse et génération.

6.1 L'analyse : Le processus d'analyse consiste à démarrer de la structure de surface (locution ou texte écrit) pour arriver à la structure profonde équivalente. Plus précisément, l'ordinateur doit pouvoir interpréter un texte écrit dans une langue naturelle afin d'en obtenir une représentation formelle. Pour ce faire, il doit pouvoir analyser le langage à différents niveaux : morphologique, syntaxique, sémantique, pragmatique, l'ordinateur doit donc faire appel à des connaissances linguistiques importantes (grammaires, dictionnaires, etc.) ainsi qu'à des connaissances pragmatiques (contexte immédiat, expérience, connaissances du monde, etc.).

6.2 La génération: Le processus de génération est le processus inverse : il consiste à passer de la structure profonde à la structure de surface. Plus précisément, l'ordinateur doit être capable de passer de la représentation formelle d'un énoncé à une formulation en langue naturelle. Ceci implique la même connaissance que pour l'analyse, mais avec des problèmes spécifiques à l'énonciation.

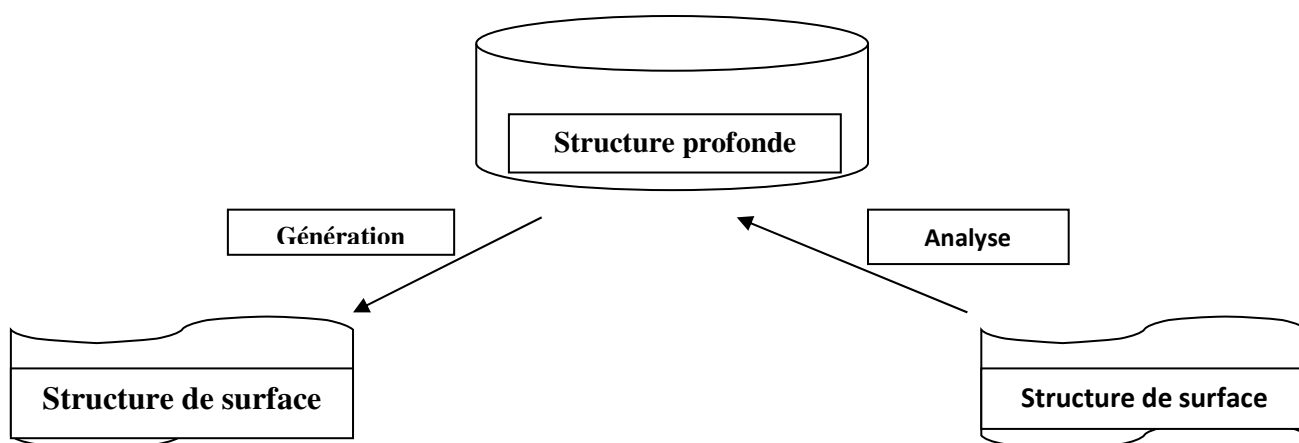


Figure 6. .Analyse-Génération

Remarque: il est possible qu'une application de TALN appartienne en même temps aux deux domaines : analyse et génération ; c'est le cas par exemple d'une application de traduction automatique.[12]

7. Conclusion

Dans ce chapitre nous avons exposé l'essentiel du background permettant de comprendre le travail effectué pour ce projet.

Chapitre 02

Le mot et la phrase arabe



Introduction

Ce chapitre expose l'existant c'est-à-dire il explique les éléments de la langue arabe afin de pouvoir proposer une grammaire génératrice implémentable dans un analyseur syntaxique. Cette grammaire prend en considération les deux types de phrase en arabe a savoir : les phrase fondamentalistes et les phrase non fondamentalistes.

1.alphabet Arabe :

Les lettres de l'alphabet en langue arabe sont les lettres de base à partir desquelles la langue est formée et à partir desquelles la parole est formée en langue arabe, et ce ne sont que des lettres originales sans sous-lettres, et très probablement les dictons sont vingt-huit lettres.

- ✓ L'alphabet arabe se compose de 28 lettres.
- ✓ L'arabe s'écrit et se lit de droit à gauche contrairement aux écritures latines. **Figure7**
- ✓ La construction d'un mot se fait en collant les lettres une à une de droite à gauche.
- ✓ En arabe chaque lettre de l'alphabet s'écrit différemment selon sa position dans le mot[13].



Figure 7.les alphabet arabe

2. le mot Arabe :

Le mot est la pierre angulaire sur laquelle l'écrivain et le porte-parole s'appuient sur l'ingénierie et la construction de leurs mots, c'est-à-dire l'établissement de phrases qui communiquent, comprennent et répondent les locuteurs aux énoncés. Les linguistes ont donné une définition du concept mot comme suite :

2.1. Lemot

C'est la plus petite unité sur laquelle la langue est construite et comprend trois sections : nom, verbe et particule.

2.2 Le mot arabe (الكلمة العربية) :

Un mot arabe est un mot, comme défini ci-dessus, qui répond aux deux conditions suivantes :

- ✓ Tous ses caractères sont alphabets arabes nus ou avec diacritiques.
- ✓ Les mots d'origine arabe sont divisés à leur tour en deux sous-catégories :

a- Mots arabes dérivés : Ce sont les verbes et les noms qui sont construits selon

les règles de dérivation arabe.

b- Mots arabes fixes : Ce sont un ensemble de mots moulés par les Arabes, autrefois, et ne respectent pas les règles de dérivation arabe. La plupart de ces mots ne sont ni fixes ni les noms des verbes, la plupart d'entre eux sont des mots fonctionnels (particules) comme les pronoms, les prépositions, conjonctions, mots interrogatifs, et autres. Ils peuvent être considérés comme la meilleure colle qui lie les mots de la phrase arabe ensemble. La catégorie des mots arabes fixes contient un nombre limité.

Les mots arabisés : les mots arabisés sont des noms empruntés à des langues étrangères (peut-être avec un certain ajustement phonétique pour convenir à la prononciation arabe) et sont devenues communs [10].

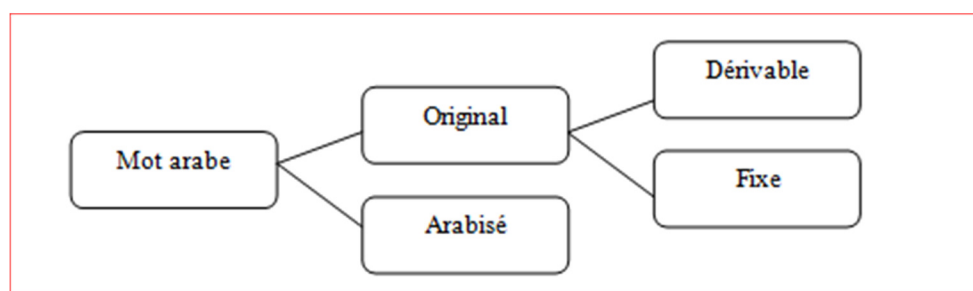


Figure 8. Classification du mot arabe.

3. Classification par rapport au rôle syntaxique :

L'arabe considère 3 catégories de mots à savoir :

3.1. Le verbe (الفعل) :

Le verbe est un mot qui se conjugue, et qui indique un état ou une action faite ou subie par le sujet.

En arabe, on désigne un verbe par sa forme conjuguée la plus simple, c'est-à-dire la troisième personne masculin singulier de l'accompli.

Cette forme équivaut à la forme de base qu'est en français l'infinitif. Par exemple : كَتَبَ - kataba = "il écrivit" > écrire

On utilise le verbe فَعَلَ - fa3ala = "il fit" > faire comme paradigme de toutes les formes verbales possibles.

Il varie donc selon les temps, les modes et les voix auxquelles il se conjugue. Il varie aussi selon qu'il est de forme simple ou dérivée. Certains verbes de racines irrégulières (consonnes redoublées, avec hamza ou semi-consonne و ou ي) ont une conjugaison qui varie selon des règles spécifiques. La plupart des verbes sont trilitères, mais certains sont quadrilatères

La conjugaison des verbes dépend de plusieurs facteurs :

- Le temps (accompli, inaccompli).
- Le nombre du sujet (singulier, duel, pluriel).
- Le genre du sujet (masculin, féminin).
- La personne (première, deuxième et troisième)
- Le mode (actif, passif)

Place du verbe dans la phrase :

En principe, il vaut mieux placer le verbe en tête de phrase. $yaktubakarîmunrisâlatan$ - $يَكْتُبُ$
 $كَرِيمٌ رِسَالَةً$ = Karim écrit une lettre.

Si le sujet est indéterminé, il doit suivre le verbe.

En grammaire arabe, une proposition verbale est une proposition qui commence par un verbe.

Le verbe peut cependant aussi précéder le sujet : $karîmunyaktuburisâlatan$ - $كَرِيمٌ يَكْتُبُ رِسَالَةً$ =
Karim écrit une lettre.

Si le verbe précède son ou ses sujets, il s'accorde en genre, mais reste au singulier. Il ne se met au duel ou au pluriel que si le ou les sujet(s) le précèdent, s'ils sont sous-entendus ou remplacés par des pronoms personnels. Lorsque les sujets ne sont pas de même genre, le verbe se conjugue à sa forme non marquée, le masculin.

3.2. Le nom (الاسم) :

Le **nom** est un mot qui peut varier en genre et en nombre, et qui peut se décliner. Il désigne le plus souvent un être, un objet, une notion, ...Il peut être un nom commun, un nom propre, un adjectif qualificatif, un participe, un pronom démonstratif ou personnel.

Il est déterminé (précédé de l'article ou suivi d'un complément du nom), ou pas

3.2. 1 le nombre :

En arabe, il existe trois nombres : singulier, duel et le pluriel :

a) Singulier - مُفْرَد :

Le singulier est la forme non marquée de la plupart des noms. Il existe des formes de singulier pour la conjugaison des verbes à toutes les personnes. L'emploi d'un nom indéterminé et au singulier est la traduction courante de l'article "un(e)" du français : دِينَارٌ - dînârûn = un dinar.

b) duel - مُثنَى:

En arabe, le duel indique que l'on parle de deux éléments. Le duel du nom (ou de l'adjectif, du participe, ...) se forme en ajoutant au singulier le suffixe 'âni - وَنٌ au nominatif et ayni - يْنِ au cas direct et indirect. Le و disparaît si le nom est déterminé par un complément du nom ou par un pronom affixe. [14]

c) pluriel - جَمْع :

En arabe, le pluriel indique que l'on parle de trois éléments ou plus (deux éléments étant exprimés par le duel). Le pluriel est marqué par différentes formes pour les noms, les pronoms et les verbes. Cependant, quand un nom au pluriel désigne autre chose que des êtres humains, tout ce qui s'y rapporte (verbe (placé avant ou après le sujet), pronom personnel, démonstratif, relatif, épithète, attribut, ...) se met généralement au féminin singulier. C'est ce qui s'appelle le "pluriel d'êtres non doués de raison" (animaux, choses concrètes ou abstraites). Les formes de pluriel ne s'utilisent de fait que pour les êtres humains. Pour les noms eux-mêmes, il existe deux catégories de **marques du pluriel** : les pluriels externes et les pluriels internes.

Seul l'usage permet de connaître le pluriel d'un nom, en général. Il est indiqué dans le dictionnaire et doit être appris. [14]

d) Les pluriels externes :

Consistent à ajouter au singulier un affixe du pluriel.

Les principales catégories de mots qui prennent un pluriel externe sont les participes (actifs ou passifs), les noms d'artisans (فَعَال - fa33âl), les noms-adjectifs de relation, formés avec le suffixe iyy - يِّ.

- Le pluriel externe masculin :

se forme en ajoutant au masculin singulier le suffixe ûna - وَنَ au nominatif et îna - يِنَ au cas direct et indirect. Dans le dictionnaire, il est indiqué par وَن.

Le وَن disparaît si le nom est déterminé par un complément du nom ou par un pronom affixe.

- Le pluriel externe féminin :

se forme en ajoutant au singulier le suffixe âtun - تَات au nominatif et âtin - تَاتِ au cas direct et indirect. Dans le dictionnaire, il est indiqué par تَات.

Si le singulier est terminé par un tâ' marbûTa, celui-ci est préalablement supprimé.

e) Les pluriels internes - جَمْعُ التَّكْسِيرِ :

Consistent à modifier la structure interne du singulier. Il en existe de nombreux types.

Un même mot peut avoir plusieurs pluriels. Chemises = قَمِيصَاتُ / قَمِيصَاتُ / قَمِيصَاتُ

* Si un adjectif a un pluriel interne, celui-ci est identique pour le masculin et le féminin. [14]

3.3. Les particules (الحروف)

Ce sont principalement les mots outils comme les conjonctions de coordination et de subordination.

Les particules sont classées selon leur sémantique et leur fonction dans la phrase, on en distingue plusieurs types (introduction, explication, conséquence, ...). Elles jouent un rôle important dans l'interprétation de la phrase, elles servent à situer des faits ou des objets par rapport au temps ou au lieu, elles jouent également un rôle clé dans la cohérence et l'enchaînement d'un texte.

Comme exemple de particules qui désignent un temps قبل, فـ, نـ pendant, avant, après, un lieu حيث, أو, ou de référence هؤلاء,.... ces particules seront très utiles pour notre traitement, elles peuvent avoir des préfixes et suffixes ce qui rajoute une complexité quant à leur identification . [10]

4. La phrase arabe

La phrase est une petite unité qui peut donner un message complète au récepteur. La leçon de grammaire se base dans toute sa recherche sur la phrase et sa structure ; ses énormes et sa nature.

4.1 Définition de la phrase arabe

C'est un énoncé composé au minimum de deux mots ou plus qui engendre un sens compréhensible. La phrase arabe est basée sur le fondement de la prédication (الاستتار), dont deux objets élémentaires qu'on définit (المسند) et elmusnad-ilayh (المسند اليه).

La phrase est une petite unité qui peut donner un message complet au récepteur. La grammaire, dans toute recherche, traite la structure de la phrase, sa nature et ses normes. [15]

5. classification de la phrase par rapport à ces éléments

Pour la phrase arabe on peut distinguer deux types de phrase des phrases en forme naturel sans utilisation de mécanismes d'avancement dit en Arabe Et-Takdime. Donc de ce fait on se retrouve avec deux formes à savoir le cas de l'avancement dit non-fondamentaliste et le cas normal dit fondamentaliste.

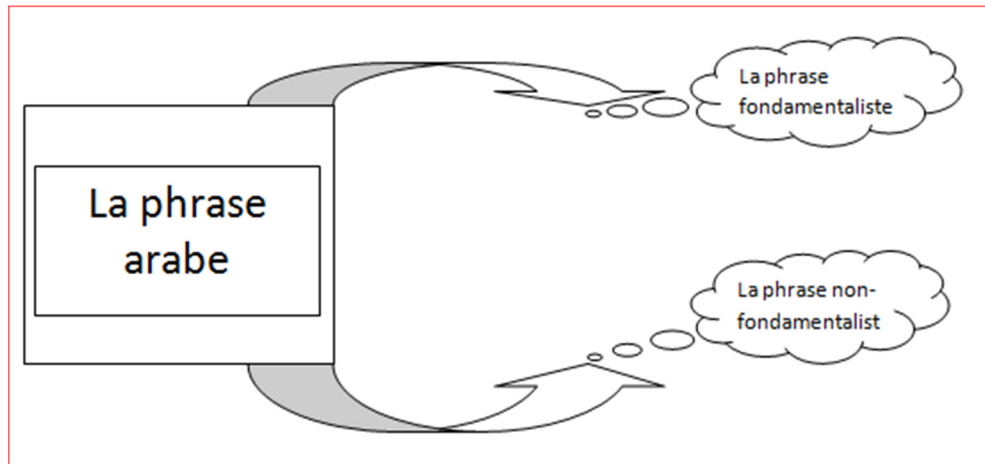


Figure.10 les types de phrase arabe

5.1 Les phrase fondamentaliste (الجملة الأصولية)

Un orateur peut faire des jugements sur un ensemble de mots successif, et jugé s'ils composent une phrase correcte ou une phrase incorrecte dans sa langue, La phrase correcte est appelée phrase fondamentaliste (c'est-à-dire, la phrase origines de la Langue) et la phrase incorrecte est appelée phrase non origine (non-fondamentaliste).

La règle ne se limite pas au fondamentaliste de la phrase, Il prévoit des degrés divers en termes de considération de la phrase, car les phrases non-fondamentaliste varient en fonction de l'étendue de leur écart par rapport à la grammaire : Le degré de non-fondamentaliste est lié au niveau auquel la règle à laquelle la phrase déviée appartienne[10]

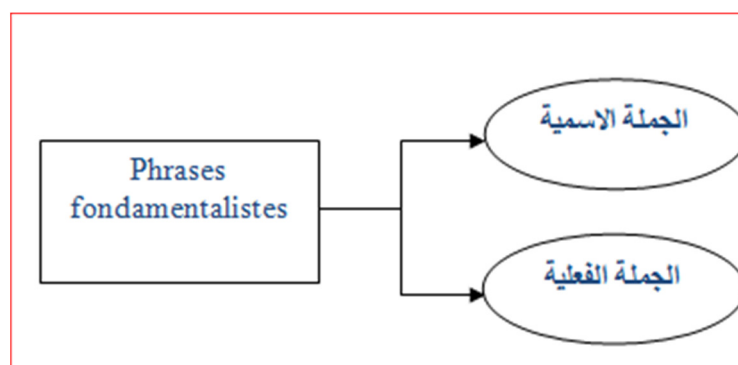


Figure 11 : Répartition formelle de la phrase fondamentaliste.

5.2 Les phrase non fondamentaliste (الجملة غير الأصولية)

La langue arabe comme toute langue est régie par un système de structuration.

Ce système syntaxique permet de produire deux types de forme syntaxique dont l'un est fondamentaliste et l'autre est non fondamentaliste. Ce dernier est l'objet de notre présent travail de recherche.

La phrase non fondamentaliste figure en deux forme :

- phrase nominale (جملة اسمية)
- Phrase verbale (جملة فعلية)

Le schéma de la figure suivante montre les éléments de la phrase et les changements de position dans la phrase qui peuvent surgir.

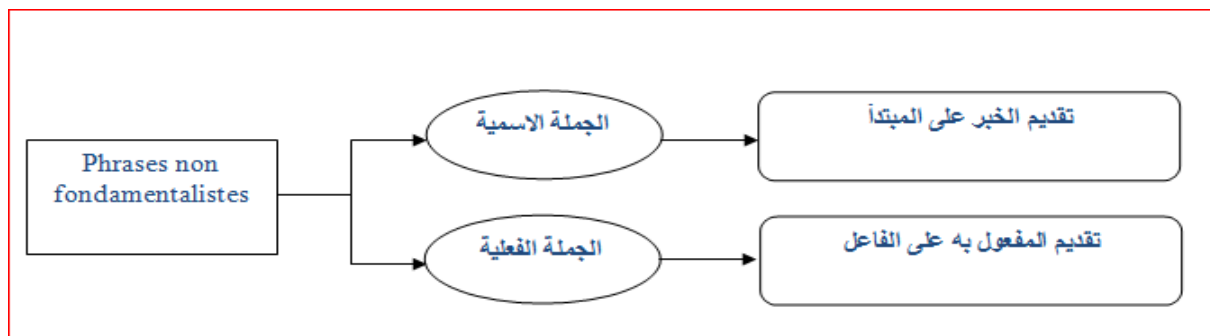


Figure 12 : Répartition formelle de la phrase non fondamentaliste.

5.2.1 La phrase nominale

- تقديم الخبر على المبتدأ

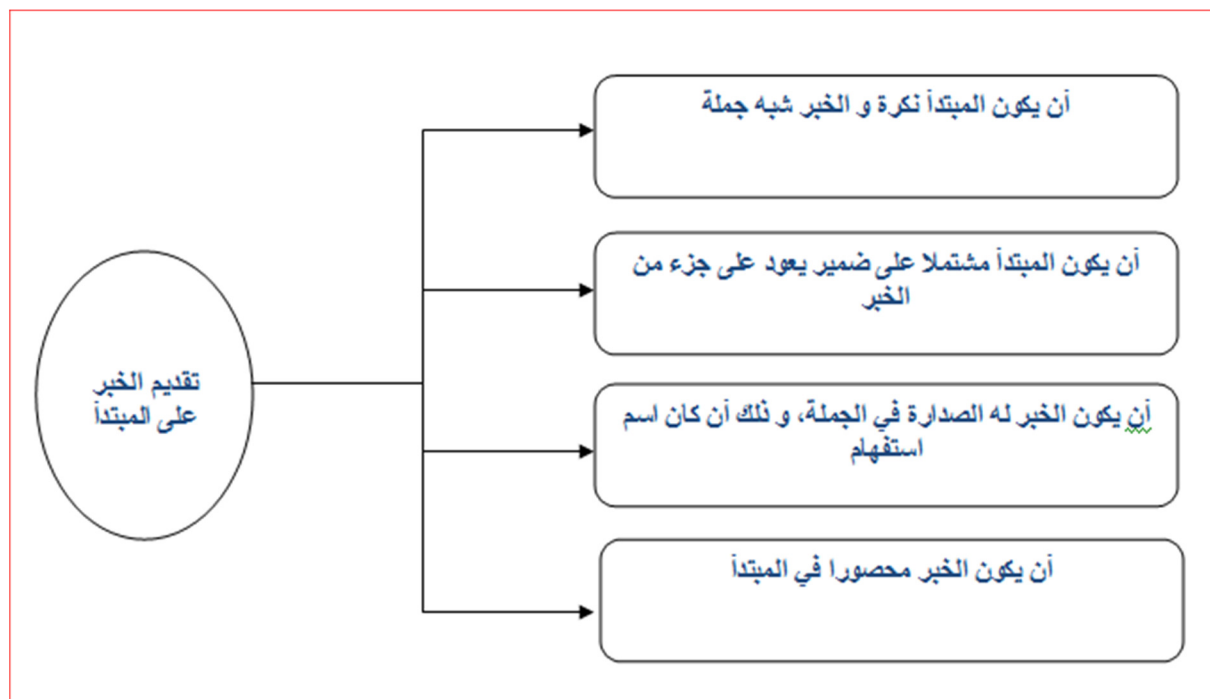


Figure 13 : L'ensemble des critères auquel régit la phrase nominale non fondamentaliste.[16]

5.2.2 La phrase verbale

(1)تقديم المفعول به على الفاعل

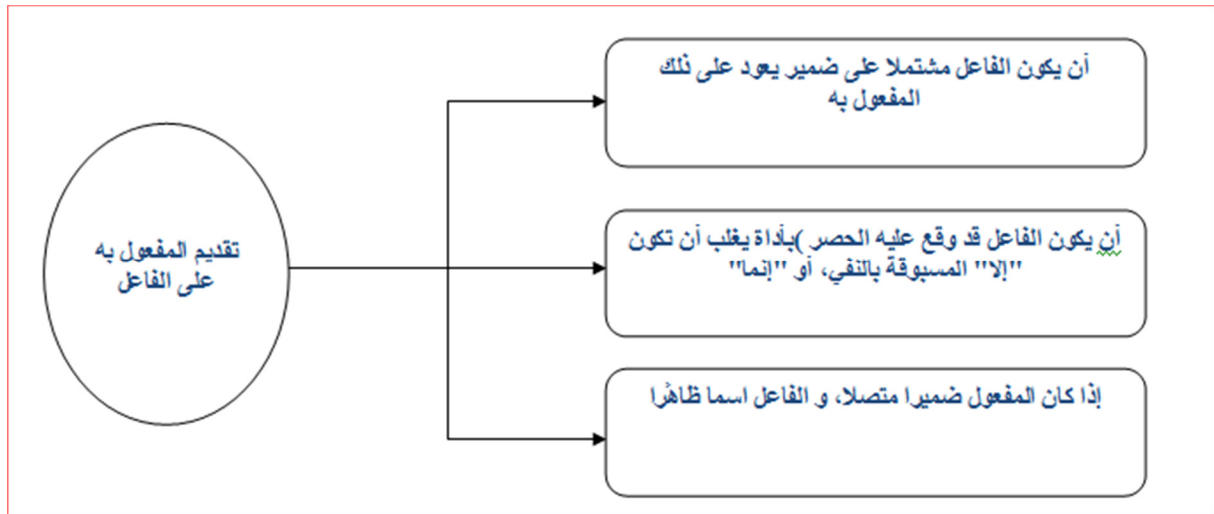


Figure 14 : L'ensemble des critères auquel régit la phrase verbale non fondamentaliste.

Conclusion

En bref, notre apport par rapport au travail déjà fait en 2018 se résume en la prise en compte de la phrase non fondamentaliste en commençant par leur intégration comme règle a la grammaire proposée pour les phrases fondamentaliste jusqu'à la prise en compte pour la segmentation et le Parsing.

Chapitre 03

La grammaire arabe



1. La grammaire arabe

Le mot dans la langue arabe peut être considéré comme une occurrence un nœud dans un graphe de dépendance représentant l'organisation hiérarchique, des classes existantes pour les rôles syntaxiques comme suit : le mot est appelé en arabe "Kalima", ce modèle comprend les trois nœuds qui sont ("فعل " verbe), ("اسم"Nom), et ("حرف " particule). Et tous ces termes ont une dépendance avec d'autres pour créer la hiérarchie des concepts



Figure.15 le principe de la hiérarchisation des classes de mots arabe

2.Le formalisme utilisé

SHA (Le *Subsumption hierarchical attribute*) : est un formalisme en cours d'évaluation, il permet de rendre la structure de représentation une symbolisation manipulable et significative en même temps. Après une représentation sur un graph de dépendance basé sur la relation de subsumption

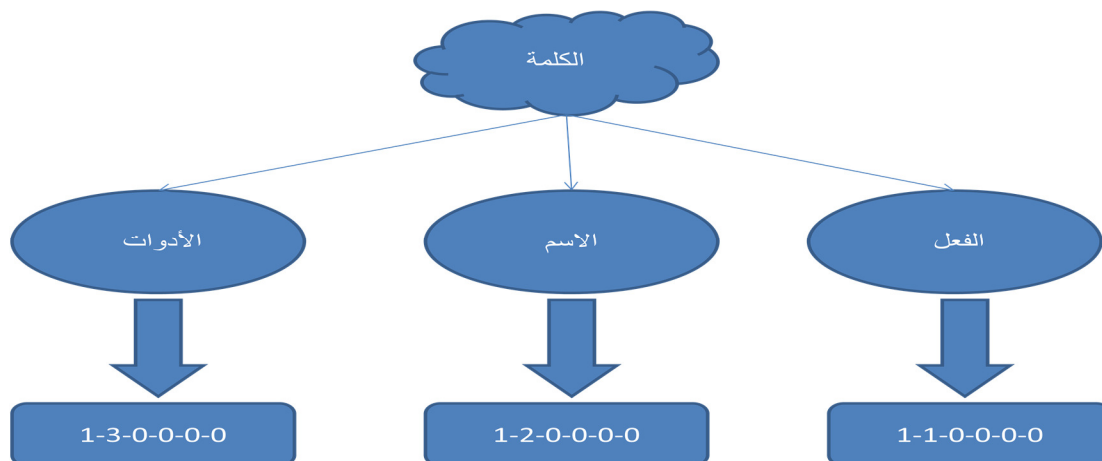


Figure16 : classification du mot arabe

Le verbe est en relation avec ces héritiers comme le montre la figure suivante :

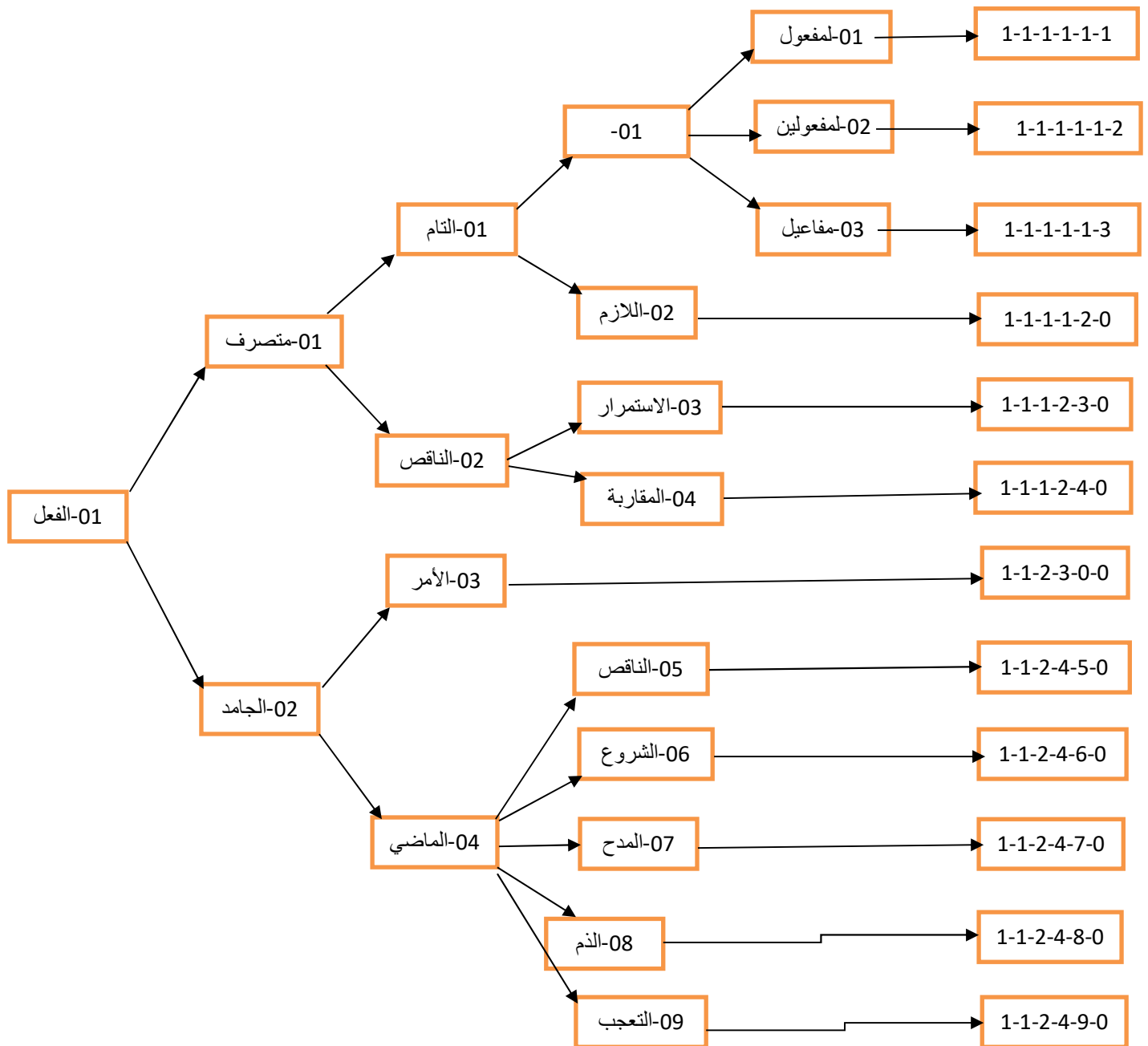


Figure.17 : Classification du verbe

De la même façon on peut décrire le nom comme dans la figure qui suit

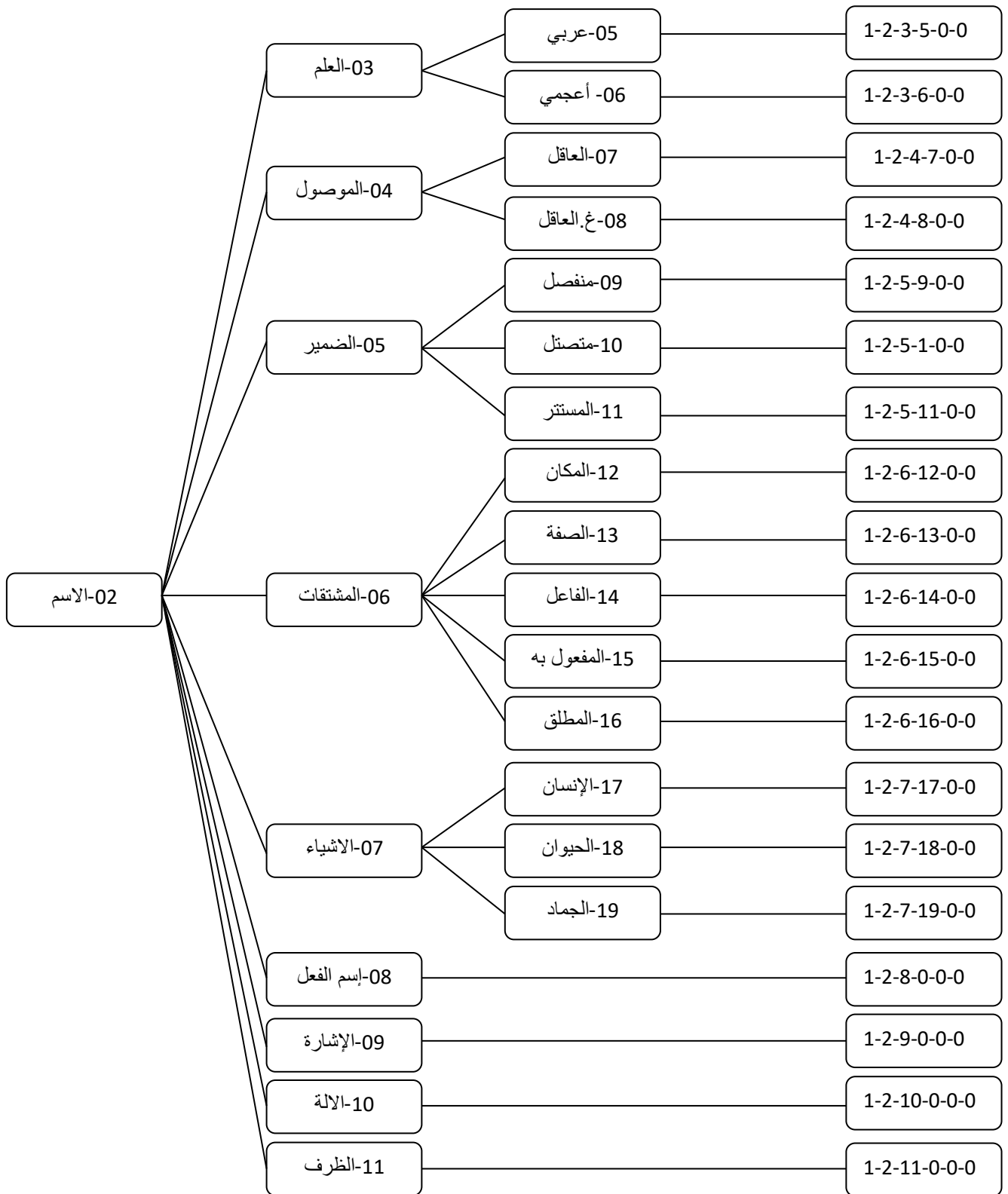


Figure.17 : Classification du nom

De la même façon en peut décrire la particule comme dans la figure qui suit

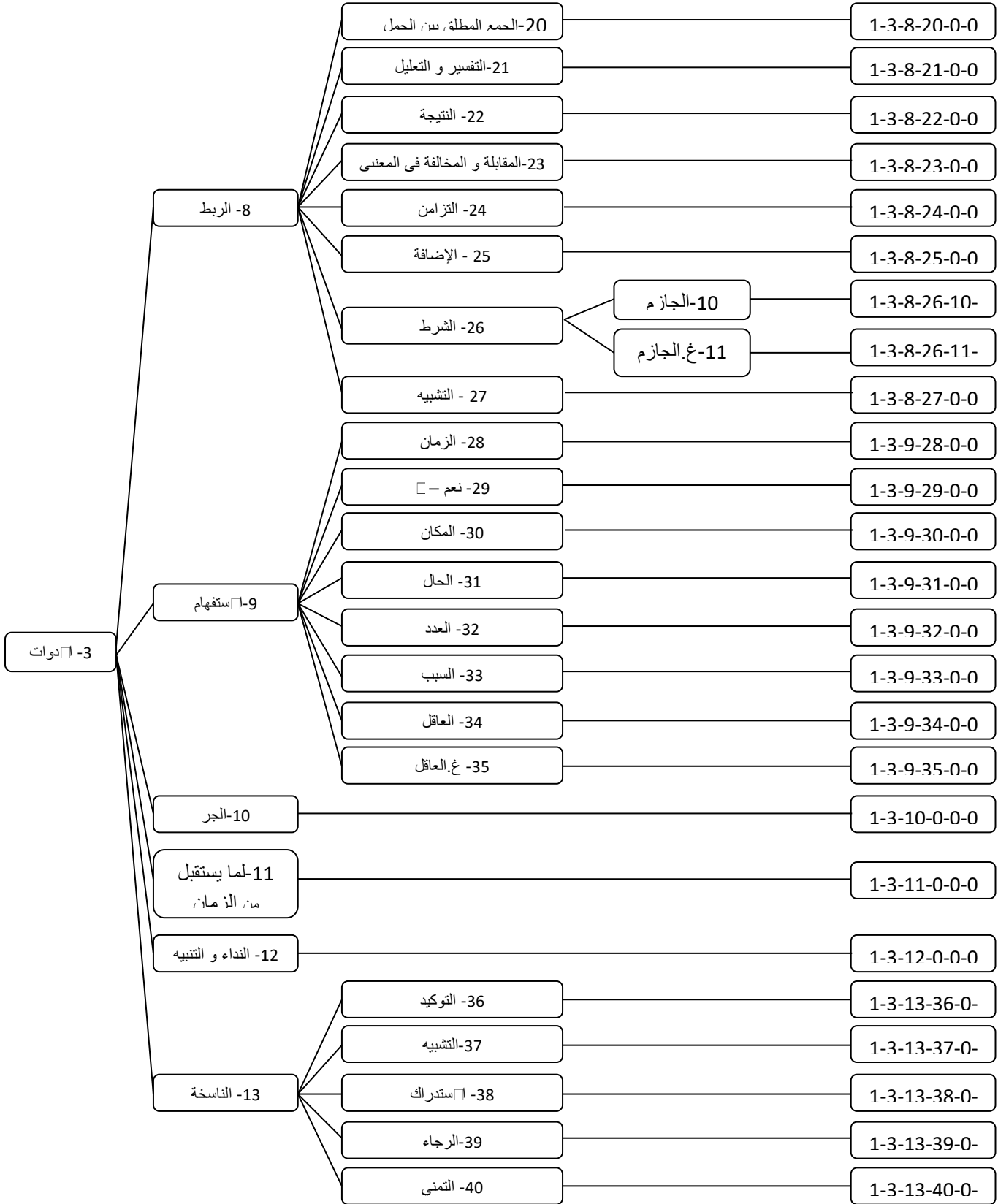


Figure.19 : Classification de la particule (الحروف او الأدوات) arabe.

3. Architecture globale du système morphosyntaxique

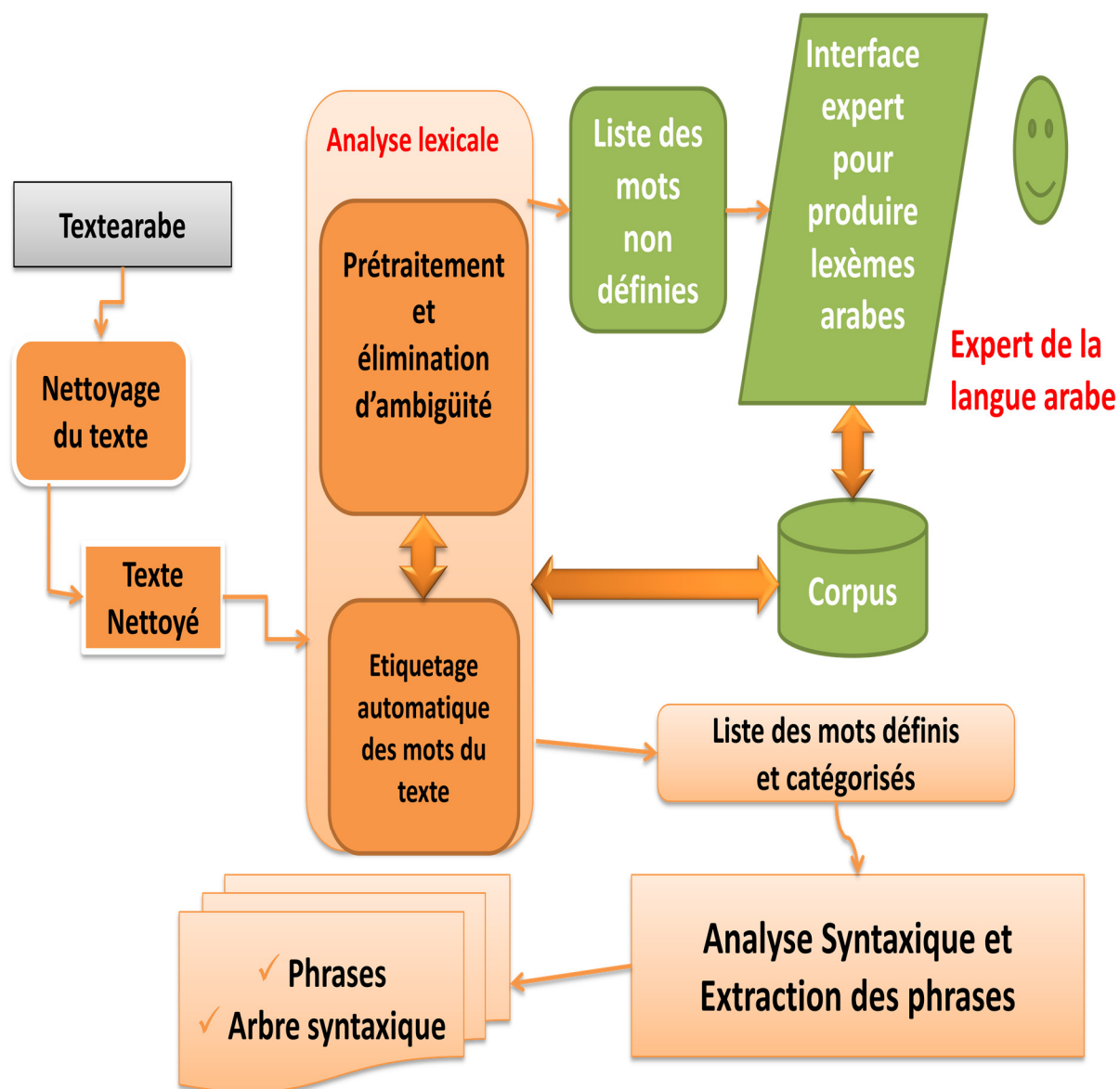


Figure 20 : Architecture globale du système morphosyntaxique.

Le système présenté dans ce présent travail a pour objet la segmentation des textes arabe. Cette segmentation est du type texte-phrases et aussi vers la liste des mots définis et étiquetés par leurs vecteurs appropriés, obtenus lors de l'étape de l'analyse lexicale dont chacun de ces vecteurs renvoie à une catégorie de mots arabes, la chose qui nous va permettre de savoir quel rôle syntaxique peut jouer ce mot pour appliquer les normes de la grammaire et délimiter chaque phrase dans le texte donné.

3.1 Les démarche du système

Le texte concerné par l'analyse de ce système est considéré comme un texte soigné conforme à la grammaire de l'arabe, on a proposé l'exécution suivante :

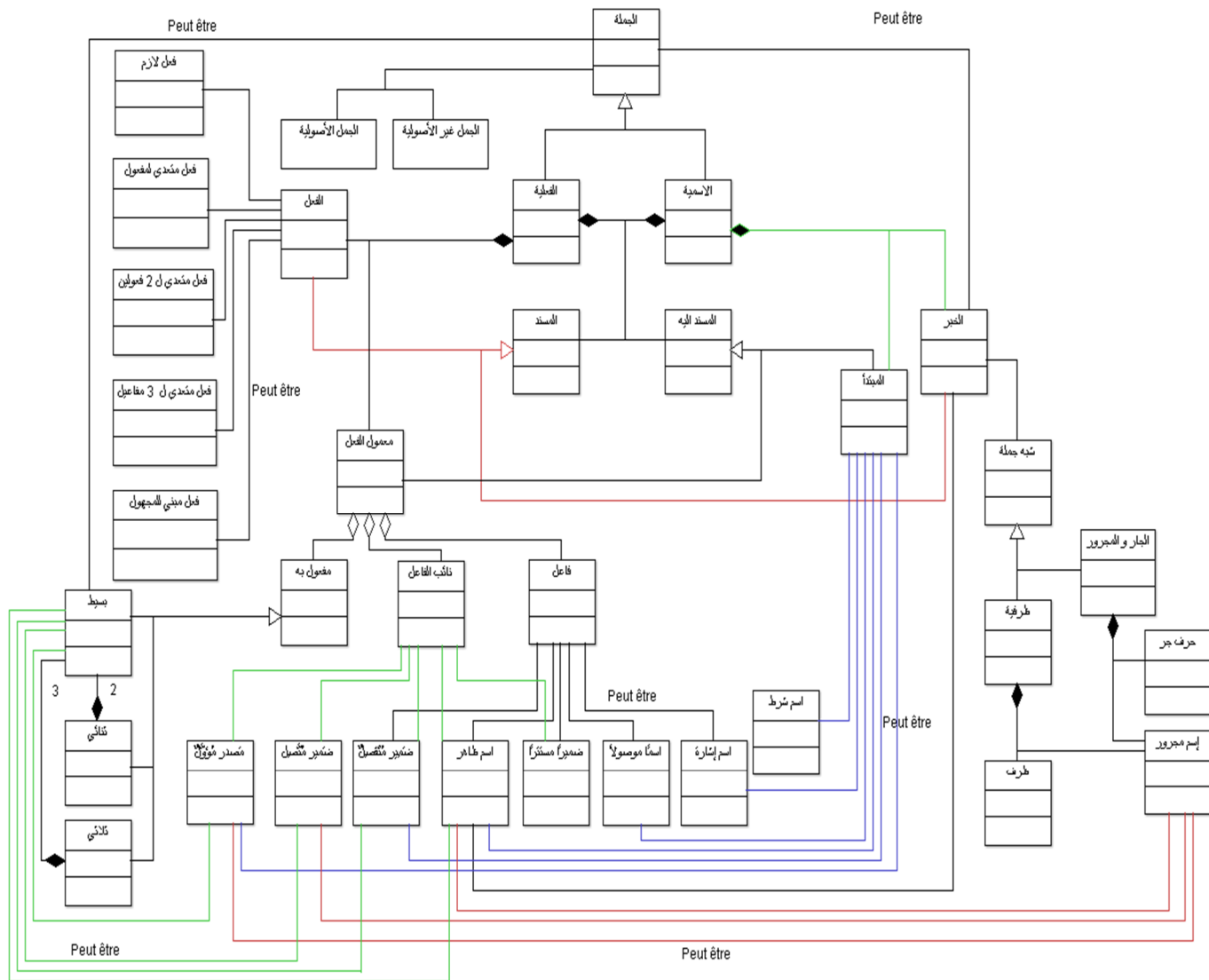


Figure 21 : Diagramme de class modélisant la phrase Arabe

5. Grammaire

Les outils mathématiques les plus utilisés dans la modélisation linguistique sont de type logico-algébrique. La première impulsion de modélisation des langues en termes algébriques vient sans doute de Zellig Harris (dont Chomsky fut un élève), qui a développé un modèle transformationnel dans lequel les phrases sont obtenues à partir de phrases noyaux subissant une suite de manipulations algébriques. Mais très vite c'est Chomsky lui-même qui assoit définitivement la domination des modèles algébriques en syntaxe avec le programme générativiste. La plupart des modèles syntaxiques qui suivront emprunteront les caractéristiques de base du formalisme chomskyen, en particulier la définition d'une grammaire comme un ensemble de règles de production s'appliquant à un vocabulaire constitué du lexique (vocabulaire terminal) et de symboles désignant les différents types de syntagmes. La suite des règles de réécriture utilisées pour engendrer une phrase peut être

symbolisée par un arbre qui fournit donc une représentation équivalente de la structure Syntaxique

5.1 Construction de la grammaire propre

5.1.1 Définition

Soit une grammaire $G = (N, T, R, S)$ tel que :

- T est le vocabulaire terminal.
- N est le vocabulaire non terminal
- R est un ensemble de règles de production
- $S \in N$ est le symbole de départ ou axiome.

Grammaire G est dite propre lorsque :

- si $X \rightarrow \varepsilon \in R$ alors $X = S$ et S ne figure pas en partie droite de règle.
- $\forall X \in N, L(X, G) = L(G, X) \neq \emptyset$
- G n'admet aucune règle de la forme $X \rightarrow Y$ avec $Y \in N$.
- $\forall X \in V$, il existe un mot $w \in V^+$ contenant le symbole X et tel que $S \xrightarrow{*} w$ [14].

5.1.2 Théorème

Tout langage hors-contexte non vide admet une grammaire hors-contexte propre [14].

La preuve de ce théorème tient dans le fait qu'on peut « nettoyer » une grammaire hors-contexte. [14]

5.1.3 Nettoyage de la grammaire

Le nettoyage de grammaire se fait en 4 étapes (à réaliser dans cet ordre):

- supprimer les symboles inutiles (optionnel, pour réduire la taille), en supprimant d'abord les improductifs, puis les inaccessibles,
- rendre la grammaire ε -libre
- supprimer les productions simple
- supprimer les symboles inutiles (en se limitant aux inaccessibles si on a déjà supprimé les improductifs).

a- Étape 1

La grammaire ne contient pas des symboles inutiles.

b- Étape 2

Rendre la grammaire ε -libre, Création d'un nouvel axiome : S' puis supprimer les

ϵ -règles, chose qui n'existe pas dans la grammaire courante.

$S' \rightarrow A / \epsilon$

$A \rightarrow B / C / Z / M / P / L / O ;$

$B \rightarrow aD / aDF / bR / CDEE / dDEEE / eG / fGE / gGEE / hD / hDE / hF / iDH / nIJ / jIJ / kIJ / lIJ / mIJ ;$

$R \rightarrow DE / EDP / \zeta p D ;$

$D \rightarrow o / p / q / r / s ;$

$E \rightarrow o / p / t / K / A ;$

$G \rightarrow o / p / t / K / A ;$

$K \rightarrow \alpha \zeta ;$

$P \rightarrow wBB ;$

$H \rightarrow o / q ;$

$C \rightarrow IJ / L / O / F\$T / v I ;$

$T \rightarrow p / \epsilon ;$

$M \rightarrow \lambda A ;$

$I \rightarrow o / k / v / r / q / w / t ;$

$J \rightarrow o / F / C / B ;$

$F \rightarrow \omega N / xN ;$

$N \rightarrow o / k / p ;$

$L \rightarrow \beta IJ / \beta FI ;$

$O \rightarrow \gamma o / \gamma oo ;$

$Z \rightarrow J\&I / bE\&D / A ;$

c- Etape 3

Supprimer les productions simples, Il s'agit de supprimer toutes les productions de forme :

$X' \rightarrow Y'$ tel que X', Y' sont des vocabulaires non terminales.

d- Etape 4

La vérification des règles inutiles (inaccessible), nous a conçu la forme suivante :

- $S' \rightarrow aD / aDF / bDE / cDEE / dDEEE / eG / fGE / gGEE / hD / hDE / hF / iDH / jIJ / kIJ / lIJ / mIJ / nIJ / IJ / \beta IJ / \beta FJ / \gamma o / \gamma oo / \varepsilon$
- $A \rightarrow aD / aDF / bDE / cDEE / dDEEE / eG / fGE / gGEE / hD / hDE / hF / iDH / jIJ / kIJ / lIJ / mIJ / nIJ / IJ / \beta IJ / \beta FJ / \gamma o / \gamma oo$
- $D \rightarrow o / p / q / r / s$
- $E \rightarrow o / p / t / \alpha z / aD / aDF / bDE / cDEE / dDEEE / eG / fGE / gGE / hD / hDE / hF / iDH / jIJ / kIJ / lIJ / mIJ / nIJ / IJ / \beta IJ / \beta FJ / \gamma o / \gamma oo$
- $F \rightarrow \omega N / x N$
- $G \rightarrow o / p / t / \alpha z / aD / aDF / bDE / cDEE / dDEEE / eG / fGE / gG / hD / hDE / hF / iDH / jIJ / kIJ / lIJ / mIJ / nIJ / IJ / \beta IJ / \beta FJ / \gamma o / \gamma oo$
- $H \rightarrow o / q$
- $I \rightarrow o / \alpha z / v / r / q / w / t$
- $J \rightarrow o / F / aD / aDF / bDE / cDEE / dDEEE / eG / fGE / gGEE / hD / hDE / hF / iDH / jIJ / kIJ / lIJ / mIJ / nIJ / IJ / \beta IJ / \beta FJ / \gamma o / \gamma oo$
- $M \rightarrow \lambda A$
- $N \rightarrow o / \alpha z / p$

Cette dernière étapes nous a permet de finaliser avec une grammaire propre.

5.1.4 Vérification de la régularité de la grammaire propre

Analysant cette grammaire, nous a paru que la majorité de ses règles acceptent la définition de la grammaire de type3 dite régulier, par contre, il existe certaines règles qui ne la respectent pas, dont on va les ajustés pour essayer d'aboutir à une grammaire régulier.

Ces sont les suivant :

$S' \rightarrow v1 DS / v2 P' / v3 Q' / v4 R / v5 G' / v6 T' / v7 U / v8 M' / v9 V / v10 W / v11 W / v12 W / v13 W / v14 W / W / O2 N' / O4 Y / O5 S' / X N' / z' Z' / wBB$

Pour voire mieux le problème de ces règles on procède au décodage de ces derniers :

النص (S') ← فعل متعدي 1 فاعل مفعول /

فعل لازم فاعلشبهه جملة /

فعل متعدي 2 فاعل مفعول مفعول /

فعل متعدي 3 فاعل مفعول مفعول مفعول /

فعل مبني للمجهول متعدي 2 نائب فاعل مفعول /

فعل مبني للمجهول متعدي 3 نائب فاعل مفعول مفعول /

فعل أمر فاعل مفعول /

فعل المدح الذم فاعل المخصوص المدح الذم /

الفعل الناقص المبتدأ الخبر /

فعل الاستمرار المبتدأ الخبر /

فعل المقاربة المبتدأ الخبر /

فعل الرجاء المبتدأ الخبر /

فعل الشروع المبتدأ الخبر /

المبتدأ الخبر

On remarque bien que le problème de ces règles est apparu dans la succession directe de plusieurs non terminaux comme il est illustré dans la règle décodée. Afin de résoudre ce problème on a proposé la solution suivante, développée en deux étapes :

- On a distingué que les non terminaux : **فاعل** et **المبتدأ** peuvent être remplacés par leurs dérivations dans tous les règles, car elle produit que des terminaux , sont montrées comme suit:

— **فاع** ————— ل ← اسم ظاهر / ضمير متصل / اسم موصول / اسم إشارة / ضمير مستتر
 — **المبتدأ** ————— دأ ← اسم ظاهر / مصدر مؤولاً / اسم استفهام / اسم إشارة / اسم موصول / اسم شرط / ضميراً منفصل

On a choisi la règle suivante comme un exemple explicatif de cette opération :

(1) النص (S') ← المبتدأ الخبر

En effectuant la démarche de remplacement sur la règle 1, on obtient le résultat suivant :

- النص (S') ← اسم ظاهر الخبر /

مصدرًا مؤولاً الخبر/

اسم استفهام الخبر/

اسم إشارة الخبر/

اسم موصول الخبر/

اسم شرط الخبر/

ضميرًا منفصل الخبر/

- Par contre les non terminaux « مفعول » et « نائب الفاعل » leurs dérivations sont sous la forme de :

$$A \rightarrow \alpha B$$

$$B \rightarrow \alpha \Sigma$$

Tel que : A, B \in VN ;

$$\alpha \in V^*T \quad \text{et} \quad \Sigma \in V^*N$$

Quant à ce type de problème, on a aperçu que la résolution de ce dernier nous mène vers une grammaire de type 3. la chose que nous a motivé pour aller chercher plus dans les différentes représentations du compliment « المفعول » avec les verbes transitifs a deux ou trois compliments. On a repéré l'idée affirmée par les grammairiens, qui disent :

- le premier complément est un sujet par ça sémantique
«يقول النحاة ان المفعول الأول فاعل في المعنى» [10]
- le plus courant est qu'il possible de remplacer le sujet par le premier compliment si le verbe été transitif a plusieurs compliment.
«و ان الأصل و الكثير أيضا ان يقام مقام الفاعل المفعول به الأول اذا كان الفعل متعديا لأكثر من مفعول به» [10]

Ce qui signifie que les cas dont le verbe est transitif a deux compliments, le premier complément ne sort pas du cas du sujet et ça nous a permis d'exclure le cas du deuxième complément qui se présente sous forme de phrase.

Et comme le «نائب فاعل» accepte lui aussi tous les cas du compliment, alors il subit les même variations que le compliment.

Pour faire face à cette nouvelle parvenue dans la grammaire, on a fractionnée les règles du «مفعول» et «نائب فاعل».

- مفعول ————— اول — اسم ظاهر / ضمير متصل / ضمير مُنْفَصِل / مصدر مُؤَوَّل /
الجملة

- $E \rightarrow o/p/t/\alpha z/aD/aDF/bDE/cDEE/dDEEE/eG/fGE/gGEE/hD/hDE/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o$

Remplacer par :

- **مفعول** ← اسم ظاهر / ضمير متصل / ضمير منفصل / مصدر مؤول
- $E \rightarrow o/p/t/\alpha z$
- **مفعول مركب** ← اسم ظاهر / ضمير متصل / ضمير منفصل / مصدر مؤول / الجملة
- $E' \rightarrow o/p/t/\alpha z/aD/aDF/bDE'/cDEE'/dDEEE'/eG'/fGE'/gGEE'/hD/hDE'/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o$

Et aussi :

- **نائب فاعل** ← اسم ظاهر / ضمير متصل / ضمير منفصل / مصدر مؤول / الجملة
- $G \rightarrow o/p/t/\alpha z/aD/aDF/bDE/cDEE/dDEEE/eG/fGE/gGEE/hD/hDE/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o$

Est Remplacé par :

- **نائب فاعل** ← اسم ظاهر / ضمير متصل / ضمير منفصل / مصدر مؤول
- $G \rightarrow o/p/t/\alpha z$
- **نائب فاعل مركب** ← اسم ظاهر / ضمير متصل / ضمير منفصل / مصدر مؤول / الجملة
- $G' \rightarrow o/p/t/\alpha z/aD/aDF/bDE'/cDEE'/dDEEE'/eG'/fGE'/gGEE'/hD/hDE'/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o$

On ajoutant les quatre nouvelles règles suscitées, nous donne une nouvelle grammaire propre, représentée comme suit :

- $S' \rightarrow aD/aDF/bDE'/cDEE'/dDEEE'/eG/fGE'/gGEE'/hD/hDE'/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o/\varepsilon$
- $A \rightarrow aD/aDF/bDE'/cDEE'/dDEEE'/eG/fGE'/gGEE'/hD/hDE'/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o$
- $D \rightarrow o/p/q/r/s$
- $E \rightarrow o/p/t/\alpha z$
- $E' \rightarrow o/p/t/\alpha z/aD/aDF/bDE'/cDEE'/dDEEE'/eG'/fGE'/gGEE'/hD/hDE'/hF/iDH/jIJ/kIJ/lIJ/mIJ/nIJ/IJ/\beta IJ/\beta FJ/\gamma o/\gamma o o$
- $G \rightarrow o/p/t/\alpha z$

- $G' \rightarrow o / p / t / \alpha z / a D / a D F / b D E / c D E E' / d D E E E' / e G' / f G E' / g G E E' / h D / h D E' / h F / i D H / j I J / k I J / l I J / m I J / n I J / I J / \beta I J / \beta F J / \gamma o / \gamma o o$
- $F \rightarrow \omega N / x N$
- $H \rightarrow o / q$
- $I \rightarrow o / \alpha z / v / r / q / w / t$
- $J \rightarrow o / F / a D / a D F / b D E' / c D E E' / d D E E E' / e G' / f G E' / g G E E' / h D / h D E' / h F / i D H / j I J / k I J / l I J / m I J / n I J / I J / \beta I J / \beta F J / \gamma o / \gamma o o$
- $M \rightarrow \lambda A$
- $N \rightarrow o / \alpha z / p$

Comme on a déjà cité en haut, cette grammaire respecte toujours la classification de type 2 (grammaire hors-contexte). Mais on parcourant ses règles on a remarqué que cette grammaire a la possibilité de se dérivé vers une grammaire de type 3 (grammaire régulier), la réalisation de cette dérivation nous exige la substitution des non terminaux suivant : E (المفعول), D (الفاعل), J (مبتدأ) et G (نائب الفاعل) par leur terminant (Nb : les non terminant concernés par la substitution produisent que des terminaux), puit éliminé les non terminaux concernés de la grammaire .on prend comme exemple la règle suivante :

- $S' \rightarrow d D E E'$

Remplaçant les non terminant D et E on obtient les règles suivantes :

- $S' \rightarrow d o o o E' / d o o p E' / d o o q E' \dots \text{ect.}$

On voit bien que les nouvelles règles respectent la forme des règles d'une grammaire de type3. Il est opportun de signaler qu'en illustrant toutes les possibilités on soutire un nombre important des règles à partir de celle-ci, qui est équivalent à : $53=125$ règles.

Cet exemple nous a donné une idée sur le nombre totale qu'on va l'obtenir si on allonge cette transformation pour toutes la grammaire, la chose que nous a éloigné de continuer dans cette voie.

On conséquence de ce qu'on a vue précédemment, on peut dire que cette grammaire propre peut être modélisé avec un automate à état fini.

5.2 Construction de la grammaire modifiée

Pour donner une description précise des règles permutantes de construire les phrases correctes d'une langue on a adopté le formalisme chomskyen pour modéliser les règles syntaxiques de la langue arabe et comme première phase on est arrivé à concevoir une grammaire initiale qui

interprète la majorité des règles grammaticales qui représentent un point commun entre les grammairiens arabes.

Cette grammaire était construite à base de la classification grammaticale de la phrase arabe.

La grammaire initiale proposée est définie comme suit :

Le vocabulaire est déterminé par l'ensemble : T =

{ 3، فعل مبنی للمجهول متعدي 2، فعل مبنی للمجهول متعدي 1، فعل مبنی للمجهول متعدي 3 فعل متعدي }
 فعل أمر، فعل المدح الذم، أفعال الاستمرار، أفعال المقاربة، أفعال الرجاء، أفعال الشروع، الفعل الناقص ،
 سم ظاهرا ، ضمير متصل ، اسم موصول ، اسم إشارة ، ضمير مستتر ، ضمير مُنْفَصِلٌ ، اسم استفهام ، اسم شرط
 ، ظرف ، فعل مضارع ، حرف مصدري ، حرف ناسخ ، حرف جر ، حرف نداء ، حرف الاستفهام ، أداة الشرط ،
 { فعل ، اسم ، أداة إستفهام ، أداة إستثناء }

Les catégories syntaxiques sont

{ الجملة ، نائب فاعل ، جملة فعلية ، جملة إسمية ، فاعل ، مفعول ، شبه جملة ،
 المخصوص المدح الذم ، المبتدأ ، الخبر ، مصدر مُوَوَّلٌ ، النسخ ، جملة الاستفهام ، مجرور ، النداء ، جملة الشرط ،
 فاعل ومفعول ، أداة نفي ، مضاف إليه }

Notons ici que la prise en compte des phrases non-fondamentalistes à pousser a la modification de la grammaire déjà définit et on signale que les règles qui figure en couleur et on Italique.

Les règles permettant de combiner des éléments du vocabulaire et des catégories syntaxiques pour engendrer des syntagmes corrects sont les suivantes : où le symbole « → » est une abréviation de « peut être composé de ».

- الجملة ← جملة فعلية / جملة إسمية / جملة الاستفهام / جملة الشرط / النسخ / النداء / أداة النفي Z
- جملة فعلية ← فعل متعدي R1 / فعل متعدي 2 فاعل مفعول مفعول / فعل متعدي 3 فاعل مفعول مفعول مفعول
 مفعول/فعل لازم فاعل/فعل لازم فاعلشبه جملة/فعل مبنی للمجهول متعدي 1
 نائب فاعل/فعل مبنی للمجهول متعدي 2 نائب فاعل مفعول / فعل مبنی للمجهول متعدي 3 نائب فاعل
 مفعول مفعول / فعل أمر فاعل/ فعل أمر فاعل مفعول / فعل أمر شبه جملة / فعل المدح الذم
 فاعل المخصوص المدح الذم/ الفعل الناقص المبتدأ الخبر / فعل الاستمرار المبتدأ الخبر /فعل المقاربة
 المبتدأ الخبر / فعل الرجاء المبتدأ الخبر/ فعل الشروع المبتدأ الخبر
- فاعل ← اسم ظاهرا/ ضمير متصل / اسم موصول/ اسم إشارة / ضمير مستتر
- مفعول ← اسم ظاهرا/ ضمير متصل / ضمير مُنْفَصِلٌ / مصدر مُوَوَّلٌ / الجملة
- نائب فاعل ← اسم ظاهرا/ ضمير متصل / ضمير مُنْفَصِلٌ / مصدر مُوَوَّلٌ/ الجملة
- مصدر مُوَوَّلٌ ← حرف مصدري فعل مضارع

- **جملة الشرط ← أداة الشرط جملة فعلية جملة فعلية**
- **المخصوص_ المدح_ الذم ← اسم ظاهر / اسم موصول**
- **جملة_ إسمية ← المبتدأ_ الخبر / النسخ / النداء/شبه جملة اسم T / أداة استفهام مبتدأ**
- **جملة_ الاستفهام ← حرف_ الاستفهام جملة_ إسمية / حرف_ الاستفهام جملة_ فعلية**
- **المبتدأ ← اسم_ ظاهرا / مصدرًا مؤولًا / اسم_ استفهام / اسم_ إشارة / اسم_ موصول / اسم شرط/ ضميرًا منفصل**
- **الخبر ← اسم_ ظاهرا/شبه جملة/جملة_ إسمية /جملة_ فعلية**
- **شبه جملة ← حرف_ جرمجور/ظرفمجرور**
- **مجرور ← اسم_ ظاهرا / مصدر_ مؤول / ضمير متصل**
- **النسخ ← حرف_ ناسخالمبتدأ_الخبر / حرف_ ناسخشبه جملةالمبتدأ**
- **النداء ← حرف_ نداءاسم_ ظاهرا / حرف_ نداءاسم_ ظاهراسم_ ظاهرا**
- **Z ← خبر أداة الاستثناء مبتدأ / فعل متعدي I مفعول أداة استثناء فاعل / جملة**
- **T ← ضمير متصل / ε**
- **R ← فاعل مفعول به / مفعول به فاعل ضمير متصل / فعل ضمير متصل فاعل**

La catégorie syntaxique de départ est la phrase « الجملة ».

Notons que La grammaire considérée ne prend pas en compte certains aspects de l'arabe, comme les accords de genre et de nombre car ces derniers sont considérés comme des traitements sémantiques.

5.3 Codage de la grammaire modifiée

Afin de faciliter les traitements et l'application des règles nettoyages de la grammaire on a procédé comme suit :

- Associer à chaque catégorie syntaxique une lettre majuscule de l'alphabet français.
- Associer a chaque élément de vocabulaire une lettre minuscule de l'alphabet français plus les symboles {α,β,ω,γ,λ,ç,\$,ù,£}.

Catégorie syntaxique	La lettre associée
الجملة	A
جملة_ فعلية	B
جملة_ إسمية	C
فاعل	D

مفعول	E
شبه_جملة	F
نائب_فاعل	G
المخصوص_المدح_الذم	H
المبتدأ	I
الخبر	J
مصدر_مؤول	K
النسخ	L
جملة_الاستفهام	M
مجرور	N
النداء	O
جملة الشرط	P
فاعل و مفعول	R
اداة النفي	Z`
مضاف اليه	T

Tableau 1 : codage de la catégorie syntaxique.

Vocabulaire	Lettre associé
فعل_لازم	A
فعل_متعدي1	B

فعل_متعدي2	C
فعل_متعدي3	D
فعل_مبني_للمجهول_متعدي1	E
فعل_مبني_للمجهول_متعدي2	F
فعل_مبني_للمجهول_متعدي3	G
فعل_أمر	H
فعل_المدح_الذم	I
أفعال_الاستمرار	J
افعال_المقاربة	k
افعال_الرجاء	l
افعال_الشروع	m
الفعل_الناقص	n
اسم_ظاهرا	o
ضميرٌ متصل	p
اسم_موصول	q
اسم_إشارة	r
ضميرٌ مستتر	s
ضَويِرٌ مُنْفَصِلٌ	t
اسم_استفهام	v

اسم_ شرط	w
ظرف	x
فعل_ مضارع	z
حرف_ مصدرى	α
حرف_ ناسخ	β
حرف_ جر	ω
حرف_ نداء	γ
حرف_ الاستفهام	λ
اداة الشرط	W
فعل	Ç
اسم	\$
اداة الاستفهام	U
اداة الاستثناء	£

Tableau 2 : codage du vocabulaire.

Après substitution des catégories syntaxiques et les éléments du vocabulaire on a obtenu la grammaire suivante :

A → B / C / Z / M / P / L / O ;

B----aD / aDF / bR / CDEE/ dDEEE / eG / fGE / gGEE / hD / hDE / hF / iDH / nIJ / jIJ / kIJ / lIJ / mIJ ;

R----DE /EDP / ç p D ;

D---- o / p / q / r / s ;

E----- o / p / t / K / A ;

G----- o / p / t / K / A ;

K---- αç ;

P-----wBB ;

H-----o / q ;

C-----IJ / L / O / F\$T / v I ;

T----- p / ε ;

M----λA ;

I---o / k / v / r / q / w / t ;

J---- o / F / C / B ;

F----- ωN / xN ;

N----- o / k / p ;

L-----βIJ / βFI ;

O-----γo / γoo ;

Z-----J£I / bE£D / A ;

Cette grammaire est hors contexte. Mais on peut voir que la grammaire contient des productions simples A produit B, On aimerait pouvoir travailler avec des grammaires qui ne présentent pas ce genre de problème. Donc cela nous a pousser à réfléchir à une grammaire propre.

Les règles de production d'analyseur LL(1) nous pousse à réfléchir à défauts qui dérangent l'implémentation à savoir :

- La récursivité à gauche qu'on peut facilement vérifier qu'elle n'existe pas.
- L'intersection des *firsts* des production de la même règle qu'on a éliminée en utilisant la technique de factorisation et la substitution.
- L'intersection Firsts Follow qui n'existe pas après vérification

Ce travail nous a permis d'aboutir à la grammaire suivante :

- الجملة _____ ← جملة فعلية / جملة اسمية / جملة الاستفهام / جملة الشرط / النسخ / النداء / أداة النفي Z
- جملة فعلية ← فعل متعدي R 1 / فعل متعدي 2 فاعل مفعول مفعول / فعل متعدي 3 فاعل مفعول مفعول
- مفعول مفعول / فعل لازم H/ فعل مبني للمجهول متعدي 1 □ نائب فاعل / فعل مبني للمجهول متعدي 2 □ نائب فاعل مفعول / فعل مبني للمجهول متعدي 3 □ نائب فاعل مفعول مفعول / فعل أمر Q / فاعل مفعول / فعل المدح الذم فاعل المخصوص المدح الذم / فعل الاستمرار المبتدأ الخبر / فعل المقاربة المبتدأ الخبر / فعل الرجاء المبتدأ الخبر
- H ← فاعل _____ ل G
- G ← شبه جملة/ع
- Q ← فاعل _____ ل IQ / شبه جملة/ع
- IQ ← مفعول/ع
- فاعل _____ ل ← اسم ظاهراً / ضمير متصل / اسم موصول / اسم إشارة / ضمير مستتر
- مفعول _____ ول ← اسم ظاهراً L / ضمير متصل / ضمير مُنْفَصِلٌ L / مصدر مُوَوَّلٌ L / الجملة الفعلية/ اسم موصول الخبر/ اسم إشارة الخبر/ شبه جملة L □ برغ
- □ نائب فاعل _____ ← مفعول
- مصدر مُوَوَّلٌ ← حرف مصدري فعل مضارع
- جملة الشرط ← أداة الشرط جملة فعلية جملة فعلية
- المخصوص المدح الذم ← اسم ظاهر / اسم موصول / جملة فعلية
- جملة اسمية ← المبتدأ الخبر / شبه جملة اسم T
- المبتدأ _____ دأ ← اسم ظاهراً / مصدرًا مؤولاً / اسم إشارة / اسم موصول / ضميراً منفصلاً
- الخبر _____ ر ← اسم ظاهراً P1 / شبه جملة P2 / جملة فعلية
- P1 ← الخبر/غ
- P2 ← اسم T / غ
- شبه جملة ← حرف جر مجرور / ظرف مجرور
- مجرور ← اسم ظاهراً / مصدر مُوَوَّلٌ / ضمير متصل
- النسخ ← حرف □ اسخ Y
- Y ← المبتدأ الخبر / شبه جملة مبتدأ
- النداء ← حرف □ داء اسم ظاهراً W
- W ← اسم ظاهر/غ
- متعدي 3 فاعل مفعول مفعول مفعول / فعل لازم H / فعل مبني للمجهول متعدي 1 □ نائب فاعل / فعل مبني للمجهول متعدي 2 □ نائب فاعل مفعول / فعل مبني للمجهول متعدي 3 □ نائب فاعل مفعول مفعول / فعل أمر Q / فعل المدح الذم المخصوص المدح الذم / الفعل الناقص / فعل الاستمرار المبتدأ الخبر / فعل المقاربة المبتدأ الخبر / فعل الرجاء / فعل الشرع المبتدأ الخبر
- C ← اسم موصول/ جملة فعلية/ اسم إشارة/ ضميراً منفصلاً/ مصدر مُوَوَّلٌ B1/K
- K ← اسم ظاهراً K1 ضميراً منفصلاً B1

- ←K1 مفعول
- ←B1 فاعل _____ ضمير متصل / اداة استثناء مبتدا
- ←B2 اداة استثناء مبتدا/خبر
- ←T ضمير متصل/غ
- ←R اسم موصول مفعول / اسم إشارة مفعول / ضمير مستتر مفعول/ضميراً منفصل فاعل/ ضمير متصل/ مصدر مؤول فاعل/ ضمير متصل/ اسم ظاهر/ V1 ضمير متصل/ V2
- ←V1 ضميراً منفصل/ مصدر مؤول/ اسم ظاهر V1
- ←V1 ضمير متصل/غ
- ←V2 ضميراً منفصل/ مصدر مؤول/ اسم ظاهر/ V1 جملة الاستفهام ← حرف الاستفهام P
- ←P جملة فعلية /لمبتدا/ X شبه جملة مبتدا
- ←X خبر/غ
- اسم موصول ← موصول جملة فعلية

Conclusion :

Donc c'est cette grammaire qu'on implémenter dans le but de pouvoir générer ou analyser un taux très important de phrase arabe soigné. L'implémentation nous permet de concrétiser un analyseur syntaxique purement arabe et qui permet de faire une explicitation des éléments de la phrase similaire a el-Iêrab.

Chapitre 04

Implémentation

Introduction

Au terme des développements théoriques concernant la contribution et l'apport scientifique au domaine du TALA, nous traiterons dans ce chapitre la phase de l'implémentation qui constitue l'ultime volet de notre mémoire. En ce sens, nous allons aborder en premier la présentation de l'environnement matériel et logiciel de ce présent travail. Ensuite, nous exposerons des échantillons de résultats obtenus et enfin, nous terminerons par montrer des textes arabes et leurs segmentations syntaxiques effectuées en utilisant ce système que nous avons développé.

1. Environnement matériel

En vue de réaliser ce projet, nous avons disposé de :

Un ordinateur de type HP équipé d'un microprocesseur Intel(R) CORE i3, possédant 4,00 Go de RAM et de 500Go d'espace de disque.

2. Environnement logiciel

Avant d'entamer l'implémentation de l'architecture conceptuelle de notre système, nous allons tout d'abord spécifier les outils utilisés qui nous sont utiles et un bon choix de par les avantages qu'ils offrent.

Choix du système d'exploitation

Notre application a été développée sous le système d'exploitation Windows 10, mais comme elle est développée en langage java, elle peut être intégrée dans n'importe quel autre système d'exploitation supportant la machine virtuelle java (Windows 98/00, Linux,...).

Le langage de programmation

Pour le choix de programmation pour notre système, nous avons opté pour le langage JAVA et cela pour de nombreuses raisons :

- JAVA est un langage orienté objet simple, qui réduit le risque des erreurs d'incohérences.
- Il est indépendant de toute plateforme, il est possible d'exécuter des programmes JAVA sur tous les environnements qui possèdent une Java Virtual Machine (JVM).
- Il est doté d'une riche bibliothèque de classes, comprenant la gestion des interfaces graphiques (fenêtres, menus, graphismes, boîtes de dialogue, contrôles), la programmation multithreads (multitâches) et la gestion des exceptions.

- Il permet un accès aux bases de données simplifié soit à travers la passerelle JDBC-ODBC ou à travers un pilote JDBC spécifique au SGBD.
- Il est caractérisé aussi par la réutilisation de son code ainsi que la simplicité de sa mise en œuvre.

NetBeans est un environnement de développement intégré (IDE) pour Java, placé en open source par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License). En plus de Java, NetBeans permet également de supporter différents autres langages, comme Python, C, C++, XML et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages web).

NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X et Open VMS.

NetBeans est lui-même développé en Java, ce qui peut le rendre assez lent et gourmand en ressources mémoires [17].

3. Implémentation

Nous venons de réaliser une application pour segmenter les textes arabes qui intègre aussi une possibilité d'enrichissement du corpus .cette application nous vas permettre l'expérimentation des règles de la grammaire.

3.1 Fenêtre principale

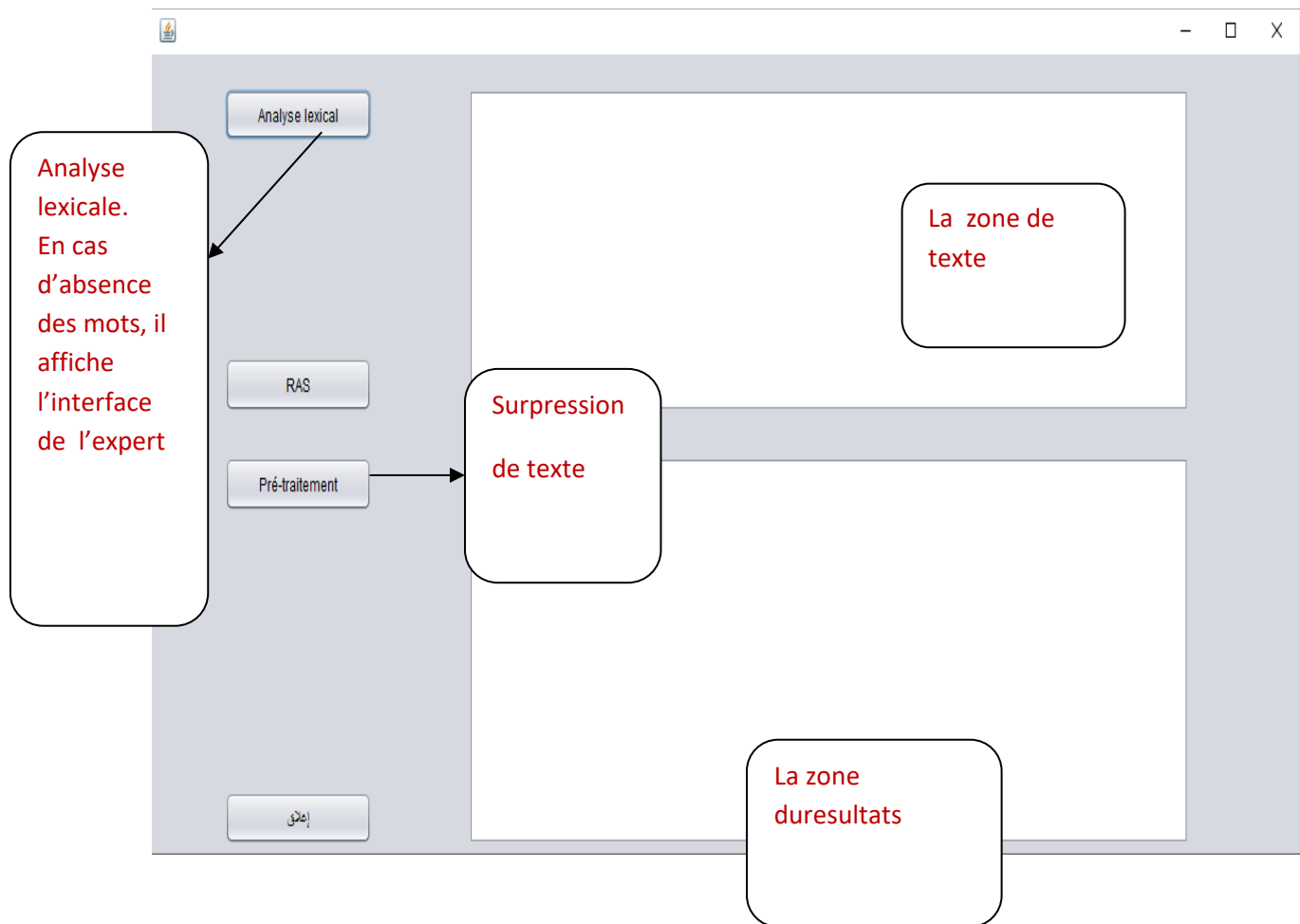


Figure 22 : l'interface principale.

Cette fenêtre offre al utilisateur la possibilité ce traiter un texte arabe syntaxiquement correct a condition d' avoir utilisé des termes existantes dans le corpus lexico-syntaxique. Dans le cas ou un terme n' est pas déjà définie et catégoriser nous devons passer a l' interface expert et proposer les variantes du mot a l' expert pour faire entrer les mots corrects et supprimer les suggestions incorrectes.

3.2 Interface expert



Figure 2 : l'interface experte pour classer les nouveaux mots apparus.

4. Résultat

A l'issue de la finalisation de notre application de traitement automatique de la langue arabe, les résultats obtenus indiquent certains objectifs atteints. Ces derniers se figurent dans l'insertion des textes arabes qui subiront par la suite à une analyse lexicale permettant la segmentation à des unités syntaxique (Phrases). Puis, elle accède à la détermination de l'étiquetage consistant à accorder à chaque mot un vecteur en vue de déterminer sa position en phrase.

4.1 Analyse syntaxique et étiquetage

Texte arabe	كَتَبَ عَلِيُّ الرِّسَالَةَ . حَسِبَ الْمُتَحَدِّثُ الْأَمْرَ سَهْلًا . مَا نَاجِحٌ إِلَّا الْمَجْدُ . يَسُوْقُ السِّيَّارَةَ صَاحِبُهَا . أَذَبْنَا عَلِيَّ .
prétraitement	كَتَبَ عَلِيُّ الرِّسَالَةَ / حَسِبَ الْمُتَحَدِّثُ الْأَمْرَ سَهْلًا / مَا نَاجِحٌ إِلَّا الْمَجْدُ / يَسُوْقُ السِّيَّارَةَ صَاحِبُهَا / أَذَبْنَا عَلِيَّ /
Etiquetages	[1-1-1-1-1-1, 1-2-0-0-0-0, 1-2-0-0-0-0, 0-0-0-0-0-0 , 1-1-1-1-1-2, 1-2-0-0-0-0, 1-2-0-0-0-0, 1-2-0-0-0-0, 0-0-0-0-0-0 , 1-3-0-0-0-0, 1-2-0-0-0-0, 1-3-0-0-0-0, 1-2-0-0-0-0, 0-0-0-0-0-0 , 1-2-3-4-5-6, 1-2-0-0-0-0, 1-2-0-0-0-0, 1-2-8-10-0-0, 0-0-0-0-0-0 , 1-1-1-1-1-1, 1-2-8-10-0-0, 1-2-0-0-0-0, 0-0-0-0-0-0]

Tableau 3 : prétraitement et étiquetage d'un petit texte.

Texte arabe	بينما كان يتحدث غاليلاني بهدوء على طاولته ، بدأت ألوح بنظري في أرجاء الغرفة ، رأيت العديد من الصور والذكريات ،
Prétraitement	بينما كان يتحدث غاليلاني بهدوء على طاولته / بدأت ألوح بنظري في أرجاء الغرفة / رأيت العديد من الصور والذكريات /
Mots non reconnus par le corpus	بدأت بينما ألوح بنظري الصور والذكريات

Tableau 4 : prétraitement d'un petit texte, contient des mots non reconnus par le corpus.

4.2. Segmentation du texte



Figure 24 : Segmentation après un prétraitement réussi.

5. Discussion et commentaire sur les résultats

Les essais sur les formes de grammaire ont donné des résultats très satisfaisants mais sur les phrases fondamentales seulement ce qui nous a permis de constater une efficacité conditionnée par le fait de la non existence de phrases non fondamentales. L'effort que nous avons fourni nous ont permis de surpasser ce conditionnement pour traiter des cas non traitable avant et parfois cause d'erreur de

traitement a cause du manque de prise en charge. A signaler que les modifications sur la grammaire nous ont faciliter la tâche.

Le SHA nous permet de s'appuyer sur une base solide qui a permet de réduire l'ensemble de l'alphabet terminale a un ensemble de 80 classe définis par des vecteurs simple et significatif et leur sémantique était tellement riche à permettre à pouvoir définir une grammaire génératrice manipulable pour maitriser le phénomène syntaxique et a un degré moins la sémantique.

Conclusion :

Les résultats obtenus sur des échantillons de phrase qui représente des exemples de phrase fondamentalistes ou non fondamentaliste permettent de généraliser car c'est la forme de la phrase qui compte.

L'importance de la forme figure du fait que le programme traite la phrase dans ça forme vectorielle codifiée en SHA donc toute les phrases ayant le même vecteur sont prise en considération.

Le traitement des phrases similaire est logiquement similaire donc ce qui et valable pour la forme et valable pour toute les variantes possible ayant cette même forme sauf que cela et conditionné par l'existence du terme dans le corpus lexico-syntaxique.

Conclusion Et Perspectives

Conclusion Et Perspectives

Le développement d'outils de TAL pour les dialectes de l'arabe se heurte à l'absence de ressources pour ces derniers. Comme conséquence d'une situation de diglossie, il existe une variante de l'arabe, l'arabe moderne standard, pour laquelle de nombreuses ressources ont été développées et ont permis de construire des outils de traitement automatique de la langue. Etant donné la proximité des dialectes de l'arabe, avec l'arabe moderne standard, une voie consiste à réaliser une traduction surfacique du dialecte vers l'arabe moderne standard afin de pouvoir utiliser les outils existants pour l'arabe standard. Nous décrivons dans cet mémoire une architecture pour une telle traduction et nous l'évaluons sur les phrases arabe

En perspective nous envisageons un travail de remplissage des ressource utilisé car du coté évaluation nous avons examiné les possibilités de généralisation et nous pouvons en fin confirmé la conformité de cette généralisation car ce n'est pas des teste d'individus ou d'occurrence de phrase ou forme syntaxique c'est des testes sur des forme de phrase ou des classes de phrases qui peuvent représenter la totalité des forme possible a un degré très satisfaisant.

ces travaux commencent à être exploités dans l'équipe dans le cadre d'autres applications que la traduction automatique. dans lesquels le choix des pictogrammes se fait en levant les ambiguïtés lexicales.

Ce type d'application est utile pour permettre d'établir une communication alternative par exemple pour des personnes en situation de polyhandicapé ou pour des personnes

maîtrisant pas ou peu la langue ou bien faire une application pour la traduction par phrase. En plus ,nous avons constaté viron notre travail l'existence d'une ambiguïtés d'origine syntaxique du genre (Isme-dhahire Isme-dhahire)ce cas la que jouer le rôle de (mobtadae khabar) comme il peut représenté (modhafa modhafa_ilyhe) comme il peut représenté d'autre cas alors en perspective ,on envisage la résolution de ce problème par l'intégration d'une représentation des signe de vocalisation ,ajoute à la représentation vectorielle .

En fin, nous espérons que notre travail soit une véritable contribution au domaine de recherche sur le traitement automatique de la langue Arabe

Bibliographie

[1] : *ARTICLE Auteur(s) : Mona Diab, Nizar Habash, Imed Zitouni*, 31 juillet 2017, *RÉF H2017V3 ou bien le site qui contient l'article* : tal-58-3.sciencesconf.org/?forward-action=index&forward-controller=index&lang=fr.consulté le 15/03/2019.

[2] : http://dictionnaire.sensagent.leparisien.fr/Traitement%20automatique%20du%20langage%20naturel/fr-fr/#cite_note-0 consulté le 15/03/2019.

[3] Mohamed Hédi Maâloul , Approche hybride pour le résumé automatique de textes. Application à la langue arabe, Thèse de doctorat en informatique, soutenue le 18 décembre 2012, Université Aix –Marseille.

[4] Siham Boulaknadel, Traitement Automatique des Langues et Recherche d'Information en langue arabe dans un domaine de spécialité : Apport des connaissances morphologiques et syntaxiques pour l'indexation, Thèse de doctorat en informatique, soutenue le 18 octobre 2008, Université de Nantes, France.

[5] Traitement automatique du langage naturel, 29.06.2017, Article, web maketing.

[6]

[7] X. TANNIER, TRAITEMENT AUTOMATIQUE DU LANGAGE NATUREL POUR L'EXTRACTION ET LA RECHERCHE D'INFORMATIONS, Mars 2006, Ecole Nationale Supérieure des Mines de Saint-Etienne Centre G2I 158, Cours Fauriel , Centre Génie Industriel et Informatique (G2I), France. xavier.tannier.free.fr/publications/files/Tannier_RR_NLP.pdf

[8] [11] Mohamed Khaled KHELIF, Web sémantique et Mémoire d expériences sur l analyse du transcriptome, thèse doctorat, 4 avril 2006 , Ecole Doctorale Sciences et Technologies de l'Information et de la Communication (S.T.I.C) , UNIVERSITE de Nice-Sophia Antipolis . tel.archives-ouvertes.fr/tel-00328114/file/These_Mohamed_Khaled_KHELIF.pdf

[9] Brigitte Bigi ,TALN Informatique, avril 2006 .

http://www.lpl-aix.fr/~bigi/Doc/big2006_01_TALN.pdf

[10] BENDEBICHE Rafiq BOUDJENANE Salah (2017-2018). Etiquetage et segmentation des textes arabes basée sur la classification syntaxique des mots l'obtention du diplôme de master en informatique. Département informatique .université de Naama.

[11]Yvon ,Une petite introduction au Traitement Automatique des Langues Naturelles , Support du cour.

[12] Mourad Loukam . http://www.loukam.net/TALN_Chap1.pdf consulté le 18/02/2019

[13] <https://www.espacefrancais.com/la-langue-arabe/> consulté le 17/03/2019

[14] <https://monsieurboursier.fr/arabe/grammaire/index.html> consulté le 17/03/2019 .

[15] http://fll.univ-biskra.dz/images/pdf_revue/pdf_revue_09/naima%20saadia.pdf consulté le 20/3/2019

[16] [<https://www.d-booker.fr/content/72-postgresql>].consulté le 28/03/2019.

[17] <https://www.techno-science.net/definition/5346.html>. Consulté le 28/03/2019

Annexe 01 : tableau des intersections des *firsts* sur la grammaire

	اداة استفهام	اداة شرط	اداة نفي	اداة استثناء	اداة نسخ	اداة نداء	فعل م م 1	فعل م م 2	فعل م م 3	فعل اللازم	فعل م م 1
جملة	جملة الاستفهام	جملة الشرط	Z		النسخ	النداء	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
جملة فعلية						R		فاعل مفعول مفعول	فاعل مفعول مفعول	H	نائب فاعل
جملة اسمية											
جملة استفهام	P										
H											
G											
Q											
Q1							مفعول	مفعول	مفعول	مفعول	مفعول
فاعل											
مفعول							جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
نائب فاعل							جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
مصدر مؤول											
جملة الشرط		جملة فعلية									
مخصوص مدح ذم		جملة فعلية									
مبتدأ											
خبر							جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
P1							خبر	خبر	خبر	خبر	خبر
P2											
شبه جملة											

Annexe

	فعل م م م 2	فعل م م م 3	فعل أمر	فعل المدح و ذم	فعل ناقص	فعل مقارنة	فعل الاستمرار	فعل الشروع	فعل الرجاء
جملة	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
جملة فعلية	نائب_فاعل مفعول	نائب_فاعل مفعول مفعول	Q	مخصوص_مدح_ذم	مبتدأ خبر	مبتدأ خبر	مبتدأ خبر	مبتدأ خبر	مبتدأ خبر
جملة اسمية									
جملة استفهام									
3C									
9									
Q									
Q1	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول
فاعل									
مفعول	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
نائب_فاعل	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
مصدر_مؤول									
جملة_الشرط									
مخصوص_مدح_ذم									
مبتدأ									
خبر	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية	جملة فعلية
P1	خبر	خبر	خبر	خبر	خبر	خبر	خبر	خبر	خبر
P2									
شبه_جملة									

Annexe

	اداة استفهام	اداة شرط	اداة نفي	اداة استثناء	اداة نسخ	اداة نداء	فعل م م 1	فعل م م 2	فعل م م 3	فعل اللازم	فعل م م 1
مجرور											
نسخ					٩٧						
نداء						٦٥					
٦٥								فاعل مفعول مفعول	فاعل مفعول مفعول	٦٢	نائب_فاعل
٩٧							٤				
٤											
٦٢											
٦٢١							مفعول	مفعول	مفعول	مفعول	مفعول
٦٢١				مبتدأ							
٦٢٢				مبتدأ			خير	خير	خير	خير	خير
٩٧											
٩٧											
٩٧١											
٩٧١											
٩٧٢											
٩٧							جملة_فعلية	جملة_فعلية	جملة_فعلية	جملة_فعلية	جملة_فعلية
٩٧							خير	خير	خير	خير	خير
اسم_موصول											

Annexe

	فعل م م م م 2	فعل م م م م 3	فعل أمر	فعل المدح و ذم	فعل ناقص	فعل مقارنة	فعل الاستمرار	فعل الشروع	فعل الرجاء
مجرور									
نسخ									
نداء									
٧٥	نائب_فاعل مفعول	نائب_فاعل مفعول مفعول	م	مخصوص_مدح_ذم	مبتدأ خبر	مبتدأ خبر	مبتدأ خبر	مبتدأ خبر	مبتدأ خبر
٧٦									
٧٧									
٧٨									
٧٩	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول	مفعول
٨٠	خبر	خبر	خبر	خبر	خبر	خبر	خبر	خبر	خبر
٨١									
٨٢									
٨٣									
٨٤									
٨٥									
٨٦	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر	جملة_فعلية خبر
٨٧									
اسم_موصول									

Annexe 02 : tableau des *firsts* des productions

N_terminale	Les firsts
الجملة الاسمية	<p>"1-2-0-0-0-0" اسم ظاهر</p> <p>"1-2-3-5-0-0"</p> <p>"1-2-3-0-0-0"</p> <p>"1-2-3-6-0-0"</p> <p>"1-2-6-0-0-0"</p> <p>"1-2-6-12-0-0"</p> <p>"1-2-6-13-0-0"</p> <p>"1-2-6-14-0-0"</p> <p>"1-2-6-15-0-0"</p> <p>"1-2-6-16-0-0"</p> <p>"1-2-7-0-0-0"</p> <p>"1-2-7-17-0-0"</p> <p>"1-2-7-18-0-0"</p> <p>"1-2-7-19-0-0"</p> <p>"1-2-10-0-0-0"</p> <hr/> <p>اسم موصول</p> <p>"1-2-4-0-0-0"</p> <p>"1-2-4-7-0-0"</p> <p>"1-2-4-8-0-0 "</p> <hr/> <p>ضمير</p> <p>"1-2-5-0-0-0"</p> <p>"1-2-5-9-0-0"</p> <hr/> <p>الإشابة</p> <p>"1-2-9-0-0-0"</p> <hr/> <p>شبه جملة</p> <p>1-3-14-0-0-0</p> <p>1-2-11-0-0-0</p> <p>-----</p>
جملة الشرط	<p>1-3-12-26-0-0</p> <p>1-3-12-26-10-0</p> <p>1-3-12-26-11-0</p>
جملة الإلتفهام	<p>1-3-13-0-0-0</p> <p>1-3-13-28-0-0</p> <p>1-3-13-29-0-0</p> <p>1-3-13-30-0-0</p> <p>1-3-13-31-0-0</p> <p>1-3-13-32-0-0</p> <p>1-3-13-33-0-0</p> <p>1-3-13-34-0-0</p> <p>1-3-13-35-0-0</p>
الجملة الفعلية	1-1-1-1-1-1

Annexe 02 :

	1-1-1-1-1-2 1-1-1-1-1-3 1-1-1-1-1-4 1-1-1-1-1-5 1-1-1-1-1-6 1-1-1-1-2-0 1-1-1-2-3-0 1-1-1-2-4-0 1-1-2-3-0-0 1-1-2-4-5-0 1-1-2-4-6-0 1-1-2-4-7-0 1-1-2-4-8-0 1-1-2-4-9-0 1-1-1-1-1-0 1-1-1-1-0-0 1-1-1-0-0-0 1-1-2-0-0-0 1-1-2-4-0-0 1-1-0-0-0-0 1-1-1-2-0-0	
جملة الترخ	1-3-17-0-0-0 1-3-17-36-0-0 1-3-17-37-0-0 1-3-17-38-0-0 1-3-17-39-0-0 1-3-17-40-0-0	الناسخ
Z (النفى)	1-3-18-0-0-0	
النداء	1-3-16-0-0-0	

الجملة الفعلية

N_terminal	Les firsts
R (فعل متعدي 1)	1-1-1-1-1-1
(فعل متعدي 2) فاعل مفعول مفعول	1-1-1-1-1-2
(فعل متعدي 3) فاعل مفعول مفعول مفعول	1-1-1-1-1-3
(فعل اللازم) فاعل	1-1-1-1-2-0
(فعل اللازم) فاعل شبه جملة	1-1-1-1-2-0
(فعل م 1) نائب فاعل	1-1-1-1-1-4
(فعل م 2) نائب فاعل مفعول	1-1-1-1-1-5
(فعل م 3) نائب فاعل مفعول مفعول	1-1-1-1-1-6
(فعل أمر) فاعل	1-1-2-3-0-0
(فعل أمر) فاعل مفعول	1-1-2-3-0-0
(فعل أمر) شبه جملة	1-1-2-3-0-0
(فعل المدح الذم) فاعل المخصوص بالمدح	1-1-2-4-7-0 1-1-2-4-8-0
(فعل ناقص مقارنة استمرار الشروع الرجاء) مبتدأ خبر	1-1-2-4-9-0 1-1-2-4-5-0 1-1-2-4-6-0 1-1-1-2-3-0 1-1-1-2-4-0

Annexe 02 :

جملة فعلية	1-1-1-1-1-1 1-1-1-1-1-2 1-1-1-1-1-3 1-1-1-1-2-0 1-1-1-1-2-0 1-1-1-1-1-4 1-1-1-1-1-5 1-1-1-1-1-6 1-1-2-3-0-0 1-1-2-3-0-0 1-1-2-3-0-0 1-1-2-4-7-0 1-1-2-4-8-0 1-1-2-4-9-0 1-1-2-4-5-0 1-1-2-4-6-0 1-1-1-2-3-0 1-1-1-2-4-0
------------	--

فاعل

N_terminal	Les firsts
اسم ظاهر	"1-2-0-0-0-0" "1-2-3-5-0-0" "1-2-3-0-0-0" "1-2-3-6-0-0" "1-2-6-0-0-0" "1-2-6-12-0-0" "1-2-6-13-0-0" "1-2-6-14-0-0" "1-2-6-15-0-0" "1-2-6-16-0-0" "1-2-7-0-0-0" "1-2-7-17-0-0" "1-2-7-18-0-0" "1-2-7-19-0-0" "1-2-10-0-0-0"
الإشارة	"1-2-9-0-0-0"
اسم موصول	"1-2-4-0-0-0" "1-2-4-7-0-0" "1-2-4-8-0-0"
(ضمير متصل)	"1-2-5-10-0-0"
(ضمير مستتر)	"1-2-5-11-0-0"

المفعول = نائب فاعل

N_terminal	Les firsts
اسم ظاهر	"1-2-0-0-0-0" "1-2-3-5-0-0" "1-2-3-0-0-0" "1-2-3-6-0-0" "1-2-6-0-0-0" "1-2-6-12-0-0" "1-2-6-13-0-0" "1-2-6-14-0-0" "1-2-6-15-0-0" "1-2-6-16-0-0" "1-2-7-0-0-0" "1-2-7-17-0-0" "1-2-7-18-0-0" "1-2-7-19-0-0" "1-2-10-0-0-0"
(ضمير متصل)	"1-2-5-10-0-0"
(ضمير منفصل)	"1-2-5-9-0-0"
مصدر مؤول	مصدر مؤول
جملة	اسم ظاهر "1-2-0-0-0-0" "1-2-3-5-0-0" "1-2-3-0-0-0" "1-2-3-6-0-0" "1-2-6-0-0-0" "1-2-6-12-0-0" "1-2-6-13-0-0" "1-2-6-14-0-0" "1-2-6-15-0-0" "1-2-6-16-0-0" "1-2-7-0-0-0" "1-2-7-17-0-0" "1-2-7-18-0-0" "1-2-7-19-0-0" "1-2-10-0-0-0" "1-2-9-0-0-0" "1-2-4-0-0-0" "1-2-4-7-0-0" "1-2-4-8-0-0" "1-2-5-9-0-0" "1-2-11-0-0-0" "1-3-10-0-0-0" حرف مؤول مصدرية 1-1-1-1-1-1 1-1-1-1-1-2 1-1-1-1-1-3 1-1-1-1-2-0 1-1-1-1-2-0 1-1-1-1-1-4 1-1-1-1-1-5 1-1-1-1-1-6 1-1-2-3-0-0 1-1-2-3-0-0

Annexe 02 :

	1-1-2-3-0-0 1-1-2-4-7-0 1-1-2-4-8-0 1-1-2-4-9-0 1-1-2-4-5-0 1-1-2-4-6-0 1-1-1-2-3-0 1-1-1-2-4-0
--	--

Annexe 03 : Des exemples traités par la grammaire.

✓ ما محمد إلا رسول.

✓ ما في الدار الا محمد.

✓ السماء صافية.

✓ محمد في الدار.

✓ في الدار محمد.

✓ يا محمد النبي.

✓ اذا ذَهَبَ فَآزَ.

✓ ذَهَبَ محمد الى الدار.

✓ محمد ذَهَبَ.

