

**UNIVERSITE SAAD DAHLEB DE BLIDA**

**Faculté des sciences**

**Département de l'informatique**



**MEMOIRE DE FIN D'ETUDES**

Master en informatique

Traitement automatique de la langue

**UNE FORMALISATION DE LA THEORIE NEO-KHALILIENNE PAR  
LA GRAMMAIRE DE PRE-GROUPE DE LAMBEK  
– CAS D'ETUDE : LA TECTONIE NOMINALE**

Par

**Mohamed KACIMI**

**Lyes SAADOUDI**

Devant le jury composé de :

- |   |                         |           |
|---|-------------------------|-----------|
| - | Enseignant, U. de Blida | Président |
| - | Enseignant, U. de Blida | Examineur |
| - | Mr. ABBACHE Ahmed       | Encadreur |

Blida, Décembre 2020

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

# DEDICACE

À nos très chers parents,

Nul mot, Nulle dédicace ne pourra suffisamment exprimer tout notre respect, notre considération et notre amour pour les sacrifices auxquels vous avez tant consentis pour notre instruction.

Pour votre générosité et votre bonté.

Et dont le présent ce travail en est une modeste récompense.

Nous vous exprimons toute notre affection

Nous dédicaçons tout aussi sincèrement notre mémoire

À toute notre chère famille,

À nos chers professeurs,

À nos chers amis,

À nos chers collègues,

À tous ceux qui nous ont aidés de près ou de loin,

Nous dédions ce modeste travail

# REMERCIEMENTS

Avant tout, nous remercions Dieu le tout puissant en qui nous avons trouvé la force, le courage et la volonté pour la réalisation ce modeste travail.

*« Celui qui ne remercie pas les gens, ne remercie pas Allah »*

[Authentique Hadith]

Aussi,

Nous remercions Mr. Abbache Ahmed de nous avoir donné l'opportunité d'accomplir cette œuvre par leurs conseils judicieux.

On tient également à remercier nos enseignants qui nous ont dispensés durant deux ans de master, leurs précieux conseils et orientations,

Enfin, merci à nos familles, nos chers mères et pères pour le soutien et leurs encouragements qu'ils n'ont eu de cesse de nous apporter.

# RESUME

La théorie néo-khalilienne a pour point de départ la découverte, est une théorie linguistique contemporaine basée sur le travail des contributions des premiers grammairiens arabes; il a été fondé par le linguiste algérien Abderrahmane Hadj Salah. Dans la théorie néo-khalilienne les unités de la langue ne sont pas nécessairement des segments (ou marginalement des accents). IL existe des dénotants abstraits aussi importants que les dénotants segmentaux ou accentuels. Exemple : le schème et la racine d'un élément nominal ou verbal : chacun d'eux dénote un sens en lui-même : la synthèse des deux dénotants donne un segment dont le sens résulte également de la synthèse des deux sens abstraits (et non de leur amalgame ou de leur juxtaposition).

La grammaire de pré-groupe est un modèle algébrique de grammaire sous la forme d'un pré-groupe, fondé par le mathématicien canadien Joachim Lambek. L'idée d'une grammaire de pré-groupe est d'attribuer à chaque mot un ou plusieurs types.

La tectonie comme entité sémiologico-grammaticale est constituée d'un noyau qui forme la racine du schème générateur syntaxique et d'incrémentes qui s'ajoutent à ce noyau en postposition ou en antéposition, selon des positions structurelles déterminées, pour former des séquences dérivées.

L'objectif du travail consiste à concevoir, à réaliser et à évaluer un système d'analyse de textes Arabes basée sur la grammaire de pré-groupe permettant de reconnaître automatiquement les tectonies nominale. Le système reçoit en entrée des documents textuels et produit en sortie une liste des tectonies nominale.

---

**Mot clés** : la théorie néo-khalilienne , Langue arabe, grammaire de pré-groupe, La tectonie, le schème et la racine

# ABSTRACT

The neo-Khalilian theory has as its starting point the discovery, is a contemporary linguistic theory based on the work of the contributions of the first Arab grammarians; it was founded by the Algerian linguist Abderrahmane Hadj Salah. In Neo-Khalilian theory the units of the language are not necessarily segments (or marginally accents). There are abstract denotants as important as segmental or accentual denotants. Example: the diagram and the root of a nominal or verbal element: each of them denotes a meaning in itself: the synthesis of the two denotants gives a segment whose meaning also results from the synthesis of the two abstract meanings (and not of their amalgam or their juxtaposition).

Pre-group grammar is an algebraic model of grammar in the form of a pre-group, founded by Canadian mathematician Joachim Lambek. The idea of a pre-group grammar is to assign one or more types to each word.

Tectonia as a semiologico-grammatical entity consists of a nucleus which forms the root of the syntactic generator scheme and increments which are added to this nucleus in postposition or in anteposition, according to determined structural positions, to form derived sequences.

The objective of the work is to design, realize and evaluate a system of analysis of the Arabic phases based on the grammar of pre-group allowing to check if a given word chain is a well formed sentence by performing a calculation on the matching type string

**Keywords:** Néo-Khalilian théory , Arabic Language, Pre-group grammar, , diagram and the root .

## ملخص

النظرية الخليلية الجديدة هي نقطة البداية للاكتشاف ، وهي نظرية لغوية معاصرة تقوم على عمل مساهمات النحويون العرب الأوائل. أسسها اللغوي الجزائري عبد الرحمن حاج صلاح. في نظرية الخليلين الجدد ، وحدات اللغة ليست بالضرورة أجزاء (أو لهجات هامشية). هناك دلالات مجردة لا تنقل أهمية عن دلالات قطعية أو مميزة. مثال: الرسم التخطيطي وجذر العنصر الاسمي أو اللفظي: يشير كل منهما إلى معنى في حد ذاته: يعطي توليف اثنين من الدلالات شريحة ينجم معناها أيضًا عن تركيب المعنيين المجردين (وليس من ملغهم أو تجاورهم).

قواعد ما قبل المجموعة هي نموذج جبرية لقواعد اللغة في شكل مجموعة سابقة أسسها عالم الرياضيات الكندي يواكيم لامبك. فكرة قواعد ما قبل المجموعة هي تعيين نوع واحد أو أكثر لكل كلمة. يتكون (tectonie) ككيان شبه نحوي من نواة تشكل جذر مخطط المولد النحوي والزيادات التي تتم إضافتها إلى هذه النواة في الوضع التالي أو في الوضع المضاد ، وفقًا للمواضع الهيكلية المحددة ، لتشكيل تسلسلات مشتقة. الهدف من العمل هو تصميم وبناء وتقييم نظام تحليل للنصوص العربية على أساس القواعد النحوية قبل المجموعة التي تسمح بالتعرف التلقائي على العوامل الاسمية. يتلقى النظام لمستندات نصية كمدخلات ويخرج قائمة للعوامل الاسمية.

**الكلمات المفتاحية:** النظرية الخليلية، اللغة العربية ، قواعد ما قبل المجموعة ، المثال ، الجذر

# TABLE DES MATIÈRES

REMERCIEMENTS

RESUME

LISTE DES FIGURES

LISTES DES TABLEAUX

LISTE DES ABREVIATIONS

INTRODUCTION GENERALE ..... 12

CHAPITRE I : TRAITEMENT AUTOMATIQUE DE LA LANGUE ARABE.. 15

**1.1 Introduction** ..... 15

1.1.1 Traitement Automatique de la Langue (TAL) ..... 15

1.1.2 Niveaux Traitement Automatique de la Langue ..... 16

**1.2 Traitement automatique de la langue arabe (TALA)** ..... 18

1.2.1 La langue arabe : ..... 18

1.2.2 L'alphabet et les caractères spéciaux ..... 19

1.2.3 Les voyelles brèves ..... 22

**1.3 Morphologie arabe** ..... 23

1.3.1 Structure d'un mot ..... 24

1.3.2 Catégories des mots ..... 25

1.3.3 Morphologie flexionnelle : ..... 27

1.3.4 Les éléments essentiels de la morphologie arabe ..... 30

**1.4 La structure d'une phrase arabe** ..... 34

1.4.1 La syntaxe de la langue arabe ..... 35

1.4.2 L'analyse syntaxique ..... 37

**1.5 Conclusion** ..... 41

CHAPITRE II : LA THEORIE NEO-KHALILIENNE ET LA GRAMMAIRE

DE PREGROUPE ..... 43

**2.1 Introduction** ..... 43

**2.2 Théorie Néo-khalilienne (TNK)** ..... 43

2.2.1 Les Concepts Fondamentaux: ..... 44

2.2.2 La définition de la tectonie ..... 44

2.2.3 Les Structures Syntaxiques Fondamentales ..... 49

**2.3 Grammairs de pregrroupe** ..... 51

2.3.1 Propriétés fondamentales ..... 51

2.3.2 Pré-grammaire ..... 52

2.3.3 Contractions et Extensions ..... 52

2.3.4 Le pré-groupe libre ..... 53

2.3.5 Pré-groupes pour la linguistique ..... 53

**2.4 Conclusions** ..... 57

<b>CHAPITRE III : POUR L'ANALYSE UNE APPROCHE ALGEBRIQUE DE LA STRUCTURE DE TECTONIE NOMINALE EN ARABE .....</b>	<b>59</b>
<b>3.1 Introduction .....</b>	<b>59</b>
<b>3.2 La machinerie algébrique.....</b>	<b>59</b>
<b>3.3 Types de base .....</b>	<b>60</b>
<b>3.4 La structure du nom arabe.....</b>	<b>61</b>
<b>3.5 La structure des lexies nominales en arabe .....</b>	<b>62</b>
<b>3.6 Les formes catégories du nom en arabe .....</b>	<b>62</b>
3.6.1 Les noms primitifs .....	63
3.6.2 Les noms dérivés.....	63
<b>3.7 Les formes de la tectonie nominale en arabe .....</b>	<b>66</b>
<b>3.8 Expression formelle du nom en arabe notée Dgncd(N) .....</b>	<b>66</b>
3.8.1 un nom défini : .....	67
3.8.2 un pronom : .....	67
3.8.3 un nom propre : .....	68
3.8.4 un annexion (dans le cas d'une annexion) .....	68
<b>3.9 Approche basée sur les notions de «' mal terme régissant » et «ma'mul terme régi ».....</b>	<b>69</b>
3.9.1 R = kana et autre homologue .....	69
3.9.2 R = inna et autre homologue .....	69
<b>3.10 Conclusions .....</b>	<b>70</b>
<b>CHAPITRE IV: EXPERIMENTATIONS ET IMPLIMENTATION .....</b>	<b>71</b>
<b>4.1 Introduction : .....</b>	<b>72</b>
<b>4.2 Les outils utilisés pour l'implémentation .....</b>	<b>72</b>
4.2.1 L'environnement de programmation: .....	72
4.2.2 NetBeans IDE 8.2: .....	72
<b>4.3 D'analyse de reconnaissance d'une tectonie nominale .....</b>	<b>73</b>
<b>4.4 Déroulement.....</b>	<b>74</b>
<b>4.5 Conclusion.....</b>	<b>77</b>
<b>4.6 Conclusion et perspective .....</b>	<b>78</b>

# LISTE DES FIGURES

<i>Figure 1: Présente La Pluridisciplinaire De TAL</i> .....	16
<i>Figure 2: Les niveaux de traitement</i> .....	17
<i>Figure 3: Répartition géographique de la langue arabe</i> .....	19
<i>Figure 4: les voyelles brèves</i> .....	22
<i>Figure 5: les singes alchadda et alsukun</i> .....	22
<i>Figure 6: signe du tanwin.</i> .....	23
<i>Figure 7: structure du mot arabe أتتذكروننا</i> .....	24
<i>Figure 8: catégories de mot</i> .....	25
<i>Figure 9: la structure d'une phrase nominale [23]</i> .....	36
<i>Figure 10: exemple d'une phrase verbale</i> .....	37
<i>Figure 11: Les différents niveaux d'analyse d'un texte.</i> .....	38
<i>Figure 12: Technique de la boîte [35]</i> .....	39
<i>Figure 13: Un arbre est constitué de branches et de noeuds</i> .....	39
<i>Figure 14: Le syntagme verbale</i> .....	40
<i>Figure 15: SHEME GENERATEUR DE LA LEXIE NOMINALE</i> .....	47
<i>Figure 16: SHEME GENERATEUR DE LA LEXIE VERBALE CAS DE L'ACCOMPLI</i> .....	48
<i>Figure 17: structure du nom arabe بالكتاب</i> .....	61
<i>Figure 18: NetBeans</i> .....	72
<i>Figure 19 : interface globale</i> .....	74
<i>Figure 20: l'interface globale du programme</i> .....	75
<i>Figure 21: le résultat final de cas vrai</i> .....	76
<i>Figure 22: le résultat final de cas faux</i> .. .....	77

# LISTE DES TABLEAUX

<i>Tableau 1: les 28 lettres arabe.....</i>	<i>20</i>
<i>Tableau 2: les lettres salaire et Lunaires [6].....</i>	<i>21</i>
<i>Tableau 3: exemple de variation de lettre ق/ qāf. ....</i>	<i>21</i>
<i>Tableau 4: structure du mot arabe.....</i>	<i>24</i>
<i>Tableau 5: conjugaison pour le verbe 'dessin' rasama', رسم' .....</i>	<i>29</i>
<i>Tableau 6: Exemples de schèmes appliqués au mot عمل Travailler [32]. ....</i>	<i>31</i>
<i>Tableau 7: quelque dérivation du verbe " 15] " كتب].....</i>	<i>32</i>
<i>Tableau 8: un exemple des préfixes [29]. ....</i>	<i>33</i>
<i>Tableau 9: un exemple des suffixes divisés selon leurs types [29]. ....</i>	<i>34</i>
<i>Tableau 10: Les noms primitifs dans la langue arabe. ....</i>	<i>63</i>
<i>Tableau 11: Les noms dérivés dans la langue arabe. ....</i>	<i>64</i>
<i>Tableau 12: Pronoms libres du singulier.....</i>	<i>65</i>
<i>Tableau 13: Pronoms libres du duel.....</i>	<i>65</i>

# LISTE DES ABREVIATIONS

<b>TAL</b>	<i>Traitement automatique des langues.</i>
<b>TALN</b>	<i>Traitement automatique des langues naturelles.</i>
<b>TALA</b>	<i>Traitement automatique de la langue arabe.</i>
<b>TNK</b>	<i>Théorie Néo-khalilienne .</i>
<b>CRTDLA</b>	<i>Centre de Recherche pour le Trésor De La langue Arabe.</i>
<b>SN</b>	<i>Syntagme nominale.</i>
<b>SV</b>	<i>Syntagme verbale.</i>
<b>SADJ</b>	<i>Syntagme adjectivale.</i>
<b>Det</b>	<i>Déterminant.</i>
<b>AS</b>	<i>Arabe Standard</i>

## À PROPOS DU PROFESSEUR:

Le Dr Abdel-Rahman Al-Haj Saleh est né dans la ville d'Oran en 1927, et ses études étaient en Egypte Et Bordeaux Paris, vous obtenez tabriz de Paris, puis doctorat en linguistique également de l'Université de Paris II a occupé plusieurs postes scientifiques, il a été professeur à l'Université de Rabat de 1961 à 1962 à l'Université d'Algérie.

Il est ensuite devenu directeur de l'Institut des sciences linguistiques d'Algérie, puis directeur du Centre de recherche scientifique pour la promotion et le développement de la langue arabe.

Avec la langue arabe en l'an 2000 après JC, a été nommé par le président de la République algérienne Abdelaziz Bouteflika comme président et était un membre important de tous les séminaires qui ont eu lieu dans les universités algériennes.

Abd al-Rahman al-Haj Saleh est décédé le 5 mars 2017 à l'hôpital Ain al-Naajh de la capitale  
Production scientifique et publications:

Les productions scientifiques d'Abd al-Rahman al-Hajj Saleh se sont multipliées et ses publications se sont multipliées

Pour les chercheurs arabes et non arabes, ce sont diverses recherches et études linguistiques, notamment:

- Etudes scientifiques spécialisées (en arabe et en français "71 recherches et études dans divers domaines Et anglais) en 2002 AD.
- Glossaire des sciences de la langue (avec participation), Alexo Arabization Coordination Office, 1992.
- Linguistique arabe et linguistique générale (en deux volumes), Algérie.
- Article «Language» et «Knowledge» dans le Department of Islamic Knowledge, New Edition, Leiden.
- Linguistique et phonétique arabes, in Linguistique et linguistique arabes appliquées, New - - yourk, 1987.
- Recherche et études en sciences de la langue en deux parties (arabe, français et anglais) en Algérie.
- Quatre articles: Hebron bin Ahmed, Al-Afash, Ibn Al-Sarraj et Al-Suhaili, dans Encyclopedia of Arab Flags (L'Organisation Arabe pour l'éducation, la culture et la science).

## *Introduction Générale:*

---

Chaque génération a ses génies, chaque langue a ses pionniers, et la langue arabe, comme toutes les autres langues du monde, a connu tout au long de sa glorieuse histoire scientifique de nombreux scientifiques et chercheurs arabe musulman.

Ils ont apporté des contributions précieuses dans de nombreux domaines de la connaissance et de la science, en particulier dans le domaine de la linguistique. Par conséquent, cette recherche cherche à mettre en évidence certains de ces faits historiques, en toute objectivité, afin de ne pas rester cachés aux yeux des gens, en particulier des étudiants et des diplômés des instituts de langues du monde arabe, en s'arrêtant à un certain nombre de chercheurs et de linguistes contemporains et non arabes qui ont étudié notre patrimoine linguistique arabe et l'ont analysé et apprécié mieux, et ils sont allés Ils la prennent sous de nouvelles perspectives et suggèrent plusieurs théories linguistiques, qui sont des deuxième et troisième lectures d'un immense patrimoine linguistique qui n'a pas été pleinement révélé.

Le projet "DAKHIRA AL Arabiya", est une initiative de la ligue arabe fondée sur le principe de la participation d'institutions scientifiques et culturelles de chaque pays membre, dont le centre de recherche Scientifique et Technique pour le Développement de la Langue Arabe (CRSTDLA) qui apporte sa pierre à cette audacieuse idée qui est une banque de données textuelles anciennes et modernes couvrant le patrimoine culturel arabe.

Ce travail expose un procédé utilisent la notion de grammaire de pré-groupe pour décrire la structure d'une tectonie nominale dans un fragment de la langue l'arabe moderne standard basée sur la théorie Néo-Khalilienne, est théorie linguistique contemporaine basée sur le travail des contributions des premiers grammairiens arabes; elle a été fondé par le linguiste algérien Abderrahmane Hadj Salah.

L'objectif du travail consiste à concevoir, à réaliser et évaluer un système d'analyse de textes arabes selon la grammaire de pré-groupe permettant de reconnaître automatiquement les tectonies nominales.

Notre mémoire est ainsi organisé :

- **Un premier chapitre intitulé " *Traitement automatique de la langue arabe* ",**

nous donnons un aperçu sur la langue arabe et exposons les différents aspects de sa morphologie ainsi que ses caractéristiques.

- **Dans le deuxième chapitre intitulé " La théorie néo-khalilienne et la grammaire de pré-groupe "**, Nous présentons un bref et précis aperçu sur l'adéquation scientifique et didactique de la théorie Néo-Khalilienne. Nous allons également présenter une formalisation de certain phrase arabe par les Grammaires catégorielles de " J. Lambek
- **Puis, en troisième chapitre** concerne pour l'analyse une approche algébrique de la structure des tectonie nominales en arabe.
- **En fin, le quatrième chapitre :**" Expérimentations et Implémentation ", se prêtera à exposer les expérimentations et les résultats obtenus ainsi que l'implémentation d'une interface de prédiction de formalisation.
-

# **Chapitre-I**

## ***Traitement Automatique de la Langue Arabe (TALA)***

## **1.1 Introduction**

La langue est un outil central dans notre vie sociale et professionnelle. Il s'agit d'un support pour véhiculer, entre autres, des idées, des opinions et des sentiments ainsi que pour persuader, demander des informations, donner des ordres, etc.

L'intérêt pour la langue du point de vue de l'informatique a débuté au début même de l'informatique, notamment dans le cadre des travaux dans le domaine de l'intelligence artificielle ; on assiste alors à la naissance du TAL.

La vague d'Internet entre le milieu des années 1990 et le début des années 2000 a été un moteur très important pour le TAL et les domaines dérivés, notamment celui de la recherche d'information (RI) et de la classification qui sont passés d'un domaine marginal et limité au seul domaine de la grande entreprise, à la recherche d'information à l'échelle d'Internet, dont le contenu ne cesse de s'élargir [2].

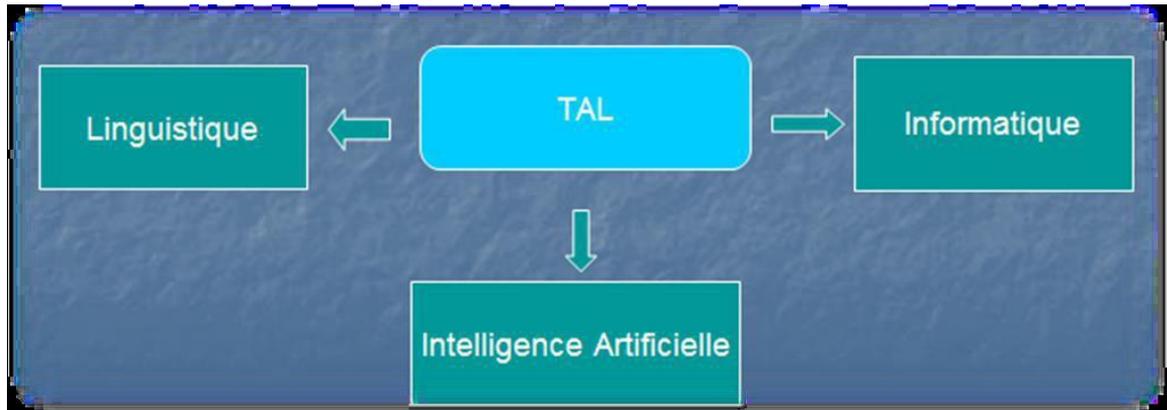
Au cours de ce chapitre nous présenterons d'une manière brève le TAL et le TALA tout en décrivant les particularités de la langue arabe ainsi que certaines de ses propriétés morphologiques et syntaxiques.

### **1.1.1 Traitement Automatique de la Langue (TAL)**

Le Traitement Automatique de la langue naturelle (TALN) ou des langues (TAL) est une discipline à la frontière de la linguistique, de l'informatique et de l'intelligence artificielle.

Elle concerne la conception de systèmes et techniques informatiques permettant de manipuler le langage humain, dont le principal objectif est la conception et le développement de programmes capables de traiter de manière automatique des données linguistiques c'est-à-dire des données exprimées dans une langue dite naturelle.

Ces dernières décennies le traitement automatique des langues a connu une véritable ascension que ce soit sur le plan scientifique mais aussi socio-économique et cela par l'émergence de plusieurs firmes et de produits spécialisés. On parle aujourd'hui : de Traduction automatique, de correction automatique d'orthographe, de résumé automatique, d'interrogation de base de données en langues naturelle, ....etc.



**Figure 1: Présente La Pluridisciplinaire De TAL**

La réalisation de n'importe quelle application parmi celles citées précédemment passe principalement par différents niveaux (lexicale, morphologique, syntaxique, sémantique et pragmatique) mais aussi par le développement de plusieurs modules importants, où la réussite de l'application dépend pleinement de la performance de ces modules [3].

### 1.1.2 Niveaux Traitement Automatique de la Langue

On va essayer de citer brièvement dans cette section les différents niveaux de traitement nécessaires pour parvenir à une compréhension complète d'un énoncé en langage naturel.

La figure 02 schématise ces différents niveaux de traitements. Ces niveaux se superposent ; chacun apportant des problèmes spécifiques à

résoudre relatifs à un niveau donné. En s'appuyant sur un découpage méthodologique classique dans le domaine de la linguistique cela nous donne la hiérarchie suivante :

- ✓ **La phonétique** concerne l'étude des sons et prosodies (variations).
- ✓ **La phonologie** concerne l'étude de Phonèmes.
- ✓ **La morphologie** concerne l'étude de la formation des mots et de leurs variations de forme.
- ✓ **La syntaxe** consistant à extraire les relations grammaticales que les mots et groupes de mots entretiennent entre eux.
- ✓ **La sémantique** se consacre au sens des énoncés.
- ✓ **La pragmatique** prend en compte le contexte d'énonciation.

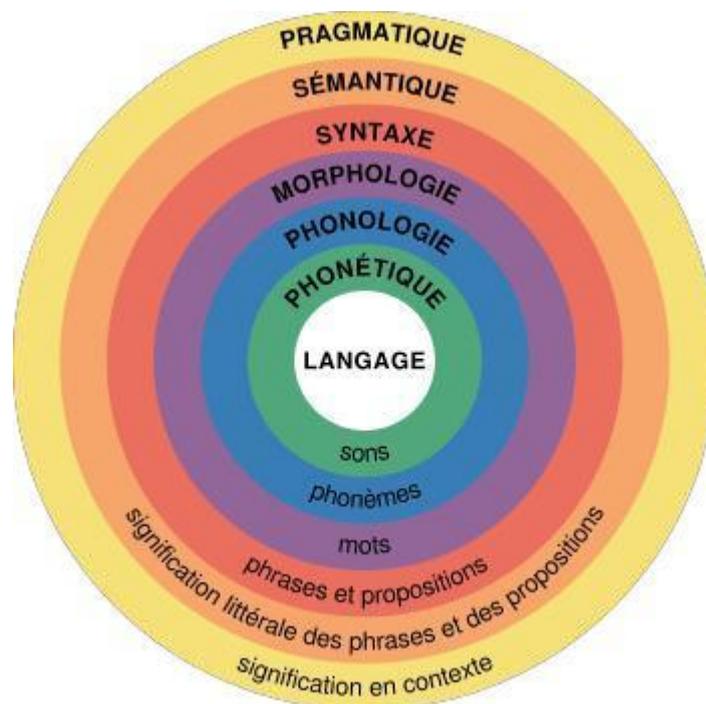


Figure 2: Les niveaux de traitement

## 1.2 Traitement automatique de la langue arabe (TALA):

Le traitement automatique de la langue arabe est une discipline en pleine expansion, et dans laquelle on constate de plus en plus de recherches et de technologies qui portent un intérêt aux spécificités de cette langue et proposent des outils nécessaires au développement de son traitement automatique.

Le traitement automatique de l'arabe est un domaine de recherche stimulant. Il 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.

Par ses propriétés morphologiques et syntaxiques la langue arabe est considérée Comme une langue difficile à maîtriser dans le domaine du traitement automatique de la langue.

Les recherches pour le traitement automatique de l'arabe ont débuté vers les années 1970. Les premiers travaux concernaient notamment les lexiques et la morphologie [4].

Avec la diffusion de la langue arabe sur le Web et la disponibilité des moyens de manipulation de textes arabes, les travaux de recherche ont abordé des problématiques plus variées comme la syntaxe, la traduction automatique, l'indexation automatique des documents, la recherche d'information, la catégorisation des textes etc.

Le domaine du Traitement Automatique des Langues (TAL) appliqué à l'arabe a fait ces 15 dernières années des progrès considérables, mais il reste un grand chemin à faire pour pouvoir rivaliser avec d'autres langues comme le français et l'anglais.

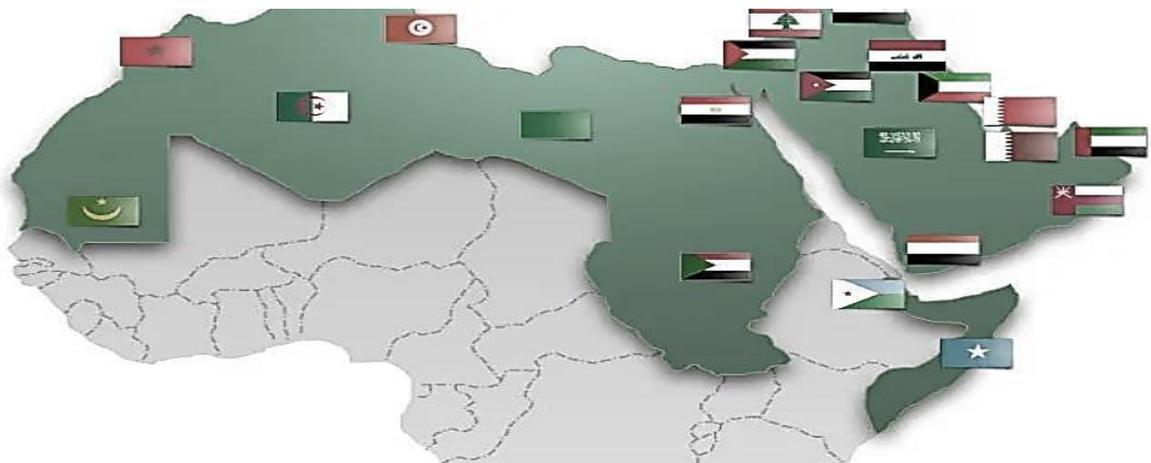
### 1.2.1 La langue arabe :

L'arabe (en arabe : العربية , al-'arabīya') est une langue afro-asiatique de la famille des langues sémitiques. Avec un nombre de locuteurs estimé entre 315 et 375 millions de personnes au sein du monde arabe et de la diaspora arabe. L'arabe est de loin la langue sémitique la plus parlée.

La langue arabe est originaire de la péninsule arabique, où elle devint au VIIe siècle la langue du Coran et la langue liturgique de l'islam.

L'expansion territoriale de l'empire arabe au moyen âge a conduit à l'arabisation au moins partielle sur des périodes plus ou moins longues du Moyen-Orient, de l'Afrique du Nord et de certaines régions en Europe (péninsule Ibérique, Sicile, Crète, Chypre, territoires où elle a disparu, et Malte, où le maltais en constitue un prolongement particulier).

Parlée d'abord par les arabes, cette langue qui se déploie géographiquement sur plusieurs continents s'étend sociologiquement à des peuples non arabes, et est devenue aujourd'hui l'une des langues les plus parlées dans le monde. C'est la langue officielle de plus de vingt pays et de plusieurs organismes internationaux, dont l'une des six langues officielles de l'Organisation des Nations Unies .



**Figure 3: Répartition géographique de la langue arabe [5]**

### 1.2.2 L'alphabet et les caractères spéciaux

Système graphique ou L'écriture en arabe est différente et distinctive des autres langues étrangères où elle présente des caractéristiques uniques que nous allons expliquer ces fonctionnalités dans les étapes suivantes :

L'arabe s'écrit et se lit cursivement (écriture dont les lettres sont reliées les unes avec les autres) de droite à gauche, en utilisant un alphabet de 28 lettres [2]. Il n'y a pas de différence entre les lettres manuscrites et les lettres imprimées, et les notions de lettre majuscule et lettre minuscule n'existent pas, l'écriture est donc monocamérale [5].

L'alphabet arabe est de type abjad ; il contient 28 lettres (25 consonnes et 3 voyelles longues « ا », « و », « ي ») (Tableau 1.1).

Lettre arabe	Correspondant français	Prononciation	Lettre arabe	Correspondant français	Prononciation
ا	A	Alef	ك	K	Kāf
ب	B	Bā	ل	L	lām
ت	T	Tā	م	M	mīm
ث	Th	Thā	ن	N	nūn
ج	J	Jīm	ه	H	hā
ح	H	ḥā	و	W	wāw
خ	Kh	Khā	ي	Y	yā
د	D	Dhal			
ر	R	Rā			
ز	Z	zāy			
س	S	sīn			
ش	Sh	chīn			
ص	S	ṣād			
ض	D	ḍād			
ط	T	ṭā			
ظ	Z	ẓā			
ع	‘	ʿayn			
غ	Gh	ghayn			
ف	F	fā			
ق	Q	qāf			

Tableau 1: les 28 lettres arabe

Les 28 consonnes arabes sont divisées en deux groupes :

- 14 consonnes solaires qui assimilent le « ل » de l'article.
- 14 consonnes lunaires qui n'assimilent pas le « ل » de l'article

Les solaires se prononcent en double, comme par exemple avec le mot « soleil » شمس (chams), au lieu de prononcer, el-chams, on prononce ech-chams, car la lettre ش (chin), est une lettre solaire.

Les lettres lunaires, se prononcent normalement et simplement pour elles-mêmes, c'est-à-dire sans les doubler. Par exemple avec le mot « lune », on prononce القمر, el-qamar tout à fait normalement, parce que la lettre ق (qaf) est une lettre lunaire (Tableau 2) .

Solaires	Lunaires
ت ث د ذ ر ز س ش ص ض ط ظ ل ن	ع غ ف ق ك م ه و ي

**Tableau 2: les lettres solaire et Lunaires .**

Toutes les consonnes se lient entre elles sauf « (waw, reh, zain, dal, thal), ( و , ز , ش , د , ذ ) » celles qui ne se joignent jamais à gauche [5]. Exemple, Chevalier s'écrit فارس /fâris/.

Le système de signes diacritiques des points joue un rôle de premier ordre. Il y a des lettres qui ne se distinguent que par la présence, le nombre et la position de points. En effet, les lettres { ع , ه , ف , ق , ت , ب , ث , ن , ي , ح , ج , خ , د , ذ , ر , ز , س , ش , ص , ض , ط , ظ , ع } sont notées au moyen des glyphes { ح , ر , س , ص , ط , ع , ه , ي } .

Chaque lettre possède une forme spécifique en fonction de sa position dans un mot (au début, au milieu ou à la fin) ou si elles sont utilisées de façon isolée , Il résulte 78 formes graphiques à partir des 28 lettres. Le Tableau 1.3 montre les variations de la lettre ق (qāf) :

Isolée	Finale	Médiane	Initiale
ق	قى	قه	قه
افتراق/iftirāq/ Séparation	بريق /barīq/ Glamour	مقفل/muqfal/ Verrouillé	قوي/qawiy/ Forte

**Tableau 3: exemple de variation de lettre ق/ qāf.**

La lettre Hamza prend différentes formes suivant sa voyellisation et la voyellisation de la lettre qui la précède ex : ( لأئؤ إأء ) que pour les lettres ي et ة qui symbolisent respectivement ي et ة .

Les véritables liaisons entre les lettres arabes doivent être des ponts concaves curvilignes, qui sont susceptibles de s'étendre. Cette propriété appelée kashida, ou TaTwil, Tamdid, Madda, MaT, Al'iTala, etc.,

### 1.2.3 Les voyelles brèves

Pour une meilleure précision de la prononciation, des signes ont été inventés. Il s'agit de trois voyelles brèves qui s'ajoutent aux consonnes. Ces trois voyelles brèves sont [5] :



Figure 4: les voyelles brèves

**Fatha** : « َ » est symbolisée par un petit trait sur la consonne et se prononce comme un [a] français.

**Damma** : « ُ » est symbolisée par un crochet au-dessus de la consonne et se prononce comme un [u] français.

**Kasra** : « ِ » est symbolisée par un petit trait au-dessous de la consonne et se prononce comme un [i] français.

### I.2.4 Signes orthographiques

Il y a sept signes orthographiques qui s'ajoutent aux consonnes, et elle les plus courants comme:

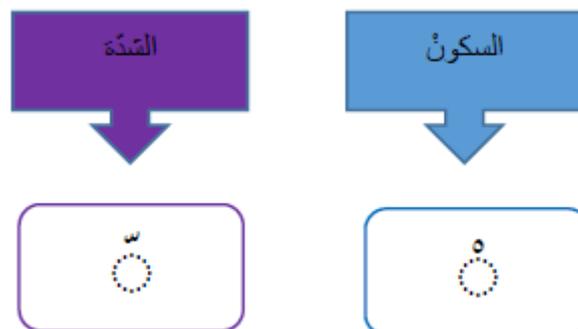


Figure 5: les signes alchadda et alsukun

**Sukun** : « ْ » ce signe indique qu'une consonne n'est pas suivie (ou muet) par une voyelle. Il est noté toujours au-dessus de la consonne « مَكْتَبٌ /maktab » [5].

**Chadda :** « َ » Le signe de la chadda peut être placé au-dessus de toutes les consonnes en position non- initiale. La consonne qui la reçoit est alors analysée en une séquence de deux consonnes identiques [9] « مَكَّنْ / makkana », « يُمَكِّنُ / yumakkinu », « مَكِّنْ /makkin ».

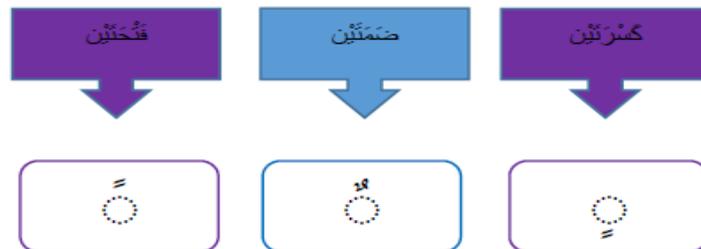


Figure 6: signe du tanwin.

### Tanwin :

Le signe du tanwin est ajouté à la fin des mots indéterminés. Il est en relation d'exclusion avec l'article de détermination ال placé en début de mot . Ce qui se traduit par l'ajout du phonème /n/ au niveau phonétique . Les symboles du tanwin sont au nombre de trois et sont constitués par le :

- [an] : « َ » pour les Fathatan ( ت /tan).
- [in] : « ِ » pour les Kasratan ( ت /tin).
- [un] : « ُ » pour les Dammatan ( ت /tun).

**Wasla :** « ْ » quand la voyelle d'un Alif au commencement d'un mot doit être absorbée par la dernière voyelle du mot qui précède .

**Madda :** « ّ » la madda (prolongation) se place sur l'Alif pour indiquer que cette lettre tient lieu de deux alifs consécutifs ou qu'elle ne doit pas porter le Hamza [5].

## 1.3 Morphologie arabe

Le lexique arabe comprend trois catégories de mots : verbes, noms et particules. Les verbes et noms sont le plus souvent dérivés d'une racine à trois consonnes radicales. Une famille de mots peut ainsi être générée d'un même concept sémantique à partir d'une seule racine à l'aide de différents schèmes.

Ce phénomène est caractéristique à la morphologie arabe. On dit donc que l'arabe est une langue à racines réelles à partir desquelles on déduit le lexique arabe selon des schèmes qui sont des adjonctions et des manipulations de la racine [14].

### 1.3.1 Structure d'un mot

Les mots peuvent avoir une structure composée, résultat d'une agglutination de morphèmes lexicaux et grammaticaux. En arabe un mot peut représenter toute une proposition. La représentation suivante schématise une structure possible de mot complexe. Notons bien que la lecture se fait de droite à gauche.

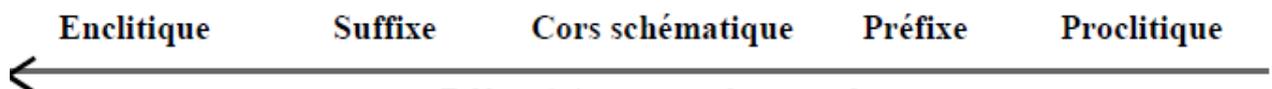


Tableau 4: structure du mot arabe.

- Les proclitiques sont des prépositions ou des conjonctions.
- Les préfixes et suffixes expriment des traits grammaticaux, tels que les fonctions de noms, le mode du verbe, le nombre, le genre, la personne... etc.
- Les enclitiques sont des pronoms personnels.
- Le corps schématique représente la base de mot « radicale » [5].

#### Exemple :

<p style="font-size: 1.2em; font-weight: bold;">أَتَتَذَكَّرُونََنَا</p> <p>"Est ce que vous vous souvenez de nous ?"</p>				
نا [nā]	ون [ona]	تَذَكَّرُ [Tathakaru]	ت [ta]	أ [a]
<b>Enclitique</b>	<b>Suffixe</b>	<b>Cors schématique</b>	<b>Préfixe</b>	<b>Proclitique</b>
Pronom suffixe complément	Suffixe verbal exprimant le	Dérivé de la racine تَذَكَّرُ selon le schème تَفَعَّلَ.	Préfixe verbal du temps de l'inaccompli	Conjonction d'interrogation

Figure 7: structure du mot arabe أَتَتَذَكَّرُونََنَا [5]

### 1.3.2 Catégories des mots

L'arabe considère 3 catégories de mots :



Figure 8: catégories de mot [5]

#### Nom :

Toute unité lexicale référant à un sens indépendant du temps, regroupe : les adjectifs féminin et masculin ; les noms démeritent, les noms prolongés ainsi que les noms réduits, les noms communs et les noms propres, les pronoms et leurs types (connectés et séparés), les noms de périphrases, les noms du verbe, les noms de voix .

#### Verbe :

La plupart des mots en arabe, dérivent d'un verbe de trois lettres, par conséquent chaque verbe est la racine d'une famille de mots. Ces mots sont obtenus en en ajoutant des suffixes ou des préfixes à leur racine. Le paradigme de la conjugaison est déterminé par certaines valeurs liées au genre, mode et structure morphosyntaxique du verbe. Ces valeurs peuvent être résumées comme suit :

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) .

**Par exemple :** La conjugaison de verbe ( ك + ت + ب K+T+B donne le verbe كتب KaTaBa (Écrire)), le temps : accompli, le nombre du sujet : singulier, le genre du sujet : masculin, la personne : troisième personne, le mode : passif.

### Les mots outils :

Un mot outil est une entité invariable exprimant un sens dépendant de compréhension. La langue arabe contient un nombre limité ne dépasse pas 80 éléments, ils se nommaient en arabe les particules de sens ( حروف المعاني ) par contre l'alphabet arabe se nommait les particules de construction ( حروف المباني ).

Les particules de sens sont de type : unitaire, binaire, tertiaire, quaternaire ou quintette, Elles jouent un rôle important dans l'articulation et l'interprétation de la phrase ainsi la cohérence et l'enchaînement d'un texte .

On peut distinguer plusieurs types :

- Préposition : exemple ( حتى , من , عن , على , الى , ب , ك ).
- Particules de coordination : exemple ( او , ثم , ف , و ).
- Particules interrogatives : exemple ( أ , هل , ما ).
- Particules d'affirmation : exemple ( بلى , اجل , نعم ).
- Particules de négation : exemple ( لا , لن ).
- Particules distinctives : exemple ( اي ).
- Particules relatives : exemple ( ما ).
- Particule de subjonctif : exemple ( لِكَيْ , لِكَيْ , كَيْ , لِ , لَنْ , يَ حَتَّى ).
- Particule d'apocopé : exemple ( ما , ما ل , لا , لِ , لَمْ ).
- Particules de futur : exemple ( س , ان , سوف , لن ).
- Particules conditionnelles : exemple ( انْ , انْ , انْ ).
- Particule d'appel : exemple ( يَا , أ , أَلَا ).
- Les conjonctions de subordination : ( بَيْنَمَا , حَيْثُمَا ).
- Les quantificateurs : exemple ( لِكُلِّ , بَعْضِ ).
- Les adverbes : exemple ( قَدْ , ارْأَيْ , دَأْبٌ ).

### 1.3.3 Morphologie flexionnelle :

La langue arabe est aussi une langue flexionnelle, ce type de morphologie est employé, principalement, pour la déclinaison des noms et la conjugaison des verbes, des indices d'aspect, de temps, de mode, personne, de nombre, de cas, de mode, etc., qui sont présentés généralement sous forme de préfixes et suffixes [6].

- Le mode des verbes : par exemple, pour le verbe " ذهب " (aller), les formes à l'accompli sont repérables à l'aide de leurs suffixes tel que " ذهبت " (je suis allé) ou de leurs préfixations telles que " أذهب " (je vais).

- La fonction des noms à l'aide des suffixations tels que " رجلان " (deux hommes au nominatif) ou " رجلين " (deux hommes à l'accusatif ou génitif) .

#### Flexions des noms :

La déclinaison des noms se fait selon les règles suivantes :

- **Le féminin singulier** : On ajoute le ة, exemple صغير petit devient صغيرة petite
- **Le féminin pluriel** : De la même manière, on rajoute pour le pluriel les deux lettres, ات, exemple صغير petit devient صغيرات petites.
- **Le masculin pluriel** : Pour le pluriel masculin on rajoute les deux lettres ين ou ون dépendamment de la position du mot dans la phrase (sujet ou complément d'objet), exemple : الراجع revenant devient الراجعين ou الراجعون revenants .
- **Le dual** : c'est un nom conjugable, il caractérise deux choses ou deux personnes par l'insertion de (ان , An) en cas de nominatif, et (ين , yn) en cas de l'accusatif et génitif. Dans le cas des noms défectueux ou qui se termine par (و , w) ou (ي , Y) ; la terminaison se transforme devant la suffixation de dual, par exemple (مجرى , MajeraY, chemin) se transforme en (مجریان , MajerayaAne, chemins) [].
- **Les cinq noms** : sont des noms bilitères qui ? leur voyelle finale se prolonger quand ils sont définis par un complément : « حم أخ، أب ، ، فو , نو , فو », (Aabun, Aakhun, Hamun, fuw, thuw \_père, frère, le beau-père, bouche, propriétaire).
- **Le Pluriel irrégulier** : Il suit une diversité de règles complexes et dépend du nom. Exemple : طفل un enfant devient أطفال des enfants .

Le phénomène du pluriel irrégulier dans l'arabe pose un défi à la morphologie, non seulement à cause de sa nature non concaténative, mais aussi parce que son analyse dépend fortement de la structure comme pour les verbes irréguliers [6].



- **L'inaccompli subjonctif (منصوب)** : utilisé en proposition subordonnée s'il est précédé par une particule de subjonctif (voir la section les mots outils), il se caractérise par une désinence (فتحة) [fathat] et par des flexions courtes [6].

Par exemple :

(**يُرِيدُ أَنْ يَكْتُبَ** رسالة) il veut écrire une lettre).

- **L'inaccompli apocopé (مجزوم)** : il précède par une particule d'apocopé (voir la section les mots outils), Il se caractérise par l'absence de désinence (سكون) [sukun] et par des flexions courtes [6]. Par exemple :

(**لَمْ يَكْتُبْ**) il n'a pas écrit).

**Remarque** : dans la particule d'apocopé que « mâ – ما », le verbe l'accompli devenir exprimé au passé [6]. Par exemple :

(**مَا كَتَبَ**) il n'a pas écrit).

- **Inaccompli futur - مستقبل** : la conjugaison d'un verbe au futur nécessite d'ajouter l'antéposition au début du verbe conjugué à l'inaccompli. En ajoutant l'antéposition à notre exemple : س on obtient (سيفتح il ouvrira), qui désigne le futur ; on peut également ajouter l'antéposition سوف on obtient « سوف يفتح il va ouvrir » [6].

#### ❖ L'impératif - أمر :

il exprime l'ordre, le commandement, la défense ou l'exhortation et dont les éléments n'existent qu'à la 2ème personne au singulier, féminin duel et pluriel. La voyelle finale /i/ caractérise l'impératif (est structuré sur le soukoun) ou sur l'élimination du noun et de la lettre défectueuse du verbe non sain[8]. Dans le tableau suivant nous donnons un exemple de conjugaison pour le verbe « رسم » rasama 'dessin' :

'أَنْتِ' (tu)	'أَنْتِ' (tu)	'أَنْتُمَا' (vous-2)	'أَنْتُمْ' (vous)	'أَنْتُنَّ' (vous)
أَرْسُخْ	أَرْسُيْ	أَرْسُمَا	أَرْسُمُوا	أَرْسُنَّ

Tableau 5: conjugaison pour le verbe 'dessin' rasama', رسم'

❖ **Voix active** - المبني للمعلوم :

Tous les verbes mentionnés ci-dessus sont des verbes actifs.

❖ **Voix passif** - المبني للمجهول :

Le passif se distingue de l'actif par la vocalisation. Dans les textes non vocalisés, il est le plus souvent indiqué par la voyelle Damma placée sur la première lettre du mot, afin de le différencier de l'actif. L'accord du verbe passif avec son sujet grammatical se fait selon les mêmes règles que celui du verbe actif .

A l'**accompli**, le passif se caractérise par un kasra sous l'avant dernière radicale, et par un Damma sur toutes les consonnes qui précèdent l'avant-dernière radicale et qui n'ont pas de sukûn. Kataba - كَتَبَ = il a écrit → kutiba - كُتِبَ = il a été écrit .

A l'**inaccompli**, le passif se caractérise par un fatHa sur l'avant-dernière radicale, et un Damma sur le préfixe pronominal, le reste ne subissant pas de changement. Yaktubu - يَكْتُبُ = il écrit → yuktabu - يُكْتَبُ = il est écrit .

Le passif du **futur** se caractérise, pour tous les verbes, par un fatHa sur l'avant-dernière radicale, et un Damma sur le préfixe pronominal, le reste ne subissant pas de changement. Sayaktubu - سَيَكْتُبُ = il écrira → yuktabu - سَيُكْتَبُ = il sera écrit .

### 1.3.4 Les éléments essentiels de la morphologie arabe

La langue arabe a une morphologie riche et différente, par rapport aux langues occidentales. L'analyse morphologique d'un mot arabe, consiste principalement à déterminer la structure générale de ce mot, s'il existe, et les autres éléments utilisés pour construire ce mot (les affixes, les modèles).

Les éléments essentiels de la morphologie de la langue arabe sont [1] :

- **Le schème**

L'Arabe est une langue sémitique dont la morphologie fonctionne sur le croisement des racines (généralement trilitères) avec un nombre de formes déterminées pour produire sa gamme de vocabulaire. Ces formes, appelées moules ou schèmes, sont essentiellement constituées par une suite de voyelles et consonnes qui viendront se placer avant, après et entre les lettres de la racine afin de former les différentes formes des flexions verbales et nominales.

Une famille de mots peut ainsi être générée d'un même concept sémantique à partir d'une seule racine à l'aide de différents schèmes. Ce phénomène est caractéristique à la morphologie arabe. Le Tableau 3 donne quelques exemples de schèmes appliqués sur le mot « عمل ». On peut ainsi dériver un grand nombre de noms, de formes et de temps verbaux [8].

Schèmes	EML	عمل	Notion de travailler
<b>R1â-R2i-R3</b>	EaMiL	عَامِلٌ	ouvrier
<b>R1a-R2a-R3a</b>	EaMaLa	عَمَلَ	a travaillé
<b>maR1R2aR3</b>	maEMaL	مَعْمَلٌ	atelier
<b>R1uR2iR3a</b>	EuMiLa	عُمِلَ	a été travaillé

**Tableau 6: Exemples de schèmes appliqués au mot عمل Travailler [8].**

Les lettres en majuscule (Ri) désignent les consonnes de base qui composent la racine. Les voyelles (â, a, i,..) désignent les voyelles et les consonnes en minuscule (m,..) sont des consonnes de dérivation utilisées dans les schèmes. La majorité des verbes arabes ont une racine composée de 3 consonnes. L'arabe comprend environ 150 schèmes ou patrons dont certains plus complexes, tel le redoublement d'une consonne ou l'allongement d'une voyelle de la racine, l'adjonction d'un ou de plusieurs éléments ou la combinaison des deux [8].

• **Les racines « الجذر » :**

Une racine est purement consonantique, elle est formée par une suite de trois ou quatre (ou même cinq pour les noms) consonnes formant la base du mot. La racine est un élément important dans les langues dérivationnelles. En effet, à chaque racine correspond un champ sémantique et à l'aide de différents schèmes, on peut générer une famille de mots appartenant à ce champ sémantique .

Par exemple, la racine « كتب » (il a écrit) a la signification de base « écrire ». Plusieurs mots sont dérivés à partir de cette racine [6].

	La racine "كتب"(écrire)			
Verbes	كتب	Il a écrit	تكتب	Il écrit
	كتبا	Nous avons écrit	يكتبون	Ils écrivent
	كتبت	Elle a écrit	تكتب	Tu écris
	تكتبون	Vous écrivez	نكتب	Nous écrivons
Noms	كاتب	Ecrivain	كتابة	Ecriture
	كتاب	Livre	مكتوب	Ecrit
	مكتب	Bureau	اكتتاب	Enregistrement

Tableau 7: quelque dérivation du verbe كتب.

- **Les affixes**

Les affixes sont des lettres qui s'ajoutent au début (les préfixes) ou à la fin des mots arabes (les suffixes). En général, Ils sont utilisés pour accorder aux mots des éléments syntaxiques. Ils marquent l'aspect verbal, le mode, les propriétés transitives, etc. Ils sont aux alentours de 150 [7].

- **Les préfixes**

Dépendent des mots auxquels ils s'attachent. En effet, la plupart des mots arabes commencent par le préfixe « التعريف ال », « al altâryif » « l'article de définition » qui est utilisé en tant que terme déclaratif. Pour cela, il y a trois types de préfixes. Premièrement, les préfixes nominaux qui sont réservés pour les noms et les adjectifs. Deuxièmement, les préfixes verbaux qui sont réservés aux verbes. Et troisièmement, les préfixes généraux qui sont utilisés indépendamment du type des mots. La Tableau présente des exemples de chaque type de préfixes .

Type	Les préfixes			
	Nom en français	Signification	Nom arabe	Transcription
Préfixes nominaux	L'article de définition	Le	ال	Al (Lamltaarif)
	Les prépositions	Avec	ب	B
		pour	ل	L
		comme	ك	K
.....	.....	.....	.....	
Préfixes verbaux	La particule du futur	Sera	س	S
	Les particules du subjonctif	Pour	ل	L
	.....	.....	.....	.....
Préfixes généraux	Les conjonctions de coordination	Et	ف	F
	L'article d'interrogation	Est-ce-que	أ	A

**Tableau 8: un exemple des préfixes [9].**

Les préfixes peuvent s'enchaîner dans un mot pour former des préfixes composés qui peuvent atteindre jusqu'à quatre lettres (<وبال , wubal, et avec le|la|les>, <وال , wual, et le|la|les>, <بال , bal, avec le|la|les>, <كال , kal, comme le|la|les>, etc.).

Dans ce cas, certains préfixes ne peuvent prendre que la première position, l'article d'interrogation <أ, a> par exemple, d'autres peuvent prendre n'importe quelle position, l'article de définition <التعريف ال , al altâryif, l'article de définition > par exemple (Tableau 1.8) [9].

### Les suffixes

Il y a deux types de suffixes, les suffixes verbaux et les suffixes nominaux. Les premiers dépendent de la transitivité et de la personne conjuguée (Tableau 1.9). Les suffixes nominaux indiquent la flexion casuelle du nom (nominatif, accusatif, et génitif), le genre (masculin et féminin), le nombre (singulier, duel et pluriel), etc [1].

Type	Nombre	Suffixes		
		Signification	Nom en arabe	Transcription
Première personne	Singulier	Moi/mon	ني	Nyi
	Duel/Pluriel	Nous/Notre	نا	Na
Deuxième personne	Singulier	Toi/ton	ك	K
	Duel	Votre/vous	كما	Kma
	Pluriel	Votre/vous	كم	Km
		Votre/vous	كن	Kn
Troisième personne	Singulier	Lui/son	ه	H
	Duel	Eux/leur	هما	Hma
	Pluriel	Eux/leur	هم	Hm
		Eux/leur	هن	hn

**Tableau 9: un exemple des suffixes divisés selon leurs types [9].**

#### 1.4 La structure d'une phrase arabe

En langue arabe, il y a deux types de phrases : les phrases verbales et les phrases nominales. Les phrases verbales servent à indiquer un évènement ou une action. Elles commencent par un verbe suivi d'un sujet et d'un complément ; ce dernier est optionnel. La phrase verbale est celle que l'on rencontre le plus souvent dans l'expression courante. Par exemple الطالب جلس est traduit par l'étudiant s'est assis .

Les phrases nominales en arabe ne contiennent pas de verbe ; il est sous-entendu. En arabe, le verbe être est implicite, les verbes ne sont pas obligatoires pour construire une phrase. Les phrases nominales sont constituées d'un sujet et d'un attribut (adjectif qualificatif, complément circonstanciel, ...). Les phrases nominales sont affirmatives.

L'utilisation des phrases verbales, comme les phrases nominales, dépend du contexte dans lequel la phrase est exprimée. Dans le style journalistique, les phrases verbales sont les plus utilisées .

## Flexibilité de l'ordre des mots dans une phrase simple

L'ordre des mots dans une phrase simple en arabe est relativement flexible. Nous disposons, généralement, d'un libre choix du terme que nous voulons mettre en valeur, en tête de phrase ; la même phrase peut être ordonnée comme [7] :

### a) Verbe + sujet + complément :

(« سافر الرئيس الى بريطانيا », saāfara al-rraīyis ilaḡ briytaāniyaā)

Le président s'est rendu en Grande Bretagne.

### b) Sujet + verbe + complément :

(« الرئيس سافر الى بريطانيا », al-rraīyis saāfara ilaḡ briytaāniyaā)

le président qui s'est rendu en Grande Bretagne.

### c) Complément + verbe + sujet :

(« إلى بريطانيا سافر الرئيس », ilaḡ briytaāniyaā saāfara al-rraīyis)

C'est en Grande Bretagne que le président s'est rendu [7].

## 1.4.1 La syntaxe de la langue arabe

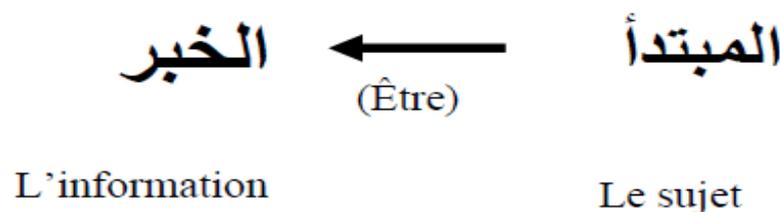
Elle étudie la structure correcte des phrases en analysant l'ordre des unités lexicales. La définition de la phrases, qui convient à l'anglais et au française, ne peut s'appliquer à tous les schémas de phrase de la langue arabe.

### Les types de phrase

#### ❖ Les phrases nominales

En langue arabe, une phrase nominale suit les règles suivantes :

- Elle commence soit par un nom ou soit par un pronom.
- Elle est composée de deux parties : la première partie appelée mubtada (المبتدأ) est le sujet de la phrase. La deuxième partie, appelée khabar (الخبر), apporte une information sur le sujet.



**Le mubtada : « المبتدأ »**

Le sujet en arabe al-mubatada est toujours composé d'un nom ou d'un pronom. Il est toujours au cas nominatif « مرفوع ». Le mubtada est le sujet de la phrase. Il peut représenter une personne ou une chose.

**Le khabar : « الخبر »**

Le khabar quant à lui, n'a pas l'obligation d'être composé d'un nom ou d'un pronom mais lui aussi est toujours au cas nominatif – مرفوع

Exemple : « محمد مجتهد »

محمد مجتهد

المُبتدأ - Le sujet - الخبر - L'information

Figure 9: la structure d'une phrase nominale [10].

### ❖ Les phrases verbales

Elles servent à indiquer un évènement ou une action. Elles commencent par un verbe suivi d'un sujet et d'un complément ; ce dernier est optionnel. La phrase verbale est celle que l'on rencontre le plus souvent dans l'expression courante, Elle est composée de 3 parties [10] :

- **Le Verbe** : Appelé en arabe « الفعل »

C'est la partie qui décrit une action effectuée dans le passé, le présent, ou le futur.

- **Le Sujet** : Appelé en arabe « الفاعل »

C'est la personne ou la chose effectuant l'action décrite par le Verbe.

- **Le Complément** : Appelé en arabe « المفعول ».

Et il y a deux genres :

- Direct : Il correspond à la personne ou la chose sur laquelle s'applique le Verbe.
- Indirect : C'est le complément qui indique le temps, le lieu, les circonstances... etc.[10].

**Remarque :**

Le verbe est la seule partie **obligatoire** pour former une phrase verbale. Les deux autres parties peuvent ne pas être présentes.

Exemple : « يذهب محمد الى الجامعة » .

يذهب محمد إلى الجامعة  
Le Complément Le Sujet le verbe  
Muhammad va à l'université

Figure 10: exemple d'une phrase verbale.

#### ❖ La phrase composée :

La phrase composée est une des phrases qui connecte plus d'une phrase, afin que le sens et l'idée de cette phrase soient conservés [8]

### 1.4.2 L'analyse syntaxique

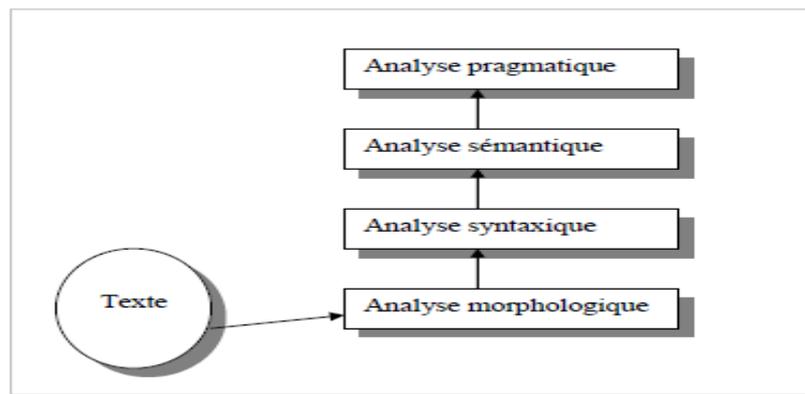
Le Traitement Automatique des Langues Naturelles (TALN) a fait l'objet de développement important ces dernières années.

L'analyse et la génération sont deux aspects différents pour le TALN écrite.

L'analyse est une suite de traitements (morphologique, syntaxique, sémantique,...), elle consiste à construire une représentation formelle du texte en entrée, cette représentation doit être facile à manipuler par la machine.

Formellement, la langue écrite est un ensemble de chaînes de caractères. On convient d'appeler MOT toute chaîne de caractères délimitée par des blancs. Par conséquent on admet qu'une phrase est constituée par une succession de mots. A partir des séquences de chaînes de caractères, différents niveaux d'analyses (traitement) peuvent être envisagé. On parle dans la littérature d'analyse morphologique, d'analyse syntaxique, d'analyse sémantique, et d'analyse pragmatique.[11]

Dans le cadre de cette thèse, nous nous intéressons plus particulièrement au niveau syntaxique Il convient donc, dans cette optique, de proposer dans ce qui suit un rapide survol des différents niveaux d'analyse.



**Figure 11: Les différents niveaux d'analyse d'un texte.**

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 vocaliques et grammaticales, relatives à la non voyellation des mots, pose des difficultés au niveau de l'analyse syntaxique. ainsi, une phrase, en absence de la voyellation, peut être interprétée et traduite selon plusieurs interprétations qui sont toutes syntaxiquement correctes.

Plusieurs méthodes d'analyse syntaxique se sont développées, mais la plus célèbre est sans doute la notion de grammaire formelle.

### **Représentations formelles de la phrase**

Plusieurs techniques sont utilisées pour représenter la phrase, notamment : [12]

#### **❖ Technique de la boîte**

Diverses formes de schématisation de la phrase ont été utilisées par les linguistes formels. Le linguiste Charles F.Hockett a commencé par proposer d'utiliser une boîte :

العالم يسير نحو التطور					
يسير نحو التطور			العالم		
التطور		يسير نحو		عالم	ال
تطور	ال	نحو	يسير	عالم	

Figure 12: Technique de la boîte [12]

Il est toutefois plus fréquent dans la tradition de N. Chomsky d'utiliser un parenthésage.

[Ph [SN [Dét ال ] [N m عالم ]][SV [V ير س ] [SN[Dét ال ] [N تطور ]]]]

❖ Représentation des arbres [12]

La représentation la plus employée est sans aucun doute celle de l'arbre syntaxique communément appelé indicateur syntagmatique

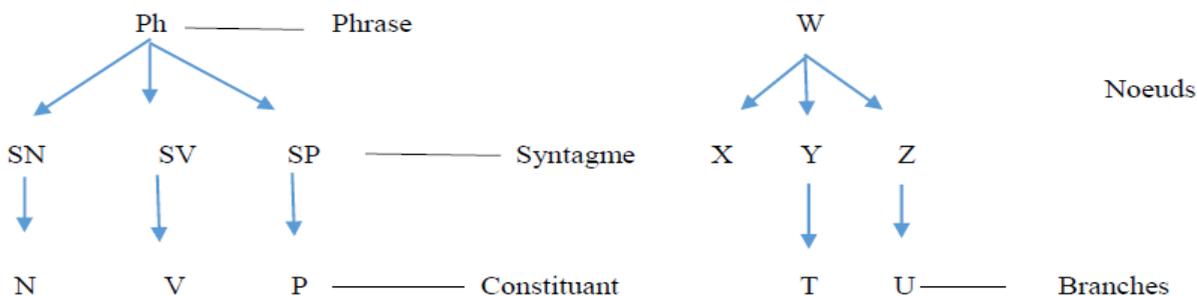
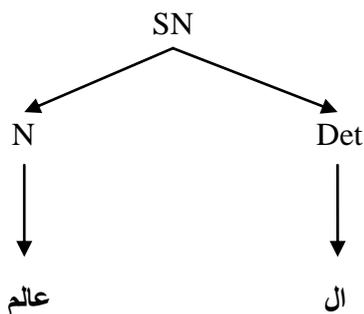


Figure 13: Un arbre est constitué de branches et de noeuds.

**Le syntagme nominal : SN**

- Le syntagme nominal est constitué en premier lieu d'un noyau nominal qui apparaît seul dans le cas de noms propres ou de pronoms.
- Ce noyau peut être accompagné d'un déterminant dans le cas des noms communs :



## Le syntagme verbal : SV

Le syntagme verbal se constitue d'abord d'un noyau verbal :

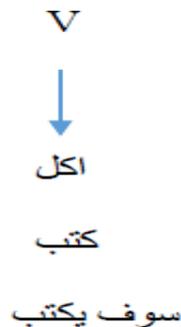


Figure 14: Le syntagme verbale

Il peut également y avoir des expansions au groupe verbal (SN, SP, SAdv, SA, proposition relative) Etant donnée une phrase, on peut donc l'analyser à l'aide des règles de la grammaire.

## Les stratégies d'analyse

### ❖ Analyse ascendante

L'analyse ascendante est également appelée « analyse dirigée par les données » car elle part des données que constituent les mots de la phrase à analyser. L'analyse ascendante essaie de combiner les éléments dans une phrase en entrée de différentes manières jusqu'à aboutir à un arbre qui couvre tous les mots de la phrase. Elle commence par les mots d'une phrase et procède par regroupement des mots en des composants de plus en plus larges jusqu'à ce que le tout soit regroupé sous la même racine (O'Donnell, 1994).

### ❖ Analyse descendante

Une analyse descendante (dite dirigée par les hypothèses) commence par la racine de l'arbre, puis construit les différents branchements jusqu'à atteindre les feuilles de l'arbre (généralement des mots). Théoriquement, une telle stratégie se base sur l'idée de produire, en utilisant une grammaire générative, toutes les structures syntaxiques possibles (O'Donnell, 1994).

L'analyse descendante permet d'éviter de construire des constituants qui, en fin d'analyse, se révéleraient impossibles à rattacher. Par contre, elle ne permet pas des analyses partielles; toute la phrase doit être reconnue (Aloulou, 2005).

## **1.5 Conclusion**

L'arabe, comme toutes les langues naturelles, est caractérisée par un ensemble de particularités créant des difficultés et des problèmes qu'il faut prendre en considération lors d'un traitement automatique comme l'ambiguïté, l'absence de voyelles, l'ordre des mots dans la phrase ...etc. Ce qui a conduit à l'absence d'un programme complet pour l'utiliser dans différentes applications.

Pour le chapitre suivant, nous allons en apprendre davantage sur l'une des théories appelée la théorie Néo-Khalilienne et les grammaires de pré-groupe qui permettent de développer le système TAL (traitement automatique de la langue).

# **Chapitre-II**

**La théorie néo-khalilienne et  
la grammaire de pré-groupe**

2

## 2.1 Introduction

Il a y eu beaucoup d'intérêt pour l'étude de la langue arabe ces dernières années. En plus de certains problèmes liés aux langues européennes (dont le traitement automatique est plus avancé) la langue arabe rencontre des difficultés liées à la structure des mots telle que l'agglutination et l'absence de voyelles à l'écrit. Ainsi la plupart des travaux se sont concentrés sur l'analyse morphologique et son utilisation dans les applications telles que la recherche d'information et les systèmes de questions/ réponse (QAS). Ces études ont montré que la langue arabe est fortement dérivative et flexionnelle dans laquelle la morphologie joue un rôle significatif.

Le traitement automatique de la langue arabe n'a débuté effectivement qu'a partir des années 1980 et ces dernières années ont vu des grands efforts avec des analyses en linguistique 'moderne' [Sproat, R.2007] et traitement informatique. Les premiers travaux des pionniers ont débuté depuis 1970 [Hlal, Y., 1979, 1985a] cité dans [Dichy, J. & all, 2007].

Dans ce qui suit nous présentons les efforts de formalisation en vue d'un traitement automatique de cette langue tels que la formalisation par les prégroupes, la théorie Néo-Khalilienne et diverses approches de traitement.

## 2.2 Théorie Néo-khalilienne (TNK)

L'approche « **Néo-khalilienne** » est un modèle linguistique présenté par le professeur Hadj-Salah qui est une formalisation des travaux des premiers linguistes arabes dont AL-Khalil Ibn Ahmed. Cette approche a donné lieu à plusieurs recherches avec des applications en informatique dans le cadre de magister au CRTDLA (Centre de Recherche pour le Trésor De La langue Arabe)

➤ Cette théorie repose sur les niveaux d'analyse *lexicale et syntaxique* est elle est définit par le professeur Hadj-Salah comme étant le *niveau centrale* de l'analyse qui se situe **entre** le lexique et la syntaxe. Les concepts fondamentaux sont « asl (origine) », « far' (pluriel furu') », « mithal (modèles linguistiques)<sup>1</sup> », «mawdhi' (position) », «qayis (pluriel maqayis : règles) » et «lexie». Pour les définitions voir [Hadj Salah A., 2003] cité dans [Loukam, M., 1999]

**2.2.1 Les Concepts Fondamentaux:**

Dans le cadre de l’analyse des résultats, nous avons exploité un certain nombre de concepts propres à la théorie néo-khalilienne que nous allons présenter dans ce qui suit:

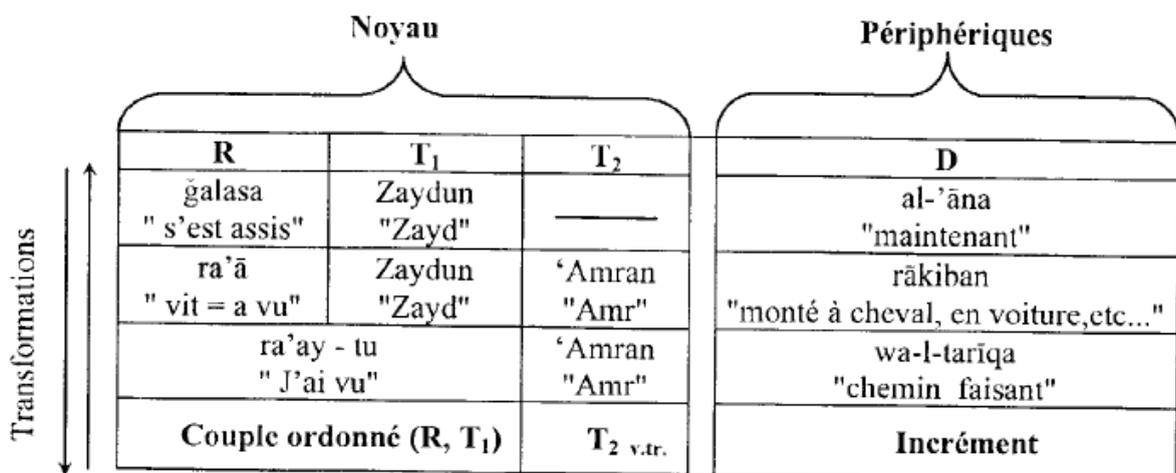
- **La racine** : représentée par l’ensemble des éléments formant les éléments de base de la kalima.
- **Le schème** : le moule qui donne forme à la kalima.
- **La kalima** : le résultat de la combinaison de la racine et du schème.
- **Le qiyas** : nous entendons par qiyās, la mise en correspondance des éléments entre eux pour déduire le schème auquel ils appartiennent.[13]

**2.2.2 La définition de la tectonie**

La tectonie comme entité sémiologique-grammaticale est constituée d’un noyau qui forme la racine du schème générateur syntaxique et d’incrémets qui s’ajoutent à ce noyau en postposition ou en antéposition, selon des positions structurelles déterminées, pour former des séquences dérivées.

Elle est représentée par deux types de Tectonie :

**La tectonie verbale: [14]**



La tectonie nominale : [14]

Noyau			Périphériques
R (régissant)	T <sub>1</sub> (terme régi 1)	T <sub>2</sub> (terme régi 2)	D (déterminant)
∅	Zaydun "Zayd"	ġālisun [est] "assis"	hunā "ici"
inna "Certes"	Zaydan "Zayd"	ġālisun [est] "assis"	al-yawma "aujourd'hui"
kāna "était"	Zaydun "Zayd"	ġālisun "assis"	ya-qra'u "en train de lire"
ḥasibtu "Je croyais"	Zaydan "Zayd"	ġālisun "assis"	ġalaṭan "par erreur"
A'lamtu 'Amran J'ai informé Amr [que]	Zaydan "Zayd"	ġālisun [est]"assis"	fawran "immédiatement"

Phrase minimale  
P<sup>l</sup>. de départ  
↑  
Transformations  
↓

La première colonne contient, comme elle le fait remarquer, un mot, un lexie ou une tectonie qui a un effet sur le reste des éléments ou de composition dans les deuxième et troisième colonnes.

Par conséquent, il est appelé le régissant (عامل), de sorte que l'élément de la deuxième colonne ne peut être présenté d'aucune façon en fonction de son régissant. C'est selon SIBAWAYH 1<sup>er</sup> terme régi (معمول اول) Lui et son régissant doivent être un mari arrangé.

Quant au deuxième terme régi (T<sub>2</sub>) trouvé dans la troisième colonne, il peut progresser sur tous les éléments et la position du régissant (R) peut être dépourvue de tout élément parlé, comme représenté par le signe de vacance (∅), comme c'est le cas avec la phrase vérifiée au début du tableau, "الجو غائم", qu'ils appellent le début "الإبتداء". Autrement dit, le manque syntaxique de dépendance à l'égard de tout élément antérieur à celui-ci, pas seulement le début de la phrase et il peut s'agir d'un mot» comme un mot ou un morphème : "إن" et "كان" "C'était" comme dans la phrase vérifiée "إن الجو غائم" ou "كان الجو غائما"

Il peut s'agir d'un lexie comme «حسبت» ou d'une tectonie comme "اعلمت زيدا", représenté par les deux phrases vérifiées du tableau. Ils sont respectivement: "حسبت الجو غائما" et "اعلمت زيدا الجو غائما" [14]

Il découle de ce qui précède que le régissant (R) selon des premiers grammairiens, son rôle comme mentionné précédemment - ne se limite pas à l'effet syntaxique, mais s'étend plutôt à la

relation entre les composants ou les éléments de la phrase, et ce lien, comme l'expliquent les anciens, est de deux types:

- ✓ **Un lien constructif ( ربط بناء )** : il relie les éléments de base de la phrase à un ensemble de relations, dont la plus importante est la relation d'efficacité dans laquelle le sujet est basé sur le verbe.
- ✓ **Connecter une ligature ( ربط وصل )**: relie les éléments non essentiels les uns aux autres. Ou entre eux et les éléments de base, comme la relation entre l'ajout et l'ajout: et l'adjectif et la description .. et la relation de ces éléments non essentiels avec les éléments de base.[14]

#### **HEME GENERATEUR DE LA LEXIE NOMINALE :**

«Lexie » est une unité fondamentale extensible à l'infini contrairement à la « kalima» qui n'admet qu'un nombre limité d'incrément. La lexie est définie à partir des critères de « ibtida» et «infisal (isolabilité) d'une séquence dans le discours

La lexie est une unité fondamentale qui se situe à l'intersection de la syntaxe et du lexique et même à l'intersection du syntaxique et du communicationnel puisque toute tranche ainsi définie peut fonctionner aussi comme un message minimal. La lexie minimale comme point de départ de l'analyse est le niveau intermédiaire entre le segment signifiant de l'unité syntaxique. À partir du niveau central, l'analyse du linguiste s'orientera vers le bas pour rechercher les segments signifiants et leurs composantes, ou vers le haut pour voir comment les minimaux séquentiels s'intègrent dans les structures syntaxiques. C'est une possibilité de génération «tafri» de séquences dérivées «furu», à partir d'une séquence primitive «asl» par laquelle est adjoint à la droite ou à la gauche de cette séquence des incréments significatifs «zawa'id) avec la capacité «tamakkun» et« tasaruf (variabilité) ». Cela permet d'établir des distinctions entre séquence isolable qui est d'une grande importance pour la langue arabe (langue agglutinante).

Certaines séquences ont un «tamakkun» parfait. Elles peuvent recouvrir n'importe quelle ligne d'incrément par exemple : «ism, gins». D'autres le sont moins par exemple les noms propres se font recouvrir d'articles de critères «d'infira' (ibtida, infisal) ». [15]

Les « tamakkun» permettent de remonter dans «lafdh» (la forme signifiante) en première unité. Les «mawadi's» déterminent toujours formellement les fonctions grammaticales des éléments qui les occupent. Toute tranche phoniques qui apparaît dans l'un des «mawadi's» est considérée comme «kalima segment». Ces signifiants s'obtiennent par élaboration du schème générateur de lexie ou d'une unité syntaxique (il s'agit d'un élément de la lexie)

L'approche adoptée se fait comme suit : partant du «Asl» (nom ou verbe), nous obtenons les différentes variantes de celui-ci par transformation, et ce par incrémentation en antéposition et/ ou en post-position placées dans des positions bien précises, formant ainsi à chaque niveau d'autres lexies voir Figure 15 et Figure 16.

**Explication**

Soit le mot posé un nom singulier« ism mufrad» comme «asl = kita :b». Par incrémentation de part et d'autre du mot «kitab» les ajouts 'valables' effectués ne concernent que le nom et sont positionnés de telle sorte qu'ils répondent 'valablement' à la syntaxe. Ces ajouts sont 'valables' à l'aide de «mitral et «qiyas» (de modèles préalablement établis).

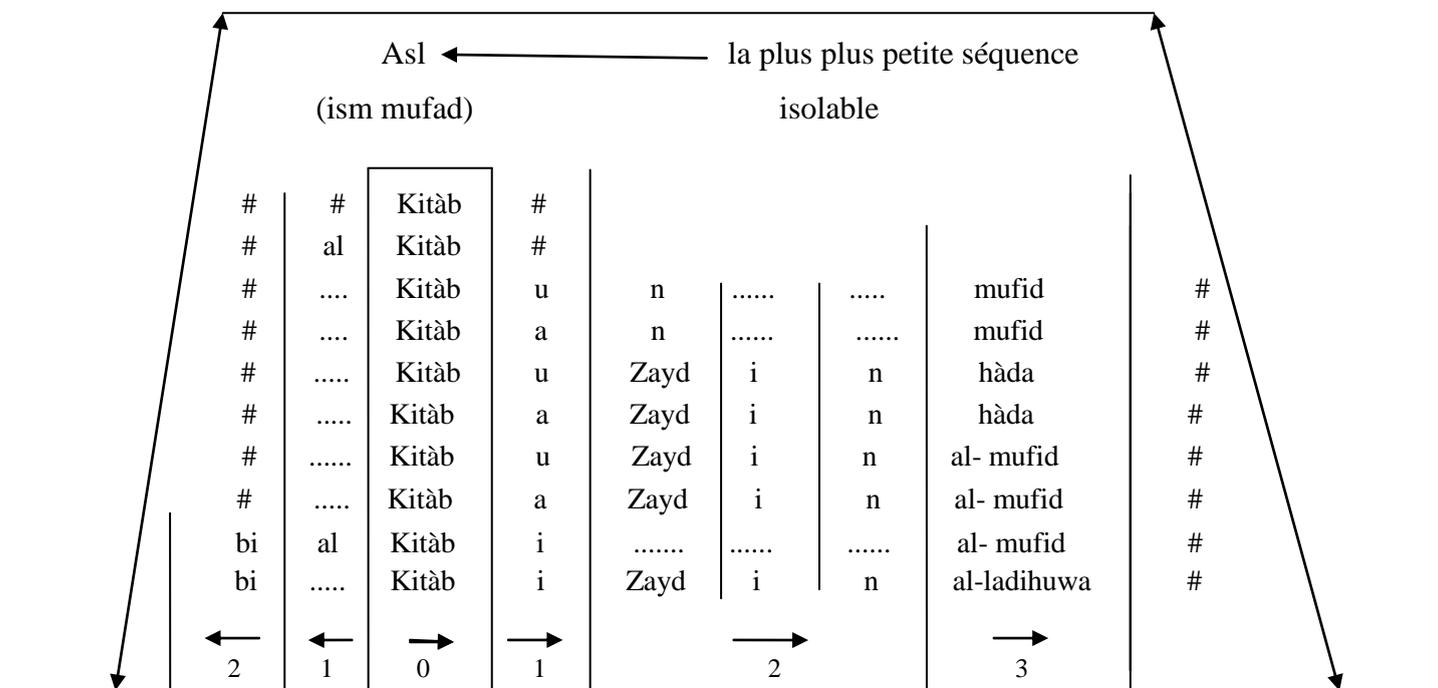


Figure 15: SCHEMA GENERATEUR DE LA LEXIE NOMINALE [15]

Pour le verbe, la même méthode est appliquée avec des variations en rapport à la conjugaison. Il est fait appel à la variation interne du noyau par «bina».

Syntaxiquement cette approche utilise les notions de «' mal terme régissant » et «ma'mul terme régi. Dans cette approche « ' mal » permettra l'accroissement de la longueur de la lexie en respectant les flexions casuelles «Eli 'rab».[15]

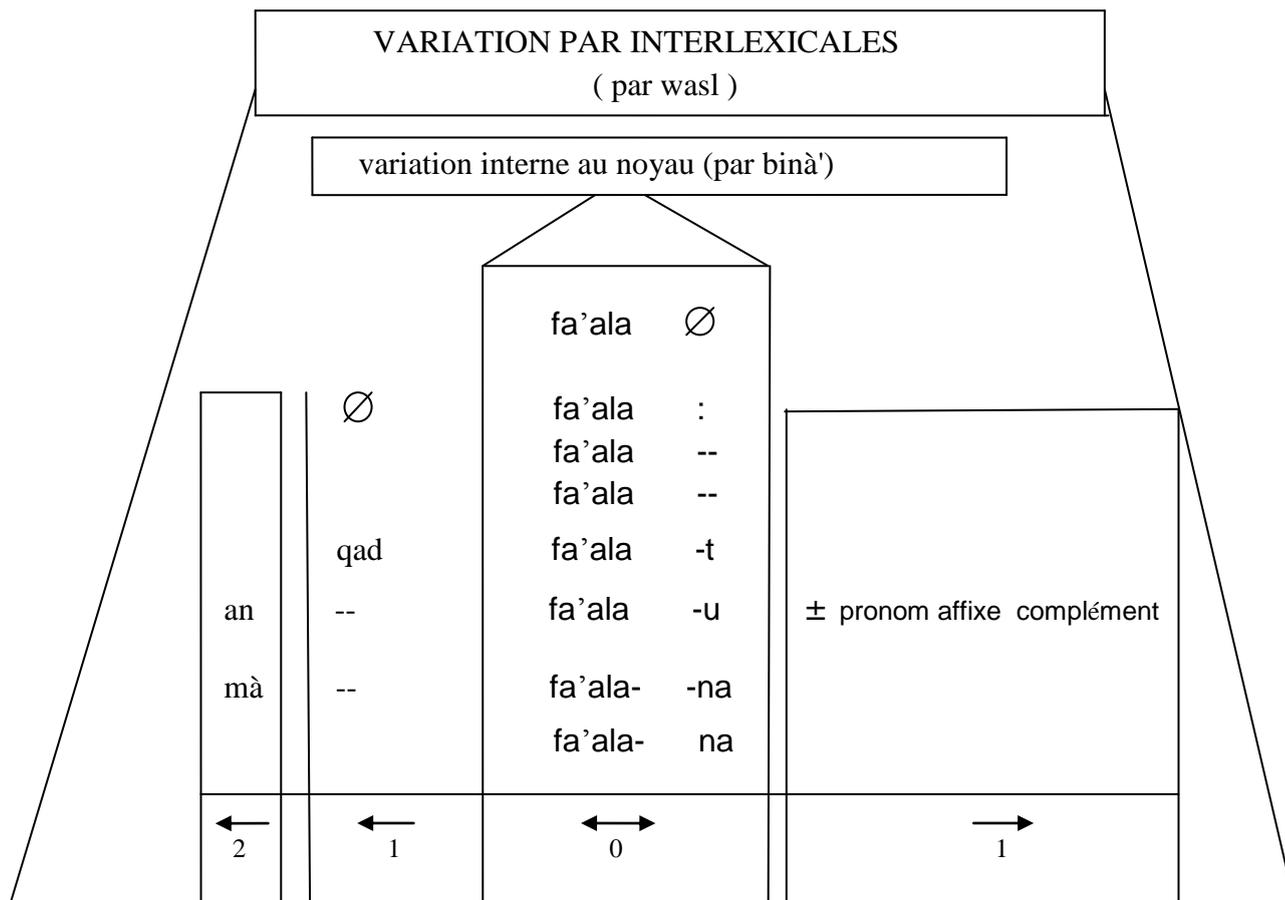


Figure 16: SHEME GENERATEUR DE LA LEXIE VERBALE CAS DE L'ACCOMPLI

### Concept de tectonie:

La théorie néo-khalilienne est une théorie linguistique fondée par le linguistique algérien Abderrahmane Hadj-Salah (1979,2011) dans la cadre d'une interprétation globale de nahw al-arabi ( النحو العربي ) (qui veut dire littéralement la grammaire arabe). Nous proposons de décrire les structures syntaxiques de l'arabe en utilisant la tectonie, définie dans le cadre de la théorie néo-khalilienne comme l'unité linguistique composée de termes régissants (R), de terme régis (T<sub>i</sub>) et d'unités périphériques (D). [14]

$$\left[ \left( R \longrightarrow T_1 \right) \pm T_2 \right] \pm S$$

Où **R** : est le régissant

**T<sub>1</sub>** : 1<sup>ier</sup> terme régi. »

→ : Le premier opérateur se lie lier terme régi (T<sub>1</sub>) à son régissant (R) de dépendance, de sorte que ce dernier ne peut pas progresser sur lui.»

**T<sub>2</sub>** : 2<sup>ième</sup> terme régi.

± : Une référence à la présence ou à l'absence du second régissant(R<sub>2</sub>) .

**S** : élément non essentiel.

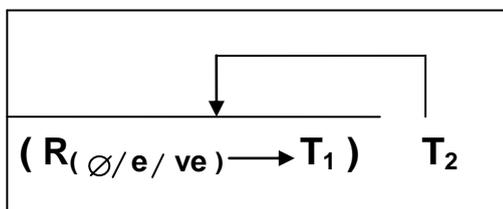
↖ : Une flèche indiquant que 2<sup>ième</sup> terme régi (T<sub>2</sub>) peut être en avance sur le régissant (R) et son 1<sup>ier</sup> terme régi (T<sub>1</sub>).

### 2.2.3 LES STRUCTURES SYNTAXIQUES FONDAMENTALES :

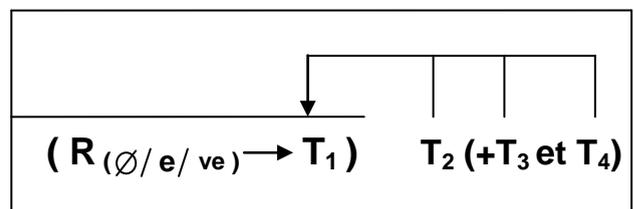
a) : Une structure basée sur l'btida' (الإبتداء) (le haml 'ala I-ism ci-dessus) que nous représenterons par la formule ci-après:

( e= exposant; ve = verbe exponentiel ; v= verbe , nv = nom ou adjectif verbal ) :

#### I-FORMULE A : [13]



Ou



Où: 1. [ R ( ≠ ∅ ) = 0 ] → [1] ou [2]

2. ( T<sub>i>1</sub> = 0 ) ↔ bina' 0

3. R<sub>ve</sub> = Kalima ou lexie ou tectonie , Exemple:

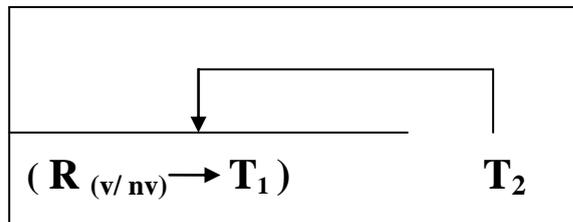
- Kalima	Kāna	Zaydun	qä'iman .... #
- Lexie	hasibtu	Zaydan	qä'iman .... #
- Tectonie	'a'lamtu ' Amran	Zaydan	qä'iman .... #
	$R_{ve}$	$T_1$	$T_2$

avec  $R_{ve}$  ( $T_2$  pronom affixe)  $\leftrightarrow$  tectonie dont le contenu coïncide avec une lexie verbale comportant un incrément pronominal (classe de Kana) -

ou 1<sup>er</sup> terme possible d'une tectonie  $\leftrightarrow$  classe de [ hasibtu ] et [ 'a'lama ]. [13]

**b) : Une structure basée sur un verbe ( non exponentielle ) ( le haml 'ala l-fi'1 ) qu' on peut représenter ainsi:**

**II -Formule B :**



Ou :

1. ( $R = 0$ )  $\leftrightarrow$  bina' 0
2. ( $T_2 = 0$ )  $\leftrightarrow$  1<sup>er</sup> terme possible d'une tectonie (couple ordonné :  $R \rightarrow T_{1+}$  actualisable sans  $T_2$ )
3. ( $T_2 =$  pronom affixe)  $\leftrightarrow$  tectonie dont le contenu coïncide avec une lexie verbale comportant un incrément pronominal. [13]

## 2.3 Grammairs de pregroupe

Les grammaires de pré-groupe fournissent une approche informatique des langues naturelles en attribuant à chaque mot du dictionnaire mental un type, à savoir un élément du pré-groupe généré librement par un ensemble partiellement ordonné de types de base. Dans ce travail, nous essayons de présenter aux linguistes une grammaire anglaise précédant le groupe en examinant les exemples les plus anciens de Chomsky.

### 2.3.1 Propriétés fondamentales

Nous rappelons la définition des pré-groupes, la construction d'un pré-groupe librement généré par [Lambek 1999] et la caractérisation géométrique des dérivations dans les grammaires de pré-groupe sous forme de graphiques de sous-liens. Un monoïde récommandé  $\langle P, 1, \cdot, \rightarrow \rangle$  est un ensemble  $P$  avec un élément distingué  $1$ , a opération binaire  $\cdot$ . Est une relation binaire  $\rightarrow$  satisfaisant pour tout  $a, b, c, u, v \in P$

- $1 \cdot a = a = a \cdot 1$
- $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- $a \rightarrow a$
- $a \rightarrow b$  et  $b \rightarrow c$  implique  $a \rightarrow c$
- $a \rightarrow b$  implique  $u \cdot a \cdot v \rightarrow u \cdot b \cdot v$ .

Le point dénotant la multiplication est généralement omis. On dit que deux éléments  $a$  et  $b$  ne sont pas comparables si  $a \not\rightarrow b$  et  $b \not\rightarrow a$ . Et la notification d'un pre-groupe, un monoïde ordonné dans lequel chaque élément  $b$  un adjoint gauche  $b^l$  tel que :  $b^l b \leq 1 \leq b b^l$

Et un droit adjoint  $a^r$  tel que :  $b b^r \leq 1 \leq b^r b$

À partir des types de base, nous construisons des types simples :  $\dots, b^{ll}, b^l, b, b^r, b^{rr}, \dots$ .

Un type composé est une chaîne de types de base. [39]

### 2.3.2 Pré-grammaire

Développer une grammaire de pré-groupe pour un langage naturel consiste en deux étapes principales :

- (i) attribuer un ou plusieurs types (de base ou composé) à chaque mot du dictionnaire.
- (ii) vérifier la grammaticalité et la phrase d'une chaîne de mots par un calcul sur les types correspondants.

Où les seules règles impliquées sont : Contractions, postulats d'ordre prenant la forme  $\alpha \rightarrow \beta$  (types de base  $\alpha, \beta$ ) et conditions appropriées introduites dans le lexique, appelés métarules. La collection  $F(A)$  de tous les types de composés est partiellement ordonnée par la commande héritée de  $A = \{\Pi_1, \Pi_2, \Pi_3, S_1, S_2, P_1, P_2, O, \dots\}$  et par les règles suivantes [17] :

### 2.3.3 Contractions et Extensions

Un type composé est une chaîne de types de base. La collection  $F(A)$  de tous les types de composés est partiellement ordonnée par la commande héritée de  $A$  et par les règles suivantes :

- Contractions :  $a^l a \rightarrow 1, a^l a^l \rightarrow 1, a^r \rightarrow 1, \text{etc.},$
- Extensions :  $1 \rightarrow a a^l, 1 \rightarrow a^l a^l, 1 \rightarrow a^r a, \text{etc.}$

On dérive

- $a \rightarrow b$  si et seulement si  $b^l \rightarrow a^l$  si et seulement si  $b^r \rightarrow a^r,$
- $a \rightarrow b$  si et seulement si  $a b^r \rightarrow 1$  si et seulement si  $b^l a \rightarrow 1.$

On comprendra que la relation d'ordre de  $F(A)$  satisfait la règle de substitution : Si  $\alpha \leq \beta$  alors  $\gamma \alpha \delta \leq \gamma \beta \delta.$

$F(A)$  est notre hiérarchie de types ; techniquement parlant, il s'agit du pré-groupe libre généré par  $A$ , mais le lecteur sur le plan mathématique peut l'ignorer. Nous allons assigner des types composés aux mots anglais et autres entités grammaticales. Pour vérifier qu'une chaîne de mots, disons de type  $\alpha$ , est une phrase, disons un énoncé de type  $S$ , nous devons vérifier que  $\alpha \leq S.$  [17]

### 2.3.4 Le pré-groupe libre

Le pré-groupe libre  $P(B)$  généré par un ensemble partiellement ordonné de types de base  $B$  est caractérisé dans [Lambek 1999] en tant que monoïde libre généré à partir d'un ensemble de types simples  $\Sigma$  consistant en des jonctions itérées de types de base

$$\Sigma = \{a^{(z)} : a \in B, z \in \mathbb{Z}\}.$$

Nous appelons  $a$  la base et  $Z$  l'itérateur du type simple  $t = a^{(z)}$ . Les types de base  $a \in B$  sont identifiés avec  $a^{(0)} \in \Sigma$  et sont donc inclus dans les types simples. Les éléments de  $P(B)$  sont appelés types, ils sont de la forme.

$$a_1^{(z_1)} \dots a_K^{(z_K)},$$

Où  $a_1, \dots, a_K$  sont des types de base et  $z_1, \dots, z_K$  sont des entiers. L'unité 1 indique la chaîne vide et la multiplication est identique à la concaténation. Les jointures gauche et droite d'un type sont définies par :

$$\begin{aligned} (a_1^{(z_1)} \dots a_K^{(z_K)})^l &= a_K^{(z_K-1)} \dots a_1^{(z_1-1)} \\ (a_1^{(z_1)} \dots a_K^{(z_K)})^r &= a_K^{(z_K+1)} \dots a_1^{(z_1+1)}. \end{aligned}$$

Par conséquent, pour un type de base  $a$  nous avons :

$$a^{ll} = a^{(-2)}, a^l = a^{(-1)}, a = a^{(0)}, a^r = a^{(1)}, a^{rr} = a^{(2)}, \text{ etc. [17]}$$

### 2.3.5 Pré-groupes pour la linguistique

Nous reprenons ici un exemple linguistique dû à Lambek (Lambek, 2001). Ces exemples utilisent les types de base suivants :

$\pi_i$  :  $i$  - la personne au singulier (tous les pluriels en anglais moderne étant donnés comme type 2 à vous).

$S_i$  : une phrase au présent ( $i=1$ ) ou au passé ( $i=0$ ).

$S$  : une phrase correcte,  $S_1 \leq S$  et  $S_2 \leq S$ .

$P_i$  : un participe, avec  $i = 1$  pour le présent et  $i = 2$  pour le passé.

$O$  : Complément d'objet direct.

$q$  : question en oui ou non.

$q_0$  : question.

$n$  : un terme, avec  $n \leq \pi_3$  et  $n \leq O$ . [16]

Dans l'intérêt d'une introduction douce, nous absorbons pour le moment le fait que A n'est qu'un ensemble, c'est-à-dire que l'ordre est discret.

$$A = \{\Pi_1, \Pi_2, \Pi_3, S_1, S_2, P_1, P_2, O, \dots\} .$$

Les exemples de phrases suivants traduiront l'idée principale :

❖ She goes

of type  $\pi_3 (\pi^r_3 S_1) = (\pi_3 \pi^r_3) S_1 \leq S_1,$

❖ you had gone

of type  $\pi_2 (\pi^r_2 S_2 P^1_2) P_2 = (\pi_2 \pi^r_2) S_2 (P^1_2 P_2) \leq S_2,$

❖ I saw her

of type  $\pi_1 (\pi^r_1 S_2 O^1) O = (\pi_1 \pi^r_1) S_2 (O^1 O) \leq S_2,$

❖ he is seeing me

of type  $\pi_3 (\pi^r_3 S_1 P^1_1) (P_1 O^1) O = (\pi_3 \pi^r_3) S_1 (P^1_1 P_1) (O^1 O) \leq S_1,$

❖ we were seen

of type  $\pi_2 (\pi^r_2 S_2 O^{ll} P^1_2) (P_2 O^1) = (\pi_2 \pi^r_2) S_2 (O^{ll} (P^1_2 P_2) O^1) \leq S_2,$

Ici le tiret dénote une trace de Chomsky.

Bien que les traces ne soient pas vraiment nécessaires pour notre description, nous les avons insérées pour les comparer à la littérature.

Le double l,  $O^{ll}$  est la manière dont la théorie des types réfléchit les traces. Nous notons que les

formes finies des verbes go, see, have et be, peuvent être organisées dans matrices de conjugaison (en ignorant le subjonctif presque obsolète) :[17]

<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">go</td><td style="padding: 2px 10px;">go</td><td style="padding: 2px 10px;">goes</td></tr> <tr><td style="padding: 2px 10px;">went</td><td style="padding: 2px 10px;">went</td><td style="padding: 2px 10px;">went</td></tr> </table>	go	go	goes	went	went	went	of type $\pi_k^r S_i$ ,
go	go	goes					
went	went	went					
<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">see</td><td style="padding: 2px 10px;">see</td><td style="padding: 2px 10px;">sees</td></tr> <tr><td style="padding: 2px 10px;">saw</td><td style="padding: 2px 10px;">saw</td><td style="padding: 2px 10px;">saw</td></tr> </table>	see	see	sees	saw	saw	saw	of type $\pi_k^r S_i O^1$ ,
see	see	sees					
saw	saw	saw					
<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">have</td><td style="padding: 2px 10px;">have</td><td style="padding: 2px 10px;">has</td></tr> <tr><td style="padding: 2px 10px;">had</td><td style="padding: 2px 10px;">had</td><td style="padding: 2px 10px;">had</td></tr> </table>	have	have	has	had	had	had	of type $\pi_k^r S_i P_2^1$ ,
have	have	has					
had	had	had					
<table style="border-collapse: collapse; width: 100%;"> <tr><td style="padding: 2px 10px;">am</td><td style="padding: 2px 10px;">are</td><td style="padding: 2px 10px;">is</td></tr> <tr><td style="padding: 2px 10px;">was</td><td style="padding: 2px 10px;">were</td><td style="padding: 2px 10px;">was</td></tr> </table>	am	are	is	was	were	was	of type $\pi_k^r S_i P_1^1$ or $\pi_k^r S_i O^{II} P_2^1$
am	are	is					
was	were	was					

Le choix dans la dernière assignation de type dépend du fait que le verbe auxiliaire soit utilisé pour l'aspect progressif ou la voix passive, respectivement. De plus, nous avons les formes non finies :

- Going, Gone de type  $P_i$
- Seeing, Seen de type  $P_i O^1$
- Being, Been de type  $P_i O^{II} P_2^1$

Une fois que nous avons décidé de permettre aux types de base d'être commandés, nous pouvons en introduire d'autres.

**S** : pour les déclarations où le temps n'a pas d'importance,

**$\pi$**  : pour les pronoms nominatifs quand la personne n'a pas d'importance,

Avec :  $\pi_k \leq \pi$  ,  $S_i \leq S$ .

### Phrase question

Nous faisons la distinction entre les questions oui ou non de type  $q_1$  au présent et  $q_2$  au passé. Pour traiter les questions-réponses, nous introduisons le type  $q'$  pour la classe de toutes les questions.

Nous supposons que le temps d'une question n'est pas pertinent, qu'il s'agisse d'une demande actuelle d'informations sur le présent ou le passé. Lorsque le temps d'une question oui-ou-non n'est pas pertinent, nous lui attribuons un nouveau type de base  $q$  et postulons [18] :

$$q_1 \leq q, q_2 \leq q, q \leq q'$$

Voici des exemples de questions oui ou non :

Does he see her ?

Did he see her ?

De types respectifs :

$$(q_1 \text{ il } \pi_3^1) \pi_3 (i O^1) O$$

$$(q_2 \text{ il } \pi^1) \pi_3 (i O^1) O$$

### Phrase nominale

Jusqu'ici, les seules phrases nominales prises en compte étaient les pronoms de type  $\pi_k$  ( $k = 1,2,3$ ) et les pronoms accusatifs de type  $o$ , tels que *me* et *her*. Il faut ajouter à ces noms des noms tels que *Jane*, *Napoléon Bonaparte*, *l'inspecteur Morse*, etc., de type  $n$ , et il faut préciser :

$$n \leq \pi_3, n \leq O,$$

La plupart des phrases nominales sont construites à partir de noms, où nous distinguons :  
Noms de masse de type  $m$ , par exemple *water*, *pork* ;

Compter les noms de type  $c$ , par ex. *apple*, *pig* ;

Les pluriels de type  $p$ , par ex. *police*, *pigs*.

Les noms de masse et les pluriels peuvent apparaître sans déterminants [18].

Nous devons donc nous assurer que :

$$m \leq \pi_3, m \leq O, P \leq \pi_2, P \leq O,$$

### Phrase négation

La négation grammaticale d'une phrase anglaise s'obtient en introduisant le mot *not* de type  $xx_1$ , où  $x = i, j, P_1$  ou  $P_2$ , à un endroit approprié.

Cela devrait représenter :

- She does not go,
- She is not going,
- I was not seen—,
- You have not gone,
- We will not be going ;

Par exemple, le troisième exemple de phrase a le type :

$$\pi_1 (\pi^r_1 S_2 O^{ll} P^l_2) (P_2 P^l_2) (P_2 O^l) \leq S_2 O^{ll} O^l \leq S_2,$$

Et le cinquième a le type :

$$\pi_2 (\pi^r S_1 j^l) (j j^l) (i P^l_1) P_1 \leq S_1, \quad \text{puisque } \pi \leq \pi \quad \text{et} \quad i \leq j.$$

Nous avons tenté une formulation de la grammaire anglaise basée sur une affectation de types à des mots anglais, les types étant des éléments d'un pré-groupe, un monoïde ordonné dans lequel chaque élément possède à la fois un adjoint gauche et un adjoint droit [18].

## 2.4 Conclusions

La présente étude donne un bref sur que çu l'adéquation scientifique et didactique de la théorie Néo- khalilienne, élaborée par le linguiste Algérien, Abderrahmane Hadj Salah, tout en mettant l'accent sur les composantes de l'adéquation scientifique telle que (La tectonie nominale et la tectonie verbale .....).

La tectonie comme entité sémio logico-grammaticale est constituée d'un noyau qui forme la racine du schème générateur lexical et d'incrémentes qui s'ajoutent à ce noyau en postposition ou en antéposition, selon des positions structurelles déterminées, pour former des séquences dérivées. elle est représentée par deux types de lexie : La tectonie nominale et La tectonie verbale.

La grammaire de pré-groupe est un modèle algébrique de grammaire sous la forme d'un pré-groupe, fondé par le mathématicien canadien Joachim Lambek.

L'idée d'une grammaire de pré-groupe est d'attribuer à chaque mot un ou plusieurs types et vérifier si une chaîne de mots donné est une phrase bien formée en effectuant un calcul par dérivation chaîne de type correspondant.

# ***Chapitre-III***

**Pour l'analyse une approche algébrique de la  
structure des tectonies nominales en arabe**

3

### 3.1 Introduction

Les grammaires de pré-groupe ont été utilisées pour étudier la structure des phrases dans un certain nombre de langues européennes. Nous nous tournons ici vers l'arabe comme notre premier exemple de langue non européenne (et non indo-européenne). Nous utilisons la notion de grammaire de pré-groupe pour décrire la structure d'une lexie verbale dans un fragment de la langue l'arabe moderne standard basée sur la théorie Néo-Khalilienne, qui est une théorie linguistique contemporaine basée sur le travail des contributions des premiers grammairiens arabes ; elle a été fondée par la linguiste algérien Abderrahmane Hadj Salah.

### 3.2 La machinerie algébrique

Nous avons étudié les formes nominales finies ailleurs et montré qu'elles peuvent toutes être calculées à partir d'une expression formelle  $D_{gncd}(N)$ :

où  $g = m$  ou  $f$  est le genre (masculin ou féminin),  $n = 1, 2$  ou  $3$  est le nombre (singulier, double ou pluriel),  $c = 1, 2, 3$  désigne les trois cas nominatif, accusatif et un troisième cas qui ressemble au latin génitif ou ablatif, selon le contexte, et  $d = 1, 2$  désigne respectivement de nom déterminé ou indéterminé )

lorsque  $c = 1 / 2 / 3$  respectivement.

Pour les noms communs, nous distinguons trois cas:

$$D_{gnc1}(N) \longrightarrow \text{al} + D_{gnc}(N)$$

$$D_{g1c2}(N) \longrightarrow D_{g1c} N + \mathbf{n}$$

$$D_{g2c2}(N) \longrightarrow D_{gC} N:$$

les types des articles doivent être classés comme suit:

$$\mathbf{AL} : n_{gnc1} n_{gnc}^1 ;$$

L'idée principale est d'attacher à chaque mot, parfois aussi à des morphèmes, un ou plusieurs types. Pour commencer, on suppose qu'un ensemble partiellement ordonné de types de base est donné, par exemple, **R** et **S** pour les tectonies nominales.

A partir des types de base, on forme des types simples : on obtient ainsi  $a^l$  et  $a^r$ , que nous appelons les adjoints gauche et droit de  $a$  respectivement. Pour les langues européennes modernes, les doubles accolades  $a^{ll}$  et  $a^{rr}$  étaient également nécessaires, mais les jointures itérées ne semblent pas jouer un rôle dans le fragment d'arabe que nous avons étudié jusqu'à présent.

Les types à attacher aux mots sont des chaînes de types de base.

- Par exemple :

.  $a^l a \rightarrow 1$ ,  $aa^r \rightarrow 1$ , Où 1 est la chaîne vide.

Par exemple:

- ❖ Mohamadun mojtahidun

De type :  $(R) n^L n^{LL} = R(n^l n^{ll}) \rightarrow R(1) = R$

Ce qui implique que cette expression est une tectonie nominale.

### 3.3 Types de base

Ce travail utilise les notions de grammaire de pré-groupe pour analyser la lexie du verbe arabe, comme suit :

- $\Pi_J$  = le J-ème pronom de la personne, où  $J = 1, 2M, 2F, 3M, 3F, 4, 5M, 5F, 6M, 6F, 7, 8M, 8F$  représente les trois personnes singulières suivies des trois personnes plurielles suivi des trois personnes du duel, m signifie masculin et f signifie féminin.

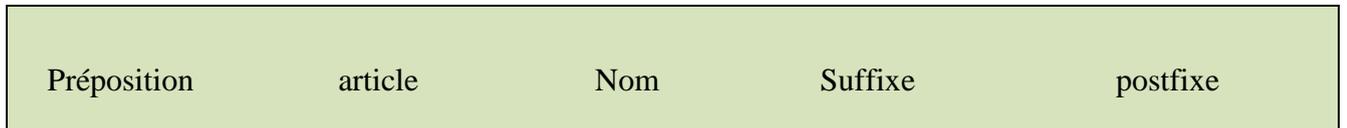
- R = déclaration.

- $S_I$  = énoncé dans le i -ème régime, où  $I = 0$  pour le passé et  $I = 1$  pour le présent,  $I = 2$  pour l'impératif et  $I = 4$  pour le futur.

- $O_J$  = objet direct, où  $J = 1, 2M, 2F, 3M, 3F, 4, 5M, 5F, 6M, 6F, 7, 8M, 8F$  indique trois personnes au singulier suivi des trois personnes du pluriel suivi des trois personnes du duel, m signifie masculin et f signifie féminin.

### 3.4 La structure du nom arabe

Les formes nominales rendent compte de l'agglutination à droite des prépositions aux noms, du caractère déterminé des noms et de leurs déclinaisons. Elles ont la structure suivante:



Exemple de la structure du nom arabe :

بالكتاب qui est phonétiquement [Bi] [Al] [Kita:] [b:]

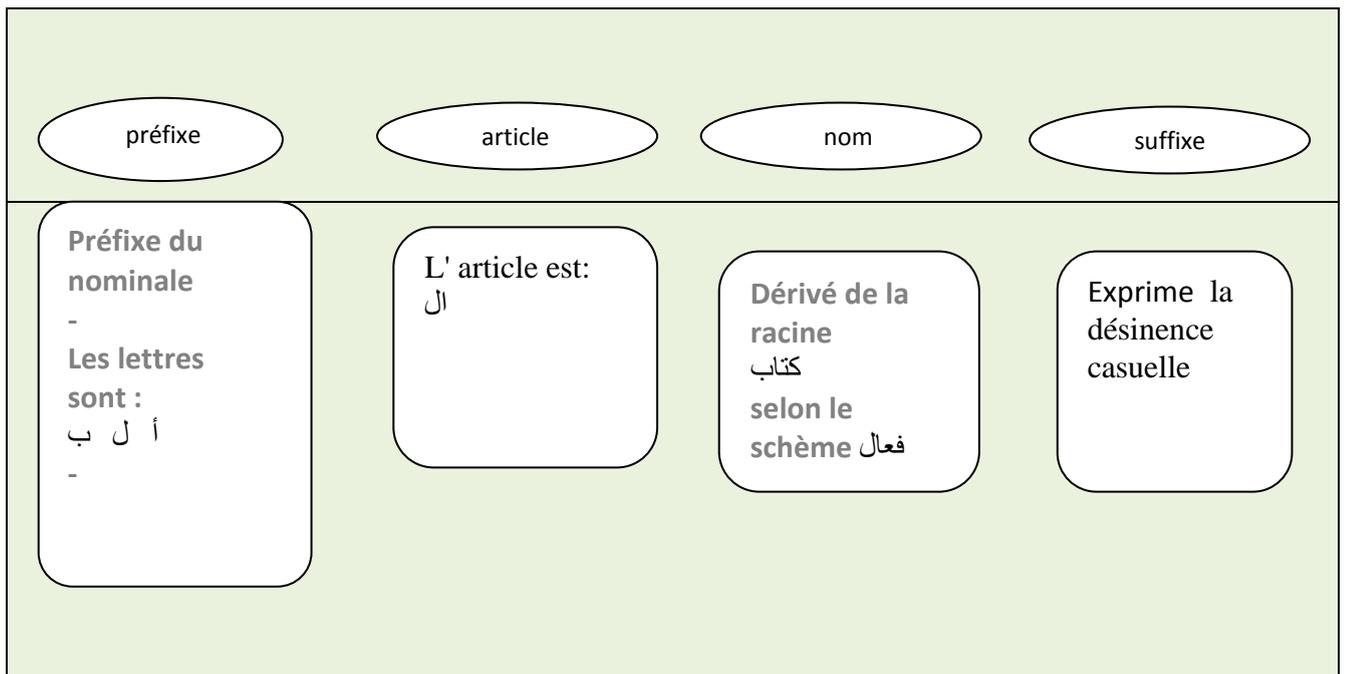
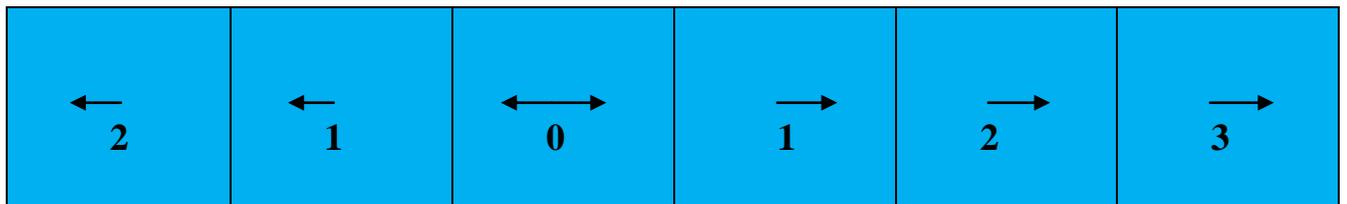


Figure 17: structure du nom arabe بالكتاب

### 3.5 La structure des lexies nominales en arabe

Elle est basée sur ce que nous avons étudié au chapitre 2 en la théorie néo-khalilinne ( Schème du Le ḤADD général du verbe ). on peut conclure la représentation suivante schématise une structure des lexies verbales en arabe. Notons bien que la lecture se fait de gauche à droite.



- ❖ La position  $\overleftrightarrow{0}$  position du ism qui fonctionne comme élément centrale ou noyau de la lexie outils.
- ❖ La position  $\overleftarrow{1}$  est la position du déterminant identifiant ( ادوات التعريف ) précède immédiatement l'élément centrale
- ❖ La position  $\overleftarrow{2}$  est la position de l'éexposant annectif(hurouf eljar): cet élément peut donc précéder immédiatement le noyau que ce dernier soit déterminé ou non .
- ❖ La position  $\overrightarrow{1}$  : est la position de la désinence casuelle : suit immédiatement le nouau centrale,
- ❖ La position  $\overrightarrow{2}$  : la position du déterminant par annexion direct (مضاف اليه).
- ❖ La position  $\overrightarrow{3}$  : la position du déterminant caractérisant : ce déterminant peut être une sifa dérivée (adjectif codérivé du verbe) ou un mawsoul avec un sila ( relatif + séquence introduite par ce relatif )

### 3.6 Les formes catégories du nom en arabe:

En Arabe, les noms peuvent être classés en 3 sous catégories: les noms primitifs, les noms dérivés et les nombres . Cette classification se base sur le rattachement du nom à une racine verbale ou non. On présentera dans ce qui suit un panorama de ces différents sous ensembles. Pour chaque sous-catégorie on donnera un tableau montrant les fonctions grammaticales que peuvent occuper ces noms ainsi que le type catégoriel qui leur correspond.

### 3.6.1 Les noms primitifs:

Cette sous-catégorie regroupe l'ensemble de noms arabes qui ne sont pas dérivés d'un verbe. Ces noms gardent la même forme indépendamment du contexte dans lequel ils sont énoncés (voir le Tableau VII). Ces noms peuvent être: des noms propres, des pronoms personnels, des prénoms démonstratifs, . .

	Fonctions Grammaticale	Exemples
Noms Primitifs	Noms propres	سَلْمَى ، تَمِيمٌ ، مَهْدِي
	Pronoms personnels (ضمير منفصل)	أنا، نحن، هم، ت
	Prenoms relatifs (إسم موصول)	الذي، التي
	Pronoms démonstratifs	هذا، هذه، ذلك، هؤلاء
	Prépositions	في، على، فوق، تحت
	Les conjonctions de coordination	ثم، و
	Adverbes	يث، أينح
	Quantificateurs	كل، بعض
	Noms Communs	سيارة، شجرة
	Noms du verbe	أه، أف
	Noms de voix	كخ، نخ
	Pronoms interrogatifs	متى، كيف، أين
	Pronoms allusifs	كم، كي

Tableau 10: Les noms primitifs dans la langue arabe.

### 3.6.2 Les noms dérivés:

Comme nous l'avons énoncé ci-dessus, cette sous-catégorie contient l'ensemble de noms qui sont dérivés de la racine d'un verbe donné. Ces noms peuvent avoir différentes fonctions grammaticales comme le montre le tableau VU!.

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

Tableau 11: Les noms dérivés dans la langue arabe.

### Pronoms dans la langue arabe:

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

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

- X E {1,2,3} : les chiffres 1,2 et 3 font référence respectivement la première la deuxième et la troisième personne.
- y E {m, f, t/J} : les lettres m et f désignent respectivement les genres masculin et féminin. Dans quelques cas, y peut avoir la valeur nulle (t/J).
- Z E {s, d, p} : Les lettres s, p, et d désignent respectivement le singulier et le pluriel et le dual.

Personne	Genre	Pronom	type catégorielle
1 <sup>ère</sup>	Masculin	ANA	$\pi_{m1} \rightarrow \Omega_{m1}$
	Féminin	أنا	
2 <sup>ème</sup>	Masculin	ANTA	$\pi_{m2} \rightarrow \Omega_{m2}$
	Féminin	أنتِ	
3 <sup>ème</sup>	Masculin	HOUA	$\pi_{m1} \rightarrow \Omega_{m1}$
	Féminin	هي	

Tableau 12: Pronoms libres du singulier

Personne	Genre	Pronom	type catégorielle
1 <sup>ère</sup>	Masculin	NAHNOU	$\pi_{m1} \rightarrow \Omega_{m1}$
	Féminin	نحن	
2 <sup>ème</sup>	Masculin	ANTOUMA	$\pi_{m2} \rightarrow \Omega_{m2}$
	Féminin	مائنت	
3 <sup>ème</sup>	Masculin	HOUMA	$\pi_{m2} \rightarrow \Omega_{m2}$
	Féminin	هُما	

Tableau 13: Pronoms libres du duel

### 3.7 Les formes de la tectonie nominale en arabe

Elle est basée sur ce que nous avons étudié au chapitre 2 en la théorie néo-khalilienne. La langue arabe a deux types de tectonie : tectonie verbale et tectonie nominale. Le prototype de tectonie verbale commence avec juste un verbe conjugué, par contre la forme de la tectonie nominale en arabe selon la théorie néo-Khalilienne est définie par la structure suivante :

$$\left[ \overbrace{\left( R \longrightarrow T_1 \right) \pm T_2} \right] \pm S$$

avec

**R** : est le régissant (العامل)

**T<sub>1</sub>** : 1<sup>ier</sup> terme régi. »

→ : Le premier opératoire se lie au premier terme régi (T<sub>1</sub>) à son régissant (R) de dépendance, de sorte que ce dernier ne peut pas progresser sur lui.»

**T<sub>2</sub>** : 2<sup>ième</sup> terme régi.

± : Une référence à la présence ou à l'absence du second régissant (R<sub>2</sub>) .

**S** : élément non essentiel.

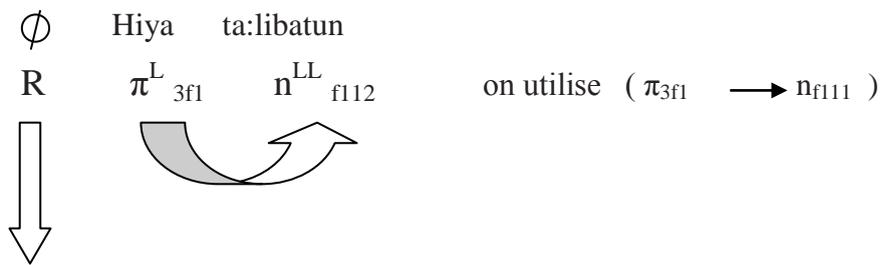
↖ : Une flèche indiquant que 2<sup>ième</sup> terme régi (T<sub>2</sub>) peut être en avance sur le régissant (R) et son 1<sup>ier</sup> terme régi (T<sub>1</sub>).

### 3.8 Expression formelle du nom en arabe notée Dgncd(N) :

Le sujet en arabe al-mubatada est toujours composé d'un nom ou d'un pronom. Il est toujours au cas nominatif مرفوع. Le mubtada est le sujet de la phrase. Il peut représenter une personne ou une chose. Il est décrit ou précisé par le khabar de la phrase. Par exemple, dans la phrase :

al jawo ra:imun « al jawo » est le mubtada – sujet et «ra:imun» est le khabar – description ou précision sur le sujet. Le mubtada peut être composé de :



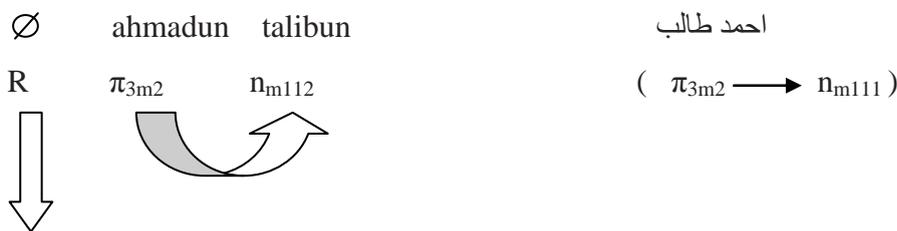


R : tectonie nominale.

Ici, la tectonie est nominale, elle commence par un pronom Hiya (ou il) qui remplace le mubtadae, suivie par un khabar.

### 3.8.3 un nom propre :

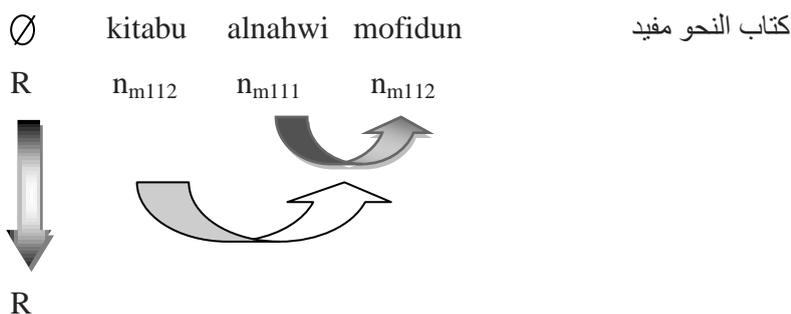
La 3<sup>ème</sup> forme de la tectonie nominale en arabe, commence par un nom propre ( احمد ) suivi d'un sujet (khabar). Prenant l'Exemple :



R : tectonie nominale.

### 3.8.4 un annexion (dans le cas d'une annexion) :

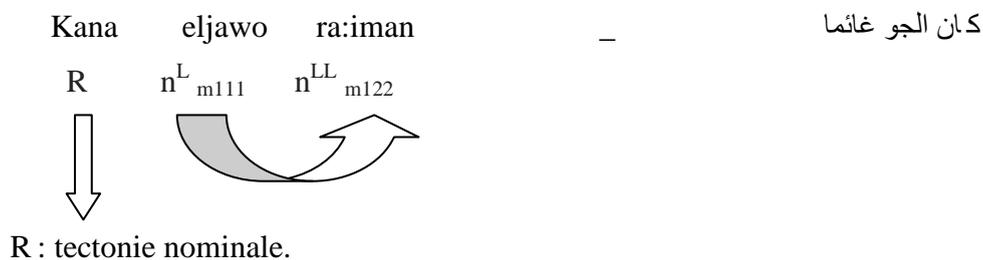
La 4<sup>ème</sup> forme de la tectonie nominale en arabe, commence par un annexant ( كتاب ) suivi d'un sujet. Prenant l'Exemple :



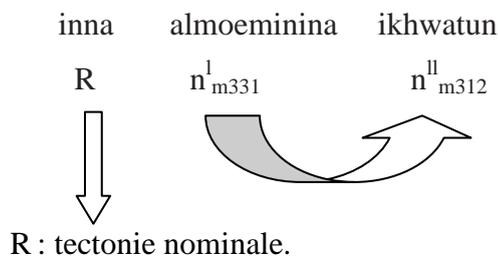
### 3.9 Approche basée sur les notions de « ' mal terme régissant » et « ma'mul terme régi ».

Syntaxiquement cette approche utilise les notions de « ' mal terme régissant » et « ma'mul terme régi ». Dans cette approche « ' mal » permettra l'accroissement de la longueur de la tectonie nominale en respectant les flexions casuelles «Eli 'rab».

#### 3.9.1 R = kana et autre homologue :



#### 3.9.2 R = inna et autre homologue :



#### Quelques exemples qui ne sont pas une tectonie nominale :

si on prend les exemples suivant :

- |                                |   |              |                     |
|--------------------------------|---|--------------|---------------------|
| kana                           | + | Aljawo       |                     |
| $(R n^I_{m111} n^{II}_{m122})$ |   | $(n_{m111})$ | → $R n^{II}_{m122}$ |
- |                                |   |              |                     |
|--------------------------------|---|--------------|---------------------|
| Al                             | + | moeminouna   |                     |
| $(R n^I_{m331} n^{II}_{m312})$ |   | $(n_{m331})$ | → $R n^{II}_{m312}$ |

### **3.10 Conclusions**

La principale contribution de ce travail est de développer la première phase de l'analyseur de grammaire pré-groupe pour Langue arabe basée sur l'analyse formelle des tectonies nominales en arabes introduites par hadj Salah.

Le travail construit des approches algébrique de la structure des tectonies nominales en arabe, contient des noms (propre , pronom , déterminé et annexion), chaque nom contient une approche algébrique unique, variant en fonction de l'expression, le genre ,le nombre et la détermination .

Pour le chapitre suivant nous présenterons les différentes étapes de la partie d'implémentation de cette approche, pour qu'on explique le langage qui utilise dans l'implémentation et le développement d'environnement.

# **Chapitre-IV**

***Expérimentations et Implémentation***

### **4.1 Introduction :**

Dans ce dernier chapitre, nous allons présenter la partie implémentation de notre projet. Nous commençons tout d'abord par la présentation de l'environnement de développement, en décrivant le langage de programmation utilisé et ensuite en détaillant les différents outils et ressources utilisés dans chaque étape. Puis nous montrons les différentes étapes de déroulement de l'application et enfin les résultats obtenus.

### **4.2 Les outils utilisés pour l'implémentation**

#### **4.2.1 L'environnement de programmation:**

Dans cette partie, nous allons d'abord définir ce qu'est un langage de programmation. Nous verrons ensuite brièvement l'histoire de ce langage, et puis nous allons présenter l'environnement de développement

#### **4.2.2 NetBeans IDE 8.2:**



**Figure 18: NetBeans**

NetBeans IDE est un environnement de développement intégré gratuit, open source (IDE) qui vous permet de développer des applications bureautiques, mobiles et Web. L'IDE prend en charge le développement d'applications dans diverses langues, y compris Java, HTML5, PHP et C++.

L'IDE offre un support intégré pour le cycle de développement complet, de la création de projet au débogage, au profilage et au déploiement. L'IDE s'exécute sous Windows, Linux, Mac OS X et d'autres systèmes basés sur UNIX.

L'IDE fournit un support complet pour les technologies JDK 7 et les améliorations Java les plus récentes. C'est le premier IDE qui prend en charge JDK 7, Java EE 7 et JavaFX 2. L'IDE prend pleinement en charge Java EE en utilisant les dernières normes pour Java, XML, Web services et SQL et prend pleinement en charge le GlassFish Server, l'implémentation de référence De Java EE.Scene Builder .

## Matériaux utilisés

- **La machine 1**

Nom de l'ordinateur : HP -PC

Processeur : Intel(R) Core(TM) i5-2520U CPU @ 2.50GHz 2.50 GHz

Mémoire installée (RAM) : 6.00 Go

Type du système : Système d'exploitation 64 bits

Windows 7 professionnel

- **La machine 2**

Nom de l'ordinateur : ACER -PC

Processeur : Intel(R) Core(TM) i3-4005U CPU @ 1.70GHz 1.70 GHz

Mémoire installée (RAM) : 4.00 Go

Type du système : Système d'exploitation 32 bits

Windows 10.

### 4.3 D'analyse de reconnaissance d'une tectonie nominale

Pour l'implémentation de notre solution, nous avons opté pour le langage NetBeans IDE et nous avons utilisé comme environnement de développement qui nous permet d'utiliser différentes API (Application Programming Interfaces) à partir de plusieurs langages de programmation comme Java et C++...etc.

Nous avons utilisé dans notre application quelques bibliothèques pour assurer certaines fonctionnalités. Citons :

#### 4.4 Déroulement

Dans cette section, nous allons expliquer les différentes étapes de déroulement de notre programme qui débute par entrer une structure de tectonie nominale arabe et permet d'extraire en sortie, la vérification que la tectonie est nominale .

La figure qui suit, illustre l'interface globale de notre système :

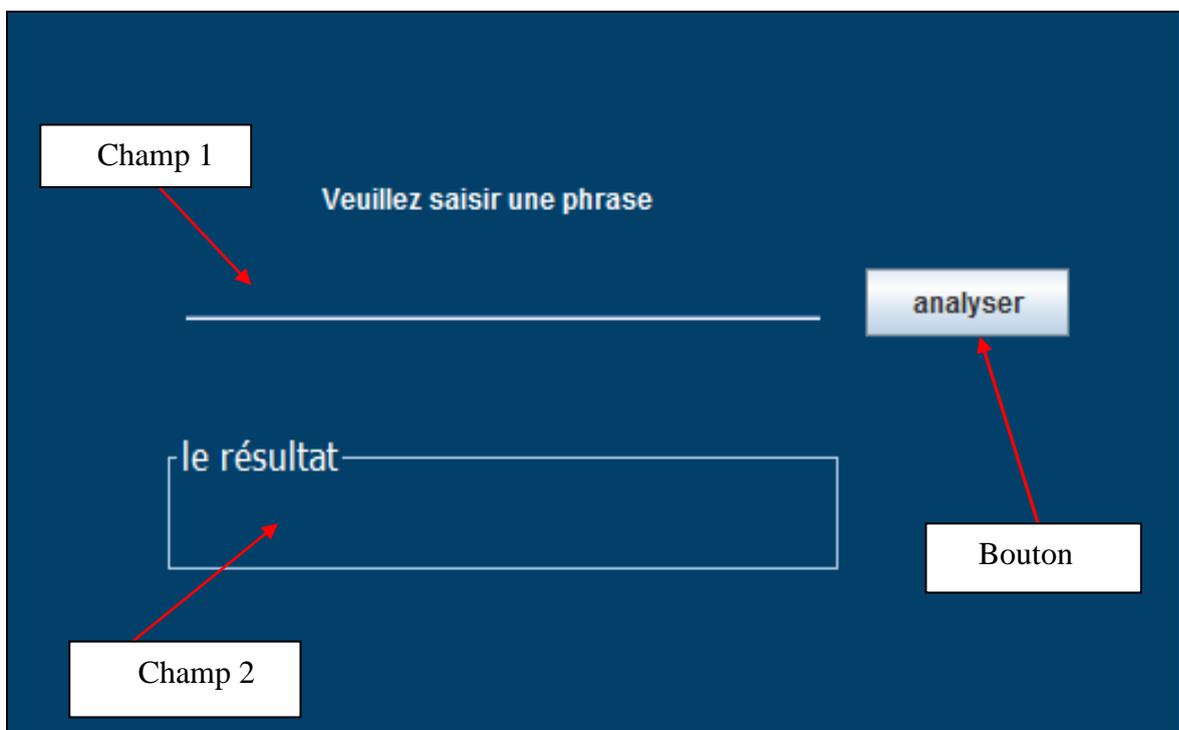


Figure 19 : interface globale

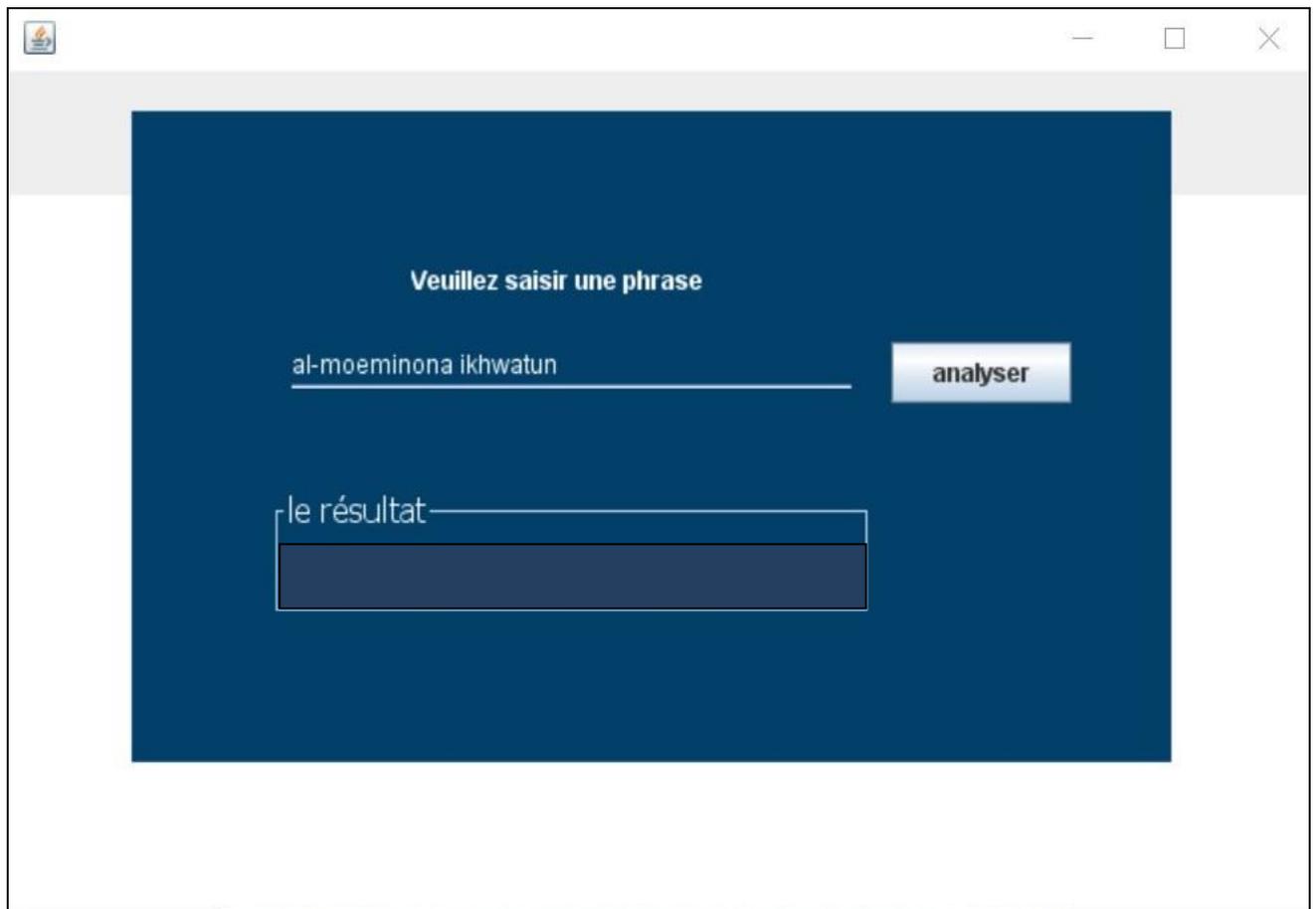
Champ 1 : pour entrer la tectonie nominale que nous allons analyser dans notre système.

Bouton : pour exécuter le processus analyse de la structure des tectonies nominales en arabe.

Champ 2 : pour afficher le résultat de l'analyse de la structure des tectonies nominales en arabe.

- **Etape 1** : entrée la tectonie nominale

Dans le champ 1 nous allons écrire la tectonie nominale qui sera analysée pour que la tectonie soit divisée en fonction de sa structure. La figure ci-dessous explique cette étape :



**Figure 20:** l'interface globale du programme.

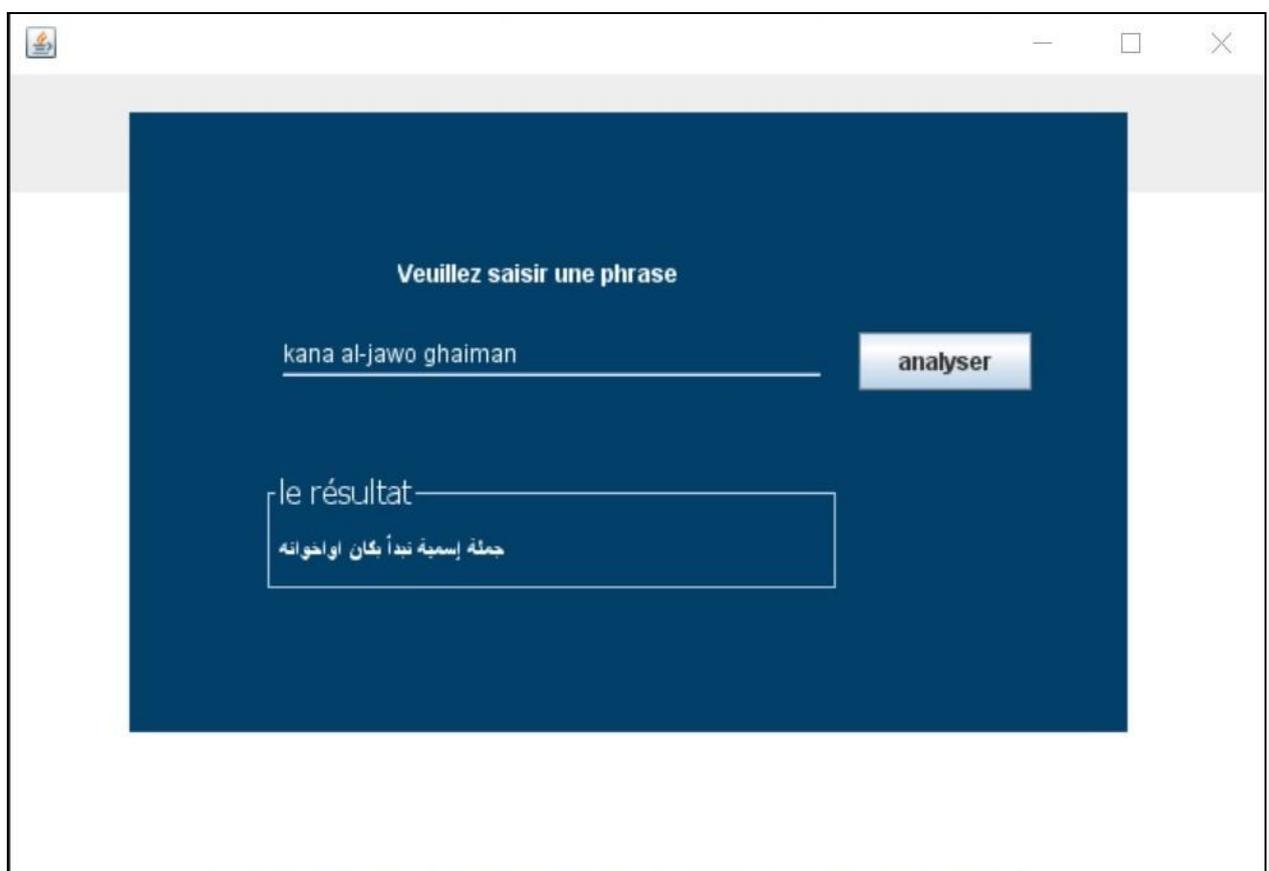
- **Etape 2** : Analyse de la structure de la tectonie nominale

Dans cette étape en appuyant sur le bouton analyse que donner le résultat final et pour afficher ce dernier il faut suivre les phases suivantes :

- **Etape 3 : résultat final**

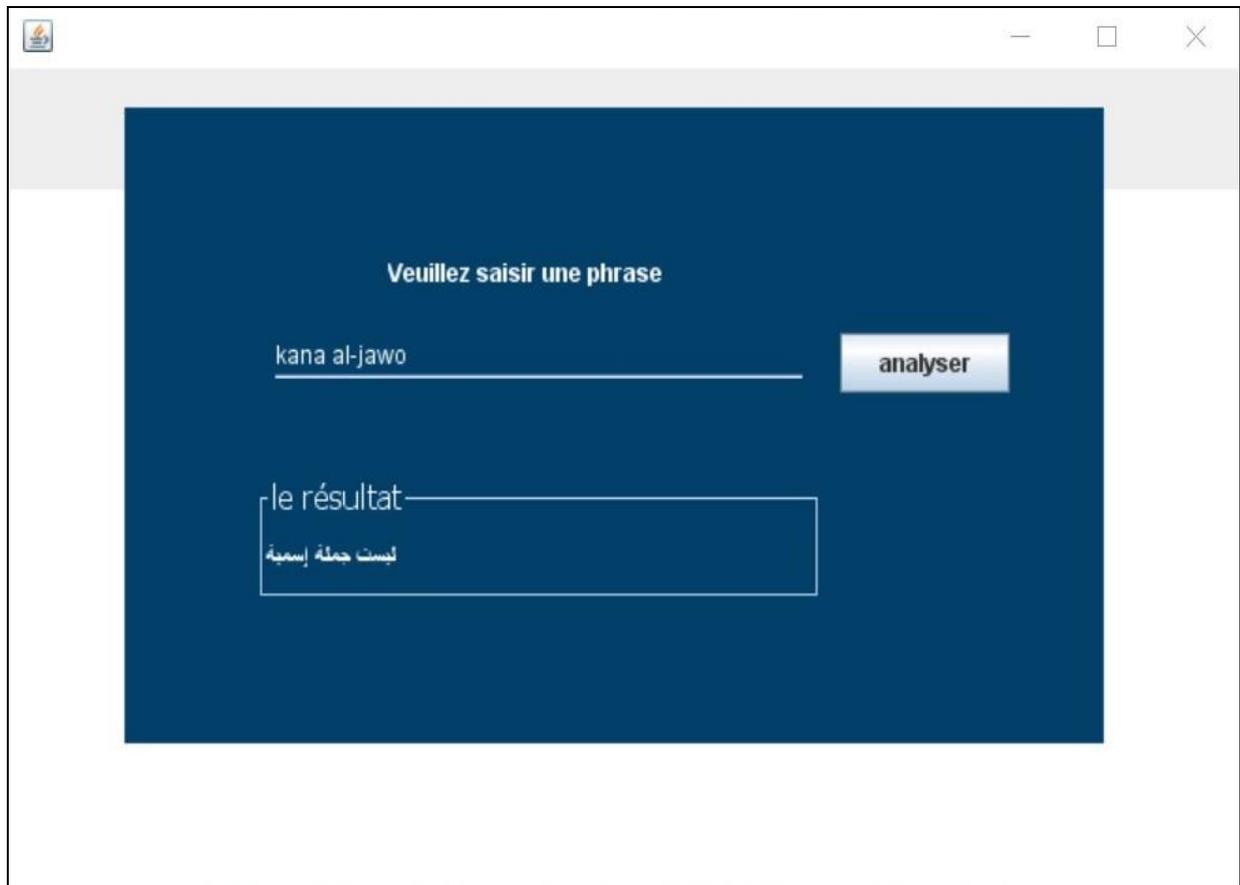
Le champ 2 est permet d'afficher l'étape finale de notre programme analyse pour donner les deux cas :

Le 1<sup>er</sup> cas où le structure de la tectonie est vrai, et son analyse dans la figure 4. qui exprime ce cas :



*Figure 21: le résultat final de cas vrai.*

Et la figure 4. Suivant présente le 2 ème cas où la structure de la tectonie est fausse



*Figure 22: le résultat final de cas faux..*

## **4.5 Conclusion**

Nous avons présenté dans ce chapitre l'environnement de développement ainsi que le langage de programmation . Le NetBeans IDE c'est le langage utilisé pour implémenter et développer notre programme. En dernier, quelques captures sont présentées pour expliquer le programme d'analyse de lexie verbele.

Nous concluons que notre système d'analyse qui donne la validation de la structure de la lexie verbale et son analyse de bons résultats.

## 4.6 Conclusion et perspective

Cet ouvrage aborde le problème de la formalisation de certains phrase dans la langue arabe.

Nous avons commencé par présenter l'environnement de développement de notre système ainsi que les différentes interfaces graphiques qui à travers lesquelles nous pouvons superviser les différents traitements du système.

Notre système consiste à concevoir, à réaliser et à évaluer un système d'analyse des phases arabe basée sur la grammaire de pré-groupe permettant de vérifier si une chaîne de mots donnée est une phrase bien formée en effectuant un calcul sur la chaîne de types correspondante

Le système réussit à vérifier toutes les phrases formelles et les considère comme une phrase bien formée (Tectonie nominale). Le système a également été testé pour des phrases grammaticales erronées, le système rejette ces phrase car elle ne fait pas partie du langage généré par un pré-grammaire définie dans le lexique.

Le travail réalisé au cours de cette thèse ouvre un certain nombre de perspectives :

- analyser plus de phrases grammaticales arabes telles que les types du noms comme le genre , le nombre .... .
- l'analyse complète du graphe de liens de réduction du mouvement clitique après le processus d'analyse.

# *Bibliographie*

- [1] Abd El Salam AL HAJJAR,"extraction et gestion de l'information a partir des documents arabes", thèse de Doctorat, UNIVERSITE PARIS VIII, SAINT DENIS ,2010.
- [2] Mohamed Zakaria Kurdi , Traitement automatique des langues et linguistique informatique 2, Volume 2 [8]
- [3] CHERAGUI Mohamed Amine Conception et Réalisation d'un lemmatiseur hybride de texte arabe Université Ahmed Draya Adrar Algérie .
- [4] Bilel Bahloul, méthode pour l'analyse automatique d'opinion de la langue arabe
- [5] MAROUF ARAIBI Mohammed el amine, ABDELMOUMENE Imene L'automatisation de la segmentation thématique de textes arabes, 2016/2017
- [6] Ed-dariouache Adnane," Etude et réalisation d'un analyseur morphologique de la langue Arabe", université Sidi Mohamed Ben Abdellah" ,24 juin 2015.
- [7] Slim MESFAR," Analyse Morpho-Syntaxique Automatique et Reconnaissance des entités nommées En Arabe Standard", thèse de Doctorat présenté en 24 Novembre 2008, Université De Franche-Comté.
- [8] Fouad Soufiane Douzidia, Résumé automatique de texte arabe, Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de M.Sc en informatique, Université de Montréal, Septembre, 2004
- [9] Aissa Bourezg, « Implémentation d'une méthode d'analyse morpho-lexicale pour la langue Arabe basée sur la position des lettres », Université Mohamed Boudiaf - M'sila,
- [10] <https://www.al-dirassa.com/lessons/la-phrase-nominale-et-la-phrase-verbale-en-arabe-lecon-8/> , al-dirassa heure :21 :00, date : 08/10/2019
- [11] Mostapha Al-Glayini, " , " العربية الدروس جامع" livre édité en 2007 en Bierut, Lebanon.
- [12] Une brève introduction à l'analyse syntaxique en grammaire générative.  
<http://linguistes.com/syntaxe/roberge/>.
- [13] Hadj Salah Abderrahmane, *Linguistique arabe et linguistique générale, essai de méthodologie et d'épistémologie du 'ilm al-'arabiya* (Thèse de doctorat, dactylographié, Université de la Sorbonne, Paris, 1979)
- [14] ABDERAHMANE HADJ-SALAH , l'inguistique et phonétique arabe(2),Centre de Recherche Scientifique et Technique Pour le Developpement De La Langue.
- [15] Fatma Zohra BELKREDIM,Formalisation de la Langue ArabeApplication à la Causalité , THÈSE Présentée pour l'obtention du grade de: DOCTEUR d'ÉTAT en INFORMATIQUE,L'Université des Sciences et de laTechnologie« HouariBoumedienne».

- [16] Ahmad M. Abd Al-Aziz, PARSING ARABIC VERB PHRASES USING PREGROUP GRAMMARS , British University in Egypt (BUE), Cairo, Egypt.
- [17] J. Lambek and A. Preller. 2003. An algebraic approach to the german noun phrase. *Linguistic Analysis*, 31:3–4.
- [18] [Buszkowski] Wojciech Buszkowski, Lambek Grammars based on pregroups, in: P. de Groote et al., editors, Logical Aspects of Computational Linguistics, LNAI 2009, Springer, 2001
- [19] Moroz, Katarzyna, (2010) “Savateev-style parsing algorithm for pregroup grammars”, In Proceedings of Formal Grammar 14, Lecture Notes in Computer Science, Vol. 5591.
- [20] PDF العامل وبناء التراكيب اللغوية العربية في ضوء النظرية الخليلية الحديثة لعبد الرحمن الحاج صالح
- [21] البنى النحوية العربية , سلسلة علوم اللسان عند العرب للدكتور عبد الرحمن الحاج صالح
- [22] PDF نظرية العامل في ضوء النظرية الخليلية الحديثة . محمد امين مولوج المركز الجامعي , تيبازة
- [23] ABDERAHMANE HADJ-SALAH , l'inguistique et phonétique arabe(2), Centre de Recherche Scientifique et Technique Pour le Developpement De La Langue.
- [24] BENAÏSSA Bedr-Eddine , Construction semi-automatique d'ontologies à partir de textes arabes, Mémoire de magister en « informatique » Option « Intelligence Artificiel et Aide à la Décision » Université Abou Bakr Belkaid.
- [25] Aïssa Bourezg, « Implémentation d'une méthode d'analyse morpho-lexicale pour la langue Arabe basée sur la position des lettres », Université Mohamed Boudiaf - M'sila,