

Ordre...../F.S.S.A/UAMOB/2018

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**  
**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE**  
**SCIENTIFIQUE**  
**UNIVERSITE AKLI MOHAND OULHADJ-BOUIRA**



Faculté des Sciences et des Sciences Appliquées  
Département d'Informatique

**Mémoire de fin d'étude**

Présenté par :

**AISSAOUI Rahma**  
**ALLOUACHE Chahira**

En vue de l'obtention du diplôme de **Master 02** en :

Filière : **INFORMATIQUE**

Option : ingénierie des systèmes d'informations et logiciels

**Thème :**

**Extraction des relations sémantique entre les mots en arabes :**  
**Reconnaissance des relations verbe-sujet-objet**

Devant le jury composé de :

Mr.	Dr.	UAMOB	Président
Mr.	Dr.	UAMOB	Examinateur
Mr. ZERROUKI Taha	Ma.	UAMOB	Encadreur

**Année Universitaire 2018/2019**

## *Remerciements :*

*Avant tout, nous remercions notre Dieu « Allah » le seul et unique le tout puissant, qui nous guide, nous protège et qui nous a aidé pour dépasser toutes les difficultés que nous avons rencontré et nous a donné la force, la volonté et la patience pour finir ce travail malgré tout.*

*A nos parents ...*

*Au terme de ce mémoire, nous voudrions remercier toutes les personnes qui ont contribué de près ou de loin à sa réalisation.*

*En premier lieu, nous aimerions remercier notre promoteur, Mr. ZERROUKI Taha, qui nous a fait l'honneur de diriger ce travail et ses précieux conseils furent d'un apport considérable. Aussi nous tenons à lui reconnaître le temps.*

*Pour finir notre liste, nous remercions énormément nos familles qui ont été avec nous tout au long de ce stage. Sans oublier nos amies qui nous ont aidé pendant cette période sans citer de noms pour n'oublier personne.*

## *Dédicace :*

*Je dédie ce modeste travail :*

*A celui qui m'a toujours appris comment réfléchir avant d'agir, à Celui qui m'a soutenu tout au long de ma vie scolaire, à Celui qui n'a jamais épargné un effort pour mon bien. Mon cher père.*

*A celle qui est toujours à côté de mon cœur, à celle qui m'a appris le vrai Sens de la vie, à celle qui n'a hésité aucun moment m'encouragé. Ma Chère mère.*

*Je dédie aussi ce travail à mes sœurs :*

*Salîha, Faiza, Nassima, Salwa, Hadjer, Fatima et Chaima.*

*Pour leur tendresse infinie ..... Je vous aime à l'infinie.*

*A toute ma famille grande et petite.*

*A mon cher binôme Chahira.*

## *Dédicace :*

*J'ai le grand honneur de dédier ce travail à :*

*A mon père Mohemad qui ma toujours soutenu moralement.*

*A l'être le plus cher de ma vie, ma mère Ouahiba Khalfoun qui a été toujours là pour moi avec sa douceur, sa patience et ses encouragements.*

*A mes frères et mes adorables sœurs Mina et Nouna.*

*A toute ma grande famille Allouache et Khalfoun.*

*A tous mes amis, tous ceux qui sèment le bonheur dans mon chemin.*

*A mes cher amis, mon binôme Rahma.*

Sommaire.....	<i>i</i>
Liste des figures.....	<i>vii</i>
Liste des tableaux .....	<i>viii</i>
Introduction générale.....	01

## **CHAPITRE I : LES RELATIONS SEMANTIQUES ENTRE LES MOTS**

I.1. Introduction .....	03
I.2. La Sémantique .....	03
I.3. Les différents niveaux d'analyse d'un texte .....	03
I.3.1. Analyse morphologique .....	04
I.3.2. Analyse syntaxique .....	04
I.3.3. Analyse sémantique .....	05
I.3.4. Analyse pragmatique .....	06
I.4. La déférence entre la Sémantique et Syntaxique .....	07
I.4.1. Liaison entre marques syntaxiques et sémantiques des mots .....	07
I.5. Les types de relations sémantiques entre les mots .....	07
I.5.1. Relations hiérarchiques et d'inclusion .....	07
a. L'hyponymie et l'hyperonymie .....	07
b. La relation partie-tout .....	08
c. Relations d'équivalence/opposition .....	08
Synonymie .....	08
Antonymie.....	08
Homonymie et Polysémie.....	09
I.6. Buts de la sémantique .....	09
I.7. Réseau sémantique .....	10
I.7.1. Définition .....	10
I.7.2. Exemple d'un réseau sémantique. ....	10
I.8. Conclusion .....	11

## **CHAPITRE II : LES ASPECTS DES RELATIONS ET LES RESSOURCES LINGUISTIQUES SIMILAIRES ET UTILES**

II.1. Introduction.....	12
II.2. L'énoncé et la sémantique .....	12
II.3. Définition de la langue arabe .....	13
II.4. Composition de la phrase arabe .....	13
II.4.1. La phrase verbale .....	13
II.4.2. La phrase nominale .....	14
II.5. L'ordre des mots dans la phrase verbale dans l'arabe .....	14
II.5.1. Ordres libres .....	14
II.5.2. Ordres restreints .....	15
II.5.3. L'ordre verbe / sujet .....	16
II.5.4. Le statut de sujets (نائب فاعل) .....	16
L'étude de l'élément (نائب فاعل) .....	16
II.5.5. L'ordre des compléments par rapport au verbe .....	17
II.6. Les différents aspects du verbe .....	17
II.6.1. L'aspect du temps .....	17
II.6.2. L'aspect fonctionnel .....	18
II.6.3. L'aspect sémantique .....	19
II.7. Difficultés de l'analyse automatique de textes arabe .....	19
II.7.1. Segmentation des textes arabes .....	19
II.7.2. L'analyse morphologique .....	20
II.7.3. L'analyse syntaxique.....	21
II.8. Les ressources linguistiques similaires.....	22
II.8.1. Wiktionnaire.....	22
II.8.1.1. Architecture de Wiktionnaire.....	22
a. Prétraitement et Extraction des définitions.....	23
b. Analyse des vocabulaires.....	24
c. Extraction des relations sémantiques.....	25

d. Création de la base lexicale .....	26
II.8.1.2 Utilité de cette approche.....	26
II. 8.2. Théorie de Structure Rhétorique.....	27
II.8.2.1 L'analyse rhétorique de textes Arabe.....	27
a. Segmentation du texte.....	27
b. Application des règles rhétoriques .....	27
c. Construction des arbres RST .....	28
II.8.3. Architecture de cette approche.....	28
II.9. Conclusion.....	29

### **CHAPITRE III : CONCEPTION**

III.1. Introduction .....	31
III.2. Description de l'architecture de l'approche .....	31
III.2.1. Le module de segmentation .....	31
III.2.2. Le module d'analyse.....	31
III.3. Description de l'algorithme de l'approche .....	32
III.3.1. Description détaillée de l'algorithme.....	33
III.3.1.1. Segmentation de texte .....	33
III.3.1.1.1. Règles de segmentation .....	33
A. Segmenter le texte en unités textuelles (phrases .....	33
Le point .....	33
Le point d'interrogation .....	33
Le point d'exclamation .....	33
B. Segmentation des phrases en sous-phrases .....	34
a. Le point-virgule .....	34
b. La virgule .....	34
C. Segmentation des phrases en unités minimales .....	37
III.3.1.1.2. Les tagger .....	37
Liste du tag .....	37

III.3.1.2. Analyse de l'approche .....	38
III.3.1.2.1. Application des règles de grammaires .....	38
Les verbes incomplètes « الأفعال الناقصة » .....	38
Détection du relation sujet et verbe .....	39
Le sujet est un nom ou un pronom .....	39
Le sujet « أن اسمها وخبرها » .....	40
Le sujet « فعل » + « ما » .....	41
Le sujet « التاء المتصلة بالفعل » .....	41
Le sujet est « نون النسوة », « يا الفاعلية », « واو الجماعة », « ألف الاثنين », « نا الفاعلية » .....	04
Détection de la relation objet et verbe .....	43
Les règles détection l'objet .....	44
L'objet est « ضمائر النصب المنفصلة » .....	44
المفعول به ضمير متصل بالفعل .....	45
III.3.1.2.2. Utilisation d'une base de données .....	46
A. Construction du réseau sémantique .....	47
III.4. Conclusion .....	49

## CHAPITRE IV : IMPLEMENTATION

IV.1. Introduction .....	50
IV.2. Environnement de travaille .....	50
IV.2.1. Environnement de matériel .....	50
IV.2.2. Environnement logiciel .....	50
PyScripter .....	50
Langages utilises .....	50
IV.3. Présentation graphique .....	51
Relation Sémantique verbe-sujet-objet .....	51
Exemples sur la fonction de boutons «voir les relations sémantiques» .....	52
Exemples sur la fonction de boutons «voir le réseaux sémantiques».....	52
IV.4. Déroulement de l'approche .....	53



IV.4.1. Les résultats .....	53
IV.4.1.1. Les erreurs .....	54
IV.4.2. Evaluation et test .....	56
IV.4.2.1. Précision .....	56
IV.4.2.2. Rappel .....	56
IV.4.2.3. F-mesure .....	57
IV.5. Conclusion .....	57

<b>Figure II.1.</b> Les informations contenues dans le fichier XML pour l'entrée حَاسُوب dans le Wiktionnaire arabe	24
<b>Figure II.2.</b> Exemple d'une base lexicale	26
<b>Figure II.3.</b> Architecture de RST	29
<b>Figure IV.1.</b> Capture d'écran de l'interface graphique «relation sémantique verbe-sujet-objet»	51
<b>Figure IV.2.</b> Résultat du bouton «voir les relations sémantiques»	52
<b>Figure IV.3.</b> Fonction de boutons «voir le réseaux sémantiques»	52
<b>Figure IV.4.</b> Quelques exemples des phrases utilisés	53
<b>Schéma I.1.</b> Les différents niveaux d'analyse d'un texte	03
<b>Schéma I.2.</b> Un exemple de représentation arborescente de la phrase « حفظ الولد القرآن »	05
<b>Schéma I.3.</b> Exemple de réseau sémantique, emprunté à H. Gallaire, 1985	10
<b>Schéma II.1.</b> Architecture de l'approche	23
<b>Schéma III.1.</b> L'architecture globale de notre système	31
<b>Schéma III.2.</b> Exemple de segmentation de texte	33
<b>Schéma III.3.</b> Exemple de segmentation des phrases en sous-phrases	35
<b>Schéma III.5.</b> Déterminer le type de verbe	38
<b>Schéma III.6.</b> Le sujet est un nom ou un pronom	39
<b>Schéma III.7.</b> Le sujet « أن اسمها وخبرها »	39
<b>Schéma III.8.</b> Le sujet « ما وفعلها »	40
<b>Schéma III.9.</b> Le sujet « التاء المتصلة بالفعل »	41
<b>Schéma III.10.</b> Le sujet est «terminaison de verbe»	42
<b>Schéma III.11.</b> Déterminer le type de verbe	43
<b>Schéma III.12.</b> L'objet est «ضمائر النصب المنفصلة»	44
<b>Schéma III.13.</b> L'objet est « ضمير متصل بالفعل »	45
<b>Schéma III.14.</b> Extrait relation sémantique selon BDD utilisée	46

<b>Schéma III.15.</b> Détermination le type de relation	<b>47</b>
<b>Schéma III.16.</b> Détermination le type de relation	<b>47</b>
<b>Schéma IV.1.</b> Pour rechercher type de relation selon une BDD.	<b>55</b>



<b>Tableau I.1.</b> Nouvelle proposition construite à partir de la 1 <sup>ère</sup> proposition.	<b>06</b>
<b>Tableau II.1.</b> Définitions obtenus à partir de l'entrée dans la figure 2	<b>24</b>
<b>Tableau II.2.</b> Exemple de résultat de l'analyseur morphosyntaxique de définition du mot الحاسوب.	<b>25</b>
<b>Tableau II.3.</b> Exemple de représentation des relations sémantiques.	<b>26</b>
<b>Tableau III.1 :</b> un ensemble de POSTag.	<b>37</b>
<b>Tableau IV.1.</b> Quelques exemples de résultat des phrases utilisés.	<b>53</b>
<b>Tableau IV.2.</b> Les erreurs de programme.	<b>53</b>
<b>Tableau IV.3.</b> Pourcentage de résultat de « relation verbe_sujet » et « relation verbe_objet ».	<b>54</b>
<b>Tableau IV.4.</b> Des résultats d'évaluations.	<b>56</b>

# ***INTRODUCTION GENERALE***

## Introduction générale :

Le traitement automatique des langues naturelles (TALN) est un domaine à la frontière de la linguistique et l'informatique et de l'intelligence artificielle ; elle concerne la conception et le développement des programmes et techniques informatiques capables de traiter de façon automatique des données exprimées dans une langue. Les applications liées au TALN ont fait l'objet d'une attention toute particulière depuis plusieurs décennies.

Certaines langues ont été privilégiées comme le français et l'anglais, où les recherches sont concentrées sur ces dernières pour proposer des outils pour des applications TALN. D'autres, telle que la langue Arabe poursuivent les recherches et les travaux afin de proposer des outils robustes de traitement permettant de proposer des applications TALN pour ces langues.

La langue Arabe, à laquelle nous nous intéressons est classée comme étant la cinquième langue au monde avec plus de 50 000 sites arabes sur le web et plus de 320 millions locuteurs, manque d'outils capable de traitement automatique de cette langue. Par ailleurs, nous assistons à un accroissement des contenus textuels en arabe, surtout en ligne.

Notre travail a pour objectif principal de implémenter un système d'extraction des relations verbes-sujets-objets à partir d'un corpus des textes arabes afin de créer un dictionnaire de relations sémantiques qui permet la réutilisation de telle ressources.

Notre mémoire est structuré come suit :

- Le premier chapitre : « **les relations sémantiques entre les mots** » porte sur le contexte général du projet, introduit généralité sur les relations sémantiques entre les mots.
- Le deuxième chapitre : « **les aspects des relations et les ressources linguistiques similaires et utiles** » nous étudions les aspects relations verbes-sujets-objets dans la langue arabe. Ensuite, Nous tâchons de recenser les ressources linguistiques similaires et utiles.
- Le troisième chapitre : « **conception** » nous traitons la conception du projet : C'est une phase de réflexion et d'étude sur le choix des meilleurs moyens ainsi que sur la méthodologie à suivre et les concepts à utiliser.
- Le quatrième chapitre : « **implémentation** » nous présentons notre application dédiée. Cette partie aborde la phase finale du projet. A ce niveau nous avons

# Introduction générale

---

essayé d'appliquer toutes les démarches et les concepts étudiés pendant la troisième partie. Cette dernière partie est réalisée en suivant le cycle de vie d'un projet depuis son démarrage jusqu' à son achèvement.

Enfin, nous terminons ce mémoire par une conclusion.



***CHAPITRE I :***  
***LES RELATIONS SEMANTIQUES***  
***ENTRE LES MOTS***

## I.1. Introduction:

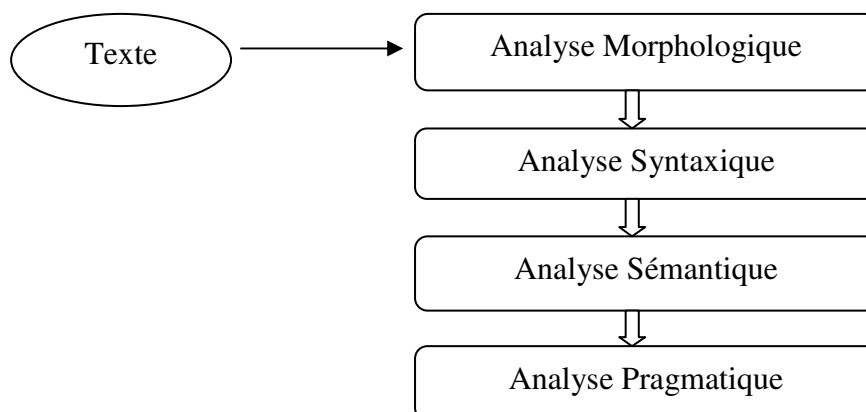
La sémantique occupe une position importante dans le traitement de la langue naturelle. Il n'est pas facile de réaliser des traitements profonds des textes sans informations suffisantes sur la sémantique des termes et les relations sémantiques entre les mots constitutifs des textes, pour cela les chercheurs du traitement automatique de la langue arabe n'ont pas écarté les questions de Synonymie et d'Antonymie et d'autres relations sémantiques, en raison de leurs valeurs ajoutées pour résoudre plusieurs problèmes comme l'analyse automatique des textes, la compréhension des textes, la traduction automatique [1].

## I.2. La Sémantique:

Le mot sémantique est dérivé du grec et veut dire « signifier, indiquer ». Il a été repris à la fin du XIX<sup>ème</sup> siècle par le linguiste français Michel Bréal, auteur du premier traité de sémantique. Le sens véhiculé par le langage est à repérer soit au niveau du mot; soit au niveau de la phrase. La sémantique lexicale étudie le sens indépendamment du contexte d'utilisation pendant que la sémantique pragmatique étudie le sens contextualité (phrase prononcée par un locuteur particulier dans un lieu et à un moment précis) [2].

## I.3. Les différents niveaux d'analyse d'un texte :

Le processus du traitement automatique des données linguistiques nécessite différents niveaux d'analyses. On parle dans la littérature d'analyse morphologique, d'analyse syntaxique, d'analyse sémantique et d'analyse pragmatique. Ces niveaux peuvent être couplés de différentes méthodes : séquentielle, parallèle, distribuée, etc. le schéma 1.1 illustre les différents niveaux d'analyse d'un texte [3].



**Schéma I.1.** Les différents niveaux d'analyse d'un texte.

# Chapitre I les relations sémantiques entre les mots

---

Dans ce qui suit nous allons décrire brièvement les différents niveaux d'analyse d'un texte en langue naturelle.

## **I.3.1 Analyse morphologique :**

L'analyse morphologique est indispensable pour tout système de traitement automatique de la langue naturelle, cette analyse permet de regrouper les mots en classes utilisables par les autres niveaux d'analyse. La définition de ces classes varie en fonction des traitements envisagés. A chaque classe on associe une étiquette appelée catégorie grammaticale ou catégorie lexicale. L'analyse morphologique permet de reconnaître une chaîne de caractère comme étant un mot de la langue. A chaque chaîne est associé :

- une classe lexicale (grammaticale) décrivant la fonction syntaxique du mot. Exemple : verbe, substantif, adjectif, etc....
- certaines variables. Ce sont des compléments de la description de la fonction syntaxique du mot. Exemple : le genre, le nombre pouvant prendre les valeurs singulier, pluriel, etc.... [3].

L'analyse morphologique suppose de construire un ensemble de règles que nous appelons une grammaire et un dictionnaire. La grammaire contient des règles qui contrôlent la composition des formes à partir des éléments contenus dans le dictionnaire [3].

## **I.3.2 Analyse syntaxique :**

Un langage formel est défini par sa grammaire, alors que la langue naturelle ne l'est pas. En effet une langue n'est pas définie par sa syntaxe, car cette dernière est écrite postérieurement et ne présente qu'une approximation, d'où on parle de modèle syntaxique. C'est cette approximation qui fait que l'analyse syntaxique n'est pas précise et pose des difficultés [3].

Plusieurs méthodes d'analyse syntaxique se sont développées, mais la plus célèbre est sans doute la notion de grammaire formelle. Une grammaire formelle est présentée sous la forme d'un ensemble de règles de dérivation, ces règles exprimant la structure des entités syntaxiques telles que la phrase (PS), le groupe nominal (GN), le groupe verbal (GV) etc [3].

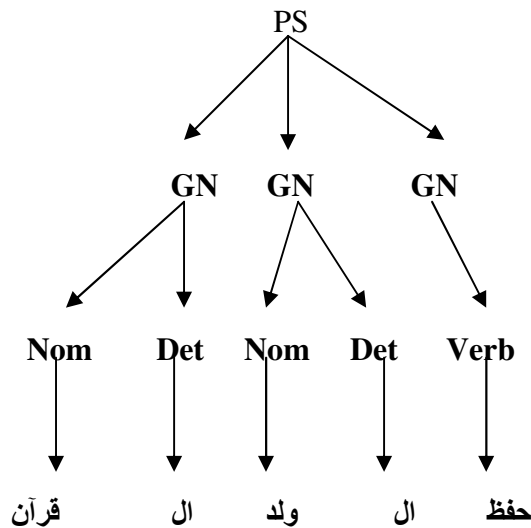
# Chapitre I les relations sémantiques entre les mots

Pour exprimer par exemple qu'une phrase est composée d'un groupe nominal et d'un groupe verbal, on utilise la règle  $PS \rightarrow GN + GV$ . Et qu'un groupe nominal est composé d'un déterminant et d'un nom, on utilise la règle  $GN \rightarrow Det + Nom$ . A l'aide de cet ensemble de règles, il est donc possible d'analyser un certain nombre de phrases [3].

Les constituants regroupant plusieurs mots (syntagme) possèdent donc leurs propres catégories (syntagme nominal, syntagme verbal, syntagme prépositionnel, etc.).

La structure de ces constituants peut être représentée sous la forme d'un arbre (appelé arbre syntaxique) ou par d'autres systèmes de représentation comme par exemple le système de parenthèses. Toutefois la représentation la plus utilisée reste la représentation sous la forme d'arbre ou représentation arborescente. Dans cette représentation les branches successives décrivent la décomposition en constituants, chaque nœud a une étiquette qui correspond à un syntagme ou à une catégorie et à chaque nœud terminal est associé un mot (ou item lexical). [3].

Le schéma 1.2 montre un exemple de représentation arborescente de la phrase « حفظ الولد القرآن ».



**Schéma I.2.** Un exemple de représentation arborescente de la phrase « حفظ الولد القرآن ».

### I.3.3 Analyse sémantique :

L'analyse sémantique des énoncés s'appuie sur une analyse syntaxique préalable. Elle cherche à construire une représentation formelle permettant des raisonnements et donc d'inférer de nouvelles informations à partir des informations présentes dans l'énoncé [3].

## Chapitre I les relations sémantiques entre les mots

Parmi les représentations, on trouve la logique des propositions. A l'aide de connecteurs logiques (comme la conjonction «  $\wedge$  », la disjonction «  $\vee$  », la négation «  $\neg$  » etc.), on peut former à partir des propositions de nouvelles propositions complexes [3].

Exemple :

Proposition	Négation
أكل آدم التفاحة	لم يأكل آدم التفاحة

**Tableau I.1.** Nouvelle proposition construite à partir de la 1<sup>ère</sup> proposition.

La logique des propositions ne s'intéresse pas au contenu des propositions mais seulement à leurs valeurs de vérité. Comme de nombreux phénomènes ne peuvent pas être représentés en logique des prédicats [3].

Une autre forme de représentation appelée « *réseaux sémantiques* » a été proposée, son principe consiste à représenter la connaissance sous la forme d'un graphe (ou réseau) de concepts. Les nœuds représentent les concepts et les arcs les relations entre ces concepts [3].

Plusieurs types de relations entre concepts existent comme : EST-UN, SORTE-DE, EST-PARTIE-DE, etc [3].

L'utilisation de ces concepts passe par des outils de navigation dans le graphe, afin de comprendre le sens de la phrase et les relations entre les différents mots qui la constituent [3].

Les réseaux sémantiques ont été étendus pour améliorer la représentation et l'inférence des connaissances [3].

### I.3.4 Analyse pragmatique :

« La pragmatique concerne l'étude de l'environnement d'une phrase, au moment où elle est émise ; elle découle de l'idée qu'une phrase (un énoncé) ne peut prendre tout son sens que si on la (le) replace dans son milieu d'origine ; c'est la prise en compte de toutes les conditions de production d'une phrase, tant il est vrai qu'un acte linguistique effectif ne peut avoir lieu qu'à l'intérieur d'une certaine situation de communication. » [3].

Ce niveau d'analyse recouvre tout ce qui est lié à l'implicite dans la communication. C'est donc le niveau qui pose le plus de problème à concevoir et par conséquent il est beaucoup plus complexe à établir, ce qui explique qu'il n'existe que peu de réalisations opérationnelles, et qui ne concerne que des applications limitées. On est donc encore loin de savoir construire des analyseurs pragmatiques pour le TALN [3].

## I.4. La déference entre la Sémantique et Syntaxique :

### I.4.1. Liaison entre marques syntaxiques et sémantiques des mots :

La signification n'est pas donnée avec évidence dans un texte. C'est pourquoi, si nous voulons trouver une procédure objective qui nous permette d'établir et de classer les significations, nous devons nous appuyer sur les propriétés d'un texte (ou d'un ensemble de phrases correctes ne constituant pas un texte, si on s'en sert comme d'un matériau brut), qui nous sont livrées par la simple observation et qui reflètent d'une manière suffisamment complète les caractères sémantiques qui nous intéressent mais qui sont dissimulées à l'observation directe [4].

D'une manière générale tout dans un texte, y compris sa structure phonologique, est conditionné et organisé par la signification; c'est pourquoi celle-ci se reflète sous une forme de nombreuses fois médiatrice, même sur le seul plan phonologique. Mais décrire la signification d'après les marques phonologiques d'un texte serait des plus inadéquats dans la mesure même où la structure sémantique de ce texte est séparée de sa structure phonologique par un très grand nombre d'étages intermédiaires [4].

## I.5. Les types de relations sémantiques entre les mots:

Les relations sémantiques entre les unités lexicales structurent le lexique sur le plan paradigmatique. Elles sont de deux types :

- Relations hiérarchiques et d'inclusion lorsqu'elles concernent des unités qui n'ont pas le même rang (hyponymes et hyperonymes, relation partie-tout) ;
- Relations d'équivalence et d'opposition lorsqu'elles concernent des unités de même rang (synonymes, antonymes, co-hyponymes) [5].

### I.5.1. Relations hiérarchiques et d'inclusion :

#### a. L'hyponymie et l'hyperonymie:

C'est une relation qui correspond très exactement, dans un autre vocabulaire, à la relation entre genre et espèce. L'hyperonyme est le terme qui désigne le genre, par exemple الزهرة, et l'hyponyme celui qui désigne l'espèce : توليب, الزنبق, etc. sont ainsi des cohyponymes de الزهرة. Si un terme hyponyme peut dans tout contexte remplacer n'importe lequel de ses

## Chapitre I les relations sémantiques entre les mots

---

hyponyme, l'inverse n'est pas vrai : j'ai cueilli une rose implique bien que j'ai cueilli une fleur, tandis que j'ai cueilli une fleur peut renvoyer à toutes les espèces [6].

La relation d'hyponymie est une relation fondamentale dans les taxinomies qui, dans le meilleur des cas, structurent les champs sémantiques. Elle est à la base de la définition l'lexicographique dite par genre prochain et différence spécifique : Carotte : plante (genre prochain) cultivée pour ça racine comestible (différence spécifique) [6].

### b. La relation partie-tout :

Une relation qui n'est généralement pas prise en compte, bien qu'elle ait une importance certaine aussi bien dans le domaine de la syntaxe que dans le domaine de la lexicographie. Elle a en effet une incidence sur la syntaxe, entre autres sur la syntaxe de l'anaphore. Considérons par exemple : شرّاع/فارب, ظفر/أصابع [6].

Relation d'hyponymie et relation de partie à tout ont donc dans ce cadre un fonctionnement analogue [6].

### c. Relations d'équivalence/opposition:

#### ✚ Synonymie :

La synonymie s'établit entre des termes équivalant, substituables dans un même contexte : فناء/ساحة.

On distingue généralement les synonymes absolus, qui sont théoriquement substituables dans n'importe quel contexte, et les synonymes approchants ou partiels qui ne peuvent l'être que dans certains, comme aborder et négocier qui ne sont interchangeable que construits avec virage [6].

En réalité, les synonymes absolus ne se rencontrent pas en dehors des nomenclatures scientifiques (ictère, cholémie, hépatite) et la majorité des synonymes, s'ils présentent une signification générale commune, sont séparés par des «idées accessoires » et des nuances qui leur confèrent des emplois spécifiques [6].

#### ✚ Antonymie :

S'établit, elle, entre deux termes de sens contraire : heureux et malheureux. Otto Duchàcek classe les antonymes sur le modèle des synonymes en antonymes absolus(il s'agit

## Chapitre I les relations sémantiques entre les mots

---

de mots qui n'ont qu'un sens et qui s'excluent) comme présent et absent (celui qui n'est pas présent es absent et réciproquement) : antonymes partiels lorsque «les unités lexicales respectives ne s'opposent à une autre unité lexicale que par une partie de leurs contenus sémantiques» comme libertin qui s'oppose tantôt à chaste, religieux ou croyant ; antonymes approximatifs qui se définissent l'un par rapport à l'autre (ابن و بيع, شراء و ابن) ; les antonymes impropres comme ليل و يوم, qui sont souvent mis en opposition, sans être réellement contraires [6].

### Homonymie et Polysémie :

Les relations que nous avons jusqu'ici examinées s'établissent entre des termes du lexique dont aussi bien le signifiant que le signifié diffèrent. Avec l'homonymie, il s'agit cette fois de relations entre des termes dont le signifiant est identique [6].

Deux termes sont en effet homonymes s'ils sont formellement identiques, phoniquement (ce sont des homophones) : thermes et terme, et éventuellement graphiquement (ce sont des homographes) : terme (mot) et terme (fin), mais présentent des sens différents. L'homonymie fait donc intervenir des relations sémantiques. En effet, si sur les exemples donnés, il est clair que les mots présentent des sens qui ne sont pas reliés, il est parfois difficile de déterminer si deux sens sont ou non apparentés. L'homonymie ne peut ainsi se traiter indépendamment du problème de la polysémie [6].

### I.6. Buts de la sémantique :

La sémantique est l'étude scientifique de la signification. En tant que telle, elle exige des techniques et des outils particuliers, et des méthodes scientifiques. Nous verrons cependant que, contrairement à la phonologie et à la syntaxe, où il existe un accord assez général sur les méthodes à utiliser, la sémantique se caractérise par une grande diversité d'approches.

La complexité de la sémantique s'explique aussi par le fait qu'elle met en jeu plusieurs niveaux de données, allant du sens des mots, à celui des phrases, aux relations sémantiques entre phrases dans le discours, et aux relations pragmatiques qui mettent en jeu l'utilisation du langage dans les situations diverses [7].



## I.7. Réseau sémantique :

### I.7.1. Définition :

Les réseaux sémantiques constituent une classe de structures de données représentables facilement dans un ordinateur. Ils sont utilisés pour cette raison en Intelligence Artificielle comme des modes de représentation de connaissances [8].

### I.7.2. Exemple d'un réseau sémantique :

En Intelligence Artificielle, la notion de réseau a été reprise pour la représentation des connaissances et son stockage en mémoire. Il existe plusieurs formes de réseaux (par exemple de réseau sémantique, emprunté à H. Gallaire, 1985 (voir le schéma 3)). Certains modèles de réseaux peuvent même constituer des modèles de machines « hautement parallèles », les nœuds étant des automates communiquant avec les voisins auxquels ils sont reliés dans le réseau [8].

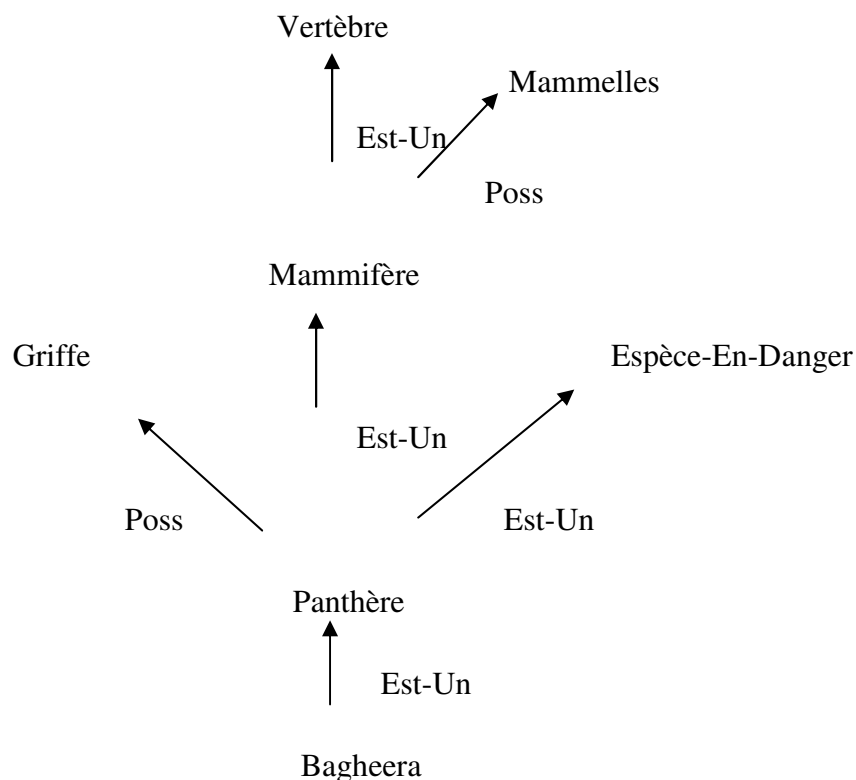


Schéma I.3. Exemple de réseau sémantique, emprunté à H. Gallaire, 1985.

### **I.8. Conclusion :**

Enfin, nous avons défini dans ce chapitre la relation sémantique et nous avons détaillé les différentes types de cette relation entre les mots et les niveaux d'analyse d'un texte.

Par les outils de l'analyse du discours ont permis que, de nos jours, la sémantique, qui est l'élément constitutif de la langue, soit au centre des sciences du langage.

***CHAPITRE II :***  
***LES ASPECTS DES RELATIONS***  
***ET LES RESSOURCES***  
***LINGUISTIQUES SIMILAIRES***  
***ET UTILES***

## **II.1. Introduction :**

On considère la sémantique comme l'étude scientifique des langues naturelles, en éclairant les conditions de l'intercompréhension linguistique et l'énoncé n'est pas comme la phrase grammaticale, il est donc naturel de commencer par présenter globalement la relation entre l'énoncé et la sémantique et langue arabe qui liés au contexte de notre étude. Ainsi, ce chapitre est organisé de la manière suivante. Dans la section 4 nous donnons 4 Composition de la phrase arabe. La section 5 est dédiée l'ordre des mots dans la phrase verbale dans la langue arabe et les différents aspects du verbe. Enfin, nous terminons ce chapitre par les difficultés de l'analyse automatique de textes arabe et quelque ressource similaire de notre ressource que nous allons traduire dans ce chapitre, et ces utilités et nous conclure bravement avec notre méthode de travaille.

## **II.2. L'énoncé et la sémantique :**

La sémantique est l'étude scientifique du sens, ou de la signification. Cette science nous invite à découvrir un monde dans lequel nous pratiquons notre langue quotidiennement ou nous exprimons et transmettons la compréhension de manière complète mais par fois incomplète car nous arrivons pas de donner de sens correcte à un énoncé et nous ne pouvons pas l'expliquer comme nous le désirons, bien que l'énonciation et le produit de cet acte, l'énoncé, a une grande importance pour la sémantique [9].

L'énoncé n'est pas comme la phrase grammaticale, qui est correcte, complète et parfaite et porte dans tous les contextes le même sens. La portée de l'énoncé va de simple exclamation au paragraphe entier. Il peut se manifester sous des formes incorrectes et imparfaites et porter de différentes significations dans des contextes différents, ou on peut encore nous laisser lui donner de diverses significations [9].

Les champs d'investigations de la sémantique peuvent êtres groupés en deux rubriques: l'une est l'étude de la langue naturelle et le produit linguistique que la langue nous représente, ou en d'autres termes, les spécificités des relations entre la parole et le discours; l'autre est l'étude de la signification de la langue en tant que forme et son influence en tant que contenu sur la forme de la langue [9].

### II.3. Définition la langue arabe :

La langue arabe véhiculaire est divisée en arabe classique (AC) et arabe standard moderne (ASM), la première est la langue des textes saints de l'islam : le coran et le hadith et du patrimoine culturel, littéraire et scientifique de la civilisation arabo-musulmane [10].

Cependant l'ASM est la langue officielle du monde arabe actuellement, elle est utilisée dans l'enseignement et dans les médias. La différence entre l'AC et l'ASM consiste dans le lexique l'ASM utilise un lexique plus vaste et plus moderne que celui de l'AC côté de la grammaire l'ASM a abandonné l'utilisation de quelques formes compliquées de l'ancienne grammaire [10].

### II.4. Composition de la phrase arabe :

Les grammairiens arabes ont distingué deux types de phrases: la phrase verbale et la phrase nominale.

#### II.4.1. La phrase verbale :

Les grammairiens arabes considèrent qu'une phrase verbale est celle qui est constituée d'au moins deux éléments ; fā'il 'le sujet' et le alfi'l' 'le verbe'. Elle exprime une action qui a eu lieu à un certain temps, produit par un certain sujet et conduit, s'il y a lieu, vers un certain objet. Le pivot de la phrase verbale est constitué par le verbe et l'agent [11].

La phrase verbale est composée des constituants suivants :

- Un complément d'objet direct (cod) :

La phrase sera constituée par un verbe, un sujet et un complément d'objet direct. Il est possible de trouver l'objet direct précédé par le sujet, lorsque le complément est un mot plus bref que le sujet. On peut changer l'ordre de ces mots et le sens de la phrase reste inchangé [11].

**Exemple :**

V + s + cod

*kataba alwaladu addarsa* (كتب الولد الدرس)

Ou v + cod + s

*kataba addarsa alwaladu* (كتب الدرس الولد)

- Un complément d'objet indirect : la phrase est constituée par un verbe transitif, un sujet et par deux compléments d'objet dont l'un est direct et l'autre est indirect. Ce dernier est introduit par une préposition et a un lien étroit avec le verbe [11].

**Exemple:**

V + s + cod + cos

*Arsala alwaladu risâlatan ilā ṣâḥibihi* أرسل الولد الرسالة إلى صاحبه

**II.4.2. La phrase nominale :**

Les grammairiens arabes considèrent que la phrase nominale est constituée de deux éléments principaux, le thème *almutad'a* et le propos *habar*, sans qu'ils soient liés l'un à l'autre par un verbe [11].

La phrase nominale a la fonction d'exprimer une qualité, une attitude ou un état appartenant à quelqu'un ou à quelque chose. Donc elle ne peut pas exprimer le temps car elle est dépourvue de verbe [11].

**II.5. L'ordre des mots dans la phrase verbale dans l'arabe :**

Ils se divisent en deux courants. Le premier considère que l'ordre des mots de la phrase arabe est libre : en effet, plusieurs ordres linéaires des composants de la phrase sont possibles, c'est-à-dire que le verbe, le sujet et le complément peuvent se placer en différentes positions. Le deuxième courant, quant à lui, considère que l'ordre des mots est strict. Nous allons détailler le point de vue de chaque courant.

**II.5.1. Ordres libres :**

Voici quelques structures qui sont considérées comme possibles :

V + s + cod

*kataba alwaladu ddarsa* (كتب الولد الدرس)

V + cod + s

*kataba addarsa alwaladu* (كتب الدرس الولد)

S + v + cod

*alwaladu kataba addarsa* الولد كتب الدرس

S + cod + v

*Alwaladu addarsa kataba* الولد الدرّس كتب

Cod + v + s

*addarsa kataba alwaladu* ( الدرّس كتب الولد )

Le sujet الولد est au cas nominatif et marqué par la flexion /u/. Il est ainsi le sujet du verbe كتب. Le mot الدرّس est au cas accusatif et marqué par la flexion /a/ et il est ainsi considéré comme le complément d'objet direct. Les constituants de ces exemples ont les marques flexionnelles qui déterminent leur fonction grammaticale, ce qui les rend acceptables.

### II.5.2. Ordres restreints :

Le deuxième courant estime que certaines interversions des constituants principaux de la phrase ne sont pas toujours possibles. La modification en genre ou en nombre du sujet peut rendre certaines permutations ambiguës voire agrammaticales [11].

En effet, l'ensemble des permutations dans la phrase (كتب الولد الدرّس) n'a pas généré de phrases agrammaticales et ambiguës, étant donné qu'elles comportent les marques flexionnelles de leur fonction grammaticale. Lorsque le sujet est un nom commun, au duel ou au pluriel, nous constaterons que d'autres constructions deviendront impossibles. Si un sujet est fléchi au duel ou au pluriel et précède un verbe, il doit alors changer de forme pour que la phrase devienne grammaticale [11].

#### Exemple :

V + s + cod

*Kataba alawalādu addarsa* كتب الأولاد الدرّس

S + v + cod

*Alawlādu kataba addarsa* ( الأولاد كتب الدرّس )

La non-pertinence de ces exemples vient du fait que la forme verbale ne porte pas la marque adéquate à son sujet. Il est ainsi nécessaire de procéder à un accord en genre et en nombre pour rendre les phrases grammaticales [11].

Voici le résultat de l'application de l'accord en genre et en nombre :

s + v + cod

*alwaladān katabā addarsa* الولدان كتب الدرّس

s + v + cod

*الأولاد كتبوا الدرس Alawlādu katabū addarsa*

### II.5.3. L'ordre verbe / sujet :

Nous affirmons que l'élément sujet a été étudié par les grammairiens en tant qu'élément principal de la phrase et ils repèrent que dans un énoncé comportant un verbe il est nécessaire pour que cet énoncé ait un sens. Et même dans les cas d'usage où le sujet n'est pas exprimé par un terme apparent, identifier le sujet est important pour l'opération communicative. Pour faciliter ce repérage les grammairiens ont cherché dans la problématique de l'ordre et ont présenté l'ordre (verbe-sujet) comme étant l'ordre de base. Ainsi, dans l'usage courant de la langue, le sujet prend sa position juste après le verbe [12].

Dans la phrase "ابتلى إبراهيم ربه" le sujet est séparé de son verbe. Cette séparation est possible grâce à la présence d'un pronom personnel renvoyant au sujet. Par contre elle est impossible dans "ضرب غلامه زيد" [12].

### II.5.4. Le statut de sujets (نائب فاعل):

#### L'étude de l'élément (نائب فاعل) :

Est considéré comme un complément pour le sens (مفعول في المعنى). En effet, en arabe il existe deux façons de bâtir le verbe:

**A)** - a la forme active, le verbe est dit (مبني للمعلوم), c'est à dire accordé à un sujet prononcé dans la phrase et connu par les interlocuteurs. Dans l'exemple "نصح العالم الأمير", le verbe (نصح) « a conseillé » est employé à la forme active puisque le sujet (العالم) est « le savant » est mentionné dans l'énoncé [12].

**B)** - a la forme passive le verbe est dit (مبني للمجهول) accordé à un inconnu, dans ce cas le sujet n'est pas mentionné dans la phrase, il est soit inconnu par le récepteur uniquement soit par les deux interlocuteurs à la fois. Le verbe dans ce genre d'usage est bâti sur le (مفعول) et non pas sur le (فعل). Dans "نُصح الأمير" « a été conseillé » (نُصح) est utilisé à la forme passive étant donné que le sujet n'est pas mentionné dans l'énoncé. " [12].

### II.5.5. L'ordre des compléments par rapport au verbe :

Les grammairiens ont traité la question de l'ordre des compléments dans le cadre des possibilités de réarrangement de la phrase qu'offre l'arabe. Etant donné que la phrase verbale



est considérée la plus répandue dans l'usage de la langue, celle-ci fournit au locuteur la possibilité d'employer une multitude de compléments [12].

Nous savons que pour les composants de base, la position de la grammaire traditionnelle est claire : le (عامل) prend une position avancée par rapport au (المعمول) et le (المسند) doit être placé avant le (المسند إليه), ainsi le sujet (الفاعل) est avant l'objet (المفعول), le verbe est par ailleurs avant le sujet. Quand aux compléments, l'élément le plus proche du verbe est le complément d'objet externe [12].

Nous schématisons cet ordre de la manière suivante :

**ph = vrb + suj + obj**

Il paraît que la tradition grammaticale arabe va en faveur de cet ordre et le considère comme l'ordre de base de la phrase arabe. Les opérations de déplacement sont réglementées par des critères syntaxiques (comme on l'a déjà signalé dans "l'ordre des mots dans la phrase arabe") [12].

### II.6. Les différents aspects du verbe :

Pour les grammairiens arabes le verbe est un élément essentiel dans la construction de la phrase. Associé au sujet, il constitue le noyau central de cette phrase. Autour de ce noyau s'ordonnent les autres éléments. Pour cette raison il est classé parmi les composants de base de la phrase [12].

#### II.6.1. L'aspect du temps :

Le temps, c'est la relation entre ce dont on parle et le moment où l'on parle : un procès est présenté comme antérieur, simultané ou postérieur au moment de l'énonciation [12].

L'aspect c'est la façon dont un procès se déroule dans le temps. Si le procès se déroule dans la période de temps concerné par l'énonciation, l'aspect est accompli : accompli [12].

Veut dire donc « s'accomplir » (dans la dite période). S'il est présenté comme la trace, dans cette période, d'un accomplissement antérieur, l'aspect est accompli [12].

En arabe, il paraît que la question du temps en général évoque certaines remarques, car les chercheurs trouvent que les deux structures (فعل) et (يفعل) ne peuvent en aucun cas englober

en toute précision les détails temporels. Pour cela, dans l'usage de la langue, le verbe arabe se présente souvent accompagné de certaines prépositions et d'indications de temps. Citons à titre d'exemple les deux particules (سا) et (سوف) [12].

En effet, les deux temps (مستقبل / الماضي) ne posent pas de problème, ainsi le verbe accompli indique que l'action est achevée dans le passé alors que le verbe inaccompli indique qu'elle est entrain de se réaliser sans être accomplie. Mais certains chercheurs affirment que la question du temps dans le verbe arabe est ambiguë [12].

Ils trouvent que le (المستقبل) futur est le même que l'inaccompli et sa structure dépourvu de la lettre (سين) renvoie au (المضارع). Ils affirment également que le (الأمر) l'impératif qui représente la structure par laquelle on ordonne au sujet de réaliser une action, rattaché par les grammairiens à l'inaccompli n'indique pas, par sa structure, un temps bien déterminé [12].

### II.6.2. L'aspect fonctionnel :

Parler de l'aspect fonctionnel, il s'agit donc d'étudier le verbe par rapport au rôle qu'il joue dans la phrase. Nous avons signalé que le verbe est un élément central autour du quel les éléments de la phrase verbale s'ordonnent [12].

En effet, dans la grammaire arabe classique, définir le verbe par son rôle dans l'opération prédicative était le fruit d'une recherche dans la structure de la phrase [12].

Dans l'ouvrage intitulé al-mūğaz, ibn al-ssarrāğ rejoint l'opinion de ibn ya'īs. Il nous rapporte la citation suivante : « le verbe est ce qui est prédicat et qui ne peut pas être prédicande » (الفعل ما كان خيرا ولا يجوز أن يُخبر عنه). Ceci dit que le verbe ne peut être que prédicatif. C'est l'élément qui émet l'information (الخبر) [12].

Nous trouvons cette même tendance chez al-ssayūṭī dans al-'ašbāh wa -l-nnazā'ir quand il dit : « la preuve que tous les verbes sont masculins est que lorsque nous informons sur les noms (par le biais des verbes) nous ne le ferons que par l'action qui existe dans le verbe étant le (المصدر).les verbes servent donc à informer sur les noms [12].

Les chercheurs contemporains affirment que quand il y'a un verbe, celui-ci est toujours le nœud central de la phrase. « L'élément central autour duquel s'organise l'énoncé est appelé prédicat » ils remarquent en même temps que le verbe est l'élément indispensable à la constitution d'un énoncé assertif fini. Le verbe est donc le syntagme susceptible de fournir à lui seul un énoncé complet. Il a un rôle dominant dans la construction de la phrase [12].

### II.6.3. L'aspect sémantique :

La notion de la sémantique correspond en langue arabe au sens lexical (المعنى الدلالي), celui-ci est défini par opposition au sens syntaxique. Dans l'usage de la langue, les noms et les verbes sont les deux catégories grammaticales les plus fondamentales. Sémantiquement ils représentent l'opposition conceptuelle maximale [12].

De point de vue sémantique, le verbe arabe est très riche grâce à son système de dérivation qui permet d'exprimer l'intensité, le but, la réciprocité, le factif ainsi que les réfléchies passives et les différents sens indispensables pour assurer la fonction communicative. Grâce à ce système souple de dérivation, l'arabe peut créer des formes dérivées du verbe par modification des voyelles, par doublement de la deuxième radicale, par adjonction et même par intercalation d'affixes [12].

L'un des chercheurs occidentaux affirme : « l'arabe possède un procédé original pour exprimer par dérivation des procès de plus en plus nuancés par rapport au sens de racine, représenté le plus souvent par le verbe trilitère que nous connaissons, et que nous appelons conventionnellement « de première forme » [12].

Il consiste à construire sur la racine, grâce à des préfixes, à des infixes ou à des redoublement, selon des schémas immuables, des « formes verbales dérivées » exprimant toujours la même nuance de sens par rapport à la racine, ou au verbe de forme [12].

## II.7. Difficultés de l'analyse automatique de textes arabe :

### II.7.1. Segmentation des textes arabes :

La grande majorité des applications de traitements du langage naturel présuppose le découpage de textes en phrases, nous parlons de reconnaissance de frontières des phrases qui est considérée comme l'unité centrale des processus du traitement du langage naturel, comme par exemple l'étiquetage [3].

On reconnaît comme phrase la suite des mots qui se trouvent entre des signes de ponctuation dits majeurs tels que le point, le point d'exclamation, le point d'interrogation et d'autres qui précèdent ou suivent ces signes[3].

Les travaux de segmentation des textes arabes ne sont pas nombreux et ceci pour plusieurs raisons, parmi lesquelles : la richesse de cette langue et sa complexité grammaticale, l'indisponibilité de documents numériques, en effet la majorité des textes numériques se trouvant sur le net sont scannés et sont présents sous format image, etc [3].

Le problème de segmentation pour l'arabe réside de la richesse grammaticale de cette langue. En effet, une phrase peut être structurée de différentes manières sans pour autant changer le sens de la phrase [3].

En effet, la liaison entre les phrases se fait en utilisant les conjonctions de coordinations, tels que le waw, و le fa, ف etc. Par ailleurs, ces particules n'ont pas uniquement le rôle de conjonctions, mais elles peuvent jouer d'autres rôles. Par exemple, le waw و peut être utilisé pour exprimer un adjectif ou bien pour la liaison entre les compléments d'objets etc. Ces particules peuvent aussi représenter une partie du mot comme par exemple وهم qui peut se lire wahmon (imagination) ou bien wa hom (et + ils) [3].

A ces difficultés de segmentation s'ajoute le problème de non voyellation, qui rend, dans certains cas, les caractéristiques morphologiques du mot ambiguës. Ainsi par exemple le mot فهم peut se lire : fahmon, (un nom: compréhension) ou bien fahima, (un verbe: il a compris), ou encore fa hom ("ils" précédé de la conjonction de coordination "ف") et par suite lors de l'analyse de ce mot, nous pouvons lui attribuer les trois catégories différentes: nom, verbe ou pronom précédé d'une conjonction de coordination [3].

Les règles de segmentations peuvent être classées en trois principales classes : les règles basées sur les signes de ponctuation, les règles basées sur les conjonctions de coordination et les règles basées sur certains mots connecteurs (exemples : إذا, بينما, لكن) [3].

### II.7.2. L'analyse morphologique :

L'opération de l'analyse morphologique tient à étudier la forme d'un mot en faisant une analyse interne de la structure de ce dernier. Le but étant de décomposer un mot à des éléments plus petits (préfixes, suffixes, etc.) Selon des règles de combinaison relatives à ces derniers [13].

A proprement parler, l'analyse morphologique ne fait que la séparation et l'identification des morphèmes semblables aux mots préfixés (comme les conjonctions "wa l

و" et "ف | ا", etc.) Des prépositions préfixées (comme "ب | ا" et "ل | ا", l'article défini "ال", etc.), des suffixes de pronom possessif [13].

L'analyse morphologique détermine un schéma possible. Les préfixes et suffixes sont trouvés en enlevant progressivement des préfixes et des suffixes et en essayant de faire correspondre toutes les racines produites par un schème afin de retrouver la racine [14].

Le problème principal de cette analyse réside dans l'agglutination et l'absence de voyelles. Pour l'agglutination et contrairement aux langues latines, en arabe, les pronoms, les prépositions, les articles, les conjonctions, et autres particules collent aux noms, verbes, adjectifs et particules auxquels ils se rapportent. Comparé au français, un mot arabe peut parfois correspondre à une phrase française [14].

Cette caractéristique engendre une ambiguïté morphologique au cours de l'analyse. Ainsi, la reconnaissance des unités lexicales qui composent une unité morphologique n'est pas toujours facile à détecter [3].

Le problème est de reconnaître que la bonne segmentation réside ainsi, dans la difficulté de distinction entre un proclitique ou enclitique et un caractère original du mot l'absence de voyelles pose un autre problème important [3].

En effet, les mots non voyelles engendrent beaucoup de cas ambigus au cours de l'analyse (e.g. Le mot non voyelle " فصل " pris hors contexte peut être un verbe au passé conjugué à la troisième personne du singulier "il a licencié | فصل", ou un nom masculin singulier "chapitre/ saison ' فصل", ou encore une concaténation de la conjonction de coordination "puis | ف" avec le verbe " صل " : impératif du verbe lier conjugué à la deuxième personne du singulier masculin) [3].

### II.7.3. L'analyse syntaxique :

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 morpho-syntaxique) et fournit en sortie une structure hiérarchisée des groupements structurels et des relations fonctionnelles qui unissent les groupements [3].

Enfin, il est à signaler que les ambiguïtés vocaliques et grammaticales, relatives à la non voyellations des mots, pose des difficultés au niveau de l'analyse syntaxique. Ainsi, une phrase, en absence de la voyellations, peut être interprétée et traduite selon plusieurs interprétations qui sont toutes syntaxiquement correctes [3].

### **II.8. Les ressources linguistiques similaires :**

#### **II.8.1. Wiktionnaire :**

Wiktionnaire est un dictionnaire multilingue, universel et librement diffusable. Depuis son lancement officiel par Jimmy Wales et Larry Sanger le 15 janvier 2001. Le nom de «Wiktionnaire» se compose de deux termes «wiki» et «dictionnaire». Un wiki est une application Web permettant l'édition collaborative simplifiée de pages [15].

Un dictionnaire est un lexique contenant l'ensemble des mots d'une langue ou d'un domaine d'activité fournissant pour chacun une définition et d'autres informations linguistiques [15].

La version arabe de Wiktionnaire est lancée le 24 mai 2004, elle contient au 11 mars 2012 plus de 48,409articles [15].

#### **II.8.1.1. Architecture de Wiktionnaire :**

Cette approche est composée de 3 parties, la première est une phase de préparation (prétraitement et extraire les définitions), la deuxième phase est l'analyse des vocabulaires de ces définitions ainsi l'extraction des relations sémantiques. La dernière phase consiste à créer une base lexicale. Le schéma 1 montre l'architecture de cette approche [15].

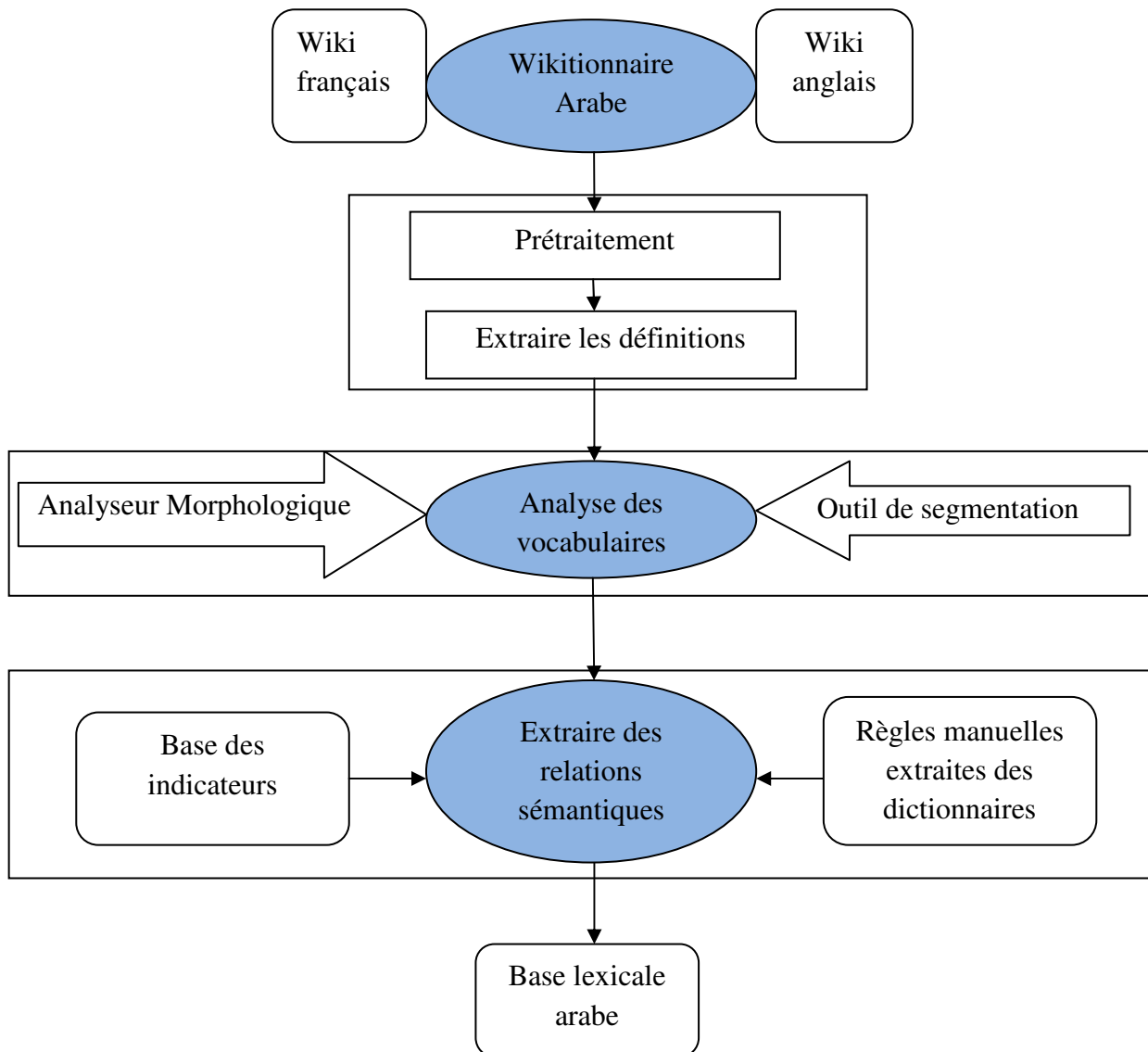


Schéma II.1. Architecture de l'approche.

#### a. Prétraitement et Extraction des définitions :

L'extraction des informations lexico-sémantiques disséminés dans les bases de connaissances collaboratives nécessite des outils d'accès automatiques à son fichier XML (Fig. II.1).

Ces outils sont disponibles pour les langues comme l'anglais, l'allemand et le portugais, dans cet objectif nous avons développé un outil pour analyser la structure du fichier XML pour la langue arabe. Qui peut supprimer les lettres latines, les chiffres, les caractères spéciaux,... [15].

```

<page>
<title>حَاسُوب</title>
.....
<textxml:space="preserve">wikipedia
=عَرَبِيَّة=
==المعاني==
# الحَاسُوب اسم مذكر يُجمع جمع تكسير على [[حَوَاسِب]]
ألة [[اللكترونية]] تقوم بعمليات [[حِسَابِيَّة]] سريعة بواسطتها يمكن للألة القيام بمهام مختلفة
على [[البيانات والمعلومات]]. [[استخدام حديث]] #
== المرادفات ==
# "1": [[حَاسِبِيَّة]]
# "2": [[أَلِيَّحَاسِب]]
.....
</page>

```

**Figure II.1.** Les informations contenues dans le fichier XML pour l'entrée حَاسُوب dans le Wiktionnaire arabe.

Cet outil peut exporter toutes les définitions et les transformer en format convivial où chaque ligne contient l'entrée, sa partie de discours et sa définition. Comme le montre l'exemple suivant [15] :

الحاسوب	اسم	مذكر يُجمع جمع تكسير على حواسيب
الحاسوب	اسم	ألة الكترونية تقوم بعمليات حسابية سريعة بواسطتها يمكن للألة القيام بمهام مختلفة على البيانات والمعلوماتية استخدام حديث
الحاسوب	اسم	حاسبة
الحاسوب	اسم	حاسب آلي

**Tableau II.1.** Définitions obtenus à partir de l'entrée dans la figure 2.

### b. Analyse des vocabulaires :

Pour extraire les relations sémantiques à partir des définitions. Nous avons utilisé l'outil de segmentation AraSeg [16], pour découper les textes des définitions en unités lexicales: paragraphe, phrase, syntagme, mot graphique, mot-forme, morphème,...etc. Ensuite l'analyseur morphosyntaxique [17] utilise les différentes ressources de connaissance de la langue arabe pour extraire les lemmes et les classes grammaticales des mots constituant les définitions, le résultat de ces outils est présenté dans le tableau suivant [15] :



Mot	Lemme	Catégorie grammaticale
آلة	آلة	اسم, مبتدأ
الالكترونية	إلكترون	اسم, خبر
تقوم	قام	فعل
عمليات	عمل	اسم مجرور
.....	.....	.....

**Tableau II.2.** Exemple de résultat de l'analyseur morphosyntaxique de définition du mot الحاسوب.

**c. Extraction des relations sémantiques :**

L'extraction des relations sémantiques est basée sur plusieurs informations morphosyntaxiques, marqueurs et des règles inspirées du dictionnaire. La phase actuelle est très importante car elle affecte les résultats du système. Elle a pour but de déterminer la relation sémantique entre les entrées de Wiktionnaire et les mots constituant leurs définitions. Ce système fait l'extraction des relations à base des indicateurs et des règles conçues manuellement. Les indicateurs et ces règles ont été inspirés à partir des dictionnaires (القاموس المحيظ, لسان العرب [12]). Les indicateurs sont présentés sous forme d'une base contenant par exemple les pronoms personnels comme (هو، هي...) et autres indicateurs comme (مرادفه، .....)، (معناه، ضده). Cependant les règles représentent les structures des phrases [15].

Cette partie donne 8321 relations sémantiques induites à partir de 25037 Définitions. Les relations sont représentées sous forme de triplets (A, R, B) dont A est un mot dans la définition, B est l'entrée de Wiktionnaire et R est le nom de la relation [15].

L'entrée de wiktionnaire « Représenté par le caractère B »	Relation « représenté par le caractère R »	Mot dans la définition « représenté par le caractère A »
الحاسوب	synonyme	آلة حاسبة
العين	synonyme	مُعَايَنَة
العين	Antonyme	نهر
.....	.....	.....

**Tableau II.3.** Exemple de représentation des relations sémantiques.



pour extraction des informations morphosyntaxiques est un avantage et un défi car son contenu est en augmentation permanente [15].

### II. 8.2. Théorie de Structure Rhétorique :

La théorie des structures rhétoriques est l'une des théories qui a été appliquée avec succès dans l'analyse et la compréhension automatique de textes. C'est une théorie qui repose sur une analyse rhétorique, c'est-à-dire une vision linguistique de la structuration des textes. La RST a été développée dans les années 1980 à l'Institut de Sciences de l'Information de l'Université de Californie du Sud par un groupe de chercheurs intéressés par la génération de langage naturel : William Mann, Christian Matthiessen et Sandra Thompson, avec l'apport de Cecilia Ford, Barbara Fox et Peter Fritts [3].

Le but de la RST était dès le départ de construire une théorie adaptée non seulement à la description, mais surtout à la génération du discours, dans l'optique du traitement automatique du langage. La théorie a rapidement été adoptée par toute une communauté de chercheurs (Hovy, Scott et de Souza). De plus, alors qu'elle avait été créée par des informaticiens, cette théorie s'est trouvée très utile aux linguistes s'intéressant aux problèmes de structuration du texte [3].

Mann et Thompson ont défini un cadre théorique qui permet de saisir la cohérence d'un texte. La RST est un outil descriptif, analytique qui peut être utilisé pour un grand nombre de types de textes, mais Mann et Thompson se limitent au départ à l'analyse du discours narratif. L'objet premier de la RST est donc d'étudier des textes *écrits*, organisés intentionnellement. La RST étudie les relations entre différentes portions d'un texte, et cela à tous les niveaux de sa structure. Afin de pouvoir subdiviser un texte et de déterminer les relations qui relient ses parties, on doit définir pour ce texte une unité de subdivision élémentaire ainsi que ce qui permet de lier ces différentes subdivisions entre elles [3].

#### II.8.2.1 L'analyse rhétorique de textes Arabe :

##### a. Segmentation du texte :

- ✚ Segmenter le texte en unités textuelles (phrases) : selon les signes de ponctuation (le point, le point d'interrogation, le point d'exclamation) [3].
- ✚ Segmentation des phrases en unités minimales : en se basant sur les indicateurs prototypiques (Cues phrases) [3].

##### b. Application des règles rhétoriques :

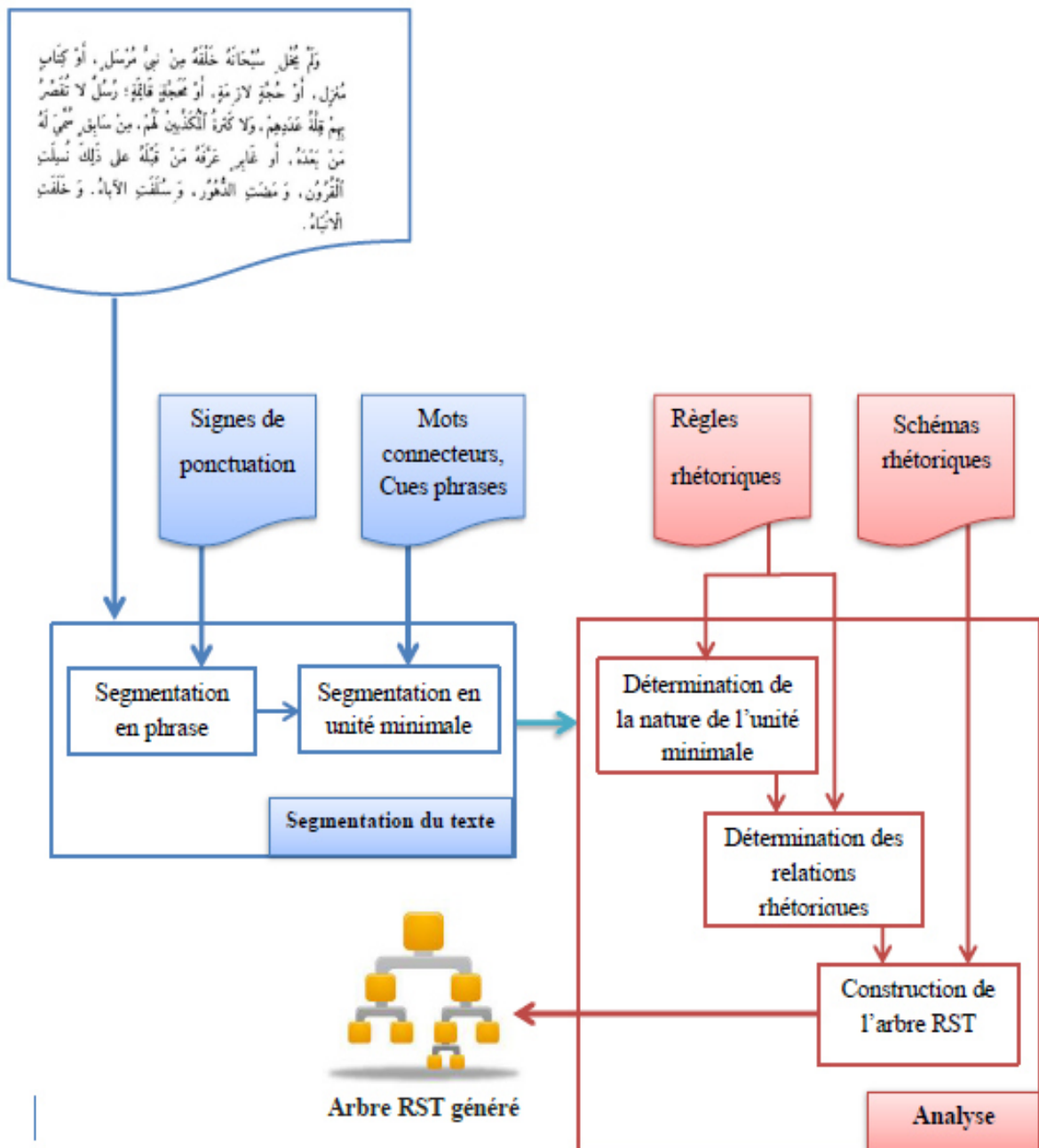
- ✚ Détermination du segment Noyau et celui Satellite : selon des règles sur les Cues phrases [3].

- ✚ Détection des relations rhétoriques : à partir la Cues phrases indiquer dans l'étape précédent (1) et un ensemble des contraintes sur les unités (Noyau et Satellite) [3].

**c. Construction des arbres RST :**

Une fois l'étape de détection du type des unités minimales et des différentes relations rhétoriques existantes est achevée, nous ajoutons à notre technique les schémas rhétoriques afin de spécifier la composition structurale du texte et construire l'arbre RST [3].

**II.8.3 Architecture de cette approche :**



**Figure II.3.** Architecture de RST.

### **II.9. Conclusion :**

Nous avons cherché dans ce chapitre à présenter de façon globale la relation sémantique verbes-sujets et verbe-objets et nous présenter aussi les différents aspects du verbe dans la langue arabe et précisément difficultés de l'analyse automatique de textes arabe et sa particularité par rapport aux autres langues et nous présentons quelques ressource similaire de notre travaille et quand va basé sur elles prochainement pour développer notre application. Par la suite, on présente les moyens et les outils possibles pour développer l'application, ce que on va présenter dans le chapitre suivant.

# ***CHAPITRE III: CONCEPTION***

### III.1. Introduction :

Une phrase simple a besoin d'un sujet, d'un verbe et parfois d'un objet. Afin de construire un arbre syntaxique il faut déterminer les verbes, les sujets et les objets. La compréhension de la phrase nécessite la reconnaissance de ces entités.

Par fois la classe grammaticale ne suffit pas pour détecter la relation entre le verbe et le nom qui le suit, par exemple la phrase " الحوت يصطاد " peut être ambiguë, car la baleine peut être sujet ou bien objet ou bien vice-sujet (فاعل، نائب فاعل، مفعول به) dans cette phrase, car la baleine peut chasser et peut être pêché (يصطاد الحوت ويُصطاد).

Par contre dans un autre exemple " النفط يستخرج ", on peut éliminer la fonction de sujet, car le pétrole ne peut qu'être extrait, donc il peut prendre la fonction objet ou bien vice-sujet (dans la voie passive).

Rappelons que le but de cette étude est d'implémenter un système d'extraction des relations verbes-sujets-objets à partir d'un corpus des textes arabes afin de créer un dictionnaire de relations sémantiques qui permet la réutilisation de telle ressources. Ce chapitre vise principalement à décrire l'architecture de notre système.

### III.2. Description de l'architecture de l'approche :

Nous décrivons, dans ce qui suit, les différents modules de cette architecture.

#### III.2.1. Le module de segmentation :

Ce module permet de segmenter un fichier au format texte d'abord en unités textuelles, en unités minimales, ensuite en sous-phrases.

#### III.2.2. Le module d'analyse :

Il permet de détecter les relations sémantiques qui relient deux unités minimales adjacentes entre elles. Ce module délivre un réseau sémantique qui qualifie les relations détectées entre les unités textuelles.

Ainsi, l'architecture globale de notre système est illustrée par la **Le schéma III.1.**

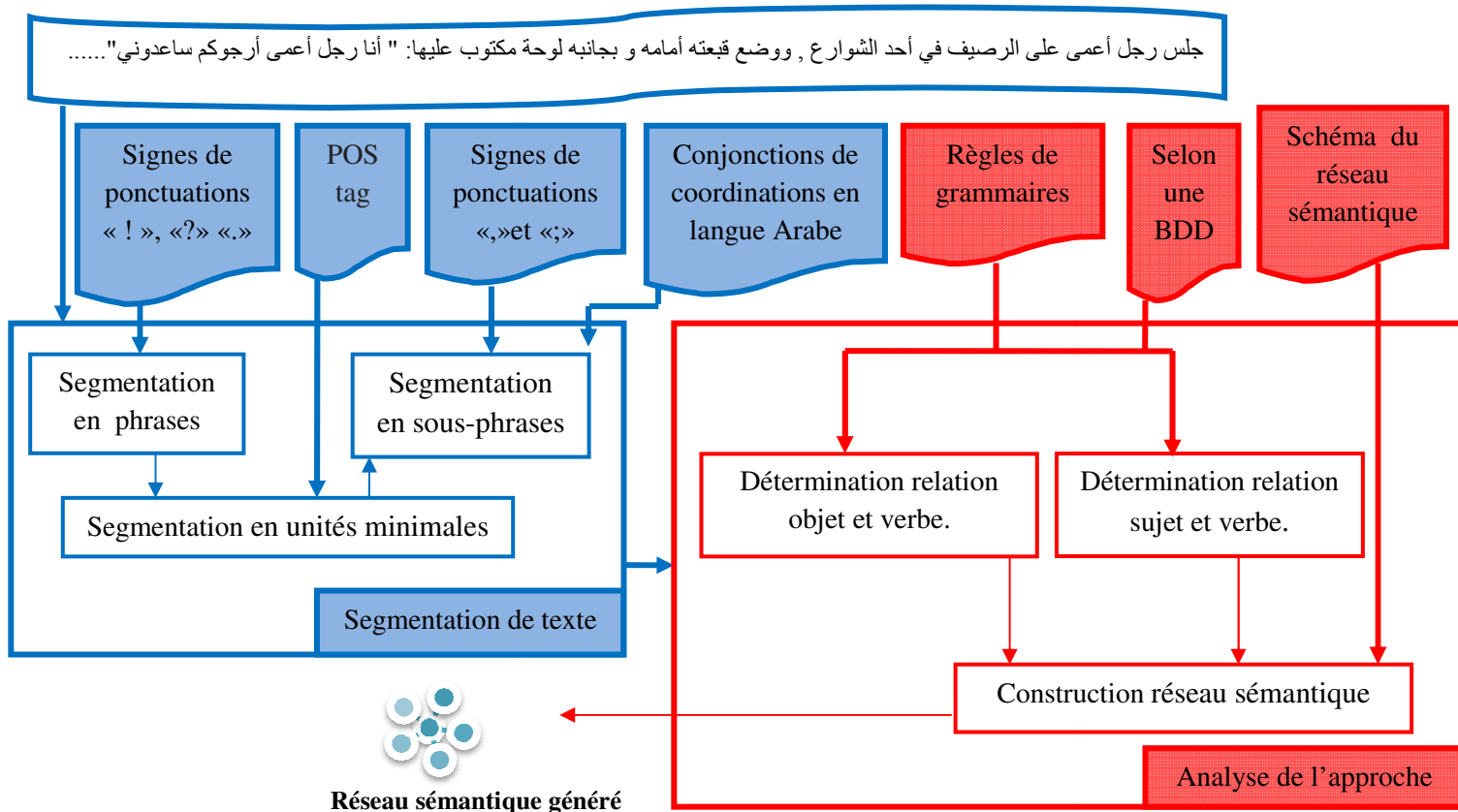


Schéma III.1. L'architecture globale de notre système.

### III.3. Description de l'algorithme de l'approche :

Notre outil d'extraction des relations sémantiques entre les mots arabes spécialement les relations entre sujets, verbes et objets met en œuvre les étapes décrites dans l'algorithme III.1 ci-dessous :

<b>Entrées :</b> texte Arabe	
<b>Sortie :</b> réseau sémantique	
<b>Début</b>	
1.	Segmentation du texte : <ul style="list-style-type: none"> <li>✓ segmenter le texte en unités textuelles (phrases) en se basant sur les signes de ponctuations « ! », « ? » et « , »</li> <li>✓ segmentation des phrases en unités minimales en se basant sur les POS tag.</li> <li>✓ segmentation des phrases en sous-phrases en se basant sur les conjonctions de coordinations en langue Arabe et les signes de ponctuations « , » et « ; » si existe.</li> </ul>
2.	Analyse de l'approche: <ul style="list-style-type: none"> <li>✓ Détection de relation verbe et sujet.</li> <li>✓ Détection de la relation verbe et objet.</li> <li>✓ Construction le réseau sémantique.</li> </ul>
<b>Fin</b>	
<b>Algorithme III.1</b> d'extraction des relations sémantiques entre les mots arabes.	



### III.3.1. Description détaillée de l'algorithme:

Si on vient à reconnaître le sujet (l'objet) par rapport au verbe, ça veut dire qu'on a bien avancé dans la sémantique. Dans ce qui suit nous allons décrire les trois grandes étapes :

#### III.3.1.1. Segmentation de texte :

La segmentation d'un texte est une étape fondamentale pour son traitement automatique ; son rôle consiste à diviser en unités dont l'un possède le sujet et l'autre l'objet de verbe.

Nous adoptons consiste à segmenter le texte source selon plusieurs niveaux : paragraphes, phrases et sous-phrases et nous nous appuyons sur un ensemble de règles permettant de tenir en considération.

##### III.3.1.1.1. Règles de segmentation :

Les règles conçues peuvent être classées en deux principales classes : les règles basées sur les signes de ponctuation, règles basées sur certains mots connecteurs (exemples : لكن, بينما, إذ, هكذا).

#### A. Segmenter le texte en unités textuelles (phrases):

Cette étape consiste à hiérarchiser et à structurer le texte source en différentes unités (phrases) tel que nous basons sur les signes de ponctuation (le point, le point d'interrogation et d'exclamation) par rapport à un texte arabe.

- **Le point** : est fait lorsque le sens de la parole est complet, lorsque la phrase est terminée, et la fin des paragraphes.
- **Le point d'interrogation** : est placé à la fin de la phrase d'interrogation.
- **Le point d'exclamation** : est placé à la fin des phrases expressives, qui expriment la joie, l'étonnement et la prière.

<b>Entrées</b> : texte Arabe <b>Sortie</b> : phrases Arabe	
<b>Début</b>	<b>SI</b> existe un signe de ponctions «.», « ? » ou « ! » <b>ALORS</b> segmenter le texte en phrase
<b>Fin</b>	
<b>Algorithme III.2</b> segmentation de texte en phrases.	

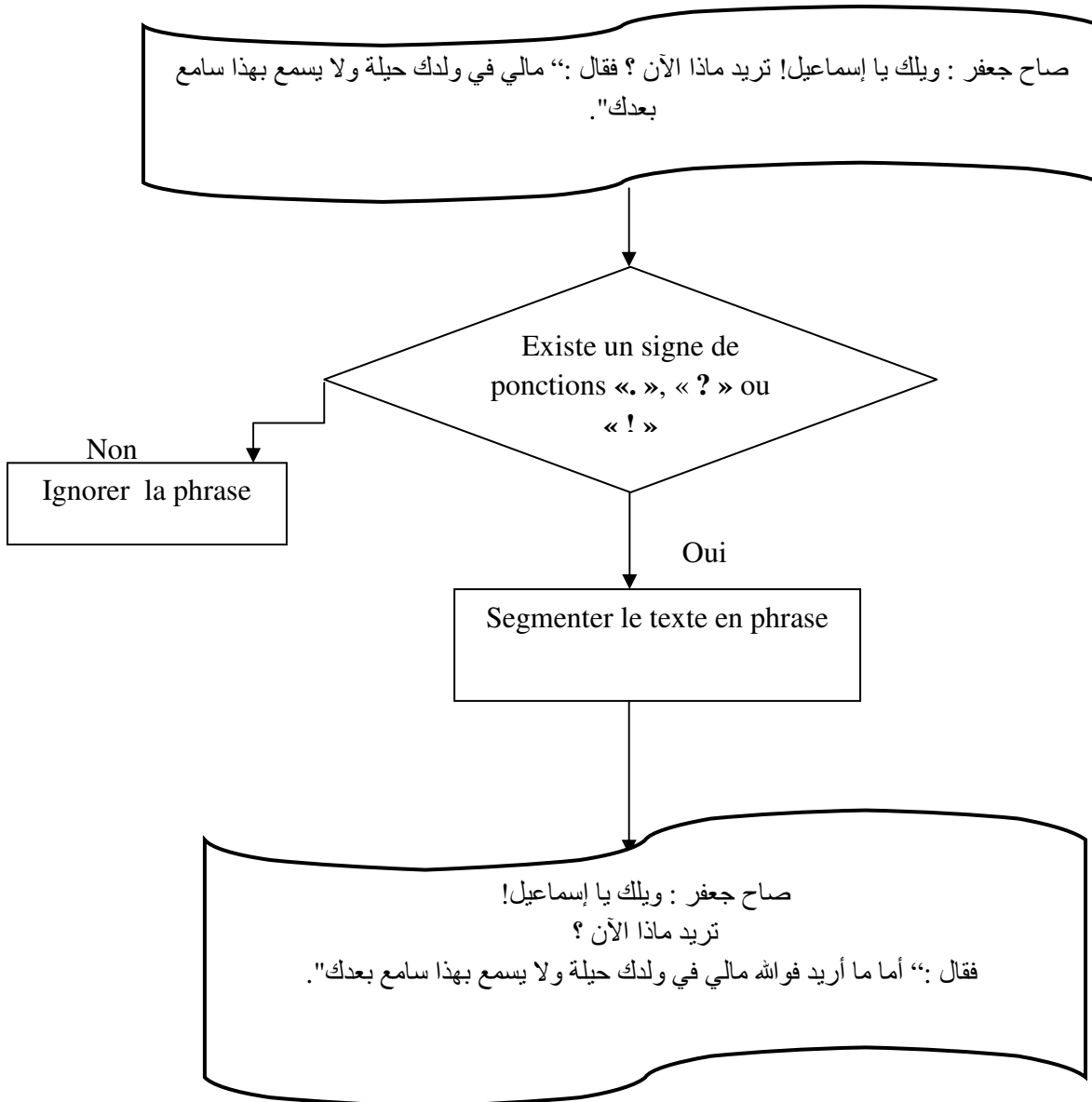


Schéma III.2. Exemple de segmentation de texte.

### B. Segmentation des phrases en sous-phrases :

La dernière étape de la segmentation est la segmentation des phrases en sous-phrases qui sont reliées entre elles par :

\*Les signes de ponctuation (le point-virgule et la virgule) :

**a. Le point-virgule :** est placé après une phrase qui est une cause ou un effet d'une autre phrase.

**b. La virgule :** est placée entre les phrases avec une signification connexe, après l'appel, et avant que les exemples soient énumérés, c.-à-d. avant le mot « comme ».

\*Les coordinateurs des phrases en langue Arabe):

La conjonction de coordination est un procédé syntaxique qui permet de coordonner deux noms ou deux phrases. Les trois conjonctions de coordination les plus connues sont :

- a. **فَ** exprimant un laps de temps très bref entre les deux actions: **دخل فجلس**; il est entré et [toute de suite] s'est assis.
- b. **وَ** exprimant un laps de temps moins bref:  
**دخل وجلس** ; il est entré et s'est [un peu de temps après] assis.
- c. **ثُمَّ** exprimant un laps de temps plutôt long entre les deux actions:  
**دخل ثم جلس** il est entré [ensuite] il s'est assis.

<b>Entrées :</b> texte Arabe <b>Sortie :</b> texte segmenté en phrase	
<b>Début</b>	<b>SI</b> nombre verbe dans la phrase $\geq 2$ <b>ALORS</b> chercher un signe de ponctuation « ; » ou « , » <b>SI</b> existe le signe de ponctuation <b>ALORS</b> segmenter le texte <b>SINON</b> rechercher un conjoncteur de coordination <b>SI</b> existe le conjoncteur de coordination <b>ALORS</b> segmenter le texte <b>SINON</b> mettre un signe de ponctuation et Segmenter le texte
<b>Fin</b>	
<b>Algorithme III.3</b> Segmentation des phrases en sous-phrases.	

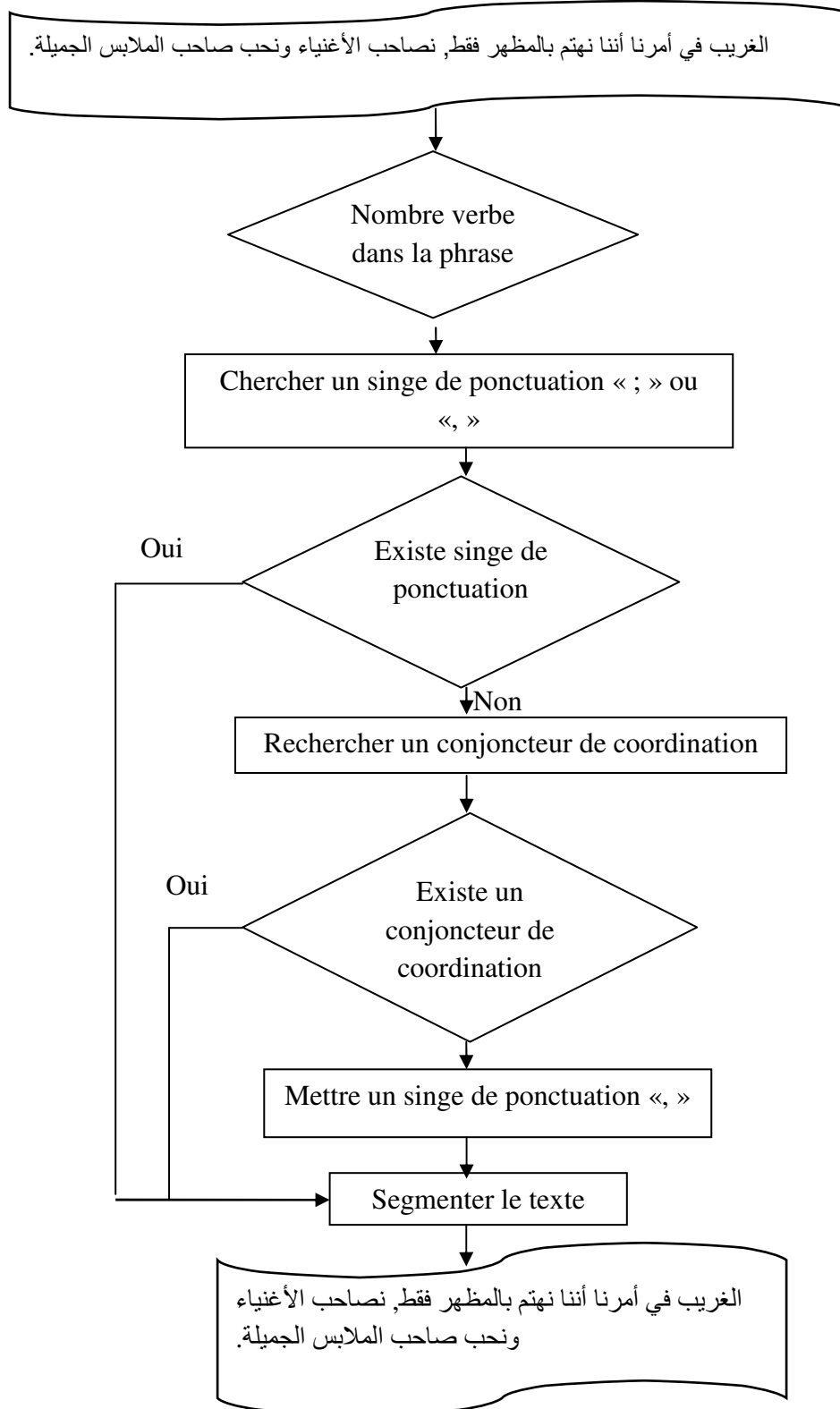


Schéma III.3. Exemple de segmentation des phrases en sous-phrases.

### *C. Segmentation des phrases en unités minimales :*

La deuxième étape de la segmentation est la segmentation des phrases en unités minimales en se basant sur les POS tag.

#### **III.3.1.1.2. Les tagger :**

Le processus de classification des mots dans leurs parties du discours et de leur étiquetage en conséquence, est connu sous le nom de marquage de partie de discours, de marquage POS ou simplement de marquage. Les parties du discours sont également appelées classes de mots ou catégories lexicales. Dans ce chapitre, nous mettons l'accent sur l'exploitation des balises et le marquage automatique du texte [18].

L'étiquetage de partie de parole consiste à classer les jetons de mots dans leur partie de parole respective et à les étiqueter avec l'étiquette de partie de parole.

Le balisage est effectué en fonction de la définition du mot et de son contexte dans la phrase [19].

#### ❖ Liste du tag :

<i>Le tag</i>	Description
<b>CC</b>	Conjonction de coordination
<b>DT</b>	Déterminant
<b>IN</b>	Préposition ou subordonnée conjonctive
<b>JJ/ JJS</b>	Adjectif superlatif
<b>NNP/NNPS</b>	Nom propre au pluriel
<b>NN/NNS</b>	Nom, pluriel
<b>PRP/ PRP\$</b>	Pronom personnel / Pronom possessif
<b>VB</b>	Verbe, infinitif
<b>VBD</b>	Verbe, passé
<b>VBG</b>	Verbe, nom verbal ou participe présent
<b>VBN</b>	Verbe, participation passée
<b>VBP</b>	Verbe, non-3rd person singular present
<b>VBZ</b>	Verbe, 3ème personne au singulier, présent
<b>DTNN</b>	Nom déterminé, singulier ou massif
<b>DTNNP</b>	Nom propre déterminé, singulier
<b>DTJJ</b>	L'adjectif du déterminant

<b>NNP</b>	Nom propre au singulier
<b>NN</b>	Nom, singulier ou massif
<b>JJ</b>	Adjectif

**Tableau III.1** : un ensemble de POSTag [22].

<b>Entrées</b> : phrase	
<b>Sortie</b> : liste des éléments de la phrase	
<b>Début</b>	<b>POUR</b> chaque mot dans la phrase Donner leur POSTag <b>SELON</b> le contexte.
<b>Fin</b>	
<b>Algorithme 3.4</b> Algorithme de POSTag.	

### III.3.1.2. Analyse de l'approche:

#### III.3.1.2.1. Application des règles de grammaires :

##### ❖ *Les verbes incomplètes* « الأفعال الناقصة » :

En arabe, il existe des verbes appelés « الأفعال الناقصة أو النواسخ » qui entrent dans la phrase nominale « ترفع المبتدأ فيسمى اسمها وتنصب الخبر ويسمى خبرها »

Par exemple : « كان خالدٌ مريضًا » → « خالدٌ مريضٌ ».

Pour ce la, on prend en considération que tous les verbes n'est pas un « فعل ناقص »

<b>Entrées</b> : éléments de la phrase dans une liste	
<b>Sortie</b> : verbe incomplet	
<b>Début</b>	<b>SI</b> verbe n'existe pas dans une <i>Liste</i> «af3alNa9isa» <b>ALORS</b> déterminer la position de verbe dans la phrase chercher le sujet et l'objet
<b>Fin</b>	
<b>Algorithme III.5</b> Algorithme de déterminer le type de verbe.	

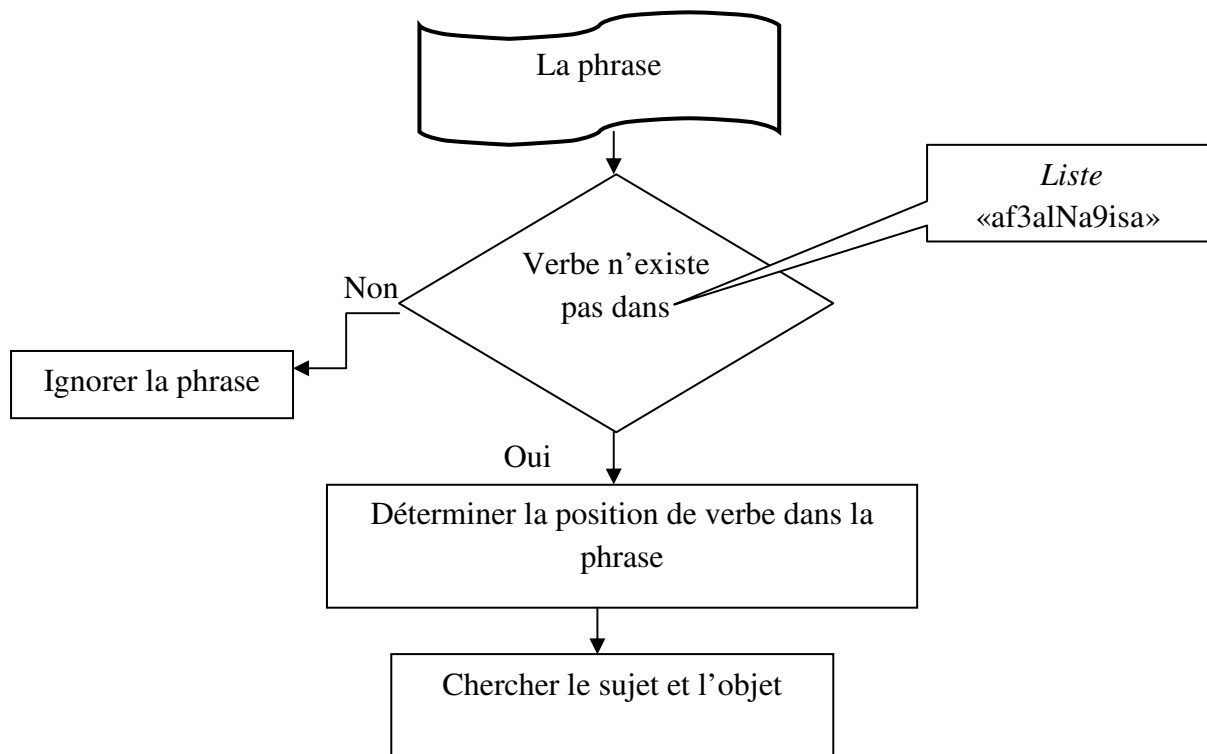


Schéma III.5.déterminer le type de verbe.

✓ **Détection du relation sujet et verbe :**

Pour chaque règle utilisée on donne un algorithme suivi par un schéma :

- *Le sujet est un nom ou un pronom :*

<b>Entrées :</b> éléments de la phrase dans une liste	
<b>Sortie :</b> verbe et leur sujet	
<b>Début</b>	Déterminer le temps de verbe Déterminer l'ordre des éléments de la phrase
<b>Fin</b>	
<b>Algorithme III.6.</b> Algorithme si le sujet est un nom ou un pronom.	

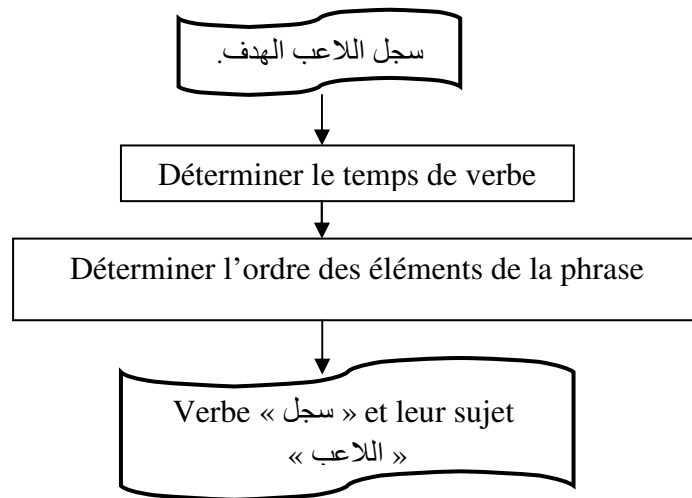


Schéma III.6. Le sujet est un nom ou un pronom.

- *Le sujet « أن اسمها وخبرها » :*

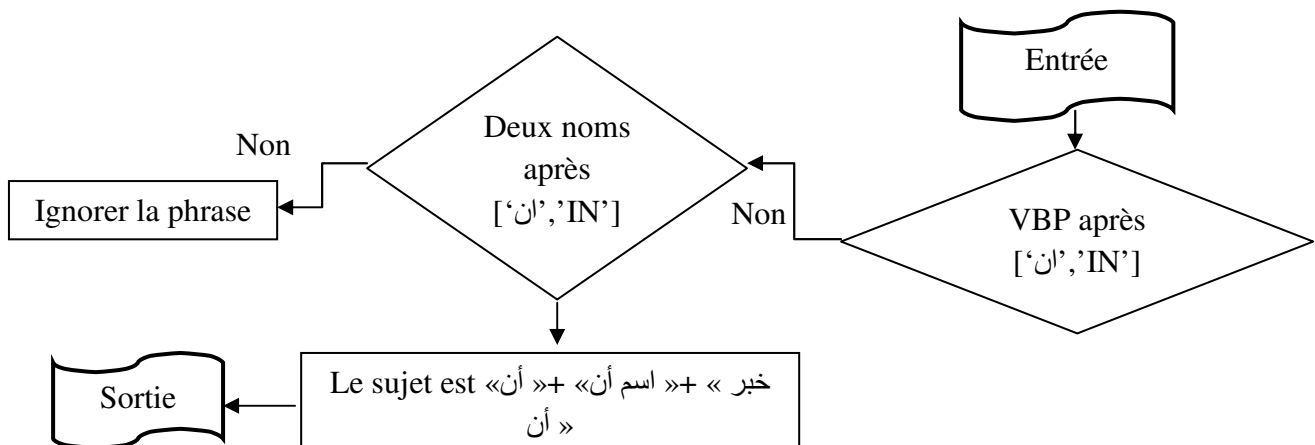
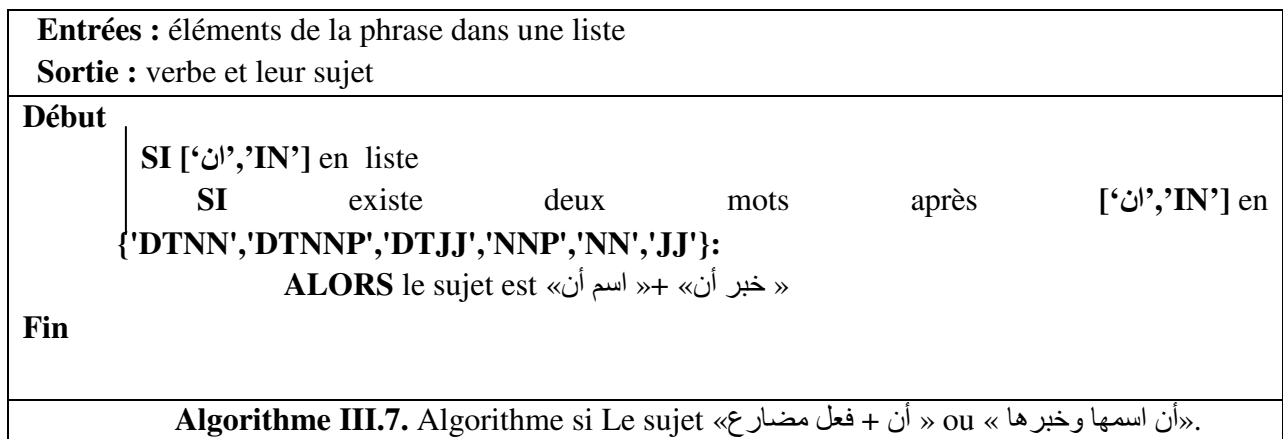


Schéma III.7. Le sujet « أن اسمها وخبرها ».



- *Le sujet « فعل » + « ما » :*

<b>Entrées :</b> éléments de la phrase dans une liste
<b>Sortie :</b> verbe et leur sujet
<b>Début</b>
SI existe mot après [‘ما’, ‘WP ‘] en {‘DTNN’,‘DTNPN’,‘DTJJ’,‘NNP’,‘NN’,‘JJ’}: ALORS le sujet est «nom de verbe»
<b>Fin</b>
<b>Algorithme III.8.</b> Algorithme si le sujet « فعل » + « ما ».

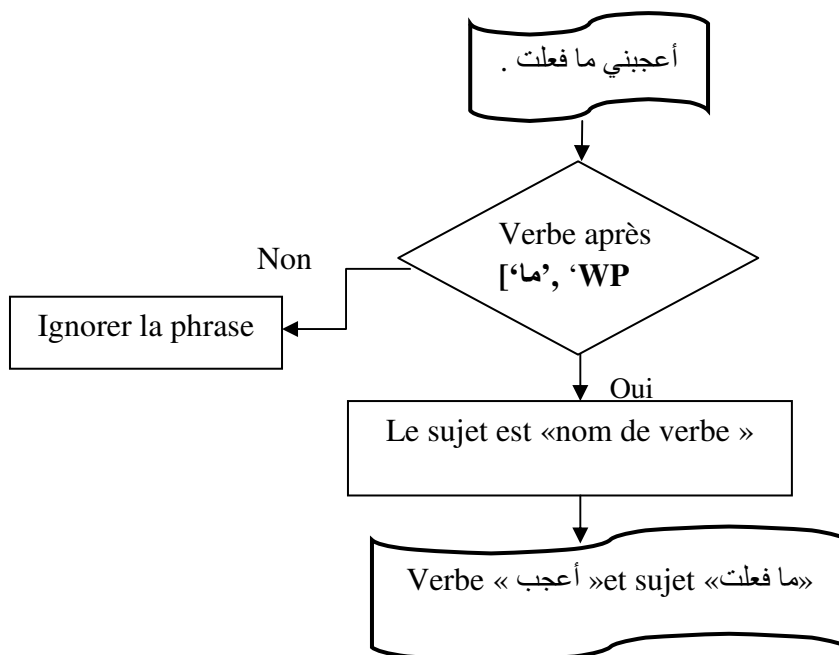


Schéma III.8. Le sujet « ما وفعلها ».

- *Le sujet « التاء المتصلة بالفعل » :*

<b>Entrées :</b> éléments de la phrase dans une liste
<b>Sortie :</b> verbe et leur sujet
<b>Début</b>
SI VBD ET terminé avec « ت » SI le mot terminer avec « ة » ALORS le mot est le sujet SINON SI le mot est dans une liste «Noms Féminins » ALORS le mot est le sujet SINON ALORS Le sujet « التاء المتصلة بالفعل »
<b>Fin</b>

Algorithme III.9. Algorithme si le sujet « التاء المتصلة بالفعل ».

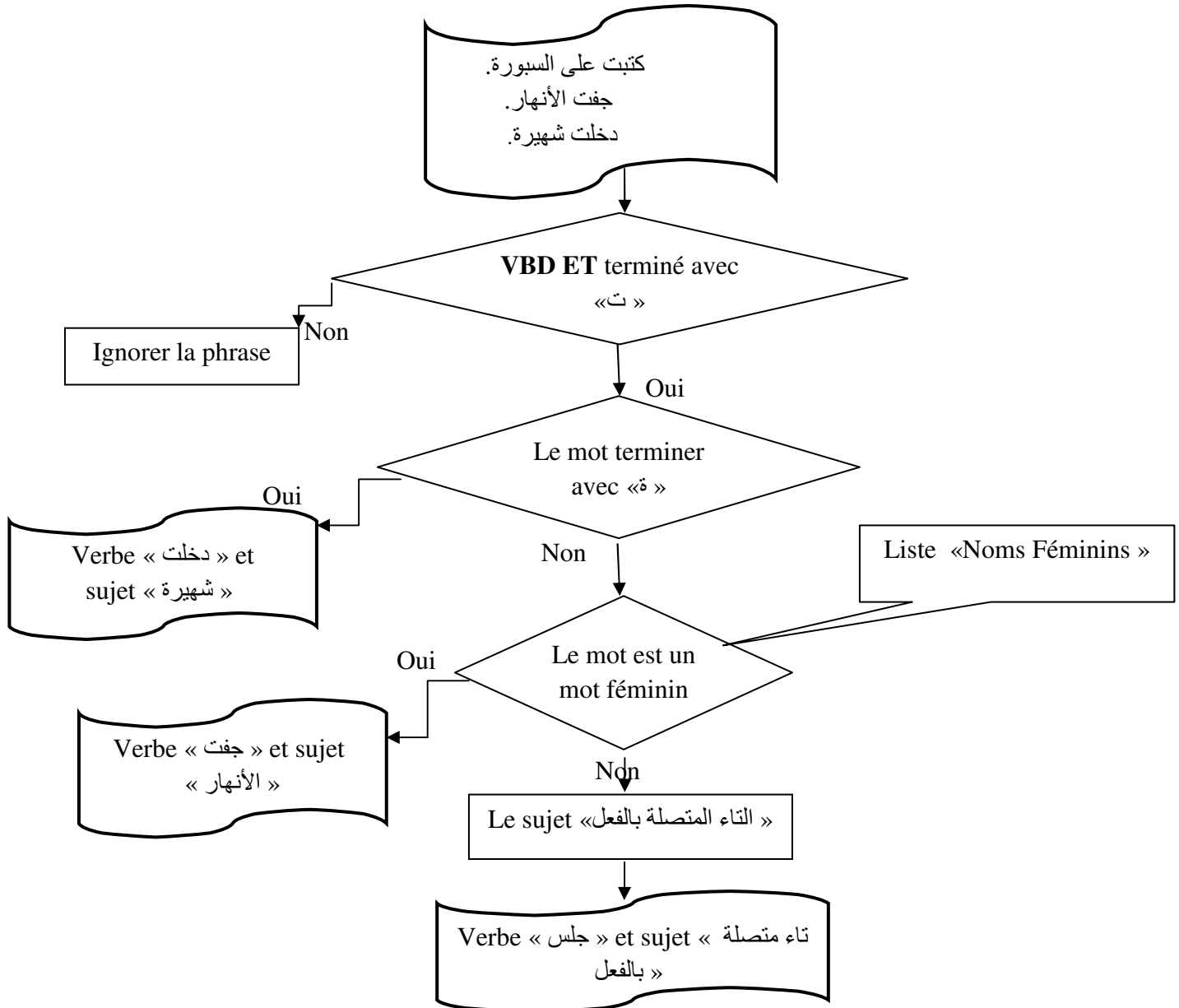
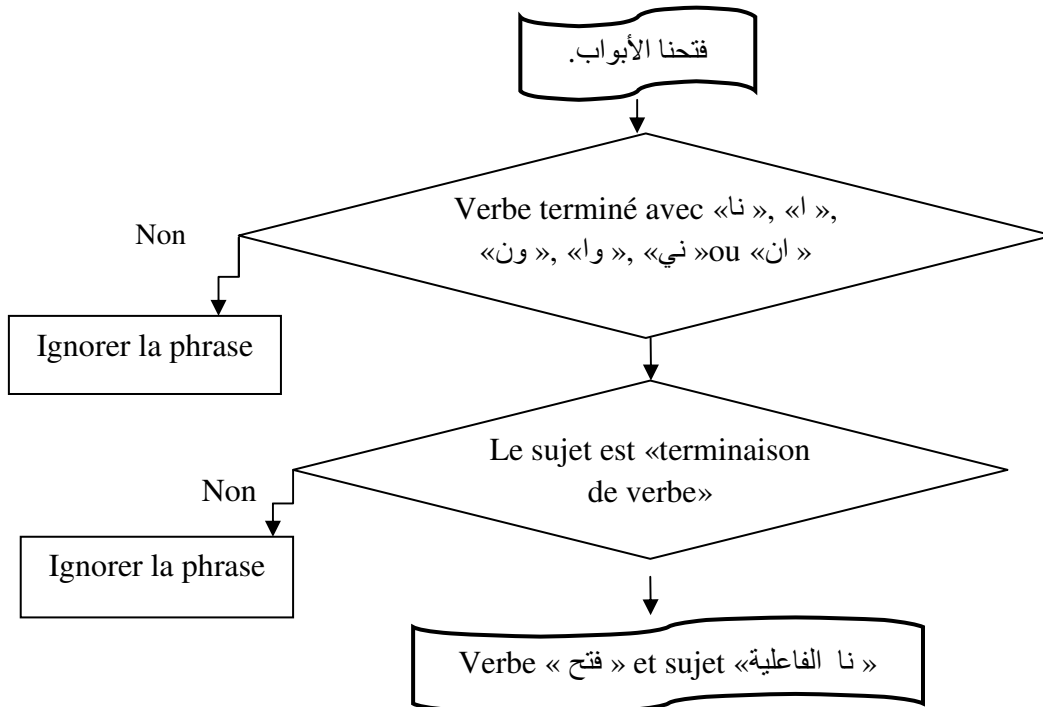


Schéma III.9. Le sujet « التاء المتصلة بالفعل ».

- *Le sujet est «نا الفاعلية», «ألف الإثنين», «واو الجماعة», «يا الفاعلية», «نون النسوة» :*

<b>Entrées :</b> éléments de la phrase dans une liste
<b>Sortie :</b> verbe et leur sujet
<b>Début</b>
SI verbe terminé avec «نا», «ا», «ون», «وا», «ني» ou «ان» : ALORS le sujet est « <i>terminaison de verbe</i> »
<b>Fin</b>
<b>Algorithme III.10.</b> Algorithme si le sujet est «نا الفاعلية», «ألف الإثنين», «واو الجماعة», «يا الفاعلية», «نون النسوة».



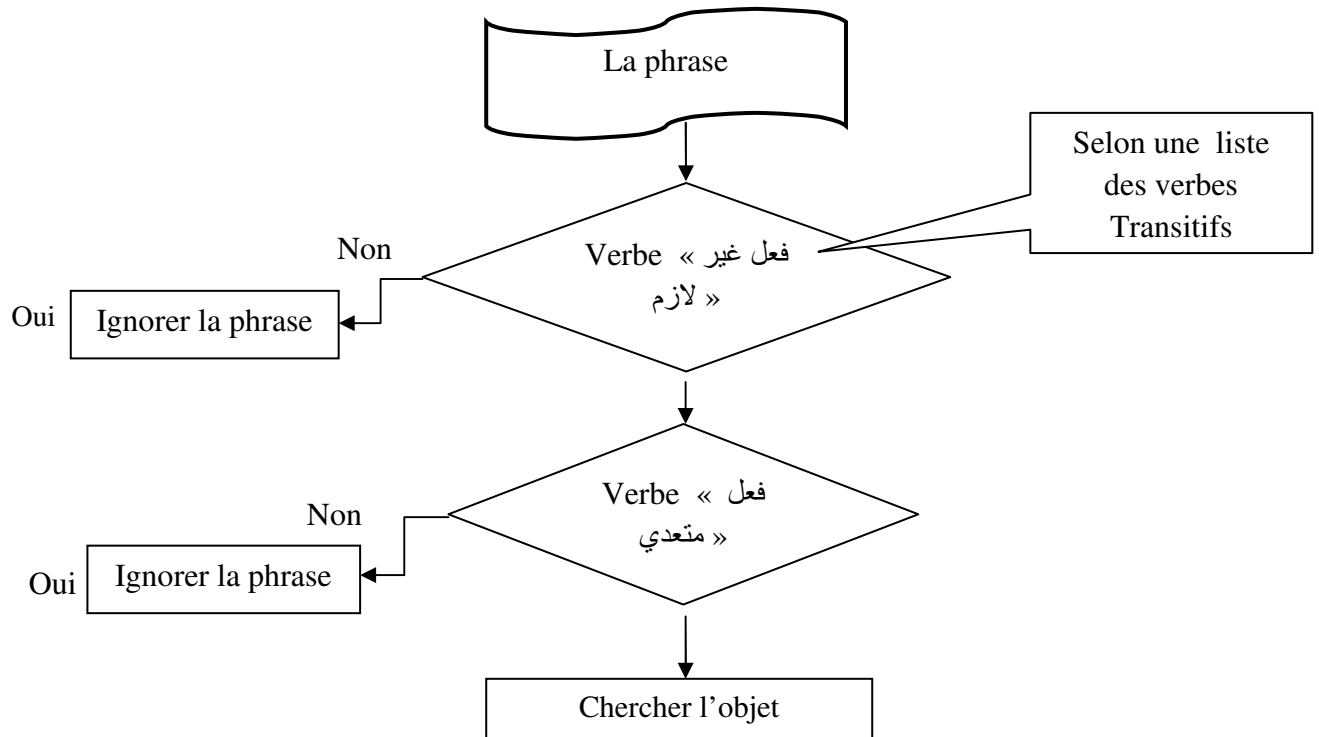
**Schéma III.10.** Le sujet est «terminaison de verbe».

Notez que tous les *sorties (couples verbe\_sujet)* sont mit dans une base de données pour les réutilisés.

✓ **Détection de la relation objet et verbe :**

Avant d'extraire l'objet de verbe il faut tout d'abord déterminer c'est le verbe est transitif ou pas. Dans la langue Arabe il y est deus types de verbe transitif : *فعل متعدي لمفعول* و*الفعل المتعدي لمفعولين*

<b>Entrées :</b> éléments de la phrase dans une liste	
<b>Sortie :</b> verbe transitif	
<b>Début</b>	SI verbe « فعل غير لازم » SI verbe « فعل متعدي » ALORS chercher l'objet
<b>Fin</b>	
<b>Algorithme III.11.</b> Algorithme de déterminer le type de verbe.	



**Schéma III.11.** Déterminer le type de verbe.

✓ **Les règles détection l'objet :**

Pour chaque règle utilisée on donne un algorithme suivi par un schéma :

- **L'objet est « ضمائر النصب المنفصلة » :**

<b>Entrées :</b> éléments de la phrase dans une liste	
<b>Sortie :</b> verbe et leur objet	
<b>Début</b>	SI « ضمير » de « ضمائر النصب المنفصلة » dans la phrase ALORS l'objet est « ضمير »
<b>Fin</b>	
<b>Algorithme III.12.</b> Algorithme si l'objet est « ضمائر النصب المنفصلة ».	

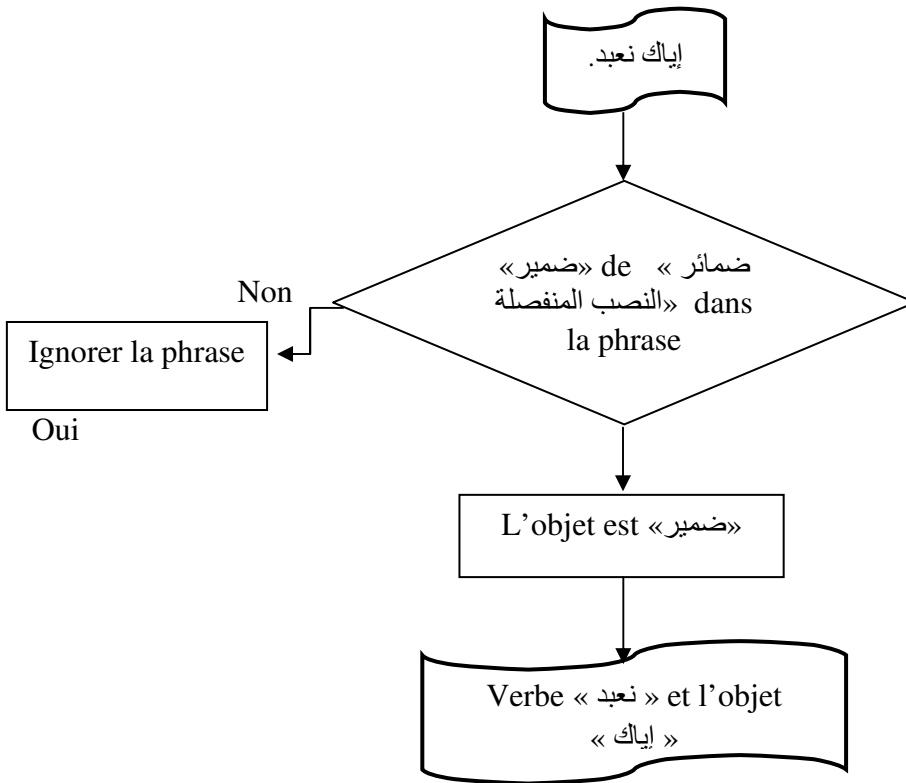


Schéma III.12. L'objet est «ضمائر النصب المنفصلة».

- **المفعول به ضمير متصل بالفعل :**

<b>Entrées :</b> éléments de la phrase dans une liste	
<b>Sortie :</b> verbe et leur objet	
<b>Début</b>	SI اتصل ضمير بالفعل ALORS l'objet est « ضمير متصل بالفعل »
<b>Fin</b>	
<b>Algorithme III.13.</b> Algorithme si l'objet est « ضمير متصل بالفعل »	

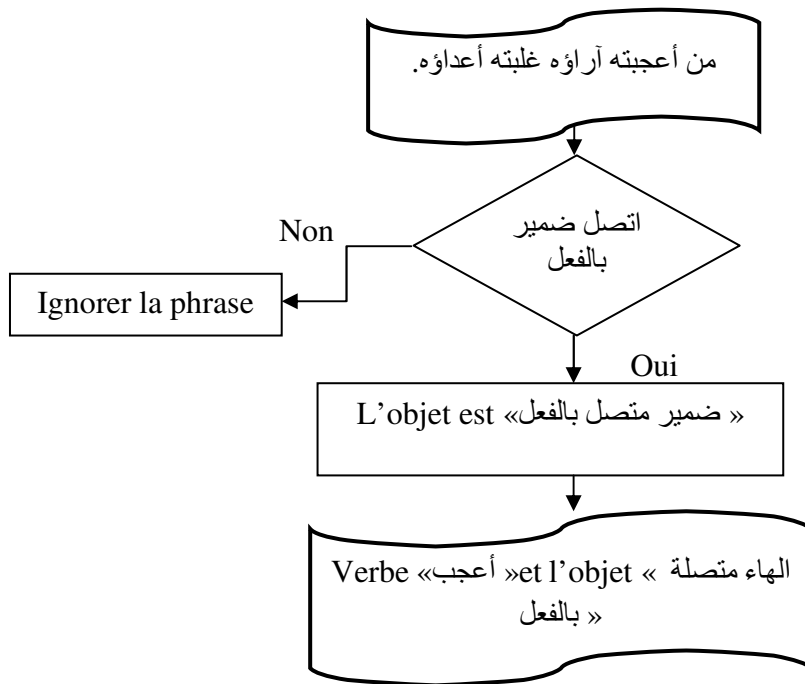


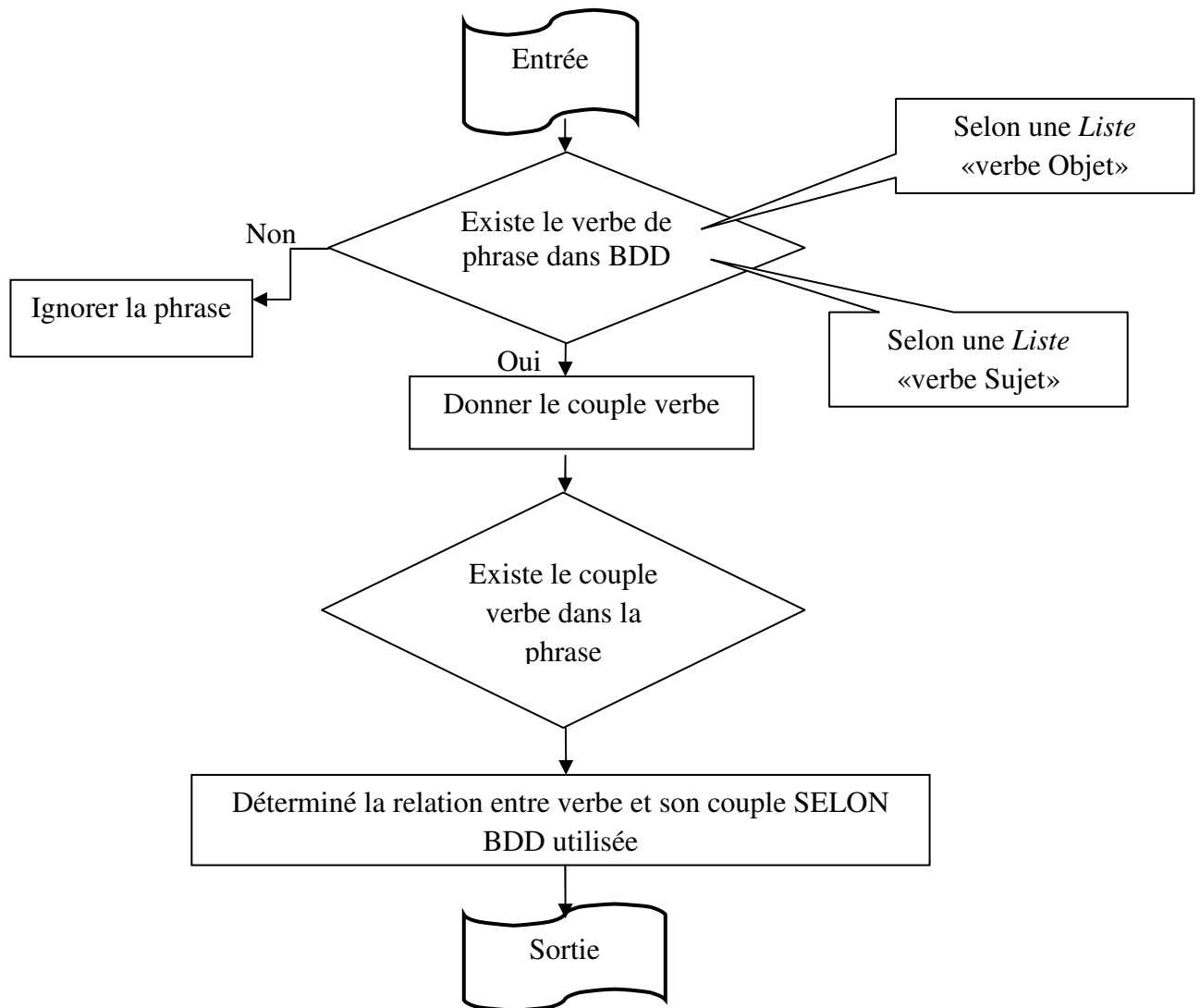
Schéma III.13. L'objet est « ضمير متصل بالفعل ».

Notez que tous les *sorties (couples verbe\_objet)* sont mit dans une base de données pour les réutilisés.

### III.3.1.2.2. Utilisation d'une base de données :

Dans certains cas, l'extraction de sujet de phrase est semblable à l'extraction d'objet. Ex le sujet possible est « محصور بإيلا » comme on trouve l'objet « محصور بإيلا » et tant que segmentation des textes avec NLTK éliminé « حركات الإعراب » on confond entre le sujet et l'objet. Pour éviter de tels cas, nous nous appuyons sur une base de données remplie manuellement, pour commencer. Ensuite, nous nous basons sur les résultats extraits mentionnés ci-dessus pour déterminer le sujet d' où objet.

<b>Entrées :</b> verbe	
<b>Sortie :</b> couple de verbe en BDD	
<b>Début</b>	<p><b>SI</b> existe le verbe de phrase dans BDD (<i>BDDverbe_sujet</i> ou <i>BDDverbe_objet</i>)          Donner le couple verbe</p> <p><b>SI</b> existe le couple verbe dans la phrase  <b>ALORS</b> déterminé la relation entre verbe et son couple <b>SELON</b> BDD utilisée</p>
<b>Fin</b>	
<b>Algorithme III.14.</b> Algorithme pour extrait relation sémantique selon BDD utilisée.	



**Schéma III.14.** Extrait relation sémantique selon BDD utilisée.

**A. Construction du réseau sémantique:**

Une fois l'étape de détection du type des unités minimales et des différentes relations sémantiques existantes est achevée, donc le réseau sémantique arbre pourrait être construit.

Le processus de construction du réseau sémantique est caractérisé par les propriétés suivant :

- ✓ D'un ensemble de nœuds étiquetés : représentant les sujets ou les objets.
- ✓ D'un ensemble de liens orientés et étiquetés entre ces nœuds : représentant généralement le type de relation entre deux nœuds.

<b>Entrées :</b> couple (verbe, nom)
<b>Sortie :</b> Réseau sémantique généré
<b>Début</b>
Mit le verbe et le nom dans des <b>NODEs</b> relie entre eux avec une <b>flèche étiqueté</b> <b>SI</b> nom est un sujet <b>ALORS</b> la flèche étiquetée est « علاقة فاعلية » <b>SINON</b> la flèche étiquetée est « علاقة مفعولية »
<b>Fin</b>
<b>Algorithme III.15.</b> Algorithme de déterminer le type de relation.

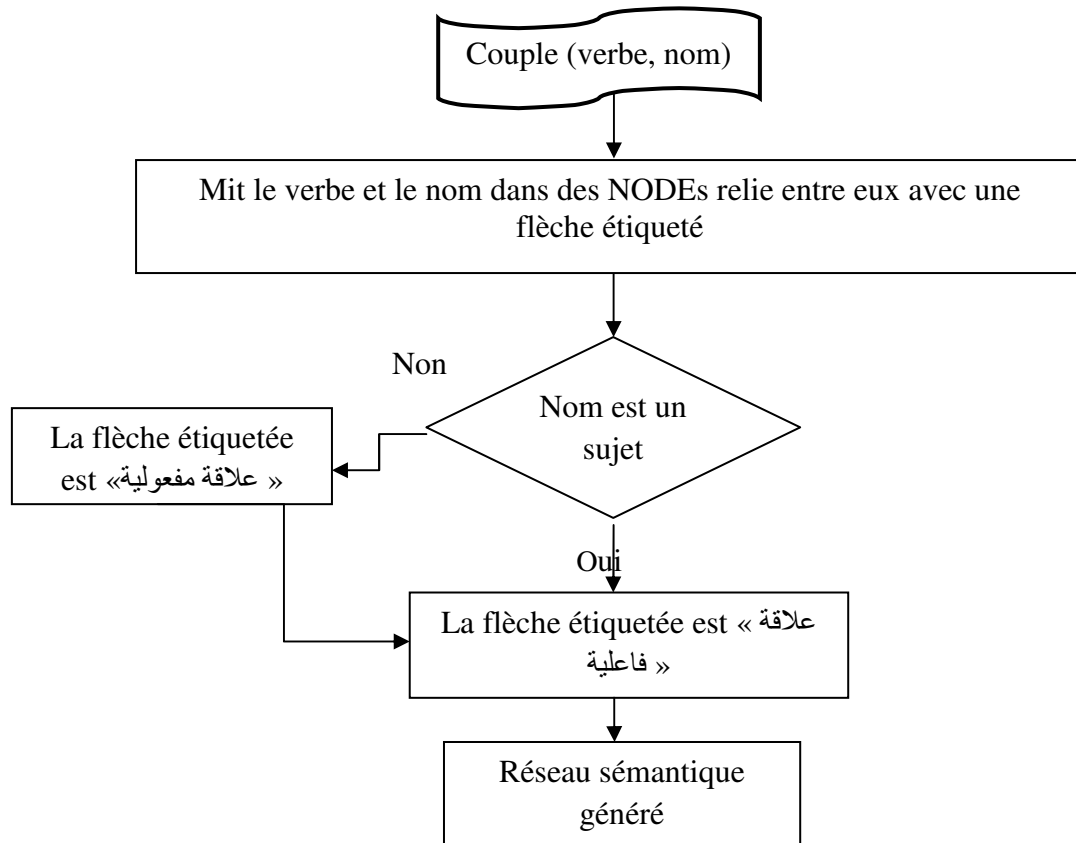


Schéma III.15. Détermination le type de relation.

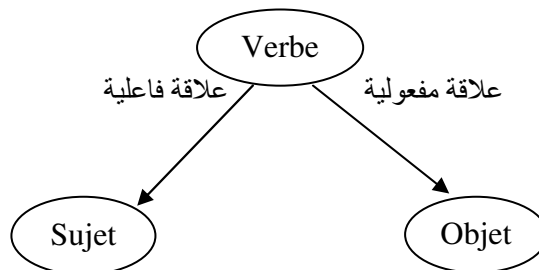


Schéma III.16. Détermination le type de relation.



### **III.4. Conclusion :**

Dans ce chapitre, nous avons conçu et documenté le code que nous devons produire. Dans cette phase, toutes les questions concernant la manière de réaliser le système à développer ont été élucidées.

Le produit obtenu est un modèle graphique (ensemble de schémas et algorithmes) prêt à être codé. Dans le chapitre suivant nous allons étudier en détails les outils et les langages utilisés durant la phase de construction.

***CHAPITRE IV :***  
***IMPLEMENTATION***

### IV.1. Introduction :

Rappelons que le but de cette étude est d'implémenter une application d'aide à l'analyse sémantique de textes Arabe et extraire des relations entre les mots (verbe, sujet et objet). Dans cette section, nous allons présenter la plate de forme de développement et l'ensemble des ressources intégré que nous avons utilisé pour la réalisation de notre travail.

### IV.2. Environnement de travaille :

#### IV.2.1. Environnement de matériel :

Pour nous avons développé notre logiciel dans un environnement WINDOWS en utilisant un ordinateur dont les caractéristiques techniques, sont les suivantes :

- ✓ Toshiba
- ✓ Processeur Intel « I3 »
- ✓ RAM 4 Go
- ✓ Disque dur 500 Go

#### IV.2.2. Environnement logiciel :

L'outil utilisé pour la réalisation de ce travail est :

➤ *PyScripter* :

Est un logiciel libre Python et un Environnement de développement (EDI) sur Windows. A l'origine, il s'agissait d'un EDI léger conçu pour fournir une solution de script fiable pour les applications Delphi. Au fil du temps, il a évolué pour devenir un EDI Python autonome et complet. Il est construit en Delphi à l'aide de P4D et est extensible à l'aide de scripts Python. Le fait d'être construit dans un langage compilé le rend plus léger que d'autres EDI. Actuellement, il n'est disponible que pour le système d'exploitation Microsoft Windows. Il a été créé par Kiriakos Vlahos [20].

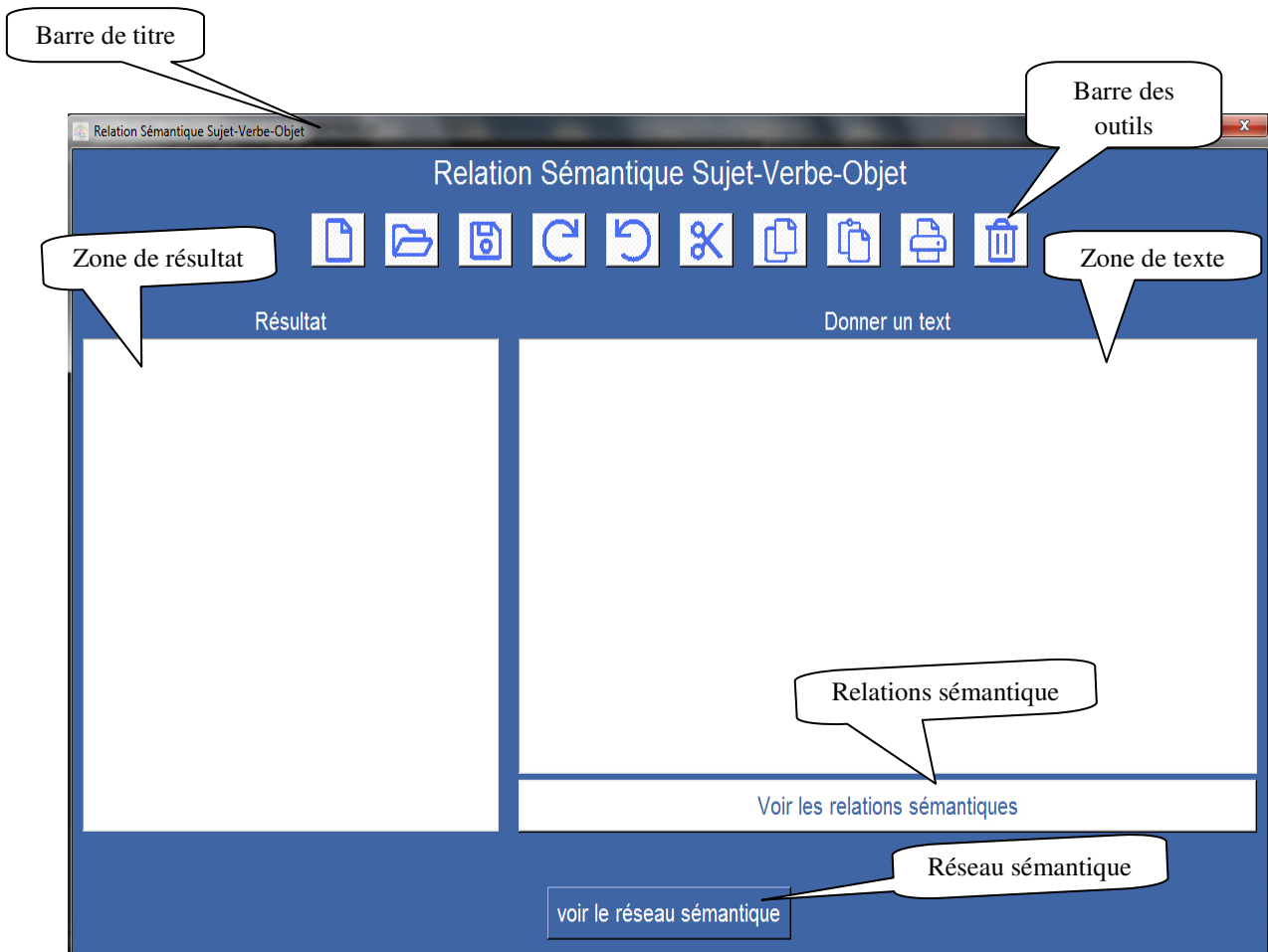
➤ **Langages utilisés :**

Python est un langage de programmation puissant et facile à apprendre. Il dispose de structures de données de haut niveau et permet une approche simple mais efficace de la programmation orientée objet. Parce que sa syntaxe est élégante, que son typage est dynamique et qu'il est interprété, Python est un langage idéal pour l'écriture de scripts et le développement rapide d'applications dans de nombreux domaines et sur la plupart des plateformes [21].

### IV.3. Présentation graphique :

Nous présentons dans cette section des capteurs d'écran, qui représentent notre Application

➤ **Relation Sémantique verbe-sujet-objet :**



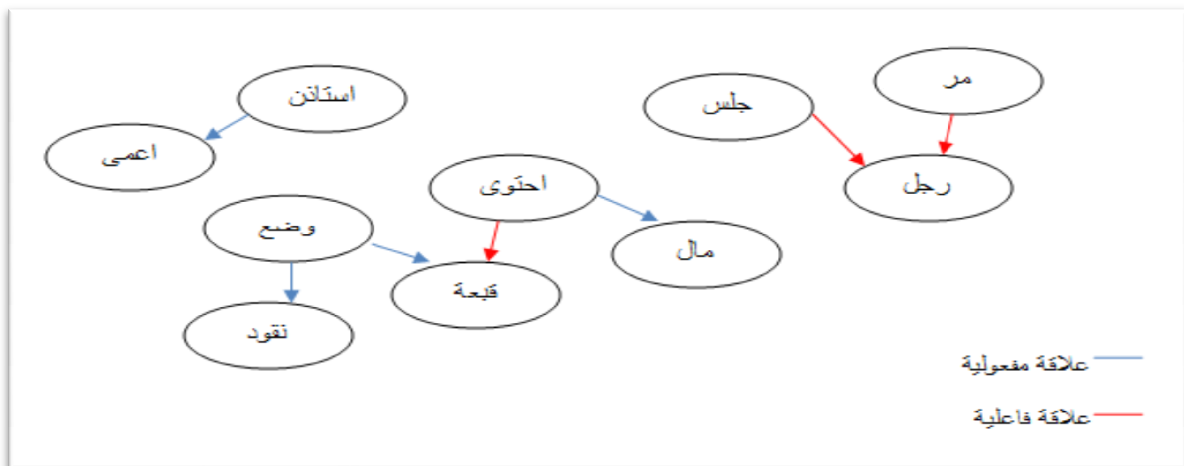
**Figure IV. 1.** Capture d'écran de l'interface graphique «relation sémantique verbe-sujet-objet»

➤ **Exemples sur la fonction de boutons «voir les relations sémantiques»:**



**Figure IV.2.** Résultat du bouton «voir les relations sémantiques»

➤ **Exemples sur la fonction de boutons «voir le réseaux sémantiques»:**



**Figure IV.3.** Fonction de boutons «voir le réseaux sémantiques»

### IV.4. Déroulement de l'approche :

On présente dans cette partie le déroulement de travail de notre système d'extraction :

#### IV.4.1. Les résultats :

Pour tester notre application nous avons utilisé 110 phrases :

شَرِبَ عَبَّاسٌ عَصِيرَ الْفَوَاكِه.
رَتَّبَتْ عَائِشَةُ الْخُرْفَةَ.
فَرَحَتْ زَيْنَبُ بِمَلَابِسِهَا الْجَدِيدَةِ.
يَفْرَحُ الْإِنْسَانُ بِمَنْظَرِ الْأَزْهَارِ.
سَافَرَ عَلِيٌّ بِالطَّائِرَةِ.
تَحَنُّنٌ تَجَلَّسُ فِي الْحَدِيقَةِ.
إِبْرَاهِيمُ يَكْتُبُ الرِّسَالَةَ.
أَتَسْعَرُ أَنْتَكَ مَرَهَقٌ جَدًّا يَا فَتَى ؟
تَبْدُو لِي كَذَلِكَ.
وَعَيْنَاكَ الضَّيْقَتَانِ تَزِيدَانِ مِنْ حَدَّتِكَ.
كَلَّمَا اكْتَمَلَتْ تِلْكَ الْعَقِيدَةُ الَّتِي تَحَلُّو وَجْهَكَ.
أَهْدَأُ رَجَاءً.
كَلَّمَا نَاولَكَ أَحَدَهُمْ نَظِيرَ عَمَلِكَ.
غَدًّا حِينَ سَتَكْبُرُ.
سَيَّةٌ مِنْ زُرْقَةٍ عَيْنِيهِ مَا لِي السَّوَادُ.
غَابَتْ ضِحْكُوتُهُ الْجَمِيلَةُ الْآسِرَةَ.
عَمِي: يَقُولُ أَبِي إِنَّكَ بَطْلٌ.
عَلِمْتَ أَنْتَكَ عَائِدٌ.
وَجَعَلْتُ الْجَمِيعَ يَرَى صُورَتَكَ بِالْبِرَّةِ الْحَسْرَةَ
جِيمِسَ لَمْ يَسْتَطِعْ أَنْ يَفْرَحَ بِكَلَامِ الصَّغِيرِ.
إِبْرَائِيْلُ يَقْصِدُ الضَّيْفُ.
بَابٌ مِنْ طَرَقَتْ ؟
كَمْ مِنْ دَارٍ يَحْتِ ؟
هَلَّا التَّدْخِيْنَ تَرَكْتَ.

Figure IV.4. Quelques exemples des phrases utilisés

Et le texte suivant pour le test :

جلس رجل أعمى على رصيف في أحد الشوارع، ووضع قبعته أمامه، وبجانبه لوحة مكتوب عليها: "أنا رجل أعمى، أرجوكم ساعدوني"، فمر رجل إعلانات بالشارع الذي يجلس فيه الأعمى، فوجد أن قبعته لا تحتوي سوى على القليل من المال، فوضع بعض النقود في القبعة، ثم -دون أن يستأذن الأعمى- أخذ اللوحة التي بجانبه وكتب عليها عبارة أخرى، ثم أعادها إلى مكانها وغادر. بدأ الأعمى يلاحظ أن قبعته امتلأت بالنقود، فعرف أن السبب هو ما فعله ذلك الرجل بلوحته، فسأل أحد المارة عما كتب على اللوحة، فكانت الآتي: "إننا في فصل الربيع، ولكنني لا أستطيع رؤية جماله!"

## Chapitre IV Implémentation

Résultat extrait		Résultat attendue		Phrases
Relation verbe – objet	Relation verbe-sujet	Relation verbe –objet	Relation verbe-sujet	
خلق, شجرة	خلق, الله	خلق, شجرة	خلق, الله	خَلَقَ اللهُ الْأَشْجَارَ.
بدا, صوم	بدا, عبد	بدا, صوم	بدا, عبد اللطيف	بَدَأَ عَبْدُ اللَّطِيفِ الصُّومَ.
شرب, ماء	شرب, خالد	شرب, ماء	شرب, خالد	شَرِبَ خَالِدُ الْمَاءَ.
استقبل, قبلة	استقبل, مصلي	استقبل, قبلة	استقبل, مصلي	اسْتَقْبَلَ الْمُصَلِّي الْقِبْلَةَ.
دخل, مطعم	دخل, مسافر	دخل, مطعم	دخل, مسافر	الْمُسَافِرُ يَدْخُلُ الْمَطْعَمَ.
أدى, زكاة	أدى, صائم	أدى, زكاة	أدى, صائم	أَدَى الصَّائِمُ زَكَاةَ الْفِطْرِ.
استخرج, جواز	استخرج, خالد	استخرج, جواز	استخرج, خالد	خَالِدٌ اسْتَخْرَجَ جَوَازَ السَّفَرِ.
فحص, مريض	فحص, طبيب	فحص, مريض	فحص, طبيب	تَفَحَّصَ الطَّبِيبُ الْمَرِيضَ.
جال, شارع	جال, زائر	جال, شارع	جال, زائر	الزَّائِرُ يَجُولُ فِي شَوَارِعِ الْمَدِينَةِ.
قضى, وقت	قضى, صديق	قضى, وقت	قضى, صديق	الْأَصْدِقَاءُ يَقْضُونَ وَقْتًا سَعِيدًا فِي الرَّحْلَةِ.
أحب, زهرة	أحب, أخ	أحب, زهرة	أحب, أخ	يُحِبُّ أَخِي الْأَزْهَارَ الْبَدِيعَةَ.
/	/	/	مال, شيء	شيء من زرقه عينيه مال إلى السواد.
زار, مدينة	زار, صلاح	زار, مدينة	زار, صلاح الدين	يَزُورُ صَلاَحَ الدِّينِ الْمَدِينَةَ لِأَوَّلِ مَرَّةٍ.
غفر, ذنب	غفر, الله	غفر, ذنب	غفر, الله	إِنَّمَا يَغْفِرُ الذُّنُوبَ اللهُ.
رتب, غرفة	رتب, عائشة	رتب, غرفة	رتب, عائشة	رَتَّبَتْ عَائِشَةُ الْعُرْفَةَ.
كتب, رسالة	كتب, ابراهيم	كتب, رسالة	كتب, ابراهيم	إِبْرَاهِيمُ يَكْتُبُ الرِّسَالَةَ.
ترك, تدخين	ترك, أنت	ترك, تدخين	ترك, أنت	هَلَا التَّدخينِ تَرَكْتَ.
ناول, نظير	ناول, أحد	ناول, نظير	ناول, أحد	كَلَّمَا نَاوَلَكَ أَحَدَهُمْ نَظِيرَ عَمَلِكَ.
قصد, إياي	قصد, ضيف	قصد, إياي	قصد, ضيف	إِيَّايَ يَقْصِدُ الضَّيْفُ.

**Tableau IV.1.** Quelques exemples de résultat des phrases utilisés.

### IV.4.1.1. Les erreurs :

erreurs	cause
استبدال همزة قطع بهمزة وصل: بدا, ادى, احب, اخ	Segmentation de texte avec Nltk « supprime (همزة قطع) . »
شيء من زرقه عينيه مال إلى السواد	Nltk ne fait pas la déférence entre le verbe (مال: انحراف الشيء) et le nom (مال: نفود).

**Tableau IV.2.** Les erreurs de programme.

## Chapitre IV Implémentation

Relation verbe_sujet			Relation verbe_objet		
Nbr Résultat attendue	Nbr Résultat extrait	Pourcentage	Nbr Résultat attendue	Nbr Résultat extrait	Pourcentage
110	100	90.90	110	98	89.09

**Tableau IV.3.** Pourcentage de résultat de « relation verbe\_sujet » et « relation verbe\_objet »

Dans le **tableau IV.1.** On trouve le pourcentage de résultats de la relation (verbe-sujet) et la relation (verbe-objet) après avoir fait un test sur notre programme afin d'évaluer notre travail on a eu les résultats suivants :

1/ Dans la relation verbe-sujet on a mis 110 phrases comme exemples dans notre programme, ce dernier a pût déterminer 100 phrases correctes et les phrases restantes n'a pas pût trouver de solution donc on est sûr un pourcentage de 90,90%.

2/ Dans la relation verbe-objet on a commencé le test avec 110 phrases aussi dont 98 notre programme les a résolu et a échoué dans 12 phrases sans, donc on est sûr un pourcentage de 89,09%.

- ❖ Dans les cas où nous somme confrontés à la confusion dans la détermination du sujet et l'objet « si la phrase contienne : *« ما، الا، انما، اما، وحرف الفاء و اي »* (Schéma 1) nous dépendons sur la base de données remplie manuellement au début puis nous la remplissons avec le programme.

Ex1 « *ما أفاد المريض إلا الدواء* » Ex2 « *ما أكل المريض إلا الدواء* » dans ces cas nous somme confrontés à la confusion dans la détermination du sujet et l'objet car le nom après le verbe de « *ما* » peut être l'objet (ex1) et peut être le sujet (ex2).

<b>Entrées :</b> phrase	
<b>Sortie :</b> continu de la phrase	
<b>Début</b>	SI ['الا', 'IN'], ['ما', 'WP'], ['انما', 'RP'], ['اما', 'RP'], ['ف', 'RP'], ['كم', 'WRB'], ['من', 'IN'], ['اي', 'QUANT'] dans la phrase
	ALORS recherché le type de relation SELON une BDD
<b>Fin</b>	
<b>Algorithme 3.</b> Pour rechercher type de relation selon une BDD	



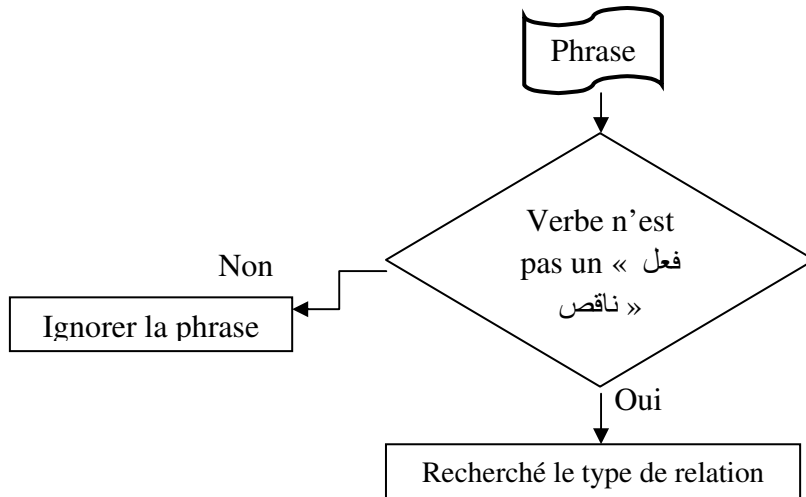


Schéma IV.1. Pour rechercher type de relation selon une BDD.

### IV.4.2. Evaluation et test :

L'évaluation d'un modèle et une évaluation afin d'assurer de la capacité de généralisation en dehors des données d'entraînement, cette évaluation permettrait de qualifier le comportement du modèle appris sur les données non utilisées lors de l'apprentissage, enfin sur les exemples d'entraînement ou autres exemples réservés pour le test. L'évaluation est liée aux différentes mesures comme la *précision*, *précision* et *f-mesure*.

#### IV.4.2.1. Précision :

Le rapport entre le nombre d'exemples correctement classés et le nombre total des exemples testés [22].

$$\text{Précision} = \frac{\text{nbr de documents correctes}}{\text{nbr documents correctes} + \text{nbr documents incorrectes}}$$

#### IV.4.2.2. Rappel :

La capacité d'un système à sélectionner tous les documents correctement classés de la collection [23].

$$\text{rappel} = \frac{\text{nbr de document correctement}}{\text{nbr total de document testés}}$$

### IV.4.2.3. *F*-mesure :

Une mesure qui combine la précision et le rappel est leur moyenne harmonique, nommée *F*-mesure ou *F*-score (**en**) :

$$f = 2. \frac{(\text{précision} * \text{rappel})}{(\text{précision} + \text{rappel})}$$

	Précision	Rappel	F-mesure
Les donnes d'entraînement	0.9	0.9	0.9
Les donnes de test	0.9	0.9	0.9

**Tableau IV.4.** Des résultats d'évaluations.

### IV.5. Conclusion :

L'objectif de ce chapitre est d'illustrer la dernière partie du projet qui est la phase de réalisation dans laquelle nous avons présenté et décrit notre application d'analyse. Quelques écrans de l'application ont été également présentés.

### **Conclusion générale et perspective :**

Dans ce mémoire, nous avons abordé le thème «Extraction des relations sémantiques entre les mots en arabes : Reconnaissance des relations verbe-sujet-objet ».

Nous avons étudiés l'analyse sémantique des textes arabe pour extraire les relations entre les mots, pour faire ça nous avons organisé notre travail selon trois étapes principales :

- D'abord, nous avons récolté et préparé toutes les données nécessaires : les questions et les réponses...
- Ensuite, nous avons développé une méthode pour la segmentation et la détermination des relations sémantiques existe entre les mots. Après la segmentation du texte en se basant sur les POSTag de chaque tokens, on simplifie, réduit et transforme chaque phrase du texte en un triplet : (Sujet, Verbe, Objet) supposés porteurs de l'information des phrases dont ils sont extraits. En fin, nous avons terminé par une représentation hiérarchique sous forme d'un réseau sémantique.

Nous considérons que nous avons réussi à réaliser une grande part des objectifs de ce projet, et que nous avons fait de bons choix concernant les outils d'implémentation, de ce fait notre travail constituera une très bonne piste pour d'autres projets avenir.

Des perspectives proposées ci-après sont adorables dans une continuation de recherche. Nous espérons :

- Augmenter la taille de la base de données concernant la relation verbe\_sujet et verbe\_objet.
- Développement du programme afin d'utilisée les textes accentuées.
- Amélioration des relations sémantiques pour schématiser tous les relations présentes dans le texte et non pas dans une seule phrase.
- Améliorer la décomposition du NLTK pour faire la différence entre le verbe et le nom dont ils ont les mêmes lettres

## Les références bibliographiques

---

- [1] : **Gouiaa M.** ; «Modélisation conceptuelle de conversations à partir de l'analyse de textes de dialogues transcrits: analyseur sémantique et générateur de la structure conversationnelle ».1998, Université Laval, Canada.
- [2] : **Elimam A.** ; «Introduction à la sémantique, cours de 1<sup>ère</sup> année médecine», 2018-2019, Université d'Oran.
- [3] : **Hafsaoui I., Boudjema F.**; «Analyse sémantique de textes arabes en se basant sur la structure rhétorique», 2016, Thèse de Master, Université Djilali Bounaâma, Algérie.
- [4] : **Tamine-Gardes J.** ; «Introduction à la lexicologie (suite) : les relations lexicales», 2015, L'Information Grammaticale, N. 16, 1983. pp. 38-40.
- [5] : Les types de relations sémantiques entre les mots, disponible sur : [https://souad-kassim-mohamed.blog4ever.com/semantique-lexicale-2#\\_ftn1](https://souad-kassim-mohamed.blog4ever.com/semantique-lexicale-2#_ftn1), 28-06-2019, 02 : 02 .
- [6] : **Ūrij Derenikovič A.** ; «Syntaxe et sémantique», 2018, Langages, 4<sup>e</sup> année, n°15, 1969. La linguistique en URSS. pp. 57-66.
- [7] : Le but de la sémantique, disponible sur : <http://post.queensu.ca/~lessardg/Cours/215/chap8.html>, 25-05-2019, 10 : 45 .
- [8] : **Descles J-P.** ; «Réseaux sémantiques : la nature logique et linguistique des relateurs», Langages, 22<sup>e</sup> année, n°87, 1987, Sémantique et intelligence artificielle, pp. 55-78.
- [9] **Anlambilim ve Edimbilim D.** ; «La linguistique, la sémantique et la pragmatique», 2011, Université de Ayse Eziler Kiran Hacettep, Turquie.
- [10] : **Doumi N.**; « Extraction de connaissances à partir du texte »,2016-2017, Thèse de Doctorat, Université de Sidi bel Abbes, Algérie.
- [11] : **Alhoti Sh.** ; «Formalisme tag pour l'arabe en vue d'applications didactiques et implantation web», 2010, Thèse de doctorat, Université de Franche-Comté, France.
- [12] : **Souid H.** ; «Les compléments comme déterminants sémantiques du verbe», 2013, Thèse de doctorat, Université de Paris III La Sorbonne Nouvelle, France.
- [13]: **Chaaben N., Hadrich Belguith L., Ben Hamadou A.** « The morph 2 new version: a robust morphological analyzer for Arabic texts», 2010, In actes des 10èmes journées internationales d'analyse statistique des données jadt'2010, Rome, Italy.

## Les références bibliographiques

---

- [14] **Debili F., Souissi E.** ; «Etiquetage grammatical de l'arabe voyelle ou non. In proceedings of the workshop on computational approaches to Semitic languages, Semitic », 1998, Stroudsburg, pa, USA.
- [15] **Bakhouche A., Tlili-Guiassa Y.** ; «Extraction des relations sémantiques à partir du Wiktionnaire Arabe», Laboratoire LRI/Equipe SRF, Université Badji Mokhtar, Annaba, Algérie.
- [16] **Mouelhi Z.** ; «AraSeg : un segmenteur semi-automatique des textes arabes In JADT 2008 : 9es Journées internationales d'Analyse statistique des Données Textuelles.
- [17] **El-Shishtawy T., El-Ghannam F.**; «An Accurate Arabic Root-Based Lemmatizer for Information Retrieval Purposes. In IJCSI» (International Journal of Computer Science Issues), pages 58-66, Vol. 9, Issue 1, No 3, January 2012
- [18]: Categorizing and Tagging Words, disponible sur le site: <https://www.nltk.org>, 26-10-2019, 19: 35.
- [19]: Alphabetical list of part-of-speech tags used in the Penn Treebank Project, disponible sur le site: [https://www.ling.upenn.edu/courses/Fall\\_2003/ling001/penn\\_treebank\\_pos.html](https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html), 28-10-2019, 14: 25.
- [20] : Python (langage) disponible sur : <http://www.python.org> consulter le 27-11-2019, 22 : 12.
- [21] : **Fuchs P. ; Poulain P.** ; «Cours de python », 2019, Université Paris Diderot-Paris 7, France.
- [22] : **Aid A.** ; «Classification, data mining», 2017, Université AOM Bouira, Algérie.
- [23] : **Boughanem M.** ; «Cour Evaluation des performances dans les SRI», 2008, France.

**Certaines bases de données que nous avons utilisées dans notre travail :**

Base de données de (أفعال ناقصة) :

كنت
كنا
كنت
كنت
كنتما
كنتم
كنتن
كان
كانت
كانا
كانتا
كانوا
كن
صرت
صرنا

**Tableau 1.**base de données de (أفعال ناقصة).

Base de données de (فعل متعدي) :

ظن
خال
زعم
حسب
حجا
جعل
عد
هب
علم
رأى
ألفى
وجد
درى
تعلم
جعل

**Tableau 2.** Base de données de (فعل متعدي).

## Base de données de radical d'un verbe :

حمد
ذهب
مكن
كتب
كلم
أكل
جلس
عمل
أخذ
طلع
عرف
كتب
فعل
جمع
نظر

Tableau 3. base de données de radical d'un verbe.

## Base de données (الفعل المعتل الناقص, الأجوف و المثال) :

الناقص	الأجوف	المثال
صلى	دام	وَأَد
نفي	بات	وَجَد
بقي	بان	وَحَد
حمى	دار	وَهُم
سعى	زار	يُنْس
نسي	صام	يُنْع

Tableau 4. Base de données (الفعل المعتل الناقص, الأجوف و المثال).

Base de données de (فعل لازم) :

لازم
نام
نجح
نشا
نضر
نظر
هبط
هدف
هرب
هلع
ورف
وقف
ولج
ولد
يسر

Tableau 5. Base de données de (فعل لازم).

Base de données de (الأسماء المؤنثة) :

مدن
قرى
نار
دار
شمس
صحراء
يد
عين
خود
فردوس
نفس
سكين
سماء
غيوم
أرض

Tableau 6. Base de données de (الأسماء المؤنثة).



## Résumé :

Une phrase simple a besoin d'un sujet, d'un verbe et parfois d'un objet. Afin de construire un arbre syntaxique il faut déterminer les verbes, les sujets et les objets. La compréhension de la phrase nécessite la reconnaissance de ces entités.

La sémantique est l'étude scientifique du sens, ou de la signification. Cette science nous invite à découvrir un monde dans lequel nous pratiquons notre langue quotidiennement ou nous exprimons et transmettons la compréhension de manière complète mais par fois incomplète car nous n'arrivons pas de donner de sens correcte à un énoncé et nous ne pouvons pas l'expliquer comme nous le désirons bien que l'énonciation et le produit de cet acte, l'énoncé, sont de première importance pour la sémantique.

L'objectif de ce travail, est d'implémenter un système d'extraction des relations verbes-sujets-objets à partir d'un corpus des textes arabes afin de créer un dictionnaire de relations sémantiques qui permet la réutilisation de telle ressources.

**Mot clés :** Sémantique, Relations verbes-sujets-objets, Arabe.

## Abstract :

A simple sentence needs a subject, a verb and sometimes an object. In order to build a syntax tree one must determine verbs, subjects and objects. The understanding of the sentence requires the recognition of these entities. Semantics is the scientific study of meaning. This science invites us to discover a world in which we practice our language daily or we express and transmit the comprehension in a complete but sometimes incomplete way because we do not arrive to give a correct meaning to a statement and we can't explain it as we desire even though this has a great importance for semantics. The objective of this work is to implement a system of extraction of verb-subject-object relationships from a corpus of Arabic texts to create a dictionary of semantic relations that allows the reuse of such resources.

**Key words:** Semantics, verb-subject-object relationships, Arab.

## ملخص:

تحتاج الجملة البسيطة إلى فعل, فاعل و مفعول به, ومن أجل تحديد البناء اللغوي للجملة يجب علينا تحديد الفعل, الفاعل والمفعول به. كما يتطلب فهم الجملة العربية معرفة هاته الكيانات.

الدلالات هي الدراسة العلمية للمعنى, تدعونا إلى اكتشاف عالم نمارس فيه لغتنا يوميا أو نعبر عن الفهم وننقله بطريقة صحيحة ولكن أحيانا غير مكتملة, لأننا لا نستطيع إعطاء معنى صحيح للجملة ولا نستطيع شرحها كما نريد, مع أن هذا له أهمية كبيرة في الدلالات.

الهدف من هذا العمل هو تطبيق نظام لاستخراج العلاقات الفاعلية (فعل+فاعل) والعلاقات المفعولية (فعل+مفعول به), من خلال مجموعة من النصوص العربية لإنشاء قاموس لهاته العلاقات يسمح بإعادة استخدامها من جديد.

**الكلمات المفتاحية :** الدلالات, علاقة فاعلية, علاقة مفعولية, اللغة العربية.