République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Djilali BOUNAÂMA Khemis Miliana

Faculté des Sciences et de la Technologie

Département de Mathématiques et d'Informatique

Mémoire Présenté

Pour l'obtention de diplôme de

Master en Informatique

Spécialité : Ingénierie du logiciel

Titre :

Analyse Sémantique de textes Arabes en se basant sur la structure rhétorique

Réalisé par : HAFSAOUI IMANE.
BOUDJEMA FATIHA.

Soutenu publiquement le : Juin /2016.

devant le jury composé de :

Président : Mr HANICHE.

Encadreur : Melle AICHA LABABO.

Examinateur1 : Mr BAHLOUL.

Examinateur2 : Melle BOUDALLI.

Année Universitaire 2015/2016

Remerciements

Notre cursus universitaire tend à sa fin, en cette heureuse occasion Nous remercions d'abord
Allah le tout puissant de nous avoir donné
La force, la sante et la volonté pour arriver à accomplir ce modeste travail.

Et puis nos parents qui ont été la cause et le support de notre vie,
Que dieu les protège.

Nous tenons à exprimer notre profonde gratitude et nos remerciements Et considération à notre promotrice Aicha Lababou Pour sa disponibilité à l'accomplissement de cette noble mission.
Notre parfaite estime, notre considération et notre profonde reconnaissance s'adressent à tous les enseignants du UKM,
Qui étaient très passions, sages et coopératives avec nous,
Durant toutes ces cinq années d'études
Pour leur précieux conseil.
Aussi, les membres du jury pour nous avoir honorés en acceptant de juger notre travaille.
Enfin toute personne qui a participé de près ou de loin à l'achèvement de ce travail, tous nos amies,
nos collègues, à toute la promo 2016
Merci de tout cœur.

Sommaire

Chapitre 02 : Théorie de la Structure Rhétorique

Chapitre 03 : Description du système et testes

Liste des abréviations

RST : Théorie de la Structure Rhétorique.

TALN : Traitement Automatique des Langues naturelles.

TALA : Traitement Automatique de la Langues Arabe.

Liste des tableaux

Liste des figures

الملخص

في الوقت الحاضر، المعالجة الالية للغات الطبيعية (TALN)، وبالتحديد اللغة العربي، هو موضوع الكثير من الابحاث فيما يتعلق بالنماذج النحوية الخاصة باللغة، ومن جهة اخرى، بتصميم وتنفيذ برامج الحاسوب في مختلف مجالات التطبيقات. حيث يساهم هذا البحث في هذه التطورات.

نظرية البنية البلاغية (RST)هو موضوع الدراسة منذ عقدين من الزمن، تستند هذه النظرية على التحليل البلاغي، أي الرؤية اللغوية لهيكلة النصوص، بغية الكشف عن العلاقات الدلالات اللفظية والعلاقات المعتمدة التي تربط بين الجمل. وقد تمت دراسة نظرية البنية البلاغية في عدة لغات. وكان استخدام هذه النظرية في اللغة العربية موضوع بعض الاعمال.

أثبتت نتائج جيدة في عدة تطبيقات كالبحث عن المعلومات، التلخيص وانشاء النصوص باستخدام RST. ويهدف هذا العمل تقديم أداة للتحليل البلاغي للنصوص العربية.

الكلمات المفتاحية: نظرية البنية البلاغيه، العلاقات البلاغية، العلامات/المؤشرات اللغوية، المعالجة الالية اللغات الطبيعية.

Résumé

A l'heure actuelle, le Traitement Automatique des Langues Naturelles (TALN) et plus précisément de la langue Arabe, fait l'objet de nombreux travaux en ce qui concerne d'une part, la modélisation linguistique propre à la langue, d'autre part, la conception et la réalisation de logiciels pour divers domaines d'applications. Notre recherche participe à ces développements.

La théorie de la structure rhétorique (RST) est un sujet de recherche depuis deux décennies. Cette théorie repose sur une analyse rhétorique, c'est-à-dire une vision linguistique de la structuration des textes, afin de détecter les relations sémantiques et les relations intentionnelles qui existent entre les phrases. La théorie de structure rhétorique a été étudiée sur plusieurs langues. L'emploi de cette théorie à la langue arabe a fait l'objet de quelques travaux.

Différentes applications telles que la recherche d'information, le résumé de texte et la génération de texte ont prouvé leurs bons résultats à l'aide de la RST. Ce travail vise à proposer un outil pour l'analyse rhétorique de texte Arabe.

Mots clés : Traitement Automatique des Langues Naturelles, La théorie de la structure rhétorique(RST), la langue Arabe. L'analyse rhétorique. Relations rhétoriques, le connecteur Arabe, Cohérence du texte

Abstract

At the present time, the *automatically* processing Natural language, and more specifically of the Arabic language, is the object of many work with regard to on the one hand, the linguistic modeling specific to the language, on the other hand, the design and implementation of software for various areas of applications. Our research is involved in these developments.

Rhetorical structure theory (RST) has been a subject of research since two decades. This theory bases on an analysis rhetoric, i.e. a linguistic vision of the structuring of the texts, in order to detect the semantic relations and the intentional relations, which exist between the sentences. Rhetorical structure theory has been studied on several languages. The employment of this theory on Arabic language has never been addressed.

Different applications have been proven to give good results when they use RST as a base. This work aims to propose a tool for the analysis rhetoric of Arabic text.

Key words : Rhetorical Structure Theory (RST), parsing Arabic texts, rhetorical relations, Arabic cue phrases, Arabic text coherence.

Introduction générale

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 de 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.

A ce jour, le traitement et l'exploitation de ces ressources documentaires présentent encore un défi pour les chercheurs dans le domaine du traitement automatique des langues naturelles. C'est ainsi que nous proposons d**ans le cadre de ce mémoire** un outil pour l'analyse sémantique d'un texte Arabe. Cette analyse se base sur une théorie qui a fait ces preuves dans le domaine du TAL, il s'agit de la théorie de la structure rhétorique (Rhetorical Structure Theory : RST, en anglais). La RST donne une vision linguistique de la structuration des textes afin de détecter les relations sémantiques et les relations intentionnelles qui existent entre les segments textuels.

Ce mémoire est structuré comme suit :

Pour situer notre étude dans la chaîne du traitement automatique d'une langue, nous passons en revue les différents niveaux d'analyse d'une langue naturelle dans **le chapitre 1**, où nous présentons aussi quelques caractéristiques de la langue arabe et nous abordons les différents problèmes du traitement automatique des textes arabes.

Le **chapitre** 2 sera consacré à décrire la théorie de la structure rhétorique de façon générale

et l'application de cette théorie à la langue Arabe, de façon particulière.

Dans le **chapitre 3** nous décrivons le système d'analyse de textes arabes que nous avons développé ainsi que des résultats que nous avons obtenus.

Enfin, nous terminons ce mémoire par une conclusion.

Chapitre 1

Le traitement automatique des langues naturelles

1. Le traitement automatique des langues naturelles

1.1 Introduction

Le traitement automatique des langues (TAL) ou encore Traitement automatique du langage naturel (TALN) a fait l'objet de développement important ces dernières années. Ce mémoire s'inscrit dans le domaine du TALN, tout particulièrement le traitement automatique de la langue Arabe (TALA). Il est donc naturel de commencer par présenter les concepts liés au contexte de notre étude. Ainsi, ce chapitre est organisé de la manière suivante. Dans la section 1.2 nous donnons des généralités sur le domaine du TALN. La section 1.3 est dédiée au TAL Arabe. Les particularités de la langue Arabe sont aussi abordées ainsi que les difficultés liées au traitement automatique de la langue Arabe. Enfin, nous terminons ce chapitre par une conclusion.

1.2 Traitement Automatique des Langues naturelles (TALN)

Le traitement automatique des langues (TAL) ou encore Traitement automatique du langage naturel (TALN) sont des termes encore méconnus du grand public alors qu'ils représentent des traitements informatiques dont les applications deviennent incontournables dans notre monde actuel.

En effet, qui n'a jamais entendu parler de programmes informatiques capables de comprendre le langage des humains et de fournir une réponse adaptée sans qu'aucune intervention extérieure ne soit nécessaire ? Même si les robots humanoïdes sont encore bien loin des mythes véhiculés par la science-fiction, il n'en demeure pas moins que des applications telles que les logiciels d'aides intelligents comme ceux que l'on peut trouver sur certains sites web sont eux déjà bien présents dans notre époque.

Les termes TAL et TALN représentent de manière générale les traitements informatiques capables de "traiter" de manière automatique le langage naturel.

Bien plus que de la simple reconnaissance de termes (ou mots clés), le TAL a pour objectif de "comprendre" le sens des phrases, les idées qui s'en dégagent et ce de manière à pouvoir "traiter" de la manière la plus optimale et la plus naturelle d'un point de vue humain ces phrases.

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. la figure 1.1 illustre les différents niveaux d'analyse d'un texte.

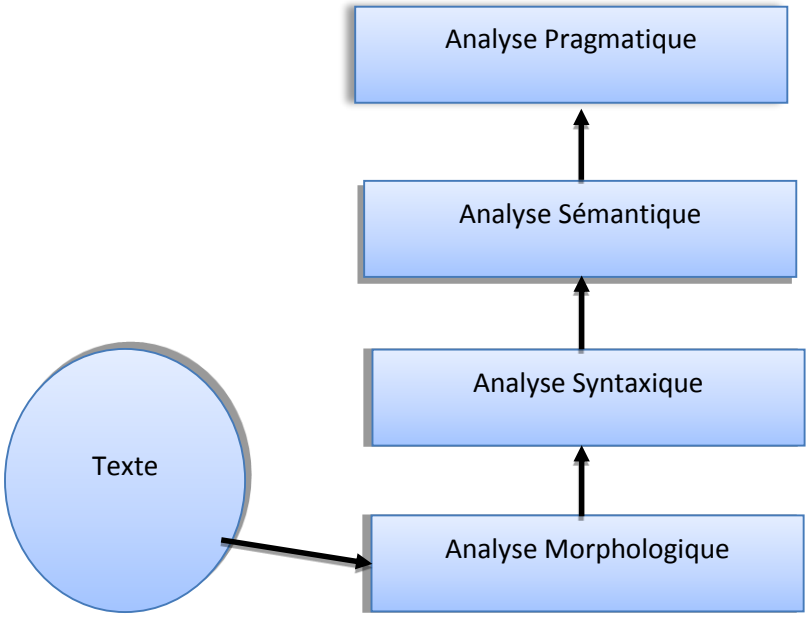

Figure 1.1 : Les différents niveaux d'analyse d'un texte.

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

1.2.1 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.… [1].

Ces variables sont à définir sur des bases linguistiques et en fonction des besoins des traitements envisagés.

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.

Cette démarche permet de séparer la grammaire de la lexicographie d'où la possibilité d'enrichir le dictionnaire au fur et à mesure de son utilisation.

Le dictionnaire, élément commun à la plupart des systèmes de traitement automatique des langues naturelles, conditionne dans une large mesure la stratégie et par conséquent la qualité du système. Il contient généralement le « vocabulaire » de la langue naturelle traitée, permettant ainsi d'identifier les différents éléments constituant un texte.

1.2.2 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.

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 (PH), le groupe nominal (GN), le groupe verbal (GV) etc.

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

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).

La figure 1.2 montre un exemple de représentation arborescente de la phrase «حفظ الولد القرآن».

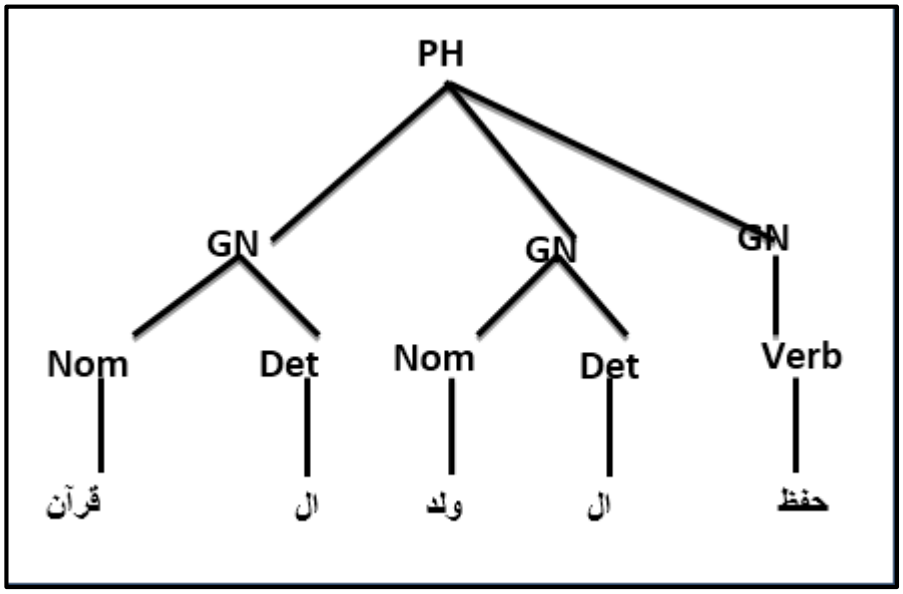

Figure 1.2: Arbre syntaxique

Deux techniques d'analyse sont possibles :

- l'analyse descendante consiste à partir de la racine (dans l'exemple précédent le nœud « PH ») et essayer toutes les dérivations pour aboutir à la phrase à analyser. Si après avoir essayé toutes les dérivations et on est toujours en situation de blocage, alors la phrase ne fait pas partie du langage engendré par la grammaire.
- L'analyse ascendante consiste à partir de la phrase à analyser et appliquer les règles de dérivation à l'envers pour remonter à la racine. Si après avoir essayé toutes les possibilités et que l'élément racine n'a pas été retrouvé, alors la phrase ne fait pas partie du langage engendré par la grammaire.

1.2.3 Analyse sémantique

« Dans un système de traitement automatique, l'analyse du sens des phrases consiste généralement à en extraire une représentation simplifiée, stylisée, de type logicomathématique, qui va permettre des calculs et raisonnements ultérieurs. » [2].

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. Parmi les représentations, on trouve la logique des propositions. A l'aide de connecteurs logiques (comme la conjonction « ∧ », la disjonction « ∨ », la négation « ⌉ » etc.), on peut former à partir des propositions de nouvelles propositions complexes.

Exemple :

proposition	Négation
لم يأكل آدم التفاحة	أكل آدم التفاحة

Figure 1.1: Nouvelle proposition construite à partir de la 1ère 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.

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

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.

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

1.2.4 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 réalisation opérationnelle, et qui ne concerne que des applications limites. On est donc encore loin de savoir construire des analyseurs pragmatiques pour le TALN.

1.3 Traitement Automatique de la Langue Arabe

L'arabe est parlé dans plus de 22 pays, du Maroc jusqu'à l'Iraq et dans toute la péninsule arabe [4]. C'est la première langue pour plus de 250 millions de personnes et la deuxième pour 40 millions. L'arabe, langue du Coran, est devenue la langue d'une civilisation et ne sert plus seulement à désigner les seuls habitants de la péninsule arabe qui la parlaient. On distingue l'arabe classique et l'arabe moderne.

L'arabe classique est la forme littéraire utilisée par tous pour les besoins de l'écriture et de l'imprimerie. C'est aussi la langue de la religion pour tous les musulmans, quelle que soit par ailleurs leur langue dialectale.

L'arabe moderne, dérivé de l'arabe classique, est la langue de la presse, des débats politiques, des textes scientifiques et de plus en plus celle des textes littéraires profanes. Parlé dans la plupart des pays arabes, l'arabe moderne n'est en revanche presque jamais la langue des échanges quotidiens.

Depuis le début des travaux en Traitement Automatique du Langage Naturelle sur la langue Arabe, plusieurs études ont poursuivi des directions de recherche diverses. On peut notamment distinguer les approches numériques s'appuyant sur des probabilités et les statistiques et des approches symboliques liées à la théorie des langages formels.

1.3.1 Particularités de la langue arabe

L'alphabet de la langue arabe compte 28 consonnes. L'arabe s'écrit et se lit de droite à gauche les lettres changent de forme de présentation selon leur position (au début, au milieu ou à la fin du mot). La figure montre les variations de la lettre ق(qaf). Toutes les lettres se lient entre elles sauf (ا ,و ,ر ,ز ,د ,ذ)qui ne se joignent pas à gauche.

Isolée	Initiale	Médiane	Finale
ق	قـ	ـقـ	ـق
	قرآن	القرآن	غسق

Figure 1.4 : différentes écritures de la lettre « qaf » en différentes positions dans le mot et comme lettre séparée. [1].

Toutefois, la langue Arabe présente quelques particularités dont nous citons :

a) Absence de voyelles

Un mot arabe s'écrit avec des consonnes et des voyelles. Les voyelles sont ajoutées au-dessus ou au-dessous des lettres (ـَ ,ـِ ,ـُ ,ـْ). Elles sont nécessaires à la lecture et à la compréhension correcte d'un texte, elles permettent de différencier des mots ayant la même représentation. Cependant, les voyelles ne sont utilisées que pour des textes sacrés (Coran, hadith) et didactiques. Les textes courants rencontrés dans les journaux et les livres n′en ne comportent habituellement pas .De plus, certaines lettres comme Alif «أ» peuvent symboliser le «آ» «إ» «أ» ou «ٱ» ; de même que pour les lettres « ى» et «ه» qui symbolisent respectivement «ي» et «ة» [11].

L'absence de voyelles (les sans voyelles) dans les textes arabes génère plusieurs cas d'ambiguïtés et des problèmes lors de l'analyse automatique. En effet, l'ambiguïté grammaticale augmente si le mot est sans voyelle. Cela est dû au fait qu'un mot non voyelle possède plusieurs voyelles possibles, et pour chaque voyelle est associée une liste différente de catégories grammaticales [5]. La figure donne un exemple pour les mots كتب et مدرسة.

Mot sans voyelles	1ère Interprétation		2ème Interprétation		3ème Interprétation	
كتب	كَتَّبَ	Il a écrit	كُتِبَ	Il a été écrit	كُتُبْ	Des livres
مدرسة	مَدْرَسَة	école	مُدَرِسَةَ	enseignante	مُدَرِسَةَ	enseignée

Figure 1.5 : ambiguïté causée par l'absence de voyelles pour les mots كتب et مدرسة [1].

[1] Fouad Soufiane Douzidia , Résumé automatique de texte arabe, Septembre, 2004

b) Catégories des mots

L'arabe considère 3 catégories de mots :

- **Le verbe** : entité exprimant un sens dépendant du temps, c'est un élément fondamental auquel se rattachent directement ou indirectement les divers mots qui constituent l'ensemble.
- **Le nom** : l'élément désignant un être ou un objet qui exprime un sens indépendant du temps.
- **Les particules** : entités qui servent à situer les événements et les objets par rapport au temps et l'espace, et permettent un enchaînement cohérent du texte.

c) Structure d'un mot

En arabe un mot peut signifier toute une phrase grâce à sa structure composée qui est une agglutination d'éléments de la grammaire, la représentation suivante (figure 1.6) schématise une structure possible d'un mot.

Post fixe	Suffixe	Corps schématique	Préfixe	Antéfixe

Figure 1.6 : structure d'un mot en langue Arabe[2].

Notons que la lecture et l'écriture d'un mot se fait de droite vers la gauche.

- Antéfixes sont des prépositions ou des conjonctions.
- Préfixes et suffixes expriment les traits grammaticaux et indiquent les fonctions : cas du nom, mode du verbe et les modalités (nombre, genre, personne,…)
- Post fixes sont des pronoms personnels.

Exemple 1 :

le mot arabe :أَتَتَذَكَّرُونَنَا

Ce mot exprime la phrase en français : "Est ce que vous vous souvenez de nous ?″

La segmentation de ce mot donne les constituants suivants :

[2] Fouad Soufiane Douzidia , Résumé automatique de texte arabe, Septembre, 2004

نَا	وذَ	تَذكّر	تُ	أ
Post fixe	**Suffixe**	**Corps schématique**	**Préfixe**	**Antéfixe**
pronom suffixe complément du nom.	suffixe verbal exprimant le pluriel.	dérivé de la racine : ذكر selon le schème $taR_1aR_2aR_3a$	préfixe verbal du temps de l'inaccompli.	conjonction d'interrogation

d) Irrégularité de l'ordre des mots dans la phrase :

L'ordre des mots en arabe est relativement libre. D'une manière générale, on met au début de la phrase le mot sur lequel on veut attirer l'attention et l'on termine sur le terme le plus long ou le plus riche en sens ou en sonorité. Cet ordre provoque des ambiguïtés syntaxiques artificielles, dans la mesure où il faut prévoir dans la grammaire toutes les règles de combinaisons possibles d'inversion de l'ordre des mots dans la phrase [6].

La figure 1.7 donne un exemple de combinaisons possibles d'inversion de l'ordre des mots dans la phrase [6].

Verbe + sujet + complément	فعل + فاعل + متمم	Est allé le garçon à l'école	ذهب الولدُ إلَى المدرسة
Sujet + verbe + complément	فاعل + فعل + متمم	Le garçon est allé à l'école	الولدُ ذهب إلَى المدرسة
Complément + verbe + sujet	متمم + فعل + فاعل	A l'école est allé le garçon	إلَى المدرسة ذهب الولدُ

Figure 1.7 : Exemple de combinaisons possibles d'inversion de l'ordre des mots dans une phrase[3].

e) Absence de ponctuation régulière :

La langue arabe n'est pas basée principalement sur les signes de ponctuations et les marqueurs typographiques ; il est à noter que ces derniers ne sont pas utilisés de façon régulière dans les textes arabes actuels, et même dans le cas où ils y figurent, ils ne sont pas gérés par des règles précises d'utilisation. [7]. Par ailleurs, nous pouvons trouver tout un paragraphe arabe ne contenant aucun signe de ponctuation à part un point à la fin de ce paragraphe. Il convient de

[3] Fouad Soufiane Douzidia , Résumé automatique de texte arabe, Septembre, 2004

noter que la présence des signes de ponctuation ne peut pas guider la segmentation comme c'est le cas pour d'autres langues latines, telles que le français ou l'anglais. Ainsi, la segmentation de textes arabes doit être guidée non seulement par les signes de ponctuations et les marqueurs typographiques mais aussi par des particules et certains mots tels que les conjonctions de coordination, etc. [7].

f) Problèmes d'agglutination de mots :

Contrairement à la plupart des langues latines, en arabe, les articles[4], les prépositions[5], les pronoms[6], etc. se collent aux adjectifs, noms, verbes et particules auxquels ils se rapportent3. Comparé au français, un mot arabe peut parfois correspondre à toute une phrase. Par exemple, le mot arabe **[أتتذكروننا]** correspond en français à la phrase **[Est-ce que vous vous souvenez de nous ?]**. Cette caractéristique(Agglutination) peut engendrer une ambiguïté au niveau morphologique. En effet, il est parfois difficile de distinguer entre une proclitique[7] ou enclitique[8] et un caractère original du mot.

Exemple 2 : le caractère "و" dans le mot "il est arrivé | وصل" est un caractère original alors que dans le mot "et il a ouvert | وفتح", il s'agit plutôt d'une proclitique [8].

g) Problèmes d'interprétation syntaxique :

Les problèmes recensés dans les paragraphes précédents, touchant la voyellation, la segmentation, l'agglutination, etc., peuvent mener à des interprétations syntaxiques distinctes d'une même phrase ou d'un même extrait de texte. Considérons l'exemple de la phrase :

Cette phrase peut être interprétée et traduite selon les trois interprétations suivantes qui sont toutes syntaxiquement correctes :

ولد العالم والباحث في مصر

- **وُلِدَ العالم والباحث في مصر** → Le savant chercheur est né en Egypte.

[4] Les articles : par exemple " ال".
[5] Les prépositions sont : في, مع, لن, عن, على, حتى, من, الى, لِ, ك, بـ
[6] Le pronom personnel en arabe est isolé ou affixé. Isolé, il correspond en français à : moi, toi, etc.
[7] Les proclitiques représentent des conjonctions mono-consonnes (و, ل) des prépositions (لِ, بـ) un préverbe(س) indiquant le futur, un article) (ال qui permet la détermination d'un nom, etc.
[8] Les enclitiques sont les compléments de pronomكما. ك. هم

- **وَلَدُ العالم والباحث في مصر**→ Le fils du savant chercheur est en Egypte.
- **وَلَدَ العالم والباحث في مصر**→ *Le savant chercheur a engendré en Egypte*).

1.3.2 Difficultés de l'analyse automatique de l'Arabe

a) Segmentation des textes arabes

La grande majorité des applications de traitements du langage naturel présuppose le découpage de textes en phrases. Cette tâche est depuis très longtemps automatisée, on parle alors de reconnaissance de frontières des phrases. La phrase est considérée comme l'unité centrale des processus du traitement du langage naturel, comme par exemple l'étiquetage. 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 précédent ou suivent ces signes.

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. 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.

De plus, la langue arabe n'est pas appuyée principalement sur les signes de ponctuations et les marqueurs typographiques (comme c'est le cas des langues latines) ; ainsi nous pouvons trouver tout un paragraphe ne contenant aucun signe de ponctuation à part un point à la fin.

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 وpeut 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 وهمqui peut se lire wahmon (imagination) ou bien wa hom (et + ils). 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 فهمpeut 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.

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 : اذا ,لكن ,بينما).

b) 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. À proprement parler, l'analyse morphologique ne fait que la séparation et l'identification des morphèmes semblables aux mots préfixés (comme les conjonctions "wa | و" et "fa | ف", etc.) des prépositions préfixées (comme "bi | ب" et "li | ل", l'article défini"ال", etc.), des suffixes de pronom possessif [9].

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. 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 (voir exemple 1, § 1.3.1) [10].

Cette caractéristique engendre une ambiguïté morphologique au cours de l'analyse. Ainsi, la reconnaissance des unités lexicales qui composent une unité morphologique n'est pas toujours facile à détecter. Le problème est de reconnaître que la bonne segmentation réside ainsi, dans la difficulté de distinction entre un proclitique ou enclitique et un caractère original du mot L'absence de voyelles pose un autre problème important. 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 |ٰفصَلَ", 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).

c) 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.

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

1.4 Conclusion

Nous avons cherché dans ce chapitre à présenter de façon globale le domaine de traitement automatique de langue Natural et précisément le traitement automatique de la langue Arabe et sa particularité par rapport aux autres langues. Dans le prochain chapitre intitulé «**La Théorie de la Structure Rhétorique**», il sera consacré à décrire la théorie de la structure rhétorique de façon générale et l'application de cette théorie à la langue Arabe, de façon particulière.

Chapitre II

La Théorie de la Structure Rhétorique

2. La Théorie de la Structure Rhétorique

2.1 Introduction

La façon de comprendre et d'analyser un texte dépend fortement du domaine qu'il traite, mais aussi de son type ; on distingue essentiellement des textes descriptifs, des textes argumentatifs et des textes narratifs. Jusqu'aux années quatre-vingt, la plupart des travaux existants, tant en intelligence artificielle qu'en linguistique ou en psycholinguistique, se sont limités aux récits.

La première idée utilisée pour mettre en évidence la structure d'un texte a consisté à tenter de décrire des structures globales de textes et à déterminer comment le texte précis analysé cadre avec une de ces structures préétablies. Ce qui a abouti à des théories décrivant les structures rhétorique d'un texte dont la théorie de la structure rhétorique (Rhetorical Structure Theory) (MANN et THOMPSON (1986) et MANN et THOMPSON (1987)), à laquelle nous nous intéressons surtout à la structure de texte, à l'analyse rhétorique.

Ce chapitre est consacré à la description de la RST. Nous commençons dans la section 1.2 par donner le principe de la RST. Nous décrirons après, section 1.3, les relations rhétoriques. La section 1.4 est dédiée à la description de la représentation graphique d'un texte lors d'une analyse RST, les arbres rhétoriques. Enfin, nous terminerons ce chapitre par une conclusion.

2.2 Théorie de Structure Rhétorique

La théorie des structures rhétoriques est l'une des théories qui a été appliqué 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 génération de langage naturel : William Mann, Christian Matthiessen et Sandra Thompson, avec l'apport de Cecilia Ford, Barbara Fox et Peter frites.

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.

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. Dans ce qui suit, nous définissons les notions clés de cette théorie.

2.2.1 Les critères de la RST

Comme nous l'avons dit, la RST étudie les relations entre différentes segments 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 segments, on doit définir pour ce texte une unité de subdivision minimales ainsi que ce qui permet de lier ces différentes subdivisions entre elles.

- ***Unité de texte*** : est un fragment de texte soudé de taille quelconque qui peut être lui-même subdivisé en segments de texte plus petits. Un segment peut donc avoir une taille supérieure à la phrase, ou bien beaucoup plus petite que la phrase, mais doit garder une « intégrité fonctionnelle indépendante », c'est-à-dire qu'il ne peut s'agir d'une partie de phrase prise au hasard et n'ayant pas de sens prise seule.
- ***Les relations rhétoriques*** sont des constructions abstraites qui relient deux segments de texte entre eux « ces deux segments ne devant pas se chevaucher » et déterminent les relations sémantiques et pragmatiques qui tiennent entre eux. Les segments de texte pouvant être de taille quelconque, une relation rhétorique pourra relier deux parties minimales d'un texte (inférieures à la phrase) aussi bien qu'un ensemble de phrases.

2.2.2 La nucléarité

Mann et Thompson ont défini la notion de nucléarité en désignant par *noyau* le segment important et par *satellite* le segment secondaire.

Noyau : Le noyau est celui des éléments que relie une relation qui est le plus indispensable à la compréhension du texte. Si on supprime le noyau, alors le texte devient incompréhensible. Prenons une phrase complète « comportant un noyau et un satellite » et privons-la de son noyau :

P 2.1 [Le train va bientôt partir] **N**, [bien qu'il soit déjà en retard] **S**.

P 2.2 [bien qu'il soit déjà en retard] **S**.

Privé de son noyau, l'exemple ne veut plus rien dire, ce qui montre bien l'importance du noyau pour la compréhension du texte.

Satellite : Le satellite apporte une information supplémentaire, il vient compléter le noyau. Il est donc suppressible, sa suppression n'empêchant pas la compréhension du texte. En reprenant l'exemple de la section précédente et en éliminant le satellite, on s'aperçoit que la compréhension n'est pas bloquée, et on voit bien que le satellite ne faisait qu'ajouter une information supplémentaire au noyau.

P 2.3 [Le train va bientôt partir] **N**

2.2.3 Relations rhétoriques

La RST se concentre à faire la reconnaissance des relations qui relient le segment minimal avec celui composé : la reconnaissance repose sur une interprétation. Cette interprétation se base sur une analyse syntaxique du contenu des segments du texte afin de repérer les différents marqueurs lexicaux, aussi appelés indicateurs ou connecteurs de relations rhétoriques, c'est-à-dire décrire les relations qui lient deux parties du texte dont l'un est noyau et l'autre satellite. Ainsi, les relations peuvent être repérées via des indicateurs linguistiques.

D'autre part, la plupart des relations rhétoriques sont binaires et relient un noyau à un satellite : ce sont des ***relations mononucléaires***. En revanche, certaines relations, sont dites ***multi-nucléaires*** du fait qu'elles font l'union entre plusieurs noyaux car aucun segment de discours ne paraît plus important que l'autre.

P 2.4 [Le cambrioleur a fracturé la serrure]**N**, [s'est introduit dans la maison] **N** [et s'est emparé de son butin] **N**.

2.2.4 Connecteurs (Cues phrases)

Un connecteur ou indicateur (cue phrases en anglais) est un terme permettant de relier deux ou plusieurs segments (phrase, énoncé, proposition...) d'un discours, parmi lesquels on compte par exemple *parce que*, *donc*, *puis*, *etc.* De cette façon, un connecteur contribue à constituer des unités

plus complexes à partir de ces unités simples que sont les segments en marquant des relations structurales et sémantiques entre les différentes entités du discours.

On considère que les connecteurs participaient à la cohérence du texte en reliant entre eux différents segments du texte. Or les connecteurs sont des indices linguistiques lexicaux qui permettent, en participant à la connexité du texte, d'en retrouver la structure assez facilement. C'est pourquoi ils sont très utiles pour faire une analyse textuelle avec la RST : quand un connecteur est présent, il devient beaucoup plus aisé de trouver quelle relation rhétorique tient entre les deux segments de discours qu'il relie que s'il n'était pas présent. En effet, un connecteur a rarement la possibilité de signaler plus d'une relation rhétorique alors que quand on n'a pas de connecteurs entre deux propositions, les interprétations sont plus nombreuses. Les connecteurs sont donc des indices linguistiques de la présence d'une relation rhétorique, et en cela ils sont utiles pour retrouver la structure d'un texte.

2.2.5 Représentation graphique : les arbres rhétoriques RST :

La structure rhétorique d'un texte se présente toujours sous la forme d'un arbre, les segments de texte élémentaires apparaissant au niveau de ses feuilles et les segments plus importants au niveau de ses nœuds. Les schémas sont l'unité de base permettant de construire un arbre rhétorique. Les schémas sont des structures abstraites composées de :

- un petit nombre de segments de discours.
- une relation qui lie entre eux ces segments.
- la relation qu'entretient le noyau avec les autres segments.

Ils permettent de représenter graphiquement une relation qui lie deux ou plusieurs segments, et pour ce faire, on trouve dans un schéma les éléments graphiques suivants :

- les ***lignes horizontales*** représentent les segments de discours.
- une ***ligne verticale*** (ou ***diagonale***) permet de repérer le noyau (un segment sans ligne verticale étant un satellite).
- les ***arcs*** qui joignent deux lignes horizontales sont les relations qui tiennent entre deux segments, la flèche indiquant le sens de la relation.

Il existe 5 types de schémas, chacun correspondant à une configuration particulière entre segments de discours selon une relation mononucléaire ou une relation multi-nucléaire.

Figure 2.1: Relations mononucléaires [9].

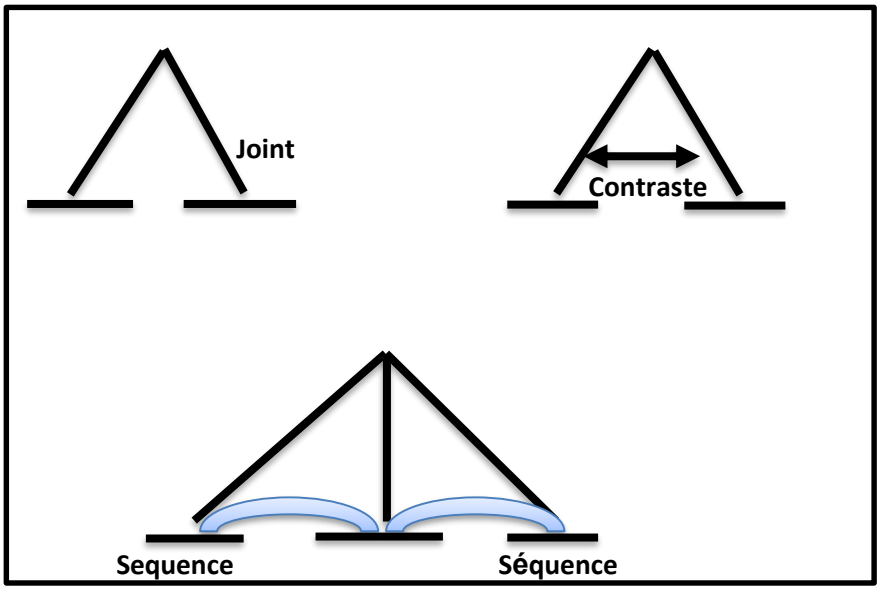

Figure 2.2: Relations multi nucléaires[10].

2.2.6 Relations de base de la RST (Mann et Thompson, 1987)

Mann et Thompson n'ont volontairement pas défini un nombre fixe de relations, mais ils ont simplement proposé un ensemble de relations de base[11] qui permet selon eux d'analyser la plupart des textes anglais. Les définitions de ces relations peuvent être modifiées et leur nombre peut varier selon les personnes qui utilisent la théorie. On peut en effet rajouter ou modifier des relations pour suivre les besoins d'une analyse : on peut imaginer le cas où aucune des relations de base ne convient pour qualifier une relation particulière rencontrée dans un texte.

[9]: http://www.sfu.ca/rst/07french/introduction.html.
[10] http://www.sfu.ca/rst/07french/introduction.html.
[11] Dans MANN et THOMPSON (1987), ils proposent 21 relations de base.

Mann et Thompson laissent donc l'ensemble des relations ouvert, et c'est cette extensibilité du nombre de relations qui permet une analyse beaucoup plus fine d'un texte que si l'ensemble de base était fermé. De cette façon on peut créer une nouvelle relation rhétorique lorsqu'on trouve un lien sémantique entre deux segments d'un texte auquel les relations existantes ne peuvent s'appliquer.

Le tableau 2.1 donne les relations de base définit dans (Mann et Thompson, 1987).

N°	**Nom de relations**	**Lieu d'effet**
1.	Élaboration	Noyau +Satellite
2.	Circonstance	Noyau +Satellite
3.	résolution	Noyau +Satellite
4.	Cause volontaire	Noyau +Satellite
5.	Résultat volontaire	Noyau +Satellite
6.	Cause non volontaire	Noyau +Satellite
7.	Résultat non volontaire	Noyau +Satellite
8.	But	Noyau +Satellite
9.	Condition	Noyau +Satellite
10.	Autrement	Noyau +Satellite
11.	Interprétation	Noyau +Satellite
12.	Évaluation	Noyau +Satellite
13.	Ajustement	Noyau +Satellite
14.	Résumé	Noyau +Satellite
15.	Preuve	Noyau
16.	Antithèse	Noyau
17.	Concession	Noyau
18.	Motivation	Noyau
19.	Facilitation	Noyau
20.	Justification	Noyau
21.	Arrière-plan	Noyau

Tableau 2.1: liste des relations rhétoriques de base [12].

2.2.7 L'analyse rhétorique d'un texte

Description détaillée de la méthode :

Si on vient à reconnaitre le noyau par rapport au satellite, ça veut dire qu'on a bien avancé dans la sémantique. Dans ce qui suit nous allons décrire les trois grandes étapes :

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 statut de noyau et l'autre celui de satellite.

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

La deuxième étape de la segmentation est la segmentation des phrases en unités minimales qui sont relie entre elles par des connecteurs (Cues phrases). Ces connecteurs sont utilisé pour premièrement relie deux unité minimale successive dans le texte, et deuxièmes pour déterminer les relations existe entre ces deux unités.

Toutefois, ces connecteurs peuvent être répertoriés en deux types : déclencheurs et complémentaires. Les connecteurs déclencheurs énoncent des concepts importants. Les connecteurs complémentaires sont recherchés dans un espace défini à partir de connecteur (dans le voisinage du connecteur). Ils peuvent ainsi agir dans le contexte afin de confirmer ou d'infirmer la relation rhétorique énoncée par le connecteur déclencheur.

2. **Application des règles rhétoriques :**

Les règles sont des faits rhétoriques qui ont comme objectif la reconnaissance d'une relation rhétorique. Chaque règles est formé par le nom de la relation qu'il indique, son connecteur déclencheur, sa liste de connecteur complémentaires de validation qui peuvent être rencontrés au voisinage de l'indicateur, la position du connecteur déclencheur indique qui peut être au début, au milieu de l'unité de texte et enfin l'unité(s) minimale(s) retenu(es).

L'application des règles rhétoriques à un double but : déterminer la nature des segments (noyau ou satellite) et détecter les relations rhétoriques entre ces segments.

2.1 Détermination du segment Noyau et Satellite

Cette étape consiste à repérer les connecteurs principaux dans les phrases déjà segmentées et à préciser leurs positions dans l'unité minimale afin d'appliquer les règles rhétoriques, en cherchant les connecteurs complémentaires**.**

Dans cette étape, nous allons donner pour chaque unité minimale un statut qui indique l'importance de cette unité par rapport à la phrase ou pour lui donner plus d'importance par rapport à une autre unité minimale. Le statut peut être un noyau ou un satellite.

2.2. Détection des relations rhétoriques

Cette étape consiste à chercher les indices complémentaires de validation au voisinage de l'indicateur principal, c'est-à-dire le segment qui contient l'indicateur principal et le segment qui le précède. C'est l'indicateur principal qui signale la relation rhétorique entre ces deux segments et c'est le rôle des indices complémentaires de confirmer ou non cette relation et de valider aussi le statut des deux segments.

Toutefois, la détection d'une relation se fait généralement par l'application d'un ensemble de contraintes sur le noyau, le satellite et sur la combinaison du noyau et du satellite.

3. Construction de l'arbre RST

Une fois l'étape de détection du type des unités minimales et des différentes relations rhétoriques existantes est achevée, donc le RS-arbre pourrait être construit. Le processus de construction de la RS-arabe est caractérisé par les propriétés suivant :

- Une arborescence de texte est un arbre binaire dont les feuilles indiquent des unités textuelles élémentaires.
- Chaque nœud est associé à *un statut* (noyau ou satellite), *un type* (le nom de relation rhétorique lier deux segments de texte adjacents entre eux), et un ensemble de saillance ou promotion (l'ensemble des unités qui constituent la partie la plus importante du texte).

Exemple : Pour illustrer la façon dont se présente un arbre rhétorique, nous présentons l'analyse du texte ci-dessous[12], illustré par la figure 2.3.

[12] Cette exemple a été extrait du site : http://www.sfu.ca/rst/07french/introduction.html. un titre et un résumé extraits du début d'un article du *Scientific American*.

[Le lactose et la lactase] :

[Le lactose est le sucre du lait] **s2**, [que l'enzyme lactase dégrade] **s3**. [Par absence de lactase, beaucoup d'adultes ne peuvent digérer le lait] **s4**. [Dans les populations consommant du lait, les adultes disposent de plus de lactase, peut-être grâce à la sélection naturelle] **s5**.

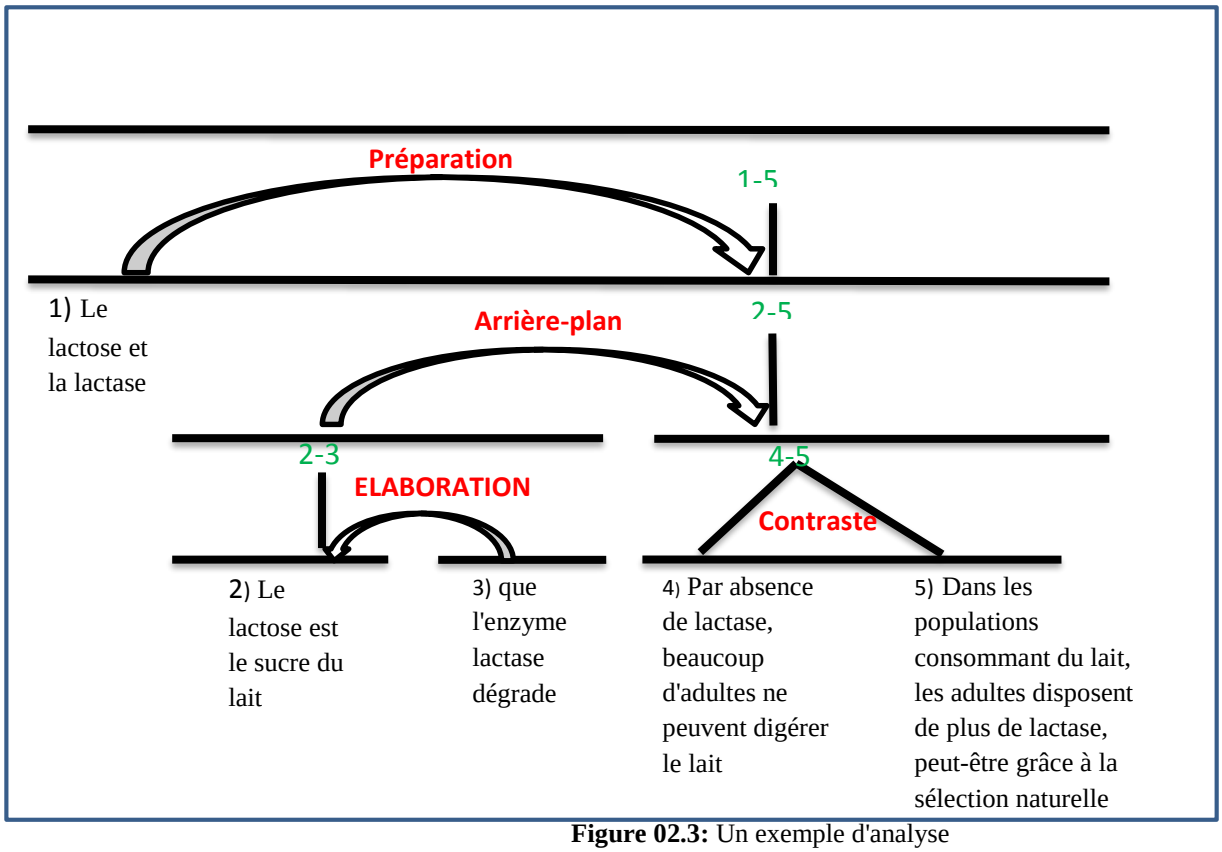

Figure 02.3: Un exemple d'analyse

Nous allons expliquer brièvement comment se fait l'analyse en reprenant le texte de la figure 2.3 ci-dessous :

Nous prenons le texte dans son entier et nous le séparer en cinq parties. Les deux premières unités (celles qui expliquent les termes lactose et lactase) ont pour but de faciliter la compréhension du reste du texte. Également, que l'unité (3) donne un détail supplémentaire sur le contenu ou, plus généralement, la situation présenté dans l'unité (2). Enfin, que les unités (4) et (5) sont en relation neutre de contraste. Chacune de ces observations est formulée en termes d'intention de l'auteur : ainsi, par exemple, l'intention de l'auteur est que le lecteur reconnaisse que les situations exprimées dans (4) et (5), identiques à plus d'un titre, diffèrent d'une manière que l'auteur souhaitait précisément formuler.

On continue de la même façon jusqu'à ce qu'on n'ait plus que des segments élémentaires, ne pouvant plus être décomposés.

2.3 Analyse rhétorique en langue Arabe

Dans cette partie nous présentons l'application de la RST à la langue Arabe. Nous présentons notamment les relations rhétoriques pour la langue arabe, ainsi que l'analyse d'un texte arabe.

Il faut noter que la RST a été appliqué avec succès dans le traitement automatique des textes indo-européens (l'anglais, le français, etc.). Par contre, peu de travaux sont consacrés à l'application de cette théorie aux applications ANLP.

En effet, et tels que nous l'avons présenté dans le chapitre 1, vue ses propriétés morphologiques et syntaxiques, la langue arabe est considérée comme une langue difficile à maîtriser dans le domaine du TAL. Les travaux sur le traitement automatique de l'arabe ont toujours fait face à des problèmes variés (que nous résumons dans la section 1.3.2 du chapitre 1) de la langue arabe issus de l'absence de voyelles, agglutination, irrégularité de l'ordre des mots dans la phrase, absence de ponctuation régulière, etc. Ces problèmes, rendent le traitement automatique de la langue arabe une tâche difficile.

2.3.1 Relations Rhétoriques pour la langue Arabe

Nous avons vu dans les sections précédentes que la phase la plus importante de l'analyse d'un texte par la RST est l'identification des relations rhétorique à retenir sur la base d'un ensemble de d'indicateurs (Cues phrases).

La plupart des travaux étaient consacrés à la langue anglaise. Du fait que l'analyse a été faite sur des corpus anglais, les relations rhétoriques qui ont été identifiées peuvent servir dans l'analyse et la compréhension de textes anglais . Mais il n'y a aucune garantie que le même ensemble de relations peut être appliqué directement à d'autres langues, telle que l'Arabe. En raison des différences entre la langue arabe et anglaise, nous pouvons dire avec un certain degré de certitude, que les relations rhétoriques anglaises ne peuvent pas être utilisées dans leurs formes actuelles pour le texte arabe.

Nous citons dans la suite de cette section les travaux de Al Sanie et al,. [13, 14] Qui sont les premiers à s'intéresser à adapter la RST à la langue Arabe.

Pour identifier les relations rhétoriques arabes Al Sanie et al,. Se sont basés sur l'approche suivante :

- Etude des relations rhétorique de la langue Anglaise.
- Etude d'un corpus arabe pour identifier de nouvelles relations selon la rhétorique Arabe.
- Etude des connecteurs Arabe (Cues phrases) pour déterminer les relations qui les lient.

Etude des relations rhétorique de la langue Anglaise : Al Sanie et al,. ont suivi l'organigramme de la figure 2.4 pour déterminer les relations rhétorique pour l'Arabe à partir de ceux de l'Anglais.

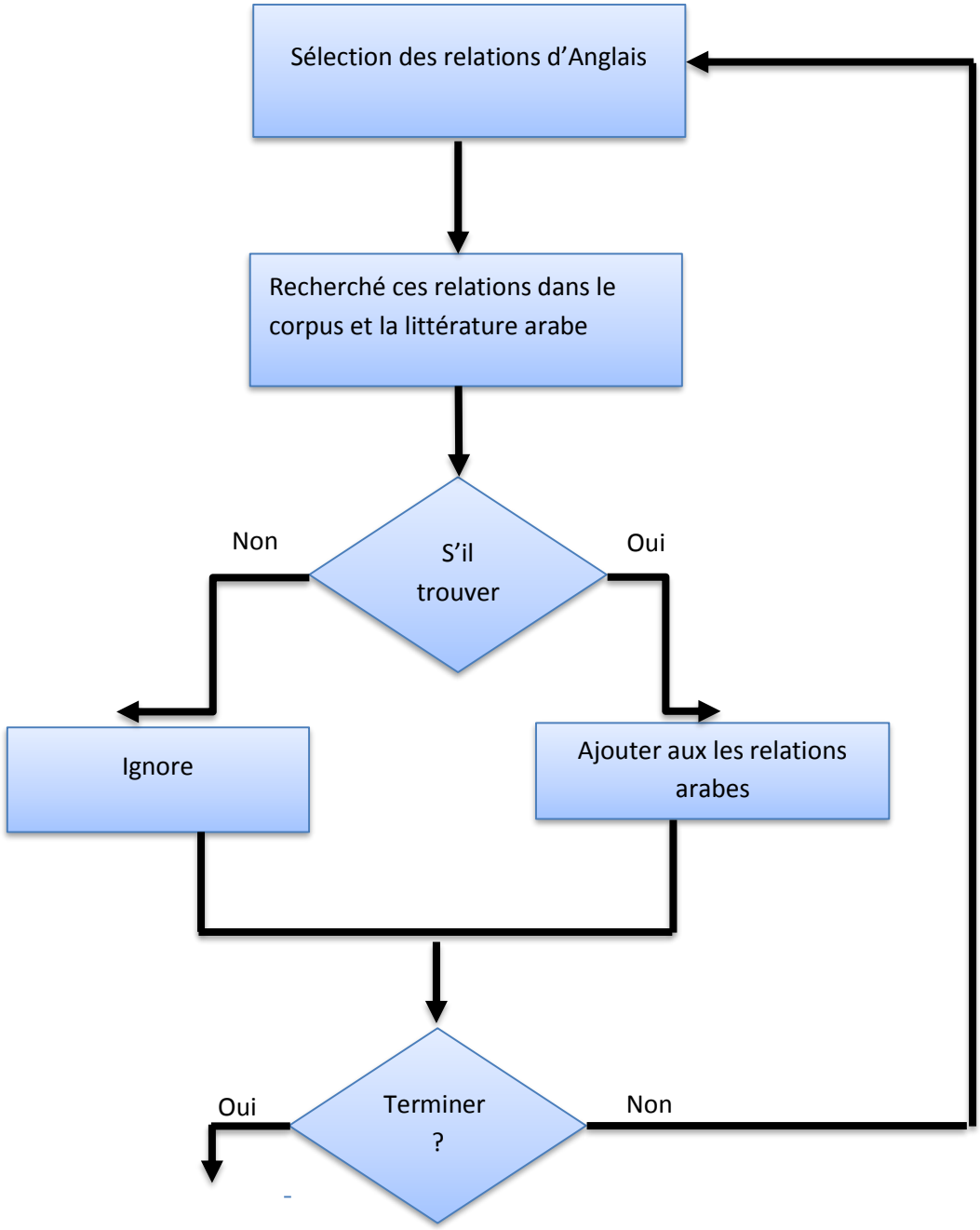

Figure 2.4 processus d'extraction de relations Arabe à partir de relations de l'Anglais [13].

Tout d'abord, Al Sanie et al ont extrait certaines des relations arabes à partir des relations anglaises. Le processus comporte trois étapes. Ils ont sélectionné une relation anglaise, puis analysé les références de rhétorique et de la littérature arabe [20, 21, 22] pour cette relation, ils ont également analysé le corpus en arabe pour voir si cette relation est explicitement existé ; si oui, la relation est ajoutée à la liste des relations arabes ; Sinon, la relation est ignoré.

Etude d'un corpus arabe pour identifier de nouvelles relations selon la rhétorique Arabe

En raison des différences dans les concepts de rhétorique entre les deux langues, les auteurs ont étudié des corpus arabe pour en extraire quelques relations rhétoriques arabes qui reflètent l'essence des textes arabes. Les auteurs ont trouvé que ces relations tombent sous deux catégories :

- **Catégorie 1 :** Les connecteurs qui relient les clauses, mais aussi des mots.
- **Catégorie 2 :** Les connecteurs qui relient les clauses seulement.

Les exemples suivants expliquent les deux catégories :

Exemple 3 : La phrase suivante comprend la relation " وجار ومجرور" qui relie une clause avec un mot.

- وجدت محمدا **في** السيارة.

Exemple 4 : La phrase suivante comprend la même relation, mais dans ce cas, le connecteur relie deux clauses.

- وجدت محمدا **في** بيت اخيه.

Exemple 5 : La phrase suivante comprend une relation« استدراك» qui relie deux clauses.

- سأذهب اليوم الى العمل، و**لكني** لن احضر الاجتماع

Les auteurs ont retenus les relations de la catégorie 2.

Etude des indicateurs, Cue phrases Arabe pour déterminer les relations qui les lient

: Dans la troisième phase, les auteurs [13.14] ont parcouru le corpus afin d'obtenir les mots qui sont considérés comme des connecteurs linguistiques (cue phrases) pour la langue arabe. Selon l'étape précédente, les auteurs ont retenus les connecteurs (soulignés) des relations de la catégorie 2.

À la fin de cette approche à 3 étapes, Al Sanie et al,. ont pu aboutir à un ensemble de 11 relations rhétoriques Arabe qui sont présentées dans le tableau 2.2 ci-dessous :

Nom de relations en Arabe	Nom de relations en Anglais
الشرط	condition
العطف	joint
التفسير	interprétation
الاستدراك	antithèses
التعليل	justification
التوكيد	confirmation
الترتيب	séquence
النتيجة	résultat
التمثيل	exemple
القاعدة	base
التفصيل	explanation

Tableau 2.2: Liste des relations rhétoriques ainsi que les indicateurs (Cues phrases) Arabe d'après [13.14].

2.3.2 L'analyse rhétorique de textes Arabe

L'analyse RST d'un texte Arabe est basée sur les mêmes étapes que prouvent les textes anglais ou français.

Nous présentons brièvement ces étapes :

1. **Segmentation du texte :**

- segmenter le texte en unités textuelles (phrases) : selon les signes de ponctuations(le point, le point d'interrogation, le point d'exclamation) ;
- segmentation des phrases en unités minimales : en se basant sur les indicateurs prototypiques (Cues phrases).

2. **Application des règles rhétoriques :**

- Détermination du segment Noyau et celui Satellite : selon des règles sur les Cues phrases.
- Détection des relations rhétoriques : à partir la Cues phrases indiquer dans l'étape précèdent (1) et un ensemble des contraintes sur les unités (Noyau et Satellite).

3. **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.

2.4 Conclusion

La RST est une théorie créée pour la linguistique informatique, et elle a été rapidement adoptée par toute une communauté de chercheurs de différents horizons, même pour des analyses purement linguistiques. **Mann et Thompson** proposent un certain nombre de relations rhétoriques de base qu'ils estiment être suffisant pour décrire la plupart des textes écrits en anglais. La question qui nous concerne est : qu'en est-il de l'application de cette théorie, la RST à la langue Arabe. Nous avons présenté dans ce chapitre comment que les chercheurs [13,14] ont pu adapter les travaux sur les autres langues, notamment la langue anglaise, sur la langue Arabe. Les résultats de ces travaux ont permis d'identifier un ensemble de relations pour l'Arabe ainsi qu'un ensemble d'indicateurs (qui peuvent varier d'une application à une autre) permettant ainsi de faire une analyse rhétorique sur des textes Arabes. Ces travaux et leurs résultats nous ont servis de base pour la réalisation de notre système qui, rappelons-le, est un outil pour l'analyse rhétorique d'un texte Arabe en se basant sur la théorie de la structure rhétorique : RST. Le prochain chapitre est consacré à la présentation de notre système.

Chapitre III

Description du système et testes

3. Description du système et testes

3.1 Introduction

Rappelons que le but de cette étude est d'implémenter un outil d'aide à l'analyse sémantique de textes Arabe en se basant sur la théorie de la structure rhétorique. Ce chapitre vise principalement à décrire l'architecture de notre système ainsi que les fonctionnalités de l'outil réalisé à travers quelques captures d'écrans.

3.2 Architecture globale du système

Notre outil d'analyse sémantique de textes arabe en se basant sur la RST, met en œuvre les étapes décrites dans l'algorithme 3.1 ci-dessous.

Entrées : texte Arabe
Sortie : Arbre RST pour le texte en entrée.

Début

1. Segmentation du texte
   - segmenter le texte en unités textuelles (phrases).
   - segmentation des phrases en unités minimales en se basant sur les indicateurs prototypiques (cues phrases).
2. Application des règles rhétoriques
   - Détermination du segment Noyau et celui Satellite
   - Détection des relations rhétoriques.
3. Construction de l'arbre RST

Fin

Algorithme 3.1 étapes d'analyse rhétorique d'un texte Arabe

Ainsi, l'architecture globale de notre système est illustrée par la figure 3.1.

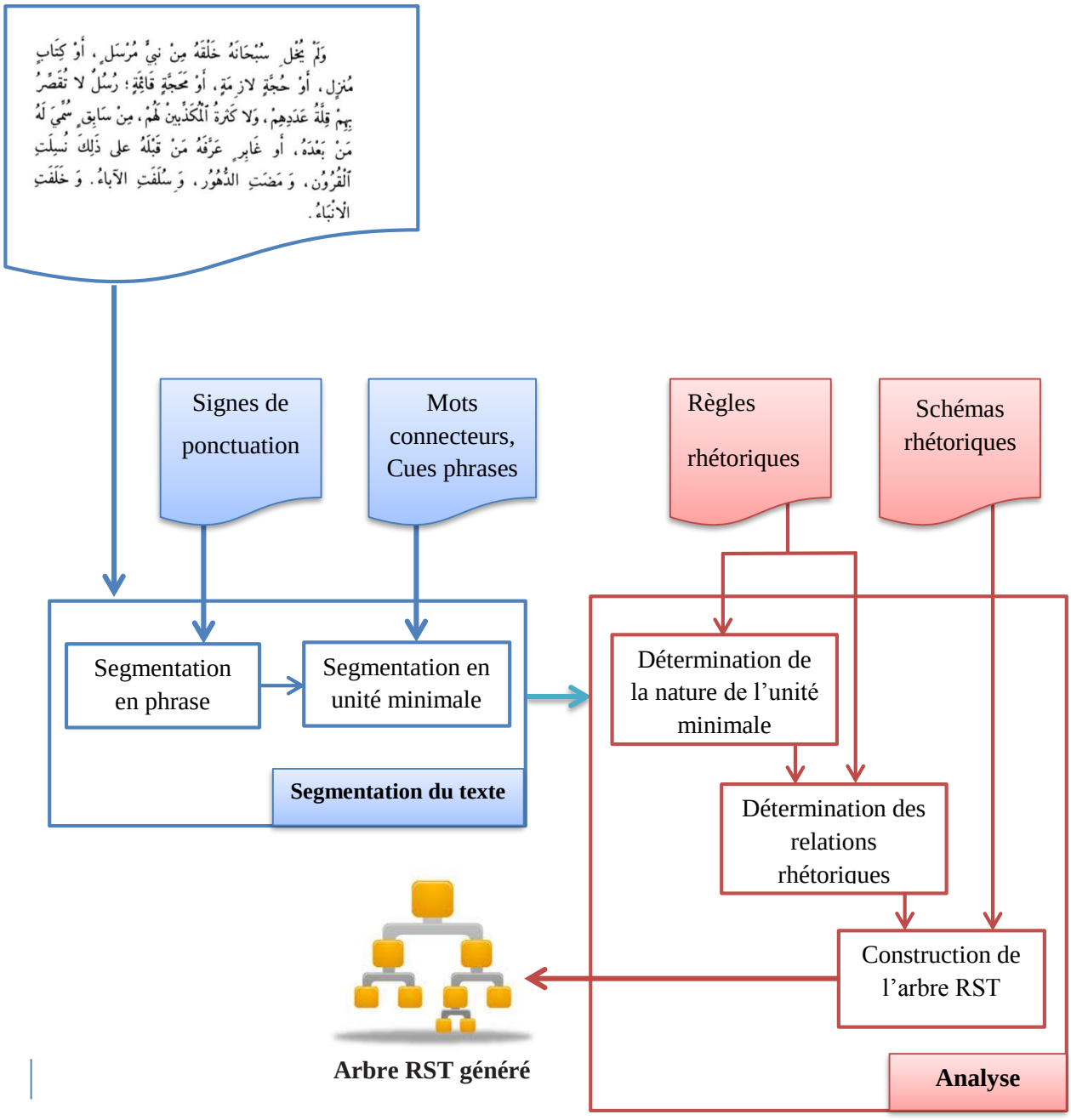

Figure 3.1 : Architecture du système .

3.2.1 Description de l'architecture de l'application :

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

Le module de segmentation : ce module permet de segmenter un fichier au format texte d'abord en unités textuelles, ensuite en unités minimales.

Le module d'analyse : il permet de détecter les relations rhétoriques qui relient deux unités minimales adjacentes entre elles. Ce module délivre un arbre RST qui qualifie les relations détectées entre les unités textuelles.

3.2.2 Implémentation :

Nous avons développé une interface graphique principale en JAVA sous Eclipse gérant en amont tous les traitements nécessaires aux tâches demandées par l'utilisateur afin de générer une analyse selon les données en entrée. Cette interface regroupe les traitements dont nous avons développé dans les parties précédente, à savoir : segmentation, analyse rhétorique, détermination de l'arbre RST.

Dans cette interface principale, le système présente à l'utilisateur les différentes étapes d'analyse du texte.

Pour effectuer une analyse, l'utilisateur n'a qu'à sélectionner le texte qu'il désire suivre les étapes décrites dans la zone cadrée en rouge à gauche, figure 3.2.

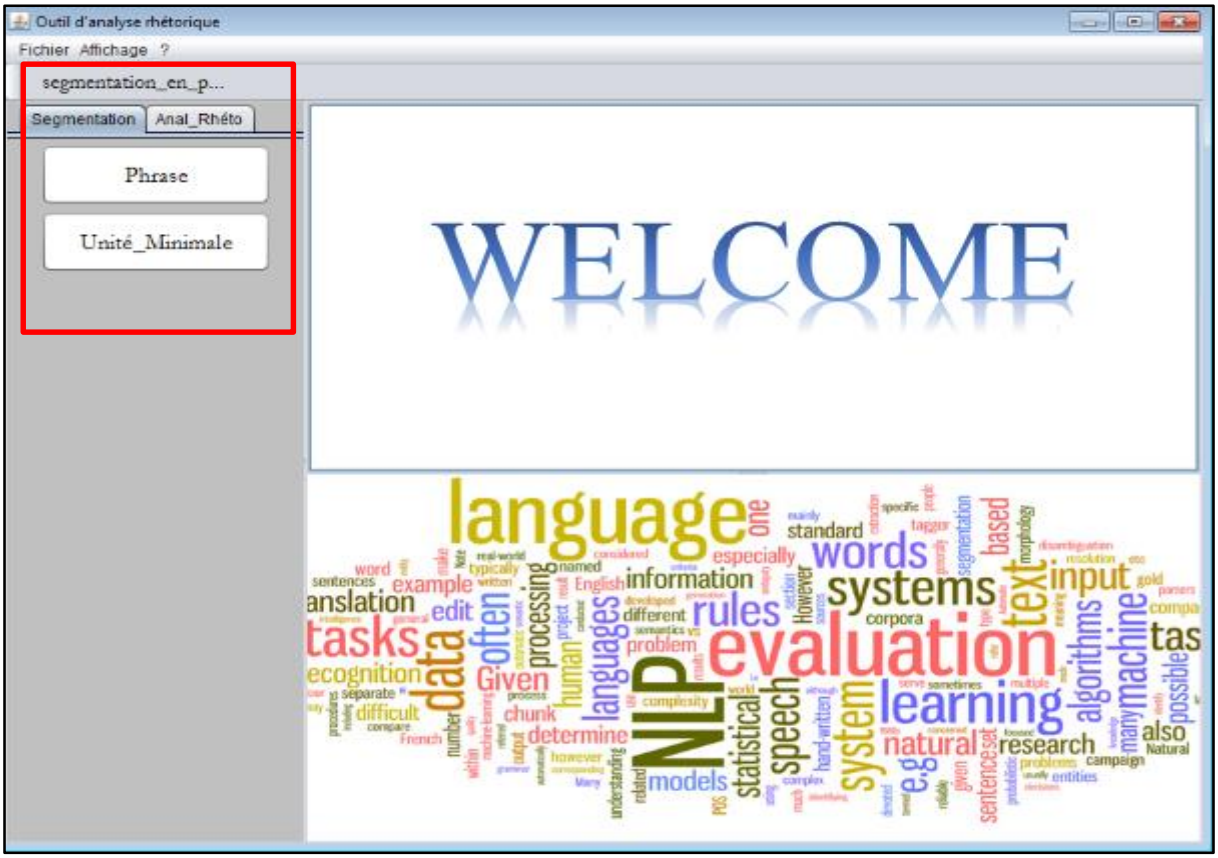

Figure 3.2 : Interface principale du système.

4.3.1 **Chargement du texte :**

L'étape de chargement d'un texte nécessite deux étapes à savoir l'identification de la source du texte 1 et sa sélection (voir figure 3.3). En effet, le système traite des textes de formats de fichiers textes(TXT).

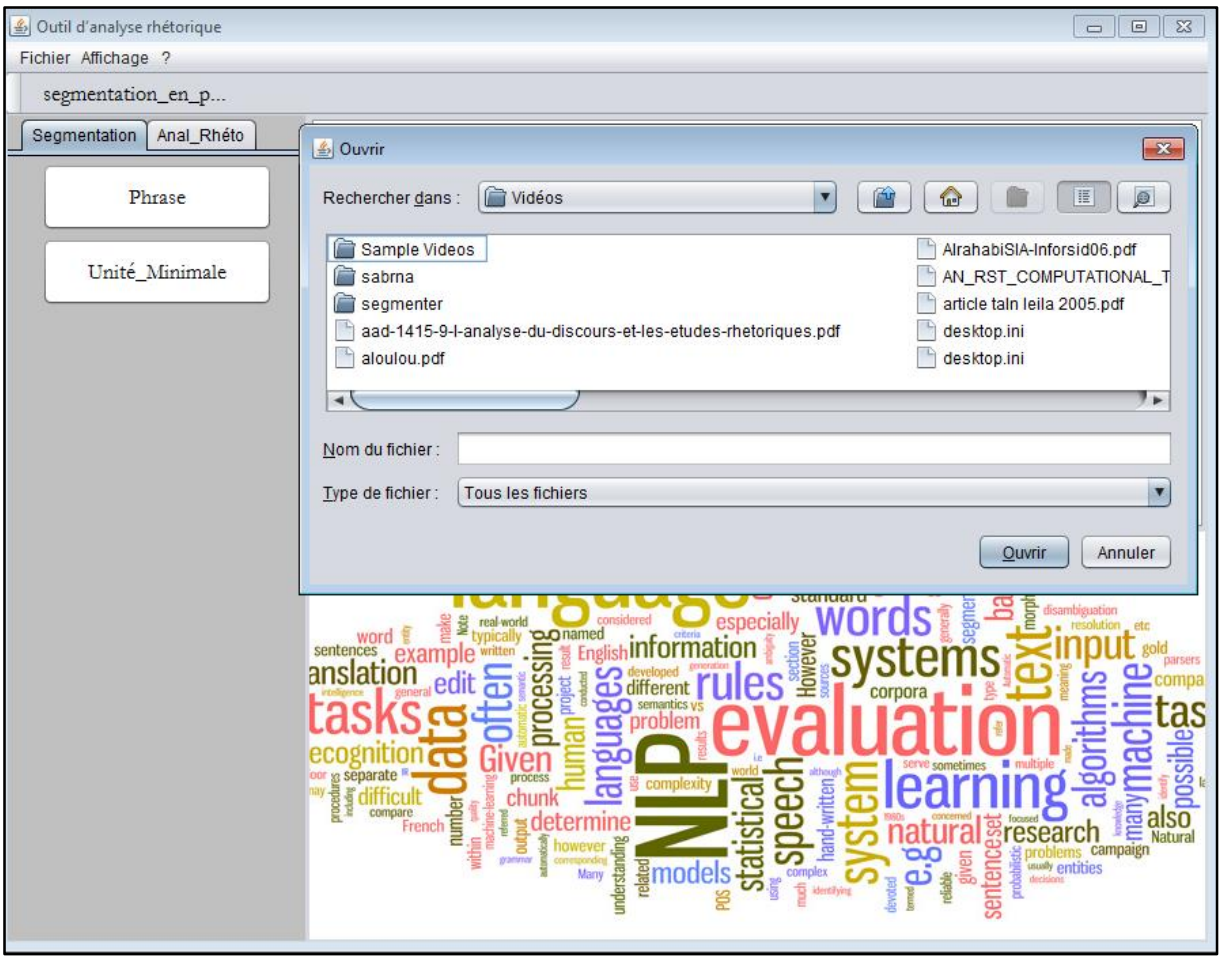

Figure 3.3 : Interface de chargement du texte.

4.3.2 Segmentation du texte :

L'étape de segmentation du texte que nous adoptons consiste à segmenter le texte source selon plusieurs niveaux : paragraphes, phrases et unités minimales.

Nous nous appuyons sur un ensemble de règles permettant de tenir en considération.

A) 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 : بينما, لكن, إذ, هكذا.)

1) Règles basées sur les signes de ponctuation :

Dans ce qui suit, nous présentons un exemple de règles pour quelques signes de ponctuation tout en signalant quelques difficultés de conception ou d'application de ces règles.

1.1) ***Le point :*** Généralement le point marque la fin de phrases sauf pour quelques cas, citons à titre d'exemple[13] :

❖ من المنتظر أن يكتمل مشروع خط أنابيب نابوك البالغ طوله 3300كيلومترا في 2014/08/12 بتكلفة تُقدر بـ 7.9مليارات.

CG			M	CD			
			.	Chiffre	Caractère		

SI un point est précédé d'un chiffre. ET SI ce chiffre est précédé d'un seul caractère. ALORS le point ne marque la fin de la phrase.

1.2) **Les deux points « : »**

Les deux points sont généralement utilisés dans un segment textuel commençant par l'un des verbes : سأل, قـال, etc. Ce dernier peut être précédé d'une conjonction de coordination (waw , وfa , ف thomma, ثم etc.) suivis ou non de guillemets et dans les dialogues ils peuvent être suivis d'un retour chariot (/r). Pour tous ces cas les deux points ne peuvent pas être des délimiteurs de phrases car le sens de la phrase n'est pas terminé. *Exemple :*

❖ فقال له أحمد: هل يعني هذا أنّه سلّ لا يوجد له شفـاء؟

CG			M	CD			
		/r	:				**فقال**

[13] Notons que **CG** : Contexte Gauche**; CD** : Contexte Droit ; **M**:Marqueur.

SI le contexte gauche de deux points commence par le mot فقال. ET SI les deux points sont suivis de /r. ALORS les deux points ne marquent pas la fin de la phrase.

2) Règles basées sur les mots connecteurs :

Les auteurs de segmentation selon des mots connecteurs, nous a permis de détecter certains mots qui peuvent jouer le rôle de connexion entre deux phrases. Nous présentons ci-dessous quelques mots et les règles correspondantes

2.1) La conjonction de coordination (حتّى) : Hatta حتّى est peut connecter entre deux phrases s'il est suivi de إذا ou ..إن

Il est à noter que إن peut être suivi d'un pronom relatif et devient :,إنه, إنها , إنهما، إنهم ، إنهنّ .

Exemple :

- ومكث بيدبا أيّاما كثيرة في سجنه حتىّ إذا كانت ليلة من اللّيالي سهد فيها الملك فغرق في التّفكير في أمور الفلك.

CG			M	CD			
		إذا	حتّى				

2.2) Les connecteurs (لكن, لكنّ) :

Les connecteursلكن lakin, lakinna, suivi ou non d'un pronom relatif (i.e لكنّه ,لكنّهم، لكنّها, لكنّني لكنّكم, لكنّ كن, لكنّ هن.) marque toujours le début d'une nouvelle phrase.

Exemple :

هناك بعض التقارير الذي تأكد انشاء قسم جديد في الكلية ولكن لم يحدد بعد موعدا لذلك.

CG			M	CD			
			لكن				

La figure 3.4 présente un exemple de résultat délivré par cette étape.

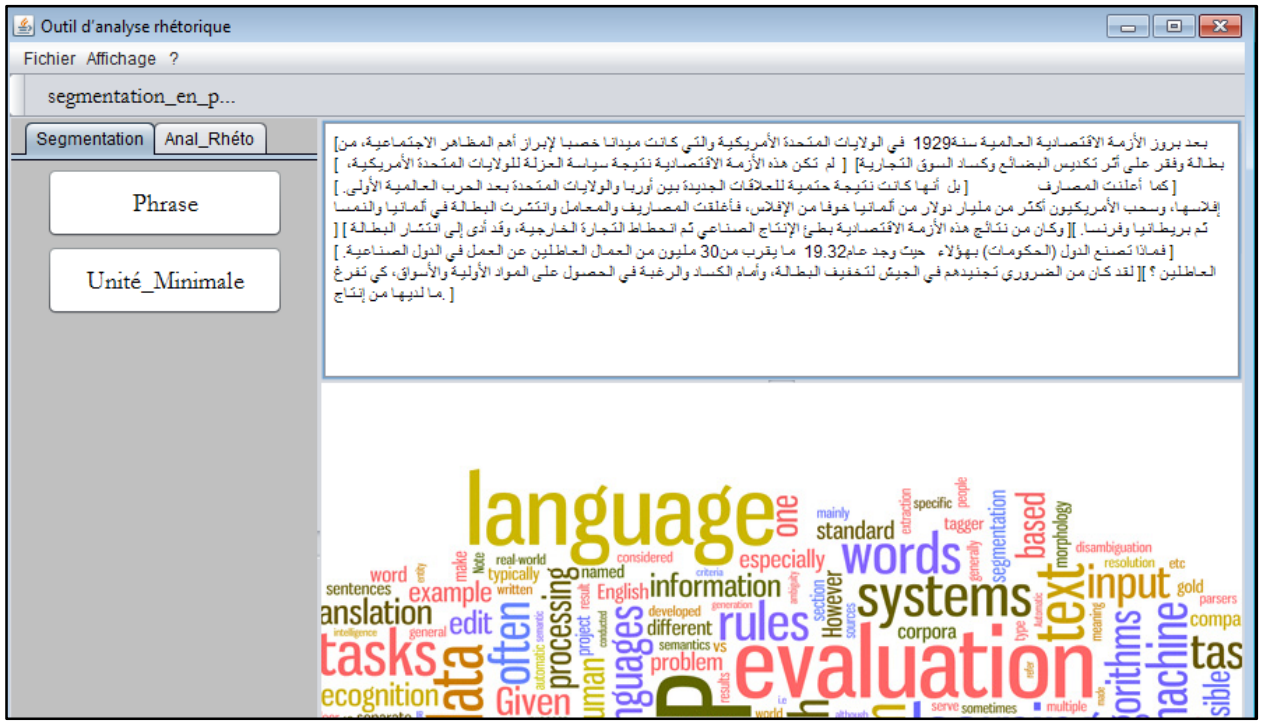

Figure3.4 : Intarface de segmentation du texte en phrases et unité minimale.

4.3.3) Analyse rhétorique du texte

L'étape de l'analyse rhétorique a pour but de comprendre le texte. Ainsi, son objet est de fournir un cadre d'interprétation pour la structure discursive du texte source d'une part et d'autre part de le d'écrire sous une représentation hiérarchisée qui met en évidence les différentes structures visuelles du texte et les propriétés qu'elles entretiennent entre elles.

Cette étape utilise le principe de la Théorie des Structures Rhétoriques définie par les chercheurs Mann et Thompson [15] afin de générer une articulation cohérente entre les différentes unités minimales du texte source.

Le système procède à l'analyse rhétorique de texte comme suit :

1) **Détermination de la relation rhétorique et de la nature de l'unité minimale.**

Cette phase a un double objectif ; premièrement de lier deux unités minimales adjacentes entre elles, dont l'une possède le statut de noyau et l'autre a le statut noyau ou satellite, et deuxièmement de déterminer les relations rhétoriques qui existent entre les différentes unités minimales juxtaposées d'un même paragraphe.

Chaque relation est définir par un ensemble de connecteur, l'apparition de l'un de ces connecteur indique une relation qui relier deux unités.

Notons que la détermination des relations rhétoriques n'est pas basée seulement sur l'apparition de connecteur, mais il y a aussi des contraintes sur les unités minimales doit être vérifier pour dire que ce connecteur indique ou définir une telle relation.

Le résultat du module d'analyse rhétorique est un fichier texte qui présente la totalité des phrases du texte avec les relations qui les relient (voir le Figure 3.5).

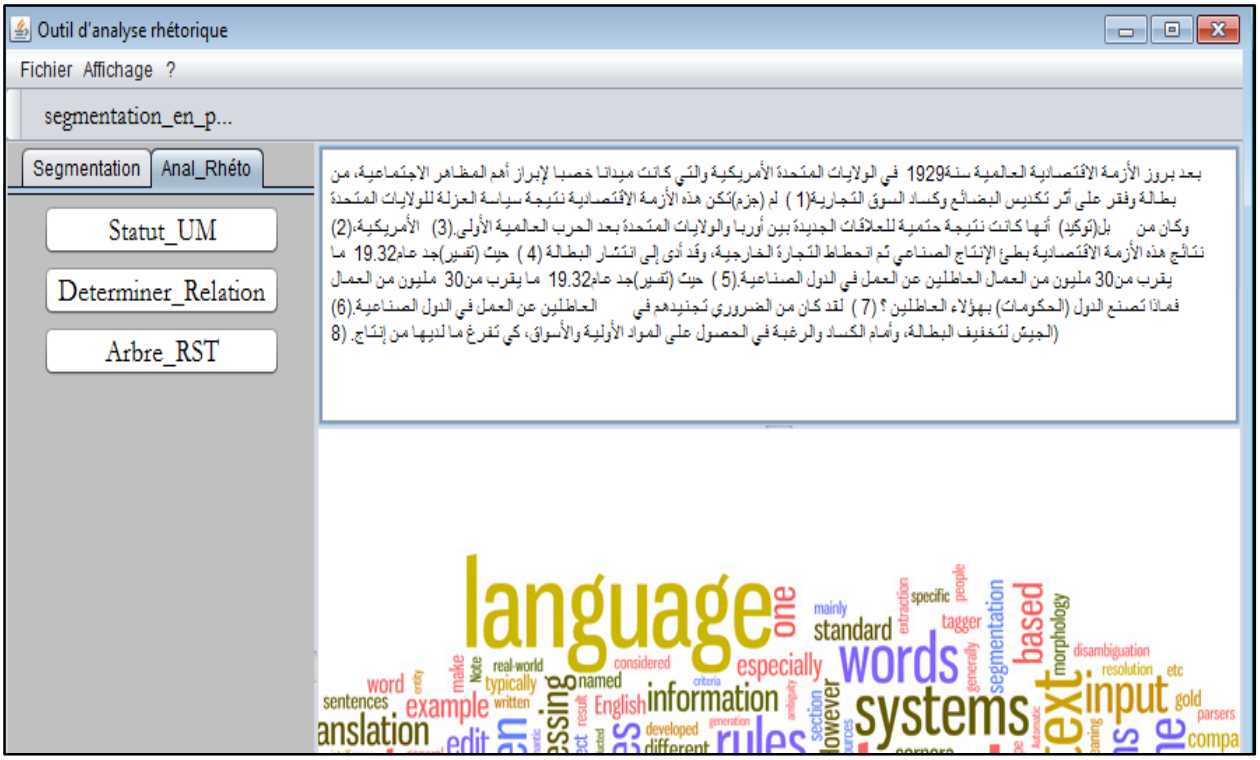

Figure 3.5 : Interface de Détermination de la relation rhétorique.

2) Détermination de l'arbre RST

Cette phase consiste à créer l'arbre RST le plus descriptif qui décrit l'organisation structurelle du texte source quel que soit le niveau hiérarchique de ce dernier, et cela en prenant en considération les différentes contraintes des liens entre unités minimales du texte. Ainsi, l'arbre RST se caractérise par sa capacité de connecter récursivement, par le biais d'une relation de discours qui ne peut relier que des segments de texte adjacents, les unités minimales et les segments de texte plus larges ainsi construits selon un ordre cohérent et informatif.

Il est à signaler que cette phase de création de l'arbre RST a fait l'objet de plusieurs travaux de recherche. Parmi ces travaux nous pouvons citer les recherches de Marcu. [16].

Le résultat est enregistré dans un fichier texte qui englobe les unités minimales (noyau et/ou satellite), les relations rhétoriques qui les relient et le niveau hiérarchique de leurs emplacements dans l'arbre.

3.3 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 outil d'analyse rhétorique. Quelques écrans de l'application ont été également présentés.

Conclusion générale

Dans cette étude, nous avons adapté l'outil « Outil d' Analyse Rhétorique » pour l'analyse sémantique des textes arabe, pour ce faire nous avons organisé notre travail selon trois étapes principales. D'abord, nous avons récolté et préparer toutes les données nécessaires : les questions et les réponses… ensuite, nous avons développé une méthode pour la segmentation et la détermination des relations rhétoriques existe entre les unités minimales de réponses, base sur la RST. En fin, nous avons terminé par une représentation hiérarchique des réponses sous forme d'un arbre rhétorique.

Il nous est apparu que la détermination des relations rhétorique entre les unités minimales des réponses, au niveau analyse rhétorique, est une étape cruciale pour obtenir par la suite une représentation hiérarchique de réponses de bonne qualité. Nous avons proposé une technique simple de détermination du relations rhétorique base sur un ensemble de Cues phrases (mots connecteurs) introduisissent par Al Sanie et al, et un ensemble des contraintes et conditions sur les unités minimales.

Au final, nous avons obtenue de bons résultats au niveau d'analyse rhétorique, comme la montre l'évaluation de notre outil. C'est pour quoi, on peut conclure que nous avons atteint à peu de frais le niveau d'analyse rhétorique présenté dans l'état de l'art.

D'un point de vue générale, l'utilisation de la RST pour l'analyse sémantique de texte arabe reste un domaine très ouvert et présente des marges de progression importante, de faits de la richesse morphologique de cette langue.

La RST est l'un des méthodes ou approches de base souvent considéré comme triviale dans des langues comme l'anglais, reste moins utilisable pour le traitement automatique de la langue arabe, ou de grandes améliorations peuvent elle encore être apportées à ce domaine.

Références bibliographiques

Références bibliographiques

[1]. Alain MERLE ; Un analyseur pré-syntaxique pour le levée des ambiguïtés dans des documents écrits en langue naturelle : Application à l'indexation automatique ; Thèse, institut national polytechnique de Grenoble, septembre 1982.

[2]. Jean VERONIS, Informatique et Linguistique 1 ; unité d'enseignement INF Z18, université de Provence, centre informatique pour les lettres et sciences humaines, 2001.

[3]. J-H. JAYEZ, Compréhension automatique du langage naturel le cas du groupe nominal en français, Masson, 1985.

[4]. Kees Versteegh. The arabic language. Edinburgh University Press, 2001. (Cité en page 47.)

[5]. Lamia Hadrich Belguith. Traitement des erreurs d'accord de l'arabe basé surune analyse syntagmatique étendue pour la vérification et une analyse multicritère pour la correction. Thèse, Faculté des Sciences de Tunis, 1999. (Cité en pages 49, 56, 58et 82.).

[6]. Lamia Hadrich Belguith, Leila Baccour et Mourad Ghassan. Segmentation de textes arabes basée sur l'analyse contextuelle des signes de ponctuations et de certaines particules. Actes de la 12ème conférence sur le Traitement Automatique des Langues Naturelles TALN'2005 - Dourdan France, vol. Vol. 1, pages 451–456, 2005. (Cité en pages xiii, 50, 51, 54, 55, 56, 82, 96, 99 et 122.)

[7]. Lamia Hadrich Belguith, Chafik Aloulou et Abdelmajid Ben Hamadou. MASPAR : De la segmentation à l'analyse syntaxique de textes arabes. In CÉPADUÈS Editions, editeur, Revue Information Interaction Intelligence I3, volume 7, pages 9 – 36, http ://www.revue-i3.org/, mai 2008. 2008. ISSN : 1630-649x. (Cité en pages 51, 59 et 82.)

[8]. Lamia Hadrich Belguith et Nouha Chaaben. Analyse et désambiguïsation morphologiques de textes arabes non voyellés. In 13ème conférence sur le Traitement Automatique des Langues Naturelles (TALN'2006), pages 493–501, Leuven- Belgique, 10-13 avril 2006 2006. (Cité en pages 50, 52 et 100.)

[9]. Nouha Chaaben, Lamia Hadrich Belguith et Abdelmajid Ben Hamadou. The MORPH2 new version : A robust morphological analyzer for Arabic texts. In Actes des 10èmes journées internationales d'analyse statistique des données JADT'2010, Rome, Italy, 2010. (Cité en pages 52, 96 et 122.)

[10]. Fathi Debili et Emna Souissi. Etiquetage grammatical de l'arabe voyell ou non. In Proceedings of the Workshop on Computational Approaches to Semitic Languages, Semitic '98, pages 16–25, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics. (Cité en pages 52 et 58.)

[11]. Xu, J., Fraser, A., and Weischedel, R. Empirical studies in Strategies for Arabic retrieval. In Sigir 2002. Tampere, Finland: ACM, 2002.

[12]. Waleed Al-Sanie . Towards an Infrastructure For Arabic Text Summarization Using Rhetorical Structure Theory. 2005.

[13]. Al-Sanie, W., Touir, A., & Mathkour, H. (2005). Towards a Rhetorical Parsing of Arabic Text. International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce 1086–1091.

[14]. Mathkour, H. I., Touir, A. a., & Al-Sanea, W. a. (2008). Parsing Arabic Texts Using Rhetorical Structure Theory. Journal of Computer Science, 4(9), 713–720.

[15]. MANN, W. C. et THOMPSON, S. A. (1987). «Rhetorical structure theory: a theory of text organisation ». ISI Reprint Series ISI / RS-87-190, Information Sciences Institute. 2, 5, 20, 24, 25, 27, 28, 35, 41

[16]. Marcu, D. (2000b) The theory and practice of discourse parsing and summarization, The MIT press.

[17]. **Sébastien Druon** . «Projet de taxinomie des connecteurs du français pour le traitement automatique : l'exemple des consécutifs **»** Mémoire de maîtrise de Sciences du Langage Université Michel de Montaigne -- Bordeaux III

[18]. Mann, W.C., Matthiessen, C.M.I.M. & Thompson, S.A. (1992) Rhetorical Structure Theory and Text Analysis, *Discourse Description: Diverse linguistic analyses of a fund-raising text. ed. by W. C. Mann and S. A. Thompson*, pp. 39-78.

[19]. MANN, W. C. et THOMPSON, S. A. (1987). «Rhetorical structure theory: a theory of text organisation ». ISI Reprint Series ISI / RS-87-190, Information Sciences Institute. 2, 5, 20, 24, 25, 27, 28, 35, 41.

[20]. Aubadah, M.I., 1983. Al-Jumlah Al-Arabiah. 1st Edn., Munshat Al-Ma'aref, Alexandria, Egypt.

[21]. Gabawah, F., 1972. Iraab Al-Jumal wa Ashbah AlJumal. Dar Al-Qalam Al-Arabi, Damscuss, Syria (in Arabic). www.qalamarabi.com/About_dar.php.

[22]. Al-Ansari, I.H. (2003) Mugni Al-Labeeb An Kutub Al-Aareeb, *Al-Maktabah Al-Asriah for publishing and printing*.

Annexes

Annexes

1) Les connecteur d'arabre :

les connecteurs	Nom de relation rhétorique
,من دون , ولو ,دون	شرط
لكن, وإن, بل , أما , لكني , لكنا, لكنهم , لكنه , ولكن, إلا أن, إلا أنه, إلا أنها, إلا أنهم, غير أن , غير أنه, غير أنها, ليس لأن, ليس لأنه, ليس لأنهم, ليس لأنها, ليس لأنهما, ليس لأنك, ليس لأنكم, ليس لأنكما, ليس لأنكن	استدراك
فهذا, كذلك, وكذلك, كما, إضافة, إضافة إلى, وإضافة إلى	تفصيل
الا , ما عدا , سوى, لم	استثناء
رغم, وإنما, على رغم, ليس إلا, إلا, لاسيما, لقد, إن, إنها, فأن, أنه, انهم, انها, انما, أكد, وأكد	توكيد
ربما, وربما, فربما	تقليل
و , ف, ثم, كما أن, لكي, إما, لا, لم يمض على,	ربط
فإن	قاعدة
لا , لم , لن, دون, ولم, ليس, ليسوا	نفي
مثلا , على سبيل المثال, مثل, فمثلا	تمثيل
بمعنى ان , حيث ان, لأن, أي, لذلك	تفسير
نتيجة لذلك , نتيجة لهذا, نستنتج, لذا , و منه , اذن ,و عليه,فلذا	استنتاج
ويجزم, يجزم	جزم
فهي	تعريف
ثم , من ثم, ف	ترتيب
على الارجح , ربما, بنسبة اكبر , اغلب الظن, لعل, ولعل, ولعله, لعله, ولعلها, ولعلهم, لعلها, لعلهم, ربما, وربما, والأرجح, الأرجح	ترجيح
أو	امكان
إلا	حصر
خاصة , بالاخص , على وجه الخصوص, خصوصا, خصيصا	تخصيص
من أجل, كي, حتى	تعليل