

Entre linguistique et informatique

Textes réunis et préfacés par :
Pr. Mohamed Bouattour



**Laboratoire Langage
et Traitement Automatique**
Faculté des lettres et sciences humaines de Sfax
Université de Sfax

*Laboratoire Langage et Traitement Automatique
Faculté des lettres et Sciences Humaines – SFAX*

Entre linguistique & Informatique

Textes réunis et préfacés par
Pr. Mohamed Bouattour

Med Ali Editions



Titre : Entre linguistique et informatique

Textes réunis et préfacés par : Pr. Mohamed Bouattour

Tous Droits réservés ©

Editeur

1- C A E U – Med Ali Editions

Rue Med Chaabouni Imm Zarka yamama – 3027 Sfax Tunisie

Tél : 216 74 407 440 / Fax : 216 74 407441

Email : edition.medali@tunet.tn

Site : www.edition-medali.tn

2- Laboratoire Langage et traitement automatique

Factulté des Lettres et Sciences Humaines de Sfax

Route de L'aéroport Km 4,5 – Sfax Tunisie

Tél : 216 74 670 544 / Fax : 216 74 670 540

ISBN :978-9973-33-524-1

N° Editeur :876-624/17

Impression :La Maghrébine pour l'impression - Tunis

SOMMAIRE

Mohamed MSALMI : Le projet MOOC FOFLE en Tunisie. Enseigner et former au supérieur avec le numérique	13
Fathia DAOUES : Conception d'un dictionnaire électronique bilingue pour une lecture labiale.....	31
Saloua KAMMOUN : La traduction automatique des structures complétives du et vers le français et l'arabe par «Google Translate»: régularités, anomalies et tentatives d'amélioration	43
Fatma BEN DHAOU : La linguistique de corpus: extraction, manipulation et traduction automatique des locutions prépositives dans le discours législatif...	53
Raja HAMDI : Importance des collocations dans les systèmes de traduction automatique des langues.....	73
Tawfik KHIL : Itinéraire de la traduction juridique automatique de l'arabe vers le français: Le paradigme «شهادة»	85
Elouni Najeh : Émotions et discours entre instantané et pré-construction: les émotions dans les tweets	99
Bougobba Amal : FrameNet et le traitement de l'ambiguïté sémantique (Cas de verbe sentir)	111
Ezzedine BOUHLEL : L'enseignement du français par les TIC	123

Ifaoui Imen : La complexité du traitement automatique des paraphrases:	137
Rym SALEM : «L'apport de TreeTagger au concordancier AntConc: examen du fonctionnement de la construction «savoir que» dans le discours journalistique»	149
Siwar MABROUK : Traduction automatique des collocations nom + adjectif.....	167
Thouraya Fadhel : Causes des problèmes de traduction automatique français-arabe des temps verbaux.....	181
Najeh El-Ouni : Les écrits électroniques: faire du face à face avec de l'écrit	
Hela SAIDANI : L'apport d'Excel en linguistique: Le cas du syntagme nominal...	211
Sana Abdelhamid : Traitement automatique des anaphores associatives méronymiques.....	227

L'APPORT D'EXCEL EN LINGUISTIQUE: LE CAS DU SYNTAGME NOMINAL

Hela SAIDANI

LLTA, helasaidani1991@gmail.com

Résumé:

La langue et l'informatique sont devenus indissociables de nos jours. En effet, grâce aux collaborations entre linguistes et informaticiens, le traitement d'un corpus donné est devenu automatique, d'où le «Traitement Automatique des Langues» ou TAL, ce qui permet une analyse plus rapide et, surtout, précise. Les corpus sur lesquels les chercheurs travaillent sont multiples. Nous nous proposons de travailler sur le syntagme nominal qui peut être étendu par une expansion. L'expansion est, traditionnellement, un adjectif, un syntagme prépositionnel, un SN, une subordonnée relative ou une conjonctive. Nous suggérons de travailler sur deux corpus; le premier, littéraire et l'autre, scientifique, afin de voir le fonctionnement dudit SNE et pour relever certaines régularités qui lui sont propres. Pour ce faire, nous aurons essentiellement recours au logiciel *Antidote*⁽⁴⁵⁾ et nous insérerons, par la suite, les données que nous obtiendrons dans le tableur de Microsoft *Excel*⁽⁴⁶⁾, sachant que dans ce tableur, il existe un sous-menu: le *sous-total*, permettant l'obtention de résultats rapides, mais aussi précis. Nous rendrons également compte des différentes structures du syntagme nominal étendu et nous verrons combien d'expansions le noyau nominal peut avoir. Nous ferons un descriptif des expansions nominales en partant des cas les plus réguliers vers les cas les moins réguliers, voire disparates.

Mots clés:

syntagme nominal, expansion, Antidote, Excel, statistiques, sous-totaux

45 – *Antidote* HD version 2.1, 2009 Druide informatique inc.

46 – Microsoft Office Excel 2010.

Abstract:

Nowadays, linguistics and computer science have become inseparable. In fact, thanks to the collaboration of linguists and computer scientists, the treatment of a corpus has become automatic, thus the "Natural Language Processing" or NLP, provides faster and more accurate analysis. Up until now, several studies have been conducted on corpus. We intend to work on the noun that can be expanded. The expansion is traditionally an adjective, a prepositional, a noun, a relative clause or a conjunctiva. We suggest to work on two corpora; the first, literary, and the other scientist, to see the execution of the noun and to raise regularities. To do so, we first use *Antidote* software and then the data which will be exported as a worksheet of Microsoft Excel, knowing that in this spreadsheet, there is one submenu: the *subtotal*, to obtain not only quick, but also precise results. Then, we will report the different structures of the extended noun and see how many expansions a nominal core can have. Last but not least, we will describe nominal expansions from the most regular cases to the least regular cases.

Keywords:

noun, extension, Antidote, Excel, statistics, subtotals

I. Introduction

La présente étude traite de la présentation du travail sur Excel en linguistique. Il s'agit d'une étude cherchant à caractériser l'utilité du tableur Excel dans un travail de recherche. Nous prendrons ici le cas du SN. La *Grammaire traditionnelle* différencie le SN minimal du SN étendu (désormais SNE) dont le nom (désormais N) noyau est précédé ou suivi d'une expansion qui, conformément au principe de la suite la plus longue, peut être un adjectif (désormais Adj.), un syntagme prépositionnel (désormais SP), un SN, une subordonnée relative (désormais REL.) ou une conjonctive (désormais CONJ.). Ces expansions sont, syntaxiquement, effaçables, donc facultatives. Cependant, il arrive que l'expansion donne au N qu'elle caractérise une exactitude et se révèle parfois essentielle. Il s'agit, dans cet article,

d'une étude qui tend à déterminer les différentes expansions du N. Nos corpus sont constitués d'un texte littéraire, intitulé *La Parure* de Maupassant (1884), et d'un article scientifique intitulé *Une étoile à la luminosité étrange surprend les scientifiques*, article que nous avons extrait d'Internet⁽⁴⁷⁾. La problématique relative à ce sujet est la suivante : quel est l'apport d'Excel pour un linguiste ? Afin d'apporter des réponses à cette question, nous répartirons notre travail comme suit. D'abord, nous nous intéresserons à la préparation des corpus, c'est-à-dire aux étapes nécessaires et préalables à l'analyse des corpus. Par la suite, nous parlerons de l'utilisation d'*Excel* en linguistique ainsi que de son apport dans un travail de recherche et nous prendrons, à titre d'exemple, le cas du SNE. Enfin, nous relèverons les statistiques et les régularités des SNE dans les deux textes.

II. Préparation des corpus

La préparation des corpus est la partie préalable à l'analyse des SNE dans le texte littéraire et l'article scientifique. Disposer d'une version numérique des textes est nécessaire pour un traitement automatique des SNE dans les deux textes choisis à titre d'exemple.

II.1. Conversion des fichiers

Vu que le travail est effectué avec la version 6 du logiciel *Antidote*, nous avons dû convertir les fichiers⁽⁴⁸⁾ *Word* (extension *.docx), en des *textes bruts* (extension *.txt), étant donné que le menu *Antidote* ne s'affiche pas sur *Word*. Les versions ultérieures d'*Antidote* s'intègrent automatiquement à toutes les composantes de Microsoft Office (désormais MS). Il est indispensable de s'assurer aussi bien du type de fichier que de la version du logiciel avec lequel le travail sera effectué.

47 – <http://www.journaldelascience.fr/espace/articles/etoile-luminosite-etrange-surprend-scientifiques-4850> consulté le 30 janvier 2016.

48 – Les deux fichiers correspondent au texte littéraire et à l'article scientifique.

II.2. Antidote

Antidote est un logiciel dans lequel sont implémentés plusieurs dictionnaires, mais également un logiciel correcteur, détecteur des classes grammaticales, des différentes conjugaisons, etc. dans un texte donné. Grâce à l'un de ses menus, *Inspection*, nous pouvons extraire, dans le cas qui nous intéresse, les SN/SNE de chaque texte. Mieux encore, le logiciel distingue les SN principaux des SN secondaires qui se trouvent dans les SP, par exemple, en raison du catégoriseur qui y est implémenté. Après conversion du fichier, nous cliquons sur le bouton *correcteur* qui s'affiche par défaut dès que le texte en format. *txt* est ouvert, et le logiciel *Antidote* s'ouvre avec une fenêtre de trois colonnes :

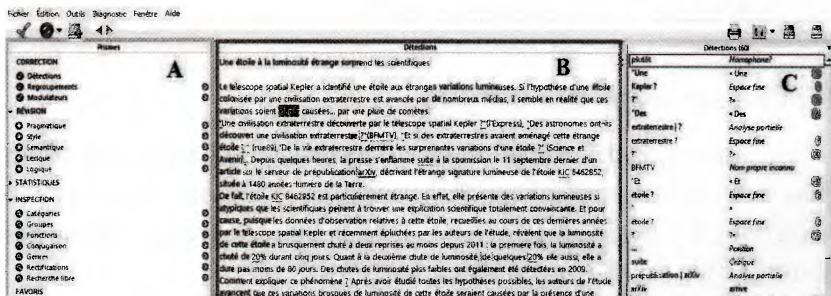


Figure 1: La fenêtre d'Antidote

La première colonne (A) indique les différents menus dont chacun effectue une opération. Quant à la deuxième colonne (B), elle affiche le texte choisi. Enfin, la troisième colonne (C) affiche le résultat de l'opération que nous avons demandé d'effectuer. En cliquant sur *Groupes*, nous sommes face à quatre options (*Nominaux*, *Prépositionnels*, *Verbaux* et *Parenthèses*). Vu que 'Groupes *Nominaux*' est la première option, elle est cochée par défaut. Ainsi, les deux dernières colonnes changent selon les paramètres que nous avons indiqués dans la première colonne. Commence alors la sélection des SN(E) et leur transfert sur un fichier *Word*.

II.3. MS Word

MS *Word* est un «programme de traitement de texte» (Mathias, 2010 : 18) et est utile vu qu'on peut non seulement y saisir un texte, mais aussi y faire des opérations selon le résultat qu'on veut obtenir. Par exemple, si nous avons une liste de mots séparés par une espace et nous voulons que chaque mot soit sur une ligne, il suffit d'aller au menu *Remplacer* et de remplacer l'*espace* par un *^p*, symbole du retour à la ligne, opération dont nous nous passons étant donné que le logiciel *Antidote* arrange chaque syntagme sur une ligne à part. Ce qui nous intéresse, c'est de répartir chacune des composantes des SNE que nous avons extraits d'*Antidote* dans une colonne. Notons que, quand bien même *Excel* serait indispensable en traitement automatique, *Word* l'est autant. C'est pourquoi, avant de passer à *Excel*, certains processus doivent être effectués en amont sur *Word*, tels que le remplacement des espaces par une *tabulation* afin que chaque mot puisse occuper une position dans le tableur *Excel*. La *tabulation* permet à un syntagme, composé d'une suite de mots séparés par une espace, d'être converti en tableau. Autrement dit, chaque mot constituant un syntagme se situera dans une colonne. Par exemple, pour un syntagme comme :

La parure de Madame Forestier

La sera dans une colonne, c'est-à-dire que, selon la *Grammaire distributionnelle*, *la* occupera la position de déterminant (désormais Dét.), *parure*, noyau du SNE (désormais N0), sera dans la position du N, *de*, dans la position de préposition (désormais Prép.), et *Madame Forestier* sera dans la position du N propre (désormais Npr.) :

Dét.	N0	Prép.	Npr.
<i>La</i>	<i>parure</i>	<i>de</i>	<i>Madame Forestier</i>

Il arrive que certains mots soient précédés d'une apostrophe, comme :

l'hypothèse d'une étoile colonisée par une civilisation extraterrestre

Vu que nous n'avons remplacé que les espaces par le symbole ^t, des syntagmes comme *d'élégance* et *d'esprit* seront dans une seule colonne. L'opération de remplacement que nous avons effectuée ne permettra pas à ces mots d'être répartis sur des cellules. Il faut, dans ce cas, refaire la même opération en remplaçant toutes les apostrophes (') par ^t. Ainsi, quand nous transférons les SN/SNE sur *Excel*, chaque mot occupera une cellule et chaque SN/SNE sera placé sur une ligne.

III. MS Excel et son utilité en linguistique

«Excel est un tableur extrêmement puissant. Il ne permet pas seulement de faire des calculs simples, mais il contient une multitude de fonctions et d'autres outils d'analyse permettant des calculs des plus complexes.»⁽⁴⁹⁾. Généralement, ce tableur est utilisé par des informaticiens, des comptables, etc. mais il est tout à fait possible de l'utiliser en linguistique. Avec l'évolution exponentielle de l'informatique, les textes littéraires, journalistiques, etc. sont devenus numérisés et, conséquemment, leur traitement automatique est possible et les opérations par lesquelles les informaticiens, les comptables, etc. passent, sont les mêmes pour un chercheur en linguistique, sauf qu'au lieu de calculer des chiffres, celui-là calculera des mots. Dans le cas qui nous intéresse, nous relèverons les statistiques des N et leurs expansions dans le texte littéraire et l'article scientifique⁽⁵⁰⁾.

III.1. La répartition des SNE sur Excel

Pour la *Grammaire distributionnelle*, un énoncé ou un syntagme est composé de positions. Chaque position est occupée par un mot qui est le résultat d'une sélection parmi les mots d'un paradigme. Cette configuration est transposable dans le tableur *Excel* avec ses

49 – <http://www.excel-online.net> consulté le 06 juillet 2016.

50 – Notons qu'abstraction sera faite des SN minimaux.

colonnes qui correspondraient aux paradigmes et aux cellules des lignes qui correspondraient aux positions. Or les différentes catégories grammaticales ne sont pas réparties dans l'ordre des colonnes. En effet, chaque colonne peut contenir différentes catégories, ce qui ne correspond pas au découpage en paradigmes.

Tableau 1: *Quelques exemples de plus qu'une catégorie grammaticale sur une même colonne*

Colonne 1	Colonne 2	Colonne 3	Colonne 4
Leur (Dé.t.)	finesse (N)	native (Adj.)	
Leur (Dé.t.)	souplesse (N)	d' (Prép.)	Esprit (N)
Leur (Dé.t.)	seule (Adj.)	hiérarchie (N)	

Comme le montre le tableau ci-dessus, nous notons la présence du N, de l'Adj. et du Dé.t. dans la même colonne (2). Dans la colonne 3 également, l'Adj., la Prép. et le N sont dans la même colonne. Il faut donc réaligner les constituants des différents syntagmes afin que chaque colonne reçoive les constituants d'une même catégorie comme l'exemple suivant :

Tableau 2: *Réalignement des constituants des syntagmes nominaux*

Dét.	Adj. antép.	N0	Adj. postp.	Prép.	Dét.	N1
Leur		finesse	native			
Leur		souplesse		d'	Ø	Esprit
Leur	seule	hiérarchie				

Organiser la liste des SNE manuellement prendrait beaucoup de temps. C'est pourquoi certaines étapes qui nous feront gagner du temps doivent être effectuées en amont, la première étape étant l'ajout d'une ligne supérieure qui, à son tour, doit être filtrée. À cet effet, aller au menu *Données*, puis, cliquer sur *Filtrer* est nécessaire afin d'effectuer le tri et pour remettre chaque classe grammaticale dans un même paradigme.

La deuxième étape consiste en le figement de la ligne supérieure⁽⁵¹⁾ en allant au menu *Affichage*, en cliquant sur le sous-menu *Figurer les volets* et en sélectionnant l'option *Figurer la ligne supérieure*.

III.2. La longueur de certaines expansions

L'expansion d'un N donné peut être longue, comme le SP, qui se compose d'une Prép. et d'un SN/SNE. Dans ce cas, il est inutile de répartir les expansions longues sur plusieurs colonnes. Cependant, vu que la tabulation des SNE a été effectuée sur *Word*⁽⁵²⁾, nous avons besoin que chaque SP, chaque REL. et chaque CONJ. soit dans une cellule unique. La 'concaténation' de certaines cellules s'avère, dans ce cas, essentielle. Tout d'abord, nous mettons les expansions longues (SP, REL., CONJ., etc.) dans une nouvelle colonne : nous plaçons les mots introducteurs (Prép., pronoms relatifs, conjonctions, etc.) de chaque catégorie sur la même colonne. Par la suite, nous ajoutons une nouvelle colonne à gauche pour y saisir la formule de concaténation. Certainement, les expansions ne sont pas toutes de la même longueur (par exemple, il y a des SP avec des SN minimaux et d'autres avec des SNE). Il faut donc noter jusqu'à quelle colonne la plus longue expansion prend fin.

Soit l'expansion la plus longue relative au N *hypothèse* :

[l'hypothèse] *d'une étoile colonisée par une civilisation extraterrestre*

Transférés dans Excel, chaque mot occupera sur la même ligne une cellule à part.

51 – La ligne supérieure comporte les dénominations des catégories grammaticales (N, Adj., etc.).

52 – Voir *supra*: §1.3.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		l' hypothèse		d' une étoile	colonisée	par une civilisation	extraterrestre					

Figure 2: Transfert du SNE dans Excel

Dans la cellule (ici C2), ajoutée pour recevoir l'expansion dans sa totalité, nous introduisons la formule de concaténation :

= CONCATENER(D2;» «; E2;» «; F2;» «; G2;» «; H2;» «; I2;» «; J2;» «; K2)

où D, E, F, G, H, I, J et K correspondent aux colonnes remplies par les mots séparés et 2 à la ligne de l'exemple.

SOMME X ✓ f =CONCATENER(D2;" «; E2;" «; F2;" «; G2;" «; H2;" «; I2;" «; J2;" «; K2)												
	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		l' hypothèse ; K2)	d' une étoile	colonisée	par une civilisation	extraterrestre						

Figure3: Introduction de la formule de concaténation

Pour obtenir le même résultat pour les lignes suivantes, il suffit de copier automatiquement la formule en glissant vers le bas le petit carreau noir situé en bas à droite de la cellule C2 jusqu'à atteindre la dernière expansion. Nous obtenons, ainsi, dans la colonne C, les expansions dans leur totalité et à partir de la colonne D, les mêmes expansions réparties sur plusieurs cellules et qui sont désormais inutiles.

III.3. Problème d'effacement des cellules inutiles

Les cellules inutiles doivent être supprimées. Cependant, leur simple suppression entraîne automatiquement le vidage de la colonne de concaténation C dont le contenu est lié, comme résultat de l'opération de concaténation, aux contenus des cellules D et suivantes. Pour conserver le résultat de la colonne C, une manipulation assez simple est requise. Il suffit de sélectionner les cellules concernées, cliquer avec le bouton droit de la souris sur le bord de la sélection, maintenir

le clic enfoncé et faire un glisser-déplacer des cellules d'abord ailleurs (en déplaçant la souris) puis en revenant au même endroit. Un petit menu apparaît alors où il faut cliquer sur le choix Copier ici les valeurs seules⁽⁵³⁾. Commence alors le travail avec le *sous-total*.

IV. Le sous-total

Le menu *sous-total* permet non seulement de gagner du temps, mais aussi d'obtenir des résultats exacts qui ne pourraient probablement pas être obtenus manuellement.

IV.1. Application du *sous-total*

Pour effectuer un *sous-total*, il faut aller au menu *Données*, puis, cliquer sur *Sous-total*. Une petite fenêtre s'affiche avec trois options. Ce qui nous intéresse réellement, ce sont les options n°2 (sélectionner nombre) et l'option n°3 (cocher toutes les catégories) pour obtenir le *total* et le *sous-total* de chaque catégorie qui correspond à une colonne.

IV.2. Le *sous-total*: opération peu efficace

Il est important de vérifier les résultats obtenus étant donné que le *sous-total* ne nous donne pas combien de N sont suivis de 2 Adj. et combien sont suivis de 3 Adj., par exemple. Dans ce cas, certains calculs doivent être effectués manuellement. Dans le résultat que nous avons obtenu, sur 34 N expansés par des Adj., nous obtenons les sous-totaux suivants :

34 N + 1Adj.	2 N + 2Adj.	1 N + 3Adj.
---------------------	--------------------	--------------------

Ces résultats sont erronés dans la mesure où il n'y a que 34 N suivis d'Adj., et les sous-totaux affichent 37 N⁽⁵⁴⁾. Autrement dit, 3 N sont en

53 – <http://www.astuceshebd.com/2011/09/excel-remplacer-une-formule-par-son.html> consulté le 10 juin 2016.

54 – Voir le résultat obtenu ci-dessus : $34 + 2 + 1 = 37$.

surplus. Une fois le calcul manuel effectué, le résultat exact sera comme suit :

32 N + 1 Adj.	1 N + 2 Adj.	1 N + 3 Adj.
----------------------	---------------------	---------------------

V. Statistiques et régularités

Le dernier axe de cet article traitera des statistiques et des régularités relevées après extraction des SN et des SNE, après nettoyage des corpus, notamment la suppression des SN minimaux et de certaines redondances de syntagmes, etc. et après le passage par certaines étapes sur le tableur *Excel*, étapes qui nous ont permis de calculer le nombre des SNE dans les deux textes, mais également de dénombrer chaque type d'expansion pour exposer les structures les plus régulières dans chacun des textes.

V.1. Statistiques

Les statistiques du texte littéraire et de l'article scientifique sont présentées par ordre de fréquence⁽⁵⁵⁾. Les SNE qui sont présents dans *la Parure* sont au nombre de 84. Quant aux SNE qui figurent dans l'article scientifique, ils sont au nombre de 34 seulement. Nous remarquons que le texte littéraire inclut plus de SNE que l'article scientifique.

V.2. Régularités relevées

Les structures régulières quant aux SNE dans *la Parure* sont la structure [N + 1Adj.] au nombre de 32, suivie de la structure [N + 1SP] au nombre de 23. Les structures restantes représentent une minorité. D'un autre côté, nous dénombrons, dans l'article scientifique, 10 structures pour [N + 1SP] et 9 structures pour [N + 1Adj.].

55 – Voir tables n°1 et n°2.

V.3. La disproportion des SNE et résolution du problème de décalage

Il existe un décalage de résultats dans la mesure où le nombre de mots dans le texte littéraire est quatre fois supérieur au nombre de mots dans l'article scientifique, ce qui signifie que les statistiques auxquelles on a abouti sont erronées⁽⁵⁶⁾. Nous avons donc été amenée à diviser le nombre des SNE obtenus dans le texte littéraire par 4⁽⁵⁷⁾ et, il s'est avéré que les SNE dans l'article scientifique sont plus nombreux (34 SNE) que ceux qui sont dans le texte littéraire (21 SNE). Ainsi, les deux structures les plus fréquentes qui figurent dans *la Parure* sont devenues inférieures (8 pour la structure [N + 1 Adj.] et 5,75 pour la structure [N + 1 SP]) aux structures régulières relevées dans l'article scientifique (10 pour la structure [N + 1 SP] et 9 pour la structure [N + 1 Adj.])⁽⁵⁸⁾.

VI. Tableaux

Dans cette section apparaîtront les différents tableaux qui se rapportent à certains éléments que nous avons abordés tout au long de l'article.

Table 1: Nombre des SNE dans le texte littéraire

Structures	Nombre de structures
N + 1 Adj.	32
N + 1 SP	23
Adj. + N	9
N + 1 REL.	7

56 – Le texte littéraire compte 2765 mots et l'article scientifique compte 706 mots. Il y a donc un énorme décalage entre les deux textes, ce qui rend incorrect le résultat obtenu.

57 – En divisant 2765 par 4, nous obtenons le nombre de mots de l'article scientifique (691,25). Donc, pour obtenir un résultat correct, il suffit de diviser les 84 SNE de *La Parure* par 4.

58 – Voir table n°3.

Structures	Nombre de structures
Adj. + N + SP	4
N + Adj. + REL.	3
N + SP + REL.	2
N + Adj. + SP	2
N + SN	2
Adj. + N + 1 REL.	1
N + Adj. + 2 SP	1
N + 3 Adj. + SP	1
N + 2 Adj. + 2 REL.	1
N + REL. + SN + REL.	1
N + CONJ	1
TOTAL	84

Table 2: Nombre des SNE dans l'article scientifique

Structures	Nombre de structures
N + 1 SP	10
N + 1 Adj.	9
Adj. + N	4
N + SN	2
N + SP + Adj.	2
N + 2 Adj.	2
Adj. + N + SP	1
N + loc. Adj. + SP + REL.	1
N + SP + 3 Adj.	1
Adj. + N + Adj. + SP	1
N + Adj. + SN	1
TOTAL	34

Table 3: L'existence de certaines structures dans un texte et leur absence dans l'autre et vice-versa

Structures	<i>La Parure</i>	<i>L'étoile KIC</i>
N + 1 Adj.	8	9
N + 1 SP	5,75	10
Adj. + N	2,25	4
N + 1 REL.	1,75	4

Structures	<i>La Parure</i>	<i>L'étoile KIC</i>
Adj. + N + SP	1	1
N + SN	0,5	2
N + Adj. + REL.	0,75	0
N + SP + REL.	0,5	0
N + Adj. + SP	0,5	0
Adj. + N + 1 REL.	0,25	0
N + Adj. + 2 SP	0,25	0
N + 3 Adj. + SP	0,25	0
N + 2 Adj. + 2 EL	0,25	0
N + REL. + SN + REL.	0,25	0
N + CONJ	0,25	0
N + SP + Adj.	0	2
N + 2 Adj.	0	2
N + loc. Adj. + SP + REL.	0	1
N + SP + 3 Adj.	0	1
Adj. + N + Adj. + SP	0	1
N + Adj. + SN	0	1
TOTAL	21	34

VII. Conclusion

Pour conclure, nous dirons que l'objectif de cette recherche a été d'étudier les préalables de l'analyse de deux corpus différents avec le tableur *Excel*, et nous avons pris, à titre d'exemple, le cas du SNE. Il convient de noter que les logiciels informatiques ainsi que certaines applications, qui se trouvent dans MS, en l'occurrence le *sous-total*, sont loin d'être infaillibles. Le travail manuel demeure essentiel pour tout type d'étude. Par ailleurs, les logiciels informatiques ne sont pas tous conçus pour traiter tout type de sujets en linguistique. Effectivement, pour travailler sur les SN, par exemple, *Antidote* et *Excel* suffisent. Un logiciel tel que *TreeTagger*⁽⁵⁹⁾ ne nous aurait pas permis de faire le travail que nous avons effectué avec *Antidote*. C'est possible, mais y

59 – *Tree Tagger* est logiciel catégoriseur qui peut être implémenté dans d'autres logiciels et qui, grâce à des formules qui contiennent des abréviations des catégories grammaticales, permet à certains logiciels de reconnaître les classes grammaticales dans un texte donné.

recourir nous aurait pris du temps. Nous aurions pu extraire tous les N des deux textes, mais certains N font partie des SN secondaires, comme ceux qui se trouvent dans les SP. Comme nous ignorons d'emblée la nature des composantes de ces SNE, il nous est impossible de saisir toutes les structures dans la requête du logiciel *AntConc*⁽⁶⁰⁾.

Références

- *Antidote HD version 2.1, 2009* *Druide informatique inc.*
- Microsoft Office Excel 2010.
- <http://www.journaldelascience.fr/espace/articles/etoile-luminosite-etrange-su-prend-scientifiques-4850> consulté le 30 janvier 2016.
- Les deux fichiers correspondent au texte littéraire et à l'article scientifique.
- <http://www.excel-online.net> consulté le 06 juillet 2016.
- Notons qu'abstraction sera faite des SN minimaux.
- La ligne supérieure comporte les dénominations des catégories grammaticales (N, Adj., etc.).
- Voir *supra* : §1.3.
- <http://www.astuceshebdo.com/2011/09/excel-remplacer-une-formule-par-son.html> consulté le 10 juin 2016.
- Voir le résultat obtenu ci-dessus : $34+2+1=37$.
- Voir tables n°1 et n°2.
- Le texte littéraire compte 2765 mots et l'article scientifique compte 706 mots. Il y a donc un énorme décalage entre les deux textes, ce qui rend incorrect le résultat obtenu.
- En divisant 2765 par 4, nous obtenons le nombre de mots de l'article scientifique (691,25). Donc, pour obtenir un résultat correct, il suffit de diviser les 84 SNE de *La Parure* par 4.
- Voir table n°3.

60 – *AntConc* est un concordancier utilisé dans la linguistique de corpus, conçu par le Professeur Laurence Anthony.

[XV] FRANÇOIS, J. & GHÉRISSI, Y., (2012), «Pour une linguistique orientée outils: la polysémie du verbe compter et les genres textuels», *Cahier du CRISCO*, n°34.

- *TreeTagger* est logiciel catégoriseur qui peut être implémenté dans d'autres logiciels et qui, grâce à des formules qui contiennent des abréviations des catégories grammaticales, permet à certains logiciels de reconnaître les classes grammaticales dans un texte donné.

- *AntConc* est un concordancier utilisé dans la linguistique de corpus, conçu par le Professeur Laurence Anthony.

[^{XV}] FRANÇOIS, J. & GHÉRISSI, Y., (2012), «Pour une linguistique orientée outils : la polysémie du verbe compter et les genres textuels», Cahier du CRISCO, n°34.