Lidil
Revue de linguistique et de didactique des langues

53 | 2016
Phraséologie et genres de discours

Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : l'apport des arbres lexico-syntaxiques récurrents

Using Recurrent Lexico-Syntactic Trees to Study Semantico-Rhetoric Routines in Academic Writing

Agnès Tutin et Olivier Kraif

Édition électronique
URL : http://journals.openedition.org/lidil/3966
DOI : 10.4000/lidil.3966
ISSN : 1960-6052

Éditeur
UGA Éditions/Université Grenoble Alpes

Édition imprimée
Date de publication : 30 mai 2016
Pagination : 119-141
ISBN : 978-2-84310-326-1
ISSN : 1146-6480

Référence électronique
Agnès Tutin et Olivier Kraif, « Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : l'apport des arbres lexico-syntaxiques récurrents », *Lidil* [En ligne], 53 | 2016, mis en ligne le 01 janvier 2017, consulté le 29 octobre 2020. URL : http://journals.openedition.org/lidil/3966 ; DOI : https://doi.org/10.4000/lidil.3966

Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : l'apport des arbres lexico-syntaxiques récurrents

*Agnès Tutin et Olivier Kraif**

Résumé

Les écrits scientifiques se caractérisent par un sociolecte présentant des propriétés linguistiques spécifiques, notamment sur le plan phraséologique. Nous nous intéressons ici aux routines sémantico-rhétoriques, parfois appelées *patrons*, *tournures*, ou *motifs*, par lesquelles les scripteurs s'inscrivent dans une « communauté de discours ». Après avoir défini plus précisément les propriétés linguistiques de ces routines, notre étude aborde les aspects méthodologiques liés à leur mise en évidence dans une approche de linguistique de corpus outillée. Nous étudions ainsi les résultats d'une méthode fondée sur l'exploitation d'annotations syntaxiques en dépendance. Nous montrons que l'extraction d'arbres lexico-syntaxiques récurrents ouvre des perspectives intéressantes dans le domaine, les résultats extraits étant à la fois moins bruités, plus complets et mieux structurés, ce qui permet de mieux associer des routines telles que [*il est {frappant/intéressant/important} de {constater/noter/observer/voir}*] à des fonctions rhétoriques récurrentes dans le corpus. Quelques exemples de ces routines sémantico-rhétoriques intégrant des verbes de constat sont présentés.

Abstract

Scientific writing is characterized by a sociolect with specific features, and in particular phraseological features. This study is dedicated to semantico-phraseological routines, called in French patrons (viz. "patterns"), tournures *or* clichés*, and by which writers become part of a discourse community. We first define the linguistic properties of a routine and explore methodological aspects for identifying these routines by means of corpus linguistics techniques. We show that the technique of identifying recurring lexico-syntactic trees based on treebanks offers interesting perspectives as results appear simultaneously less "noisy",*

* LIDILEM, Université Grenoble Alpes.

*more complete and better structured. The latter is especially pertinent when examining routines associated with specific rhetorical functions, such as [*il est {frappant/intéressant/important} de {constater/noter/observer/voir}*]. Examples of semantico-rhetorical routines involving stative verbs are also discussed.*

1. Introduction

Les écrits scientifiques, en particulier les articles de recherche, se caractérisent par un sociolecte présentant des propriétés linguistiques spécifiques à tous les niveaux de description linguistique (Swales, 1990 ; Rinck, 2010 ; Tutin & Grossmann, 2014). La phraséologie y est particulièrement présente, sous différentes formes (Tutin, 2014) : 1) à travers les collocations (Gledhill, 2000 ; Pecman, 2004), par exemple des expressions comme *résultats encourageants*, *valider des hypothèses* ; 2) à travers des marqueurs de discours complexes (Tran, 2014), comme les séquences lexicales indiquant la structuration (*dans un premier temps*, *en conclusion*) ou la reformulation (*en d'autres termes*) et les marqueurs de point de vue comme *jusqu'à un certain point*, *en grande partie* (Grossmann & Tutin, à paraitre) ; 3) à travers les routines sémantico-rhétoriques, qui correspondent à des fonctions rhétoriques spécifiques de l'écrit scientifique, par exemple, le patron [{*nous*/*on*/*cet article*} {*reprenons*/*utilisons*}DET {*concept*/*modèle*/*conception*} de X] qui indique le positionnement scientifique par rapport aux pairs (Sandor, 2007 ; Tutin, 2010).

Dans cet article, nous souhaitons principalement explorer les routines phraséologiques du 3e type en réfléchissant à des procédures heuristiques permettant de les mettre en évidence, grâce à des méthodes du TAL exploitant des corpus arborés et le repérage de cooccurrences complexes (Kraif & Diwersy, 2012 ; Kraif & Diwersy, 2014). Nous faisons l'hypothèse que les méthodes basées sur des corpus arborés, en particulier la méthode d'extraction itérative des arbres récurrents, sont très prometteuses pour mettre en évidence les motifs, en particulier par comparaison à des approches purement séquentielles (voir, entre autres, Quiniou et coll., 2012), comme la méthode des segments répétés (Salem, 1986). Nous pensons en effet qu'en s'affranchissant de la linéarité et en exploitant les relations syntaxiques, les méthodes basées sur les corpus arborés permettent l'extraction de motifs plus abstraits[1], plus

1. Par « plus abstrait », nous entendons plus proche d'un traitement sémantique.

adaptés à notre objectif linguistique. Nous appliquerons la méthode aux articles scientifiques de sciences humaines et nous nous intéresserons en particulier aux routines phraséologiques associées aux verbes de constat (*voir*, *constater*, *noter*, *observer*, *remarquer*), déjà étudiés dans le cadre du projet Scientext[2], et particulièrement intéressantes du fait de leur forte dimension dialogique, qui se traduit par des propriétés syntaxiques spécifiques (incises, propositions en *comme*) (Grossmann & Tutin, 2010 ; Grossmann, 2014). Nous souhaitons observer comment cette classe sémantique s'insère dans des motifs lexico-syntaxiques récurrents, afin d'en étudier les fonctions discursives et rhétoriques.

Dans un premier temps, nous préciserons les contours de notre objet d'étude, les routines, ainsi que les méthodologies habituellement mises en œuvre pour les identifier et les caractériser. Dans la section suivante, nous détaillerons la méthode d'extraction automatique des arbres récurrents que nous comparerons à l'approche par segments répétés. Nous présenterons ensuite une extension de la méthode en l'appliquant à la classe des verbes de constat, et nous terminerons par une analyse des résultats de cette méthode, à partir d'un corpus d'articles scientifiques.

2. Les routines : des séquences aux configurations lexico-syntaxiques

2.1. Les routines sémantico-rhétoriques

Les *routines*, parfois aussi appelées *patrons*, *motifs* ou *tournures*, intéressent depuis un certain temps la phraséologie dont le champ d'investigation s'étend maintenant au-delà de la lexicographie vers l'étude des textes et des discours (cf. Legallois & Tutin, 2013). Bien que les notions qui nous intéressent ne soient pas encore parfaitement stabilisées (cf. introduction de ce numéro), nous pensons, comme Legallois (2012), que ces séquences « sont de formidables points d'entrée dans les textes, pour observer des relations intertextuelles ou thématiques » (p. 50). Par l'étude de ces séquences, on cherche ainsi à repérer les routines d'écriture, parfois banales, qui font la spécificité d'un genre et par lesquelles les scripteurs s'intègrent dans une « communauté de discours », pour reprendre les termes de Swales (1990). Ces travaux peuvent relever de l'analyse du discours, par exemple le repérage et l'étude de « routines

2. cf. <http://scientext.msh-alpes.fr/> (consulté en septembre 2015).

discursives » emblématiques des rapports d'éducateurs (Née et coll., 2014 ; ce numéro) ou de marqueurs ayant une fonction textuelle et discursive (Longrée & Mellet, 2013), de l'étude du style (par exemple, le repérage des motifs syntaxiques dans la poésie, Quiniou et coll., 2012) ou du repérage des clichés des romans sentimentaux (Legallois et coll., ce numéro), ou encore des travaux de linguistique appliquée sur les *blocs lexicaux* des écrits académiques débouchant sur des applications didactiques (Biber et coll., 2007). Les approches employées pour mettre en évidence ces séquences se réclament souvent de méthodes d'exploration de corpus inductives (*corpus-driven*), c'est-à-dire sans modèle linguistique à priori (cf. Quiniou et coll., 2012), par opposition aux approches fondées sur corpus (*corpus-based*), qui visent à vérifier des hypothèses linguistiques, bien que dans les faits, les approches soient souvent mixtes[3].

Dans les différentes études réalisées, les routines ou autres séquences présentent généralement les caractéristiques suivantes : 1) elles sont linéaires, avec parfois des « trous » (*gaps*) dans les séquences, par exemple le motif « des * plus * que » repéré dans le corpus de poésie de Quiniou et coll. (2012) ; 2) elles comportent des suites de mots, de lemmes ou de traits morpho-syntaxiques (cf. les motifs d'*itemsets* de Quiniou et coll., 2012) ; 3) elles ne constituent pas nécessairement des constituants syntaxiques classiques, par exemple le bloc lexical à fonction discursive *if we look at* de Biber (2007) ; 4) elles sont caractérisées par une fonction discursive, textuelle, ou rhétorique spécifique.

Pour notre part, nous inclurons les routines sémantico-rhétoriques dans la classe des expressions appartenant à la « phraséologie étendue[4] » (Legallois & Tutin, 2013), qui sont propres à un type de discours (cf. Tutin, 2013). Nous les délimiterons de la façon suivante :

3. Les approches sont souvent mixtes pour au moins deux raisons. D'une part, une sélection des éléments est souvent effectuée au moins en partie manuellement : les extractions automatiques des séquences produisent souvent beaucoup de bruit et un repérage manuel est généralement indispensable. D'autre part, les outils de traitement automatique utilisent des prétraitements linguistiques qui se basent sur des modélisations linguistiques : modèles de catégories syntaxiques, analyse syntaxique de dépendance dans notre méthode ; les corpus ne sont donc pas bruts mais déjà analysés linguistiquement.
4. Cette notion est très proche de la notion de « phraséologie large » chez Bolly (2011).

— sur le plan formel, ces expressions constituent des énoncés récurrents, souvent construits autour d'un verbe ; les expressions nominales ou adjectivales ne sont en principe pas des routines, même si des expressions de ce type peuvent être intégrées à des routines ;
— sur le plan sémantique, les éléments de la routine sont construits autour d'un prédicat et d'éléments remplissant différents rôles sémantiques, comme un agent, un objet ou un lieu ; les routines peuvent se réaliser à l'aide d'un matériel lexical varié mais la structure sémantique reste assez stable ;
— en ce qui concerne les fonctions de ces routines, on observe qu'elles remplissent une fonction discursive et/ou rhétorique spécifique.

Les genres institués et très structurés comme les écrits scientifiques sont friands de ce type de formulations routinières qui peuvent exprimer un ensemble de fonctions assez spécifiques (cf. Tutin, 2014) dont nous donnons quelques exemples :

— fournir une preuve à l'aide d'un fait : *comme nous le voyons sur ce tableau, …*; *nous pouvons observer que…*;
— marquer le contraste ou la comparaison : *Contrairement à Parker (1990), nous…*; *Notre étude diffère de Parker (1990)…*;
— établir une filiation scientifique et académique : *À la suite de Parker (1970), nous…*; *Nous reprenons la définition de Parker (1979)…*; *Notre modèle reprend les travaux de Parker (2010)*;
— définir une problématique : *Notre article traite de la phraséologie scientifique*; *l'objet de notre article est la phraséologie scientifique*.

On relèvera ici la diversité du lexique employé. Le caractère préconstruit de ces expressions relève à la fois de la structure sémantique stable de la formule et de la fonction sémantico-rhétorique associée. Dans le cadre de cet article, nous nous intéresserons particulièrement aux routines construites autour des verbes de constat.

2.2. Quelle méthode heuristique pour mettre en évidence les routines ?

Le repérage de ces routines se fait souvent manuellement. Depuis quelques années toutefois, les outils de la linguistique de corpus offrent

des méthodes automatiques qui présentent plusieurs avantages. Elles permettent d'abord de quantifier les expressions et de donner une assise statistique à l'intuition, mais aussi de faire émerger des motifs parfois insoupçonnés.

Parmi ces approches, la méthode des segments répétés, simple à mettre en œuvre, a souvent été utilisée. Les résultats qu'elle génère comportent un bruit considérable, même si les segments sont filtrés par des seuils de fréquence importants, des seuils de spécificité ou des seuils de dispersion[5]. Les éléments générés ne forment en effet pas nécessairement des constituants syntaxiques ou au moins des suites de mots faisant sens. À titre d'expérimentation, nous avons extrait les segments répétés comportant les verbes *constater*, *noter*, *observer*, *remarquer* et *voir* dans un corpus de 300 articles scientifiques de sciences humaines (30 articles x 10 disciplines), qui apparaissent au moins 15 fois dans au moins 3 disciplines sur 10 (pour limiter le bruit, les signes de ponctuation ont été écartés et la taille des segments varie de 3 à 7 unités graphiques). La méthode génère 30 segments[6] (par exemple, *à voir avec*, *à voir de*, *à voir le*, *aller le voir*, *comme on le avoir voir*, *constater que le*) dont la moitié apparait peu pertinente.

L'observation des résultats montre bien le bruit auquel doit faire face le linguiste qui cherche à repérer les segments pertinents, surtout s'il est à la recherche de routines rhétoriques et discursives propres à l'écrit scientifique. Il est confronté à des informations redondantes (certains sous-segments sont inclus dans d'autres segments, par exemple *de constater que* inclus dans *de constater que le*), à des informations à priori peu pertinentes (par exemple *un* dans *on observe un*), à des informations de sous-catégorisations ou liées au figement peu en rapport avec l'étude pragmatique et discursive (*de constater que*, *donner à*

5. Le seuil de spécificité calculera la spécificité des expressions dans le corpus analysé. Le critère de dispersion permet de vérifier la présence d'un élément dans *n* sous-corpus ou *n* tranches du corpus.
6. Par ordre alphabétique, la liste des segments extraits : *à voir avec*, *à voir de*, *à voir le*, *aller le voir*, *comme on le avoir voir*, *constater que le*, *de constater que*, *de constater que le*, *de noter que*, *de voir le*, *donner à voir*, *être de constater*, *être de constater que*, *être intéresser de noter*, *être voir comme*, *il falloir noter*, *le avoir voir*, *le on observer*, *noter que le*, *nous le avoir voir*, *observer dans le*, *observer que le*, *on constater que*, *on le avoir voir*, *on le voir*, *on ne voir*, *on observer un*, *on pouvoir voir*, *on voir que*, *remarquer que le*.

voir). Pour notre objectif, les éléments suivants apparaissent cependant pertinents :

— la présence du modal *pouvoir* associée au verbe de constat ;

— la structure *(comme) on le voit/l'a vu* tout à fait typique de ces verbes de constat ;

— la présence du *nous/on* marqueur de dialogisme interlocutif, au sens de Bakhtine, dans la fonction de co-constat (cf. 4.4).

Si ces séquences linéaires permettent bien de mettre en évidence certaines routines propres à l'écrit scientifique, des modèles plus abstraits basés sur la syntaxe de dépendance nous paraissent toutefois plus adaptés pour notre objectif, pour plusieurs raisons :

— ils s'affranchissent de la linéarité : ainsi, des suites comme *étonnant résultat*, *résultats étonnants* et *résultats vraiment étonnants* auront la même structure profonde qu'une structure N-Adj, représentée de la façon suivante.

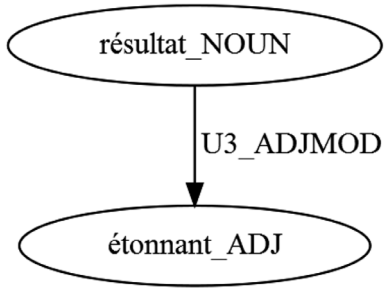

Figure 1. – Schéma syntaxique de dépendance pour le GN *résultats étonnants*.

Cela permet de regrouper davantage d'éléments dans les mêmes classes en évitant la dispersion des résultats.

— ils n'associent que des éléments qui sont en relation syntaxique les uns par rapport aux autres ; cela évite l'extraction de séquences comme *observer que le* puisque dans une grammaire de dépendance, le mot *le* n'est en relation qu'avec le substantif qui suit et non avec la conjonction *que* ; cette restriction permet de réduire drastiquement le bruit ;

— les routines extraites sont proches de routines sémantico-rhétoriques comportant des dépendances syntaxiques pouvant assez

facilement être transformés en relations sémantiques ; des relations syntaxiques profondes peuvent aussi être utilisées pour mettre en équivalence des relations syntaxiques de surface, par exemple la relation profonde objet (*effectuer une analyse*), un passif (*l'analyse est effectuée*) ou un passif réduit (*l'analyse effectuée*) (Hagège & Roux, 2003), ainsi que cela a été effectué par Sylvain Hatier[7] sur notre corpus avec l'analyseur *XIP*.

Dans la suite de cet article, nous présenterons la méthode employée, la comparerons au repérage des segments répétés et effectuerons une analyse des extractions effectuées.

3. Méthode itérative d'extraction des arbres lexico-syntaxiques récurrents

La méthode mise au point est basée sur la notion de « cooccurrence syntaxique » (pour reprendre les termes de Evert, 2007) qui caractérise une association statistique significative reliant deux mots par une relation syntaxique, par exemple (*jouer* →OBJ→ *rôle*). Si l'on s'affranchit de la notion de cooccurrence de surface pour se concentrer sur cette co-occurrence syntaxique, l'observation de segments répétés ou de motifs séquentiels récurrents[8] prend une autre forme, et aboutit au repérage de patrons dans un espace de structures hiérarchiques : il s'agit cette fois d'extraire des sous-arbres récurrents, dont la fréquence est suffisamment élevée pour traduire un degré d'association significatif sur le plan statistique (entre les éléments qui les composent).

Pour mettre en œuvre ce type d'observation, nous avons utilisé l'extraction des lexicogrammes telle qu'elle a été implémentée dans un outil baptisé *Lexicoscope* (Kraif & Diwersy, 2012, 2014). Un lexicogramme est un tableau permettant d'identifier, dans un espace de relation syntaxique déterminé (qui peut couvrir toutes les relations ou seulement un sous-ensemble de celles-ci), quels sont les collocatifs les plus significatifs. Par exemple, pour *constater*, on obtient les collocatifs

7. Par exemple, dans le cas du passif *l'analyse est effectuée* une relation DEEPOBJ-SUBJ a été ajoutée, en post-traitement de *XIP*, entre *effectuer* et *analyse*.
8. C'est-à-dire de suites récurrentes d'éléments pouvant intégrer des places vides.

syntaxiques suivants, triés par degré d'association décroissant (avec la mesure du rapport de vraisemblance notée *loglike*[9], cf. Dunning, 1993) :

Pivot	Collocatif	Relations	cooc	fréq. pivot	fréq. colloc	disp.	loglike
constater_VERB	on_PRON	SUBJ	197	3587	19935	10	832,8242
constater_VERB	pouvoir_VERB	OBJ U3_OBJQUE VMOD ~OBJ ~VMOD	83	3587	47172	10	97,3820
constater_VERB	frappant_ADJ	~ADJMOD	10	3587	113	8	85,5988
constater_VERB	nous_PRON	SUBJ	34	3587	8912	10	82,5248
constater_VERB	être_VERB	AUXIL OBJ ~U3_DE_VMOD U3_OBJQUE ~VMOD VMOD U3_OBJ ~COORDITEMS ~U3_A_VMOD	207	3587	220868	10	80,3868
constater_VERB	différence_NOUN	U3_DEEPOBJ ~NMOD OBJ SUBJ VMOD	23	3587	5473	6	59,7038
constater_VERB	en effet_ADV	U3_ADVVMOD	15	3587	2091	7	53,5993
constater_VERB	intéressant_ADJ	~ADJMOD	10	3587	973	4	42,5441
…							

Tableau 1. – Extrait d'un lexicogramme pour le verbe *constater* (toutes relations confondues).

Outre les statistiques fréquentielles et les mesures d'associations choisies, ce lexicogramme contient des informations sur les relations syntaxiques mises en jeu, ainsi que sur la *dispersion*, qui indique le nombre de sous-corpus où la cooccurrence a été identifiée. Cette donnée est utile pour cibler des phénomènes généraux, partagés par l'ensemble

9. On trouve de nombreuses mesures dans la littérature, comme l'information mutuelle spécifique, le *t-score*, le Dice, ou le χ^2. Le *loglike*, proche du χ^2, présente l'avantage de ne pas surpondérer les évènements de basse fréquence, comme l'information mutuelle spécifique. C'est une mesure fiable dans de nombreux cas de figure (Evert, 2007). Par défaut, on retient les cooccurrences qui obtiennent un score supérieur à 10,83, ce qui correspond à une probabilité inférieure à 1/1000 d'obtenir le tableau de contingence par le seul jeu du hasard.

des sous-corpus étudiés, certaines récurrences pouvant être très saillantes dans une petite partie du corpus (voire un seul document) sans pour autant avoir de portée générale.

L'architecture du *Lexicoscope* permet d'étudier les collocatifs pour des pivots simples, mais aussi pour des arbres, nommés *pivots complexes*, comparables à ce que Rainsford et Heiden (2014) nomment des *keynodes*[10]. À partir de ces fonctionnalités, nous avons développé une méthode d'extraction itérative des arbres récurrents, entièrement automatisée, qui fonctionne de la manière suivante :

1. On part d'un pivot initial (mot simple ou arbre) ;
2. On en extrait le lexicogramme ;
3. Tous les collocatifs dépassant un certain seuil de cooccurrence et de mesure d'association (ici le *loglike*) sont rattachés au pivot, avec la relation concernée, pour former des arbres augmentés ;
4. On réitère l'étape 2 en reprenant ces nouveaux arbres comme pivot ; le processus est répété tant que l'on obtient, pour augmenter les arbres, des collocatifs dépassant les seuils de significativité, et que les arbres extraits n'ont pas dépassé une certaine longueur (paramétrable : dans la suite, la longueur sera fixée à 8 éléments).

La figure 2 ci-après illustre la méthode itérative d'extraction des arbres récurrents. Les éléments extraits par la méthode sont appelés *arbres lexico-syntaxiques récurrents* (nous noterons désormais ALR).

Notons que le *Lexicoscope* permet également de travailler avec des classes lexicales, par exemple $OnNous=(*on, nous*), ou $CONSTAT=(*constater, voir, noter, remarquer, observer*). Les collocatifs appartenant

10. Ces derniers auteurs utilisent toutefois un formalisme de requête un peu différent, celui de TigerSearch (cf. <www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/TIGERSearch/doc/html/QueryLanguage.html>, consulté en 2015), le formalisme utilisé dans le *Lexicoscope* reprenant celle de l'interface Scientext (Falaise et coll., 2012). Dans ce dernier, on exprime les tokens entre <…>, puis les relations de dépendance sous la forme de triplets (relation,gouverneur,dépendant). Par exemple, pour « il est frappant de constater » on a la requête : <l=il,c=PRON,#2>&&<l=être,c=VERB,#1> && <l=frappant,c=ADJ,#3> && <l=de, c=PREP,#4> && <l=constater, c=VERB,#5> ::(ADJMOD,3,5) (OBJ,1,3) (PREPOBJ,5,4) (SUBJ,1,2) (U3_DEEPATTRSUJ,2,3)

à une classe ainsi définie sont assimilés à celle-ci, ce qui permet de regrouper leurs statistiques au niveau de la classe.

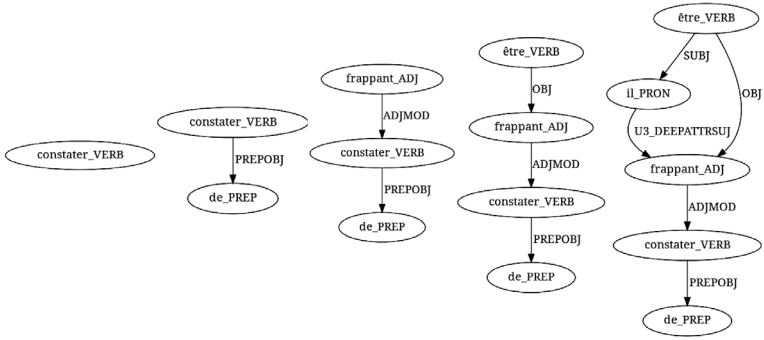

Figure 2. – Exemple d'extraction itérative d'arbres récurrents, en partant du pivot *constater*.

4. Observations : les ALR autour des verbes de constat

Nous avons appliqué cette méthode à un corpus arboré (noté SCIEN) contenant 500 articles scientifiques d'environ 5 millions de mots, rassemblés dans le cadre des projets Scientext et Termith (Hatier et coll., 2014), et annoté syntaxiquement à l'aide de *XIP* (Aït-Mokhtar et coll., 2002). Ce corpus est partitionné en 10 domaines : anthropologie (ANT), économie (ECO), histoire (HIS), géographie (GEO), linguistique (LIN), psychologie (PSY), sciences de l'éducation (SCE), sciences de l'information (SCI), sciences politiques (SCP), sociologie (SOC). Le calcul de dispersion est basé sur cette partition.

Pour notre expérimentation, nous sommes partis de cinq verbes fréquents de constat qui nous intéressent : *constater*, *voir*, *noter*, *remarquer*, *observer*.

Dans un premier temps, les cooccurrences lexicales les plus significatives de ces verbes sont extraites en utilisant un seuil de 25 occurrences, une répartition d'au moins 3 disciplines sur 10 et un rapport de vraisemblance (*loglike*) supérieur à 10,83. Ces cooccurrences binaires sont ensuite étendues avec la méthode des sous-arbres récurrents avec des seuils de 40, 35, 25, 15, 5, 5 et 5 occurrences, correspondant respectivement aux différentes itérations de l'algorithme. Le tableau suivant présente les expressions apparaissant au moins 15 fois.

Expressions (projection linéaire des lemmes des ALR)	**Fréquence**	**Taille de l'expression**
avoir constaté pouvoir	19	3
avoir été observé	35	3
avoir rien à voir	18	4
avoir vu le jour	17	4
comme nous le avoir vu	29	5
comme on le avoir vu	36	5
comme on le voir	38	4
comme on peut voir	17	4
constater que avoir	19	3
constater que être	61	3
donner à voir	70	3
être que constater	61	3
être que observer	56	3
être que voir	57	3
il falloir y voir	17	4
force est de constater	26	4
il être à noter	18	4
il falloir noter	64	3
il falloir remarquer	24	3
il falloir voir	53	3
il intéressant de noter	27	4
laisser à voir	25	3
le avoir vu	30	3
noter que avoir	28	3
noter que être	99	3
nous aller le voir	20	4
nous aller voir	49	3
nous avoir observer	27	3
nous le voir	36	3
nous pouvoir constater	20	3
nous pouvoir remarquer	18	3
observer on être	29	3
observer que être	56	3
on aller le voir	22	4
on aller voir	36	3
on avoir vu	109	3
on constater que être	21	4
on le voir	131	3

on noter que être	15	
on pouvoir constater	53	3
on pouvoir noter	97	3
on pouvoir observer	39	3
on pouvoir remarquer	68	3
on pouvoir remarquer être	15	4
on pouvoir voir	226	3
on pouvoir y voir	34	3
on remarquer avoir	16	3
on remarquer que être	23	4
on venir voir	31	3
on voir pas	22	3
on voit apparaître	17	3
on voit bien	20	3
on voit donc	24	3
on voit ici	19	3
on voit que est	28	4
on voit que pouvoir	19	4
on y voir	29	3
permettre de constater	20	3
pouvoir être observée	26	3
qu'on voit	48	3
que a vu	25	3
que avons notée	28	3
que l'on observe	35	4
que nous observer	18	3
remarquer on avoir	16	3
remarquer que être	64	3
se être voir	25	3
voir le article	26	3
voir le travaux	38	3
voir ce qui	31	3
voir en annexe	15	3
voir le jour	63	3
voir que être	57	3
voir tableau p.	17	3

Tableau 2. – Les arbres récurrents autour des verbes de constat (fréq. >= 15 – Pour une meilleure lisibilité, seuls les lemmes projetés des ALR sont représentés, en se basant sur les réalisations de surface les plus fréquentes).

4.1. Variations syntaxiques

Bien que de très nombreux arbres récurrents correspondent à des expressions linéaires sans variation de l'ordre des éléments (donc à priori facilement identifiables à l'aide de segments répétés ou de motifs séquentiels), on constate qu'ils permettent d'identifier des routines plus exposées aux inversions syntaxiques, quand elles contiennent par exemple des pronoms clitiques, des conjonctions ou des adverbes. C'est le cas de l'exemple suivant (nous noterons désormais <entre chevrons> les expressions projetées qui représentent en fait un ALR) :

<Il falloir y voir> (17 occurrences) :

SCP-12572 : Sans doute **faut-il y voir** un effet de la dissociation, peut-être plus marquée en économie qu'ailleurs, entre une discipline – instrument de connaissance et une discipline – instrument de pouvoir.

SOC-12253 : **Il faut** à nouveau **y voir**, mais pas seulement, cette nécessaire adaptation à la complexité d'un système d'enseignement que l'expérimentation contribue à accentuer encore.

ANT-2214 : La chose est remarquable, car la jeune fille jouit en pays vouté, depuis aussi longtemps qu'on s'en souvienne (**il** n'**y faut voir** aucune trace d'une quelconque « modernité »), d'une grande liberté dans le choix de son époux.

Par ailleurs, notons que même si de nombreuses expressions sont contigües, le recours aux relations syntaxiques permet de franchir facilement les incises et autres insertions, sans grever les calculs.

LIN-12573 : En effet **il est intéressant**, écrit-il, **de constater** que tous les emplois de clé ont systématiquement l'interprétation : 'ouvrir un chemin' ou 'ouvrir un contenant', sous- prédicats de 'ouvrir l'accès à un lieu'.

D'après nos observations, les relations verbe-objet peuvent impliquer des *gaps* assez importants à l'intérieur des ALR. Pour l'ALR <on/nous $CONSTAT DET différence>, on trouve par exemple :

SCE-8002 : En revanche, en ce qui concerne ceux de quatrième et de cinquième années, on **observe** pour les tests relatifs à la compréhension en lecture **une différence** significative en faveur des élèves promus.

LIN-11321 : **On observe**, par rapport aux premières, **les différences** suivantes :

Ces cooccurrences auraient difficilement pu être identifiées sans recourir à la syntaxe, étant donné la distance des unités reliées.

4.2. Mise en œuvre des classes

Si l'on réunit maintenant ces verbes en une classe $CONSTAT et que l'on effectue des extractions, on obtient encore davantage de configurations intéressantes, dont certaines n'étaient pas repérées précédemment.

Les arbres récurrents suivants par exemple intègrent clairement la routine de constat comme un élément d'une argumentation, comme l'indiquent les connecteurs logiques *ainsi* ou *en effet* qui les accompagnent fréquemment :

<On $CONSTAT en effet> (16 occurrences) :

ECO-8516 : **On remarque en effet** que, pour les hommes comme pour les femmes, le diplôme joue d'autant plus que la qualification est élevée.

SCI-10116 : **En effet**, **on voit** bien que la relation de proximité instituée par le talk-show dépasse largement le seul « ethos d'humanité ».

<On peut $CONSTAT ainsi> (15 occurrences) :

SCI-8533 : **On peut voir ainsi** émerger le film interactif, dans lequel la vidéo serait délinéarisée voire documentée à l'aide d'autres ressources.

HIS-246 : **On peut ainsi voir** une représentation de Luther portant un crucifix et la Bible avec, derrière lui, un cygne représentant Hus lisant également la Bible.

SCE-5837 : **Ainsi**, par exemple, a**-t-on pu voir** des listes de mots auxquels on pouvait se référer le jour suivant.

On constate par ailleurs que certains arbres récurrents sont assez bien partagés par les verbes de la classe $CONSTAT. C'est le cas des arbres suivants (les verbes sont listés par fréquence décroissante) :

<il est intéressant de $CONSTAT> : *noter*/*constater*/*observer*/*remarquer*/*voir*

<comme on pouvoir le $CONSTAT> : *voir*/*constater*/*remarquer*/*noter*

<il est à $CONSTAT> : *noter*/*remarquer*/*observer*/*voir*.

À l'opposé, d'autres expressions semblent n'admettre qu'un seul élément de la classe. C'est le cas de : *nous allons le voir* (20) ou *donner à voir* (70). Si ce dernier cas était prévisible, étant donné le figement manifeste de l'expression, c'est moins évident pour *nous allons le voir*, qui est transparente et ne présente pas de figement particulier.

Notons que le calcul des statistiques au niveau de la classe $CONSTAT permet par ailleurs de récupérer, subsumées par certains

arbres récurrents, des variantes peu fréquentes, qui n'auraient pas franchi nos seuils de significativité si elles avaient été considérées isolément.

Par exemple, pour l'arbre <il est à $CONSTAT>, on a 18 occurrences de *il est à noter*, mais une seule occurrence de *il est à observer* et de *il est à remarquer* :

> HIST-3862 : **Il est à remarquer** que le mouvement qui porte les nobles à se hisser dans la direction du capitalisme agro-industriel s'accélère dans les années 1880/90 lorsque les ravages de la crise agraire incitent à définir de nouveaux rôles économiques.
>
> SOC-2183 : Si l'on se rapporte à l'étude statistique menée à partir des questionnaires envoyés aux artistes, **il est à observer** que, pour la diffusion sur les scènes nationales entre 1998 et 2000, les structures qui ont le moins de chances d'être conviées sont les compagnies indépendantes non subventionnées par le ministère de la Culture, puisqu'elles étaient lors de notre enquête 94,1 % à n'avoir présenté aucun spectacle.

Ces observations nous permettent donc d'identifier, pour le même ALR, à la fois la forme canonique d'un motif et ses variantes possibles attestées dans le corpus.

4.3. Expansion des ALR

Nous avons observé précédemment que la classe $CONSTAT construite à priori était cohérente avec de nombreux arbres récurrents extraits du corpus : mais si ces arbres peuvent inclure n'importe quel élément de la classe, on peut supposer que la réciproque est vraie — à savoir que tous les éléments susceptibles d'apparaitre dans l'arbre à la même position sont sémantiquement apparentés. En d'autres termes, on peut supposer que certains des ALR identifiés sont sémantiquement typés, en rapport avec cette classe verbale et seulement celle-ci. Par exemple, si l'on recherche les verbes qui cooccurrent avec l'arbre <il est intéressant de +V> (avec une cofréquence et une dispersion supérieure ou égale à 3), on trouve : {*noter*/*constater*/*observer*/*remarquer*/*comparer*/*relever*/*voir*}. Seul le verbe *comparer* porte un sens spécifique qui l'éloigne, dans une certaine mesure, de la classe $CONSTAT initialement étudiée.

Partant d'un ALR représentant une expression polylexicale déterminée, il est donc possible d'effectuer une expansion de celui-ci pour identifier les routines plus générales dont cette expression n'est qu'une réalisation parmi d'autres. Pour explorer cette hypothèse, à l'issue de notre algorithme d'extraction itérative, nous avons mis en place

une recherche automatique des paradigmes liés aux ALR les plus longs (contenant 4 mots et plus). On trouve une série d'expansions intéressantes :

<il être intéressant de noter> :

<il être intéressant de {noter/constater/observer/remarquer/comparer/relever/voir}>

<il être {frappant/important/intéressant/possible} de noter>

<on $CONSTAT un différence> :

<on $CONSTAT un {différence/effet/tendance}>

<comme on $LE avoir $CONSTAT> :

<comme {on/nous} $LE avoir $CONSTAT>.

Ces résultats permettent d'identifier les ALR dans leur généralité, en vue d'en donner une schématisation abstraite tenant compte des traits sémantiques correspondant aux paradigmes identifiés. Sous l'expression *il est frappant de constater* on trouve par exemple un ALR que l'on pourrait noter (cf. section suivante) :

<il est ADJ^{saillance} de V^{constat}>.

Les structures obtenues associant schéma lexico-syntaxiques et classes de lexèmes sont assez proches de la notion de collostructions élaborée par Stefanowitch et Gries (2003). La différence réside surtout dans la méthode, les schémas d'ALR étant élaborés de façon incrémentale, à partir d'une association lexicale préférentielle puis étendue.

4.4. Des ALR aux routines sémantico-rhétoriques

Les ALR extraits, en particulier à l'aide des classes de verbes de constat, apparaissent tout à fait spécifiques du genre considéré. Pour être des routines sémantico-rhétoriques selon notre point de vue, il faut toutefois que ces expressions soient accompagnées d'un fonctionnement spécifique dans le discours scientifique. L'observation des contextes montre que nombre de ces ALR sont effectivement employés dans des contextes caractéristiques d'interaction avec le lecteur, qu'il s'agisse d'appuyer la validité de la démonstration ou d'attirer l'attention du lecteur sur un phénomène spécifique (cf. aussi Grosmann & Tutin, 2010 ; Grossmann, 2014). Par manque de place, nous ne présenterons ici que les routines les plus caractéristiques. Ces routines suivent une proposition de modélisation esquissée dans Tutin (2010, 2014), pour laquelle

on essaie de mettre en évidence, à la façon des « cadres sémantiques » de la *Frame Semantics* de Fillmore et coll. (2003), des cadres rendant compte à la fois des alternances lexicales et syntaxiques, mais aussi des rôles énonciatifs et sémantiques des participants.

Une première routine impliquant le constat est associée à une fonction dialogique de co-constat, servant à apporter une validité de la preuve en indiquant les faits au lecteur. La routine intègre le lecteur par le biais du pronom inclusif *on*/*nous*, et met aussi en œuvre l'objet et le lieu du constat. Le « lieu » est à ici envisager au sens large, car il peut s'agir aussi bien aussi de preuves amenées dans des figures ou des tableaux, que de portions de texte antérieures ou ultérieures, ou de références externes.

> PSYCHO-12545 : **Comme on peut le voir dans le tableau I**, chez les participantes en comparaison descendante, **il y a une tendance à la diminution de l'identification plus forte lorsque leur cible de comparaison est une autre femme (M = - . 44) plutôt qu'un homme (M = - . 08)**.

> ECO-12602 : L'analogie avec la discrimination du troisième degré est immédiate. **Nous avons déjà vu à la fin de la Section 2 que l'interdiction de pratiquer la discrimination du troisième degré entre marchés à élasticités différentes a des conséquences ambigües sur le bien-être**.

> ECO-9630 : **À partir des graphiques 1 et 2, nous pouvons remarquer que le niveau d'indemnisation de départ joue de manière différente dans les deux réglementations**.

Cette routine peut être modélisée de la façon suivante :

> **(*comme*) agent:*on*/*nous* VCONSTAT:*voir*/*constater*/*observer*/*remarquer*/*noter* objet (lieu:*figure*/*section*/REF. BIBLIO)**
> => apporter une preuve en impliquant le lecteur dans le co-constat

De façon intéressante, l'emploi énonciatif du verbe de constat accompagné des pronoms *nous*/*on* se différencie fortement d'autres genres de discours, comme celui des rapports d'éducateurs (Cf. Née et coll., 2014) où le *nous constatons*/*observons*, renvoyant exclusivement à l'auteur éducateur ou à la communauté des éducateurs, n'inclut pas le lecteur, mais est employé pour apporter un élément de diagnostic qui justifiera la préconisation des rédacteurs.

> « nous constatons que Damien n'utilise plus de grossièretés dans son langage » (exemple de Née et coll., 2014, p. 503).

Une deuxième routine caractéristique implique uniquement le verbe *voir* (employé concurremment à l'abréviation *cf.*). Cette routine de renvoi très typique de l'écrit scientifique parait à priori assez éloignée de la première. Elle inclut toutefois aussi le lecteur (par l'infinitif jussif, proche d'un impératif) et est aussi utilisée pour apporter une preuve de la validité de l'analyse ou du raisonnement. Elle est également associée au même type de lieu que la première routine, qui peut être une référence bibliographique, une section antérieure ou ultérieure du texte, une figure.

> SC-EDU-4812 : Une importante partie du travail consiste (ensuite) à regrouper les quelque 1 500 professions observées en 10 catégories socioprofessionnelles **(voir tableaux en annexe 2)**.
>
> SC-ECO : Les enseignements de la plupart des modèles théoriques convergent pour affirmer qu'une hausse des allocations chômage entraine une aggravation du chômage (**voir** par exemple **LAYARD et alii [1991], MANNING [1993], CAHUC et ZYLBERBERG [1996], PISSARIDES [1990]**).

La modélisation de cette routine est :

> ***Voir*/*cf* lieu:*figure*/*section*/REF. BIBLIO…**
> => Renvoyer le lecteur à une référence, section, etc. Contribue à la validité de la preuve.

Enfin, une troisième routine remarquable, déjà signalée plus haut, vise à signaler au lecteur l'intérêt particulier d'un fait. La routine est introduite par un adjectif de « saillance » et un verbe de constat qui souligne la pertinence d'un fait, comme dans les exemples suivants :

> PSYCHO-11430 : Enfin, **il est intéressant de noter** que, même parmi les enfants qui réussissent les deux épreuves, aucun ne présente un score de réussite plus élevé en production qu'en révision.
>
> SC-POL-5239 : Or, **il est frappant d'observer** que, dès la deuxième semaine d'émeutes, dans l'Ouest notamment, une série de villes qui constituent les lieux d'installation des grandes familles noires a connu des violences [11].

Cette routine est modélisée de la façon suivante :

> ***Il être* AdjSaillance:*frappant*/*intéressant*/*important* de Vconstat:*voir*/*observer*/*noter***
> => Mettre en évidence un fait saillant pour le lecteur.

Le repérage automatique des ALR, associé à l'observation des contextes textuels, permet ainsi de mettre en évidence quelques fonctionnements rhétoriques propres et constitutifs du genre. Cette première esquisse doit bien entendu encore être affinée en prenant davantage en compte le fonctionnement de ces routines dans l'argumentation, en particulier en lien avec les marqueurs de discours.

5. Conclusion et perspectives

Au terme de cette expérimentation, la méthode lexico-syntaxique développée parait donc adaptée pour une procédure heuristique d'identification des routines sémantico-rhétoriques. L'approche n'est pas miraculeuse, mais présente des avantages certains sur les méthodes plus classiques (par exemple, l'extraction de segments répétés, qui se limite à la reconnaissance de cooccurrences lexicales contigües en surface, ou les motifs séquentiels qui peuvent intégrer des « trous »). En s'affranchissant de la linéarité, notre approche extrait davantage d'éléments, se situe à un niveau d'analyse syntaxique plus abstrait et permet de capter des configurations syntaxiques plus complexes.

Elle apparait particulièrement intéressante quand elle manipule des classes de mots. De fait, à la différence des motifs séquentiels ou des segments répétés, les ALR présentent des configurations *structurées*, qui les rendent plus propices aux généralisations : en remplaçant un nom par une classe lexicale, on peut regrouper sous le même arbre un très grand nombre de réalisations possibles. On évite ainsi d'aggraver les deux tendances que nous avons identifiées pour l'extraction des segments répétés : la forte redondance des sorties, qui présentent parfois la même expression sous des dizaines de variantes possibles, et le bruit important (expressions fragmentaires ou non pertinentes). En outre, comme nous l'avons vu, les classes lexicales liées aux ALR identifiés peuvent non seulement être fixées à priori pour servir les objectifs de l'étude, dans une perspective *corpus-based*, mais peuvent également être induites par les données selon une méthodologie *corpus-driven*, tout en restant cohérentes.

Plusieurs extensions intéressantes de la méthode pourraient être envisagées en réintroduisant la dimension linéaire. Tout d'abord, la ponctuation et la position dans la phrase pourraient être mieux prises en compte. On a vu en effet que plusieurs arbres récurrents étaient particulièrement utilisés en incise. Par ailleurs, si l'on veut analyser la fonction

argumentative de certaines de ces routines, la position textuelle ou dans le paragraphe sera également à considérer.

RÉFÉRENCES BIBLIOGRAPHIQUES

AÏT-MOKHTAR, Salah, CHANOD, Jean-Pierre & ROUX, Claude. (2002). Robustness beyond Shallowness: Incremental Dependency Parsing. *Special issue of the NLE Journal*.

BIBER, Douglas, CONRAD, Susan & CORTES, Viviana. (2004). *If You Look At…*: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, *25*(3), 371-405.

BOLLY, Catherine. (2011). *Phraséologie et collocations. Approche sur corpus en français L1 et L2*. Bruxelles : Peter Lang.

DUNNING, Ted. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, *19*(1), 61-74.

EVERT, Stefan. (2007). Corpora and Collocations. Dans A. Lüdeling & M. Kytö (dir.), *Corpus Linguistics. An International Handbook* (article 58). Berlin : Mouton de Gruyter.

FALAISE, Agnès, TUTIN, Agnès & KRAIF, Olivier. (2012). Une interface pour l'exploitation de corpus arborés par des non-informaticiens : la plateforme ScienQuest du projet Scientext. *TAL*, *52*(3), 241-246.

FILLMORE, Charles, JOHNSON, Christopher & PETRUCK, Myriam. (2003). Background to Framenet? *International Journal of Lexicography*, *16*(3), 235-250.

GLEDHILL, Christopher. (2000). *Collocations in Science Writing*. Tübingen : Gunter.

GROSSMANN, Francis & TUTIN, Agnès. (2010). Evidential Markers in French Scientific Writing: The Case of the French Verb *Voir*. Dans E. Smirnova & G. Diewald (dir.), *Evidentiality in European Languages. Empirical Approaches to Language Typology (EALT)* (p. 279-308). Berlin / New York : Mouton de Gruyter.

GROSSMANN, Francis. (2014). Verbes de constat et autres verbes « parenthétiques ». Quel statut dans l'écrit scientifique ? *Arena Romanistica*, *15*, 106-122.

GROSSMANN, Francis & TUTIN, Agnès. (À paraitre). Les adverbiaux polylexicaux d'attitude dans l'écrit scientifique. Dans *Actes du colloque « Approches théoriques et empiriques en phraséologie » des 11 et 12 décembre 2014*. Tübingen : Stauffenburg.

HAGÈGE, Caroline & ROUX, Claude. (2003). Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes. Dans *Actes*

de la conférence « Traitement automatique du langage naturel » (TALN 2013), Batz-sur-Mer.

Hatier, Sylvain, Tutin, Agnès, Jacques, Marie-Paule, Jacquey, Évelyne & Kister, Laurence. (2014). *Catégorisation sémantique des noms simples du lexique scientifique transdisciplinaire*. Communication présentée au Congrès ACFAS « Étude de lexiques à vocation particulière : approches théoriques, méthodologiques, pédagogiques et multidisciplinaires », Montréal.

Kraif, Olivier & Diwersy, Sascha. (2014). Exploring Combinatorial Profiles Using Lexicograms on a Parsed Corpus: A Case Study in the Lexical Field of Emotions. Dans P. Blumenthal, I. Novakova & D. Siepmann (dir.), *Les émotions dans le discours – Emotions in Discourse*. Berlin : Peter Lang.

Kraif, Olivier & Diwersy, Sascha. (2012). Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques. Dans *Actes de la conférence TALN 2012* (p. 399-406), Grenoble.

Legallois, Dominique (2012). La colligation : autre nom de la collocation grammaticale ou autre logique de la relation mutuelle entre syntaxe et sémantique ? *Corpus*, *11*. Disponible en ligne sur <http://corpus.revues.org/2202> (consulté le 28 mars 2016).

Legallois, Dominique & Tutin, Agnès. (2013). Présentation : Vers une extension du domaine de la phraséologie. *Langages*, *189*(1), 3-25.

Longrée, Dominique & Mellet, Sylvie. (2013). Le motif : une unité phraséologique englobante ? Étendre le champ de la phraséologie de la langue au discours. *Langages*, *189*(1), 65-79.

Née, Émilie, Sitri, Frédérique & Fleury, Serge. (2014). L'annotation du pronom « nous » dans un corpus de rapports éducatifs. *Objectifs, méthodes, résultats*. Dans *Actes des Journées internationales d'analyse statistique des données textuelles – JADT 2014* (p. 495-506).

Pecman, Mojca. (2004), *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique* (Thèse de doctorat). Université de Nice Sophia Antipolis.

Quiniou, Solen, Cellier, Peggy, Charnois, Thierry & Legallois, Dominique. (2012). Fouille de données pour la stylistique : cas des motifs séquentiels émergents. Dans *Actes des Journées internationales d'analyse statistique des données textuelles – JADT 2012)* (p. 821-833).

Rainsford, Thomas R. & Heiden, Serge. (2014). Key Node in Context (KNIC) Concordances: Improving Usability of an Old French Treebank. Dans *Actes de la 4e édition du Congrès mondial de linguistique française (CMLF)* (vol. 8, p. 2707-2718).

RINCK, Fanny. (2010). L'analyse linguistique des enjeux de connaissance dans le discours scientifique : un état des lieux. *Revue d'Anthropologie des connaissances*, *4*(3), 427-450.

SALEM, André. (1986). Segments répétés et analyse statistique des données textuelles. *Histoire & Mesure*, *1*(2), 5-28.

SÁNDOR, Ágnes. (2007). Modeling Metadiscourse Conveying the Author's Rhetorical Strategy in Biomedical Research Abstracts. *Revue française de linguistique appliquée*, *12*(2), 97-108.

STEFANOWITSCH, Anatol & GRIES, Stefan. (2003). Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics*, *10*(2), 161-198.

SWALES, John. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge : Cambridge University Press.

TRAN, Thi Thu Hoai. (2014). *Description de la phraséologie transdisciplinaire des écrits scientifiques et réflexions didactiques pour l'enseignement à des étudiants non-natifs. Application aux marqueurs discursifs* (Thèse de doctorat, sous la direction d'Agnès Tutin et Cristelle Cavalla). Université Grenoble Alpes.

TUTIN, Agnès. (2010). Showing Phraseology in Context: An Onomasiological Access to Lexico-Grammatical Patterns in Corpora of French Scientific Writings. Dans S. Granger & M. Paquot (dir.), *Elexicography in the 21st Century: New Applications, New Challenges* (p. 303-312). Louvain : Presses de l'Université de Louvain.

TUTIN, Agnès. (2014). La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques. Dans A. Tutin & F. Grossmann (dir.), *L'écrit scientifique : du lexique au discours. Autour de Scientext* (p. 27-44). Rennes : Presses universitaires de Rennes.