Lidil

Revue de linguistique et de didactique des langues

53 | 2016

Phraséologie et genres de discours

Présentation

Frédérique Sitri et Agnès Tutin

Édition électronique

URL : <http://journals.openedition.org/lidil/3914>

DOI : 10.4000/lidil.3914

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

Date de publication : 30 mai 2016

Pagination : 5-18

ISBN : 978-2-84310-326-1

ISSN : 1146-6480

Référence électronique

Frédérique Sitri et Agnès Tutin, « Présentation », *Lidil* [En ligne], 53 | 2016, mis en ligne le 01 janvier 2017, consulté le 02 novembre 2020. URL : <http://journals.openedition.org/lidil/3914> ; DOI : <https://doi.org/10.4000/lidil.3914>

Présentation

Frédérique Sitri et Agnès Tutin***

1. Introduction

Depuis quelques années, les travaux linguistiques s'intéressent beaucoup aux séquences préfabriquées, « prêtes à parler » ou « prêtes à écrire » de la langue. Pour définir brièvement l'objet qui nous intéresse, nous dirons qu'il s'agit de suites de formes et de lemmes, parfois de catégories, formant des expressions sémantiquement compositionnelles et généralement plus larges que les expressions figées ou phrasèmes classiques. Ces séquences sont généralement caractéristiques de certains « genres institués », comme définis par Maingueneau (2004), et occupent souvent des fonctions discursives et rhétoriques remarquables. Il peut ainsi s'agir de routines spécifiques à l'écrit scientifique, qui renvoient à la démonstration de la preuve (comme *on/peut le voir/constater/observer sur/dans*), ou à des routines discursives propres à des écrits professionnels tels que les rapports éducatifs (comme *X V État en/dans + nom abstrait* ou *X peut Vinf*). Précisons aussi, comme le signalent Née, Sitri et Veniard dans ce volume, que si cet engouement est assez nouveau pour l'écrit « monologique », cet intérêt est déjà bien établi en linguistique interactionnelle avec la notion de « routines conversationnelles » chez certains auteurs comme Coulmas (1981).

Plusieurs facteurs peuvent expliquer cet enthousiasme pour les séquences et formules. Tout d'abord, on observe un intérêt croissant en linguistique de l'acquisition pour le langage formulaire et la préfabrication, notions qui semblent jouer un rôle déterminant dans l'acquisition des langues comme le montrent les travaux de Wray par exemple (2005, 2013). L'intérêt pour la phraséologie, de manière générale, s'est également accru dans plusieurs courants contemporains de la linguistique, à travers les grammaires de constructions où les expressions polylexicales occupent un rôle de premier plan, à côté des modèles fonctionnels

* Modyco-Université Paris Ouest Nanterre La Défense, CNRS.

** Université Grenoble Alpes, LIDILEM.

et des approches de la linguistique de corpus (Legallois & Tutin, 2013). La linguistique outillée, de son côté, a rendu disponibles de nombreux outils permettant d'observer des attractions lexicales ou d'extraire des motifs lexico-syntaxiques sur de grosses masses de données, peu perceptibles à l'œil nu. Ces outils ont permis de dégager des séries d'observables linguistiques, jusqu'ici peu accessibles, qui ont renouvelé les problématiques linguistiques. Enfin, la question du genre (de discours), avec un accès facilité aux textes intégraux et aux outils d'explorations, devient centrale, et accorde une grande importance à ces formules récurrentes qui jouent un rôle de premier plan dans les différents sociolectes.

Il nous apparaît opportun de revenir aujourd'hui sur ces objets phraséologiques dont les définitions restent encore parfois sommaires, afin d'en préciser les contours et les caractéristiques sémantiques, syntaxiques et énonciatives, si l'on souhaite se les approprier dans la description linguistique, en particulier dans une perspective d'analyse du discours. Les aspects méthodologiques associés à cette problématique (méthodes d'extraction, outils d'exploration de corpus) sont également de première importance dans la mesure où les routines extraites sont souvent déterminées par des méthodes quantitatives et des outils de traitement de corpus, dans une démarche de linguistique outillée. Si ces méthodes sont largement utilisées dans les travaux linguistiques de langue anglaise, elles restent encore à promouvoir dans les études sur l'analyse de discours et la phraséologie en français.

Ce numéro de *Lidil* a pour objectif de préciser les liens entre les routines phraséologiques et les genres institués en proposant de réfléchir à la catégorisation des objets phraséologiques (définition, structures, modélisation) et aux méthodes élaborées pour les mettre en évidence dans les corpus. Il interroge également les liens entre routines et genres textuels, en observant les fonctions rhétoriques, les fonctions argumentatives et les fonctions discursives que ces séquences remplissent. Avant d'aborder plus précisément les contributions de ce volume, nous retraçons rapidement la genèse des notions de routine, formule et motif, sans aucune prétention d'exhaustivité, mais dans une volonté de mise en perspective des travaux présentés dans ce volume.

2. Collocations et cadres collocationnels

Depuis une trentaine d'année, on observe un intérêt croissant pour la phraséologie au sens large, principalement dans les pays de langue anglaise, avec l'influence des modèles fonctionnels et des modèles

contextualistes. De nombreux travaux en linguistique de corpus s'inscrivent dans la filiation de linguistes comme Sinclair (1991) : en développant le « principe de l'idiomaticité », ils mettent en avant l'importance du contexte dans le sens lexical. Ce point de vue fait écho aux travaux de linguistes et lexicologues « continentaux » comme Hausmann (1997) qui déclare de son côté que « Tout est idiomatique dans les langues ». Ces linguistes se sont en particulier intéressés à un concept émergent en linguistique, celui de collocation, plus solidement établi maintenant dans la communauté des phraséologues (voir Williams, 2003 ; Tutin & Grossmann, 2002). Ces associations lexicales privilégiées, représentées par des exemples comme *jouer un rôle*, *grièvement blessé* ou *remercier infiniment*, intègrent des structures généralement binaires et compositionnelles, déjà mises en évidence dans les débuts de la phraséologie par Bally avec ses « séries phraséologiques » (1909). L'intérêt pour ces expressions s'inscrit dans la prise en compte croissante de l'usage, des corpus et des mesures quantitatives dans les travaux linguistiques.

La notion de collocation est parfois élargie à la notion de « cadre collocationnel » (Renouf & Sinclair, 1991) qui, au-delà de la binarité, intègre les alternances possibles d'éléments et des « trous » possibles dans les expressions récurrentes¹. Cette notion, comme on le verra plus loin, présente des affinités avec la notion de motif.

Les régularités de ces associations ont aussi été mises en évidence par certains linguistes. Anscombe, par exemple, schématise ces régularités à travers les « matrices lexicales » qui sont des « schémas comportant des unités fixes et des variables linguistiques » (2011, p. 25), particulièrement productifs pour certaines expressions, par exemple les noms composés ou les proverbes.

En bref, des concepts linguistiques sont développés pour rendre compte d'associations lexicales fréquentes et de structures lexico-syntaxiques récurrentes, mais ces objets sont essentiellement formels. Ils n'apparaissent pas liés à des genres spécifiques, des emplois socialement situés ou des fonctions discursives ou rhétoriques précises. C'est ce qui les différencie principalement des notions de formules, motifs ou routines.

1. Pour illustrer cette notion, on peut prendre l'exemple d'un patron de déterminant complexe comme le suivant : *Un * groupe/tas/ribambelle/troupeau de/d'*.

3. Des segments répétés aux motifs et aux routines

Des notions proches de celles de motifs ou de routines ont été introduites par des linguistes s'intéressant aux spécificités de certains genres, à travers les suites récurrentes de mots.

Le terme de « segment répété » est introduit par Lafon et Salem (Lafon & Salem, 1983 ; Salem, 1987), dans le cadre de travaux lexicométriques sur le discours politique. Les segments répétés sont simplement des « suites de formes graphiques non séparées par une ponctuation forte [...] qui apparaissent plus d'une fois dans un corpus de textes ». S'ils constituent un observatoire privilégié de la répétition en discours, et en particulier de la « langue de bois » caractéristique des discours politiques, les segments répétés ne constituent pas en eux-mêmes des séquences signifiantes. L'Inventaire des segments répétés (ISR) constitue, pour reprendre les termes de Lafon et Salem, « le point de départ d'expériences statistiques intéressantes qui dépassent les descriptions lexicométriques traditionnelles parce qu'elles prennent en charge la dimension syntagmatique » (1983, p. 176). Ainsi, l'ISR peut servir de support au repérage de séquences plus ou moins figées (voir entre autres Fiala, Habert, Lafon & Pineira, 1987) et guider, en association avec d'autres méthodes d'exploration, la mise au jour de patrons lexico-syntaxiques plus lâches (voir Née, Sitri & Fleury, 2014).

La notion de segments répétés présente des similitudes avec la notion des « blocs lexicaux » (*lexical bundles*) de Biber (par exemple, 2006) qui reçoivent aussi une définition purement statistique de suites récurrentes de mots². En revanche, Biber propose une analyse fonctionnelle à posteriori de ces éléments, par exemple, dans l'écrit académique, selon leurs fonctions discursive (*If we look at, as well as the range*), modale (*I don't know, are more likely*) ou référentielle (*at the end of, on the basis of*).

La notion de « motif » en textométrie³ est assez récente. Elle s'inscrit dans une approche outillée, ancrée sur des méthodes statistiques, qui

2. [Lexical bundles] « are simply the most frequently occurring sequences of words, such as *do you want to* and *I don't know what*. These examples illustrate two typical characteristics of lexical bundles: they are usually not idiomatic in meaning, and they are usually not complete grammatical structures » (Biber, 2006, p. 134).

3. Ce terme est employé diversement en linguistique, parfois comme synonyme de *patron*, par exemple, pour indiquer des séquences lexico-syntaxiques ré-

cherche à mettre en évidence des spécificités textuelles. Ce sont principalement Longrée, Luong et Mellet qui ont développé cette notion, à partir de travaux sur des textes latins (Longrée, Luong & Mellet, 2008 ; Longrée & Mellet, 2013). L'objectif était au départ une meilleure approche topologique des textes, afin d'observer ce que le motif pouvait apporter « à l'étude de la structuration interne des textes et à [sa] caractérisation au sein d'un corpus contrastif des expressions récurrentes » (Longrée, Luong & Mellet, 2008, p. 734). Le motif est d'abord défini comme une association récurrente d'éléments dans une structure linéaire multidimensionnelle, dont certains éléments peuvent être facultatifs (Longrée, Luong & Mellet, 2008), avec une définition finalement assez proche du cadre collocationnel de Renouf et Sinclair (1991). Dans un second temps (Longrée & Mellet, 2013), la notion est affinée avec la dénomination de « motif textuel » qui représente un sous-ensemble de motifs ayant une fonction structurante ou discursive, par exemple le motif « *quae cum ita sint* » (« les choses étant ce qu'elles sont », « étant donné la situation », « dans ces conditions »), [...] motif, transitionnel et résomptif [... qui] a [...] une fonction argumentative » (Longrée & Mellet, 2013, p. 73). Les motifs ainsi définis sont appliqués aux textes et permettent de montrer le rôle qu'ils remplissent dans l'agencement discursif et textuel. La notion de « motif » est reprise et élargie chez Quiniou, Cellier, Charnois et Legallois (2012) avec la notion de « motif émergent » ainsi que dans la présente contribution de Legallois, Charnois et Poibeu, les patrons repérés prenant à la fois en compte des « trous » dans les séquences, mais aussi des catégories syntaxiques et des traits. La spécificité des motifs par rapport à d'autres genres ou sous-genres est également mise en évidence. Ainsi, Quiniou, Cellier, Charnois et Legallois (2012) repèrent des motifs propres à un corpus de poésie comme « des * plus * que » correspondant aux motifs émergents « des N plus ADJ que ». Dans ces derniers travaux, la prise en compte du genre, ou du sous-genre, apparaît déterminante.

La notion de « routine », utilisée par Née, Sitri et Veniard (2014 ; ce volume) et Tutin (2013 ; Tutin & Kraif, ce volume), semble introduire une nuance. Cette notion est utilisée pour qualifier des séquences lexico-syntaxiques « prêtes à écrire », étroitement associées à des pratiques discursives spécifiques, liées à des communautés bien délimitées,

currente, comme dans les méthodes d'exploration contextuelle, par exemple Ben Hazez, Descles & Minel (2001).

la communauté scientifique ou la communauté des éducateurs par exemple. Ces façons de dire sont étroitement liées aux genres institués (Maingueneau, 2004) et jouent un rôle important dans l'identité sociale des scripteurs. C'est en partie en s'appropriant ces routines que les scripteurs se définissent comme appartenant à une communauté de discours donnée. Il est possible également que la notion de « routines » conduise à envisager le texte du point de vue de sa production dans la mesure où les « routines » observées dans les textes peuvent renvoyer à des routines processuelles facilitant l'écriture (voir Plane, Alamargot & Lebrave, 2010).

4. Routines et analyse du discours

Cette dimension routinière de l'écrit est présente dans des travaux en analyse de discours comme ceux de Branca-Rosoff sur les écrits des peu-lettrés (lettres de soldats de la guerre 14-18 ou comptes rendus de l'époque révolutionnaire par exemple, voir Branca-Rosoff, 1990, 1997). Ces travaux mettent en évidence l'existence de patrons (appelés « formulaires ») qui, en plus de fonctions textuelles, pragmatiques et identitaires, permettent la mise en place de « routines », de « prêts-à-écrire » facilitant l'entrée dans l'écriture des peu-lettrés. On pourrait rapprocher ces caractéristiques des « routines » repérées dans les écrits professionnels (Pène, 1994). Plus généralement, l'analyse de discours est attentive depuis ses origines aux phénomènes de répétition et de redondance, car ils constituent un mode d'accès à des éléments préconstruits, « déjà-là », qu'elle cherche à mettre en relation avec plusieurs niveaux de détermination du discours. Ainsi, les travaux menés au sein du Laboratoire de Saint-Cloud sur les résolutions de congrès syndicaux mettent en évidence, à partir des inventaires de segments répétés et des concordances, plusieurs niveaux de figement : figements en langue mais aussi figements propres au genre, et figements propres au domaine (ou secteur d'activité pour reprendre le terme de Bakhtine), comme le discours syndical, le discours de la CGT par exemple (voir Fiala, Habert, Lafon & Pineira, 1987, par exemple). À l'heure actuelle, c'est le niveau du genre qui apparaît comme le plus pertinent pour rendre compte des fonctionnements discursifs, que le genre soit envisagé — entre autres — plutôt dans une perspective anglo-saxonne (Swales, Halliday), sous l'angle de sa fonction sociale et de sa visée pragmatique, ou bien, plutôt dans la filiation de Bakhtine, dans sa dimension interdiscursive. C'est la raison pour laquelle nous avons voulu, dans ce numéro de *Lidil*, inviter les

auteurs à mettre en regard la description des séquences phraséologiques avec le genre dans lequel s'inscrivent les discours étudiés.

5. Présentation des contributions

Ce numéro 53 de *Lidil* témoigne à la fois du caractère prometteur du domaine défriché, où se rencontrent différentes disciplines, différentes méthodes et différents points de vue et — en retour — de son caractère exploratoire et non encore stabilisé.

Les contributions réunies dans ce volume viennent de la terminologie, de l'analyse du discours, de la stylistique, de la phraséologie et croisent le plus souvent l'expertise dans ces disciplines avec celle de spécialistes de la linguistique de corpus et du TAL (Traitement automatique des langues). Elles portent sur des genres relevant de la sphère politique (vœux présidentiels), professionnelle (comptes rendus de bilan orthophonique, rapports éducatifs, articles scientifiques) ou littéraire (roman policier ou de science-fiction, roman sentimental).

Le caractère non stabilisé des objets phraséologiques étudiés apparaît à travers la diversité des dénominations — mais aussi des catégories — utilisées par les auteurs : motifs (Legallois, Charnois & Poibeau ; Kraif, Novakova & Sorba), routines discursives (Née, Sitri & Veniard ; Brin-Henry & Knittel), routines sémantico-rhétoriques (Tutin & Kraif), formules ou énoncés rituels (Leblanc). Cette diversité d'approches se retrouve aussi dans la diversité des modes de saisie, c'est-à-dire de mise en évidence, de ces unités. Sur ce point, ce numéro de *Lidil* fait apparaître trois grandes tendances, plus complémentaires que réellement rivales : un mode de saisie « linéaire » de type séquentiel, s'appuyant sur la liste des segments répétés ou sur des concordanciers (voir Brin-Henry & Knittel et, dans une certaine mesure, Née, Sitri & Veniard) ; un mode de saisie linéaire non séquentiel basé sur la cooccurrence (voir la méthodologie d'Alceste utilisée par Leblanc ou l'extraction de patrons (ou cadres collocationnels) dans Legallois, Charnois & Poibeau) ; un mode de saisie plus « hiérarchique » et syntaxique basé sur des méthodes qui prennent en compte les relations syntaxiques de dépendance, comme l'extraction d'arbres lexico-syntaxiques récurrents (Tutin & Kraif ; Kraif, Novakova & Sorba). On peut aussi opposer, comme le font Legallois, Charnois et Poibeau, les méthodes inductives, qui partent des listes fournies par le corpus, et les méthodes qui partent de requêtes le plus souvent autour d'un mot pivot (*difficultés* dans Brin-Henry & Knittel ou les verbes de constat dans Tutin & Kraif

par exemple). Bien souvent, il y a association de plusieurs méthodes : par exemple dans Née, Sitri & Veniard, c'est à partir de la liste des segments répétés que sont menées des recherches de cooccurrents ; dans Legallois, Charnois & Poibeu, la comparaison de la liste des segments répétés entre romans sérieux et romans sentimentaux fait apparaître le caractère extrêmement stéréotypé de ces derniers. Dans tous les cas cependant, il est clair que les objets dont il est question ici ne sont pas superposables aux segments répétés : d'une part, ils ne sont pas réductibles à une suite de formes fixes mais constituent une structure ou un schéma flexible dont certaines places peuvent être instanciées par du matériel lexical différent. D'autre part, ces objets ne sont pas seulement des associations récurrentes de type formel : ils se caractérisent par une fonction discursive et/ou textuelle au sein d'un genre de discours. Quelle que soit la méthode pour les identifier, elle fait nécessairement place à un moment ou à un autre à l'interprétation. Ainsi les différentes contributions réunies dans ce numéro caractérisent différemment en fonction de leur domaine les propriétés discursives des séquences identifiées : donner accès à des représentations sous-jacentes des scripteurs (Brin-Henry & Knittel ; Née, Sitri & Veniard) ; caractériser différentiellement un genre par des récurrences formelles (études sur le roman de SF et le roman sentimental) ; actualiser des fonctions rhétoriques (Tutin & Kraif) ou des déterminations discursives propres au genre (Née, Sitri & Veniard).

L'article de **F. Brin-Henry** et de **M.-L. Knittel** porte sur l'emploi du mot *difficulté(s)* dans un corpus de comptes rendus de bilan orthophoniques (CRBO), genre qui obéit à des normes rédactionnelles précises dans la sphère médico-sociale. Se situant dans une perspective terminologique qui prend en compte les déterminations linguistiques et l'analyse des discours, les auteurs postulent que les choix linguistiques révèlent les représentations sous-jacentes des orthophonistes sur les patients et sur leur façon de catégoriser les troubles du langage. La méthodologie, qui utilise le concordancier AntConc, combine une étude syntaxique et une étude sémantique des associations lexicales intégrant le mot *difficulté(s)*. Elle est fondée sur l'inventaire des formes, qui fait apparaître la fréquence de *difficulté(s)*, et sur une analyse des contextes droits croisant étiquetage syntaxique et étiquetage sémantique. L'étiquetage syntaxique met en évidence la richesse et la variété des dépendances de *difficulté(s)* (SP, Subordonnées, Adjectifs) ; le classement sémantique distingue les différents types d'informations sur l'identité des difficultés (par exemple informations relatives à l'envi-

ronnement, au temps, à l'incidence et à l'intensité des difficultés). Le croisement de ces deux critères de classement permet de faire apparaître des oppositions et des variations. Ainsi, on peut distinguer deux types de patrons distincts qui identifient ou spécifient les problèmes selon l'axe du « Quoi » et l'axe du « Où », patrons que les auteurs analysent comme des routines discursives car ils sont liés à des conceptions différentes des pathologies du langage chez les praticiens.

L'article de **J.-M. Leblanc** s'intéresse aux formules rituelles dans le genre des vœux présidentiels, genre du discours politique fortement codifié « constitutif d'un rituel institutionnel et politique ». Essentiellement méthodologique, cet article propose une expérimentation basée sur la méthodologie inductive d'Alceste, laquelle permet d'identifier les grandes classes thématico-sémantiques du corpus. Parmi ces classes, l'auteur s'intéresse ici à la classe qui traite du rituel et des vœux, qui est remarquable car elle est la première à émerger de la classification descendante, la plus importante sur le plan quantitatif, et essentiellement composée d'énoncés rituels. L'auteur soumet de façon récursive cette classe à une nouvelle analyse par Alceste, ce qui permet de faire apparaître plusieurs sous-classes dont la classe des formules marquant l'empathie présidentielle pour ceux qui souffrent dont la forme significative est *pens+er*. En soumettant de nouveau cette dernière classe à un traitement par Alceste, on obtient la sous-classe des énoncés où domine le lexique de la souffrance, représentée par exemple par *comment ne pas penser à ceux qui sont seuls et qui souffrent*. En menant des expérimentations avec d'autres méthodes basées en particulier sur la recherche des cooccurents spécifiques de groupes de forme (ici « penser » et « souffrance »), l'auteur montre que l'utilisation récursive d'Alceste fait apparaître de façon plus économique et plus fiable les énoncés ritualisés, quitte à affiner ensuite l'étude longitudinale par le recours à d'autres outils qui permettent de faire apparaître des variations selon la période ou selon les locuteurs.

É. Née, F. Sitri et M. Veniard proposent quant à elles de faire de la notion de « routines » une notion opératoire pour l'analyse du discours, en la définissant comme l'association d'un patron lexico-syntaxique plus ou moins figé avec des fonctions textuelles et/ou des déterminations discursives propres au genre. La démonstration prend appui sur l'analyse d'un corpus de rapports éducatifs rédigés par des éducateurs dans le champ de l'enfance en danger, dont on admet qu'ils relèvent d'un même genre. Les méthodes d'exploration textométrique (segments répétés, cooccurents) permettent de faire apparaître des patrons lexico-

syntactiques qui sont intrinsèquement liés à la double visée descriptive/évaluative du genre ou à des thèmes propres au discours des éducateurs. Deux exemples sont présentés : le premier « *dans le cadre* + spécification » illustre une routine à base prépositionnelle et renvoie à l'importance du « cadre » chez les éducateurs, tandis que le second « *et peut* + inf » est de nature syntaxique et actualise le thème du « progrès ». Les auteurs indiquent pour finir que la notion de routines ouvre l'analyse du discours sur la prise en compte du processus d'écriture.

Les trois articles suivants conjuguent approches linguistiques et méthodes du TAL (Traitement automatique des langues).

D. Legallois, T. Charnois et T. Poibeau mettent en œuvre la notion de « motifs » sur un corpus de romans sentimentaux. Ayant défini le roman sentimental comme un sous-genre romanesque « constitutivement » lié au cliché, ils présentent une méthode de repérage des patrons lexico-syntaxiques sous-jacents aux clichés effectivement réalisés dans les textes. La méthode des segments répétés, si elle permet de mettre en évidence la part considérable de la répétition et du figement dans le sous-genre « roman sentimental », doit être complétée par la méthode des motifs, qui permet le repérage de patrons moins lexicalement spécifiés car composés d'éléments dont la nature est mixte : formes fixes, lemmes et catégories morpho-syntaxiques. Sont considérés ici comme motifs « les seuls segments présentant une régularité d'ordre lexical [...] et/ou fonctionnel ». L'extraction des motifs utilise deux modes de calcul : la méthode de l'information mutuelle extrait les segments par calcul « endogène » sur le seul sous-corpus de romans sentimentaux tandis que le calcul des spécificités par la loi hypergéométrique compare la fréquence des segments du roman sentimental à celle de deux autres sous-genres, le roman policier et le roman dit « sérieux ». Sont ensuite sélectionnés les motifs révélateurs de clichés. La méthode des motifs permet donc d'extraire de façon inductive des séquences susceptibles de constituer des clichés, définis comme une « restriction du champ expressif d'une thématique ».

L'article de **A. Tutin** et de **O. Kraif** se donne pour objet les routines sémantico-rhétoriques propres aux articles scientifiques, et se veut également essentiellement méthodologique. Les routines sont ici envisagées comme des éléments prédicatifs qui remplissent une fonction discursive et/ou rhétorique spécifique à un genre. L'objectif des auteurs est de montrer l'intérêt, à côté de méthodes « linéaires » comme celles qui s'appuient sur l'inventaire des segments répétés, de modèles plus abstraits basés sur la syntaxe de dépendance et en particulier la méthode

des arbres lexico-syntaxiques récursifs (ALR). Il s'agit d'extraire les collocatifs syntaxiques les plus significatifs statistiquement d'une unité, puis de procéder de façon itérative sur les pivots complexes ainsi obtenus afin de dégager des arbres récurrents. On peut ainsi regrouper sous une même classe des séquences présentant des variations syntaxiques de surface, voire prendre comme pivot de départ une classe sémantique — ici la classe des « verbes de constat ». Sur le plan discursif, on constate que certains de ces ALR sont effectivement associés à des fonctions textuelles ou rhétoriques que l'on peut mettre en relation avec les spécificités du genre. Ainsi pour les verbes de constat, on identifie une fonction dialogique de co-constat qui vise à apporter une preuve en impliquant le lecteur, une routine impliquant le verbe *voir* permettant de renvoyer le lecteur à une référence et qui contribue à la validité de la preuve, une routine visant à signaler au lecteur un fait saillant.

Le dernier article qui compose ce numéro, celui de **O. Kraif**, de **I. Novakova** et de **J. Sorba**, met également en œuvre la méthode des ALR, mais dans une perspective un peu différente puisqu'il s'agit de caractériser les sous-genres romanesques que sont le roman policier (POL) et le roman de science-fiction (SF), au sein d'un corpus plus vaste regroupant six sous-corpus correspondant à des sous-genres romanesques. Étant donné l'objectif de la recherche et le caractère pilote de l'étude qui est présentée, il ne s'agit pas ici de partir d'un mot pôle choisi pour sa pertinence mais de procéder à l'extraction automatique de tous les ALR statistiquement significatifs dans les deux sous-genres considérés. Les premiers résultats présentés par les auteurs font apparaître des récurrences thématiques et stylistiques. Sur le plan thématique, les auteurs identifient, dans le corpus SF, le motif thématique SN + *de cristal* avec des restrictions lexicales sur le SN qui participent à la construction de la « xéno-encyclopédie » propre au genre, et, dans le roman policier, le motif thématique de la localisation qui se fait autour d'une association lexicale fixe que représente la séquence *scène de crime*. Sur le plan stylistique, est mis en évidence, entre autres, le motif stylistique de l'excès, qui se réalise sous la forme SN + *trop* + ADJ dans les deux genres, mais avec des préférences sémantiques différentes (*lèvres trop rouges* dans POL vs *ciel trop bleu* dans SF).

RÉFÉRENCES BIBLIOGRAPHIQUES

- ANSCOMBRE, Jean-Claude. (2011). Figement, idiomatité et matrices lexicales. Dans J.-C. Anscombre & S. Mejri (dir.), *Le figement linguistique : La parole entravée* (p. 17-40). Paris : Honoré Champion.
- BAKHTINE, Mikhaïl. (1979) [1984 pour la traduction]. Les genres du discours. Dans *Esthétique de la création verbale*. Paris : Gallimard.
- BALLY, Charles. (1909). *Traité de stylistique française*. Heidelberg : Winters.
- BEN HAZEZ, Slim, DESCLÉS, Jean-Pierre & MINEL, Jean-Luc. (2001). Modèle d'exploration contextuelle pour l'analyse sémantique des textes. Dans *Actes de Taln* (p. 73-82). Disponible en ligne sur <<http://lalic.paris-sorbonne.fr/PUBLICATIONS/2000-2001/BenHazez/Taln2001.pdf>> (consulté le 31 mars 2016).
- BIBER, Douglas. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam/Philadelphie : J. Benjamins.
- BRANCA-ROSOFF, Sonia. (1990). Conventions d'écriture dans la correspondance des soldats. *Mots*, 24, 21-36.
- BRANCA-ROSOFF, Sonia. (1997). Modèles de locutionnarité et effets de figement dans le discours politique de l'an II. Dans P. Fiala, P. Lafon & M.-F. Piguet (dir.), *La locution : entre lexique, syntaxe et pragmatique* (p. 285-293). Paris : Klincksieck.
- COULMAS, Florian (dir.). (1981). *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*. The Hague : Mouton.
- FIALA, Pierre, HABERT, Benoît, LAFON Pierre & PINEIRA, Carmen. (1987). Des mots aux syntagmes, figements et variations dans la Résolution générale du congrès de la CGT de 1978. *Mots*, 14, 45-87.
- HAUSMANN, Franz Josef. (1997). Tout est idiomatique dans les langues. Dans M. Martins-Baltar (dir.), *La locution entre langue et usage* (p. 277-290). Fontenay : ENS Éditions.
- KRIEG-PLANQUE, Alice. (2009). *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*. Besançon : Presses universitaires de Franche-Comté.
- LAFON, Pierre & SALEM, André (1983). L'inventaire des segments répétés d'un texte. *Mots*, 6(1), 161-177.
- 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.
- LONGRÉE, Dominique, LUONG, Xuan & MELLET, Sylvie. (2008). Les motifs : un outil pour la caractérisation topologique des textes. Dans S. Heiden, B. Pincemin, L. Vosghanian & l'équipe Lexicométrica (dir.), *Actes des*

- 9^e Journées internationales d'analyse statistique des données textuelles – JADT 2008 (p. 733-744). Disponible en ligne sur <<http://lexicometrica.univ-paris3.fr/jadt/jadt2008/pdf/longree-luong-mellet.pdf>> (consulté le 29 mars 2016).
- MAINGUENEAU, Dominique. (2004). Retour sur une catégorie : le genre. Dans J.-M. Adam, J.-B. Grize & M. Ali Bouacha (dir.), *Texte et discours : catégories pour l'analyse* (p. 107-118). Dijon : Éditions universitaires de Dijon.
- NÉE, Émilie, SITRI, Frédérique & FLEURY, Serge. (2014). L'annotation du pronom « nous » dans un corpus de rapports éducatifs. Objectifs, méthodes, résultats. Dans É. Née, M. Valette, J.-M. Daube & S. Fleury (dir.), *Actes des 12^e Journées internationales d'analyse statistique de données textuelles – JADT 2014* (p. 495-506). Disponible en ligne sur <<http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/41-JADT2014.pdf>> (consulté le 29 mars 2016).
- PÈNE, Sophie. (1994). Analyse de postes, bilan d'entretiens, écritures de procédures. Un «prêt-à-écrire» pour inscrire l'activité. *Éducation permanente*, 120, 39-57.
- PLANE, Sylvie, ALARMAGOT, Denis & LEBRAVE, Jean-Louis. (2010). Temporalité de l'écriture et rôle du texte produit dans l'activité rédactionnelle. *Langages*, 177, 7-28.
- QUINIOU, Solen, CELLIER, Peggy, CHARNOIS, Thierry & LEGALLOIS, Dominique. (2012). Fouille de données pour la stylistique : cas des motifs séquentiels émergents. Dans A. Dister, D. Longrée & G. Purnelle (dir.), *Actes des 11^e Journées internationales d'analyse statistiques des données textuelles – JADT 2012* (p. 821-833). Liège. Disponible en ligne sur <<http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Quiniou,%20Solen%20et%20al.%20-%20Fouille%20de%20donnees%20pour%20la%20stylistique.pdf>> (consulté le 31 mars 2016).
- RENOUF, Antoinette & SINCLAIR, John. (1991). Collocational Frameworks in English. Dans K. Aijmer & B. Altenberg (dir.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (p. 128-143). Harlow : Longman.
- SALEM, André. (1987). *Pratique des segments répétés. Essai de statistique textuelle*. Paris : Klincksieck.
- SINCLAIR, John. (1991). *Corpus, Collocation, Concordance*. Oxford : Oxford University Press.
- TUTIN, Agnès. (2013). 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.

- TUTIN, Agnès & GROSSMANN, Francis. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*, 7(1), 7-25.
- WILLIAMS, Geoffrey. (2003). Les collocations et l'école contextualiste britannique. Dans F. Grossmann & A. Tutin (dir.), *Les collocations : analyse et traitement* (p. 33-44). Amsterdam : De Werelt.
- WRAY, Alison. (2005). *Formulaic Language and the Lexicon*. Cambridge : Cambridge University Press.
- WRAY, Alison. (2013). Formulaic Language. *Language Teaching*, 46(3), 316-334.