

KATARZYNA GABRYSIAK

MATRICE LEXICO-SYNTAXIQUE
DE L'ÉCRIT SCIENTIFIQUE EN TANT QUE TYPE
DE DISCOURS SPÉCIALISÉ

A LEXICAL-SYNTACTIC TEMPLATE OF A SCIENTIFIC TEXT
AS A TYPE OF SPECIALIZED DISCOURSE

A b s t r a c t

The paper offers an analysis of lexical-syntactic structures typical of a scientific text, that is a text that follows quite a stable and rigid structure. A corpus-based analysis, achieved through the use of the Scientext corpora, runs across two dimensions. The first dimension is constituted by the subject matter of the text while the other dimension concerns the relation between the author of the text and the recipient. The analysis presented is a two-stage process. At the first stage, lexical-syntactic structures are singled out. The second stage is to assign those structures to the particular parts of the text, such as Introduction, Main body, Conclusion. The final objective is to arrive at a lexical-syntactic template of a scientific text.

Key words: lexical-syntactic structures; semantic motive; key-concepts; collocations.

REMARQUES PRÉLIMINAIRES

L'objectif de cet article est d'exposer des analyses visant l'établissement de la matrice lexico-syntaxique de l'écrit scientifique. Cette étude se situe dans le cadre des travaux réalisés par le groupe de recherche DiSém¹

Dr KATARZYNA GABRYSIAK – professeure adjointe au Département de linguistique romaine de l'Université pédagogique de Cracovie, membre du groupe de recherche DiSém ; adresse de correspondance – courriel : kajgab@interia.pl

¹ DiSém (Discours, Sémantique, Inférence) – groupe de recherches à l'Université pédo-

(Gabrysiak, Hajok, Muryn, Niziołek, Prażuch). Actuellement, le groupe s'occupe de l'analyse linguistique du roman policier. Les premiers résultats ont déjà été publiés (Muryn et al. 2016) et nous ont incités à élargir le champ de recherche à l'écrit scientifique. Nos récentes analyses concernant, entre autres, l'expression de la perception, de la finalité, des structures lexico-syntaxiques, de la relation émetteur-récepteur dans le texte scientifique nous ont permis de supposer l'existence d'une telle matrice. En plus, ce type de texte est construit sur un schéma discursif. Sa structure interne est stable et comporte toujours les mêmes parties textuelles qui sont mises dans un ordre précis : introduction, développement, conclusion ainsi qu'un résumé et une bibliographie. Chaque partie se caractérise par sa propre structure et joue un rôle différent. Évidemment, tous ces traits facilitent les travaux. Nos recherches s'appuient sur la méthodologie du DiSém (Muryn et al. 2016). Nous en empruntons quelques notions fondamentales telles que : motif sémantique, structure lexico-syntaxique, nébuleuse sémantique, matrice lexico-syntaxique. De plus, elles suivent la vision du rapport entre la phrase et le discours admettant que « chaque type de discours possède sa propre organisation de structures sémantiques complexes et il se distingue aussi par le choix de prédicats et d'arguments, la spécification de positions impliquées » (Muryn et al. 2016 : 3). Nous étudions donc l'écrit scientifique en tant que type de discours et plus exactement de discours spécialisé.

OBJET D'ÉTUDE

Du point de vue sociolinguistique, l'écrit scientifique peut être considéré comme une sorte de sociolecte partagé par une communauté de discours se distinguant par un langage spécifique et ayant les mêmes objectifs rhétoriques. Du point de vue purement linguistique, il constitue un type de discours. Étant donné que le concept de discours reste toujours difficile à appréhender, ses définitions se multiplient. Dans l'œuvre de Widdowson, c'est « l'utilisation d'énoncés en combinaison pour l'accomplissement d'actes sociaux » (in : Kramsch 1984 : 10). Benveniste le définit comme « toute énonciation supposant un locuteur et un auditeur et chez le premier l'intention d'influencer l'autre en quelque manière » (Benveniste 1966 : 242). Pour Jaubert, c'est « du langage en situation » (Jaubert 1990 : 22). Kerbrat-Orecchioni le définit

comme « langage mis en action » (in : Bougnoux 1993 : 219), alors que du point de vue de Maingueneau, « le discours n'est pas un objet concret offert à l'intuition, mais le résultat d'une construction [...], le résultat de l'articulation d'une pluralité plus ou moins grande de structurations transphrastiques, en fonction des conditions de production » (Maingueneau 1976 : 16). Malgré une si grande diversification, on peut en extraire quelques points communs à toutes les approches et proposer une définition la plus générale possible, à savoir : un discours est une entité complexe qui reçoit : une dimension linguistique (parce que c'est un texte), une dimension sociologique (parce que c'est une production en contexte), une dimension communicationnelle (parce que c'est une interaction finalisée) (Charaudeau 2002). En s'appuyant sur cette conception de discours, il est possible de constater que l'écrit scientifique est un discours. En effet : c'est un texte produit en contexte dans un but bien déterminé. Comme nous l'avons déjà mentionné, c'est un discours spécialisé. Le problème qui se pose concerne de nouveau sa définition. Cela résulte de la proximité naturelle entre les notions de discours spécialisé et de langue de spécialité. Tout d'abord, cette proximité est due à une relation fondamentale du discours à la langue dont l'existence a été confirmée par de nombreux chercheurs (Charaudeau & Maingueneau 2002). Ensuite, les deux notions évoquent d'une façon explicite le spécialisé ce qui les distingue de la langue et du discours généraux. Néanmoins, l'emploi du vocabulaire de nature lexico-terminologique spécialisé ne suffit pas à définir le discours comme tel. Par conséquent, le discours spécialisé paraît être généralement conçu comme un objet langagier complexe, mettant en jeu un ensemble d'éléments, de nature linguistique caractéristiques « des usages langagiers propres à l'exercice de certaines activités » (Charaudeau & Maingueneau 2002 : 540) et renvoyant donc à des éléments extra-linguistiques, de contenu cognitif notamment. L'une de ces caractéristiques est la langue de spécialité définie comme : « expression générique pour désigner les langues utilisées dans des situations de communication qui implique la transmission d'une information relevant d'un champ d'expérience particulier » (Galisson & Coste 1976 : 511). Il est évident que la langue de spécialité doit se distinguer par la présence du vocabulaire correspondant au domaine de spécialité. Étant donné que l'écrit scientifique peut concerner chaque domaine de spécialité le vocabulaire propre à ces domaines ne peut pas être à la fois un trait distinctif de l'écrit scientifique permettant de le classer comme le discours spécialisé. Alors, il faut qu'il y ait un autre lexique propre à l'écrit scientifique indépendamment du domaine qu'il représente, à savoir « un lexique transdisciplinaire, différent

de la terminologie des domaines scientifiques concernés, un lexique partagé par la communauté scientifique mis en œuvre dans la description et la présentation de l'activité scientifique qui ne renvoie pas aux objets scientifiques des domaines de spécialité mais au discours sur les objets et les procédures scientifiques ». (Tutin 2013 : 27). L'existence d'un lexique transdisciplinaire a été confirmée dans les études menées par le LIDILEM². Nous pensons que l'explication de ce phénomène réside dans la sémantique et plus exactement au niveau sémantico-syntaxique. Par conséquent, nous voulons dégager un ensemble de structures lexico-syntaxique qui soit transdisciplinaire, c'est-à-dire qu'il soit propre à chaque texte indépendamment du domaine scientifique qu'il représente. Cet ensemble est nommé matrice lexico-syntaxique.

CORPUS

Notre étude se fonde sur Scientext se composant de quatre corpus qui sont les suivants :

- le corpus d'écrits scientifiques du français qui contient 4,8 millions de mots ;
- le corpus d'écrits scientifiques anglais comportant 13,9 millions de mots et se compose surtout d'articles de recherche et de communications écrites ;
- le corpus d'écrits universitaires en anglais langue étrangère contenant 1,1 millions de mots et rassemble des travaux d'apprenants universitaires français rédigeant en anglais des textes argumentatifs ;
- le corpus d'évaluations de propositions de communications élaboré par le LIDILEM comporte 520 commentaires évaluatifs de relecteurs pour un colloque de jeunes chercheurs en sciences du langage.

MÉTHODOLOGIE

La matrice que nous cherchons s'extériorise au niveau lexico-sémantique et constitue une réalisation linguistique de la nébuleuse sémantique qui est

² LIDILEM – Laboratoire de Linguistique et de Didactique du Française Langue Etrangère et Maternelle, Université Grenoble 3-Stendhal, <http://lidilem.u-grenoble3.fr/>.

définie comme un schéma mental obligatoire pour le genre³, ne comportant que des éléments prototypiques. Autrement dit, elle est construite autour des concepts propres à un genre donné. La nébuleuse est universelle c'est-à-dire, elle ne dépend d'aucune langue (Muryn et al. 2016). Par contre, la matrice est idiomatique. Elle est construite de structures lexico-syntaxiques qui sont grammaticalement complètes et illustrent des motifs sémantiques. La notion du motif fait l'objet de nombreuses études et reçoit plusieurs définitions. Nous choisissons celle proposée par DiSém (Muryn et al. 2016) d'après laquelle le motif sémantique est un schéma de concepts obligatoires se réalisant dans une situation précise.

Notre analyse se compose de deux étapes : 1) l'extraction des structures LS, 2) leur répartition en différentes parties textuelles. Elle se déroule aussi sur deux niveaux. Le premier niveau se concentre autour du problème présenté, du sujet abordé dans un texte en question. Il s'active surtout dans le développement. Le second s'associe avec le marquage métadiscursif qui fonctionne entre l'auteur et le lecteur. Il est présent dans l'introduction et dans la conclusion. Néanmoins, il apparaît aussi dans le développement. Ce marquage a pour but de présenter un sujet, à conclure, à commenter, etc. (Gabrysiak 2016).

EXTRACTION ET RÉPARTITION DES STRUCTURES LS

L'extraction des structures LS est possible grâce aux concepts clés. Afin de les établir, il convient de se référer aux connaissances préalables concernant l'écrit scientifique. L'ensemble de ces connaissances stockées dans la mémoire s'appelle cadre qui, selon les théories cognitivistes, est un schéma mental accessible à chaque individu lui permettant de mémoriser des connaissances et d'organiser toutes les informations acquises au cours du processus de perception des objets, des situations, des événements. Grâce à ces schémas, l'individu est capable non seulement de reconnaître des concepts déjà assimilés mais aussi de traiter et de comprendre de nouvelles informations (Minsky 1975 ; Schank 1977). La notion de cadre a été aussi développée par Fillmore (Fillmore 1982). Il admet qu'il est impossible de bien comprendre la signification des mots, sans prendre en considération le

³ Vu que DiSém analyse le roman policier, il emploie la notion de genre. Pourtant, toutes les définitions sont vraies aussi pour l'écrit scientifique qui constitue un type de discours. La définition du genre ainsi que la relation entre le genre et le discours ont été présentées dans les travaux du DiSém (Muryn et al. 2016 ; Muryn, Niziołek 2016).

contexte événementiel ou situationnel dans lequel ils se situent. Chaque cadre conceptuel est caractérisé par des rôles thématiques, des propriétés, etc. appelés éléments cadres.

Une fois les structures LS dégagées, on les répartit en différentes parties textuelles : introduction, développement, conclusion, résumé etc. En plus, on les classe selon le niveau auquel elles se rapportent.

QUELQUES EXEMPLES DES LS PROPRES À L'ÉCRIT SCIENTIFIQUE

LES LS DU NIVEAU MÉTADISCURSIF :

Tableau 1 : Motif sémantique : auteur du texte – communauté scientifique – présentation de recherches – lecteur.

Dans	chapitre article communication thèse partie mémoire	nous on	présenter étaier exposer	étude recherche théorie travail	de	N <HUM> N<Institution>	sur	N<ABSTR>
	tout d'abord ensuite enfin puis			analyse modèle approche				
	∅							

Exemples :

- ◆ *Dans la première partie, nous présentons les études de Simon (2007), Perez Flores (2008) et Chiou (2008) sur les cartes de connaissances.*
- ◆ *Ensuite, nous présentons la théorie de Reichenbach (1947) dans laquelle tout temps verbal est défini par trois points.*
- ◆ *On va présenter les travaux de Y. Pesin sur cette question.*

<https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=search>

Tableau 2 : Motif sémantique : auteur du texte – lecteur – annoncer la suite – renforcement de l'argumentation.

Comme	nous on	pronom neutre le ø	voir au présent de l'indicatif, au futur	dans	chapitre article communication thèse maîtrise partie mémoire	proposition
				ø		

Exemples :

- ◆ *Comme nous allons le voir, l'appel à une réduction massive de la population mondiale, en particulier de celle des pays sous-développés, sera une constante.*
- ◆ *L'activité la plus sensible à ce contrôle moral, comme on le verra dans la troisième partie de ce rapport, est l'organisation des sorties scolaires.*

<https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=search>

Tableau 3 : Motif sémantique : auteur du texte – présentation de l'objectif – lecteur.

dans	chapitre article communication thèse partie mémoire	nous on	viser à chercher à vouloir avoir pour but de	Infinitif
tout d'abord ensuite enfin puis				N<ABSTR>
ø				

Exemples :

- ◆ *Dans cette thèse, nous visons l'amélioration de quelques algorithmes en algèbre matricielle rapide.*

- ◆ *Dans notre étude, nous cherchons à tester dans quelle mesure un très gros corpus peut contribuer à la levée des ambiguïtés de rattachement prépositionnel rattachement prépositionnel ambiguïté.*

<https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=search>

Tableau 4 : Motif sémantique : auteur – présentation de la problématique – lecteur.

dans	chapitre article communication thèse maîtrise partie mémoire ∅	Nous on	présenter étaier exposer	Pronom possessif ∅	étude recherche théorie travail	sur	N<ABSTR>
tout d'abord ensuite enfin puis ∅					analyse modèle approche	de	

Exemples :

- ◆ *Nous présentons nos recherches sur les oiseaux marins et sur le réchauffement climatique.*
- ◆ *Enfin nous présentons nos recherches sur le cas français afin d'éprouver la validité de l'hypothèse d'une européenisation de la politique d'intégration française.*
- ◆ *Par la suite, nous exposons nos analyses des données ainsi que leur encodage et leur organisation.*
- ◆ *Dans cette partie, on présente l'étude de la décohérence et la relaxation du qubit.*
- ◆ *On présentera ensuite une étude des effets des propriétés thermophysiques du milieu sur les caractéristiques dynamiques de la thermosonde.*

<https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=search>

Les structures présentées servent à organiser le discours en annonçant la suite ou en rappelant ce qui a déjà été présenté. Elles guident le lecteur dans le plan du texte. Leur valeur métadiscursive résulte de l'emploi des marqueurs discursifs mais elle peut s'exprimer aussi par l'intermédiaire du prédicat. C'est le cas de la LS (4) construite autour de *présenter*. Pourtant, si le sujet n'est plus exprimé par *nous* et en plus, le verbe est à l'indicatif présent,

la LS n'a plus de fonction métadiscursive et elle s'inscrit au niveau de la problématique (7). Dans d'autres cas, même l'effacement du marqueur ne prive pas la structure LS de cette valeur. La présence du marqueur rend la lecture et la compréhension du texte plus facile. Une autre fonction des LS est de déterminer la relation entre l'auteur du texte et la communauté scientifique. En présentant les études d'autres chercheurs, l'auteur du texte partage des connaissances et situe la problématique de ses recherches dans un contexte plus large. De plus, l'auteur peut présenter l'objectif ainsi que la problématique de ses études. Il y a des structures qui l'aident à renforcer l'argumentation en renvoyant le lecteur à une preuve confirmant une thèse (2). En ce qui concerne le placement dans un texte, les LS en question s'activent le plus souvent dans l'introduction et dans la conclusion. La seule différence réside dans le temps grammatical, à savoir dans l'introduction c'est le présent ou le futur de l'indicatif qui y domine tandis que dans la conclusion on observe l'emploi du temps du passé. Même si ces LS s'extériorisent au développement, elles ne constituent pas son trait distinctif.

LES LS DU NIVEAU DE LA PROBLÉMATIQUE :

Tableau 5 : Motif sémantique : auteur du texte – objet scientifique – source – emprunt.

nous on	s'appuyer		sur	méthodologie modèle travail approche étude	de	N <HUM> N<Institution>
	choisir					
	fonder baser	analyse étude recherche	sur			

Exemples :

- ◆ *Dans notre étude, nous nous appuyons sur le classement proposé par Booiij et De Jong.*
- ◆ *Pour évaluer la conformité des prises en charge réalisées, nous avons choisi la méthodologie de Gyssens que nous avons adaptée.*
- ◆ *Nous empruntons la méthodologie du Groupe de Travail de Marrakech qui consiste à mettre en place un processus de priorisation.*

<https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=search>

Tableau 7 : Motif sémantique : auteur du texte – problématique – lecteur.

nous on	traiter parler s'occuper	de	N<ABSTR>
	étudier analyser		
modèle travail approche étude analyse	traiter présenter		

Exemples :

- ◆ *Cette étude traite de la problématique du rythme de la parole ordinaire et emprunte ses exemples à la langue française*
- ◆ *Cette étude présente une esquisse très générale des propriétés linguistiques des titres de faits divers*
- ◆ *Nous étudions le sens du toucher et son rôle dans la manipulation fine*

<https://corpora.aiakide.net/scientext19/?do=SQ.setView&view=search>

Les structures dégagées se trouvent le plus souvent dans l'introduction. Elles servent à définir la problématique de recherche de l'auteur du texte. Lorsqu'elles apparaissent dans le développement elles peuvent présenter aussi la problématique des études d'autres chercheurs. De plus, la LS 5 demeure l'un des traits distinctifs de l'écrit scientifique. Elle exprime la relation entre l'auteur du texte et la communauté scientifique. Tutin l'appelle la filiation scientifique (Tutin 2013).

REMARQUES FINALES

Nous avons exposé les premiers résultats de nos analyses cherchant à déterminer la matrice lexico-syntaxique de l'écrit scientifique en tant que type de discours spécialisé. Selon nous, cette matrice constitue une illustration de la nébuleuse sémantique qui est universelle, à savoir elle ne dépend d'aucune langue. Par contre, la matrice est idiomatique et diffère en fonction d'une langue choisie. Pour le moment, nous nous concentrons sur les structures lexico-syntaxiques propres à la langue française. Néanmoins, notre but éloigné est de dégager une telle matrice aussi pour le polonais et l'espagnol ce qui faciliterait le travail aux traducteurs et aux chercheurs rédigeant des articles en langue étrangère.

BIBLIOGRAPHIE

- Benveniste Emile, 1966, *Problèmes de linguistique générale*, T.1, Paris, Gallimard.
- Bougnoux Daniel, 1993, *Sciences de l'information et de la communication*, Paris, Larousse.
- Fillmore Charles, 1982, « Frame semantics » [in :] *Linguistics in the Morning Calm*, Seoul, Hanshin Publishing Co.
- Gabrysiak Katarzyna, 2016, « Structures rhétorico-lexico-syntaxiques dans l'écrit scientifique ». *Neophilologica*, 28, Katowice, Wydawnictwo Uniwersytetu Śląskiego, 61-67.
- Galisson Robert, Coste Daniel, 1976, *Dictionnaire de didactique des langues*, Paris, Hachette
- Jaubert Anna, 1990, *La lecture pragmatique*, Paris, Hachette.
- Kramsch Claire, 1984, *Interaction et discours dans la classe de langue*, Paris, Hatier.
- Maingueneau Dominique, 1976, *Initiation aux méthodes de l'analyse du discours*, Paris, Hachette.
- Maingueneau Dominique, Charaudeau Patrick, 2002, *Dictionnaire d'analyse du discours*, Paris, Seuil.
- Minsky Marvin, 1975, *A Framework for Representing Knowledge. The Psychology of Computer Vision*, P. H. Winston (ed.), McGraw-Hill.
- Muryn, Teresa et al., 2016, « La Matrice rhétorico-lexico-syntaxique du roman policier » [in :] *Sbornik statej po itogam mezdunarodnoj konferencii « Āzyk i dejstvitel'nost' : naučnye čteniâ na kafedre romanskih âzykov »*, Moskva, Mpgu, 191-199.
- Muryn Teresa et al., 2016, « Scène de crime dans le roman policier : essai d'analyse lexicosyntaxique », [in :] Actes du CMLF2016 : <http://dx.doi.org/10.1051/shsconf/20162706007>
- Muryn Teresa, Niziołek M. 2016, «Pour une analyse phraséologique du roman policier», [in :] *Fraseologia, variacion y traducción*, Frankfurt am Mein, Peter Lang.
- Schank Roger, Abelson Robert, 1977, *Scripts, plans, goals, and understanding: An Inquiry into human knowledge structures*, New York, L. Erlbaum Associates distributed by the Halsted Press Division of J. Wiley and Sons.
- Tutin Agnès, Grossmann Francis, 2013, *L'écrit scientifique : du lexique au discours*, Rennes, Presse Universitaire de Rennes.
- <http://scientext.msh-alpes.fr/scientext-site/spip.php?article9> (DW : 2017).

MATRICE LEXICO-SYNTAXIQUE DE L'ÉCRIT SCIENTIFIQUE
EN TANT QUE TYPE DE DISCOURS SPÉCIALISÉ

R é s u m é

L'étude présentée dans cet article a pour objectif de dégager les structures lexicosyntaxiques propres à l'écrit scientifique. S'appuyant sur la méthodologie du groupe de recherche DiSem, l'auteur définit les principales notions indispensables pour pouvoir procéder à l'analyse qui se compose de deux étapes. Après avoir dégagé les structures en question étalées sous forme de tableaux suivies d'exemples tirés du corpus, on passe à leur distribution en fonction des parties textuelles où elles apparaissent. La dite analyse fait partie d'un projet dont le but final est de déterminer la matrice lexicosyntaxique de l'écrit scientifique en tant que type de discours spécialisé.

Mots-clés : structure lexicosyntaxique ; motif sémantique ; concept clé ; collocation.

MATRYCA LEKSYKALNO-SKŁADNIOWA TEKSTU NAUKOWEGO
JAKO TYPU DYSKURSU WYSPECJALIZOWANEGO

Streszczenie

Przedstawione w artykule badania mają na celu opracowanie matrycy leksykalno-składniowej tekstu naukowego. Autorka odwołuje się do metodologii grupy badawczej DiSem, definiując jej podstawowe pojęcia, takie jak: motyw semantyczny, struktura leksykalno-składniowa, mgławica semantyczna, matryca leksykalno-składniowa. Omawiana analiza składa się z dwóch etapów: wydobycie struktur leksykalno-składniowych, a następnie ich przypisanie do różnych części tekstu naukowego: wstęp, rozwinięcie, zakończenie itp. Struktury przedstawione są w formie tabeli i zilustrowane przykładami zaczerpniętymi z korpusu badawczego.

Słowa kluczowe: struktura leksykalno-składniowa; motyw semantyczny; koncepty klucze; kolokacja.