

IZABELLA THOMAS
ANASTASIA GALMICHE

MODÉLISATION
DU CONTEXTE DES LEXIES SPÉCIALISÉES
EN VUE DE L'ÉLABORATION D'UN SYSTÈME D'AIDE
À LA RÉDACTION SCIENTIFIQUE
DANS LE DOMAINE BIOMÉDICAL

MODELING OF THE CONTEXT
OF TERMINOLOGICAL UNITS FOR THE NEEDS
OF A SCIENTIFIC WRITING AID TOOL IN THE BIOMEDICAL FIELD

Abstract

In this paper we propose a modeling of contextual information around a terminological unit, for the needs of a scientific writing aid tool in the biomedical field. We focus more specifically on the modeling of significant phraseic contexts that we formalize as semantically characterized argument patterns. This modeling is based on a large corpus of biomedical scientific articles and relay on semantic types specified in a domain ontology, *Unified Medical Language System*.

Key words: terminology; term; scientific writing; scientific writing aid tool; context; modeling; ontology; corpus.

Dr IZABELLA THOMAS est maîtresse de conférences HDR en Linguistique et Traitement Automatique des Langues et ANASTASIA GALMICHE est chercheuse associée au Centre de Recherches Interdisciplinaires et Transculturelles CRIT-Tesnière EA 3224 à l'Université de Bourgogne Franche-Comté; adresse de correspondance : Université Franche-Comté, UFR SLHS, 30 rue Mégevand, 25030 Besançon Cedex, France; courriel : izabella.thomas@univ-fcomte.fr

Une très grosse partie de l'activité scientifique
est consacrée non pas aux manipulations
mais à la mise en forme écrite des résultats :
un chercheur dépense la moitié de son temps à écrire (...)
Ce travail d'écriture, pendant lequel il est assis à son bureau,
joue un rôle tout à fait déterminant
dans la construction de la science.

Jacobi 1993

1. INTRODUCTION ET CONTEXTE

L'impact comme la qualité d'un travail universitaire de recherche sont étroitement liées à la diffusion internationale de ses résultats. Cependant, pour de nombreux chercheurs francophones, la barrière linguistique constitue un obstacle important, confrontés qu'ils sont à la nécessité de publier en langue anglaise. Ce fait est d'autant plus vrai dans le champ des sciences dites « exactes », et notamment dans le domaine biomédical, dans lequel la publication en anglais est devenue incontournable, pour des raisons d'accessibilité des résultats à la communauté internationale des chercheurs et de notoriété des revues.

Pour cerner les difficultés qu'éprouvent les professionnels de Santé non-anglophones rédigeant des articles scientifiques en anglais, nous avons mené une enquête auprès d'experts en rédaction médicale ainsi que de médecins, chercheurs et traducteurs affiliés au Centre Hospitalier Régional Universitaire de Besançon d'une part, et à la Faculté de Médecine de l'Université de Franche-Comté d'autre part¹; l'objectif ultime de cette démarche étant de concevoir un logiciel répandant au mieux à leurs besoins². Nous avons interrogé ces professionnels sur les problématiques liées à la rédaction en anglais spécialisé, sur leurs habitudes concernant l'utilisation des outils d'aide à la rédaction et sur le type d'outils dont ils souhaiteraient disposer pour les soutenir dans le processus d'écriture en anglais biomédical.

Concernant l'usage de la langue, les personnes interrogées n'ont pas de difficulté majeure à connaître les équivalents anglais de terminologie de leurs domaines (61,5%). La problématique qui ressort de la consultation est

¹ Enquête diffusée en ligne entre 01.12.2015 et 31.01.2016 ; l'ensemble de résultats, basés sur 68 réponses, peuvent être consultés dans Thomas 2016.

² SARS, *Système d'aide à la rédaction scientifique (SARS) dans le domaine biomédical* (2015-2017), projet soutenu par le Conseil Régional de la Franche-Comté, en collaboration avec la structure fédérative de recherche FED 4234.

plutôt celle de la ‘contextualisation’ d’un terme (70%), à savoir de la difficulté de trouver/choisir des mots exacts pour l’accompagner dans un contexte phrastique donné.

Un autre éclairage donné par l’enquête concerne l’utilisation des outils d’aide à la rédaction. Sans surprise, l’ère de dictionnaires papier, qu’ils soient généralistes ou spécialisés, semble révolue³. Les utilisateurs se tournent plus volontiers vers les moteurs de recherche, surtout pour trouver des exemples « d’utilisation d’un terme ou d’une expression dans d’autres articles déjà publiés » dans leurs domaines (88%). Cependant, le recours à des outils en ligne n’est pas encore généralisé, et lorsque l’on pose la question des principaux manques ou défauts de ces outils, les chercheurs citent le plus fréquemment :

- ◆ leur manque de spécificité (outils non-adaptés au domaine ; problème de traduction lorsque la terminologie est trop spécifique ; outils intégrant trop peu l’anglais scientifique) ;
- ◆ leur manque d’adaptation au contexte de la phrase ;
- ◆ la perte de temps résultant du recours à des outils multiples, mais pas vraiment adaptés ;
- ◆ le défaut de simplicité d’utilisation et de convivialité.

Notre démarche étant résolument centrée sur les utilisateurs et leurs besoins, nous avons donc projeté d’élaborer un outil d’aide à la rédaction d’article scientifique dans le domaine biomédical sous la forme d’un moteur de recherche de termes en contexte, composé de deux parties complémentaires :

- ◆ une partie ‘dictionnaire’, qui affiche un terme (entrée du dictionnaire) avec ses différentes relations contextuelles préétablies et validées ;
- ◆ une partie ‘moteur de recherche’, qui affiche des exemples de phrases qui émanent du corpus, illustrant les différentes relations contextuelles d’un terme.

La question spécifique que nous abordons dans le présent article est celle de la méthodologie du recensement et de la modélisation de l’information contextuelle, qui prend en compte les besoins exprimés par les rédacteurs scientifiques. Du point de vue des utilisateurs, l’information doit être adaptée le plus étroitement possible au domaine, fiable quant aux sources utilisées et accessible par les professionnels rédigeant qui ne possèdent que peu de connaissances en linguistique. De ce fait, elle doit faire appel à des

³ La plupart des répondants affirme n’utiliser aucun dictionnaire papier, qu’il soit monolingue (69,49%) ou bilingue (59,09%).

concepts du domaine communément partagés et éviter un métalangage linguistique trop complexe. Finalement, elle doit être présentée de manière à la rendre facilement acceptable et manipulable par les utilisateurs.

La méthodologie que nous avons conçue se propose de répondre à l'ensemble de ces besoins. Pour l'expliquer, nous nous appuyons sur le terme 'patient', l'une des lexies spécialisées les plus fréquentes et les plus transversales dans les différents sous-domaines du secteur biomédical. Nous proposons d'abord un bref état de l'art autour de la question de contextualisation du lexique spécialisé et de sa représentation (section 2). Nous y présentons aussi quelques outils existants dans le domaine de la rédaction scientifique.

La section 3 expose la méthodologie que nous avons adaptée pour modéliser un type d'information contextuelle : les contextes phrastiques récurrents autour des lexies spécialisées. Notre démarche est basée sur un corpus d'articles scientifiques dans le domaine biomédical, *Scientext – Ecrits scientifiques en anglais*⁴ et les Types Sémantiques d'une ontologie du domaine biomédical, *Unified Medical Language System (UMLS)*⁵ développée par la National Library of Medicine.

Une conclusion et les perspectives du présent travail sont données dans le dernier chapitre.

2. TERMES EN CONTEXTE ET OUTILS D'AIDE À LA REDACTION SCIENTIFIQUE

Les termes sont caractérisés par « une syntagmatique restreinte (cooccurrences et commutations dans les limites d'un domaine spécialisé) » (Lerat 1995 : 52), ce qui rend possible l'énumération de leurs contextes significatifs. Ces contextes sont souvent définis en termes de cooccurents ou de collocations (Volanschi et Kübler 2010 ; Kübler et Pecman 2012 ; L'Homme 2013), de patrons sémantico-syntaxiques (Yan et Hatier 2016), ou de cadres sémantiques pour caractériser les termes de nature prédicative (Wandji,

⁴ Corpus élaboré par l'équipe LiCorn de l'Université de Bretagne Sud (Geoffrey Williams, Chrystel Millon), qui contient 7 564 textes (35 244 378 mots). Les textes proviennent de la maison d'édition indépendante BioMed Central et portent exclusivement sur la biologie et la médecine. Disponible en ligne : <https://corpora.aiakide.net/scientext18/?do=SQ.setView&view=corpora>, accédé le 01.06.2018.

Tous les exemples de notre article proviennent de ce corpus.

⁵ <https://www.nlm.nih.gov/research/umls/>, accédé le 05.06.2018.

L'Homme, Grabar 2013 ; L'Homme 2016). Les dictionnaires les plus complets décrivant les différentes propriétés des lexies spécialisées sont ceux développés par l'Observatoire Sens-Texte à l'Université de Montréal autour de l'équipe de Marie-Claude L'Homme, DiCoInfo⁶ et DicoEnviro⁷. Bien qu'au fur et à la mesure de leur développement, ces dictionnaires évoluent vers une présentation de l'information contextuelle plus accessible aux utilisateurs non-linguistes (L'Homme 2014), leur vocation lexicographique de dictionnaires d'apprentissage implique une structuration encore trop 'riche' pour les besoins de l'outil que nous envisageons. Enfin, il est important de mentionner deux projets français, ARTES⁸ et Scientext⁹ qui ont grandement contribué à différents aspects de l'analyse de textes scientifiques, par la mise en ligne des bases de données et de corpus d'écrits scientifiques, permettant aux utilisateurs de différents types de consultations. Cependant ces outils sont encore trop complexes et trop peu spécialisés pour s'adresser directement aux professionnels de la Santé.

Du côté des applications, même si on trouve une documentation conséquente sur la rédaction biomédicale et plus généralement, la rédaction scientifique, il n'existe que très peu de logiciels qui lui sont expressément dédiés. Les deux logiciels destinés à soutenir la rédaction biomédicale, AMADEUS (Aluisio et al. 2001) et COBWEB (Barnes et al. 2015) proposent respectivement une aide à la définition de stratégies rhétoriques et à la structuration du contenu, et ne concernent pas les problématiques purement langagières. Le logiciel SWAN (Scientific Writing AssistaNt; Lebrun 2011), quant à lui, est un logiciel de post-édition qui parcourt le texte déjà rédigé pour identifier et corriger d'éventuels problèmes liés à la rédaction de textes scientifiques de tout domaine. (Kinnunen et al. 2012).

La problématique d'aide informatisée à la rédaction scientifique dans les domaines du secteur biomédical de recherche reste encore un terrain à explorer.

⁶ <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi?ui=fr>, accédé le 05/06/2018.

⁷ <http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search.cgi>, accédé le 05/06/2018.

⁸ <https://art.es.eila.univ-paris-diderot.fr/>, accédé le 30/05/2018.

⁹ <http://scientext.msh-alpes.fr/scientext-site/spip.php?article1>, accédé le 30/05/2018.

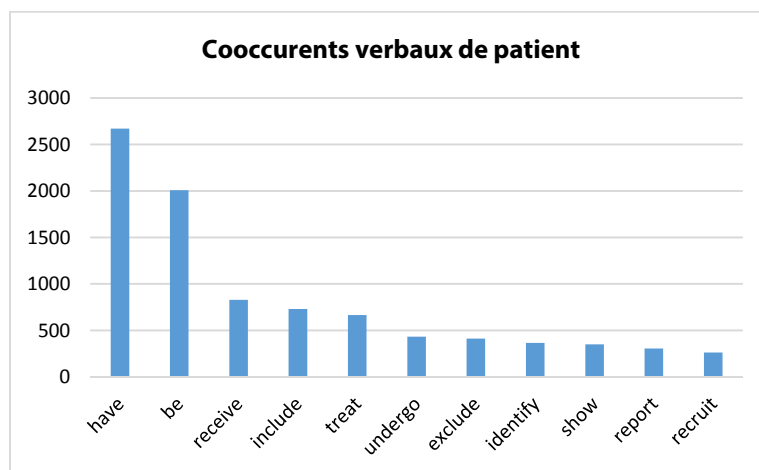
3. METHODOLOGIE D'ETABLISSEMENT DE SCHEMAS ACTENTIELS

Pour expliquer notre méthodologie de contextualisation de lexies spécialisées, nous prenons comme exemple le terme *patient*, qui fait partie du Lexique Scientifique Transbiomédical¹⁰, à savoir une liste de près de 800 termes simples établis comme les plus fréquents dans un ensemble significatif de sous-domaines du secteur biomédical, à travers l'analyse des articles scientifiques. L'information que nous voulons recenser concerne les contextes phrastiques de l'utilisation de cette lexie, que nous modéliserons sous forme de schémas actantiels, caractérisés sémantiquement à l'aide de Types Sémantiques provenant d'une ontologie du domaine. Nous explicitons ce processus dans les sections qui suivent.

3.1. RECHERCHE DE COOCCURRENTS VERBAUX

Pour recenser les différents contextes dans lesquels apparaît le lemme *patient* dans les textes médicaux, nous avons fait une première étude sur ses différentes occurrences dans le corpus *Scientext – Ecrits scientifiques en anglais*. La recherche sur le lemme *patient* dans ce corpus aboutit à 95067 occurrences ; nous les avons filtrées en focalisant notre démarche sur la relation *Verbe + patient* dans la position d'objet direct (OD)/ objet indirect (OI), ou de sujet du verbe (22147 occurrences). Nous avons ainsi obtenu une liste de 952 verbes associés au lemme *patient*, dont 335 sont des hapax et plus de 700 ont moins de 10 occurrences dans le corpus. Les verbes les plus fréquents autour du lemme *patient* sont *have* et *be* : ils totalisent plus de 21% de toutes les cooccurrences de *patient* + Verbe dans le corpus (Figure 1). La grande majorité des verbes est employée sous la forme conjuguée (87,54%), 5.92% en tant qu'infinitif et 6.54% sous la forme d'un participe ou d'un gérondif.

¹⁰ Travaux non-publiés d'Anastasia Galmiche, Mémoire Master TAL, 2017, Université de Franche-Comté.

Figure 1. Cooccurrents verbaux du lemme *patient*

L'analyse des phrases extraites du corpus contenant le terme et son co-occurrent verbal nous permet d'établir les structures actantielles. Les structures à l'infinitif (1) et au passif (2), ainsi que les structures avec le gérondif (3) sont rassemblées autour de la structure élémentaire à la voix active. On extrait aussi les structures élémentaires des structures phrastiques complexes (4), composées de plusieurs propositions subordonnées.

- (1) (...) *use of several agents combined or of fixed dose combinations treatments have the potential to bring patients to goal blood pressure (...)*
SUJET +BRING+ PATIENT +PRÉP+ OI
- (2) *All but eight patients were prescribed diuretics.*
SUJET+ PRESCRIBE +PATIENT +OD
- (3) *In a double blind, placebo control study, treating over 500 patients with OCD, clomipramine was found to be superior to placebo.*
SUJET +TREAT +PATIENT +PRÉP +OI
- (4) *Previous consensual advice that combination treatments expose patients to the increased risk of adverse events has been replaced by good evidence to the contrary.*
SUJET +EXPOSE +PATIENT +PRÉP +OI

On arrive ainsi à classer l'ensemble des structures en trois catégories principales selon que le *patient* est le premier actant (I), le second actant (II) ou le troisième actant (III) du verbe:

<u>Fonction/ Structure</u>	<u>Exemple</u>
I. <i>PATIENT</i> + VERBE + <u>OD</u>	<i>The patients underwent echocardiography (...)</i>
I. <i>PATIENT</i> + VERBE + <u>OD</u> + PRÉP + <u>OI</u>	<i>These asthma patients also showed higher bilirubin levels in the induced sputum.</i>
I. <i>PATIENT</i> + VERBE + PRÉP + <u>OI</u>	<i>Overall, 10% of sarcoidosis patients suffer from permanent sequelae.</i>
II. <u>SUJET</u> + VERBE + <i>PATIENT</i>	<i>Most previous studies have included selected patients with impaired left ventricular systolic function.</i>
II. <u>SUJET</u> + VERBE + <i>PATIENT</i> + PRÉP + <u>OI</u>	<i>Endpoint trials follow patients over a pre-defined period of time.</i>
III. <u>SUJET</u> + VERBE + <u>OD</u> + PRÉP + <i>PATIENT</i>	<i>(...) and that blood was collected from patients.</i>

3.2. MODÉLISATION SÉMANTIQUE DES ACTANTS

L'étape suivant consiste à caractériser les actants de façon sémantique pour obtenir une représentation plus explicite et plus généralisante de différentes structures actantielles. Pour cela, nous nous appuyons sur les Types Sémantiques d'une ontologie du domaine biomédical : Unified Medical Language System (UMLS)¹¹ développé par la National Library of Medicine (NLM).

3.2.1. DESCRIPTION DE L'UMLS

L'UMLS est une base de connaissances et un ensemble de logiciels permettant le développement d'applications pour la recherche et l'extraction d'informations dans le domaine biomédical. Elle a pour objectif d'aligner les différentes ressources termino-ontologiques du domaine biomédical à un niveau conceptuel, pour les rendre interopérables. Le cœur de l'UMLS, le Metathesaurus, est une collection de concepts et de termes, ainsi que de leurs relations, provenant de plusieurs vocabulaires contrôlés en différentes langues. Le Metathesaurus associe plusieurs informations à chaque concept (Figure 2).

¹¹ <https://www.nlm.nih.gov/research/umls/> , accédé le 05/06/2018.

Concept : Nom du concept UMLS [Identifiant Concept] Nom du concept
Semantic Types : Type Sémantique auquel appartient le concept Type sémantique [Identifiant Type]
Definitions : Définition du concept par source Source Définitions ()
Atoms : Instances des écritures (occurrences) dans les différentes sources Occurrence [Identifiant Atom/Source/Type du Terme/Identifiant Occurrence]
Contexts : Position du concept dans la structure d'une source Arbre/Structure de la source
Concept Relations : Relation avec d'autres concepts Type de Relation Attribut de la Relation Source Concept Identifiant Concept

Figure 2. Organisation du Metathesaurus, UMLS

Pour mieux comprendre les informations de ces différentes entrées, nous présentons l'exemple (simplifié) du concept PATIENTS (Figure 3).

Concept	[C0030705] PATIENTS
Semantic Types	PATIENT OR DISABLED GROUP [T101]
Definitions	NCI A person who receives medical attention, care, or treatment, or who is registered with medical professional or institution with the purpose to receive medical care when necessary. NCI_DICOM A patient is the subject of observations. PSY Persons under medical care. Use a more specific term if possible. Consider also CLIENTS.
Atoms	patient [A0601454/AOD/DE/0000006912] PATIENT [A0765239/CCPSS/PT/0024077] patients [A18556155/CHV/SY/0000009372}

Figure 3. Patient dans l'UMLS

Ce qui est intéressant dans cette représentation, c'est qu'elle relie une forme graphique, appelée Atom (*patient*) au Concept (PATIENTS) et à son Type Sémantique (PATIENT OR DISABLED GROUP). Nous pouvons alors l'utiliser pour attribuer les Types Sémantiques à des formes graphiques trouvées dans le corpus et obtenir ainsi une représentation sémantique de différents actants dans les structures verbales autour d'une lexie spécialisée.

3.2.2. CARACTÉRISATION SÉMANTIQUE DES ACTANTS

Nous illustrons ce processus à l'aide de l'exemple suivant :

- (5) *A recent study by Nadelman et al. followed human patients that **showed** recurrent skin lesions.*

La structure actantielle autour de la lexie *patient* dans cette phrase est :
PATIENT + **SHOW** + **OD** (recurrent skin lesions)

Le processus de caractérisation sémantique des actants se décompose en plusieurs étapes :

- i. Extraction de l'objet de la phrase → reccurent skin lesions
- ii. Recherche de l'Atome dans l'UMLS → *skin lesions*
- iii. Identification du Concept associé à l'Atome → SKIN LESION
- iv. Identification du Type Sémantique associé au Concept → DISEASE OR SYNDROME
- v. Modélisation sémantique de la structure actantielle :
PATIENT + **SHOW** + DISEASE OR SYNDROME

Dans cette modélisation, *patient* est le terme recherché par l'utilisateur ; il est représenté par sa forme lemmatisée ; c'est aussi le cas du verbe *show*, cooccurrent verbal de la lexie *patient*. L'objet est représenté par le type sémantique auquel il appartient, DISEASE OR SYNDROME.

- vi. Recherche d'autres exemples qui correspondent à cette structure dans le corpus et validation de la structure.

Cette étape a pour objectif de vérifier la pertinence de la modélisation sémantique. Si plusieurs objets renvoient vers le même Type Sémantique, nous considérons que la structure est pertinente ; ce qui est le cas pour la structure analysée :

- (6) *Patients with RAS showed a higher degree of renal insufficiency (DISEASE OR SYNDROME) compared to non stenotic patients (...)*
- (7) *Five patients showed a lymphocytic infiltration (DISEASE OR SYNDROME) (...)*

Par ailleurs, nombreux types sémantiques peuvent être associés au même actant. Dans le cas du verbe *show*, au moins 4 types d'objets sont possibles : DISEASE OR SYNDROME/ CELL OR MOLECULAR DISFUNCTION/ SIGN AND SYMPTOMS/ FINDINGS. Par conséquent, plusieurs variantes de la même structure sont recensées, accompagnées des exemples extraits du corpus :

Schémas actantiels	Exemples du corpus
<u>DISEASE OR SYNDROME</u>	<i>Five patients showed a lymphocytic infiltration (...)</i>
<u>CELL OR MOLECULAR DISFUNCTION</u>	<i>(...), asymptomatic carriers and hepatitis patients showed the A1896 mutation.</i>
Patient SHOW	
<u>SIGN AND SYMPTOMS</u>	<i>The patients show flu-like symptoms (...)</i>
<u>FINDINGS</u>	<i>Similarly, DMD patients initially show a mild phenotype (...)</i>

4. CONCLUSION ET PERSPECTIVES

Dans cet article, nous décrivons le processus de modélisation de contextes phrastiques autour de lexies spécialisées, appelé à contribuer à l'élaboration d'un logiciel d'aide à la rédaction scientifique. Pour ce faire, nous nous appuyons sur un corpus étiqueté et sur une ontologie du domaine biomédical.

Nous nous efforçons de répondre aux besoins des utilisateurs, en produisant une représentation qui ne fait appel à aucune notion grammaticale complexe et qui n'est pas surchargée d'informations métalinguistiques. Elle prend la forme de schémas actantiels incluant la lexie spécialisée recherchée par l'utilisateur, son cooccurrent verbal et les autres actants impliqués par le verbe et modélisés sous forme de Types Sémantiques. Cette représentation est, de surcroît, complétée par des exemples issus du corpus spécialisé, ce qui permet à l'utilisateur de projeter les schémas sur les phrases réelles, et de construire ainsi sa propre expression.

Cette modélisation n'est qu'une première étape de notre travail autour de la contextualisation de lexies spécialisées. Pour le moment, notre analyse est semi-automatique, alors que l'UMLS permettrait a priori d'attribuer automatiquement les Types Sémantiques à des actants auparavant identifiés

dans un corpus étiqueté. Néanmoins, il ne s'agit pas d'un processus trivial : les actants possèdent eux-mêmes une structure complexe dont il faut retirer les éléments significatifs (*skin lessions* extrait du syntagme *récurrent skin lessions*, ou *renal insufficiency* extrait du syntagme *a higher degree of renal insufficiency*, par exemple). Nous projetons aussi de recenser les structures dérivées à partir de structures élémentaires (par exemple, *hospitalised patient* à partir de « PROFESSIONAL OR OCCUPATIONAL GROUP hospitalise Patient »).

Dans un second temps, nous allons également modéliser d'autres types d'informations contextuelles, telles que les collocations (*positive patients*), les structures collocationnelles plus complexes du type Type Sémantique + Nom (« DISEASE OR SYNDROME *patient* » pour *asthma patient*, *gastric cancer patient*) ou des structures non-prédicatives, telles que Nom + Prep + Nom (« DISEASE OR SYNDROME *in Patient* », « *Patient with* DISEASE OR SYNDROME »).

Notre objectif à terme est de d'aboutir à une visualisation adéquate de ces différents contextes pour les rendre opérationnels dans la tâche d'aide à l'écriture d'un article scientifique.

RÉFÉRENCES

- Aluísio Sandra M., Barcelos Iris, Sampaio Jandir, & Oliveira Osvaldo N., 2001, « How to learn the many unwritten “rules of the game” of the academic discourse: a hybrid approach based on critiques and cases to support scientific writing » [in :] *Advanced Learning Technologies Proceedings. IEEE International Conference*, 257-260.
- Barnes Caroline, Boutron Isabelle, Giraudeau Bruno, Porcher Raphael, Altman Douglas G., & Ravaud Philippe, 2015, « Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial » [in :] *BMC medicine*, 13(1), 1.
- Kinnunen Tomi, Leisma Henri, Machunik Monika, Kakkonen Tuomo, & Lebrun Jean-Luc, 2012, « SWAN-scientific writing AssistaNt: a tool for helping scholars to write reader-friendly manuscripts » [in :] *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 20-24.
- Kübler N. & Pecman M., 2012, « The ARTE bilingual LSP dictionary: From collocation to higher order phraseology » [in :] *Electronic Lexicography*, S. Granger & M. Paquot (eds). Oxford: Oxford University Press, 187-209.
- Lebrun Jean-Luc, 2011, « Scientific writing 2.0: a reader and writer's guide », World Scientific.
- Lerat Pierre, 1995, « Langues spécialisées », PUF, Paris.
- L'Homme Marie-Claude, 2016, « Terminologie de l'environnement et Sémantique des cadres » [in :] *Congrès mondial de linguistique française (CMLF 2016)*, Tours, France.

- L'Homme Marie-Claude, 2014, « Why lexical semantics is important for e-lexicography and why it is equally important to hide its formal representations from users of dictionaries » [in :] *International Journal of Lexicography* 27(4).
- L'Homme Marie-Claude, Robichaud Benoît & Leroyer Patrick, 2013 « Encoding collocations in DiCoInfo: From formal to user-friendly representations » [in :] *Electronic Lexicography*, Oxford : Oxford University Press, 211–236.
- Thomas Izabella, 2016, « Ecrire en langues spécialisées : méthodes et outils du traitement automatique des langues au service de l'autonomie des rédacteurs », HDR, Université de Franche-Comté, UFR des Sciences du Langage, de l'Homme et de la Société.
- Yan Rui, Hatier Sylvain, 2016, « L'extraction et la modélisation de patrons lexico-syntaxiques pour leur enseignement en FLE : un exemple à partir du verbe montrer » [in :] *Linguistik Online* 78 (4).
- Volanschi Alexandra, Kübler Natalie, 2010, « Building an electronic combinatory dictionary as a writing aid tool for researchers in biology » [in :] Sylviane Granger, Magali Paquot (eds.), *e-Lexicography in the 21st Century: New Applications, New Challenges*, Louvain-la-Neuve : Presses Universitaires de Louvain, 343-355.
- Wandji Ornella, L'Homme Marie-Claude, Grabar Natalia, 2013, « Discovery of Semantic Frames for a Contrastive Study of Verbs in Medical Corpora » [in :] *Terminology and Artificial Intelligence*, TIA.

MODÉLISATION DU CONTEXTE DES LEXIES SPÉCIALISÉES
EN VUE DE L'ÉLABORATION D'UN SYSTÈME D'AIDE
À LA RÉDACTION SCIENTIFIQUE
DANS LE DOMAINE BIOMÉDICAL

R é s u m é

Dans cet article nous proposons une modélisation de l'information contextuelle autour des lexies spécialisées en vue de l'élaboration d'un système d'aide à la rédaction scientifique dans le domaine biomédical. Nous considérons plus spécifiquement la modélisation du contexte phrastique décrit en termes de schémas actantiels que nous caractérisons sémantiquement. Cette modélisation est fondée sur un grand corpus d'articles scientifiques dans le domaine biomédical. Elle s'appuie également sur les Types Sémantiques d'une ontologie du domaine, *Unified Medical Language System*.

Mots-clés : terminologie ; terme ; rédaction scientifique ; système d'aide à la rédaction scientifique ; contexte ; modélisation ; ontologie ; corpus.

MODELIZACJA KONTEKSTU LEKSYKI SPECJALISTYCZNEJ
W CELU WYPRACOWANIA SYSTEMU WSPOMAGANIA
REDAGOWANIA ARTYKUŁÓW NAUKOWYCH W BIOMEDYCYNIE

S t r e s z c z e n i e

Autorki proponują modelizację informacji kontekstualnej wokół leksemów specjalistycznych w celu wypracowania systemu wspomaganie redagowania prac naukowych w dziedzinie biomedycyny. Pod uwagę brana jest tu modelizacja kontekstu zdaniowego opisanego jako schematy aktantowe, które mają charakter semantyczny. Modelizacja ta opiera się na obszernym korpusie

artykułów naukowych z biomedycyny, a także na Typach Semantycznych ontologii danej dziedziny, *Unified Medical Language System*.

Słowa kluczowe: terminologia; termin; redagowanie prac naukowych; system wspierania redagowania prac naukowych; kontekst; modelizacja; ontologie; korpus.