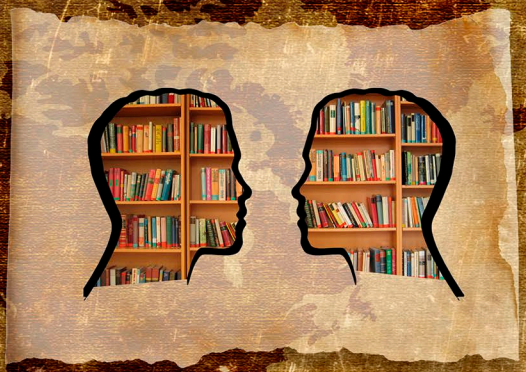
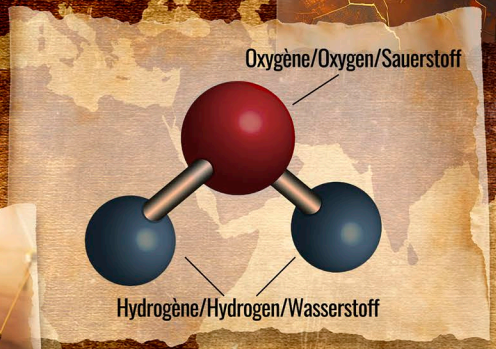
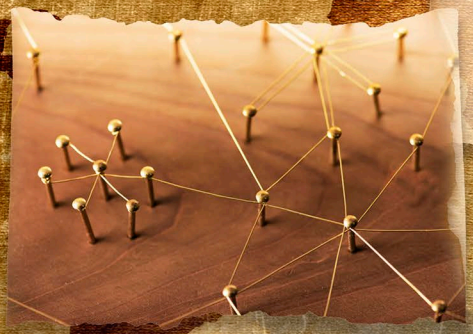
COLLOQUE INTERNATIONAL











Lexiques | Lexicons | Lexik

16–18 DÉCEMBRE 2026

NANCY
Campus Lettres et Sciences Humaines
ATILF

www.atilf.fr lexiques.sciencesconf.org lexiques@sciencesconf.org

atilf – Analyse et Traitement Informatique de la Langue Française · cnrs · Université de Lorraine · ALL Arts, Lettres et Langues Nancy · SHS Sciences Humaines & Sociales Nancy · Orion · France 2030 Initiative d'Excellence Lorraine · Lorraine Université d'Excellence · Projet Impact LUE INSIGHT · LINGUIST List · Ortho Édition · Nouveaux Cahiers d'Allemand

© ATILF | LG | 11/2025 | Illustrations Freepik

COLLOQUE INTERNATIONAL

Lexiques | Lexicons | Lexik

L'Axe disciplinaire « Lexique » du laboratoire *Analyse et Traitement Informatique de la Langue française* (ATILF – CNRS/UL – UMR 7118) organise un Colloque international consacré à la thématique « Lexiques » du 16 au 18 décembre 2026 à Nancy (France).

Cette manifestation scientifique vise à rassembler et à faire échanger chercheuses et chercheurs, juniors et expérimenté·e·s, sur des problématiques de recherche actuelles relevant d'un champ disciplinaire qui occupe une place centrale en Sciences du langage, et par-delà s'avère primordial dans la communication et dans les relations humaines.

La thématique du colloque est pensée de manière à inclure le plus largement possible des recherches touchant aux différentes (sous-)disciplines des sciences du langage, et impliquant des éclairages aussi bien théoriques et empiriques que des discussions méthodologiques dans le domaine. S'entrecroiseront, pour chacune des sous-thématiques du colloque, des approches diachroniques et synchroniques, des études de corpus ou de terrain, des analyses en contexte monolingue, bilingue et multilingue. Les interventions pourront relever, en premier lieu mais pas exclusivement, de la morphologie, de la syntaxe, de la sémantique, de l'analyse du discours et de l'étymologie.



PROGRAMME

Mercredi 16 décembre

13h00 **ACCUEIL – Mots de bienvenue** | Salle Imbs

Conférence plénière 1 | Salle Imbs **13h15**
Sören Stumpf (Université de Kassel)
Constructional idioms revisited: Vergangenheit, Gegenwart, Zukunft

Session 1 | Salle Imbs

Modération : **Yvon Keromnes** (Université de Lorraine, ATILF, CNRS)

14h00 **Agnès Tutin / Francis Grossmann** (Université Grenoble Alpes)
Phrases préfabriquées interrogatives à fonction métalinguistique hétéroadressées : quelles constructions ?

Lian Chen (Université d'Orléans) **14h30**
Approche onto-phraséologique pour l'extraction automatique des relations idiomatiques

15h00 **Session posters : présentation éclair**
Coordination : **Paola Herreño Castañeda, Maeva Sillaire, Akerke Yessenali, Louis Mathieu** (Université de Lorraine, ATILF, CNRS)

Pause **15h25**

15h40 **Alexander Ziem** (HHU), **Nina Böbel** (HHU), **Athina Sioupi** (Aristotle University of Thessaloniki), **Voula Giouli** (Aristotle University of Thessaloniki)
Towards a Verb-Lexicon Model: Greek and German in a contrastive constructional perspective

Marie-Luis Merten (Université de Mannheim) **16h10**
Alignmentmarker in der digitalen Anschlusskommunikation: Korpusanalytische Zugänge zu polyfunktionalen Ressourcen

16h40 **Francesco Caprioli** (Université de Bergame)
Entre idéologie et multilinguisme : la traduction des mots-slogans au Parlement européen (allemand-italien-français)

Pause **17h10**

17h30 **Rosa Cetro** (Université de Pise)
Phraséologie et communication publicitaire : analyse d'un corpus de slogans de produits de la marque Monoprix

Fabio Mollica / Ciro Porcaro (Université de Milan) **18h00**
Lexikalische Kreativität und argumentative Wirkung durch Automobilmarken in der metaphorischen XYZ-Konstruktion

Jeudi 17 décembre

9h30 **ACCUEIL – thé – café** | Salle Imbs

Session 2 | Salle CLSH G04

Modération : **Mathilde Dargnat** (UL, ATILF, CNRS)

10h00 **Angelina Aleksandrova** (Paris Université)
Entre lexique et phraséologie : modéliser la dénomination des mouvements sportifs

Valentina Bisconti (Université de Picardie)
Les expressions polylexicales contenant des noms de nombres : de la cardinalité idiomatique à l'expression du vague

Nesrine Raissi (Universités de Montpellier III et Bourgogne Europe)
Dire le traumatisme par les témoins du camp de Rivesaltes

12h00 **Mathilde Gaillard** (Université d'Assas)
« Reinvent your workforce » *: choix lexicaux dans deux discours spécialisés d'experts sur l'intelligence artificielle*

Laure Gautherot (Université Sorbonne Nouvelle)
‚Bettenstop' und ‚Tourismus-Notbremse': Die Lexik des Massentourismus im Mediendiskurs

Session 3 | Salle Imbs

Modération : **Rute Costa** (NOVA-CLUNL)

Nadia Alaoui (Université de Picardie Jules Verne, CERCLL-PraLing)
Développement lexical et évaluation orthophonique en contexte bilingue

Marina Gonzalez (Université de Madrid) **10h30**
The acquisition of subject-specific vocabulary and CLIL: Word production and lexical networks

11h00 **Pause**

Frédérique Henry / Marie-Laurence Knittel / Evelyne Jacquey / Sandrine Ollinger (Université de Lorraine, ATILF, CNRS) **11h30**
Étude linguistique et textométrique des termes clés du processus du bilan orthophonique

Dina Caetano Alves (Polytechnic Institute of Setúbal; Center of Linguistics, University of Lisbon)
Towards a Standardized Diagnostic Lexicon in Speech and Language Therapy: Methodological Framework and Preliminary Outcomes

Aurélie Picton (Université de Genève) / **Pascaline Duri** (Université Lyon 2) **12h30**
« Mon loup m'a attaquée… » : Regard sur l'appropriation et la mise en discours de la terminologie médicale dans les récits de patients atteints de maladies chroniques

13h00 **Buffet sur place**

Session 2 | Salle CLSH G04	Session 3 | Salle Imbs
Modération : **Éva Buchi** (UL, ATILF, CNRS)	Modération : **Frédérique Brin-Henry** (UL, ATILF, CNRS)
Elise Schürgers (Université de Liège) *Quand l'autonymie vient à la formule. L'activité métadiscursive autour de la locution « fake news » et sa portée argumentative*	**Sylvain Farge / Anais Malfatti** (Université Montpellier 3) *Définir la maladie, entre lexique et discours : une étude contrastive allemand-français sur les lexèmes désignant l'obésité*
14h00	
14h30 **Simone Gomes / Anna Ladilova / Katharina Müller** (Justus Liebig Universität Giessen) *Néologismes et emprunts en langues romanes*	**Thomas Bertin** (Université de Rouen) **/ Sofian Berrouiguet** (CHU de Brest) **/ Ambre Marie** (Université de Bretagne Occidentale) *Gestes, passages à l'acte et idées noires – désigner le suicide pour évaluer le risque de récidive*
Lana Benett (Université Grenoble Alpes) *Lexique et argumentation dans le discours climatosceptique états-unien : l'exemple d'un think tank conservateur (2021-2024)*	**Sandrine Ollinger / Benjamin Husson** (Université de Lorraine, ATILF, CNRS) *Mon voisin de réseau : quand l'annotation en sens invoque la structure des lexiques*
15h00	
15h30 Pause	
Daniel Vojtek (Université P. J. Šafárik de Košice) *Élaboration d'un corpus de terminologie grammaticale multilingue (slovaque-français-anglais-latin) : quelle méthode ?*	**Beatriz Curti Contessoto** (Université Lyon 2) *L'interdisciplinarité entre droit et santé publique : quelques réflexions pour la création d'un système conceptuel accessible au public profane*
16h00	
16h30 **Romain Delhem** (Université Paris Nanterre) *Assessing measures of productivity for derivational morphological patterns*	**Valentin Neveu** (Paris Université) *La constitution diachronique d'un lexique de spécialité par l'emprunt : le cas du law french 14e-18e*
Andreu Llòria Pascual (Université Lumière Lyon 2 / Universitat Autònoma de Barcelona) *Homonymie et polysémie à l'épreuve des emprunts : le cas de TOQUE en moyen français et en français préclassique*	**Juan Carlos Justiniano / Maria Rubio** (Université de Madrid) *Denominative variation in Spanish musical lexicon. The panhispanism of the vocabulary of the Leximus dictionary*
17h00	
17h30 **Jarvis Looi** (Université de Lorraine) *« Prochain » ou « suivant » ? Une confrontation entre théorie introspective et données de corpus*	**Veronika Lux-Pogodalla / Cristina Pacheco Diego / Nabil Gader** (Université de Lorraine, ATILF, CNRS) *Réseau lexical musical : une (re)naissance*
19h00 Dîner	

Vendredi 18 décembre

8h30 **ACCUEIL – thé - café** | Salle Imbs

Conférence plénière 3 | Salle Imbs
Hans Boas (Université d'Austin)
On the use of metalanguage in structuring lexicons
9h00

Session 4 | Salle Imbs

Modération : **Hélène Vinckel-Roisin** (Université de Lorraine, ATILF, CNRS)

9h45 **Zinaida Geylikman** (Independent Researcher)
Modeling human nouns' lexicon in diachrony: a methodological essay

Gauvain Chalchli (Université de Bordeaux)
TLFi-Dat : une modélisation chronologique du lexique français à partir des données du TLFi
10h15

10h45 **Pause**

Christine Da Silva / Sylvie Multon (Université de Lorraine)
Fluence verbale sémantique de jeunes adultes francophones : exploration de nouvelles données en vue d'un référentiel adulte
11h15

11h45 **Amparo Alcina / Fabio Badenes** (Université de Jaume)
Utilisation du modèle ONTODIC pour l'implémentation du Réseau Lexical du Français dans une base de connaissances. Le cas de l'entrée fleur I.1

Clôture **12h15**

12h45 **Buffet sur place**

Session posters

Lian Chen (Université d'Orléans) / **Huy-Linh Dao** (INALCO) / **Alexander Delaporte** (CRLAO-CNRS) / **Damien Nouvel** (INALCO)
Étude et détection automatique des néologismes en chinois et en vietnamien (2015–2025)

Christine Da Silva (Université de Lorraine) / **Marine Le Mené** (Université du Québec à Montréal) / **Caroline Masson** (Université Sorbonne nouvelle) / **Christophe Benzitoun** (Université de Lorraine, ATILF, CNRS) / **Loïc Liegeois** (Université Clermont Auvergne)
Prise en compte de la variation dans l'évaluation des compétences lexico-sémantiques des enfants : réflexions théoriques et méthodologiques

---

Henriette Huber (Université de Düsseldorf)
Word(s) in progress: A usage-based analysis of univerbation tendencies in New High German

---

Filip Kalas / **Pavol Liptak** (Université de Brastislava)
Quantifying Bilingual Terminology Extraction for Slovak Public Administration : GPT Prompting vs. RAG

Laurence Kister / **Evelyne Jacquey** / **Sandrine Ollinger** (Université de Lorraine, ATILF, CNRS)
Annotation de contextes en référence et en facettes sémantiques pour faciliter la désambiguïsation lexicale

---

Marie-Laurence Knittel (Université de Lorraine, ATILF, CNRS)
L'ambiguïté des adjectifs en *-ant* : structure argumentale et propriétés du suffixe *-ant*

Matthieu Quiniard / **Audrey Mazur** (ENS Lyon)
L'impact de la dyslexie-dysorthographie sur les pratiques lexicales : un outil de recherche et d'aide à la remédiation

---

Jozef Stefcik (Université de Bratislava)
LLM extraction tools in the Bilingual Slovak–English Terminology project of Public Administration

Table des matières

CONFÉRENCES INVITÉES 13

Hans Boas 14

Sören Stumpf 15

COMMUNICATIONS ORALES 18

Nadia Alaoui 19

Alcina Amparo / Fabio Badenes 22

Angelina Aleksandrova 25

Lana Benett 28

Thomas Bertin / Sofian Berrouiguet / Ambre Marie 31

Valentina Bisconti 34

Dina Caetano Alves 37

Francesco Caprioli 39

Rosa Cetro 42

Gauvain Chalchli 44

Lian Chen 46

Beatriz Curti Contessoto 49

Christine Da Silva / Sylvie Multon 52

Romain Delhem 55

Sylvain Farge / Anaïs Malfatti 57

Mathilde Gaillard 60

Laure Gautherot 62

Zinaida Geylikman 64

Simone Gomes / Anna Ladilova / Katharina Müller 67

Marina Gonzalez 69

Frédérique Henry /Marie-Laurence Knittel / Evelyne Jacquey / Sandrine Ollinger 70

Juan Carlos Justiano / Maria Rubio 73

Andreu Llòria Pascual 75

Jarvis Looi 79

Veronika Lux-Pogodalla / Cristina Diego Pacheco / Nabil Gader 81

Marie-Luis Merten 85
Fabio Mollica / Ciro Porcaro 87
Valentin Neveu 88
Sandrine Ollinger / Benjamin Husson 89
Aurélie Picton / Pascaline Duri 92
Nesrine Raissi 95
Elise Schürgers 97
Athina Sioupi 100
Agnès Tutin / Francis Grossmann 103
Daniel Vojtek 105
POSTERS 108
Lian Chen / Huy-Linh Dao / Alexander Delaporte / Damien Nouvel 109
Christine Da Silva /Marine Le Mené / Caroline Masson /Christophe Benzitoun / Loïc Liegeois 111
Henriette Huber 113
Evelyne Jacquey / Laurence Kister / Sandrine Ollinger 115
Filip Kalaš / Pavol Lipták 117
Marie-Laurence Knittel 119
Matthieu Quignard / Audrey Mazur 122
Jozef Štefčík 125



CONFÉRENCES INVITÉES

Hans Boas
Université d'Austin

On the use of metalanguage in structuring lexicons

This presentation discusses the status and role of metalanguage in organizing lexicons by providing a brief overview of different types of metalanguages developed for different purposes. The first part of the paper offers an analysis and critical review of A. Wierzbicka's Natural Semantic Metalanguage by investigating the various categories used for meaning descriptions and how these may relate to different types of concepts. The second part of the paper reviews a number of issues with Wierzbicka's semantic primes, specifically regarding the methodology used to arrive at primes. The second part of the paper introduces C. Fillmore's semantic frames, giving a short overview of Fillmore's 1968 "The case for case" paper by discussing his various so-called "deep cases" and showing why they were eventually abolished in the 1970s. Then, a review of Fillmore's work on semantic frames during the 1970s and 1980s forms the basis for his emerging theory of Frame Semantics and its subsequent implementation in the Berkeley FrameNet database. The third part of the paper offers a comparison of Wierzbicka's and Fillmore's approaches by discussing the "building blocks" employed for recognizing, describing, and analyzing meaning and it highlights some important differences in methodology such as the use of empirical corpus data, reproducibility, and falsifiability. One of the central themes of this talk will be how form and meaning are connected to each other and to what degree it is possible to arrive different levels of generalizations over meaning structures that may allow linguists to shed light on how meaning is constructed.

Sören Stumpf
Ludwig Maximilian University Munich

Constructional idioms revisited: Vergangenheit, Gegenwart, Zukunft

Eine zentrale Annahme der Konstruktionsgrammatik besteht darin, dass Lexikon und Grammatik nicht als strikt getrennte Bereiche zu verstehen sind, sondern vielmehr ein Kontinuum bilden. An dem einen Ende dieses Spektrums stehen volllexikalische Einheiten, am anderen schematische Strukturen. In meinem Vortrag widme ich mich Mehrworteinheiten, die zwischen diesen beiden Polen angesiedelt sind: Darunter fallen Strukturen, die einerseits lexikalisch feste Elemente enthalten, andererseits aber morphosyntaktisch offene Stellen aufweisen und die zugleich eine eigenständige Gesamtbedeutung tragen oder eine pragmatische Funktion erfüllen (z. B. [X *of you*], [X *ist ein* Y-*Ding*]). Für diese Muster kursiert in der phraseologischen und konstruktionsgrammatischen Forschung eine Vielzahl unterschiedlicher Bezeichnungen; im Vortrag verwende ich die Termini „constructional idioms“ bzw. „Phrasem-Konstruktionen“.

An der Schnittstelle von Lexikon und Syntax verortet, galten teilschematische Strukturen in der Phraseologie lange Zeit eher als randständig. In der Konstruktionsgrammatik hingegen zählen sie von Beginn an zu den zentralen Gegenstandsbereichen. Gerade im Hinblick auf das Tagungsthema „Lexiques – Lexicons – Lexik“ erweist sich ihr ambivalenter Status als besonders aufschlussreich: Phrasem-Konstruktionen sind weder als (mehrgliedrige) Lexeme noch als rein syntaktische Strukturen adäquat zu erfassen. Vielmehr vereinen sie Eigenschaften von Lexikon und Grammatik, die je nach Konstruktion variieren und in empirischer Analyse wie theoretischer Modellierung gleichermaßen zu berücksichtigen sind.

Der Vortrag skizziert zunächst die Entwicklungslinien maßgeblicher Forschungsarbeiten zu Phrasem-Konstruktionen und geht anschließend auf aktuelle Ansätze sowie zukünftige Tendenzen ein. Neben grundlegenden theoretischen Überlegungen zur Bestimmung und Abgrenzung des Untersuchungsgegenstands beziehe ich auch eigene aktuelle Korpusstudien zum Deutschen ein, die insbesondere die Dynamik von lexikogrammatischen Mustern herausstellen. Im Zentrum stehen die Fragen, wie sich teilschematische Konstruktionen innerhalb kurzer Zeit durch die rekurrente Modifikation volllexikalisierter Idiome herausbilden und wie im Kontext digitaler Kommunikation gänzlich neue Phrasem-Konstruktionen entstehen.

Der Vortrag wird in deutscher Sprache gehalten, die Präsentationsfolien sind englischsprachig.

English abstract

A central assumption of Construction Grammar is that lexicon and grammar are best understood not as strictly separate domains, but as a continuum. At one end of this spectrum are fully lexical units, at the other schematic structures. In my presentation, I will focus on multi-word expressions that lie between these two poles: these include structures that combine lexically fixed elements with morphosyntactically open slots and that, at the same time, have a general meaning or fulfil a pragmatic function (e.g. [X *of you*], [X *ist ein* Y-*Ding*]). A variety of different terms are used for these patterns in phraseological and constructionist research; in my presentation, I will use the terms ‘constructional idioms’ and ‘Phrasem-Konstruktionen’.

At the interface between lexicon and syntax, partially schematic structures were long regarded as marginal phenomena in phraseological research. In Construction Grammar, on the other hand, they have been one of the central topics from the outset. Their ambivalent status is particularly interesting with regard to the conference theme ‘Lexiques – Lexicons – Lexik’: constructional idioms cannot be adequately described either as lexemes or as purely

syntactic structures. Rather, they combine characteristics of lexicon and grammar, which vary depending on the construction and must be considered in empirical analysis and theoretical modelling.

The presentation first outlines the development of key research on constructional idioms and subsequently discusses current approaches and future trends. In addition to fundamental theoretical considerations on the definition and delimitation of the subject of investigation, I also draw on my own recent corpus-based analyses on German, which highlight the dynamics of lexico-grammatical patterns. I address, in particular, how semi-schematic constructions develop within a short period of time through the repeated modification of fully lexicalized idioms and how new constructional idioms emerge in the context of digital communication.

Ausgewählte Literatur / Selected literature

Boas, Hans C. 2010. The syntax-lexicon continuum in Construction Grammar. A case study of English communication verbs. *Belgian Journal of Linguistics* 24. 54–82.

Corpas Pastor, Gloria. 2021. Constructional idioms of 'insanity' in English and Spanish: A corpus-based study. *Lingua* 254. 1–20.

Dobrovol'skij, Dmitrij. 2011. Phraseologie und Konstruktionsgrammatik. In Alexander Lasch & Alexander Ziem (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*, 111–130. Tübingen: Stauffenburg.

Fillmore, Charles, Paul Kay & Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions. The Case of *let alone*. *Language* 64. 501–538.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. & Thomas Herbst. 2021. The *nice-of-you* construction and its fragments. *Linguistics* 59. 285–318.

Ivorra Ordines, Pedro. 2023. *Por mí como si te operas*. Constructional idioms of rejection from a constructionist approach. *Yearbook of Phraseology* 14. 89–120.

Mellado Blanco, Carmen (ed.). 2022. *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*. Berlin & Boston: De Gruyter.

Mellado Blanco, Carmen, Fabio Mollica & Elmar Schafroth (eds.). 2022. *Konstruktionen zwischen Lexikon und Grammatik. Phrasem-Konstruktionen monolingual, bilingual und multilingual*. Berlin & Boston: De Gruyter.

Schafroth, Elmar. 2024. Phraseoschablonen. In Felix Tacke & Anja Hennemann (eds.), *Konstruktionen, Kontexte, Gattungen. Constructions, contexts, genres*, 27–54. Göttingen: Vandenhoeck & Ruprecht.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.

Stein, Stephan & Sören Stumpf. 2019. *Muster in Sprache und Kommunikation. Eine Einführung in Konzepte sprachlicher Vorgeformtheit*. Unter Mitarbeit von Andrea Bachmann-Stein, Natalia Filatkina, Carina Hoff und Martin Wengeler. Berlin: Schmidt.

Stumpf, Sören. 2021. „Star Wars ist eben mehr so ein Männerding". Eine Korpusuntersuchung zur Syntax, Semantik und Pragmatik der Konstruktion [X Kopula *ein* Y-*Ding*]. *Deutsche Sprache* 49. 24–45.

Stumpf, Sören. 2024. *Wie viele Meistertitel wollt ihr noch gewinnen? FC Bayern: Ja* – Eine Textkonstruktion in den sozialen Medien zwischen Stabilität, Variabilität und Kreativität. In Sören Stumpf & Stephan Stein (eds.), *Konstruktionsgrammatik X. Textsorten und Textmuster als Konstruktionen?*, 63–87. Tübingen: Stauffenburg.

Stumpf, Sören. 2025. On the Dynamics of Constructional Idioms. A Micro-Diachronic Approach to the Entrenchment and Conventionalization of Lexico-Grammatical Patterns in German. In Sabine Arndt-Lappe & Natalia Filatkina (eds.), *Dynamics at the Lexicon-Syntax Interface. Creativity and Routine in Word-Formation and Multi-Word Expressions*, 193–240. Berlin & Boston: De Gruyter.

Ziem, Alexander (ed.). 2018. *Muster im Sprachgebrauch: Construction Grammar meets Phraseology*. Special Issue *Linguistik Online* 90.3.



COMMUNICATIONS ORALES

Nadia Alaoui
Université de Picardie Jules Verne, CERCLL-PraLing

Développement lexical et évaluation orthophonique en contexte bilingue : *Étude comparative auprès d'enfants de 24 à 32 mois bilingues arabe marocain-français et monolingues français*

La pratique orthophonique s'est longtemps appuyée sur des idéologies linguistiques valorisant la norme monolingue comme modèle de référence et négligeant la complexité des répertoires plurilingues. Cette orientation a longtemps invisibilisé la diversité linguistique, particulièrement celle des enfants bilingues, que les tests standardisés ne permettent pas d'évaluer de manière adéquate.

Les pratiques de *translanguaging* redéfinissent aujourd'hui les usages langagiers dans de nombreux contextes sociolinguistiques (Tai et Wei 2025) et le recours exclusif à une norme monolingue ne répond plus aux besoins d'une évaluation équitable et contextualisée.

C'est dans cette perspective que s'inscrit notre recherche, qui vise à interroger la pertinence des outils orthophoniques actuellement disponibles lorsqu'ils sont mobilisés auprès d'enfants bilingues. Afin de préserver une terminologie consensuelle au sein de la profession, nous avons choisi de conserver le terme « bilinguisme ».

Les recherches en acquisition du langage montrent que les enfants bilingues et monolingues suivent des trajectoires développementales globalement similaires, franchissant les mêmes étapes dans l'émergence du langage malgré des variabilités interindividuelles importantes (Petitto *et al*. 2001 ; De Houver 2006). Toutefois les données empiriques concernant le développement lexical des enfants bilingues restent contrastées (Hoff et Core 2013) : certains travaux montrent que le vocabulaire conceptuel global, (c'est-à-dire l'ensemble des lexèmes produits dans les deux langues) est comparable à celui des enfants monolingues (Zablit et Trudeau 2008). Ce type de mesure est d'ailleurs considéré comme un indicateur plus pertinent pour évaluer le lexique en contexte bilingue (Klassert *et al*.2014), D'autres en revanche, rapportent un lexique expressif plus restreint chez les bilingues (Hoff et Ribot 2017).

Et si, pour la majorité des enfants, l'acquisition se déroule sans difficulté majeure, certains présentent des retards transitoires ou des troubles développementaux du langage (TDL). Le TDL qui représente un enfant sur quatorze (Norbury *et al*. 2016) est un enjeu de santé publique (Kern et Gayraud 2010) et l'évaluation langagière dès l'âge pré-scolaire constitue un levier essentiel pour une prise en charge orthophonique précoce avec un meilleur pronostic. Les études montrent que les enfants bilingues et monolingues avec un TDL présentent un profil comparable (Paradis *et al*. 2003). Toutefois, le diagnostic clinique orthophonique de TDL en contexte bilingue reste complexe en raison du peu d'outils d'évaluation standardisés et validés pour cette population (Thordardottir 2015), et adaptés à la pratique clinique orthophonique. Cette complexité est aussi exacerbée par une forte tendance à vouloir privilégier une approche strictement quantitative au détriment de l'analyse qualitative pourtant indispensable.

Dans ce contexte, de nombreux travaux ont souligné la nécessité de repenser les critères d'évaluation afin de mieux distinguer les variations développementales typiques des signes précoces de TDL chez les bilingues (Bishop *et al*. 2017). Définir des repères développementaux spécifiques aux enfants bilingues apparaît indispensable mais demeure rare pour de nombreuses langues.

Notre étude examine les compétences lexicales de deux groupes d'enfants bilingues arabe marocain-français et monolingues français âgés de 24 à 32 mois. L'objectif principal est de décrire leurs compétences expressives et réceptives et de les comparer, afin d'apporter aux cliniciens des repères développementaux.

Trois outils orthophoniques, élaborés initialement pour des monolingues français (Cornet 2020,) ont été utilisés : un questionnaire parental, une épreuve de dénomination orale sur images et une épreuve de désignation d'images. Le recours à ce matériel n'avait pas pour objectif de mesurer les performances des enfants bilingues selon une norme de

référence monolingue, mais au contraire de révéler les biais cliniques induits par une telle approche. Sa mise en œuvre dans un contexte bilingue a conduit à questionner la pertinence des épreuves orthophoniques standardisées et à souligner la nécessité d'une adaptation des tests aux trajectoires linguistiques des enfants bilingues, tant en termes de cotation que d'interprétation.

Nous avons formulé l'hypothèse qu'une différence de taille lexicale existe entre les deux groupes, y compris en tenant compte du vocabulaire conceptuel total chez les bilingues. Cette différence serait mise en évidence à travers des épreuves directes (dénomination orale et désignation) et corroborée par les réponses recueillies via les questionnaires parentaux.

Nos premiers résultats sur le groupe de 30 à 32 mois révèlent une différence significative en faveur des enfants monolingues dans l'épreuve de production (dénomination) ; aucune différence n'est observée en compréhension (désignation). Ces tendances sont corroborées par les réponses aux questionnaires parentaux, ce qui confirme leur validité comme outil de dépistage (Kern et Gayraud 2010) à condition qu'ils soient interprétés de manière contextualisée en tenant compte du profil linguistique de l'enfant.

L'analyse qualitative des productions révèle des différences dans un réseau lexical encore en structuration : les enfants bilingues recourent plus fréquemment à des stratégies sémantiques de type surgénéralisation (Waxman et Gelman 2009), tandis que les enfants monolingues ont davantage recours à des simplifications phonologiques (McLeod *et al.* 2001).

Le recours à l'arabe à la tâche de dénomination souligne la nécessité d'une évaluation tenant compte de l'ensemble du répertoire lexical, afin de ne pas invisibiliser une partie de leurs compétences.

L'ensemble de nos résultats suggère que les performances plus faibles observées en production chez les bilingues ne traduisent pas un déficit, mais reflètent une répartition des connaissances entre deux systèmes linguistiques, et une dynamique différente d'accès lexical.

Cependant ces résultats doivent être interprétés avec prudence en raison de plusieurs limites méthodologiques : taille réduite de l'échantillon, absence de contrôle de variables confondantes comme le statut socio-économique et caractère non-représentatif de l'échantillon bilingue à l'ensemble de la population bilingue arabe marocain -français.

Malgré ces limites méthodologiques, cette étude constitue, à notre connaissance, la première exploration ciblée du développement lexical bilingue arabe marocain–français à l'aide d'un protocole combinant évaluation directe et questionnaire parental. Elle ouvre des perspectives pour la conception d'outils orthophoniques mieux adaptés au bilinguisme, en intégrant le vocabulaire conceptuel total et en adaptant les listes d'items aux spécificités linguistiques des enfants.

Les résultats obtenus confirment que l'évaluation en contexte bilingue ne peut se limiter à l'application de normes monolingues. Cette recherche contribue également à ouvrir le champ de recherche au bilinguisme arabe marocain-français, encore peu exploré en orthophonie malgré la forte représentativité de la population concernée.

Bibliographie :

Bishop D. V. M., Snowling M. J., Thompson, P. A., Greenhalgh, T. (2017). The CATALISE-2 consortium. Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), Article 10.

Cornet, D. (2020). *L'émergence du langage : Réflexions linguistiques, troubles développementaux et nouvelle approche statistique* (Thèse de Doctorat en Sciences du Langage, sous la direction de Christophe Rey). UFR LSH. Laboratoire LT2D. Université de Cergy-Pontoise. Non publié.

De Houwer, A. (2006). Le développement harmonieux ou non harmonieux du bilinguisme de l'enfant au sein de la famille. *Langage et société, 116*(2), 29-49.

Hoff, E., et Core, C. (2013). Input and Language Development in Bilingually Developing Children. *Seminars in Speech and Language*, *34*(04), 215-226.

Hoff, E., et Ribot, K. M. (2017). Language growth in English monolingual and Spanish-English bilingual children from 2.5 to 5 years. *The Journal of Pediatrics, 190*, 241-245.e1.

Kern, S., et Gayraud, F. (2010). L'inventaire français du développement communicatif : Version longue. *Médecine et enfance*. http://www.ddl.cnrs.fr/fulltext/Kern/Bovet_2005a.pdf

Klassert, A., Gagarina, N., et Kauschke, C. (2014). Object and action naming in Russian- and German-speaking monolingual and bilingual children. *Bilingualism: Language and Cognition*, *17*(1), 73-88.

McLeod, S., Doorn, J. V., et Reed, V. A. (2001). Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology*, *10*(2), 99-110.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder : Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), Article 11*57*(11), 1247-1257.

Paradis, J., Crago, M., Genesee, F., et Rice, M. (2003). French-English bilingual children with SLI : How do they compare with their monolingual peers? *Journal of Speech, Language and Hearing Research, 46*(1),113-127

Petitto, L. A., Katerelos, M., Levy, B. G., Gauna, K., Tétreault, K., et Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language, 28*(2). 453-496

Tai, K. W. H., et Wei, L. (2025). Translanguaging. In *Reference Module in Social Sciences* (p. B9780323955041005275). Elsevier. https://doi.org/10.1016/B978-0-323-95504-1.00527-5

Thordardottir, E. (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J. de Jong, et N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment*. 331-358. Bristol, Blue Ridge Summit: Multilingual Matters.

Waxman, S. R., et Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258-263.

Zablit, C., et Trudeau, N. (2008). Le vocabulaire chez les jeunes enfants libanais arabophones, francophones et bilingues. *Glossa, 103*, 36-52.

Alcina Amparo / Fabio Badenes
Université de Jaume

Utilisation du modèle ONTODIC pour l'implémentation du *Réseau lexical du français* dans une base de connaissances. Le cas de l'entrée *fleur I.1*

Les fonctions lexicales constituent un outil développé dans le cadre de la lexicologie explicative et combinatoire (LEC) (Mel'čuk et al. 1995), qui permet de formaliser des relations combinatoires et de dérivation sémantique entre des unités lexicales, tant au niveau syntagmatique que paradigmatique.

Elles ont été utilisées pour créer des ressources lexicographiques, comme le *Dictionnaire explicatif et combinatoire du français contemporain* (Mel'čuk et al. 1984) ou, plus récemment, des ressources électroniques comme le *Réseau lexical du français* (Lux-Pogodalla et Polguère 2011; Polguère 2014).

Cette dernière ressource reflète la possibilité de traiter les relations entre les unités lexicales en les intégrant dans une base de données relationnelle et en proposant une visualisation graphique sous forme de réseau, ce qui facilite la consultation des unités lexicales et leurs relations par rapport aux ressources développées précédemment, qui ne présentent qu'une distribution textuelle ou tabulaire.

Chaque unité lexicale est composée de deux éléments : l'unité proprement dite et un code alphanumérique. Chaque code renvoie à une signification précise associée à l'unité, indique que les unités numérotées partagent un trait sémantique et établit la distance sémantique qui les sépare : les chiffres romains indiquent une grande distance ; les chiffres arabes, relativement petite ; et les lettres, petite.

Par exemple, pour l'unité lexicale *fleur*, on distingue plusieurs significations (Figure 1) : *fleur I.1*, qui signifie « élément d'une plante », et *fleur I.2*, qui signifie « plante », toutes deux plus proches sémantiquement l'une de l'autre que de *fleur II*, ayant la signification de « surface ».

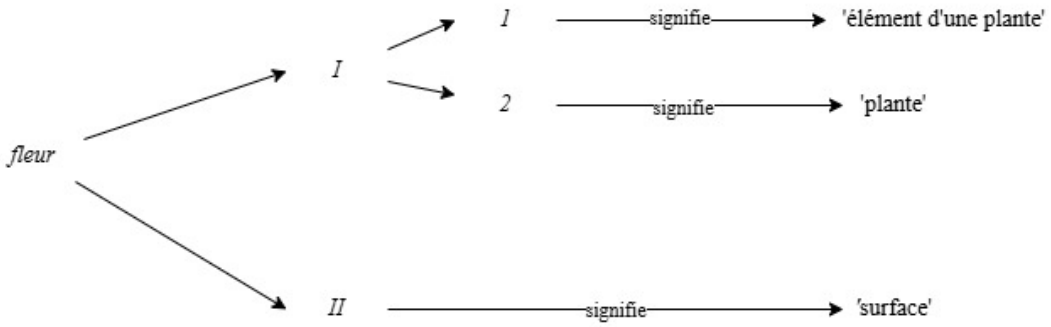


Figure 1 Schématisation des signifiés de l'unité lexicale fleur

Dans cette ressource, les fonctions lexicales relient deux unités accompagnées de leurs codes.

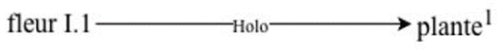


Figure 2 Relation entre fleur I.1 *et* plante[1] *formalisée à l'aide de la fonction* Holo

La fonction *Holo*, qui explicite une relation sémantique de holonymie, relie *fleur I.1* à *plante*[1](Figure 2).



Figure 3 Relation entre fleur I.1 *et* éclore II *formalisée à l'aide de la fonction* IncepFunc$_0$

IncepFunc$_0$, une fonction verbale qui indique le sens de « commencer », relie *fleur I.1* à *éclore II* (Figure 3).

Comme on peut le constater, la fonction *Holo* relie *fleur I.1* sur un plan paradigmatique, tandis qu'*IncepFunc$_0$* le fait sur un plan syntagmatique.

On voit ainsi que la LEC contribue à représenter formellement des unités et leurs relations sur ces deux plans, permettant de travailler tant au niveau sémantique qu'au niveau lexical du langage.

Dans notre groupe de recherche, nous avons travaillé pour améliorer la représentation et la formalisation des données linguistiques à l'aide de bases de connaissances. Les derniers développements de cette technologie s'appuient sur la logique descriptive, qui reprend à son tour des concepts de la théorie des ensembles (Baader et al. 2017). Les bases de connaissances structurent les domaines à l'aide de trois types d'éléments : les individus, les classes (qui regroupent les individus en fonction de leurs propriétés) et les propriétés qui décrivent et relient à la fois les classes et les individus. Cette configuration d'éléments permet une structuration des connaissances moins rigide et plus transparente que celle des bases de données.

Dans ce contexte, le modèle ONTODIC (Alcina 2025) constitue un modèle de représentation du langage et des connaissances linguistiques fondé sur la logique descriptive, intégrant les principes fondamentaux de la linguistique et de l'ingénierie des connaissances. Il permet d'expliciter les relations entre les données linguistiques et de les regrouper en catégories en fonction de leurs propriétés. La logique inhérente au modèle et la technologie de l'ingénierie des connaissances offrent un instrument permettant de vérifier la conformité des données et des postulats en appliquant un raisonnement logique. Le modèle est un outil indépendant de toute théorie linguistique, car sa fonction est de faciliter la représentation et l'enchaînement logique de ses éléments. Dans nos axes de travail, dans le cadre du projet UNE-ONTODIC, nous proposons de formaliser les postulats de différentes théories et phénomènes linguistiques à l'aide de ce modèle de représentation.

Dans ce travail, nous présentons nos avancées dans la formalisation des fonctions lexicales à l'aide du modèle ONTODIC. Conformément à ce modèle, nous partons de la distinction entre les éléments du langage (mots, syntagmes, etc.) et la conceptualisation que nous en faisons dans le système linguistique. Ainsi, nous séparons les désignations de leur(s) signification(s) et nous distinguons formellement entre les relations syntagmatiques, qui existent entre les désignations et qui se manifestent dans les textes, et les relations sémantiques, qui relient les significations associées aux désignations. Cette distinction favorise une compréhension plus précise du fonctionnement linguistique.

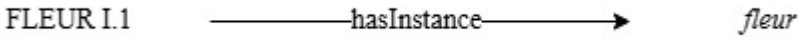


Figure 4 Relation entre le signifié FLEUR I.1 *et l'unité lexicale* fleur

Pour reprendre l'exemple de *fleur I.1*, en suivant ONTODIC, nous distinguons l'unité *fleur* et la signification qui lui est associée, FLEUR I.1.

Cela nous permet de distinguer les relations entre les unités lexicales et les relations entre les significations. La fonction *IncepFunc$_0$* explicite une relation lexicale entre les unités *fleur* et *éclore*, tandis que *Holo* explicite une relation au niveau sémantique, qui s'établit entre les significations explicitées dans FLEUR I.1 et PLANTE[1] (Figure 5).

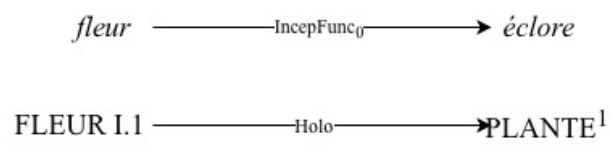


Figure 5 Formalisation des relations de fleur I.1 *selon ONTODIC*

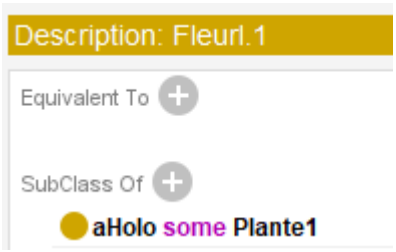


Figure 6 Implémentation de la fonction Holo *sur Protégé*

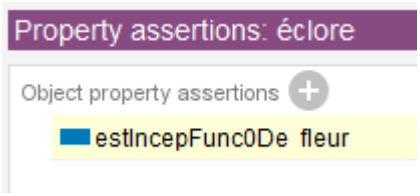


Figure 7 Implémentation de la fonction IncepFunc_0 *sur Protégé*

Nous avons implémenté ce modèle dans l'éditeur d'ontologies Protégé pour créer une base de connaissances, et nous avons vérifié sa consistance à l'aide du raisonneur de cet outil, qui a permis de détecter d'éventuelles erreurs et d'obtenir automatiquement des inférences. La Figure 6 présente la relation entre les significations. La Figure 7 montre une inférence obtenue à partir de la relation entre les unités *fleur* et *éclore*.

Le cas de *fleur I.1* nous a permis de formaliser et de mettre en œuvre une entrée de cette ressource en suivant ONTODIC, et les résultats obtenus indiquent que ce modèle permet sa formalisation et sa mise en œuvre dans une base de connaissances, ce qui permet d'analyser en détail les informations lexicales, sémantiques et linguistiques associées aux unités lexicales, ainsi que les informations inhérentes à l'usage de chacune des fonctions.

En définitive, ONTODIC peut être utilisé dans des ressources fondées sur la LEC et permet d'intégrer les fonctions lexicales comme outil d'aide à la formalisation et à la mise en œuvre de relations sémantiques et lexicales.

Bibliographie :

Alcina, Amparo. 2025. «ONTODIC. A model of linguistic knowledge representation based on description logic». *Terminology* 31 (2): 208-37. https://doi.org/10.1075/term.23030.alc.

Baader, Franz, Dresdenian Horrocks, Carsten Lutz, et Uli Sattler. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Lux-Pogodalla, Veronika, y Alain Polguère. 2011. «Construction of a French Lexical Network: Methodological Issues». 54-61. https://hal.science/hal-00686467v1/document.

Mel'čuk, Igor, André Class, et Nadia Arbatcewsky-Jumarie. 1984. *Dictionnaire explicatif et combinatoire du français contemporain : recherches lexico-sémantiques*. Les Presses de l'Université de Montréal.

Mel'čuk, Igor, André Class, et Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Duculot.

Polguère, Alain. 2014. «From Writing Dictionaries to Weaving Lexical Networks». *International Journal of Lexicography* 27 (4): 396-418.

Angelina Aleksandrova
Paris Université

Entre lexique et phraséologie : modéliser la dénomination des mouvements sportifs

L'étude du lexique spécialisé constitue depuis plusieurs décennies un champ privilégié pour interroger les relations entre langue, savoir et pratique. De nombreux travaux ont montré que les lexiques de spécialité, loin de se réduire à un simple répertoire terminologique, constituent des systèmes dynamiques où les unités lexicales reflètent et organisent des connaissances expertes (Temmerman 2000, Maniez 2002, Thoiron & Béjoint 2010 ; L'Homme, 2020 ; L'Homme et Polguère, 2008). Leur description suppose ainsi d'articuler plusieurs niveaux d'analyse (morphologique, sémantique, phraséologique et discursif) afin de rendre compte à la fois de la structuration interne des dénominations et de leur mise en usage dans des contextes situés (Legallois & Tutin, 2013 ; Corpas Pastor & Mitkov, 2019). Dans cette perspective, le lexique du sport, et tout particulièrement celui des dénominations des mouvements de fitness, offre un terrain d'observation singulier : à la croisée de la terminologie technique et du langage ordinaire (Galisson 1978, Petiot 1975, Ligas 2010, Gomez Fernandesz 2021), il se caractérise par une forte productivité et une tension constante entre standardisation disciplinaire et créativité discursive. Étudier ces dénominations revient donc à interroger la manière dont une pratique intrinsèquement ancrée dans la temporalité de l'action, façonne ses outils lexicaux[1[1]], entre contrainte normative et inventivité.

Cela étant dit, le linguiste est rapidement confronté à la difficulté de recensement de ce vocabulaire transversal, en partie liée aux conditions même de circulation des termes. Ce vocabulaire évolue dans un écosystème discursif mouvant où coexistent plusieurs communautés de pratique (fédérations, clubs, médias, applications numériques, cf. Krebs 2018) sans instance normalisatrice unifiée (Robin et al. 1987). Cette pluralité engendre des variations terminologiques, référentielles et morphosyntaxiques qui traduisent des différences de conventions plus que de véritables innovations lexicales. À cela s'ajoute une forte perméabilité interlinguistique : la présence massive de l'anglais favorise l'instabilité des usages, certaines formes francisées demeurant en concurrence avec leurs équivalents originaux (*burpee / saut de grenouille, clean & jerk / épaulé-jeté*). Enfin, dans le discours *in situ* des coachs, la tendance à l'ellipse, à la réduction ou à la reformulation pragmatique contribue à entretenir une impression d'instabilité, alors qu'il s'agit souvent d'ajustements contextuels. L'inventaire lexical du sport apparaît ainsi non comme un espace de création libre, mais comme un système en réorganisation permanente, modelé par la diversité des pratiques, la dynamique des emprunts et les contraintes interactionnelles propres au discours sportif.

L'étude proposée décrit et modélise les dénominations françaises des mouvements d'entraînement (corpus LEXMOUV-FR) qui ne relèvent pas d'une discipline sportive unique mais construisent un lexique transversal caractéristique des pratiques dites "fonctionnelles"[2[2]] qui agrègent des gestes issus de disciplines distinctes (haltérophilie, gymnastique, sports d'endurance et de force athlétique, etc.) sous une grammaire commune du mouvement. Ce lexique ne se laisse donc pas assigner à un domaine sportif fermé : il forme plutôt un répertoire interdisciplinaire de mouvements, fondé sur des logiques biomécaniques (tirer, pousser, sauter, soulever, courir, etc.) et des structures phraséologiques récurrentes (verbe + site, nom + modifieur, etc.). Au-delà des frontières disciplinaires sportives, ce lexique forme un réservoir lexical partagé entre tous les acteurs du sport, en allant des athlètes confirmés jusqu'aux pratiquants loisirs, en passant par les coachs, les préparateurs physiques et les médecins du sport.

L'objectif de notre communication sera double : (i) caractériser les patrons de formation (schèmes morpho-syntaxiques et sémantiques) qui structurent les dénominations étudiées[3[3]] ; (ii) tester la prévisibilité de nouvelles

1[1] Cf. l'action de *conventionnalisation* (Rodríguez, 2020).
2[2] Telles que le *CrossFit®* ou le *crosstraining*, regroupés sous la dénomination de fitness fonctionnel.
3[3] Selon une grille analytique *ad hoc*, construite sur la base de différents travaux (cf. notamment Cabezas-García et al. (2019), Charneau (2022), Kauffer & Keromnes (2022), Gómez Fernandez (2019), Mejri (2020), Mizouri (2020)).

dénominations par des patrons productifs. L'étude sera articulée à une observation en discours : *in situ*, les coachs recourent rarement à la dénomination « complète » (ellipses, troncations, variantes syntagmatiques), ce qui invite à relier la forme lexicale (nom canonique) aux préférences phraséologiques en usage.

Notre échantillon de travail est composé de 938 dénominations françaises (*traction supination, soulevé de terre jambes tendues*) ou hybrides (au moins un composant lexical en français est présent – *push press à la barre*, *overhead fente marchée*). Il est structuré en segments étiquetés selon quatre plans : LEM (lemme), SEM (rôle sémantique), TAG (catégorie morpho-syntaxique) et POS (rang dans la chaîne). Un format long (avec 1 segment/ligne) qui permet de calculer :

a. les motifs de construction (séquences de TAG),

b. les distributions des têtes (N_TETE) et des modifieurs,

c. les forces d'association entre têtes et différents modifieurs (matériel, sites, modalité d'exécution, etc.), et

d. la productivité des patrons (couverture des formes attestées).

Afin de rendre compte de la structuration interne du lexique et des régularités phraséologiques qui le sous-tendent, l'étude croise une description des schèmes morphosyntaxiques de formation et s'appuie sur trois mesures complémentaires d'attirance lexicale. La PMI permet l'identification des associations spécifiques et souvent rares, qui révèlent la créativité lexicale ; le t-score pondère ces résultats par la fréquence des combinaisons stables et représentatives des patrons productifs ; enfin, la log-vraisemblance-G^2 nous fournit une estimation robuste de la dépendance statistique entre les éléments têtes et les modifieurs. L'articulation de ces trois mesures permettra de dégager différents régimes d'association lexicale : les patrons canoniques, fréquents et institutionnalisés, les patrons périphériques, plus rares mais sémantiquement marqués, et les patrons émergents, porteurs d'innovation terminologique.

Enfin, l'articulation avec le discours in situ est abordée via une table de correspondance entre forme canonique et réalisation phraséologique (abréviations, ellipses, anglicismes, calques) dans un corpus de discours de coachs en ligne. On observe que des noyaux nominaux (p. ex. *développé, fentes, pompes*) fonctionnent comme des ancres, tandis que certains modifieurs (*unilatéral, latéral, nuque*) sont régulièrement récupérés ou élidés au profit de consignes procédurales (ex. *bras tendus*, *prise serrée*). L'apport attendu porte sur (i) une typologie opératoire des patrons de dénomination des mouvements de fitness, (ii) un modèle de prévisibilité fondé sur la productivité des schèmes, et (iii) une interface lexique/phraséologie rendant compte de la dynamique entre le nom du mouvement et ses réalisations en discours. Au-delà du domaine sportif, nous espérons apporter des éléments de réflexion sur la modélisation des lexiques spécialisés en contexte d'action, en montrant comment les pratiques langagières façonnent la structure et la vitalité d'un lexique professionnel en constante reconfiguration.

Bibliographie :

Cabezas-García M., León-Araúz P. (2019). « On the Structural Disambiguation of Multi-word Terms » Corpas Pastor G., Mitkov R. (dirs.), Computational and Corpus-Based Phraseology, p. 46-60.

Charneau C. (2022). « Phraséologie spécialisée et actes professionnels », dans Mangiante J.-M., Parpette C. (dirs.), Le FOS Aujourd'hui. Etat de la recherche en Français sur Objectif Spécifique., Peter Lang, Bruxelles, p. 143-161.

Colson J.-P. (2020). « XIX. La phraséologie, discipline linguistique autonome ou intégrée ? Arguments constructivistes et fondés sur les corpus », dans La phraséologie française en questions, Hermann, p. 309-321.

Colson J.-P. (2022). « Grammaire de construction et phraséologie : des corpus aux réflexions théoriques », Langages, 225, p. 19-32.

Corpas Pastor G. (2019). Computational and Corpus-Based Phraseology: Third International Conference, Europhras 2019, Malaga, Spain, September 25-27, 2019, Proceedings, Springer International Publishing AG, Cham (Lecture Notes in Computer Science Ser).

Diwersy S., Legallois D. (2019). « L'apport de la méthode des motifs aux analyses phraséologiques en discours », dans Kauffer M., Keromnes Y. (dirs.), Approches théoriques et empiriques en phraséologie, Eurogermnaistik.

Galisson R. (1978). Recherches de lexicologie descriptive : la banalisation lexicale, Nathan, Paris.

Gómez Fernandez A., Uzcanga Vivar I. (2019). « Phrasèmes dans le discours de l'alimentation et du sport: à toute vapeur, à toute vitesse », dans Fraseologia e paremiologia: Prospettive evolutive, pragmatica e concettualizzazione, Edizioni Accademiche Italiane, p. 515-534.

Kauffer M., Keromnes Y. (2022). « Aspects de la recherche actuelle en phraséologie. Présentation », Langages, 225, n° 1, p. 9-18.

Keromnes Y. (2022). « Idiomaticité, constructions et variations : approche contrastive sur corpus », Langages, 225, n° 1, p. 33-47.

Krebs G. (2018). « Sport : réalités et discours », dans Tournadre J.-F. (dir.), Pour une histoire du sport et de la jeunesse, Presses Sorbonne Nouvelle, Paris (Monde germanophone).

L'Homme M.-C. (2020). Lexical semantics for terminology, John Benjamin Publishing Company, Amsterdam ; Philadelphia.

L'Homme M.-C., Polguère A. (2008). « Mettre en bons termes les dictionnaires spécialisés et les dictionnaires de langue générale », dans Maniez F., Dury P. (dirs.), Lexicographie et terminologie : histoire de mots. Hommage à HenriBéjoint, Presses de l'Université de Lyon, Lyon, p. 191-206.

Ligas P. (2010). « Parlez-vous le sportif? La langue des sports de trouvailles en clichés », dans Dotoli G., Ligas P., Selvaggio M. (dirs.), Les langues du sport, Schena Editore, p. 41-49.

Mejri S. (2020). « Introduction. La phraséologie française en questions », dans La phraséologie française en questions, Hermann, p. 7-11.

Mizouri I. (2020). « VIII. La discontinuité polylexicale : espace d'enchaînement discursif ? », dans La phraséologie française en questions, Hermann, p. 117-140.

Petiot G. (1975). « La langue des sports », La banque des mots, n° 10, p. 155-185.

Robin M., Carrère P., Mondenard J.-P. (1989). « Le vocabulaire de l'entraînement physique », Ann. Kinésithér., 16, n° 6, p. 261-286.

Rodríguez C.D. (2020). « Révision taxonomique des unités phraséologiques lexicales », dans La phraséologie française en questions, Hermann, p. 47-58.

Temmerman R. (2000). « Une théorie réaliste de la terminologie: le sociocognitivisme », Terminologies Nouvelles, 21, p. 58-64.

Lana Benett
Université Grenoble Alpes

Lexique et argumentation dans le discours climatosceptique états-unien : l'exemple d'un *think tank* conservateur (2021-2024)

Le lexique constitue un espace privilégié d'observation des phénomènes argumentatifs : les mots ne sont pas de simples véhicules d'information, mais des instruments de positionnement par rapport aux objets du discours. Cela est particulièrement manifeste dans les discours relatifs au changement climatique, où le choix des termes joue un rôle central dans l'acceptation, ou au contraire, la négation du phénomène. En effet, alors que le changement climatique est aujourd'hui largement reconnu comme un enjeu central, il continue de faire l'objet de discours climatosceptiques qui cherchent à en relativiser l'importance ou à en décourager la prise en charge. Ces discours à dimension argumentative (au sens large ; voir Amossy, 2018) ne reposent pas uniquement sur une négation frontale de la réalité climatique. Ils présentent deux approches terminologiques :

1. La première consiste à reprendre les termes officiels recensés par les bases de données consacrées au lexique du changement climatique, souvent en les détournant ou en leur donnant un sens différent. Par exemple, *climate crisis* deviennent des mots ironisés, parfois placés entre guillemets, pour dénoncer une supposée exagération médiatique ou politique.

2. La seconde repose sur l'emploi d'une autre terminologie, en évitant ces termes officiels pour désigner et dénigrer les phénomènes liés au changement climatique, comme l'expression *climate alarmism*.

Ces emplois participent à la mobilisation de procédés rhétoriques permettant de construire le changement climatique comme un phénomène inexistant, secondaire, naturel, ou coûteux à traiter afin d'entraver l'action climatique (Lamb et al, 2020). La présente soumission, qui s'inscrirait dans la sous-thématique « argumentation » du colloque « Lexiques », s'intéresse à la manière dont ces termes apparaissent dans les publications d'un *think tank* états-unien conservateur et influent, *The Heritage Foundation*, régulièrement associé à des positions climatosceptiques. Ce choix relève du fait que les *think tanks*, qu'ils soient d'orientation conservatrice et/ou libertarienne, participent considérablement à la diffusion de discours climatosceptiques et à la délégitimation du consensus scientifique sur le changement climatique (Busch et Judick, 2021 ; Dunlap & Jacques, 2013 ; Dunlap & McCright, 2015), s'inscrivant dans une rhétorique de contre-discours (Hourcade & Wagener, 2021). La question sous-jacente qui guide donc cette étude est la suivante : comment la terminologie du discours climatosceptique construit-elle le phénomène comme un problème inexistant ou insignifiant ?

Le corpus analysé est constitué de 125 énoncés extraits de 37 articles de type *commentaries* (tribunes), publiés dans la section "*climate*" du site de la *Heritage Foundation* au cours du mandat du président Joe Biden. L'approche adoptée est qualitative et quantitative et repose sur une grille de lecture qui articule positionnements vis-à-vis du changement climatique et stratégies sceptiques.

Le choix de la période se justifie par l'importance accordée à la question climatique dans l'agenda de l'administration Biden, qui a consacré des efforts sans précédent en faveur de la transition énergétique (Collomb, 2024a ; 2024b). Nous faisons l'hypothèse que ces mesures ont créé un terrain propice aux réactions climatosceptiques à droite de l'échiquier politique et qu'elles sont présentes dans les publications de la *Heritage Foundation*.

Il existe un certain nombre de *think tanks*. Celui-ci a été sélectionné, et tout premier lieu, du fait de ses positionnements ouvertement climatosceptiques (Washington et al., 2011). De plus, il présente une catégorie "*climate*" rassemblant un grand nombre de publications en tous genres. Nous avons fait le choix des *commentaries* (tribunes) car ces derniers représentent la part la plus importante des articles publiés par le *think tank*.

Les passages sélectionnés sont ceux où le changement climatique est évoqué, de manière complète (*climate change, climate crisis, the climate*, etc.) ou partielle (*CO2 levels, temperatures, greenhouse gas emissions*, etc.). Nous avons également retenu les occurrences mentionnant les politiques de lutte climatique. Les énoncés retenus correspondent aux différentes approches sceptiques identifiées par la littérature (Capstick et Pidgeon, 2014) : un scepticisme scientifique, qui fait référence à la minimisation ou à la négation de la réalité scientifique du changement climatique, de ses causes et conséquences ; et un scepticisme « social/comportemental », qui concerne les réponses critiques aux niveaux individuel, sociétal et politique des mesures de lutte. Notre démarche s'inscrit dans une approche énonciativiste et argumentative.

Les résultats confirment que les discours climatosceptiques contemporains ne se réduisent pas à un refus frontal de la réalité du changement climatique, mais reposent aussi sur une grande variété de procédés discursifs qui visent aussi à réduire la gravité du phénomène et à critiquer les réponses mises en œuvre.

L'analyse des mots utilisés pour évoquer le changement climatique fait apparaître deux directions argumentatives (qui ne s'excluent pas mutuellement et peuvent coexister au sein d'un même passage) :

1. Des stratégies, qui, d'une part, construisent l'ethos des énonciateurs (Amossy, 2010) climatosceptiques pour renforcer leur légitimité.

2. D'autre part, des stratégies qui diminuent celle de leurs antagonistes pro-climat.

Dans le renforcement de leur propre camp, se trouvent les stratégies de polémique (Amossy, 2014) qui reposent principalement sur la polarisation, le dialogisme, la politisation du phénomène et les arguments réactionnaires des conséquences négatives (inanité, effet pervers et mise en danger) comme envisagés par Hirschman (1991). On trouve aussi les énoncés légitimant ou renforçant la voix de l'énonciateur. Ces derniers passent souvent par le recours à l'autorité scientifique : « *We are constantly reminded that we are experiencing a climate crisis, but as a climate scientist, I can tell you that's not what the science has shown us so far.* ».

Dans la catégorie diminuante du camp opposé, se trouvent les énoncés qui minimisent l'objet (le changement climatique lui-même, ses effets, les mesures de luttes mises en place, etc.) et qui passent souvent par un cadrage partiel du changement climatique, ce qui permet son euphémisation, sa recatégorisation, sa naturalisation, ou tout autre stratégie qui permette de diminuer son existence ou son importance : « *Other than modest warming, there has been little change in any kind of severe weather that can be attributed to global greenhouse gas emissions.* ». Les stratégies diminuantes passent également par la minimisation de la voix énonciative antagoniste et ont recours au dialogisme et à la polarisation.

Références bibliographiques :

Amossy, Ruth, 2014. Apologie de la polémique. *Presses Universitaires de France*. https://doi.org/10.3917/puf.amos.2014.01.

Amossy, Ruth. 2018. Introduction : la dimension argumentative du discours - enjeux théoriques et pratiques. *Argumentation et Analyse du Discours*, nº 20 (avril): 20. https://doi.org/10.4000/aad.2560.

Amossy, Ruth. 2010. La présentation de soi : Ethos et identité verbale. *Presses Universitaires de France*. https://doi.org/10.3917/puf.amoss.2010.01.

Busch, Timo & Judick, Lena. 2021. Climate Change—That Is Not Real! A Comparative Analysis of Climate-Sceptic Think Tanks in the USA and Germany. *Climatic Change* 164 (1): 18. https://doi.org/10.1007/s10584-021-02962-z.

Capstick, Stuart Bryce & Pidgeon, Nicholas Frank. 2014. What *is* climate change scepticism? Examination of the concept using a mixed methods study of the UK public. *Global Environmental Change* 24 (janvier): 389401. https://doi.org/10.1016/j.gloenvcha.2013.08.012.

Chetouani, Lamria. 2007. Les mots de la controverse sur le changement climatique. *Le Télémaque* (Caen) 31 (1): 81104. https://doi.org/10.3917/tele.031.0081.

Collomb, Jean-Daniel. 2024a. Croître et décarboner : la quadrature du cercle énergétique de l'administration Biden ? *Recherches internationales* 130 (3): 85100. https://doi.org/10.3917/rein.130.0085.

Collomb, Jean-Daniel. 2024b. La transition énergétique démocrate et la crédibilité des États-Unis. *IRIS*. https://prod.iris-vt-prod-lamp01.dcsrv.eu/ris/la-transition-energetique-democrate-et-la-credibilite-des-etats-unis/.

Dunlap, Riley E. & Jacques, Peter J.. 2013. Climate Change Denial Books and Conservative Think Tanks: Exploring the Connection. *American Behavioral Scientist* 57 (6): 699731. https://doi.org/10.1177/0002764213477096.

Dunlap, Riley E. & McCright, Aaron M. 2015. Challenging Climate Change: The Denial Countermovement. In *Climate Change and Society: Sociological Perspectives*, édité par Riley E. Dunlap et Robert J. Brulle. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199356102.003.0010.

Florea, Marie-Laure. 2020. Postures énonciatives & Dialogisme intralocutif. *Corela, Cognition, représentation, langage*, nº HS-32 (novembre): HS-32. https://doi.org/10.4000/corela.12087.

Fløttum, Kjersti, Gjerstad, Øyvind & Oloko, Francis Badiang. 2019. Les voix dans le discours climatique : essai d'une combinaison de la polyphonie avec l'analyse de contenu, l'analyse narrative et l'analyse rhétorique. *Cahiers de praxématique*, nº 73 (décembre): 73. https://doi.org/10.4000/praxematique.5747.

Gaillard, Mathilde. 2015. Entre parole d'expert et visée programmatique : les policy briefs, genre discursif propre aux think tanks américains. *ASp. la revue du GERAS*, nº 67 (mars): 67. https://doi.org/10.4000/asp.4647.

Haltinner, Kristin & Sarathchandra, Dilshani. 2021. Considering attitudinal uncertainty in the climate change skepticism continuum. *Global Environmental Change* 68 (mai): 102243. https://doi.org/10.1016/j.gloenvcha.2021.102243.

Hirschman, Albert O. 1991. *The Rhetoric of Reaction: Perversity, Futility, Jeopardy*. Harvard University Press.

Hourcade, Renaud & Wagener, Albin. 2021. Le climatoscepticisme : une approche interdiscursive. *Mots. Les langages du politique* 127 (3): 922.

IPCC. 2023. Annex I: Glossary In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 119-130, doi: 10.59327/IPCC/AR6-9789291691647.002.

Jacques, Peter J., Dunlap, Riley E. & Freeman, Mark. 2008. The organisation of denial: Conservative think tanks and environmental scepticism. *Environmental Politics* 17 (3): 34985. https://doi.org/10.1080/09644010802055576.

Jaubert, Anna. 2008. Dire et plus ou moins dire. Analyse pragmatique de l'euphémisme et de la litote. *Langue française* (Paris) 160 (4): 10516. https://doi.org/10.3917/lf.160.0105.

Lamb, William F., Mattioli, Giulio, Levi, Sebastian et al. 2020. Discourses of Climate Delay. *Global Sustainability* 3 (janvier): e17. https://doi.org/10.1017/sus.2020.13.

Medimorec, Srdan & Pennycook, Gordon. 2015. The language of denial: text analysis reveals differences in language use between climate change proponents and skeptics. *Climatic Change* 133 (août). https://doi.org/10.1007/s10584-015-1475-2.

Plantin, Christian. 2017. Types, typologies, arguments. *Travaux neuchâtelois de linguistique*, nº 65 (janvier): 6778. https://doi.org/10.26034/tranel.2017.2011.

Rabatel, Alain. 2013. Positions, positionnements et postures de l'énonciateur. *Linha D'Água* 26 (décembre): 159. https://doi.org/10.11606/issn.2236-4242.v26i2p159-183.

Washington, Haydn, Cook, John & Oreskes, Naomi. 2011. *Climate Change Denial: Heads in the Sand*, Routledge. https://doi.org/10.4324/9781849776738.

Thomas Bertin / Sofian Berrouiguet / Ambre Marie
Univ. de Rouen Normandie / CHU de Brest/ Univ. de Bretagne occidentale

Gestes, passages à l'acte et idées noires - désigner le suicide pour évaluer le risque de récidive

1. Contexte

Avec plus de 700 000 décès recensés chaque année dans le monde, le suicide demeure une cause majeure de mortalité (GBD-2021 Suicide Collaborators 2025) dont les facteurs de risque sont multiples et complexes (Baudelot & Establet 2018). La France n'est pas épargnée et la prévention du suicide reste « un enjeu majeur de santé publique » (Ministère de la Santé).

Dans ce cadre, depuis 2015, le dispositif *VigilanS* a été déployé sur l'ensemble du territoire. Il consiste en un système de « recontact et d'alerte » permettant un suivi à distance d'un patient *suicidant* (*i.e.* « ayant fait une tentative de suicide ») qui a accepté d'intégrer le dispositif au terme de son hospitalisation. Dès lors, à tout moment, il peut contacter la plateforme d'écoute. Parallèlement, *VigilanS* assure un suivi *via* l'envoi de cartes postales personnalisées et des appels téléphoniques qui doivent permettre une évaluation de son état de santé mentale (site du Ministère de la Santé).

Le dispositif fait l'objet de recherches au laboratoire LaTIM (Brest) qui voient collaborer psychiatres, spécialistes de l'IA et linguistes. Dans cette présentation, nous nous focalisons sur les **communications téléphoniques** dont l'enjeu est crucial : le *vigilanseur*, infirmier psychiatrique, doit évaluer le risque suicidaire pour accompagner et orienter au mieux le patient.

2. Cadre et enjeux de l'analyse linguistique

Nous proposons une **analyse du lexique en pratique clinique** ancrée dans la tradition de l'analyse du discours en milieu médical – cf. Ledegen & Wagener (2020) pour un exemple en prévention du suicide – et, plus spécifiquement ici, des interactions verbales (Blasco 2022).

L'entrée lexicale conduit à s'inscrire dans le cadre de l'analyse de discours à entrée lexicale où, dans un souci de prise en compte de la dynamique interactionnelle, les mots sont « amenés à être dépassés, de par leur solidarité avec la chaîne syntagmatique et avec le discours » (Née & Veniard, 2012 : 20). Ce dépassement est facilité par le recours à la notion de *chaine de référence* (Schnedecker 2021).

Nous étudions la manière dont les tentatives de suicide passées et les potentiels futurs « passages à l'acte » sont évoqués et désignés. D'une part, nous identifions les termes employés – notamment *passage à l'acte* ou *idées noires* sans négliger *suicide*, *suicidaire*, *mort* ou *mourir* – dont certains peuvent s'avérer ambigus (cf. *geste* et même *idées noires*). D'autre part, nous explorons leur contexte d'emploi au cours de l'entretien téléphonique. Nous cherchons ainsi à **saisir comment le vigilanseur évalue le risque de récidive**.

3. Des données au corpus

D'une part, le protocole de recherche a permis l'accès à une base de 3628 entretiens enregistrés du 7 mars 2019 au 13 septembre 2021. Il s'agit d'appels dits « entrants » (*i.e.* le patient est l'appelant) même si certains sont des « faux-entrants » (le patient rappelle suite à un appel manqué du vigilanseur). L'analyse qualitative ne prend pas en compte l'intégralité de ces enregistrements. Cependant, cette vaste **BASE** fait l'objet d'analyses automatiques complémentaires reposant sur la transcription et l'exploitation de modèles de classification et de langage (Glaz *et al.* 2021).

D'autre part, nous avons extrait de manière aléatoire 20 enregistrements de BASE pour constituer un **CORPUS**. Ces 20 enregistrements représentent 7 heures d'entretien avec 19 patients (16 femmes et 3 hommes) et 6 vigilanseurs.

4. Analyse

Nous présenterons une analyse en trois étapes :

Caractérisation du cadre interactionnel. Le cadre interactionnel se singularise par une très nette complémentarité entre un patient *vulnérable* (Ploog & Verdier 2025) et un professionnel de santé contraint de s'adapter à tout ce qui « survient » au cours d'un entretien marqué par la distance physique entre les inter-actants et une grande imprévisibilité. L'analyse nécessite de prendre en compte les rôles sociaux déterminés par le cadre institutionnel (Cappeau 2022) mais surtout de caractériser les relations sociale et interlocutive (Vion 1996) telles qu'elles se font jour dans CORPUS.

Découpage en séquences. Les vigilanseurs s'adaptent au tour que prend l'échange et ne suivent pas un strict protocole d'entretien. Il reste possible de mettre au jour une macro-structure type en croisant l'analyse automatique de BASE – *via* la transcription automatique (Bain *et al.* 2023) et le recours à un modèle de langage (Beno 2024) – et l'analyse manuelle de CORPUS. Pour cette dernière, dans un perspective sémantique et pragmatique, nous nous appuyons sur les notions de *séquence* (Kerbrat-Orecchioni 1998) et de *direction d'ajustement* (Searle 1982) afin de distinguer phases d'évaluation et phases de conseil.

Analyse des modes de désignation du suicide. Le séquençage met au jour trois séquences de nature évaluative qui apparaissent cruciales. La première cherche à jauger un état général en réponse à la question « comment ça va ? » (cf. Beaud *et al.* 2009). Ce questionnement anticipe l'évocation du suicide mais la fait rarement émerger directement (sauf à l'initiative, rare, du patient).

On trouve presque toujours une deuxième séquence concernant des tentatives de suicide passées dont le vigilanseur (VIG) prend généralement l'initiative de la désignation, de façon plus ou moins explicite (cf. *tentative de suicide* vs. *geste*) selon son « style interactionnel » mais aussi selon le contexte d'appel et la situation du patient (PAT) :

VIG	c'était la p- la première fois c'était: y a combien de temps\
PAT	ouh là y a trois ans
VIG	premier **geste** y a trois ans donc euh super jeune parce que t- (.) t'as seize ans hein c'est ça/

Celle-ci s'articule plus ou moins directement à une troisième consacrée au risque de récidive dont l'évaluation reste un objectif central de l'entretien. Là encore, les désignations sont plus ou moins explicites et s'opèrent parfois par ajustement lexicaux. Selon le patient, elle conduit à des réactions très diverses, parfois déroutantes, qu'une expertise clinique permet au vigilanseur d'interpréter :

VIG	d'accord (.) et euh est-ce qu'y a **des idées noires** là monsieur *** **des idées de suicide** est-ce qu'il y en a à nouveau des choses comme ça (.) est-ce que vous y pensez
PAT	un petit peu

Nous examinons les **désignations du suicide** dans ces deux derniers types de séquences :

1. À quel moment de l'entretien émergent-elles et par quels moyens lexicaux ?
2. Comment sont-elles prises en charge, interprétées et éventuellement reprises ?
3. En quoi permettent-elles potentiellement au vigilanseur d'évaluer le risque de récidive ?

Dans BASE (enrichie de métadonnées « patient »), ces désignations sont utilisées comme paramètres d'entrée pour des modèles de classification supervisée afin de mettre au jour automatiquement des corrélations entre mode de désignation du suicide et risque de récidive (Shin *et al.* 2022).

Références bibliographiques :

Bain M., Huh J., Han, T. & Zisserman A. (2023). WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. *ArXiv*, 4489-4493. https://doi.org/10.48550/arXiv.2303.00747

Baudelot C. & Establet R. (2018 / 2006). *Suicide – L'envers de notre monde*. Paris : Seuil.

Beaud L., Cahagne V. & Guyard H. (2009). « Alors ? Comment allez-vous ? » : Entrée en matière dans une consultation médicale. *Langage et société*, 126(4), 5774. https://doi.org/10.3917/ls.126.0057

Beno J. (2024). ELECTRA and GPT-4o: Cost-Effective Partners for Sentiment Analysis. *ArXiv*, abs/2501.00062. https://doi.org/10.48550/arXiv.2501.00062

Blasco M. (Éd.). (2022). *Parler à l'hôpital : Écouter ce qui est dit, décrypter ce qui se dit*. Münster : Nodus Publikationen.

Cappeau P. (2022). Le jeu du pronom *je* et des verbes. In : M. Blasco (éd.), Parler à l'hôpital. Münster : Nodus Publikationen, 63-74.

GBD-2021 Suicide Collaborators (Davis Weaver N. *et al.*) (2025). Global, regional, and national burden of suicide, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet Public Health*, 10(3). https://doi.org/10.1016/S2468-2667(25)00006-4

Glaz, A. L., Haralambous, Y., Kim-Dufor, D.-H., Lenca, P., Billot, R., Ryan, T. C., Marsh, J., DeVylder, J., Walter, M., Berrouiguet, S., & Lemey, C. (2021). Machine Learning and Natural Language Processing in Mental Health : Systematic Review. *Journal of Medical Internet Research*, *23*(5), e15708. https://doi.org/10.2196/15708

Kerbrat-Orecchioni C. (1998 / 1990). *Les interactions verbales – Tome 1*. Paris : Armand Colin.

Ledegen G. & Wagener A. (2020). « Nous ne doutons pas de votre souffrance » : Analyse pragmatique et sociolinguistique du *nous* de distanciation dans un chat de prévention du suicide. *Corpus* 21. https://doi.org/10.4000/corpus.4977

Ministère du travail, de la santé, des solidarités et des familles (Site officiel). https://sante.gouv.fr/prevention-en-sante/sante-mentale/promotion-et-prevention/la-prevention-du-suicide/ [page consultée le 22 mai 2026]

Née É. & Veniard M. (2012). Analyse du Discours à Entrée Lexicale (A.D.E.L.) : Le renouveau par la sémantique ? *Langage et société* 140(2), 1528. https://doi.org/10.3917/ls.140.0015

Ploog K. & Verdier M. (dir.) (2025). *Vulnérabilités en situation.* Éditions de l'Université de Lorraine.

Schnedecker C. (2021). *Les chaînes de référence en français*. Paris : Éditions Ophrys.

Searle J. (1982 / 1979). *Sens et expression*. Paris : Minuit.

Shin D., Kim K., Lee S., Lee C., Bae Y., Cho W., Kim M., Park H., Chie E., Kim N. & Ahn, Y. (2022). Detection of Depression and Suicide Risk Based on Text From Clinical Interviews Using Machine Learning: Possibility of a New Objective Diagnostic Marker. *Frontiers in Psychiatry* 13. https://doi.org/10.3389/fpsyt.2022.801301.

Vion R. (1996). L'analyse des interactions verbales. *Les carnets du Cediscor* 4, 1932. https://doi.org/10.4000/cediscor.349

Valentina Bisconti
Université de Picardie

Les expressions polylexicales contenant des noms de nombres : de la cardinalité idiomatique à l'expression du vague

Les expressions polylexicales sont des blocs très hétérogènes en termes de composition interne et de comportement morphosyntaxique allant d'un minimum à un maximum d'agglutination. Elles le sont aussi du point de vue des valeurs sémantiques qu'elles peuvent encoder. La recherche en lexicologie et en phraséologie (Gross 1996, Mel'cuk 1998, Wray 2002, Goldberg 2003, Grossmann & Tutin 2003, Libben & Jarema 2006, Burger *et al.* 2007, Mejri 2007, Frath 2008, Wray 2018, Tutin 2008 & 2010, Anscombre & Mejri 2011, Bouveret & Legallois 2012, Legallois & Tutin 2013, Simone & Piunno 2017, Legallois & Patard 2017) a montré que le mécanisme qui rend possible les constructions lexicales idiomatiques - *idiom principle* (Sinclair 1991) - sélectionne toute une gamme de structures et de catégories pour se réaliser.

Dans cette communication nous étudierons les expressions polylexicales contenant des noms de nombres (locutions verbales et nominales) et explorerons la capacité des nombres[4[1]] à désigner une valeur autre qu'arithmétique quand ils interviennent dans des structures polylexicales plus ou moins contraintes (Dehaene 1997, Pollmann & Jansen 2001, Bazzanella 2011, Planchon 2012). Autrement dit, les noms de nombres n'expriment pas la cardinalité mais des significations idiomatiques qui expriment des valeurs indéterminées et approximatives, parfois ambivalentes : en français *quatre* peut désigner « la petite et la grande quantité » (*quatre pelées et un tondu* vs *avoir de l'esprit comme quatre*), *trente-six (mille)* peut désigner tout à la fois « la quantité indéterminée et la presque-nullité » (*parler de trente-six (mille) choses à la fois* vs *Il n'y a pas trente-six solutions*) (Rézeau 1993). Notre hypothèse est que l'idiomaticité constitue également une ressource pour exprimer le vague (Fuchs 1986, Kleiber 1987, Gardelle & Landragin 2024), phénomène encore peu étudié dans le cadre des études romanes. L'emploi idiomatique des noms de nombres comme mécanisme pour flouter la référence et pour exprimer l'approximation, tant vers des quantités inférieures que supérieures à la valeur exacte, s'inscrit dans le phénomène plus large de l'intensité, basée sur un continuum graduel qui se distribue entre les deux pôles de l'atténuation et du renforcement (Bazzanella 2011 : 28). Cette scalarité, liée au flottement de sens, est un facteur d'enrichissement lexical (polysémie) et de changement linguistique.

Nous montrerons que l'emploi idiomatique de la cardinalité concerne l'expression de la quantité/quantification, de la totalité, de la nullité, de l'(in)détermination, de l'ordre syntaxique. Notre approche, de type contrastif, s'appuiera sur un corpus français et italien constitué à partir à partir du logiciel *Sketch Engine*, ainsi que sur le traitement lexicographique qui est offert des noms des nombres dans les dictionnaires généraux (le *TLFi* et *Le Petit Robert 2025* pour le français ; le *Grande dizionario italiano dell'uso* et le *Nuovo Devoto-Oli. Il vocabolario dell'italiano contemporaneo 2025* pour l'italien). La communication sera l'occasion de mieux cerner les phénomènes de *vagueness* et de mettre à jour une sémantique du vague.

Nota bene : Cette communication s'inscrit dans les sous-thématiques « Phraséologie » et « Polysémie et ambigüité ». La communication sera présentée en français, les supports Powerpoint/hand-out seront produits en anglais, conformément aux indications de l'appel à contributions.

4[1] En français : *trois, quatre, cinq, mille, trente-six, mille*, *trente-six mille...* ; en italien : *due, cinque, cento, mille...*

Bibliographie :

Albano Leoni F., Formigari L. éd., 2015. *Le tout et ses parties. Langue, système, structure*, *Histoire Épistémologie Langage*, 37-1.

Anscombre J.-Cl., Mejri S. éd., 2011. *Le figement linguistique : la parole entravée*, Paris, Champion.

Bazzanella C., 2011. *Numeri per parlare*, Rome-Bari, Laterza.

Bouveret M., Legallois D. éd., 2012. *Constructions in French*, Amsterdam-Philadelphia, Benjamins.

Caillies S., 2009. « Descriptions de 300 expressions idiomatiques : familiarité, connaissance de leur signification, plausibilité littérale, 'décomposabilité' et 'prédictibilité' », *L'Année psychologique*, vol. 109 (3), 463-508.

Channel J., 1994. *Vague Language*, Oxford, Oxford University Press.

Dehaene S., 1997. *The Number Sens: How The Mind Creates Mathematics*, Oxford Univ. Press.

Fauconnier G., Turner M., 2003. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*, New York, Basic Books.

Fillmore C. J., Kay P., O'Connor M. C., 1988. « Regularity and idiomaticity in grammatical constructions: The case of let alone », *Language*, 64 (3), 501-538.

Frath P. éd., 2008. *Dénomination, phraséologie et référence*, *Zeitschrift für französische Sprache und Literatur – Beiheft*, n° 35, Franz Steiner Verlag, Stuttgart.

Fuchs C., 1986. « Le vague et l'ambigu : deux frères ennemis », *Quaderni di semantica*, 2, 235-245.

—, 1996. *Les ambiguïtés du français*, Paris, Ophrys.

Gizbert-Studnicki T., 2000. « Types of Vagueness », in W. Krawietz *et al.* (dir.) *The Reasonable as Rational, On Legal Argumentation and Justification, Festschrift for Aulis Aarnio*, Duncker und Humbolt, Berlin, 135-145.

Gledhill Ch., Frath P., 2007. « Collocation, phrasème, dénomination : vers une théorie de la créativité phraséologique », *La Linguistique*, vol. 43, fasc. 1, 63-88.

Goldberg A. E., 1995. *Constructions: A construction grammar approach to argument structure*, Chicago: University Chigaco Press.

—, 2003. « Constructions: a new theoretical approach to language », *Trends in Cognitive Sciences*, 7-5, 19-224.

Gross G., 1996. *Les expressions figées en français : noms composés et autres locutions*, Paris, Ophrys.

Grossmann F., Tutin A. éd., 2003. *Les collocations : analyse et traitement*, Amsterdam, De Werelt.

Guillon B., 1990. « Ambiguity, generality, and indeterminacy: Tests and definitions », *Synthese*, vol. 85, 391-416.

Kleiber G., 1987. « Quelques réflexions sur le vague dans les langues naturelles », in *Études de linguistique générale et de linguistique latine offertes en hommage à Guy Serbat*, Paris, Bibliothèque de L'Information Grammaticale, 157-172.

—, 1999. *Problèmes de sémantique. La polysémie en questions*, Villeneuve d'Ascq, Presses universitaires du Septentrion.

Hausmann F. J., 1997. « Tout est idiomatique dans les langues », in Martins-Baltar M. éd., *La locution entre langue et usages*, Lyon, ENS Éditions, 277-290.

Hausmann F.J., Blumenthal P. éd., 2006. *Collocations, corpus, dictionnaires*, *Langue française*, n° 150.

Lamiroy B., Klein J. R., 2005. « Le problème central du figement est le semi-figement », in A. Balibar-Mrabti & C. Vaguer éd., *Le semi-figement*, *LINX*, n° 53, 135-154.

Gardelle L., Landragin F. éd., 2024. *Flou, vague et sous-détermination dans la référence, Lingvisticæ* Investigationes, vol. 47 (2).

Legallois D., Patard A. éd., 2017. *Les constructions comme unités de la langue : illustrations, évaluation, critique*, *Langue Française*, n° 194 (2).

Legallois D., Tutin A. éd., 2013. *Vers une extension du domaine de la phraséologie*, *Langages*, n° 189.

Libben G., Jarema G. éd., 2006. *The Representation and Processing of Compound Words*, Oxford, Oxford University Press.

Martin R., 1987. « Flou. Approximation. Non-dit », *Cahiers de lexicologie*, 50, 165-176.

Mejri S., 2007. « French Phraseolgy », in Burget H. *et al.* éds, *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/ An international Handbook of Contemporary Research*, Berlin/New York, De Gruyter, 682-691.

Mel'čuk I., Clas A., Polguère A., 1995. *Introduction à la lexicologie explicative et combinatoire*, Duculot, Paris/Louvain-la-Neuve.

Mel'cuk I., 2003. « Collocations, définition, rôle et utilité », in Grossmann F., Tutin A. éd., *Les collocations : analyse et traitement*, Amsterdam, De Werelt, 23-32.

Mel'cuk I., 2011. « Phrasèmes dans le dictionnaire », in Anscombre J.-Cl., Mejri S. éd., *Le figement linguistique : la parole entravée*, Paris, Champion, 41-61.

Planchon P., 2012. « L'emploi approximatif des noms de nombres de 1 à 5 », in Schnedecker C. & Armbrecht C. éd., *La Quantification et ses domaines*, Paris, Champion, 131-138.

Polguère A., 2003. *Lexicologie et sémantique lexicale*, Québec, PUM.

Pollmann M.M.W., Jansen C.J.M., 2001. « On round numbers: pragmatic aspects of numerical expressions", *Quantitative Linguistics*, 187-202.

Rézeau P., 1993. *Petit dictionnaire des chiffres en toutes lettres*, Paris, Seuil.

Schmale G., 2013. « Qu'est-ce qui est préfabriqué dans la langue ? Réflexions au sujet d'une définition élargie de la préformation langagière », in Legallois D., Tutin A. éd., *Vers une extension du domaine de la phraséologie*, *Langages*, n° 189, 27-45.

Simone R., 2017. « Word as a stratification of formats », in D'Alessandro *R. et al*. éds, *Di tutti i colori. Studi linguistici per Maria Grossmann*, Utrecht University Repository, [En ligne], http://docs.wixstatic.com/ugd/7ae1d7_2677b342f1324b1296c74bb03bc74488.pdf, 1-16.

Simone R., Masini F. éd., 2014. *Word Classes. Nature, typology and representations*, Amsterdam-Philadelphia, Benjamins.

Simone R., Piunno V., 2017. « Combinazioni di parole che costituiscono entrata. Fenomeni, rappresentazione lessicografica e aspetti lessicologici », *Studi e Saggi Linguistici*, n°20, 1-27.

Sinclair J., 1991. *Corpus, Concordance, and Collocation*, Oxford: Oxford University Press.

Tutin A., 2008. « For an extended definition of lexical combinations », in Bernal E., De Cesaris J. éd., *Proceedings of the XIIIth EURALEX International Congress*, Barcelona, IULA, 1453-1460.

Wray A., 2008. *Formulaic Language: Pushing the Boundaries*, Cambridge, Cambridge University Press.

Dina Caetano Alves
Polytechnic Institute of Setúbal; Center of Linguistics, University of Lisbon

Towards a Standardized Diagnostic Lexicon in Speech and Language Therapy: Methodological Framework and Preliminary Outcomes

The absence of a standardized diagnostic lexicon in speech and language therapy (SLT) represents a critical challenge for both clinical practice and research. While international classification frameworks such as the DSM-5, ICD-11, and ICF provide structured taxonomies in related disciplines, no equivalent resource currently exists to harmonize terminology within SLT. As a result, diagnostic processes remain fragmented across clinical domains, with inconsistencies in the use of concepts and terms, increasing the risk of diagnostic errors, limiting comparability of studies, and creating barriers in professional training.

This project addresses this gap by developing a diagnostic manual for SLT. The initiative aims to establish both macrostructural and microstructural frameworks for diagnostic entries, offering clinicians and researchers a coherent reference system. The ultimate goal is to strengthen communication across professional contexts, reduce ambiguity, and contribute to the international alignment of SLT practices.

Objectives

The project pursues four main objectives:

1. To collect, analyze, and systematize terminological data relevant to diagnostic practice in SLT.
2. To design a macrostructural architecture organizing clinical domains, disorder groups, and diagnostic categories in a transparent, hierarchical way.
3. To define a microstructural framework specifying diagnostic entries, including identification, diagnostic criteria, and descriptive information.
4. To validate the proposed structures through expert consultation, ensuring both clinical and linguistic adequacy.

Methodology

The methodological approach combines systematic review and qualitative validation.

- **Systematic Literature Review**: A review protocol guided the collection of over 1,500 sources (scientific articles, books, and grey literature) across the nine traditional domains of SLT, including audition, child and adult language, speech, fluency, voice, swallowing, orofacial motricity, and communication. From this corpus, approximately 2,450 excerpts were extracted and coded into 16 analytical categories, covering the descriptive, diagnostic, and procedural dimensions of clinical practice.
- **Overlap Management**: The review revealed recurrent overlaps among diagnostic concepts across domains. A procedure for managing these overlaps was developed, clarifying whether entries should be merged, differentiated, or cross-referenced, thus ensuring terminological consistency.
- **Macrostructural Framework**: A hierarchical architecture was drafted. Each clinical domain includes a textual description, followed by groups of disorders, each with embedded diagnostic categories represented through hierarchical diagrams. This model allows both vertical navigation (from domains to specific diagnostics) and horizontal connections (across domains with shared features).
- **Microstructural Framework**: The adopted methodology has also permitted the proposal of a detailed template for diagnostic entries. Each entry includes: I. Identification Data and II. Diagnostic Criteria. Section I presents the clinical condition, including its most equivalent terms or expressions from DSM-5 and ICD-11. Section II includes: 1. Intrinsic, Functional, Statistical and Other General Characteristics (Diagnostic Level, Non-Statistical Characteristics, Primary Signs and Symptoms, Severity, Statistical Characteristics, Associated

Characteristics, and Secondary Signs and Symptoms), 2. Comorbidities and Differential Diagnosis, 3. Aetiology and Risk Factors, 4. Prognosis and Evolution, 5. Prevalence and Incidence, and 6. Functional Considerations and Impact. This structure ensures systematic representation across domains and supports interoperability with international classification systems.

- **Focus Groups**: To complement the documental phase, a Focus Group were conducted. Experts critically reviewed preliminary diagnostic entries, evaluated terminological adequacy, and provided feedback on the clinical relevance of the proposed criteria. These sessions further explored the usability of hierarchical diagrams and the potential for integrating the manual within academic curricula and professional training contexts.

Preliminary Results

The combined approach has produced a robust draft architecture for the diagnostic manual. At the macrostructural level, six big domains have been defined (plus one general domain), each containing coherent groupings of disorders. Hierarchical diagrams have been successfully tested to represent relations among entries. At the microstructural level, diagnostic templates have been drafted and refined through iterative consultation. Terminological overlaps were reduced, with explicit cross-referencing strategies ensuring internal consistency. Focus group feedback highlighted the practical value of the manual, especially in supporting interprofessional communication and reducing ambiguities in diagnostic language.

Next Steps

Future work will focus on extending expert consultation to an external broader panel of clinicians and researchers. Consensus procedures will be applied to finalize the terminology and criteria. A validation phase will follow, testing the manual in clinical settings.

The project foresees an open publication of the manual, designed for integration into training and professional contexts, and is open to international collaborations.

Expected Impact

The development of a standardized diagnostic lexicon in SLT is expected to generate significant advances:

- **For clinical practice**: improved diagnostic accuracy, reduced errors, and clearer communication among professionals.
- **For research**: enhanced comparability of studies and more precise operationalization of diagnostic categories.
- **For training**: a reliable pedagogical tool for students and early-career professionals.
- **For international collaboration**: potential convergence with global frameworks, promoting interoperability and shared standards.

This initiative demonstrates how the systematic construction of diagnostic terminology can bridge gaps between clinical practice and research, thereby reinforcing the scientific foundations of SLT.

References

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Whitehouse, A. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721 retrieved in https://pubmed.ncbi.nlm.nih.gov/28369935

Brin-Henry, F., Costa, R., & Despres, S. (2019, July). TemPO: Towards a conceptualisation of pathology in speech and language therapy. In *Terminologie et Intelligence Artificielle – Atelier TALN-RECITAL et IC (PFIA 2019)*, Toulouse, France. https://hal.science/hal-02193859

Denman, D., Wilson, N. J., Munro, N., Kim, J.-H., Speyer, R., & Cordier, R. (2020). Factors Influencing Speech-Language Pathologists' Application of Terminology for Describing Pediatric Language Assessments. *Communication Disorders Quarterly*, *42*(4), 195-206. https://doi.org/10.1177/1525740120915176 (Original work published 2021)

Schindler, A. (2005). Terminology in speech pathology: Old problem, new solutions. *Advances in Speech Language Pathology*, *7*(2), 84–86. https://doi.org/10.1080/14417040500125327

Francesco Caprioli
Università di Bergamo | francesco.caprioli@unibg.it

Entre idéologie et multilinguisme: la traduction des mots-slogans au Parlement européen (allemand-italien-français)

Dans le discours politique, entendu comme toute expression d'acteurs et d'actrices politiques portant sur des questions politiques et visant à convaincre l'électorat (Schröter & Carius, 2009), le lexique joue un rôle fondamental pour l'argumentation : le choix des mots influe fortement sur la réception des discours (Le Bart, 1998; Niehr, 2021), pour illustration en raison des émotions qu'ils peuvent susciter avec d'autres aspects linguistiques (Schwarz-Friesel, 2013, p. 134-170, 224-232; cf. aussi Carobbio, 2020). En particulier, les politicien(ne)s recourent aux mots-slogans, c'est-à-dire des termes issus du lexique commun mais investis sémantiquement d'une charge idéologique (Burkhardt, 2003; Niehr, 2007) – ici, au sens donné par Van Dijk (1998, 2006), pour qui l'idéologie consiste en toutes représentations sociales axiomatiques définissant l'identité d'un groupe ainsi que ses croyances partagées relatives à ses conditions fondamentales, à ses modes d'existences et à ses modalités de reproduction. Les mots-slogans ne le sont pas intrinsèquement, mais doivent être interprétés en fonction du contexte d'énonciation, du moment historique où ils sont proférés et de l'appartenance politique de ceux qui les utilisent (Brambilla, 2007). Grâce à leur potentiel persuasif, lié à leur concision, les responsables politiques cherchent à exprimer des programmes entiers sans recourir à des argumentations complexes (Niehr, 2019). Ces mots peuvent ainsi être analysés à la lumière des principes du stancetaking (Du Bois, 2007; Dang-Anh, 2023), puisque le choix de tel ou tel mot dans le contexte politique dépend finalement du point de vue que le locuteur entend transmettre à son public électoral (Niehr, 2021).

À partir de ces observations, cette recherche se propose d'analyser l'utilisation des mots-slogans, et notamment leur traduction, dans un contexte politique spécifique : les discours pléniers au Parlement européen. Plus précisément, les combinaisons linguistiques de l'allemand vers l'italien et de l'allemand vers le français sont examinées. Le Parlement européen constitue un cadre particulier pour les études politolinguistiques, dans la mesure où, encore plus que dans les autres institutions de l'Union Européenne, il est soumis aux principes du multilinguisme inscrits dans le Règlement 1/1958, qui prévoit l'égalité des 24 langues officielles et impose que l'ensemble des activités des institutions de l'UE puisse toujours se dérouler dans chacune d'elles (Corbett et al., 2011; Cosmai, 2007). Au Parlement européen, où les discours pléniers se tiennent oralement, l'apport des interprètes, qui traduisent simultanément toutes les interventions dans toutes les langues officielles, est donc fondamental (Gazzola, 2006). Or, la traduction – produite en temps réel, à l'oral et impliquant un effort cognitif considérable de la part des interprètes (Riccardi, 1998, 2007), lesquels doivent en outre faire face à une vitesse d'élocution souvent peu adaptée à la traduction simultanée (Anghelli & Mori, 2022) – n'est pas toujours parfaitement fidèle (Cinato, 2016, 2019; Schröder, 2019). Au plan lexical, par exemple, les mots-slogans peuvent subir des modifications de sens altérant les intentions originales des locuteurs, sans toujours restituer ce qu'ils veulent véritablement communiquer (Beaton-Thome, 2013; Defrancq & Plevoets, 2022) ou évoquant des frames (au sens de Fillmore & Baker, 2001) différents.

Cette étude s'appuie sur les travaux d'Antonioli (2014) et de Caprioli (2025), selon lesquels les mots-slogans interagissent avec d'autres niveaux linguistiques, notamment syntaxique et prosodique : en allemand, les mots-slogans apparaissent fréquemment dans une position syntaxiquement marquée, à la périphérie droite de l'énoncé – occupant en particulier l'après dernière position (Nachfeld) sans un nouveau prédicat verbal (Vinckel, 2006; Vinckel-Roisin, 2011)– et constituent une unité intonative dotée d'un accent, de manière à être particulièrement mis en relief. Afin d'examiner la manière dont ces mots-slogans sont rendus en traduction dans les discours pléniers au Parlement européen, cette recherche examine un corpus de 100 occurrences représentatives d'occupation de l'après dernière

position par des mots-slogans dans des discours prononcés entre 2023 et 2024 en allemand par des parlementaires élu(e)s en Allemagne, ainsi que leurs traductions en italien et en français fournies par les interprètes. Les occurrences originales et traduites sont transcrites selon la méthode GAT 2 (Selting et al., 2009), permettant également des observations prosodiques.

L'analyse contrastive compare les occurrences en allemand avec leurs traductions selon différents niveaux : (i) syntaxique, en vérifiant le maintien ou non de la position marquée (cf. entre autres pour l'italien Benincà, 2001; et pour le français Pekarek Doehler et al., 2015) ; (ii) lexical, en rapprochant la traduction du mot-slogan selon la taxonomie proposée par Caprioli (2025; basant sur Anghelli & Mori, 2022) : traduction littérale, reformulation avec connotation ou acte perlocutoire équivalent, reformulation avec connotation divergente ou acte perlocutoire divergent, recours à expressions plus vagues ; (iii) prosodico-pragmatique, en confrontant les types d'unité intonatives contenant les mots-slogans sur la base de la Language Into Act Theory (Cresti, 2000; Moneglia & Raso, 2014), selon laquelle les indices prosodiques – avant tout la typologie des tons aux frontières des unités intonatives – permettent de repérer leur statut informatif et de segmenter la chaîne du discours en énoncés pragmatiquement plus ou moins (in)dépendants.

En tenant compte de ces différents niveaux d'analyse, les résultats mettent en évidence des stratégies cross-linguistiques de convergence et de divergence entre l'utilisation des mots- slogans par les parlementaires et leur traduction par les interprètes.

Références bibliographiques :

Anghelli, I., & Mori, L. (2022). Migration in EP plenary sessions: Discursive strategies for the Other construction and political Self representation in Italian to Spanish interpreter-mediated texts. In M. Kajzer-Wietrzny, A. Ferraresi, I. Ivaska, & S. Bernardini (Éds.), Mediated discourse at the European Parliament : Empirical investigations (p. 93–131). Language Science Press.

Antonioli, G. (2014). Nachfeldbesetzung und diskursive Strategien der Hervorhebung. Eine Untersuchung am Beispiel von nationalsozialistischen Politikerreden. RiCOGNIZIONI. Rivista di lingue, letterature e culture moderne, 1, 205–217.

Beaton-Thome, M. (2013). What's in a word? Your enemy combatant is my refugee. The role of simultaneous interpreters in negotiating the lexis of Guantánamo in the European Parliament. Journal of Language and Politics, 12(3), 378–399.

Benincà, P. (2001). L'ordine degli elementi della frase e le costruzioni marcate. In L. Renzi, G. Salvi, & A. Cardinaletti (Éds.), Grande Grammatica italiana di consultazione (p. 129–209). Il Mulino.

Brambilla, M. (2007). Il discorso politico nei paesi di lingua tedesca : Metodi e modelli di analisi linguistica. Aracne.

Burkhardt, A. (2003). Das Parlament und seine Sprache. Studien zu Theorie und Geschichte parlamentarischer Kommunikation. Niemeyer.

Caprioli, F. (2025). Das Nachfeld in EU-Plenardebatten über Krieg und Frieden. Eine kontrastive Untersuchung (Deutsch-Italienisch). Muttersprache, 135(2-3), 146–165.

Carobbio, G. (2020). Prosodische Realisierungen expressiver Prozeduren am Beispiel einer politischen Rede. In G. Carobbio, C. Desoutter, & A. Fragonara (Éds.), Macht, Ratio und Emotion : Diskurse im digitalen Zeitalter (p. 63–81). Peter Lang.

Cinato, L. (2016). Ricostruire il senso. Strategie traduttive nell'interpretazione simultanea dal tedesco all'italiano. In M. Selig, E. Morlicchio, & N. Dittmar (Éds.), Gesprächsanalyse zwischen Syntax und Pragmatik. Deutsche und italienische Konstruktionen (p. 245–271). Stauffenburg.

Cinato, L. (2019). Politische Persuasion im europäischen Parlament : Deutsch-Italienisch im Vergleich. Linguistik Online, 97(4), 71–88.

Corbett, R., Jacobs, F., & Shackleton, M. (2011). The European Parliament (8e éd.). John Harper Publishing.

Cosmai, D. (2007). Tradurre per l'Unione Europea. Prassi, problemi e prospettive del multilinguismo comunitario dopo l'ampliamento a est (2e éd.). Hoepli.

Cresti, E. (2000). Corpus di italiano parlato. Accademia della Crusca.

Dang-Anh, M. (2023). Politische Positionierung. Sprachliche und soziale Praktiken der Relationalitätskonstitution. In M. Dang-Anh (Éd.), Politisches Positionieren : Sprachliche und soziale Praktiken (p. 7–38). Winter.

Defrancq, B., & Plevoets, K. (2022). Ut interpres : Linguistic convergence between orators and interpreters in the European Parliament. In M. Kajzer-Wietrzny, A. Ferraresi, I. Ivaska, & S.Bernardini (Éds.), Mediated discourse at the European Parliament : Empirical investigations (p.1–27). Language Science Press.

Du Bois, J. W. (2007). The stance triangle. In R. Engebretson (Éd.), Stancetaking in Discourse : Subjectivity, evaluation, interaction (p. 139–182). John Benjamins.

Fillmore, C. J. & Baker, C. F. (2001). Frame Semantics for Text Understanding. Proceedings of WordNet and Other Lexical Resources Workshop. NAACL (p. 59–64).

Gazzola, M. (2006). La gestione del multilinguismo nell'Unione europea. In A. Carli (Éd.), Le sfide della politica linguistica di oggi. Fra la valorizzazione del multilinguismo migratorio locale e le istanze del plurilinguismo europeo (p. 15–116). FrancoAngeli.

Le Bart, C. (1998). Le discours politique. Presses Universitaires de France.

Moneglia, M., & Raso, T. (2014). Notes on Language into Act Theory (L-AcT). In T. Raso & H. Mello (Éds.), Spoken corpora and linguistic studies (p. 468–495). John Benjamins.

Niehr, T. (2007). Schlagwort. In G. Ueding (Éd.), Historisches Wörterbuch der Rhetorik (Vol. 8, p. 496–502). Niemeyer.

Niehr, T. (2019). Schlagwörter und Leerformeln in der politischen Rede. In A. Burkhardt (Éd.), Handbuch Politische Rhetorik (p. 671–688). De Gruyter.

Niehr, T. (2021). Politischer Sprachgebrauch. Lublin Studies in Modern Languages and Literature, 45(1), 75–85.

Pekarek Doehler, S., De Stefani, E., & Horlacher, A.-S. (2015). Time and Emergence in Grammar : Dislocation, Topicalization and Hanging Topic in French Talk-in-Interaction. John Benjamins.

Riccardi, A. (1998). Interpreting Strategies and Creativity. In A. Beylard-Ozeroff, J. Králová, & B. Moser-Mercer (Éds.), Translators' Strategies and Creativity (p. 171–179). John Benjamins.

Riccardi, A. (2007). L'interazione di fattori cognitivi, linguistici e pragmatici nell'interpretazione simultanea. Annali Istituto Universitario Orientale. Sezione Germanica, XVII(1-2), 659–674.

Schröder, J. (2019). Politische Rhetorik im Europäischen Parlament. In A. Burkhardt (Éd.), Handbuch Politische Rhetorik (p. 1153–1172). De Gruyter.

Schröter, M., & Carius, B. (2009). Vom politischen Gebrauch der Sprache : Wort, Text, Diskurs. Peter Lang.

Schwarz-Friesel, M. (2013). Sprache und Emotion (2e éd.).

Francke.Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K., Couper-Kuhlen, E., Deppermann, A., Gilles, P., Günthner, S., Hartung, M., Kern, F., Mertzlufft, C., Meyer, C., Morek, M., Oberzaucher, F., Peters, J., Quasthoff, U., Schütte, W., … Uhmann, S. (2009). Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). Gesprächsforschung, 10, 353–402.

Van Dijk, T. A. (1998). Ideology : A Multidisciplinary Approach. Sage.

Van Dijk, T. A. (2006). Ideology and discourse analysis. Journal of Political Ideologies, 11(2), 115–140.

Vinckel, H. (2006). Die diskursstrategische Bedeutung des Nachfelds im Deutschen : Eine Untersuchung anhand politischer Reden der Gegenwartssprache. Deutscher Universitätverlag I

GWV Fachverlage.

Vinckel-Roisin, H. (2011). De la linéarisation marquée de l'énoncé à la cohérence du discours : L'après dernière position (Nachfeld) en allemand contemporain. Syntaxe et sémantique, 12(1), 189–207.

Rosa Cetro
Université de Pise

Phraséologie et communication publicitaire : analyse d'un corpus de slogans de produits de la marque Monoprix

Le discours publicitaire contemporain affiche une double nature : d'un côté, il se configure comme « un espace de circulation de significations et de valeurs » (Soulages, 2006), de l'autre, comme « un prescripteur de biens de consommation » (*ibidem*). Cette ambivalence est intrinsèquement liée aux objectifs poursuivis par la publicité elle-même : l'attirance des consommateurs potentiels ainsi que la volonté de persuader ces derniers vers l'achat de produits, sans montrer ces objectifs de manière trop explicite. D'où l'emploi de stratégies de communication basées sur l'humour, qui visent à réduire les distances et à créer une connivence avec le consommateur.

Dans cette communication, nous nous proposons d'analyser les stratégies publicitaires adoptées dans les slogans de produits de la marque française Monoprix à partir d'un corpus d'environ 40 slogans recueillis sur le site de celle-ci au cours des cinq dernières années.

S'il est vrai que pour chaque slogan il est possible d'identifier une stratégie de communication dominante, il est tout aussi vrai que la plupart des slogans affichent une combinaison de plusieurs techniques dans le but de créer un effet ludique. Les jeux de mots basés sur la polysémie – en particulier, sur la polysémie d'acceptions (Martin, 1983) – sont très fréquents dans notre corpus, tout comme ceux qui sont basés sur la phraséologie. Cette dernière constitue un patrimoine partagé au sein d'une communauté linguistico-culturelle, et s'avère donc être une voie privilégiée pour établir une connivence avec le destinataire.

En cas d'association entre polysémie et phraséologie idiomatique, le slogan joue souvent sur une manipulation de la locution idiomatique concernée, ce qui amène à une interprétation littérale de celle-ci (voir, à titre d'exemple, « On ne vous demandera plus de changer de disque » pour une confection de disques démaquillants).

La manipulation des expressions figées, connue sous le nom de défigement volontaire (Mejri, 2005) touche non seulement les expressions figées idiomatiques, mais aussi d'autres types d'unités phraséologiques (Gonzalez-Rey, 2021), comme les collocations et les pragmatèmes. En raison de la nature complexe des unités phraséologiques, le défigement ne se limite pas à un seul niveau mais intervient sur les plans phonologique, morphologique et sémantique à la fois (comme dans le slogan d'un paquet de biscuits « Embarquettement immédiat pour le goûter », basé sur le défigement du pragmatème « embarquement immédiat pour… » par le biais de la création du néologisme *embarquettement*).

En nous situant dans une perspective phraséodidactique, nous nous interrogerons aussi sur l'intérêt de ce corpus dans le cadre de l'enseignement du français à partir d'expériences pédagogiques menées à l'Université de Pise avec des apprenant-es italophones de niveau intermédiaire.

Références bibliographiques

Gonzales rey, I., *La nouvelle phraséologie du français*, Toulouse, Presses Universitaires du Midi, 3ème éd. revue et augmentée, 2021.

Gross, G., *Les expressions figées en français* : noms composés et autres locutions, Paris, Ophrys, 1996.

Grunig, B., *Les mots de la pubblicité : l'architecture du slogan*, Paris, Presses du CNRS, 1990.

Lecler, A., « Le défigement : un nouvel indicateur des marques du figement ? », *Cahiers de praxématique* [En ligne], 46 | 2006, document 3, mis en ligne le 01 janvier 2010, consulté le 28 août 2025.
URL : http://journals.openedition.org/praxematique/596 ; DOI : https://doi.org/10.4000/praxematique.596

Martin, R., *Pour une logique du sens*, Paris, PUF, 1983.

Mejri, S., « Figement absolu ou relatif : la notion de degré de figement », *Linx* [En ligne], 53 | 2005, mis en ligne le 15 février 2011, consulté le 11 septembre 2025. URL : http://journals.openedition.org/linx/283 ; DOI : https://doi.org/10.4000/linx.283

Soulages, J.-C. , « Les stratégies humoristiques dans le discours publicitaire », *Questions de communication* [En ligne], 10 | 2006, mis en ligne le 01 décembre 2006, consulté le 18 septembre 2025.
URL : http://journals.openedition.org/questionsdecommunication/7692 ; DOI : https://doi.org/10.4000/questionsdecommunication.7692

Gauvain Chalchli
Université Bordeaux Montaigne

TLFi-Dat : une modélisation chronologique du lexique français à partir des données du TLFi

1 - Contexte

De Darmesteter (1877) à Hasselrot (1972), les recherches en morphologie dérivationnelle synchronique utilisaient systématiquement un critère de néologicité pour la sélection des données pertinentes pour la description de la compétence et de la créativité morphologique des locuteurs. Au contraire, à partir des travaux fondateurs de la morphologie dérivationnelle post-générativiste (Aronoff 1976, Corbin 1987), ce critère de néologicité n'a généralement plus été pris en compte dans la sélection des données. Depuis le développement des études de la productivité morphologique sur corpus (Baayen 1989), le critère néologique est redevenu progressivement de plus en plus important jusqu'à être reconnu à nouveau récemment explicitement comme crucial (Bonami & Thuillier 2019, Missud 2023, Villoing & Gueye 2024, Huygevelde et al. 2025).

Plusieurs méthodes existent pour contrôler le caractère néologique d'un mot. L'une d'elles, utilisable à la fois en synchronie et en diachronie, est l'estimation de l'âge des mots par la datation de leur plus ancienne attestation disponible. Ce type d'information est directement disponible pour le français dans certains dictionnaires incluant dans leurs articles une section historique ou étymologique. Cependant, ces dictionnaires sont le plus souvent en format papier d'une part et d'autre part, même quand il existe une version électronique, difficilement accessible autrement qu'en consultant un par un manuellement la section pertinente de l'article correspondant à chaque mot étudié. Dans ce contexte, il pourrait être utile pour les chercheurs d'avoir à disposition une base de données électronique centralisée permettant d'accéder facilement et à grande échelle à cette information.

Dans cette communication, nous présentons une base de données intitulée TLFi-Dat, pouvant répondre en partie à ce besoin pour le français et un exemple d'usage en morphologie dérivationnelle.

2 - La base de données

La base de données TLFi-Dat est le résultat d'une extraction semi-automatique des premières dates d'attestation à partir des 54 525 entrées lexicales principales du TLFi. Parmi toutes les sources lexicographiques existantes, la source du TLFi a été choisie pour sa disponibilité en version électronique d'une part, d'autre part en tant que production scientifique collective majeure à des fins de recherche (Imbs 1961, Pierrel 2004, Surmont 2006) et enfin pour sa qualité supérieure aux autres sources existantes sur ce point (Radermarcher 2004).

Le protocole d'extraction a été adapté à la variation des articles en termes de structure, de localisation des informations de datation et de marquage de cette information dans la rédaction des articles. Suite à ce processus d'extraction, la base de données au format CSV contient 53 267 entrées. Parmi cette liste, 3496 entrées (6.6%) n'ont aucune information de datation disponible dans le TLF et le processus d'extraction a échoué pour quelques dizaines d'entrées. Parmi les 49 771 entrées ayant une datation disponible, la base de données permet par exemple d'observer que 9% n'ont pas d'attestation antérieure au XXe siècle, que 21% sont attestées pour la première fois au XIXe siècle et 1% sont déjà attestées au Xe siècle.

Exemple d'usage

La base de données TLFi-Dat peut être utilisée pour analyser le lexique du français standard en fonction de sa stratification chronologique. Cela peut permettre par exemple de comparer facilement l'ancienneté des lexèmes avec leur fréquence, de structurer une recherche diachronique ou de comparer la difficulté d'acquisition d'un mot et son apparition historique dans l'usage. Pour illustrer l'utilité de cette base de données, nous présentons son application au

contrôle du biais d'archaïsme dans la sélection des données dans l'étude des noms en *-ion* en répliquant l'étude classique de Bonami et al (2009).

Conclusion

La base de données TLFi-Dat n'élimine pas toute recherche manuelle dans la recherche des datations puisqu'elle ne comporte que les entrées principales du TLFi et contient certaines lacunes et erreurs issues soit de la compilation initiale soit du traitement proposé. Cependant, elle permet déjà de faciliter le contrôle à grande échelle de l'ancienneté d'une grande quantité de lexèmes, ce qui peut être utile dans de nombreuses recherches, et notamment de façon cruciale en morphologie dérivationnelle. De plus, cette première version a vocation à être améliorée et enrichie grâce au contrôle des utilisateurs et à l'intégration d'autres ressources. Nous aborderons les différentes étapes envisagées pour cet enrichissement de la ressource.

Références bibliographiques :

Aronoff, M. (1976). Word formation in generative grammar. Cambridge, MA: MIT Press.

Baayen, R. H. (1989). A corpus-based approach to morphological productivity: Statistical analysis and psycholinguistic interpretation. Dissertation, Free University, Amsterdam.

Bonami O., Boyé G. & Kerleroux F. (2009). L'allomorphie radicale et la relation flexion-construction. Dans FRADIN, B., KERLEROUX, F. et PLÉNAT, M. (éds) : Aperçus de morphologie, 103-126. Presses Universitaires de Vincennes, Saint-Denis.

Bonami, O. et Thuilier, J. (2019). A statistical approach to rivalry in lexeme formation : French -iser and -ifier. Word Structure, 12(1):4-41.

Corbin, D. (1987). Morphologie dérivationelle et structuration du lexique, vol. 193, 194, Linguistische Arbeiten. Tübingen: Max Niemeyer Verlag. [two volumes].

Darmesteter, A. (1877), De la création actuelle des mots nouveaux dans la langue française et des lois qui la régissent. Paris: Vieweg [reprint 1972, Genève: Slatkine Reprints]

Hasselrot, B. (1972). Étude sur la Vitalité de la Formation Diminutive Française au XXe Siècle. Almqvist & Wiksell, Uppsala.

Huygevelde M., Kayirici R., Bonami O. & Hemforth B., « Affix rivalry in French demonyms: an experimental approach », *Lexique* [Online], Numéro spécial | 2025.

Imbs, P. (1961). La place du vocabulaire ancien dans un thesaurus de la langue française. Lexicologie et lexicographie françaises et romanes. Orientations et exigences actuelles (Strasbourg, 12–16 novembre 1957), 133-147.

Missud, A. (2023). Modélisation de la compétition entre suffixations nominalisatrices : Productivité et contraintes morphologiquesl Thèse de doctorat, Université Paris Nanterre.

Pierrel, J.-M. (2004). La préface du TLFi. https://www.atilf.fr/wp-content/uploads/projets-ressources/atilf_Preface_TLFi_Pierrel-Jean-Marie.pdf

Radermacher, R. (2004). Le Trésor de la langue française: une analyse lexicographique (Doctoral dissertation, Université Marc Bloch (Strasbourg) (1971-2008)).

Surmont (de), J. N. (2006). De la genèse à l'informatisation du Trésor de la langue française et du Grand Robert électronique. Revista de filologia románica, 23, 55-66.

Villoing, F., & Gueye, A. (2024). La nominalisation néologique en -ment du français: structures morphologiques des bases verbales. In SHS Web of Conferences (Vol. 191, p. 08005). EDP Sciences.

Lian Chen
Université d'Orléans

Approche onto-phraséologique pour l'extraction automatique des relations idiomatiques

S'inscrivant dans la continuité de travaux précédents sur l'articulation entre la phraséologie et les *multiword expressions* (MWEs) (Polguère, 2002 ; Constant, 2012 ; Constant *et al*., 2017 ; Chen, 2021 ; Savary *et al*., 2023) ainsi que sur l'ontologie (Cimiano 2006 ; Choi & Jung, 2025 ; Chen, 2025 ; Dessí, Osborne, Buscaldi, Reforgiato Recupero & Motta, 2025), cette étude propose une méthode automatique de construction d'une ontologie idiomatique visant à modéliser les relations sémantiques sous-jacentes aux unités phraséologiques françaises. Traditionnellement, la conception d'ontologies demeure un processus manuel, long et réservé aux spécialistes de l'ingénierie des connaissances. Cependant, les récents progrès en traitement automatique du langage (TAL) et en intelligence artificielle ouvrent de nouvelles perspectives pour intégrer la complexité linguistique — notamment les expressions idiomatiques et métaphoriques — dans des cadres ontologiques structurés et lisibles par machine (Han et al. 2020).

Sur le plan théorique, notre travail s'appuie sur le modèle de l'*Ontology Layer Cake* (Cimiano, 2006 ; Després & Szulman, 2008) ainsi que sur la théorie phraséologique telle que formulée par Mel'čuk (2008, 2011), Polguère (2002, 2008), Mejri (2011), Sułkowska (2016) et Chen (2021), qui définissent les idiomes comme des unités phraséologiques (UP) caractérisées par leur opacité sémantique et leur conventionalisation. En combinant ces cadres, nous visons à formaliser les idiomes non pas comme de simples unités lexicales, mais comme des relations conceptuelles reliant des formes linguistiques à des significations abstraites (par ex. : *avoir la main heureuse* ⟶ *main* ⟶ *chance*).

Sur le plan méthodologique, notre approche repose sur une architecture en quatre phases inspirée des cadres d'apprentissage ontologique tels qu'OLAF (Schaeffer et al., 2023). La première phase concerne la construction et le prétraitement du corpus : un corpus de 15,8 millions de mots issus de Wikipédia (OPUS v1.0) a été nettoyé, tokenisé et segmenté en 200 sous-ensembles équilibrés. Une liste de 750 UP classés par champs sémantiques (corps humain, animaux) a été utilisée pour guider la détection automatique. Contrairement à des versions préexistantes de Wikipédia annotées en morphosyntaxe (par ex. le corpus REDAC, 2008), notre choix s'est porté sur une extraction plus récente et non annotée afin de garantir une meilleure couverture lexicale contemporaine et une cohérence avec les outils actuels de TAL (notamment spaCy).

La deuxième phase porte sur l'extraction automatique des candidats idiomatiques à l'aide de méthodes statistiques et hybrides (TF-IDF, PMI, C-value/NC-value, RAKE), fusionnées par la méthode Reciprocal Rank Fusion (RRF). Un filtrage linguistique (analyse morphosyntaxique et dépendancielle) et une validation interactive via une interface Streamlit ont permis de constituer un lexique idiomatique consolidé.

Dans la troisième phase, les relations sémantiques entre idiomes et composantes lexicales clés (ex. *main*, *cœur*, *tête*, *loup*) ont été automatiquement extraites à partir des dépendances syntaxiques (via spaCy) et associées à des catégories conceptuelles (ex. *main* ⟶ *aide*, *cœur* ⟶ *mémoire*). Ces relations relèvent d'un modèle hybride combinant : (i) des relations conceptuelles de type ontologique (ex. partie-de, propriété-de, agentivité), (ii) des relations lexicales issues de la tradition lexicographique (ex. synonymie partielle, champ sémantique), ainsi que (iii) des relations figuratives propres à la phraséologie (métaphore, métonymie). Ces relations ont été représentées sous forme de triplets <idiome, mot-clé, concept> et visualisées dans des graphes sémantiques pondérés à l'aide de PyVis.

Le choix du concept repose sur une base conceptuelle construite à partir de regroupements sémantiques inductifs (clustering) et alignée partiellement avec des catégories issues de ressources lexicales existantes (ex. WordNet, classifications ontologiques légères), puis validée manuellement.

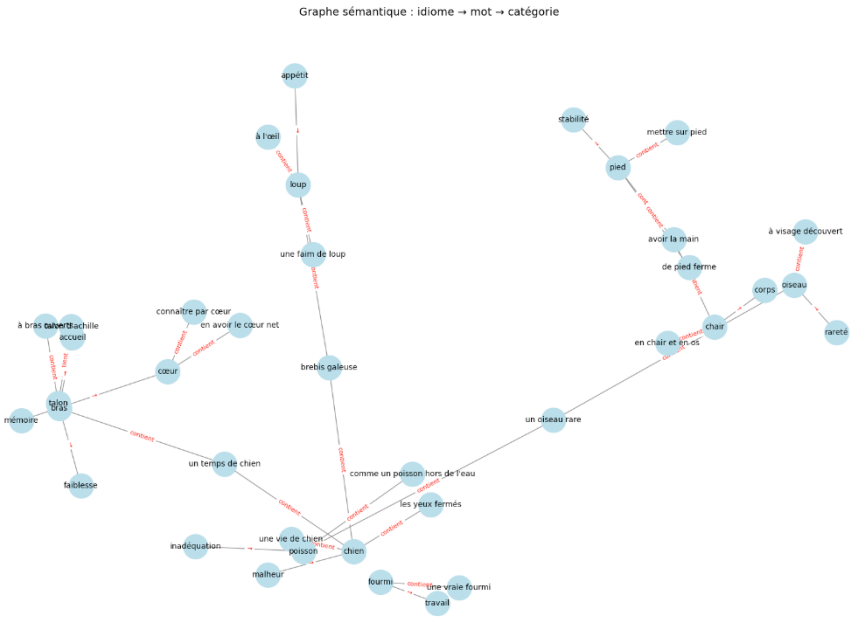


Figure 1 Graphes sémantiques et relationnels des expressions idiomatiques françaises

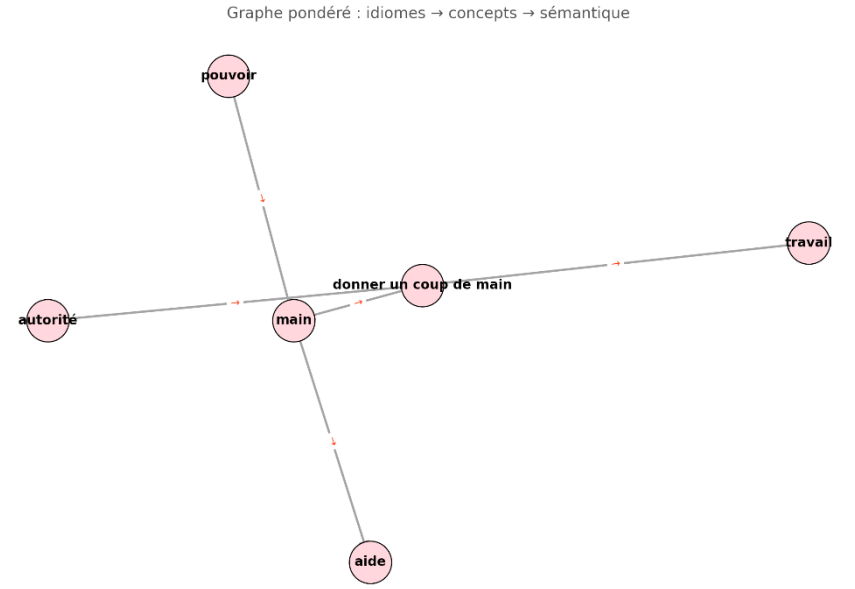


Figure 2 Graphe sémantique pondéré des expressions idiomatiques liées à la *main*

La quatrième phase repose sur l'apprentissage profond : un modèle BERT multilingue (*bert-base-multilingual-cased*) a été affiné pour classifier automatiquement les relations sémantiques entre idiomes et mots-clés. Malgré la taille modeste du jeu de données, les résultats sont encourageants (BERTScore F1 = 0,94), révélant une correspondance pertinente entre les prédictions du modèle et les interprétations humaines.

Notre travail se concentre principalement sur la méthode d'extraction automatique des relations entre unités phraséologiques, qui constitue le cœur du processus de modélisation.

Les résultats montrent que la combinaison du traitement automatique du langage (TAL), de la modélisation statistique et de l'apprentissage profond permet une extraction et une structuration automatiques du savoir phraséologique. Au-delà de ses applications lexicographiques, cette approche contribue à l'auto-lexicographie — c'est-à-dire à la construction automatisée de ressources lexicales et conceptuelles — et participe à l'articulation entre créativité linguistique et représentation des connaissances. L'objectif final de ce projet est de modéliser les unités

phraséologiques à partir des triplets extraits automatiquement, lesquels sont ensuite convertis en une ontologie OWL conforme au modèle OntoLex-Lemon, garantissant ainsi l'interopérabilité et la réutilisabilité des données au sein du Web sémantique. Les perspectives futures incluent l'extension du cadre à la modélisation ontologique bilingue (français–chinois) et à l'alignement interlinguistique des idiomes.

Bibliographie :

Chen, L. (2021). *Analyse comparative des expressions idiomatiques en chinois et en français (relatives au corps humain et aux animaux)* [Thèse de doctorat, Cergy Paris Université].

Chen, L. (2025). *Modeling and structuring of a bilingual French-Chinese phraseological dictionary: Neural automatic approach for ontology and lexicography*. In *Proceedings of eLex 2025: Electronic lexicography in the 21st century Intelligent Lexicography* (pp. 830–851). Lexical Computing CZ s.r.o.

Choi, S., & Jung, Y. (2025). Knowledge graph construction: Extraction, learning, and evaluation. *Applied Sciences, 15*(7), 3727. https://doi.org/10.3390/app15073727

Cimiano, P. (2006). *Ontology learning and population from text: Algorithms, evaluation and applications*. Springer. https://link.springer.com/book/10.1007/978-0-387-39252-3

Constant, M. (2012). *Mettre les expressions multi-mots au cœur de l'analyse automatique de textes : Sur l'exploitation de ressources symboliques externes* [Habilitation à diriger des recherches, Université Paris-Est].

Constant, M., Eryiğit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., & Todirascu, A. (2017). Multiword expression processing: A survey. *Computational Linguistics, 43*(4), 837–892. https://doi.org/10.1162/COLI_a_00302

Dessí, D., Osborne, F., Buscaldi, D., Reforgiato Recupero, D., & Motta, E. (2025). CS-KG 2.0: A large-scale knowledge graph of computer science. *Scientific Data, 12*, 964. https://doi.org/10.1038/s41597-025-04995-6

Després, S., & Szulman, S. (2008). Réseau terminologique versus ontologie. In *Proceedings of TOTh 2008* (pp. 1–19).

Han, X., Gao, T., Lin, Y., Peng, H., Yang, Y., Xiao, C., Liu, Z., Li, P., Zhou, J., & Sun, M. (2020). More data, more relations, more context and more openness: A review and outlook for relation extraction. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing* (pp. 745–758). Association for Computational Linguistics.

Mejri, S. (2011). Phraséologie et traduction. *Équivalence, 38*(1–2), 111–133.

Mel'čuk, I. (2008). La phraséologie et son rôle dans l'enseignement/apprentissage d'une langue étrangère. *Études de Linguistique Appliquée, 92*, 82–117.

Mel'čuk, I. (2011). Phrasèmes dans le dictionnaire. In J.-C. Anscombre & S. Mejri (Éds.), *Le figement linguistique : La parole entravée* (pp. 41–61). Honoré Champion.

Polguère, A. (2002). *Notions de base en lexicologie*. Ophrys.

Polguère, A. (2008). *Lexicologie et sémantique lexicale*. Les Presses de l'Université de Montréal.

Savary, A., Stymne, S., Barbu Mititelu, V., Schneider, N., Ramisch, C., et al. (2023). Parseme meets universal dependencies: Getting on the same page in representing multiword expressions. *Northern European Journal of Language Technology, 9*(1). https://doi.org/10.3384/nejlt.2000-1533.2391

Schaeffer, M., Sesboüé, M., Kotowicz, J.-P., Delestre, N., & Zanni-Merk, C. (2023). OLAF: An ontology learning applied framework. In *Proceedings of the 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES 2023)* (pp. 2106–2115). https://doi.org/10.1016/j.procs.2023.10.201

Sułkowska, M. (2016). Phraséodidactique et phraséotraduction : Quelques remarques sur les nouvelles disciplines de la phraséologie appliquée. *Yearbook of Phraseology, 7*, 35–54.

Beatriz Curti Contessoto
Université Lyon 2

L'interdisciplinarité entre droit et santé public : quelques réflexions pour la création d'un système conceptuel accessible au public profane

La traduction et l'adaptation de la terminologie spécialisée à différents contextes linguistiques et culturels représentent un défi majeur, en particulier pour les termes juridiques liés à la santé publique. En élargissant la notion de traduction à la traduction intralinguistique (cf. Finatto et Tcacenco, 2021), il s'agit aussi d'adapter un langage spécialisé vers une forme simplifiée. Dans cette perspective, une recherche interdisciplinaire menée dans une faculté de médecine au Brésil, coordonnée par une terminologue en coopération avec un centre de recherche français, vise à rendre plus accessibles les informations législatives en santé publique. Ce projet mobilise une équipe multidisciplinaire et se concentre sur un enjeu central : l'usage de ressources terminologiques (DeCS et MeSH, par exemple) pour construire un système conceptuel intelligible pour le grand public.

Au Brésil, des thèmes tels que l'avortement, la vaccination ou le contrôle d'épidémies tropicales (dengue, zika, chikungunya, fièvre jaune) suscitent de fréquents débats publics. Ces questions relèvent à la fois du droit – garant des droits et devoirs des citoyens – et de la santé publique – qui définit les politiques de prévention et d'accès aux soins (art. 25 de la Déclaration universelle des droits de l'homme ; art. 196 de la Constitution fédérale brésilienne de 1988). Cette imbrication se reflète dans la langue : les discours juridiques et sanitaires mobilisent des terminologies spécialisées, souvent opaques pour le public profane. L'accessibilité terminologique apparaît donc essentielle pour renforcer la communication entre experts et citoyens. Or, les ressources disponibles (OMS, OPAS, BIREME, ONU, Ministère de la Santé brésilien) traitent ces domaines séparément, sans toujours mettre en lumière leur interdisciplinarité. L'exemple du terme *avortement* est significatif : présent dans BIREME et UNTERM, seule cette dernière ressource souligne sa dimension interdisciplinaire (santé, droits humains, genre).

Pour répondre à ces enjeux, la recherche propose la création d'une plateforme informative, gratuite et accessible (cf. Estopà, 2021 ; Finatto et Paraguassu, 2022) sur des thèmes tels que l'avortement, l'éthique médicale, la vaccination et le contrôle des arboviroses. L'étude s'appuie sur deux approches complémentaires : la Théorie communicative de la terminologie (TCT – Cabré, 1998), qui envisage les termes sous leurs dimensions linguistique, cognitive et communicative ; et la Terminologie diachronique (TD – Dury, 1997 ; Picton, 2009 ; Zanola, 2021 ; Curti-Contessoto, 2024), qui étudie l'évolution des usages terminologiques (néologie, nécrologie, résurgence). Ces approches, complétées par la linguistique de corpus, permettent de saisir la dynamique lexicale des domaines spécialisés comme reflet de l'évolution des savoirs et des pratiques. Le projet a ainsi constitué le *LBMedCorpus*, corpus de textes législatifs brésiliens (1832-2023) incluant lois, décrets, amendements et « portarias » portant sur la vaccination, l'avortement, l'éthique médicale et les épidémies d'arboviroses. Ce corpus est organisé en sous-corpus dynamiques et statiques (cf. Dury et Picton, 2009), selon les genres textuels et les périodes historiques. Les termes ont été identifiés via Sketch Engine et validés en deux étapes : d'abord par comparaison avec des ouvrages terminographiques de référence, puis par l'analyse de leur emploi dans des contextes riches en connaissances (Knowledge Rich Contexts – KRCs, cf. Meyer, 2001 ; Barrière, 2004 ; Condamines, 2022). Ces contextes explicitent les caractéristiques ou relations conceptuelles et contribuent à la compréhension fine des termes.

L'analyse conceptuelle menée selon les normes ISO (25964-1, 2011 ; 1087, 2019) et la comparaison avec les ressources existantes ont suscité trois réflexions principales. La première concerne la neutralisation des particularités culturelles et juridiques. Les ressources internationales adoptent une approche standardisée qui gomme souvent la dimension contextuelle et socioculturelle. L'exemple de l'avortement est révélateur : défini techniquement comme l'interruption volontaire ou spontanée d'une grossesse, il reste incomplet car il n'intègre pas les dimensions sociales, culturelles et juridiques, très variables d'un pays à l'autre, y compris au sein d'une même communauté linguistique. Au Portugal,

l'avortement (*aborto*) a été légalisé en 2007 à la suite d'un référendum, et il est autorisé jusqu'à 10 semaines de grossesse dans le cadre du système national de santé. Au Brésil, en revanche, l'*aborto* demeure interdit, sauf dans des cas exceptionnels (viol, risque vital pour la mère ou anencéphalie), et reste au centre de fortes tensions religieuses, politiques et morales (cf. Curti-Contessoto et Serpa, 2025). Le terme *aborto* renvoie donc à des cadres normatifs très différents selon les pays lusophones (et au-delà), ce qui met en évidence la nécessité d'une approche fine et contextualisée pour construire un système conceptuel en terminologie sanitaire, particulièrement lorsqu'elle s'articule avec le droit.

La deuxième réflexion porte sur le manque d'accessibilité terminologique. En effet, les thésaurus et ontologies proposent le plus souvent des définitions très techniques, difficilement compréhensibles pour un public non spécialiste. Or, la santé publique concerne l'ensemble de la population, caractérisée par une grande hétérogénéité en termes de niveaux de scolarité, de conditions socio-économiques et de capacités de compréhension. Au Brésil, la difficulté est accentuée par le contexte sociolinguistique : selon l'INAF, plus de 20 % de la population est en situation d'analphabétisme fonctionnel. Une partie significative des citoyens, bien que sachant lire, peine à interpréter des textes complexes, notamment juridiques. C'est ici que la traduction intralinguistique et les principes du langage simple (ISO 24495-1, 2023) peuvent jouer un rôle clé, en reformulant un langage spécialisé vers une version simplifiée, sans perdre la rigueur conceptuelle. Toutefois, l'une des hypothèses de notre recherche consiste à questionner dans quelle mesure ces principes internationaux du langage simple s'avèrent réellement efficaces dans le contexte brésilien – une interrogation qui constituera une étape ultérieure de notre travail.

La troisième réflexion concerne la complexité de présentation des systèmes conceptuels. Les ressources existantes, bien que riches, demeurent difficilement accessibles aux non-spécialistes : organisation opaque, relations implicites, navigation peu intuitive. Le prochain défi consistera donc à concevoir un système terminologique simple, intuitif et adapté au contexte brésilien, en tenant compte notamment des usagers en situation d'analphabétisme fonctionnel.

Enfin, cette communication vise également à présenter d'autres exemples analysés au cours de la recherche, ainsi qu'un premier prototype d'organisation de ce système conceptuel. Ce prototype cherche à offrir une structuration plus simple et plus intuitive, adaptée à un public profane et, dans le contexte brésilien, également aux personnes en situation d'analphabétisme fonctionnel.

Références bibliographiques :

Barrière, C. (2004). Knowledge-rich contexts discovery. In Advances in Artificial Intelligence: 17th Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2004, London, Ontario, Canada, May 17-19, 2004. *Proceedings 17* (pp. 187–201). Springer.

Brésil. (1988). *Constitution de la République fédérative du Brésil de 1988*. Présidence de la République. http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

Cabré, M. T. (1998). *Terminology: Theory, methods and applications*. John Benjamins.

Condamines, A. (2022). How the notion of "knowledge rich context" can be characterized today. *Frontiers in Communication, 7*, 824711. https://doi.org/10.3389/fcomm.2022.824711

Curti-Contessoto, B. (2024). O(s) lugar(es) da diacronia na Terminologia: de onde partir para realizar um estudo terminológico-diacrônico hoje? *Acta Scientiarum (UEM), 45*, 1–12.

Curti-Contessoto, B., & Serpa, T. (2025). Abortion in comparative law: Examining intra-linguistic equivalence challenges in current legislative texts across Lusophone countries. *Jurisprudence – Revue critique, 10*, 135–159.

Dury, P. (1997). *Étude comparative et diachronique de l'évolution de dix dénominations fondamentales du domaine de l'écologie en anglais et en français* (Thèse de doctorat). Université Lumière Lyon 2 / Éditions du Septentrion.

Dury, P., & Picton, A. (2009). Terminologie et diachronie : vers une réconciliation théorique et méthodologique? *Revue française de linguistique appliquée, 15*(2), 31–41.

Estopà, R. (2021). Culturalización en salud: aportaciones desde la lingüística. *Cadernos de Linguística, 2*(1), 1–22.

Finatto, M. J. B., & Tcacenco, L. M. (2021). Tradução intralinguística, estratégias de equivalência e acessibilidade textual e terminológica. *Tradterm, 37*(1), 30–63. https://doi.org/10.11606/issn.2317-9511.v37p30-63

International Organization for Standardization. (2011). *ISO 25964-1. Information and documentation — Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. ISO.

International Organization for Standardization. (2019). *ISO 1087. Terminology work and terminology science — Vocabulary*. ISO.

International Organization for Standardization. (2023). *ISO 24495-1. Plain language — Part 1: Governing principles and guidelines*. ISO.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography. In D. Bourigault, C. Jacquemin, & M. C. L'Homme (Eds.), *Recent advances in computational terminology* (Vol. 2, pp. 279–302). John Benjamins.

Nations Unies. (1948). *Déclaration universelle des droits de l'homme*. Assemblée générale des Nations Unies, résolution 217 A (III). https://www.un.org/fr/about-us/universal-declaration-of-human-rights

Paraguassu, L. B., & Finatto, M. J. B. (2022). A linguistic approach to health literacy in Brazil: Terminological aspects. *Terminàlia, 25*, 14–27.

Picton, A. (2009). *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial* (Thèse de doctorat). Université Toulouse le Mirail – Toulouse II.

Zanola, M. T. (Ed.). (2021). *Cahiers de lexicologie – Terminologie diachronique : méthodologie et études de cas*. Classiques Garnier.

Christine Da Silva / Sylvie Multon
Université de Lorraine

Fluence verbale sémantique de jeunes adultes francophones : exploration de nouvelles données en vue d'un référentiel adulte

Le lexique joue un rôle clé non seulement dans la manière dont l'enfant s'approprie sa langue, mais aussi dans la capacité de l'adulte à comprendre et soutenir ses productions langagières. Cela explique l'importance donnée à l'évaluation lexicale quelle que soit la période de la vie. Pendant longtemps, les études s'intéressant à évaluer les compétences lexicales se sont limitées à déterminer la taille ou la richesse du vocabulaire (Zareva, 2005). Toutefois, les études récentes montrent la nécessité de considérer la diversité des dimensions lexicales (Best et al., 2006) comprenant notamment l'organisation du lexique. La tâche de fluence verbale (FV) est particulièrement intéressante, lorsque l'on souhaite explorer cette organisation.

Une tâche de FV implique la production de mots liée à un critère donné qu'il soit sémantique en présentant une catégorie (ex. *animaux*), ou phonologique (ex. /*p*/) avec une limite de temps (Hurks et al., 2010; Troyer, 2000). Cette tâche implique un nombre important de compétences linguistiques et cognitives dont la sélection de mots adaptée au critère donné ou encore l'inhibition d'éléments qui n'y correspondent pas (Brandeker & Thordardottir, 2023; Henry et al., 2015). Cette tâche repose donc à la fois sur des connaissances lexicales et sur les fonctions exécutives qui accompagnent l'accès aux mots. Cette tâche est principalement utilisée sur des populations adultes, voire âgées (Dubé & Elin Thordardottir, 2024), présentant des troubles neurologiques (Carpenter et al., 2021) et elle est considérée davantage pour les fonctions exécutives. Pourtant d'un point de vue lexical, elle est particulièrement pertinente, car elle permet d'apprécier particulièrement la composante langagière (Whiteside et al., 2016) et les éventuelles difficultés lexico-sémantiques (Mengisidou et al., 2020). En effet, les données de la littérature montrent une association claire entre la taille du vocabulaire et les performances obtenues à des tâches de FV, ce qui en fait une tâche pertinente, y compris dans les cas de bilinguisme (Dubé & Elin Thordardottir, 2024). La production de mots peu fréquents implique un répertoire lexical plus étendu, comme dans les tâches dites de langage spontané (Rondal, 2003). Alors que les mots fréquents et usuels sont employés par la majorité des individus, cela n'est pas le cas des mots peu fréquents qui sont limités au discours d'un nombre limité de participants (Dubé & Elin Thordardottir, 2024).

L'objectif de cette communication est double : d'une part, présenter une tâche de fluence sémantique mise en place dans le cadre d'un projet s'intéressant au profil lexico-sémantique d'enfants avec ou sans trouble développemental du langage, et d'autre part, exposer les premières données recueillies sur une population adulte en vue d'élaborer un référentiel des productions attendues.

Les données de la présente étude reposent sur les productions de vingt-cinq adultes francophones âgé.es entre 18 et 22 ans. Les sujets ont participé à une tâche de FV sémantique comprenant six items (*vêtement, animaux, soleil, maison, chanter et marcher*) qui ont été présentés séparément et dans un ordre pseudo-aléatoire avec une consigne informatisée, garantissant une standardisation de la passation. L'intégralité de la tâche a été enregistrée en vidéo. Puis, ces données ont été pseudonymisées et analysées à l'aide du logiciel ELAN (Wittenburg et al., 2006), ce qui a permis la segmentation, la transcription et l'annotation des données avec un système de codage identique pour toutes les données. Celles-ci ont été analysées en considérant le nombre des mots produits, le temps de latence pour produire le premier mot, le nombre de répétition de mots, le nombre et la taille des *clusters* sémantiques (production de séries de mots issus de la même sous-catégorie) et *switches* (changement de sous-catégories). Ces analyses ont été réalisées pour chaque item et en considérant la catégorie grammaticale des mots ciblés par la tâche (noms *vs*. verbes).

Les résultats montrent que le temps de latence moyen est inférieur à 2 secondes et varie selon la nature du mot ciblé. En effet, les items verbaux impliquent un temps de latence avant le premier mot légèrement plus élevé que les items

nominaux. Certains mots (ex. *chat*, *chien* pour l'item *animaux*) sont proposés par l'ensemble des personnes, alors que d'autres mots sont plus rares (ex. *marcel* pour les *vêtements*) et apparaissent plus tardivement dans les productions relevant ainsi un lexique plus étendu.

Les données recueillies auprès d'une population adulte mettent en évidence des variations interindividuelles importantes. Les items nominaux ont amené une production moyenne significativement plus élevée que les items verbaux. Parmi les items nominaux, *animaux* se distingue par la moyenne la plus élevée alors que *soleil* enregistre la plus faible. Cette variation est également visible par le nombre de *clusters* et de *switches* plus importants pour les items nominaux, avec une disparité au sein de ces items (ex. animaux > soleil) s'expliquant notamment par des différences taxinomique (terme générique vs. spécifique) et perceptuelle (multisensorielle vs. visuelle/thermique) entre les deux termes. Des corrélations positives entre le nombre total de mots produits et le nombre total de *clusters* mobilisés ou de *switches* montrent que plus une personne produit de mots, plus elle mobilise chacun de ces deux processus. Par ailleurs, la moitié des participant.es produisent des répétitions de mots déjà donnés. Une corrélation positive est observée entre le nombre total de mots produits et le nombre total de mots répétés, indiquant que plus une personne produit de mots, plus elle est susceptible de répéter certains d'entre eux.

L'analyse quantitative et qualitative des productions des jeunes adultes francophones dans une tâche de FV sémantique permet de relever des différences de traitement selon les mots ciblés. En effet, les productions sont plus orientées vers des évocations taxinomiques pour certains items tels que *animaux*, *vêtements* ou *maison* (Rofes et al., 2023), alors qu'elles sont davantage associatives (contextuelle, fonctionnelle, perceptuelle) pour les autres items. En outre, les participant.es mobilisent fréquemment des stratégies de regroupements sémantiques (*clustering*) et de changement de catégorie (*switching*), processus dont l'alternance permet l'efficacité en fluence verbale (Troyer et al., 1997). Enfin, la tâche de fluence sémantique présentée dans cette étude, avec l'introduction d'items verbaux, constitue un outil pertinent pour explorer le stock lexical et des stratégies de récupération de mots.

Références bibliographiques :

Best, R. M., Dockrell, J. E., & Braisby, N. R. (2006). Real-world word learning : Exploring children's developing semantic representations of a science term. *British Journal of Developmental Psychology*, *24*(2), 265-282. https://doi.org/10.1348/026151005X36128

Brandeker, M., & Thordardottir, E. (2023). Verbal fluency in bilingual school-age children : Looking at switching, clustering, and the effect of language experience. *International Journal of Bilingualism*, *27*(3), 374-393. https://doi.org/10.1177/13670069211055263

Carpenter, E., Peñaloza, C., Rao, L., & Kiran, S. (2021). Clustering and Switching in Verbal Fluency Across Varying Degrees of Cognitive Control Demands : Evidence From Healthy Bilinguals and Bilingual Patients With Aphasia. *Neurobiology of Language*, *2*(4), 532-557. https://doi.org/10.1162/nol_a_00053

Dubé, D. & Elin Thordardottir. (2024). Using semantic verbal fluency to estimate the relative and absolute vocabulary size of bilinguals : An exploratory study of children and adolescents. *Journal of Communication Disorders*, *111*, 106450. https://doi.org/10.1016/j.jcomdis.2024.106450

Henry, L. A., Messer, D. J., & Nash, G. (2015). Executive functioning and verbal fluency in children with language difficulties. *Learning and Instruction*, *39*, 137-147. https://doi.org/10.1016/j.learninstruc.2015.06.001

Hurks, P. P. M., Schrans, D., Meijs, C., Wassenberg, R., Feron, F. J. M., & Jolles, J. (2010). Developmental Changes in Semantic Verbal Fluency : Analyses of Word Productivity as a Function of Time, Clustering, and Switching. *Child Neuropsychology*, *16*(4), 366-387. https://doi.org/10.1080/09297041003671184

Macoir, J., Lafay, A., & Hudon, C. (2019). Reduced Lexical Access to Verbs in Individuals With Subjective Cognitive Decline. *American Journal of Alzheimer's Disease & Other Dementias®*, *34*(1), 5-15. https://doi.org/10.1177/1533317518790541

Mengisidou, M., Marshall, C. R., & Stavrakaki, S. (2020). Semantic fluency difficulties in developmental dyslexia and developmental language disorder (DLD) : Poor semantic structure of the lexicon or slower retrieval processes? *International Journal of Language & Communication Disorders*, *55*(2), 200-215. https://doi.org/10.1111/1460-6984.12512

Rofes, A., Beran, M., Jonkers, R., Geerlings, M. I., & Vonk, J. M. J. (2023). What Drives Task Performance in Animal Fluency in Individuals Without Dementia? The SMART-MR Study. *Journal of Speech, Language, and Hearing Research*, *66*(9), 3473-3485. https://doi.org/10.1044/2023_JSLHR-22-00445

Rondal, J. A. (2003). *L'évaluation du langage*. Mardaga.

Troyer, A. K. (2000). Normative Data for Clustering and Switching on Verbal Fluency Tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370-378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency : Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138-146. https://doi.org/10.1037/0894-4105.11.1.138

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal Fluency : Language or Executive Function Measure? *Applied Neuropsychology: Adult*, *23*(1), 29-34. https://doi.org/10.1080/23279095.2015.1004574

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Éds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. European Language Resources Association (ELRA). https://aclanthology.org/L06-1082/

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System*, *33*(4), 547-562. https://doi.org/10.1016/j.system.2005.03.005

Romain Delhem
Université Paris Nanterre

Assessing measures of productivity for derivational morphological patterns

Morphological patterns, including affixes, are notably distinguished by whether they can be used to form new words or not, a property called productivity. Among the measures used to assess productivity quantitatively are type frequency (e.g. Bybee 1985), type–token ratio (e.g. Norde & Van Goethem 2018), potential productivity (e.g. Baayen & Lieber 1991), and renewal rate (Berg 2020). All these measures entail that productivity is a scalar phenomenon (Dal & Namer 2016: 79); however, they use different variables from a set of four (types, tokens, hapaxes, and neologisms) and are therefore not expected to yield similar results. The problem is that depending on the measure used, a given affix will be considered unproductive, slightly productive, or very productive.

There are at least two ways of deciding what the best measure for productivity is:

- by comparing the measures with the definitions of productivity given by authors and seeing which measure best captures these definitions;
- by comparing them with previous assessments of the productivity of a number of affixes by certain authors.

Bauer & Huddleston (2002: 1677–1715) analyze over 130 English derivational affixes, 77 of which are explicitly assigned a degree of productivity. The authors, however, do not write what measure they used to assess the productivity of the affixes under scrutiny. The goal of this study is to see whether there are discrepancies between the productivity of English affixes as measured by the mathematical ratios given above and their perceived productivity in Bauer & Huddleston. To do so, each of the 77 affixes was given a score, based on the way its productivity (or lack thereof) was described by the authors:

- 0 for affixes described as unproductive (e.g. *–age*, *–ard*, *–ence*, *–th*);
- 1 for affixes described as "probably no longer productive", or "marginally productive" (e.g. *fore–*; *–let*, *–ment*, *–some*);
- 2 for the only affix exhibiting "some degree of productivity" (*–ette*);
- 3 for productive affixes (e.g. *intra–*, *meta–*; *–dom*, *–ese*, *–ify*);
- 4 for very or extremely productive affixes (e.g. *ex–*, *non–*, *ultra–*; *–er*, *–ism*).

These scores are not intended to quantify productivity precisely; rather, they act as categories. If a given productivity measure were in perfect agreement with Bauer & Huddleston's assessment, all the affixes with a score of 4 ("very productive") would be more productive than those with a score of 3 ("productive"), all the affixes with a score of 3 would be more productive than those with a score of 2 ("somewhat productive"), and so on.

In order to verify which productivity measure fits Bauer & Huddleston's analysis best, a search for all the 77 affixes was carried out on the Corpus of Historical American English (COHA, Davies 2010). The results were filtered for the decade 1990–1999, which is the period during which Huddleston & Pullum's (2002) grammar book was written. The variables (types, tokens, hapaxes, and neologisms) and the productivity measures were then calculated for each affix under consideration, as well as the Pearson correlation coefficient between each of these measures and the scores described above.

The data shows a very strong correlation ($\rho \approx 0.94$) between hapaxes and neologisms, which was predicted by Baayen & Lieber (1991: 813). It would therefore seem that, for a given period of time, the number of hapaxes can indeed be used with a fair amount of confidence to predict the number of neologisms if diachronic data is not available.

As for the productivity measures, the study shows that renewal rate is the ratio that has the strongest correlation ($\rho \approx 0.63$) with Bauer & Huddleston's description of the affixes. This measure (the number of neologisms coined during a given period of time divided by the number of types used on the same period) is the only one that explicitly takes into account whether the words formed through a given morphological pattern are new or not.

The renewal rate of a morphological pattern, which is its ability to create new types over time, seems to approximate Bauer & Huddleston's claims better than does potential productivity (hapaxes divided by tokens), which is a frequently used method corresponding to the probability of encountering a new type (something hapaxes approximate) in a corpus. As a result, not only do the results of the study indicate that the renewal rate is the best ratio to measure productivity, but the definition of this rate itself (i.e. the ability of a pattern to regenerate itself) seems closer to what productivity is generally understood to be.

References:

Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29. 801–843.

Bauer, Laurie & Huddleston, Rodney. 2002. Lexical word-formation. In Huddleston & Pullum (2002), 1621–1721.

Berg, Kristian. 2020. Changes in the productivity of word-formation patterns: Some methodological remarks. *Linguistics* 58 (4). 1117–1150.

Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Philadelphia: Benjamins.

Dal, Georgette & Namer, Fiammetta. 2016. Productivity. In Hippisley, Andrew & Stump, Gregory Thomas (eds.), *The Cambridge Handbook of Morphology*, 70–90. Cambridge: Cambridge University Press.

Davies, Mark. 2010. *The Corpus of Historical American English (COHA)*. <english-corpora.org/coha>

Huddleston, Rodney & Pullum, Geoffrey. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Norde, Muriel & Van Goethem, Kristel. 2018. Debonding and clipping of prefixoids in Germanic: Constructionalization or constructional change?. In Booij, Geert (ed.), *The construction of words: Advances in Construction Morphology*, 475–518. Dordrecht: Springer.

Sylvain Farge / Anaïs Malfatti
Université Montpellier 3

Définir la maladie, entre lexique et discours : une étude contrastive allemand-français sur les lexèmes désignant l'obésité

Certaines maladies chroniques présentent des enjeux définitionnels forts : leur définition, notamment à l'ère du numérique, n'appartient pas qu'aux seuls experts médicaux (Delavigne, Picton, Thibert, 2022 ; Delavigne, 2020), et les patients, souvent organisés en associations, font valoir leur vécu de la maladie, parfois présenté comme expertise propre (Pour les *fat studies*, voir Saguy & Riley, 2005). D'ailleurs, les expertises médicales et profanes diffèrent parfois grandement (Laroche & al., 2019). La présentation de la maladie, et donc les discours à leur sujet, varient largement, par exemple, entre terminologie médicale et approche profane (notamment, dans la presse : Humbert Droz, 2024, sur l'endométriose). Au demeurant, la définition même de la maladie et le regard que le monde médical porte sur elle, au-delà même de celui des profanes, peut faire l'objet de débats, de perceptions différentes (pour l'obésité, voir Carof, 2015). Terminologie et langue profane connaissent enfin des recoupements, par la circulation des discours, et tel terme médical, dans la langue profane, se charge de nouvelles représentations : par exemple, un terme médical comme *IMC* (Indice de Masse Corporelle), fait l'objet de définitions très différentes, souvent abusives (Basdevant, 2009), même s'il a été montré, pour le surpoids, que les représentations des médecins d'une aire culturelle données différaient moins de celles des profanes de la même aire qu'entre deux médecins d'aires culturelles différentes (Leeman, Fischler, Rozin, 2011). L'enjeu terminologique-lexical est tel que les institutions de santé publique publient des recommandations, à l'adresse, aussi bien, des profanes que des médecins, notamment concernant l'obésité (HAS, 2023 ; Association pour la santé publique du Québec – Collectif Vital, 2023)

En somme, la définition de la maladie est un enjeu et un objet aussi bien médical que social, elle ne naît pas dans un milieu scientifique objectif ni imperméable aux enjeux sociaux et à la circulation des discours, et le lexique en porte la trace. Nous nous proposons, dans cette contribution, d'étudier cette question à la lumière des termes désignant la forte corpulence, stade obésité, en allemand et français. Dans quelle mesure les ressources lexicales disponibles pour évoquer l'obésité dans les deux langues conditionnent-elles la perception de la forte corpulence ? Que nous apprennent les recommandations et les néologismes euphémiques sur les débats définitionnels ?

Pour répondre à cette question, nous procèderons une analyse de sémantique discursive (Lecolle, Veniard, Guérin, 2018) des lexèmes de l'allemand (fettsucht, fettleibig, adipös) et du français (obèse) exprimant l'obésité, et nous tâcherons de voir à quel point les sémantismes dans les deux langues sont porteurs de motifs différents[5[1]]. La sémantique discursive que nous proposons se fonde sur une étude préalable de sémantique lexicale des lexèmes envisagés qui s'appuie d'emblée sur des emplois de ces derniers (et non seulement des formes attestées dans des ouvrages de référence, comme les dictionnaires). Les formes sémantiques dégagées pour chaque lexème sont alors étudiées en contexte, dans l'analyse d'un corpus textuel.

Analyse lexicologique :

L'analyse sémantique préalable montrera, en français, que le terme « obésité » est polysémique et peut s'employer hors du domaine médical pour désigner un comportement coupable de veulerie et de gabegie, par exemple financière (Farge, 2024), ou un trop-plein, comme l'infobésité. Les expressions terminologiques comme *obésité morbide* sont à leur tour susceptibles de véhiculer des visions discriminantes, ce qui justifie des recommandations terminologiques à l'endroit des médecins, au Québec par exemple (Association pour la santé publique du Québec – Collectif Vital, 2023).

5[1] Pour la Théorie des Formes Sémantiques (Cadiot & Visetti, 2001), le motif est la structure fondamentale du lexème, appelée à se développer en contexte selon son profilage (classe de mot et acception) et la thématisation (l'insertion dans un contexte discursif donné).

Nous nous concentrerons sur l'interprétation de la métaphore constitutive du lexème obèse, suivant Legallois et Kwon (2005).

En allemand, inversement, plusieurs termes monosémiques co-existent, aux emplois plus ou moins actuels : *die Fettsucht*, qui ne s'emploie plus que rarement comme terme, porte une notion d'addiction en allemand contemporain ; *die Fettleibigkeit* s'emploie couramment, mais sa morphologie et ses emplois permettent de détecter un potentiel de dénomination discriminante fort, ce qui peut expliquer qu'il soit peu à peu évincé par *die Adipositas*, terme médical plus récent, à potentiel euphémique (Sablayrolles, 2016), mais qui peut se charger, dans la presse, malgré l'intention euphémique, de connotations péjoratives (López Diaz, Seoane, 2020 ; Seoane, 2016), nourrissant ainsi ce que Jamet (2017) nomme le *lexical treadmill*. Ceci explique probablement la reprise, par des médecins germanophones, d'un néologisme récent pour exprimer le surpoids, *das Mehrgewicht* (notamment dans une émission de la chaîne ARD), dont nous présenterons également les propriétés lexicales.

Analyse des lexèmes en corpus (analyse discursive)

Nous inscrirons notre analyse discursive dans le cadre de la linguistique symétrique de M.-A. Paveau (2013, 2017), à qui nous emprunterons également la notion d'emploi ajusté (2019) : la linguistique symétrique pose que le sens naît de l'interaction du locuteur avec le monde et ne peut se concevoir comme un système cognitif encapsulé. Par exemple, la signification de *obèse* est indissociable des techniques comme la balance, la mesure du poids, les régimes, etc. L'emploi ajusté, enfin, est la capacité d'un locuteur à tenir compte du contexte socio-discursif pour employer les mots, les expressions qui ne blessent pas autrui.

Partant de ces prémisses, nous analyserons un corpus (à constituer, sur les années 2020-2025) de textes de presse ou littéraires, germano- et francophones (environ 5 par langue) présentant des commentaires métalinguistiques sur les termes de corpulence : comment des locuteurs, médecins ou profanes, se situent-ils par rapport aux termes à leur disposition ? Dans quelle mesure la conscience morphologique permet-elle de prendre ses distances avec les représentations que peuvent véhiculer les lexèmes de la langue que l'on parle. En d'autres mots, en quoi le lexique participe-t-il à la constitution de la représentation de la maladie ? Comment se négocie le rapport à la corpulence en fonction des mots à disposition ? La qualité terminologique d'un lexème est-elle un gage suffisant de discours ajusté ?

Dans cette contribution, ainsi, nous entendons proposer un cadre de sémantique discursive qui permette d'interroger aussi bien le poids du sémantisme des lexèmes dans la construction des représentations (polysémie et représentations, différences culturelles dans le lexique) que le caractère limité de ce poids : c'est l'emploi en discours qui, en dernière instance, détermine la charge axiologique des mots employés. S'il est en effet important de réfléchir sur les mots et les discours qu'ils véhiculent nécessairement en langue, avant même leur emploi, il convient de montrer qu'un mot n'a jamais de sens qu'en usage. Loin de constituer une taxonomie, les mots proposent des indices pour saisir le monde en discours. (1091 mots)

Bibliographie :

Association pour la santé publique du Québec – Collectif Vital. (2023). *Trousse pour des communications saines liées au poids*. https://collectifvital.ca/files/P2-Enjeux_poids/Trousse-fran%C3%A7ais/Trousse-communication-saine-fr.pdf

Basdevant, A. (2009). Buts et abus de la définition médicale contemporaine. In J. Csergo (Éd.), *Trop gros ? L'obésité et ses représentations* (Ed. Autrement).

Cadiot, P., & Visetti, Y.-M. (2001). *Pour une théorie des formes sémantiques*. Presses Universitaires de France. https://doi.org/10.3917/puf.cadi.2001.01

Carof, S. (2015a). *Le « surpoids », un stigmate acceptable ? Représentations, discriminations et réappropriations des rondeurs féminines en France, en Allemagne et en Angleterre*. EHESS.

Delavigne, V. (2021). Que fait le numérique aux articles encyclopédiques ? : L'exemple de la vulgarisation médicale. *La linguistique*, *Vol. 57*(1), 111-131. https://doi.org/10.3917/ling.571.0111

Delavigne, V., Picton, A., & Thibert, E. (2022). Socioterminologie et terminologie textuelle : L'expertise en questions. *SHS Web of Conferences*, *138*, 04012. https://doi.org/10.1051/shsconf/202213804012

Farge, S. (2024). Décrire l'institution malade : Métaphore et culture. Une étude comparative français-allemand. In *Nommer le corps, nommer la maladie : Modèles de représentation* (Lambert-Lucas).

Humbert-Droz, J. (2024). Terminologie de l'endométriose et représentations de la maladie : Regards croisés entre presse généraliste et discours spécialisés. *SHS Web of Conferences*, *191*, 05003. https://doi.org/10.1051/shsconf/202419105003

Jamet, D. (2017). The Neological Functions of Disease Euphemisms in English and French : Verbal Hygiene or Speech Pathology? *Lexis*, *12*. https://doi.org/10.4000/lexis.2397

Kleiber, G. (1990). *La Sémantique du prototype : Catégories et sens lexical*. Presses universitaires de France.

Laroche, F., Guérin, J., Azoulay, D., Coste, J., & Perrot, S. (2019). La fibromyalgie en France : Vécu quotidien, fardeau professionnel et prise en charge. Enquête nationale auprès de 4516 patients. *Revue du rhumatisme*, *86*, 90-95.

Lecolle, M., Veniard, M., & Guérin, O. (Éds.). (2018). *Vers une sémantique discursive : Propositions théoriques et méthodologiques*. Armand Colin.

Leeman, R. F., Fischler, C., & Rozin, P. (2011). Medical doctors' attitudes and beliefs about diet and health are more like those of their lay countrymen (France, Germany, Italy, UK and USA) than those of doctors in other countries. *Appetite*, *56*, 558-563.

Legallois, D., & Kwon, S.-N. (2005). Sémantique lexicale et examen écologique de la co-occurrence. *Cahiers du CRISCO*, 18 pages.

López Díaz, M., & Seoane, A. (2020). Les euphémismes dans les médias : Entre voilements, démasquages et discours qui les traversent. Présentation. *Çédille*, *17*, 13-18. https://doi.org/10.25145/j.cedille.2020.17.02

Paveau, M.-A. (2013). *Langage et morale : Une éthique des vertus discursives*. Éditions Lambert-Lucas.

Paveau, M.-A. (2017). *Les prédiscours : Sens, mémoire, cognition*. Presses Sorbonne Nouvelle.

Paveau, M.-A. (2019). *La blessure et la salamandre. Théorie de la resignification discursive*. https://hal.science/hal-02003667

Sablayrolles, J.-F. (2016). Des procédés néologiques euphémiques et quelques domaines privilégiés d'apparition. *Les néologismes euphémiques*, 187-200.

Saguy, A., & Riley, K. (2005). Weighing Both Sides : Morality, Mortality, and Framing Contests over Obesity. *Journal of Health Politics, Policy and Law, 30*(5), 869-921.

Seoane, A. (2016). Deux néologismes par glissement sémantique : Quand l'euphémisme cristallise : *La linguistique*, *Vol. 52*(2), 271-290. https://doi.org/10.3917/ling.522.0271

Streckfuß, C., & Geiger, B. (Scénaristes). (2023, mai 15). Mein Körper, mein (Mehr-)Gewicht [Émission]. In *Mein Körper*. https://www.ardmediathek.de/video/mein-koerper-oder-ard-wissen/mein-koerper-mein-mehr-gewicht/ard/Y3JpZDovL2Rhc2Vyc3RlLmRlL2FyZC13aXNzZW4vMjAyMy0wNS0xNV8yMi01MC1NRVNa

Mathilde Gaillard
Université d'Assas

« *Reinvent your workforce* » : choix lexicaux dans deux discours spécialisés d'experts sur l'intelligence artificielle

Cette proposition s'inscrit dans le cadre de l'étude des variétés spécialisées de l'anglais et se concentre sur les choix lexicaux effectués dans les discours produits par deux milieux d'experts. Bien qu'ayant fait l'objet de plusieurs études en analyse du discours (Cussò et Gobin 2008 ; Léglise et Garric 2012 ; Oger 2021), le champ de l'expertise reste encore peu conceptualisé en anglais de spécialité et présente un intérêt pour le domaine dans la mesure où il recouvre une pratique transversale à divers milieux spécialisés. Nous proposons ici d'interroger le rôle que joue le lexique dans la construction de la visée argumentative inhérente aux discours d'experts, et notamment de l'éthos qui y est mis en œuvre.

Nous nous concentrons sur deux milieux spécialisés, les *think tanks* aux États-Unis d'une part et les consultants en management, aux États-Unis également, d'autre part. « Forme prototypique de l'expertise » (Cadiou 2006), le premier représente un réservoir d'experts dont l'objectif est de contribuer, voire d'influencer le débat politique à travers la diffusion d'expertise (Gaillard 2019). Le second a pour but de proposer des solutions clé en en main à ses clients (entreprises, institutions et gouvernements) en matière de gestion organisationnelle ou de stratégie (Monod et al. 2024), et s'engage ces dernières années dans des activités de diffusion d'expertise à un public élargi afin de faire connaître ses travaux. Dans les deux cas, il convient pour ces experts de légitimer leurs interventions souvent non sollicitées dans un marché des idées saturé (Drezner 2017). Aussi, notre hypothèse de travail est la suivante : l'enjeu de légitimation inhérent à la pratique de l'expertise se traduirait dans le discours par des choix lexicaux visant à construire un éthos d'expert crédible et solidaire des préoccupations de ses lecteurs.

Notre étude en synchronie s'appuie sur deux corpus composés des rapports d'expertise publiés par 6 organisations, *think tanks* et cabinets de conseil, sur l'impact de l'adoption de l'intelligence artificielle sur le monde du travail entre 2024 et 2025. Objet de multiples débats depuis l'apparition de ChatGPT fin 2022, l'intelligence artificielle interroge les acteurs économiques et politiques qui se tournent vers les experts afin d'en saisir les enjeux et d'en tirer profit. Le choix des mots pour appréhender ce nouveau domaine a donc toute son importance pour convaincre ces différents acteurs de la recevabilité des solutions proposées par les experts.

L'exploration des deux corpus se veut à la fois qualitative et quantitative, à l'aide de l'outil TXM (0.7.9). Deux modules d'analyse, structurel et interpersonnel, sont traités. Dans un premier temps, nous mettons en évidence le rôle des connecteurs argumentatifs (Ducrot 1980 ; Anscombre et Ducrot 1983) dans la construction d'une démonstration structurée. Nous nous concentrons sur les connecteurs à valeur de contraste (par exemple, « *however* ») qui permettent la mise en scène de l'argumentation du rapport d'expertise, à savoir l'identification d'une situation problématique nécessitant diagnostic et remédiation (Née et al. 2017). Ce diagnostic s'appuie sur des estimations chiffrées et des ordres de grandeur qui semblent objectiver les conséquences de l'intelligence artificielle sur le monde du travail, et contribuent à un apparent « parler neutre » de l'expert (Oger 2021, 268-270). Le second module d'analyse porte sur la dimension axiologique du lexique dans les rapports afin d'exposer l'« espace doxique » (Amossy 2021, 89), ou les valeurs partagées par experts et lecteurs. Un relevé manuel des marqueurs d'attitude dans les corpus est effectué. Nous montrons que le champ lexical du danger et de la rapidité participe à alimenter les inquiétudes quant à la magnitude du phénomène de l'intelligence artificielle, et ainsi à justifier la réponse d'un expert qui partage ces inquiétudes. À travers les champs de la justice sociale et de la transformation, s'élabore l'éthos d'un expert compétent, visionnaire, mais également doté d'empathie. Enfin, le choix de verbes dont le sens est changé par l'affixation du préfixe « *re* » (« *reinvent* », « *reshape* », « *reskill* ») pour véhiculer les recommandations des rapports participe d'une

rhétorique de la transformation, même modérée : le consultant, comme l'expert de *think tank*, devient partenaire et guide qui identifie les enjeux et propose des pistes d'action dans un avenir incertain (Kitay & Wright 2007).

Partant, nous espérons démontrer qu'au-delà des particularités inhérentes à chaque milieu spécialisé la pratique discursive de l'expert se traduit par des choix lexicaux similaires qui visent à légitimer ses interventions et à convaincre de leur bien-fondé.

Références bibliographiques :

Amossy, Ruth. 2021, *L'argumentation dans le discours*. 4e ed. Paris : Armand Colin.

Anscombre, Jean-Claude et Oswald Ducrot. 1983. L'argumentation dans la langue. Bruxelles : Pierre Mardaga.

Cadiou, Stéphane. 2006. « Savoirs et action publique : un mariage de raison ? L'expertise en chantier ». *Horizons stratégiques* [*En ligne*] 1/1. Consulté le 23 juin 2018. URL : <https://www.cairn.info/revue-horizons-strategiques-2006-1-page-112.htm>.

Cussò, Roser & Corinne Gobin. 2008. « Du discours politique au discours expert : Le changement politique mis hors débat ? ». *Mots. Les Langages du Politique* 88 : 5–11.

Drezner, Daniel. 2017. *The Ideas Industry*. New York : Oxford University Press.

Ducrot, Oswald. 1980. « Analyse de textes et linguistique de l'énonciation ». In *Les mots du discours*, dirigé par Oswald Ducrot, 7-56. Paris : Minuit.

Gaillard, Mathilde. 2019. *Contribution à la caractérisation du milieu des think tanks aux États-Unis à travers l'étude de leur discours et de leur culture professionnelle*. Thèse de doctorat. Paris : Sorbonne Université.

Kitay, Jim, et Christopher Wright. 2007. « From Prophets to Profits: The Occupational Rhetoric of Management Consultants ». *Human Relations* 60 (11): 1613-40. https://doi.org/10.1177/0018726707084302.

Léglise, Isabelle & Nathalie Garric. 2012. *Discours d'experts et d'expertise*. Bern : Peter Lange.

Monod, Emmanuel & Nataliia Korotkova & Sabine Khalil & Antonia Meythaler & Elisabeth Joyce. 2024. « Unveiling blind spots in consulting firms' disseminating discourse on digital transformation ». *Information Systems and e-Business Management* vol. 22(4): 759-803.

Née, Émilie, Claire Oger & Frédérique Sitri. 2017. « Le rapport : opérativité d'un genre hétérogène ». *Mots. Les langages du politique* [*En ligne*] 114. Consulté le 4 avril 2018. URL : <http://journals.openedition.org/mots/22752>.

Oger, Claire. 2021. *Faire référence, la construction de l'autorité dans les discours des institutions.* Paris : Editions EHESS.

Siroux, Jean-Louis. 2008. « La dépolitisation du discours au sein des rapports annuels de l'Organisation mondiale du commerce ». *Mots. Les langages du politique* [*En ligne*] 88/3, 13-23. Consulté le 18 mars 2014. URL : <https://www.cairn.info/revue-mots-2008-3-page-13.htm>.

Laure Gautherot
Université Sorbonne Nouvelle, Paris

‚*Bettenstop*' und ‚*Tourismus-Notbremse*': Die Lexik des Massentourismus im Mediendiskurs

Lange galt der Tourismus unter dem Begriff „ökonomische Fremdenverkehrswissenschaft" als Untersuchungsgegenstand der Wirtschaftswissenschaften (Spode, 1993: 27). Die Entwicklung in der zweiten Hälfte des 20. Jahrhunderts führt aber dazu, die Geschichte der Tourismusforschung unter dem Dach der historischen Sozialwissenschaft zu integrieren (Spode, 1993; Heuwinkel, 2023). Die einst „Ferienmenschen" (Krippendorf, 1984) genannten Urlauber:innen, die in die Sommerfrische fuhren, werden heute in den deutschsprachigen Medien als Akteure eines viel kritisierten *Sauftourismus* (1997 gebucht, dwds), *Massentourismus* bzw. *Übertourismus* (2020 im Duden gebucht) dargestellt, die sich auf das maßlose Verhalten der Tourist:innen und auf die negativen Auswirkungen für die Bevölkerung der Urlaubsorte (Umweltverschmutzung, Wohnungsnot) beziehen. Wenn das touristische Erlebnis selbst beeinträchtigt wird, lässt sich der Übertourismus aus ökonomischer Sicht den „Allmenden Problemen" zuordnen, wobei der Tourismus als eine allgemeine Ressource betrachtet wird, die nicht im Besitz einer Person, sondern der Allgemeinheit steht (Fritz/Maier, 2019).

Der Mediendiskurs über den Massentourismus schöpft aus zwei Teildiskursen: einem Expertendiskurs, der sich aus Interviews und Stellungnahmen von „Tourismusexperten" – professionellen Vertretern der Tourismusbranche und Anbietern touristischer Leistungen – zusammensetzt. Dazu zählen auch Beschreibungen des Phänomens Massentourismus von Tourismusforscherinnen und -forschern, wie beispielsweise Spode in der Reportage „Warum reist der Mensch so gern?" (rbbKultur, 17.05.2025). Der zweite Teildiskurs besteht aus Tatsachenberichten und Meinungen von Einwohnerinnen und Einwohnern der Urlaubsorte und von Reisenden, und wird als Laiendiskurs aufgefasst. Der Vortrag setzt sich als Ziel, in den beiden Teildiskursen die Rolle der Lexik bei der sachlich-neutral informierenden und der expressiv-wertenden, persuasiven Funktion der Mediensprache (Elsen/Dzikowicz, 2005; Knipf-Komlósi, 2009) zu beleuchten. Gezeigt werden soll, wie der Gebrauch von „Expertenwortschätzen" (Wichter, 1994) und „expressiven Wortbildungen" (Finkbeiner, 2023; Scherer, 2019) einem bestimmten Argumentationsmuster dient, das die Notwendigkeit einer Verhaltensänderung vonseiten der Reisenden begründet und die politische Intervention legitimiert.

Die hier eingenommene Forschungsperspektive stellt den Gebrauch der Wörter in den Mittelpunkt einer diskursorientierten Argumentationsanalyse (Römer, 2018; Wengeler, 2016; 2018). Die Diskussion um den Massentourismus in den deutschsprachigen Medien (digitale und gedruckte Presseartikel, Postings von Nachrichtensendern und Online-Zeitungen auf sozialen Netzwerken) seit 2023 liefert das Korpusmaterial unserer Analyse. Die Erholung des Tourismus nach der Corona-Pandemie wird 2023 nämlich durch das touristische Verhalten des „Revenge Travel" gekennzeichnet, 2024 wird die durch den Einsturz eines überfüllten Restaurants auf Mallorca ausgelöste Tourismuskrise zum Gegenstand des medialen Diskurses (Gautherot, i.E.). Unter dem morphopragmatischen Aspekt werden Anglizismen (z.B. *Overtourism*, *Too-rism*), Okkasionalismen wie *Bettenstop* oder *Tourismus-Notbremse,* die die „Kompositionsfreudigkeit der deutschen Sprache" (Schlücker, 2012) illustrieren, rhetorische Stilmittel wie die Flutmetapher (*Touristenstrom;* Dolomitental *quillt über*) analysiert. Unter dem semantischen Aspekt werden auch die Modifikationsverfahren der Bedeutungsverschlechterung und der „intentionale[n] Neu-Lexifizierung" (Harnisch, 2025: 221) beschrieben. Schließlich will der Vortrag aufzeigen, wie die angeführten Beispiele an der Schaffung eines Krisenbewusstseins (Römer, 2020) und eines klimaverträglichen Tourismusverhaltens beteiligt sind.

Bibliographie :

BUBENHOFER, Noah, 2009. *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse.* Berlin/New York: De Gruyter.

ELSEN, Hilke/DZIKOWICZ, Edita, 2005. „Neologismen in der Zeitungssprache", *Deutsch als Fremdsprache* 2/2005, 80–85.

FINKBEINER, Rita, 2023. „Brilleria, Hipsteria, Schlamperia: Hybride Suffigierung, Expressivität und Konstruktionsmorphologie", *Zeitschrift für Wortbildung/Journal of Word Formation* 2023/7(2), 150–190. https://doi.org/10.21248/zwjw.2023.2.103

FRITZ, Oliver/MAIER, Karin, 2019. „Wenn Reisende zur Belastung werden", *WIFO-Monatsberichte*, 2019, 92(11), 821–831.

GAUTHEROT, Laure (i.E.). „'Ein Tourist? Ich doch nicht!' Expressivität in den Kurzformen der Tourismuskrise". In: Irmtraud BEHR/Frank LIEDTKE (Hrsg.): *Gemeinsam die Krise bewältigen. Expressivität von Phrasemen und anderen Kurzformen in der Krisenkommunikation. Kongressakten der Internationalen Vereinigung für Germanistik 2025*, Graz: Grazer Universitätsverlag.

HARNISCH, Rüdiger, 2025. „Concept Creep – linguistisch". In: Dominique DIAS, Laure GAUTHEROT & Nathalie SCHNITZER (Éd.), *La langue en tension. Critiques sociétales et évolutions linguistiques. Mélanges en l'honneur d'Odile Schneider-Mizony*, Stuttgart: Ibidem Verlag, 213–234.

HEUWINKEL, Kerstin, 2023. *Tourismussoziologie*, 2. überarbeitete und ergänzte Auflage, Konstanz: UTB Verlag.

KNIPF-KOMLÓSI, Elisabeth, 2009. „Schnittstellen-Phänomene in der Zeitungssprache". In: Rita BRDAR-SZABÓ, Elisabeth KNIPF-KOMLÓSI & Attila PÉTERI (Hrsg.), *An der Grenze zwischen Grammatik und Pragmatik*, Frankfurt a.M. [u.a.]: Peter Lang, 199–211.

KRIPPENDORF, Jost, 1984. *Die Ferienmenschen – Für ein neues Verständnis von Freizeit und Reisen*, Zürich: Orell Füssli.

LEGALLOIS, Dominique/FRANÇOIS, Jacques, 2011. „Définition et illustration de la notion d'expressivité en linguistique". In: Nicole LE QUERLER, Franck NEVEU & Emmanuelle ROUSSEL (Eds.), *Relations, Connexions, Dépendances. Hommage au Professeur Claude Guimier*, Rennes: Presses universitaires de Rennes, 197–222.

RÖMER, David, 2018. „Argumentationstopoi in der Text- und Diskursanalyse – alte Pfade, neue Wege". *Tekst i dyskurs – text und diskurs*, 11(1), 117–135. https://doi.org/10.7311/tid.11.2018.05

RÖMER, David, 2020. „Diskurskritik". In: Thomas NIEHR, Jörg KILIAN & Jürgen SCHIEWE (Hrsg.), *Handbuch Sprachkritik*, Heidelberg: Metzler Verlag, 233–240.

SCHERER, Carmen, 2019. „Expressivität in der Wortbildung. Ein Überblick". In: Frank D'AVIS/Rita FINKBEINER (Hrsg.), *Expressivität im Deutschen*, Berlin/New York: De Gruyter, 49–74.

SCHLÜCKER, Barbara, 2012. „Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung". In: Livio GAETA/Barbara SCHLÜCKER (Hrsg.), *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte.* Berlin/New York: De Gruyter (= Linguistik Impulse & Tendenzen 46), 1–25.

SPODE, Hasso, 1993. „Historische Tourismusforschung". In: Heinz HAHN/H. Jürgen KAGELMANN (Hrsg.), *Tourismuspsychologie und Tourismussoziologie. Ein Handbuch zur Tourismuswissenschaft*, München: Quintessenz, 27–29.

WENGELER, Martin, 2016. „Diskursorientierte Argumentationsanalyse". In: Niehr, Thomas/Kilian, Jörg/Wengeler, Martin (Hrsg.): Handbuch Sprache und Politik. Bremen, 261–281.

WENGELER, Martin, 2018. „Diskurslinguistik als Argumentationsanalyse". In: Ingo H. WARNKE (Hrsg.), *Handbuch Diskurs*, BerlinBoston: De Gruyter, 242–264.

WICHTER, Sigurd, 1994. *Experten- und Laienwortschätze. Umriß einer Lexikologie der Vertikalität*, Tübingen: Niemeyer Verlag.

Webnachweise

„Südtirol: ‚Bettenstop' und mehr – Eine Region zieht die Tourismus-Notbremse", *Die Welt*, 20.5.2024. https://www.welt.de/wirtschaft/plus251171616/Suedtirol-Bettenstopp-und-mehr-Eine-Region-zieht-die-Tourismus-Notbremse.html (Abruf am 25.9.2025)

„Warum reist der Mensch so gern?", *rbbKultur Das Magazin*, 17.5.2025. https://www.rbb-online.de/rbbkultur-magazin/archiv/20250517_1830/traum-zeit-reise-buch-veroeffentlichung-tourismus-forscher-hasso-spode-kulturgeschichte.html (Abruf am 25.9.2025)

„Wohin die Reise geht", *Die Tageszeitung*, 24.7.2023. https://taz.de/Erholung-des-Tourismus/!5944256/ (Abruf am 5.5.2026)

Zinaida Geylikman
Independent Researcher

Modeling human nouns' lexicon in diachrony: a methodological essay

Speaking about any human being – a real or an imaginary one, someone one knows or someone who died hundreds of years ago – one has a rather large choice of linguistic labels. These include personal names, pronouns and human nouns (henceforth HN) – common nouns that refer to human beings in any context whatsoever.

As Dahlgren (1985: 381) puts it, "persons are conceived in a richer and more complex way than the physical objects". Indeed, the choice of the precise label in the given utterance is never random, and, above the descriptive function, is conditioned by both the speaker and addressee's identities, the pragmatic aim of the utterance and the overall socio-historical context. As a part of natural languages lexica, HN lexicon changes continually, its change being easily perceptible within a lifetime. This evolution can be driven by objective reality, *e.g.* the emerging of a new phenomenon, as in *blogger*, or by ideological reasons, *e.g*. new neutral terms replacing existing ones judged as negatively connotated or inappropriate, such as *disabled person* instead of *invalid*. Furthermore, most of the realities HN refer to are social constructs: one cannot become *director* or *president* unless he/she is named so (Baider & Todirascu 2018). As labels for social constructs, HN undergo semantic change along with the realities they designate: *cf.* Coady (2025) that demonstrates the challenges of grasping the meaning of a word as common as *woman* in a society marked by heated gender debates. Thus, the evolution of the HN lexicon reflects social change on multiple levels while contributing to the construction of social realities.

Different subgroups of human nouns across languages have drawn the attention of various branches of human and social sciences including linguistics: kinship terminology studies, categorization studies, gender studies, morphology, terms of address studies *etc*. Nevertheless, HN lexicon has rarely been considered as a whole. Apart from several non-systematic works (Braun 1997, Gross 2008), the recent NHUMA project highlighted the relevance of HN lexicon as a study object. Focusing mainly on contemporary state of languages, it has resulted in several programmatic publications, such as Mihatsch & Schnedecker 2015, Schnedecker & Mihatsch 2018 that demonstrate a finer semantic differentiation of HN in comparison to other lexica and propose a theoretical framework for HN studies. Nevertheless, the project has not been conclusive as for the internal unity of the lexicon.

We argue that while its unity can indeed be questioned from a strictly lexicological point of view, the relevance of HN as a lexicon goes beyond the proper linguistic level – as a cultural object that responds to and drives social evolution. Therefore, in order to model the HN lexicon and its evolution overtime one must consider not only proper linguistic features (etymology, gender, number, meaning, phraseological patterns), but also sociolinguistic features (*e.g.* referent's gender and social group), pragmatic features (*e.g.* speech act's type if applicable, power/solidarity relationship between speakers in direct discourse *etc.*), as well as the overall situation of communication (metatextual data).

This talk presents the results of the first stage of a large-scale research project that aims to investigate the interdependence between the HN lexicon and societal change through a long-term, cross-discursive analysis of semantic evolution. Its core hypothesis is that the HN lexicon both reflects and drives social transformations, embodying power dynamics and cultural narratives and shaping collective representations. To test this, the project develops an innovative methodology integrating historical lexicology, pragmatics, sociolinguistics, and discourse analysis, enriched by digital humanities and NLP tools. More precisely, it aims to produce a dynamic, multifactorial model of HN evolution on the example of Medieval French.

Medieval French (9th-15th century, GGHF 2020, vol. 1: LI) appears as a relevant period for a large-scale diachronic study of HN lexicon evolution for several reasons. Thus, it is represented by a considerable volume of digital databases, its

linguistic evolution on all levels is synthetized in grammars (Buridant 2019; GGHF 2020) and its lexica are described by numerous dictionaries (FEW, DEAF, TL, GdF, DMF, DéROM). At the same time, the corresponding period of European history witnessed considerable social changes that are extensively studied by historical science allowing to put lexical evolution in relation with social changes.

This contribution provides the results of a small-scale case study of HN in Old and Middle French (12th-15th century) texts conceived in order to test our methodological choices. Our corpus includes 3 discourse traditions (Koch 1997) of Medieval writings, each one represented by four texts from 12th to 15th century from the BFM22 database (overall 12 texts, 526024 tokens): *chivalric romance*, *hagiographies* and *chronicles*.

The analyzed dataset includes all utterances of HN in the corpus (30 words before and after the pivot). We consider as HN any common noun or nominalized adjective referring to a human being. The list of HN is obtained in two steps: based on the common nouns' index in BFM22 that allows to spot most items, then through manual reading of texts allowing to spot rare items and *ad hoc* creations. The utterances are extracted automatically through lemma search; the resulting dataset is annotated automatically (*e.g.* for textual metadata), semi-automatically (*e.g*. for grammatical gender) or manually (*e.g.* for speech act type). Later on, the dataset is explored through multiple correspondence analyses in order to identify big patterns (Glynn 2014), followed by cluster analyses that allows to detect associations between different factors. As a result, we obtain several classifications within the lexicon allowing a holistic linguistic and extra-linguistic description of the usage of the HN lexicon, its diachronic evolution and variation across discourse traditions.

Our contribution is thus meant as a methodological essay that demonstrates the relevance of HN lexicon studies in diachrony and paves the way for a large-scale project of modelling HN lexicon and its evolution in Medieval French.

Manuals, dictionaries and databases

BFM22 = Base de Français Médiéval [En ligne]. Lyon: ENS de Lyon, Laboratoire IHRIM, 2022, <txm.bfm-corpus.org>.

FEW = Wartburg, W., et al., (1922-2002), Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes, 25 vol., Bonn/Heidelberg/Leipzig-Berlin/Basel, Klopp/Winter/Teubner/Zbinden.

DMF = Dictionnaire du Moyen Français, version 2023 (DMF 2023). ATILF - CNRS & Université de Lorraine. URL: http://www.atilf.fr/dmf

DEAF = Dictionnaire étymologique de l'ancien français (DEAF). K. Baldinger/S. Dörr/F. Möhren/T. Städtler/S. Tittel/ Niemeyer, Tübingen/Presses universitaires de Québec, 1974-2019

GdF = Godefroy, Frédéric, Dictionnaire de l'ancienne langue française et de tous ses dialectes du XIe au XVe siècle, Paris, F. Viewig, 1881-1902

GGHF = Marchello-Nizia, Christiane, Bernard Combettes, Sophie Prévost, and Tobias Scheer (eds.) 2020. Grande Grammaire Historique du Français (GGHF). De Gruyter.

NHUMA = Les noms d'humains : de la description linguistique aux applications lexicographiques, https://nomsdhumains.weebly.com/

TL = Tobler, Adolf & Lommatzsch, Erhard, Altfranzösisches Wortebuch, [1928, 1930], Stuttgart-Wiesbaden, Franz Steiner Verlag, 2002.

References

Baider, F. H., & Todirascu, A. C. (2018). Deux sous-catégories de noms d'humains à définir: Les noms de statut et de fonction. In C. Schnedecker & W. Mihatsch (Eds.), Nouvelles approches en sémantique lexicale. Les noms d'humains (théorie, méthodologie, classification) (pp. 229–250). Mouton de Gruyter.

Braun, P. (1997). Personenbezeichnungen: der Mensch in der deutschen Sprache. Reihe Germanistische Linguistik, 189.

Buridant, C. (2019). Grammaire du français médiéval: XIe-XIVe siècles. ELIPHI, Éditions de linguistique et de philologie.

Coady, A. (2025). The semantic adventures of woman. Lexis, 25. https://doi.org/10.4000/14i19

Dahlgren, K. (1985). The Cognitive Structure of Social Categories*. Cognitive Science, 9(3), 379–398. https://doi.org/10.1207/s15516709cog0903_4

Glynn, D. (2014). Correspondence analysis: Exploring data and identifying patterns. In D. Glynn & J. A. Robinson (Eds.), Human Cognitive Processing (Vol. 43, pp. 443–485). John Benjamins Publishing Company. https://doi.org/10.1075/hcp.43.17gly

Gross, G. (2008). Sur le statut syntaxique des substantifs humains (p. 27). Université de Savoie. https://shs.hal.science/halshs-00411322

Huyghe, R. (2015). Les typologies nominales : présentation. Langue Française, n° 185(1), 5–27. https://doi.org/10.3917/lf.185.0005

Koch, P. (1997). Diskurstraditionen: zu ihrem sprachtheoretischen Status und ihrer Dynamik. In B. Frank, T. Haye, & D. Tophinke (Eds.), Gattungen mittelalterlicher Schriftlichkeit. Narr.

Mihatsch, W., & Schnedecker, C. (2015). Les noms d'humains: une categorie a part ? Franz Steiner Verlag.

Schnedecker, C., & Mihatsch, W. (Eds.). (2018). Les noms d'humains—théorie, méthodologie, classification: Nouvelles approches en sémantique lexicale. De Gruyter. https://doi.org/10.1515/9783110586169

Simone Gomes / Anna Ladilova / Katharina Müller
Justus Liebig Universität Giessen

Néologismes et emprunts en langues romanes : langue et société dans le contexte de la pandémie de Covid-19

Ce travail examine comment les transformations sociales survenues pendant la pandémie de Covid-19 ont produit des changements linguistiques dans l'espagnol, le français, l'italien et le portugais. En accordant une attention particulière à la néologie, cette étude analyse comment chaque société a compris la crise et formulé ses propres réponses socioculturelles durant la première année de la pandémie. En nous appuyant sur la lexicologie sociale (Matoré 1953 ; Cambraia 2013), nous analysons, à travers un corpus de presse à grande diffusion en Espagne, en France, en Italie et au Brésil (*El País* et *El Mundo*, *Le Monde*, *La Repubblica* et *Folha de São Paulo*), l'intégration, la diffusion et l'usage des termes *confinement, lockdown, quarantaine, isolement, fermeture* et *couvre-feu* (et de leurs équivalents dans les autres langues), afin d'étudier leur spécialisation contextuelle et de comprendre comment ils participent à la construction d'une réalité sociale différenciée selon le pays. Ces unités lexicales, représentatives du champ lexical (cf. Abbade 2012) des « mesures prophylactiques », témoignent d'une forte instabilité sémantique : dans certains contextes, elles se rapprochent et peuvent être utilisées comme des synonymes ; dans d'autres, elles font l'objet d'une spécialisation sémantique, se distinguant ainsi pour désigner des réalités différentes. En s'appuyant sur la dichotomie de Sablayrolles (2018 ; 2019), l'étude distingue les matrices internes, issues de la néologie endogène, de la matrice externe. Cette dernière couvre le continuum de l'emprunt, de l'importation brute (xénismes) à l'emprunt lexicalement intégré.

La constitution des corpus a reposé sur l'utilisation d'un web crawler conçu pour extraire, dans les journaux analysés, l'ensemble des articles et des commentaires de lecteurs traitant de la COVID-19. Ce dispositif, associé à des outils de programmation en Python, a permis de générer une base de données traitées ensuite à l'aide du logiciel *AntConc*. L'analyse des données a été réalisée selon une double approche : quantitative, à travers l'étude des fréquences, des colocations (*collocates*), des clusters/N-grammes et des listes de mots ; qualitative, par l'examen rigoureux des concordances pour saisir le sens des termes en contexte d'usage. En complément, nous avons eu recours aux dictionnaires de référence de chaque langue – notamment l'Aulete, le Larousse, le Littré, le DRAE et le Treccani – ainsi qu'à des dictionnaires étymologiques. Cette démarche a permis de dater l'apparition des néologismes, d'identifier les emprunts et de vérifier les glissements sémantiques survenus durant cette période.

L'étude a révélé comment chaque société a appréhendé la pandémie, reflétant des visions du monde ainsi que des positionnements politiques et idéologiques spécifiques. En français, par exemple, le terme *lockdown* est demeuré marginal et apparaît souvent entre guillemets, signalant une résistance à l'anglicisme. Cette posture est influencée par une politique linguistique qui cherche à préserver la langue des influences étrangères. À l'inverse, le portugais brésilien a rapidement intégré le terme *lockdown* dans l'usage courant, suscitant de moins en moins d'étrangeté au fil du temps, ce qui reflète une plus grande ouverture du portugais brésilien à l'incorporation d'emprunts. En italien, on observe un phénomène similaire (cf. Sgroi 2020 : 88ss.), où le terme génère des dérivés tels que *post[-]lockdown* ou *dopo[-]lockdown*. En espagnol, le terme *confinamiento* et le verbe *confinar* ont déjà été intégrés au DRAE à la fin de l'année 2020 (cf. Coll/Barité 2022 : 133) ; cependant, leur usage prédomine en Espagne, alors que l'Argentine et d'autres pays latino-américains ont privilégié *aislamiento* et *cuarentena* (cf. Adelstein/de los Ángeles Boschiroli 2022 : 96). En français, le mot *confinement* s'est imposé comme une néologie semantique majeure, corroborant l'hypothèse de Thiéry-Riboulot (2020 : 137), selon laquelle le succès de sa diffusion à partir de 2020 serait lié à sa plasticité polysémique et à sa capacité de s'articuler à divers référents : aux mesures prises (*mesures de confinement*), à la situation territoriale (*le confinement de la ville*), à la période (*période de confinement*) et aux personnes (*le confinement des malades*). Les dérivations comme *déconfinement* et la forte fréquence du terme dans les commentaires des lecteurs confirment sa diffusion dans la langue courante (cf. Reboul-Touré 2021 : 414 ; Kolářiková

2021 : 37). À l'inverse de l'emprunt *lockdown*, la vitalité de *confinement* illustre la préférence du français pour des mécanismes de création interne. En espagnol, on observe un phénomène similaire avec *confinamiento* et *desconfinamiento* ainsi qu'avec les verbes *confinar* et *desconfinar* (cf. Adelstein/de los Ángeles Boschiroli 2022 : 113). En italien, en revanche, le terme *confinamento* a été plus rarement mobilisé, ce qui, selon Thiéry-Riboulot (2020 : 135), s'expliquerait par sa proximité avec le mot italien *confine* (« frontière »). À la place, les locuteurs ont privilégié les calques *isolamento* et *chiusura* (cf. Sgroi 2020 : 93). En portugais, c'est *isolamento* qui a le mieux représenté les mesures de contention mises en œuvre au Brésil. Ce terme a incarné un fort potentiel polysémique pour exprimer la complexité du phénomène (cf. Gomes/Cordeiro 2024), évoluant dans un contexte d'intenses affrontements politiques et idéologiques qui ont marqué la gestion de la crise au Brésil. Cette polarisation aurait conduit à une fragmentation – ou « pulvérisation » – des mesures de type *isolamento/ confinamento*, orientant les débats vers des directions divergentes de celles observées en France et en Espagne. Dans ces derniers pays, les mesures ont été plus centralisées et ont bénéficié d'une plus grande adhésion de la part des gouvernants et de la population.

L'étude met en lumière l'importance de prendre en compte les facteurs historiques, sociaux et culturels dans la compréhension des phénomènes linguistiques, révélant l'indissociabilité des liens entre langue et culture. Elle réaffirme la nature vivante du langage, sa capacité à se reconfigurer face à de nouvelles réalités, structurant ainsi les changements dans les rapports sociaux et dans les perceptions du réel. En définitive, cette analyse contrastive démontre que le choix des termes – qu'ils soient issus de créations internes ou d'emprunts – n'est jamais neutre, mais constitue une véritable trace mémorielle de la manière dont chaque société a traversé la crise sanitaire.

Références bibliographiques :

Abbade, Celina M. S. 2012. *Lexicologia social. A lexemática e a teoria dos campos lexicais*, in: Isquerdo, Aparecida/Seabra, Maria Cândida (eds.). *As ciências do léxico*, Campo Grande: Editora UFMS, 141–161.

Adelstein, Andreína/de los Ángeles Boschiroli, Victoria. 2022. Spanish neologisms during the COVID-19 pandemic: Changing criteria for their inclusion and representation in dictionaries. In: Klosa-Kückelhaus, Annette/ Kernerman, Ilan (eds.). *Lexicography of coronavirus-related neologisms*, Berlin/Boston: De Gruyter, 93-124. DOI: 10.1515/9783110798081-006.

Cambraia, César N. 2013. Da lexicologia social a uma lexicologia sócio-histórica. Caminhos possíveis. *Estudos de linguagem* 21(1), 157–188.

Cartier, Emmanuel et al. 2022. Linguistic repercussions of COVID-19: A corpus study on four languages. *Open Linguistics* 8(1), 751–766. DOI:10.1515/opli-2022-0222.

Coll, Magdalena/Barité, Mario. 2022. Specialized voices in the 23rd edition of the *Diccionario de la lengua española*: Analysis of the COVID-19 field and its neologisms. In: Klosa-Kückelhaus, Annette/Kernerman, Ilan (eds.). *Lexicography of coronavirus-related neologisms*. Berlin/Boston: De Gruyter, 125-145. DOI: 10.1515/9783110798081-007.

Gomes, Simone/Cordeiro, Maryelle. 2024. Falando a língua na pandemia: mudanças sociais e linguísticas em contexto românico. *Quo Vadis Romania?*, v. 63, p. 36-57.

Koláříková, Dagmar. 2021. Reflet de la pandémie de Covid-19 dans les dictionnaires de la langue française. *Studia Romanistica* 21(2), 31-45. DOI: 10.15452/SR.2021.21.0008.

Matoré, Georges. 1953. *La méthode en lexicologie. Domaine français*, Paris: Didier.

Pietrini, Daniela. 2021. *La lingua infetta. L'italiano della pandemia*, con una presentazione di Giuseppe Antonelli. Roma: Treccani.

Reboul-Touré, Sandrine. 2021. The Crisis in Discourse: As an Event, a Discursive Semantics, and a Culture. *Zeitschrift für Literaturwissenschaft und Linguistik* 51, 399–420. DOI:10.1007/s41244-021-00211-5

Sablayrolles, Jean-François. 2018. *Néologismes et emprunts : des relations complexes*. Communication présentée au 4ème Congrès international de néologie dans les langues romanes (CINEO 2018), Lyon, France.

Sablayrolles, Jean-François (2019). *Comprendre la néologie. Conceptions, analyses, emplois.* Limoges: Lambert Lucas.

Sgroi, Salvatore Claudio. 2020. *Dal Coronavirus al Covid-19. Storia di un lessico virale*. Alessandria: Edizioni dell'Orso.

Thiéry-Riboulot, Véronica. 2020. Une étude de sémantique historique du mot confinement. *Mots. Les langages du politique* 124, 127-144.

Marina Gonzalez
Université de Madrid

The acquisition of subject-specific vocabulary and CLIL: Word production and lexical networks

Content and Language Integrated Learning (CLIL) programmes have been increasingly implemented across Europe as a way to promote multilingualism and to enhance foreign language acquisition. CLIL is "a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language" (Coyle et al., 2010, p. 1). Thus, it allows for greater exposure to English as a Foreign Language (EFL), enabling students to access a wider range of lexical resources (Merino & Lasagabaster, 2018). In fact, previous research comparing CLIL and traditional EFL students concluded that CLIL learners exhibit broader vocabulary knowledge (Canga Alonso & Arribas García, 2015) and display greater lexical richness (Agustín Llach, 2017). However, it is particularly relevant to explore the acquisition of subject-specific vocabulary (SSV), as learners need to be familiarised with this type of vocabulary when doing academic study in EFL (Nation, 2013). Hence, the main objectives of this empirical study are (1) to ascertain CLIL and traditional EFL learners' SSV size, (2) to analyse the kinds of lexical items produced, and (3) to account for similarities and differences in the organisation of SSV in their mental lexicons. For this purpose, a lexical availability task (LAT) was conducted to explore the vocabulary that 8th-grade CLIL (n = 11) and non-CLIL (n = 10) learners produce triggered by the prompt *Physical Education*. The elicited responses were analysed by using LexPro (Hernández Muñoz et al., 2023) and VocabProfilers (Heatley et al., 2002). Results indicate that CLIL learners possess a more extensive and advanced SSV than their non-CLIL peers. Moreover, CLIL learners present a more proficient lexical organisation than the traditional EFL learners, as they present more interconnected and centralised lexical networks. Therefore, these findings align with previous research (Castellano-Risco, 2019; Geoghegan, 2023; Rieder-Bünneman et al., 2022) suggesting that the CLIL approach enhances SSV growth among students.

References

Agustín Llach, M. P. (2017). The effects of the CLIL approach in young foreign language learners' lexical profiles. *International Journal of Bilingual Education and Bilingualism*, *20*(5), 557-573. https://doi.org/10.1080/13670050.2015.1103208.

Agustín Llach, M. P., & Rubio, J. (2024). Navigating the Mental Lexicon: Network Structures, Lexical Search and Lexical Retrieval. *Journal of Psycholinguistic Research*, *53*, 1-23. https://doi.org/10.1007/s10936-024-10059-8.

Canga Alonso, A., & Arribas García, M. (2015). The Benefits of CLIL Instruction in Spanish Students' Productive Vocabulary Knowledge. *Encuentro: Revista de investigación e innovación en la clase de idiomas*, *24*, 15-31. https://hdl.handle.net/10017/25338.

Castellano-Risco, I. O. (2019). La influencia del enfoque de AICLE en el vocabulario académico: El caso de alumnos de Educación Secundaria. *Cauce: Revista Internacional de Filología, Comunicación y sus Didácticas*, *42*, 135-148.

Coyle, D., Hood, P., & Marsh, D. (2010). *Content and Language Integrated Learning*. Cambridge University Press.

Geoghegan, L. (2023). The effect of CLIL on productive thematic vocabulary. *International Journal of Applied Linguistics*, *34*(2), 1-15. https://doi.org/10.1111/ijal.12506.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programmes. Available at https://www.victoria.ac.nz/lals/staff/paul-nation.aspx.

Hernández Muñoz, N., Tomé Cornejo, C., López García, M., & Bartol Hernández, J. A. (2023). *Manual de LexPro*. https://dispogram.usal.es/.

Merino, J. A., & Lasagabaster, D. (2018). CLIL as a way to multilingualism. *International Journal of Bilingual Education and Bilingualism*, *21*(1), 79-92. https://doi.org/10.1080/13670050.2015.1128386.

Rieder-Bünemann, A., Hüttner, J., & Smit, U. (2022). 'Who would have thought that I'd ever know that!': subject-specific vocabulary in CLIL student interactions. *International Journal of Bilingual Education and Bilingualism*, *25*(9), 3184-3198. https://doi.org/10.1080/13670050.2021.2020211.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language*. Cambridge University Press.

Frédérique Henry /Marie-Laurence Knittel / Evelyne Jacquey / Sandrine Ollinger
Université de Lorraine, ATILF, CNRS

Étude linguistique et textométrique des termes clés du processus du bilan orthophonique.

Le bilan orthophonique constitue une étape fondamentale engageant les compétences cliniques, terminologiques et communicationnelles de l'orthophoniste. Il sert à évaluer le caractère ordinaire, atypique ou pathologique d'un dysfonctionnement ou déficit chez une personne dans les domaines de la communication, du langage, de la voix et des fonctions cognitives et oromyofaciales. Pour Bélot & Tricot (2000) « il constitue à lui seul une charnière entre la souffrance du patient et l'élaboration du projet thérapeutique qui lui sera proposé ». (Coquet 2000) précise que « cette démarche d'évaluation débouche sur une décision de prise en charge éventuelle ».

L'orthophoniste décide du temps nécessaire, des moyens à mettre en œuvre et des outils à utiliser pour ce bilan. Il permet d'identifier et d'évaluer les difficultés, et aboutit à un diagnostic orthophonique, indispensable[6[1]] à la mise en œuvre légale de rééducation couverte en France par la sécurité sociale (Kremer & Lederlé 2020). Le diagnostic orthophonique ne se réduit pas à une simple identification nosographique mais résulte d'un processus terminologique (langagier et conceptuel) mettant en correspondance :

- Des représentations conceptuelles des pathologies chez l'orthophoniste ;
- Des observations cliniques et des résultats de tests et d'évaluations, dans un contexte spécifique à chacun (Bachman 2007) ;
- Une sélection terminologique parmi des nomenclatures et des classifications diverses.

Par ailleurs, les notions de diagnostic et d'évaluation sont soumises à des normes sociales, individuelles et collectives, et à des inférences interdisciplinaires présentes dans la coordination des soins. Elles relèvent d'une responsabilité éthique et épistémique, décrivant les mécanismes linguistiques dysfonctionnels et influant sur les réalités vécues par le patient (OMS CIF 2001).

Notre étude vise la description du bilan orthophonique d'un point de vue linguistique, terminologique et conceptuel. Pour ce faire, nous exploitons des ressources textuelles, termino-ontologiques et lexicographiques déjà constituées. À court et moyen terme, ces travaux s'intéressent aux termes en orthophonie et à leur usage dans des corpus écrits. Plus globalement, les résultats visent à alimenter une réflexion épistémique sur les sciences orthophoniques.

Nous analysons les usages linguistiques de couples Nom/Verbe qui sont révélateurs de concepts clés relatifs à chaque étape du bilan orthophonique, par exemple *diagnostic/diagnostiquer*. Ces couples sont extraits du corpus OrthoCorpus[7[2]] V3, comportant 1153 articles de la revue Rééducation Orthophonique (5,5 millions de tokens), parus entre 1997 et 2022 et rédigés par des orthophonistes et d'autres professionnels de la santé ou de l'éducation.

Nous articulons une approche linguistique (morphosyntaxique et sémantique, par l'étude des contextes droits et gauches) et une approche textométrique (analyse des cooccurrents spécifiques) comme suit :

1. Analyse linguistique :

- À l'aide du logiciel TXM (Heiden, Magué & Pincemin 2010), extraction des concordances du Nom et du Verbe (10 mots) et export sous Excel ;
- Annotation syntaxique et aspectuelle, puis sémantique des contextes immédiats (classes sémantiques déterminées par une chercheuse en orthophonie puis ajustées par le groupe).

6[1] voir le code de la santé publique, le décret n°2002-721 du 2 mai 2002 et la loi n° 2023-379 du 19 mai 2023
7[2]https://www.ortolang.fr/market/corpora/orthocorpus

2. Analyse textométrique :

- Dans le sous-corpus composé des paragraphes contenant le Nom ou le Verbe choisi, calcul des cooccurrents spécifiques du Nom ou du Verbe (calcul des spécificités : Lafon 1980) et export sous Excel ;
- Sélection des cooccurrents spécifiques de catégorie adjectif, nom et verbe ;
- Annotation sémantique des cooccurrents en cohérence avec la classification réalisée dans l'analyse des concordances.

Les annotations appliquées selon ces deux approches (syntaxiques, aspectuelles et sémantiques) sont ensuite croisées afin de produire :

- Une interprétation disciplinaire en se fondant sur les proportions obtenues pour la répartition des annotations ;
- Une interprétation linguistique mettant en valeur les différences d'usage entre nom et verbe du point de vue de l'aspect, des structures argumentales et restrictions de sélection.

Pour le corpus constitué du couple *diagnostic* (2194 occurrences) / *diagnostiquer* (217 occurrences), l'analyse linguistique des concordances montre que les thématiques principales retenues concernent le patient (pathologie, symptôme), le type de diagnostic posé ainsi que ses caractéristiques temporelles (moment de la pose). On relève également trois premiers constats :

- Les occurrences du nom *diagnostic* tendent à montrer que, en contexte, cette notion qui se suffit à elle-même : dans quasiment 40 % des cas, *diagnostic* et *diagnostiquer* ne sont suivis d'aucune autre unité lexicale ;
- Dans un quart des cas des 60 % d'occurrences restantes, le processus de diagnostic résulte en l'attribution d'une étiquette concernant le patient (ex : le patient diagnostiqué autiste, dyscalculique ou TDAH), ou plus fréquemment concernant une pathologie (ex : *diagnostic de démence sémantique)* ;
- Les verbes introduisant *diagnostic* montrent que les orthophonistes s'interrogent et décrivent le processus permettant d'aboutir au diagnostic (*proposer, faciliter un diagnostic*), y compris du côté du patient (*digérer un diagnostic*).

Du point de vue de l'approche textométrique, le calcul de spécificité appliqué aux cooccurrents de *diagnostic* donne une image des préoccupations des auteurs. La question des moyens utilisés dans la réalisation du diagnostic, du fait que le processus du diagnostic soit un questionnement (processus réflexif) et les pathologies diagnostiquées sont 3 axes massivement présents dans le corpus. Les orthophonistes décrivent principalement le type de diagnostic posé et le processus en jeu (*annonce, approfondir*), la pathologie orthophonique identifiée (*dyslexie, dysphasie*), alors qu'en comparaison, les cooccurrents significatifs dans les paragraphes rédigés par des non-orthophonistes mentionnent surtout une temporalité (*précoce*, *dépistage*).

Sur la base de ces résultats, nous étendrons les travaux à deux autres couples moins limités sémantiquement, *évaluation/évaluer* et *examen/examiner*, puis nous mettrons les analyses en lien avec les étapes du bilan, sujet qui est, par ailleurs, encore mal documenté en orthophonie.

Nous souhaitons présenter une communication orale permettant d'expliciter les enjeux de ces travaux pour l'orthophonie et pour les sciences du langage. Nous exposerons :

- Les modèles conceptuels et représentations du bilan orthophonique sur lesquels nous nous sommes appuyées et auxquels nous avons abouti ;
- La méthode de recherche employée et les résultats quantitatifs et qualitatifs obtenus pour les couples étudiés.

Références bibliographiques :

Bachman, L. F., 2007. 3. What is the Construct? The Dialectic of Abilities and Contexts in Defining Constructs in Language Assessment. In : Fox, J. ; Wesche, M. ; Bayliss, D. ; Cheng, L. ; Turner, C. E. ; Doe, C. (eds.). *Language Testing Reconsidered*. Ottawa : Les Presses de l'Université d'Ottawa. Disponible à l'adresse : https://books.openedition.org/uop/1563 (Consulté le 13 août 2025).

Fox, J. ; Wesche, M. ; Bayliss, D. ; Cheng, L. ; Turner, C. E. ; Doe, C. (eds.), 2007. *Language Testing Reconsidered*. Ottawa : Les Presses de l'Université d'Ottawa. Disponible à l'adresse : https://books.openedition.org/uop/1542 (Consulté le 13 août 2025).

Bélot, C. ; Tricot, M., 2002. Les orthophonistes et le bilan : État des lieux en juin 2000. *Glossa*, No. 80, p. 18-27. Disponible à l'adresse : https://www.glossa.fr/index.php/glossa/article/view/125 (Consulté le 13 août 2025).

Brin-Henry, F., 2012. Le diagnostic orthophonique comme clé de voûte du bilan ? In : « *Le Bilan orthophonique a sa Place : Mesurer, Comprendre, Soigner* ». *Journée de l'Association de formation continue de l'École d'orthophonie de Nantes*. Disponible à l'adresse : https://hal.science/hal-01083111v1 (Consulté le 13 août 2025).

Cattini, J. ; Duboisdindien, G., 2022. L'évaluation critériée statique formelle en orthophonie. *Glossa*, No. 133, p. 27-47. Disponible à l'adresse : https://www.glossa.fr/index.php/glossa/article/view/1351 (Consulté le 13 août 2025).

Chauvin D. ; Demouy, J., 2013. Bilan orthophonique. In : Ferrari, P.; Bonnot, O. (Dir.). *Traité européen de psychiatrie et de psychopathologie de l'enfant et de l'adolescent*, Lavoisier. Disponible à l'adresse : https://compagnon.librairiemedicale.com/pedopsychiatrique/Traite_psychiatrie_eChapitre5.pdf (Consulté le 13 août 2025).

Coquet, F., 2000. Essai de modélisation de la démarche de bilan orthophonique. *Glossa,* No. 74, p. 38-42. Disponible à l'adresse : https://www.glossa.fr/index.php/glossa/article/view/159 (Consulté le 13 août 2025).

Diepeveen, S. ; Van Haaften, L. ; Terband, H. ; De Swart, B. ; Maassen, B., 2020. Clinical Reasoning for Speech Sound Disorders: Diagnosis and Intervention in Speech-Language Pathologists' Daily Practice. *American journal of speech-language pathology*, Vol. 29, No. 3, p. 1529-1549.

Duboisdindien, G.; Cattini, J., 2024. L'expertise et le raisonnement cliniques de l'orthophoniste dans la pratique évaluative. In : Brin-Henry, F. (Dir.). *Orthophonie : identité professionnelle et expertise*, 216 pages. De Boeck Supérieur. Disponible à l'adresse : https://www.deboecksuperieur.com/numerique/9782807358348-orthophonie (Consulté le 13 août 2025).

Jones, R. ; Barraclough, K. ; Dowrick, C., 2010. When no diagnostic label is applied. *BMJ (Clinical research ed.)*, No. 340, c2683. Accessible à l'adresse : https://doi.org/10.1136/bmj.c2683 (Consulté le 13 août 2025).

Heiden, S. ; Magué, J.-Ph. ; Pincemin, B, 2010. TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In : *JADT 2010 : 10th International Conference on the Statistical Analysis of Textual Data*, p. 1021-1032. Rome, Italie. Disponible à l'adresse : http://halshs.archives-ouvertes.fr/docs/00/54/97/79/PDF/Heiden_al_jadt2010.pdf (Consulté le 13 août 2025).

Jutel, A., ; Nettleton, S, 2011. Towards a sociology of diagnosis: Reflections and opportunities. *Social Science & Medicine*, Vol. 73, No. 6, p. 793-800. Accessible à l'adresse : https://doi.org/10.1016/j.socscimed.2011.07.014 (Consulté le 13 août 2025).

Kremer, J. ; Lederlé, E., 2020. *L'orthophonie en France*, 128 pages. Paris : Presses Universitaires de France, Collection Que sais-je ?

Lafon, P., 1980. Sur la variabilité de la fréquence des formes dans un corpus, *Mots*, Vol. 1, p. 128-165.

Organisation Mondiale de la santé, 2001. Classification Internationale du Fonctionnement, du Handicap et de la Santé (CIF).

Plug, L. ; Sharrack, B. ; & Reuber, M., 2010. Seizure, Fit or Attack? The Use of Diagnostic Labels by Patients with Epileptic or Non-epileptic Seizures. In : *Applied Linguistics*, Vol. 31, No. 1, p. 94-114. Accessible à l'adresse : https://academic.oup.com/applij/article-abstract/31/1/94/182785 (Consulté le 13 août 2025).

Sims, R. ; Michaleff, Z. A. ; Glasziou, P. ; Thomas, R., 2021. Consequences of a Diagnostic Label: A Systematic Scoping Review and Thematic Framework. *Frontiers in Public Health*, 22 pages, Vol. 9. Accessible à l'adresse : https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2021.725877 (Consulté le 13 août 2025).

Woollven, M., 2021. Diagnostiquer les difficultés scolaires. Étude des outils et des processus diagnostiques de la dyslexie en France et au Royaume-Uni. *Sociologie*, Vol. 12, No. 3, p. 285-302. Accessible à l'adresse : https://shs.cairn.info/article/SOCIO_123_0285/pdf?lang=fr (Consulté le 13 août 2025).

Juan Carlos Justiano / Maria Rubio
Université de Madrid

Denominative variation in Spanish musical lexicon. The panhispanism of the vocabulary of the Leximus dictionary

Despite being specialized, musical vocabulary —like the language we use in our daily lives— also changes when we observe diatopic (dependent on geographic location), diastratic (dependent on social groups), diaphasic (dependent on communicative situations), and contextual variations. *Dinero, plata, pasta, lana,* or *billets* all refer to the same thing but from different places and perspectives. In the same way, musical vocabulary is also a "shared lexicon" that adapts to the needs and interests of everyone. This is mainly because it refers to a part of reality that continuously permeates the lives of all kinds of speakers.

And although we can speak in general terms of "specialized languages," it is advisable to attempt independent approaches to each field (Gutiérrez Rodilla, 1998; Águila Escobar, 2010; Artigas Guillamón, 1998). One could argue that each specialized language has peculiarities and unique characteristics that stem, for instance, from its own epistemological status. In this regard, the levels of specialization in each disciplinary field depend on social sanction and on their degree of integration into the culture of a linguistic community. Thus, the closer a given subject area is to the general knowledge of a society, the more vitality and penetration its vocabulary will likely have in the common language. In this sense, linguists such as Moreno Fernández (1999), Hummel (2009), Rodríguez-Piñero and García Antuña (2010), Águila Escobar (2008: 221), Ávila Martín (2011), and García Antuña (2017) have argued that there are differences between specialized languages regarding their degree of hermeticism and their semantic behavior. According to these authors, the vocabularies of the humanities or the social sciences tend to be more transparent than those of the natural, exact, or experimental sciences, due to their greater proximity to everyday speech.

One of the aspects in which this closeness of musical vocabulary to the general language is most evident is that it is also deeply subject to denominative variation, which is currently viewed as a richness rather than an obstacle to standardization by models such as Socioterminology (Gambier, 1991 and 1993; Gaudín, 1993), the Communicative Theory of Terminology (Cabré, 1999), Sociocognitive Terminology (Temmerman, 2000), or Frame-Based Terminology (Faber et al., 2005).

In this context, describing denominative variants is one of the purposes of the LexiMus dictionary, which will document diatopic variation phenomena such as the alternation between *altavoz* in Peninsular Spanish and *parlante* in Latin America to refer to the "electroacoustic device that transforms electrical current into sound" (Diccionario de la Lengua Española). It will also record cases in which certain regions of Latin America, especially in the Caribbean and the United States, use the Italianism *timpani* to refer to the musical instrument that in Peninsular Spanish is called *timbales*.

This will be one of the main challenges faced by the project LexiMus: Spanish Lexicon and Ontology of Music (ref. PID2022-139589NB-C31), a pioneering initiative in the field of Spanish musicology, developed in collaboration between the Instituto Complutense de Ciencias Musicales, the University of La Rioja, and the University of Salamanca. The LexiMus dictionary aims to describe the denominative varieties of musical vocabulary in Spanish, a language present in more than twenty countries, with enormous cultural differences and over 600 million speakers. The inclusion and labeling of all varieties of Spanish not only reflects the terminological richness of the language but also contributes to making visible different musical traditions and their corresponding manifestations across the Spanish-speaking world.

Bibliographical references:

Águila Escobar, Gonzalo (2005): *Estudio lingüístico y glosario de los términos de la arqueología* (tesis doctoral), Granada, Universidad de Granada.

_____ (2010): «El español y el lenguaje científico-técnico: nuevas consideraciones», *Nuevos caminos del hispanismo. Actas XVI Congreso de la Asociación Internacional de Hispanistas. París, del 9 al 13 de julio de 2007*, 2.

Artigas Guillamón, María del Carmen (1998): «Propuestas metodológicas para el estudio del léxico», Estudios *Románicos,* 10, pp. 9-32.

Ávila Martín, Carmen (2011): «El léxico en los diccionarios jurídicos del español», *Tonos digital. Revista de estudios filológicos*, 21.

Cabré, María Teresa (1999): *La terminología: representación y comunicación. Elementos para una teoría de base cognitiva,* Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Faber, Pamela; Márquez Linares, Carlos y Vega Expósito, Miguel (2005): «Framing Terminology: A Process-Oriented Approach», *Meta*, 50, 4.

Gambier, Yves (1991): «Travail et vocabulaires specialisés: prolegomènes à une socioterminologie», *Meta*, 36, 1, pp. 8-15.

_____ (1993): «Vers une histoire sociale de la terminologie», *Le langage et l'homme*, 28, 4, pp. 233-246.

García Antuña (2017): «Las relaciones conceptuales en terminología», *RILCE: Revista de filología hispánica,* 33, 3, pp. 1359-1384

Gaudin, François (1993): *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles,* Rouen: Publications de l'Université de Rouen.

Gutiérrez Rodilla, Bertha (1998): *La ciencia empieza en la palabra. Análisis e historia del lenguaje científico,* Barcelona, Península.

Hummel, Martin (2009): «Semantics, Terminology, and the Impact of History and Culture on Socioeconomic Terms», *Fachsprache*, 31, 3-4, pp. 109-25.

Moreno Fernández, Francisco (1999): «Lenguas de especialidad y variación lingüística», en Sierra Ayala, L.; Barrueco García, S. y Hernández, E., *Lenguas para fines específicos (VI). Investigación y enseñanza,* Alcalá de Henares, Editorial Universidad de Alcalá, pp. 3-14.

Rodríguez-Piñero, Ana Isabel y García Antuña, María (2010): «Lenguas de especialidad y lenguas para fines específicos, precisiones terminológicas y conceptuales e implicaciones didácticas», en Vera Luján, Agustín y Martínez, Inmaculada (eds.), *El español en contextos específicos: enseñanza e investigación*, 2, pp. 907-932.

Temmerman, Rita (2000): *Towards New Ways of Terminology Description. The Sociocognitive Approach,* Amsterdam, John Benjamins.

Andreu Llòria Pascual
Université Lumière Lyon 2 / Universitat Autònoma de Barcelona

Homonymie et polysémie à l'épreuve des emprunts : le cas de *toque* en moyen français et en français préclassique

Les langues vivantes sont en perpétuelle évolution, notamment à cause des situations de contact linguistique (Guilbert, 1975 : 89). Ces contacts peuvent résulter d'échanges commerciaux, de relations économiques ou politiques, de conflits, de voisinages géographiques ou encore de situations de bilinguisme. L'une des manifestations les plus significatives de ces dynamiques est l'emprunt, défini de manière générale comme « […] une unité ou un trait linguistique qui existait précédemment dans un parler B (dit langue source) et que A ne possédait pas » (Dubois *et al*., 2012, 177 : s.v. *emprunt*).

Si l'emprunt est généralement associé à l'introduction d'une forme nouvelle dans la langue d'accueil, dite langue cible, il ne faut pas oublier que « […] les langues peuvent emprunter aussi des morphèmes, des phonèmes, des accents, des sens, des tours syntaxiques » (Deroy, 1956 : 67). Identifier l'origine des emprunts demeure essentiel pour en évaluer la portée et la prédominance à chaque époque et dans chaque champ lexical (Sablayrolles & Jacquet-Pfau, 2008). Pourtant, cette tâche s'avère particulièrement ardue dans le cas des emprunts dits sémantiques, lesquels sont partiels, dans la mesure où ils modifient uniquement le contenu sémantique d'un vocable [8[1]] déjà existant. La distinction entre évolution sémantique interne et influence externe devient alors délicate, ce qui explique pourquoi les emprunts sémantiques sont rarement signalés dans les ouvrages lexicographiques. Le français préclassique (1550-1650), marqué par plusieurs emprunts de ce type à l'espagnol, illustre bien cette situation.

Durant cette période, le français a intégré de nombreux hispanismes (Treps, 2003 : 140-147 ; Duval, 2007 : 422 ; Siouffi, 2007 : 593), en partie en raison de l'attrait des Français pour la culture, la littérature et la langue espagnoles (Bardon, [1931] 1974 : 6 ; Zuili, 2016 : 15), mais surtout à cause de la puissance politique et culturelle de l'Espagne, notamment dans les domaines maritime et vestimentaire. L'analyse manuelle d'un corpus de dix traductions françaises d'œuvres du Siècle d'or espagnol – chacune publiée entre 1527 et 1633 et non encore numérisée au format texte – nous a permis de constater que plusieurs hispanismes introduits en français à cette époque correspondent à des emprunts sémantiques qui affectent des vocables eux-mêmes empruntés à d'autres langues et introduits en français peu avant l'époque préclassique. Ces hispanismes sémantiques sont rarement décrits comme tels dans les dictionnaires étymologiques : faute d'une analyse exhaustive des différentes acceptions, entendues ici comme les divisions *lexicographiques* du sens d'un vocable, certains sont intégrés à la polysémie du vocable sans mention d'emprunt ; d'autres sont traités comme des homonymes, car issus d'étymons distincts, et font dès lors l'objet d'entrées séparées ; d'autres, enfin, sont totalement absents, tant des ouvrages lexicographiques des xx^{e}-xxi^{e} siècles que des dictionnaires contemporains à l'époque. À ces difficultés s'ajoutent des problèmes d'ambiguïté sémantique, au sens où le contexte d'une occurrence ne fournit pas toujours d'indices suffisants pour déterminer laquelle des interprétations possibles d'une unité linguistique est activée. L'ensemble de ces difficultés complexifie considérablement l'étude du lexique et souligne la nécessité d'adopter une méthodologie rigoureuse.

Pour illustrer au mieux ces enjeux, notre communication orale portera sur l'analyse du substantif *toque*, dont la polysémie et l'étymologie posent plusieurs problèmes lexicographiques. Aujourd'hui, *toque* désigne un « Chapeau sans bord, de forme et de hauteur diverses, qui sert à la parure ou se porte dans l'exercice de certains métiers, comme signe de certaines dignités » (Académie, 1992-2024 : s.v. *toque*), notamment la « Coiffure en toile blanche, plus ou

8[1] Conformément à Polguère (2016), nous entendons par *vocable* l'ensemble des lexies qui sont associées à un même signifiant et qui présentent un lien sémantique évident.

moins haute selon le rang, portée par les cuisiniers, les boulangers, les pâtissiers et leur personnel » (*TLF* : s.v. *toque*). Cependant, ses attestations aux xv^e^ et xvii^e^ siècles, dont quatre occurrences dans la première traduction française de la seconde partie du *Quixote* (1618), révèlent une polysémie qui rend complexes son analyse étymologique et sa représentation dans les ouvrages lexicographiques.

Attesté pour la première fois en moyen français, *toque* apparaît sous deux entrées lexicales dans certains dictionnaires, alors que d'autres n'en proposent qu'une seule. De plus, certains sens ne sont pas répertoriés dans ces ouvrages et, pour d'autres, les exemples fournis semblent renvoyer à un sens différent. Pour clarifier l'étymologie de *toque*, retracer son évolution sémantique et désambiguïser ses sens selon le contexte d'apparition, nous avons croisé l'analyse diachronique d'occurrences textuelles avec l'examen de diverses ressources : dictionnaires d'époque, ouvrages sur le français ancien, dictionnaires étymologiques, bases textuelles et bibliothèques virtuelles. Ces ressources ont été complétées par la lecture d'ouvrages lexicographiques d'autres langues, particulièrement utiles pour l'analyse étymologique. Cette démarche a permis de mieux rendre compte de la polysémie de *toque* et de son évolution historique. Elle a révélé l'existence probable d'un usage en moyen français et en français préclassique pour désigner un 'voile utilisé comme couvre-chef' et a confirmé l'existence de deux homonymes, dont l'un résulterait d'un emprunt à l'espagnol *toca*, lui-même d'origine probablement orientale.

Au-delà de ce cas, l'étude de *toque* illustre l'intérêt de croiser corpus, approche diachronique et ressources lexicographiques pour analyser la polysémie et affiner la représentation des lexies dans les dictionnaires. Notre communication s'inscrit de fait à la croisée des axes « Polysémie et Ambigüité » et « Modélisation du lexique » du colloque international *Lexiques* : elle contribue à la réflexion théorique sur la polysémie, l'homonymie et l'ambiguïté, tout en proposant une modélisation lexicographique plus fine de *toque*.

Bibliographie :

Corpus :

Cervantes, M. de, 1616, *Segunda parte del ingenioso cauallero don Quixote de la Mancha*, Bruxelles, Huberto Antonio.

Rosset, F. (trad.), 1618, *Seconde partie de l'histoire de l'ingenieux, et redoutable cheualier, Don Quichot de la Manche*, Paris, Iacques du Clou et Denis Moreau.

Références générales :

Bardon, M., [1931] 1974, *Don Quichotte en France au xvii^e^ et au xviii^e^ siècle (1605-1815)*, Genève, Slatkine Reprints.

Bandrés, M., 2002, *La moda en la pintura: Velázquez. Usos y costumbres del siglo xvii*, Pampelune, Ediciones Universidad de Navarra.

Bernis, C., 2001, *El traje y los tipos sociales en El Quijote*, Madrid, Ediciones el Viso.

Deroy, L., 1956, *L'emprunt linguistique*, Paris, Presses Universitaires de Liège.

Dubois, J., *et al*., 2012, *Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse.

Duval, F., 2007, « Le Moyen Âge », A. Rey, F. Duval et G. Siouffi (dir.), *Mille ans de langue française*, Paris, Perrin, p. 418-422.

Guilbert, L., 1975, *La créativité lexicale*, Paris, Larousse.

Sablayrolles, J.-F., Jacquet-Pfau, C., 2008, « Les emprunts : du repérage aux analyses. Diversité des objectifs et des traitements », *Neologica : revue internationale de la néologie, 2*, p. 19-38, Paris, Éditions Garnier.

Siouffi, G., 2007, « De la Renaissance à la Révolution », A. Rey (dir.), *Mille ans de langue française*, Paris, Perrin, p. 591-594.

Treps, M., 2003, *Les mots voyageurs : petite histoire du français venu d'ailleurs*, Paris, Seuil.

Walter, H., 2009, *L'aventure des mots français venus d'ailleurs*, Paris, Laffont.

Zuili, M. (éd.), 2016, *Tesoro de las dos lenguas española y francesa. Tresor des deux langues françoise et espagnolle* [étude introductive et édition de Marc Zuili. Préface de Dominique Reyre], Paris, Champion.

Ouvrages lexicographiques et bases textuelles :

Académie, 1694, *Le dictionnaire de l'Académie Françoise*, dedié au Roy, Paris, Coignard.

Académie, 1992-2024, *Dictionnaire de l'Académie Française*, 9e éd [en ligne]. URL : https://www.dictionnaire-academie.fr/

Alcover, A., Moll, B., 2005, *Diccionari català-valencià-balear*, Palma de Majorque, Ed. Moll.

Autoridades = Real Academia Española, 1726-1739, *Diccionario de la lengua castellana* [...], Madrid, Imprenta Francisco del Hierro.

Bloch, O., Wartburg, W., 1989, *Dictionnaire étymologique de la langue française*, Paris, Presses universitaires de France.

CDH = Instituto de Investigación Rafael Lapesa de la Real Academia Española, 2013, *Corpus del nuevo diccionario histórico del español* [en ligne]. URL (version 3.1.) : https://webfrl.rae.es/CNDHE/

Cioranescu, A., 1987, *Los hispanismos en el francés clásico*, Madrid, Imprenta Aguirre.

CORDE = Real Academia Española, s.d., *Corpus diacrónico del español* [base textuelle en ligne]. URL : https://www.rae.es/banco-de-datos/corde

Corneille, T., 1694, *Dictionnaire des arts et des sciences*, Paris, Chez la veuve de Jean Baptiste Coignard.

Corriente, F., 2003, *Diccionario de arabismos y voces afines en iberorromance*, 2e éd., Madrid, Editorial Gredos.

Corriente, F., 2008, *Dictionary of Arabic and Allied Loanwords. Spanish, Portuguese, Catalan, Galician and Kindred Dialects*, Boston, BRILL.

Cortelazzo, M., Zolli, M., 1999, *Dizionario Etimologico della Lingua Italiana*, 2e éd., Bologne, Zanichelli.

Cotgrave, R., 1611, *A Dictionarie of the French and English Tongues*, Londres, Islip.

Covarrubias, S., 1611, *Tesoro de la lengua castellana o española*, Madrid, Luis Sanchez.

Crusca = Accademia della Crusca, 1612, *Vocabolario degli Accademici della Crusca*, 1re éd., Venise, Giouanni Alberti.

Crusca = Accademia della Crusca, 1691, *Vocabolario degli Accademici della Crusca*, 3e éd., Florence, Stamperia dell'Accademia della Crusca.

DCECH = Corominas, J., Pascual, J. A., 1980, *Diccionario crítico etimológico castellano e hispánico*, 3e éd., Madrid, Gredos.

DEAF = Baldinger, K. et al., 1974, *Dictionnaire Étymologique de l'Ancien français*, Québec / Tübingen / Paris, Presses de l'Université de Laval / Niemeyer / Klincksieck. URL (DEAFél - Dictionnaire Étymologique de l'Ancien français électronique) : https://deaf.hadw-bw.de/

DECat = Coromines, J., 1982, *Diccionari etimològic i complementari de la llengua catalana*, Barcelone, Curial edicions catalanes.

DEHfr = Dubois, J., Mitterand, H., Dauzat, A., 2022, *Dictionnaire étymologique et historique du français*, Paris, Larousse.

Devic, M., 1832-1888, *Dictionnaire étymologique des mots français d'origine orientale (arabe, persan, turc, malais)*, Paris, Imprimerie nationale.

DEVRI = Menard, M., 1957-2017, *Le dictionnaire diachronique du breton* [en ligne], sous la direction actuelle d'Herve Le Bihan. URL : https://devri.bzh/

DH = Academia Española, 1933-1936, *Diccionario histórico de la lengua española*, Madrid, Academia Española.

DHLE = Real Academia Española, 2013-, *Diccionario histórico de la lengua española* [en ligne]. URL : https://www.rae.es/dhle/

DHLF = Rey, A., Tomi, M., *et al*., 2012, *Dictionnaire historique de la langue française*, Paris, Le Robert.

DEHfr = Dubois, J., Mitterand, H., Dauzat, A., 2022, *Dictionnaire étymologique et historique du français*, Paris, Larousse.

Diez, F., 1887, *Etymologisches Wörterbuch der romanischen Sprachen*, Bonn, A. Marcus.

DMF = ATILF, *Dictionnaire du Moyen Français*, Nancy, ATILF-CNRS et Université de Lorraine. URL (*DMF* 2023) : http://zeus.atilf.fr/dmf/

Dozy, R., 1843, *Dictionnaire détaillé des noms des vêtements chez les Arabes*, Beirut, Librairie du Liban.

Dozy, R., 1869, *Glossaire des mots espagnols et portugais dérivés de l'arabe*, Amsterdam, Oriental Press.

Dozy, R., [1881] 1927, *Supplément aux Dictionnaires Arabes*, Leiden, Librairie et Imprimerie E.-J. Brill.

Ernault, E., 1895, *Glossaire Moyen-Breton*, Paris, E. Bouillon Éditeur.

Estienne, R., [1539] 1549, *Dictionaire Francoislatin*, Paris, Robert Estienne.

FEW = Wartburg, W., 1922-2002, *Französisches Etymologisches Wörterbuch*, Tübingen / Bonn / Heidelberg / Leipzig / Berlin / Bâle / Nancy, Mohr / Klopp / Winter / Teubne / Zbinden, ATILF-CNRS. URL : https://lecteur-few.atilf.fr/

Frantext = ATILF, 1998-2024, *Base de données textuelles Frantext*, ATILF-CNRS et Université de Lorraine [en ligne]. URL : https://www.frantext.fr/

Furetière, A., 1690, *Dictionnaire universel*, La Haye / Rotterdam, Arnout & Reinier Leers.

Gay, V., [1882-2002] 1928, *Glossaire archéologique du Moyen âge et de la Renaissance*, Paris, Librairie de la Société Bibliographique.

GDLI = Battaglia, S., 1961-2002, *Grande dizionario della lingua italiana*, Turin, Utet.

Godefroy, F., 1881-1902, *Dictionnaire de l'ancienne langue française et de tous ses dialectes du ix^e^ au xv^e^ siècle*, Paris, F. Vieweg.

Guadix, D. de, 2007, *Diccionario de Arabismos : recopilación de algunos nombres arábigos* [1593], Ahumada, I. & Moreno, M. A. (éds), Jaén, Publicaciones de la Universidad de Jaén.

Guilbert, L., Lagagne, R., Niobey, G. (dirs), 1989, *Grand Larousse de la langue française*, 2^e^ éd., Paris, Larousse.

Hornkens, H., 1599, *Recueil de dictionnaires françoys, espagnols et latins*, Bruxelles, Rutger Velpius.

Houaiss, A., 2003, *Diccionário Houaiss da Língua Portuguesa*, Lisbonne, Temas e Debates.

Huguet, E., 1925-1967, *Dictionnaire de la langue française du xvi^e^ siècle*, Paris / Nogent-le-Rotrou, Champion.

La Curne, J. B. de, 1875-1882, *Dictionnaire historique de l'ancien langage françois ou Glossaire de la langue françoise depuis son origine jusqu'au siècle de Louis XIV*, Niort, L. Favre.

LEI = Pfister, M., Schweickard, W., dirs., 1979-, *Lessico Etimologico Italiano, Wiesbaden*, Reichert. URL (*LEI digitale*) : https://lei-digitale.it/

Littré, É., 1873-1874, *Dictionnaire de la langue française*, Paris, L. Hachette.

Mistral, F., 1932, *Le trésor dóu felibrige ou dictionnaire provençal-français*, Paris, Delagrave.

Nebrija, A., 1494, *Vocabulario español-latino*, Salamanca, Juan de Porras.

Nicot, J., 1606, *Thresor de la langue francoyse, tant ancienne que moderne*, Paris, Douceur.

Oudin, A., 1645, *Le trésor des deux langues espagnolle et françoise de Cesar Oudin* [...] *corrigé & reduit en meilleur ordre par Antoine Oudin*, Paris, Antoine de Sommaville et Augustin Courbé.

Oudin, C., 1607, *Tesoro de las dos lenguas española y francesa ; Thresor des deux langues espagnolle et françoise*, Paris, Marc Orry.

Oudin, C., 1616, *Tesoro de las dos lenguas española y francesa ; Thresor des deux langues espagnolle et françoise*, Paris, Marc Orry.

Palazzi, F., Folena, G., 1992, *Dizionario della lingua italiana*, Torino, Loescher Editore.

Pallet, J., 1604, *Diccionario muy copioso de la lengua española y francesa ; Dictionnaire tres ample de la langue Espagnole et Françoise*, Paris, Chez Mathieu Guillemot.

Richelet, P., 1680, *Dictionnaire françois*, Genève, Widerhold.

TLF = Imbs, P., Quemada, B. (dir.), 1971-1994, *Trésor de la langue française : dictionnaire de la langue du xix^e^ et du xx^e^ siècle (1789-1960)*, Paris, CNRS. URL (*TLFi – Trésor de la langue française informatisé*) : http://atilf.atilf.fr/tlf.htm

TLIO = Beltrami, P., Leonardi, L., fond. / Squillacioti, P., (dirs), 1997-, *Tesoro della Lingua Italiana delle Origini*, Florence, CNR. URL : https://tlio.ovi.cnr.it/TLIO/

Vittori, G., 1609, *Tesoro de las tres lenguas Francesa, Italiana y Española. Thresor des trois langues Françoise, Italienne et Espagnolle*, Genève, Par Philippe Albert et Alexandre Pernet.

Zaccaria, E., 1927, *L'elemento iberico nella lingua italiana*, Bologne, L. Cappelli Editore.

Zingarelli, N., 2015, *Vocabolario della lingua italiana*, Bologne, Zanichelli.

Jarvis Looi
Université de Lorraine

« Prochain » ou « suivant » ?
Une confrontation entre théorie introspective et données de corpus

Le débat entre la linguistique de corpus et les méthodes introspectives a fluctué au fil de l'histoire de la linguistique. Chomsky (1979 : 57) a critiqué l'approche des linguistes de corpus des années 1950 comme n'étant que de la « collection de papillons », accumulant des exemples isolés sans réelle analyse systémique. Depuis, la linguistique de corpus a considérablement évolué, mais le clivage entre les deux approches persiste, chacune ayant ses défenseurs et ses détracteurs.

En linguistique française, l'hybridation entre introspection et données de corpus est fréquente, notamment dans les recherches portant sur la catégorie adjectivale dite des « adjectifs de troisième type » (Schnedecker, 2002). Ce terme désigne un ensemble d'adjectifs ne s'insérant pas aisément dans les deux catégories traditionnelles : les adjectifs qualificatifs et relationnels. Ici, les analyses qualitatives de données observées, souvent sans appui statistique, sont régulièrement enrichies d'exemples introspectifs pour combler les lacunes empiriques ou illustrer des usages considérés comme impossibles ou inacceptables. Cette méthode continue d'être employée dans des études récentes (Goes, 2021, 2022 ; Yamamoto, 2020).

Le présent article réexamine les variables proposées par Berthonneau (2002) dans le cadre de l'alternance synonymique entre les adjectifs *prochain* et *suivant*. Il explore la possibilité de reproduire son analyse uniquement à partir de données de corpus et de méthodes statistiques, et vise à démontrer la pertinence de ces outils dans l'étude des adjectifs de troisième type et, plus largement, en linguistique française.

En comparant les fonctions référentielles de *prochain* et de *suivant*, Berthonneau (2002) souligne une asymétrie entre les emplois temporels et spatiaux de *prochain*, ainsi qu'une convergence avec *suivant* lorsqu'il est utilisé avec des noms matériels. Avec des noms comme station (1–3), *prochain* peut être utilisé de manière interchangeable avec *suivant*, indépendamment de sa position, car *la station prochaine*, *la prochaine station* et *la station suivante* renvoient toutes à la même entité. En revanche, *prochain* et *suivant* produisent des références temporelles distinctes dans (4–6), et la position de *prochain* entraîne également des changements sémantiques.

1. Je descends à la station prochaine.
2. Je descends à la prochaine station.
3. Je descends à la station suivante.
4. Je viendrai la semaine prochaine.
5. la prochaine semaine de stage
6. Paul viendra à Noël et Jules la semaine suivante.

Pour expliquer le choix entre *prochain* et *suivant*, Berthonneau (2002) invoque plusieurs variables : le manque de saillance des entités temporelles dans la situation de communication, la séquentialité, la répétitivité et l'individualité des noms temporels, la prévisibilité des entités temporelles, et enfin, l'inclusion topologique des entités temporelles ou matérielles.

La présente étude reprend ces variables issues de l'introspection en les opérationnalisant sous forme de treize variables catégorielles. Une variable supplémentaire (le mode de production) est ajoutée pour tenir compte des différences diamésiques. L'étude s'appuie sur le CEFC, un corpus de référence équilibré de 10 millions de mots en français écrit et oral. Le volet oral, représentant 4 millions de mots issus de quatorze sources, comprend des

enregistrements récents de plus de 2 500 locuteurs francophones adultes de France, de Suisse et de Belgique et couvre divers contextes formels et informels. Le volet écrit (6 millions de mots) comprend des œuvres littéraires, des journaux, des publications scientifiques et des écrits informels comme les SMS, tweets et blogs. Les données ont été extraites avec TXM à partir des formes *prochain.e.s* et *suivant.e.s*. Le jeu de données final comprend 1.309 occurrences de *prochain* et 1.018 de *suivant*. Un échantillon représentatif de 10 % a été sélectionné aléatoirement, en respectant la proportion écrit/oral, pour une annotation manuelle.

Le choix entre *prochain* et *suivant* constituant une variable binaire, un modèle de régression logistique binomiale a été retenu pour répondre à la première question de recherche. Pour répondre à la seconde question portant sur la nature des relations entre les variables et le choix de l'adjectif, nous avons utilisé l'Analyse des Correspondances Multiples (ACM), un outil efficace pour explorer les liens entre des variables catégorielles.

La régression montre que parmi les treize variables proposées par Berthonneau, dix variables sont statistiquement associées au choix de *prochain* ou *suivant*. *Prochain* est préféré en français écrit, avec des noms temporels dans des aspects futurs, conditionnels ou passés, associés à des événements peu prévisibles. *Suivant*, en revanche, se trouve davantage dans le français parlé, avec des noms non temporels, pluriels, séquentiels et individualisés, donc plus saillants, souvent utilisés au présent.

Le critère décisif semble être la clarté du repère temporel. L'ACM révèle une opposition essentielle entre *prochain*, de nature déictique, et *suivant*, de nature anaphorique. *Prochain* est lié aux catégories temporelles de conditionnel et de futur, tandis que *suivant* est lié au présent. Cette divergence nuance les descriptions de Berthonneau, qui voyait *suivant* comme référant à la fois au futur et au présent. En réalité, ce qui détermine le choix, c'est le degré de prévisibilité de l'événement modifié : *prochain* sert à indiquer des événements attendus mais incertains, tandis que *suivant* décrit des faits jugés certains ou avérés.

En dépit de contraintes de place, cette étude montre une convergence significative entre les hypothèses introspectives de Berthonneau (2002) et les données issues des corpus et de l'analyse inférentielle. Mais l'introspection ne permet ni de quantifier les effets ni de hiérarchiser les variables ; elle peut aussi manquer certaines formes de créativité linguistique. Comme le souligne Fillmore (1992), les données de corpus engagent les chercheurs à une responsabilité totale. Par exemple, bien que Berthonneau (2002 : 104) distingue les noms temporels et matériels en fonction de leur capacité à être « considérés comme des lieux », notre étude montre que les noms de lieu sont rarement modifiés par *prochain*, aucun ne ressort comme collocatif significatif.

Notre objectif n'est pas de réfuter Berthonneau, mais de montrer comment les données de corpus et les méthodes statistiques peuvent interagir avec des hypothèses issues de l'introspection. Comme le rappellent McEnery et Hardie (2012), tester une théorie non fondée sur des corpus à l'aide de données empiriques constitue une mise à l'épreuve critique de cette théorie, et non sa simple reconfirmation.

Références bibliographiques

Berthonneau, A.-M. (2002). *Prochain/dernier* et compagnie. Les adjectifs « déictiques » à l'épreuve de l'espace ou comment circuler dans le temps, l'espace, le texte. *Langue Française*, 136, 104-125.

Chomsky, N. (1979). *Language and Responsibility: Based on Conversations with Mitsou Ronat.* Pantheon.

Fillmore, C.J. (1992). "Corpus linguistics" vs. "computer-aided armchair linguistics". In J. Svartvik (Ed.), *Directions in Corpus Linguistics: Proceedings from a 1991 Nobel Symposium on Corpus Linguistics*, (pp. 35-66), Mouton de Gruyter.

Goes, J. (2021). L'adjectif: une partie du discours éminemment syncatégorématique. *Kalbotyra*, 74, 72-87.

Goes, J. (2022). Grand et petit, de simples antonymes ? *Éla. Études de Linguistique Appliquée*, 206, 139-153.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

Schnedecker, C. (Ed.). (2002). L'adjectifs sans qualité(s). *Langue Française*, 136.

Yamamoto, D. (2020). L'adjectivité des épithètes antéposées sale et foutu. *Travaux de Lingustique*, 80, 49-61.

Veronika Lux-Pogodalla / Cristina Diego Pacheco / Nabil Gader
Université de Lorraine, ATILF, CNRS

Réseau lexical musical : une (re)naissance

La base de données LMR (Lexique Musical de la Renaissance), initiée à la fin des années 1990, constituait un travail pionnier visant à identifier et à visualiser le lexique musical en diachronie (XVe-XVIIe siècles) à partir d'exemples issus principalement d'un corpus théorique (Diego Pacheco 2022). La base recense environ 4200 termes en espagnol et 1780 termes en français ; elle leur associe (1) une catégorie grammaticale (2) d'éventuelles variations graphiques et (3) 49000 exemples issus de 9000 extraits de traités savants, traités théoriques, traités de chants, etc. Les termes, choisis par des musicologues, sont des mots simples (ex. *âme*, *abaisser*) ou des syntagmes (ex. *adoucir l'harmonie, à vide*) et la base LMR ne se limite pas à des noms, ce qui est en accord avec les propositions d'experts en terminologie comme M.-C. L'Homme (L'Homme 2004).

LMR sert de point de départ au Réseau lexical musical (RL-Mus), avec les objectifs de pérenniser la ressource, de l'améliorer en corrigeant des erreurs, et de l'enrichir.

Un réseau lexical (Polguère 2025) organise l'information lexicale de façon relationnelle. Il se présente comme un ensemble d'unités lexicales, soit lexèmes (ex. *flûte*) soit locutions (ex. ⌜ *instrument à vent* ⌝), reliées par des liens de différents types :

- liens de copolysémie, entre les unités lexicales d'un vocable polysémique (ex. *accord* 'fait d'accorder un instrument de musique' , *accord* 'groupe de plusieurs notes qui sonnent simultanément')
- liens de dérivation sémantique (ex. entre *sonner* et *sonneur*)
- liens de collocation (ex. entre *sonner* et *cornemuse*)
- liens d'inclusion formelle (ex. entre *flûte* et ⌜ *flûte de pan* ⌝)

Formellement, une conversion est donc opérée entre deux modèles, celui de la base de données LMR et celui d'un réseau lexical. La richesse du modèle cible permet une conversion sans perte d'information mais rend nécessaires des corrections manuelles dans le réseau résultant de la conversion (dit *RL-Mus initial).* Rappelons qu'une conversion autrement complexe avait précédemment été réalisée entre le modèle de WordNet (Fellbaum 1998) et le modèle réseau lexical (Gader et al. 2014), pour créer un Réseau Lexical de l'anglais. Le modèle du LMR étant plus simple que celui de WordNet, le programme informatique de conversion est également moins compliqué.

Le RL-Mus initial est un réseau atypique réduit à un ensemble d'unités isolées : les termes de LMR (ex. *à l'ouvert, à vide, abaissement, accent, accent plaintif, accentuel*). De la même manière le réseau RL-fr initial, né par importation d'une nomenclature d'amorçage (Polguère et Sikora 2013), se réduisait à un ensemble de vocables. Dans ces réseaux initiaux, aucun lien n'existe entre les unités.

Cet article présente (I) les principales corrections nécessaires et surtout (II) les perspectives d'enrichissement des données. Le travail lexicographique entamé sur le RL-Mus initial porte uniquement sur le français ; il progresse non selon l'ordre alphabétique mais par champs sémantiques ; il a commencé avec le champ sémantique des instruments de musique.

I. Corrections nécessaires

Le LMR compte de nombreuses unités polylexicales qualifiées de *syntagmes*. Certaines sont des locutions (ex. ⌜ instrument à vent ⌝, ⌜ à vide ⌝) qui ont le statut de vocables à part entière dans le RL-Mus, d'autres, comme *jouer de son flageolet* sont en fait des collocations et doivent être décrites avec des fonctions lexicales syntagmatiques.

Ainsi, un lien de collocation est tissé entre le nom *flageolet* et l'unité lexicale verbale *jouer 1*, qui est son verbe de réalisation en utilisant la fonction lexicale Real1.

Toutes les unités polylexicales du LMR seront donc requalifiées, principalement comme locutions et comme collocations.

Par ailleurs, les contributeurs du LMR ont parfois exprimé des informations jugées importantes en mésusant le modèle à leur disposition. Ainsi, les variantes graphiques ont été utilisées pour décrire des synonymes : par exemple, dans le LMR, *flûte d'angleterre* « a pour variante graphique *flûte douce* ». Dans le RL-Mus, cette information est exprimée en tissant un lien de synonymie entre les locutions ⌜ *flûte d'angleterre* ⌝ et ⌜ *flûte douce* ⌝.

II. Perspectives d'enrichissement

Le traitement de la polysémie est une tâche conséquente mais incontournable. La polysémie est bien présente dans le vocabulaire musical, et s'observe jusque dans les noms d'instruments de musique. Ainsi, le vocable *chalumeau*, comprend (au moins) deux unités lexicales différentes : *chalumeau 1* dénote un instrument à vent et *chalumeau* 2 dénote un tuyau, assez semblable au chalumeau 1, qui est une partie d'instrument, en particulier, une partie de la cornemuse. Le traitement de la polysémie conduit au tissage de liens de copolysémie dans le RL-Mus et rend nécessaire le réexamen de tous les exemples pour les associer aux différentes unités lexicales, avec la difficulté que certains exemples peuvent être ambigüs.

Par ailleurs, un modèle réticulaire invite évidemment à décrire liens de dérivation sémantique et collocations, en se posant la question de l'identification d'un sous-ensemble des liens intéressant les musicologues afin de circonscrire le travail lexicographique, dans une démarche inspirée de (L'Homme 2004, 2007, 2012). Nos observations dans le RL-Mus pourront être comparées avec celles de (Badenes Tejedo, à paraître) dans un autre domaine de spécialité (céramique).

Comme mentionnés précédemment, les premiers liens de collocation tissés dans le RL-Mus utilisent la fonction lexicale Real1 (verbe de réalisation) tandis que, prévisiblement, les premiers liens de dérivation sémantique utilisés utilisent le lien Gener (générique) : par exemple, les unités lexicales *flûte*, *chalumeau* et *cornemuse* sont mises en relation, par le lien Gener, avec la locution ⌜ *instrument à vent* ⌝. Il sera particulièrement intéressant d'observer le rôle du lien Gener dans la structuration globale du RL-Mus (Polguère 2022).

Un autre grand chantier potentiel est celui de l'ajout de définitions lexicographiques, enrichissement majeur répondant aux vœux des musicologues créateurs du LMR.

En guise de conclusion

La progression par champs sémantiques doit favoriser systématique et homogénéité dans la description lexicographique, en permettant le repérage de la polysémie régulière (ex. un vocable dont l'unité lexicale de base dénote un instrument comprend souvent une unité lexicale dénotant lla personne qui en joue), l'usage raisonné des fonctions lexicales (ex. à tous les noms dénotant des instruments de musique seront associés des verbes de réalisation), voire, l'établissement de patrons de définition.

À terme, un RL-Mus enrichi selon toutes ces pistes serait un réservoir potentiel de visualisations pour différents types d'utilisateurs – visualisations qui pourront, en particulier, prendre la forme de dictionnaires (Lux-Pogodalla et Polguère 2011).

Enfin, comme RL-Mus se distingue à plusieurs titres des réseaux lexicaux qui l'ont précédé, il amènera aussi des questions de recherche inédites :

- le RL-Mus décrit le lexique d'une langue de spécialité (musique) et uniquement celui-ci. Se posera donc la question de l'articulation de ce lexique avec la langue générale non modélisée dans le RL-Mus. Par contraste, en introduisant un vocabulaire spécialisé (chimie) dans le RL-fr, (Mikhel 2024) peut articuler le premier avec le second dans le détail, à l'intérieur de vocables polysémiques (ex. *atome*) ;
- le RL-Mus décrit un lexique ancien ;
- le RL-Mus fait cohabiter des unités lexicales de plusieurs langues dans un même réseau, ce qui appelle des connections interlingues.

Références bibliographiques :

BADENES TEJEDO, Fabio (à paraître). « Formalización de colocaciones terminológicas e implementación en una base de conocimiento ». thèse de doctorat. Université Jaume I

DANCETTE, Jeanne et Marie-Claude L'HOMME (2004). « Building Specialized Dictionaries using Lexical Functions ». In : *Linguistica Atverpiensia*, vol. 3, n° 3. pp.113-32.

DIEGO PACHECO Cristina (2022), « Musique ancienne et lexicologie. Esquisse d'une nouvelle approche disciplinaire », In : *Musique et lexique à la Renaissance. Une approche de la musique ancienne par ses mots*, dir. Cristina Diego Pacheco et Amaya S. Garcia Perez, Paris : Classiques Garnier. pp. 27-79.

GADER, Nabil, Sandrine OLLINGER et Alain POLGUÈRE (2014) « One Lexicon, Two Structures: So What Gives ? ». *Seventh Global Wordnet Conference (GWC2014)*, Jan 2014, Tartu, Estonia. pp.163-171.

INGROSSO, Francesca et Alain POLGUÈRE (2015). « How Terms Meet in Small-World Lexical Networks: The Case of Chemistry Terminology ». *Terminology and Artificial Intelligence (TIA 2015)*, Thierry Poibeau and Pamela Faber, Nov 2015, Granada, Spain. pp.167-171.

FELLBAUM, Christiane ed. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA : The MIT Press.

L'HOMME, Marie-Claude (2004), *La terminologie : principes et techniques*, Montréal : Les presses de l'université de Montréal.

L'HOMME, Marie-Claude (2012). « Using ECL (Explanatory Combinatorial Lexicology) to discover the lexical structure of specialized subject fields ». In : *Words, Meanings and other Interesting Things. A Festschrift In Honour of the 80th Anniversary of Professor Igor Alexandrovic Mel'cuk*, ed. Juri Apresjan, Igor Bugoslavsky, Marie-Claude L'Homme, Jasmina Milicevic, Alain Polguère et Leo Wanner. Jazyki slavjanskoj kultury publishers.

LUX-POGODALLA, Veronika et Alain POLGUÈRE (2011) « Construction of a French Lexical Network: Methodological Issues ». In : *Proceedings of the First International Workshop on Lexical Resources*, WoLeR 2011. An ESSLLI 2011 Workshop. Ljubljana, Slovenia. pp. 54–61.

MIKHEL, Polina (2024), « Using Lexical Networks for terminological modeling: the case of the terminology of chemistry ». In : *Cahiers de lexicologie,* vol. 1, n° 124. Varia, pp. 97-121.

MEL'CUK, Igor (1996) « Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon ». In : *Lexical Functions in Lexicography and Natural Language Processing*, ed. Leo Wanner. Amsterdam/Philadelphia : Benjamins. pp. 37–102.

MEL'CUK, Igor (2006). « Explanatory Combinatorial Dictionary ». In : *Open Problems in Linguistics and Lexicography*, ed. Giandomenico Sica. Monza : Polimetrica. pp. 225-355.

MEL'CUK, Igor and Alain Polguère (2007). *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Champs linguistiques. Brussels : De Boeck & Larcier.

MEL'CUK, Igor et Alain Polguère (2018). « Theory and Practice of Lexicographic Definition ». In : *Journal of Cognitive Science*. vol. 19, n° 4, pp. 417-470.

MEL'CUK, Igor et Alain Polguère (à paraître). *Lexical Functions : the Naked Truth*.

MEL'CUK, Igor, André Clas et Alain Polguère (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot.

POLGUERE, Alain (2025) *Lexicologie et sémantique lexicale. Notions fondamentales*. 4th edition. Montréal : Les Presses de l'Université de Montréal.

POLGUÈRE, Alain (2022). « Fishing out regular polysemy with the Gener (generic) relation ». In : *La polysémie régulière*. ed. Lucie Barque, Pauline Haas. Lexique, Vol. 31, pp. 27-54.

POLGUERE, Alain et Dorora SIKORA (2013), « Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique ». In : *Enseigner le lexique*, ed. Claudine Garcia-Debanc, Caroline Masseron, Christophe Ronveaux. Namur : Presses Universitaires de Namur, pp.35-63.

Marie-Luis Merten
Université de Mannheim

Alignmentmarker in der digitalen Anschlusskommunikation: Korpusanalytische Zugänge zu polyfunktionalen Ressourcen

Digitale Anschlusskommunikation, wie wir ihr auf zahlreichen Plattformen wie Instagram, TikTok oder YouTube in Form von Kommentarspalten begegnen, steht wiederkehrend im Zeichen des sprachlichen und multimodalen Entwurfs von Zu- bzw. Übereinstimmung, Ablehnung, Relativierung usw. (Locher & Bolander 2017; Zappavigna 2017). User:innen positionieren sich hierbei sowohl zu Inititalbeiträgen (YouTube-Videos, TikTok-Videos, Instagram-Posts etc.) als auch etwa zu bereits hinterlassenen Kommentaren (Merten 2025). Kommunikativ gesehen besonders nützlich sind in diesem Zusammenhang sogenannte (Dis-)Alignmentmarker (Heller 2016; Lanwer 2019; grundsätzlich zum Begriff des Alignments Du Bois 2007), also (häufig kurze) lexikalische Ressourcen, denen primär die Funktion zukommt, Konsens und Dissens anzuzeigen – man denke ganz typisch an *ja*, *nein*, *doch*, *stimmt* usw. Solche (Dis-)Alignmentmarker, die (im gesprochen-sprachlichen Kontext) überwiegend äußerungsinitial auftreten, werden geläufiger Weise auch als Antwort- bzw. Responsivpartikeln (Duden 2022: 895f.; Betz 2025) der Klasse der Gesprächspartikeln zugeordnet (siehe etwa Imo 2016: 110ff.). Im Gegensatz zum Terminus der (Responsiv-/Gesprächs-)Partikel hebt der Begriff des Alignmentmarkers die zentrale Funktion solcher lexikalischen Ressourcen in Kontexten des Positionierens hervor: Ihre Nutzung ist sozial-kommunikativ fundiert, ihnen kommt eine soziale Bedeutsamkeit zu. Ob als Gesprächspartikeln oder Alignmentmarker begriffen, die bisherige Forschung zu diesen „kleinen" Ausdrücken ist für die gesprochene Sprache (u.a. Imo 2017; Oloff 2017) und in besonderem Maße für digitale (Schrift-)Zusammenhänge (u.a. König 2021) lediglich überschaubar.

Angesichts dieses Desiderats widmet sich der geplante Vortrag gebräuchlichen Alignmentmarkern der digitalen Anschlusskommunikation. Hierzu soll der Fokus insbesondere auf TikTok-Kommentare gerichtet werden; die Gründe dafür sind vielfältig: (1) Verhältnismäßig kurze TikTok-Kommentare dienen primär dem (Dis-)Alignment. (2) Darüber hinaus finden hier auch solche Ressourcen Verwendung, die durch ein gewisses Maß an Innovativität gekennzeichnet sind. (3) Zudem lassen sich in den entsprechenden Kommentarspalten spannende Fälle des *lexical borrowing* am Beispiel der Alignmentmarker beobachten (ähnlich Balteiro 2018). Denn die User:innen verwenden in ihren vordergründig deutschsprachigen Beiträgen stellenweise Alignmentmarker, die aus dem Englischen stammen. Diese Gesichtspunkte (1) bis (3) lassen sich u.a. durch die folgenden Belege veranschaulichen:

1. **Same**, hatte meine Sprücheseite bis 2020/2021 sogar noch (rosenfinsternis) 😘
2. hahah **Same** bin 24/7 auf TT aber habe nix gesehen
3. **Isso** bei mir auch 😳
4. **Isso** wie kann man das nd sehen
5. **Fr** was macht die da
6. **Fr** warum ist das auf einma so berühmt 😂 aber ey hab das geliebt

Es deutet sich bereits ein grundlegender Unterschied zu „prototypischen" Responsivpartikeln an: Die hervorgehobenen Lexeme dienen nicht dem Beantworten einer im Raum stehenden Frage, vielmehr stehen das Anzeigen geteilter Erfahrungen, Erwartungen sowie Positionen und damit der Entwurf von Zugehörigkeit (zur Kommentierenden-Community) im Vordergrund.

Auf Basis eines TikTok-Kommentarkorpus, das sich aus rund 65.000 Kommentaren (Zeitraum: 2022 bis voraussichtlich 2026) zusammensetzt, werden mithin wiederkehrend genutzte Alignmentmarker – abseits von *ja* – erforscht. Im Zuge einer ersten Analyse haben sich neben *same*, *isso* und *fr* auch bereits *word*, *sure* und *safe* als aufschlussreiche

Ressourcen erwiesen. Mit Blick auf die Schwerpunkte der Tagung sollen vor allem die folgenden Aspekte reflektiert und beleuchtet werden: Die entsprechenden Lexeme zeichnet ein **divergierendes Maß an Polyfunktionalität** (im Sinne einer funktional perspektivierten Polysemie) aus; so treten etwa *fr* und *safe* ebenso als Modalpartikeln/Modaladverbien äußerungsmedial auf (etwa *Die würden **fr** so ein gutes Paar sein*) (Tagungsschwerpunkt Polysemie und Mehrdeutigkeit). Es begegnen also für die zu beleuchtenden Lexeme divergierende Muster der Äußerungsintegration (mit ebenso divergierender Funktion). Hier bietet es sich an, **konstruktionsgrammatische Ansätze** heranzuziehen, um diese Kopplungen von differenten (integrierenden) Formen und Funktionen zu modellieren (Tagungsschwerpunkt Phraseologie/Konstruktionsgrammatik). Wie ein solcher Zugang korpusanalytisch mittels Verschränkung von automatisierten Untersuchungsschritten und manueller Annotation umgesetzt werden kann, soll im Vortrag gezeigt werden. Die fokussierten Alignmentmarker (bzw. konstruktionsgrammatisch gedacht: Alignmentmarker-Konstruktionen) bilden naheliegenderweise ein Netz an Ressourcen (Stumpf & Mollica 2023), die ein funktionales Profil im Sinne eines (sehr) grundlegenden Bedeutungspotenzials teilen, deren Herkunft und soziale Verortung (z.B. speziellere jugendsprachliche Registrierung vs. „lediglich" umgangssprachliche Registrierung) sich allerdings mehr oder weniger unterscheiden (können). Auch dieser Umstand wird im Rahmen des Vortrags eingehender behandelt.

Gewünschtes Format: **Vortrag** (Vortragssprache: Deutsch, Foliensatz: Englisch)

Literatur

Balteiro, Isabel (2018): *Oh wait*: English pragmatic markers in Spanish football chatspeak. In: Journal of Pragmatics 133, 123–133.

Betz, Emma (2025): *ja*, *achja*, *eben*. Responsivpartikeln in der Interaktion. In: Nadine Proske, Thilo Weber, Monika Dannerer & Arnulf Deppermann (Hg.): Gesprochenes Deutsch. Struktur, Variation, Interaktion. Berlin, Boston: De Gruyter, 347–374.

Du Bois, John W. (2007): The stance triangle. In: Robert Englebretson (Hg.): Stancetaking in Discourse. Subjectivity, evaluation, interaction. Amsterdam, Philadelphia: John Benjamins, 139–182.

Duden (2022): Die Grammatik. Struktur und Verwendung der deutschen Sprache. Sätze – Wortgruppen – Wörter. Berlin: Dudenverlag.

Heller, Vivien (2016): „das_s VOLL verARsche hier": Aligment und Disalignment mit jugendsprachlichen Praktiken in der Unterrichtsinteraktion. In: Carmen Spiegel & Daniel Gysin (Hg.): Jugendsprache in Schule, Medien und Alltag. Frankfurt a.M.: Peter Lang, 91–108.

Imo, Wolfgang (2016): Grammatik. Eine Einführung. Stuttgart: J.B. Metzler.

Imo, Wolfgang (2017): Über *nein*. In: Zeitschrift für germanistische Linguistik 45.1, 40–72.

König, Katharina (2021): HM and EHM as discourse markers in German WhatsApp chats. In: Discourse, Context & Media, 100457.

Lanwer, Jens (2019): Alignmentmarker in norddeutscher Alltagssprache (AINA). In: Niederdeutsches Wort 59, 33–75.

Locher, Miriam A. & Brook Bolander (2017): Face work and identity. In: Christian R. Hoffmann & Wolfram Bublitz (Hg.): Pragmatics of Social Media. Berlin, Boston: De Gruyter, 407–434.

Merten, Marie-Luis (2025): Soziale Positionen – soziale Konstruktionen. Stancetaking im Online-Kommentieren. Berlin, Boston: De Gruyter.

Oloff, Florence (2017): Genau als redebeitragsinterne, responsive, sequenzschließende oder sequenzstrukturierende Bestätigungspartikel im Gespräch. In: Hardarik Blühdorn, Arnulf Deppermann, Henrike Helmer & Thomas Spranz-Fogasy (Hg.): Diskursmarker im Deutschen. Reflexionen und Analysen. Göttingen: Verlag für Gesprächsforschung, 207–232.

Stumpf, Sören & Fabio Mollica (2023): Konstruktionsfamilien im Deutschen: Begriffsverständnis, Forschungsstand und Konzeption des Bandes. In: Fabio Mollica & Sören Stumpf (Hg.): Konstruktionsgrammatik IX. Konstruktionsfamilien im Deutschen. Tübingen: Stauffenburg, 9–25.

Zappavigna, Michele (2017): Evaluation. In: Christian Hoffmann & Wolfram Bublitz (Hg.): Pragmatics of Social Media. Berlin, Boston: De Gruyter, 435–458.

Fabio Mollica / Ciro Porcaro
Université de Milan

Lexikalische Kreativität und argumentative Wirkung durch Automobilmarken in der metaphorischen XYZ-Konstruktion

In der Präsentation beschäftigen wir uns mit der metaphorischen XYZ-Konstruktion, bei der eine Entität (Y) im übertragenen Sinne verwendet wird, um eine Eigenschaft zu beschreiben, die einer anderen Entität (X) mit Bezug auf eine bestimmte Referenzdomäne (Z) zugeschrieben wird (z. B. *Das ist der Rolls-Royce unter den Geflügeln;* vgl. u.a. Turner 1991; Brdar/Brdar-Szabó 2007; Brdar 2017). Diese Konstruktion ist zwar kein klassisches phraseologisches Muster, weist jedoch mit Automobilmarken im Y-Slot eine besondere lexikalische Funktion auf. Gerade die wiederkehrende Verwendung von Ergonymen wie *Mercedes*, *Ferrari* oder *Rolls-Royce* verdeutlicht, dass Eigennamen in schematischen Strukturen nicht nur referenziell auftreten, sondern sich als Träger evaluativer und kulturell aufgeladener Bedeutungen etablieren können (vgl. Brdar/Brdar-Szabó 2007; Brdar 2020).

Ziel des Beitrags ist es, zu untersuchen, wie lexikalische Elemente – etwa Ergononyme (insbesondere Automobilmarken) – in die XYZ-Konstruktion integriert werden und welche Rolle sie dabei spielen. Im Mittelpunkt stehen folgende Fragen:

1. Wie verändert sich die Bedeutung von Automobilmarken durch Prozesse der (De-)Semantisierung und Resemantisierung, wenn sie in einer festen schematischen Konstruktion auftreten?
2. Inwiefern lässt sich ihre Verwendung zwischen Konventionalisierung und kreativer Verwendung verorten?
3. Welche Frames werden durch diese Eigennamen aktiviert und wie wirken sie sich auf die Wahrnehmung und Rezeption im Diskurs aus?
4. Welche Rolle spielen sie in argumentativen Strukturen, etwa zur Intensivierung, Persuasion oder ironischen Brechung?

Die Analyse basiert auf Korpusdaten aus dem Deutschen (DWDS-Korpus) und wird durch einen ergänzenden Blick auf das Italienische erweitert. Im Vordergrund steht die Frage, ob Automobilmarken im Italienischen eine ähnliche Rolle in der XYZ-Konstruktion einnehmen und inwieweit sich Parallelen oder Unterschiede in der Verteilung von Ergononymen zwischen den beiden Sprachen abzeichnen.

Der Beitrag versteht die XYZ-Konstruktion mit Automobilmarken als Schnittstelle zwischen Lexikon, Kultur und Diskurs. Er zeigt, dass lexikalische Kreativität nicht nur in lexikalisch gesättigten, nicht-schematischen Einheiten**,** sondern auch in schematischen Mustern/Konstruktionen sichtbar wird, die das Lexikon zugleich stabilisieren und erweitern. Damit leistet die Untersuchung einen Beitrag zu einem besseren Verständnis der Dynamik zwischen lexikalischer Modellierung, kulturellem Wissen und argumentativer Wirkung.

Literatur

Athanasiadou, Angeliki (2023): Paragon and antonomasia: Similar but/and different?. In: Lingua 288, S. 1–11.

Brdar, Mario (2017): Intensification and metonymy in some XYZ constructions: from the Bible to Einstein. In: LaMiCus – Language, Mind, Culture and Society 1, S. 110–134.

Brdar, Mario/Brdar-Szabó, Rita (2007): When Zidane is not simply Zidane, and Bill Gates is not just Bill Gates: Or, Some thoughts on online construction of metaphtonymic meanings of proper names. In: Radden, Günter/Köpcke, Klaus-Michael/Berg, Thomas/Siemund, Peter (Hg.): Aspects of meaning construction. Amsterdam: John Benjamins, S. 125–142.

Brdar-Szabó, Rita/Brdar, Mario (2020): The Bruce Willis of sandwiches: The Y of X is Y of Z construction on its journey towards a paragon model, as one way of achieving intersubjectivity. In: Diadorim 22, 2, S. 277–302.

Turner, Mark (1991): Reading Minds: The Study of English in the Age of Cognitive Science. Princeton: Princeton University Press.

Veale, Tony (2014): The ABCs of XYZs: Creativity and conservativity in humorous epithets. In: Manjaly, Jaison/Indurkhya, Bipin (Hg.): Cognition, experience, and creativity. New Delhi: Orient Blackswan, S. 135–154.

Wałaszewska, Ewa (2023): Non-literal uses of proper names in XYZ constructions: A relevance theory perspective. In: Pragmatics Quarterly Publication of the International Pragmatics Association (IPrA), S. 368–392.

Valentin Neveu
Paris Université

La constitution diachronique d'un lexique de spécialité par l'emprunt : le cas du law french 14e -18e

Le lexique juridique de l'anglais présente cette particularité d'être en grande partie emprunté au français, via ce que la tradition a appelé le law french, langue spécialisée utilisée par les juristes anglais du 14e au 18e siècle.

Il n'est pas rare qu'une langue de spécialité emprunte abondamment, comme l'illustrent les terminologies du domaine médical ou de la biologie, qui empruntent largement au grec ou au latin. Toutefois, l'emprunt direct à une langue vernaculaire, particulièrement au moyen-âge – et surtout par le biais un objet linguistique aussi singulier que le law french – demeure plus exceptionnel et interroge.

En effet, cette langue professionnelle a, par la longévité de son usage, été la principale source lexicale et le creuset du développement de la terminologie juridique anglophone, et par conséquent de la common law et de science juridique anglo-saxonne au sens large. Son influence demeure aujourd'hui sensible dans les lexiques juridiques de nombreux pays mais méconnue, notamment en France. Le détail de ces emprunts et le processus par lequel ce lexique issu du français est passé à l'anglais sont également peu étudiés.

Dans cette communication, nous nous intéresserons particulièrement à la constitution de ce lexique de spécialité dans le law french, en observant de manière diachronique le développement de quelques lexèmes, au regard du développement des concepts juridiques qu'ils portent depuis le moyen âge jusqu'à nos jours. Des termes aussi courant en anglais contemporain que jury, court, arson trouve leur origine dans le law french ou ont été largement influencé par lui. L'intérêt est plus pour les unités lexicales plus techniques : lexique et droit sont ainsi indissociables au sein d'une science humaine reposant autant sur des décisions et des concepts juridiques que sur la capacité à les formuler.

Ainsi, des termes comme tort, mortgage, fee ou encore waiver ont des origines françaises de manière plus ou moins évidente et désignent en droit anglophone des réalités spécifiques, dont l'étude sémantique diachronique peut éclairer l'origine et l'évolution comme concepts juridiques.

Par exemple, mortgage, ou waiver désignent des actes courants des droits de la propriété et de la finance anglophones, alors même que le français n'a pas conserver ces lexèmes dans sa propre terminologie. De même, tort, dont la parenté avec le substantif français tort ou la locution faire du tort peut sembler évidente à des francophones, connait par exemple un développement important, qui donne naissance à un domaine entier des droits anglo-américain : la tort law ; là où le droit français confine les emplois du lexème à quelques locutions qui relèvent autant, sinon davantage, du langage courant que spécialisé.

Ces disparités nous pousserons donc à adopter également une approche plus contrastive pour comparer le devenir de ces unités lexicales en français, en examinant leurs évolutions parallèles dans les deux langues, éclairant indirectement l'histoire du développement de deux conceptions du droit, d'apparence différente, mais pourtant apparentées. Cette approche permettra d'inscrire notre étude en diachronie longue, tout en l'ancrant à la matérialité réelle du monde juridique ancien et contemporain.

Bibliographie succincte :

Baker, John Hamilton, (1990) Manual of law french. London, Routledge, 232 p. https://doi.org/10.4324/9781315249780

Gémar, Jean-Claude. (1990). Les fondements du langage du droit comme langue de spécialité. Du sens et de la forme du texte juridique. Revue générale de droit, 21(4), 717–738.https://doi.org/10.7202/1058214ar

Laske, Caroline. (2017). Legal History Meets Lexical Semantics : Consideration : The Origin of Term and Concept. Ghent, Belgium: Ghent University. Faculty of Law and Criminology. http://hdl.handle.net/1854/LU-8537493

Laske, Caroline, (2018) Le law french, un idiome protégeant les privilèges du monde des juristes anglais entre 1250 et 1731,Corela, HS-26

Löfstedt, Leena. (2014) Notes on the beginnings of law french, Romance Philology, Fall 2014, Vol. 68, No. 2, pp. 285-337 ; Brepols; University of California Press

Sandrine Ollinger / Benjamin Husson
Université de Lorraine, ATILF, CNRS

Mon voisin de réseau : quand l'annotation en sens invoque la structure des lexiques

L'annotation de corpus textuels en sens lexicaux, connue sous le terme de *désambiguïsation lexicale* en traitement automatique des langues, est une tâche réputée difficile. Lorsque l'on demande à des annotateurs humains d'associer les occurrences d'un lemme en contexte à ses différents sens décrits dans une ressource lexicale telle qu'un dictionnaire de langue, l'accord entre les annotateurs ne dépasse pas les 80 % (Navigli, 2009 ; Veronis, 2001).

Du côté du traitement automatique, les systèmes évalués au début du siècle atteignaient une précision inférieure à 70 %, tant pour l'anglais (Navigli, 2009), que pour le français (Segond, 2000). Au cours des vingt-cinq dernières années, les systèmes ont évolué et le plafond de verre des 80 % de précision a même été brisé (Bevilacqua et al., 2021). Dans ce panorama, les modèles qui atteignent les plus hauts scores sont basés sur une architecture neuronale, couplée à l'exploitation de ressources lexicales et, plus précisément, de graphes lexicaux, tels WordNet (Fellbaum, 1998) ou BabelNet (Navigli & Ponzetto, 2010). Si les éléments de description lexicale exploités varient d'un modèle à l'autre, il est intéressant de noter que la moitié des modèles présentés par (Bevilacqua et al., 2021) utilisent les informations relationnelles contenues dans les graphes.

Dans le même temps, la modélisation du lexique sous forme de réseaux a montré son intérêt dans le cadre de l'analyse sémantique outillée. Pour le français, on soulignera notamment les travaux sur la polysémie et la délimitation d'espaces sémantiques menés par Gaume (2004) et Ploux & Victorri (1998). Ces travaux exploitent les propriétés mathématiques de réseaux lexicaux qui ont la particularité d'être des graphes dits « petit monde » (Newman, 2003; Watts & Strogatz, 1998). Ils ont notamment permis la création du *Dictionnaire électronique des synonymes* (DES) du laboratoire CRISCO (Chardon, 2020) et l'analyse psycholinguistique de la production verbale chez des enfants en cours d'acquisition langagière et des malades atteints d'Alzheimer (Duvignau et al., 2024).

En parallèle de ces travaux, qui exploitent des dictionnaires de langue préexistants pour construire les graphes lexicaux (Loiseau et al., 2011), plusieurs ressources lexicales du français ont vu le jour, directement structurées en graphes. Polguère (2009, 2014) pose ainsi les bases de la famille des Systèmes Lexicaux, objets lexicographiques dans la lignée des travaux de la Lexicographie Explicative et Combinatoire (Mel'čuk et al., 1995), dont le Réseau Lexical du français [RL-fr] (Lux-Pogodalla & Polguère, 2011) est aujourd'hui l'instance la plus avancée. Lafourcade & Joubert (2008, 2010), de leur côté, proposent la construction d'un vaste réseau de relations lexicales ou fonctionnelles par l'intermédiaire d'un jeu en ligne.

Au-delà de leur exploitation interne et de l'émergence de nouvelles connaissances sémantiques, on peut se demander si de telles ressources peuvent être projetées sur des corpus pour les enrichir et quels seraient les atouts d'un tel enrichissement. Dans le cadre de cette contribution, nous nous intéressons à la réalisation manuelle de la tâche d'annotation en sens lexicaux.

Dans ce cadre, nous proposons d'interroger la notion de *voisinage lexical* disponible dans les graphes lexicaux. Nous chercherons à établir s'il peut être utile d'intégrer ce voisinage au matériel fourni aux annotateurs lors de campagnes d'annotation. Une annotation préalable en voisins lexicaux peut-elle faciliter la tâche ou risque-t-elle d'introduire des hésitations supplémentaires ?

Du côté des sens lexicaux, nous avons choisi les inventaires de sens correspondant aux lemmes nominaux costume, crime et serviette, sélectionnés dans le RL-fr pour leurs structures polysémiques contrastées. La notion de *voisinage lexical* s'entend alors comme l'ensemble des sommets du RL-fr reliés à chacun des sens de ces trois lemmes à l'aide de fonctions lexicales (Mel'čuk & Polguère, 2021).

Du côté des contextes à annoter, nous nous tournons vers la partie en langue française du corpus Open Super-large Crawled Aggregated coRpus Community Edition [OSCAR CE], une version ouverte et versionnée du corpus OSCAR (Ortiz Suárez et al., 2019, 2020). Ce corpus, dérivé de l'initiative Common Crawl (https://commoncrawl.org), rassemble des données textuelles issues du Web. Pour notre étude, une centaine de contextes est sélectionnée aléatoirement pour chaque lemme. Chaque contexte contient une occurrence unique du lemme concerné. Les contextes sont automatiquement annotés en *voisins lexicaux* et rendus disponibles sur la plateforme INCEpTION (Klie et al., 2018). Chaque annotation comporte trois informations distinctes, extraites du RL-fr : (a) de quel(s) sens de notre inventaire l'occurrence annotée est-elle un voisin ? (b) la relation qui relie sens et voisin est-elle paradigmatique ou syntagmatique ? (c) quel est le poids sémantique de cette relation [0, 1 ou 2] ?

La Figure 1 montre un exemple d'une telle annotation. On peut y voir un énoncé contenant le lemme serviette dans lequel deux occurrences de voisins lexicaux ont été identifiées. Dans les deux cas, le sens voisin est serviette I.1 *[Pièce de tissu destiné à s'essuyer après la toilette.*] et la relation qui relie sens et voisin est syntagmatique. Le poids sémantique de la relation est plus important dans le cas de *séchez* que de *propre*.

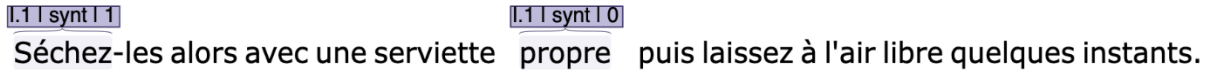


Figure 1 Annotation en voisins lexicaux de serviette

Dans un tel cas, l'annotation en voisins lexicaux oriente le choix de l'annotation en sens lexicaux de *serviette* vers le sens serviette I.1. Dans la Figure 2, en revanche, l'annotation en voisins lexicaux n'apporte aucun indice, puisque l'adjectif vêtu se trouve dans le voisinage de trois des quatre sens de costume décrits dans le RL-fr.

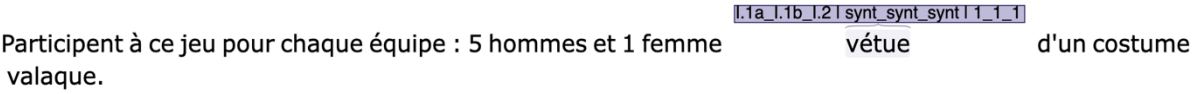


Figure 2 Annotation en voisins lexicaux de COSTUME

Après une présentation détaillée de la méthodologie mise en œuvre, nous mènerons une analyse qualitative des résultats et chercherons à établir une typologie des cas de figure rencontrés.

Références bibliographiques

Bevilacqua M, Pasini T, Raganato A, et al. (2021) Recent Trends in Word Sense Disambiguation: A Survey. In: *Twenty-Ninth International Joint Conference on Artificial Intelligence*, 9 August 2021, pp. 4330–4338. Available at: https://www.ijcai.org/proceedings/2021/593 (accessed 20 February 2024).

Chardon L (2020) L'espace sémantique du 'Dictionnaire électronique des synonymes' (DES) et les méthodes de regroupement de sens : l'exemple de 'sec'. *Syntaxe et Sémantique* 1(21): 87.

Duvignau K, Fossard M, Gaume B, et al. (2024) Approches et flexibilités sémantiques dans la construction et 'déconstruction' dynamique de la signification. *Linguagem em (Dis)curso* 7. Universidade do Sul de Santa Catarina: 371–387.

Fellbaum C (1998) *WordNet : an electronic lexical database*. Language, Speech and Communication.

Gaume B (2004) Balades aléatoires dans les Petits Mondes Lexicaux. *Information interaction intelligence* 4(2): 39–96.

Klie J-C, Bugert M, Boullosa B, et al. (2018) The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, Santa Fe, USA, June 2018, pp. 5–9. Association for Computational Linguistics. Available at: http://tubiblio.ulb.tu-darmstadt.de/106270/.

Lafourcade M and Joubert A (2008) Détermination des sens d'usage dans un réseau lexical construit grâce à un jeu en ligne. In: *TALN'08 : Traitement Automatique des Langues Naturelles*, 2008, p. 189. Available at: https://hal-lirmm.ccsd.cnrs.fr/lirmm-00358853 (accessed 26 May 2025).

Lafourcade M and Joubert A (2010) Computing Trees of Named Word Usages from a Crowdsourced Lexical Network. *Investigationes Linguisticae* 21: 39–56.

Loiseau S, Gréa P and Magué J-P (2011) Dictionnaires, théorie des graphes et structures lexicales. *Revue de Sémantique et Pragmatique* (27): 51.

Lux-Pogodalla V and Polguère A (2011) Construction of a French Lexical Network: Methodological Issues. In: *Proceedings of the First InternationalWorkshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*, Ljubljana, Slovenia, August 2011, pp. 54–61. Available at: https://hal.science/hal-00686467 (accessed 4 June 2025).

Mel'čuk I and Polguère A (2021) Les fonctions lexicales dernier cri. In: Marengo S (ed.) *La Théorie Sens-Texte. Concepts-Clés et Applications*. Dixit Grammatica. Paris: L'Harmattan, pp. 75–155.

Mel'čuk I, Clas A and Polguère A (1995) *Introduction à La Lexicologie Explicative et Combinatoire*. Paris/Louvain-la-Neuve: Duculot.

Navigli R (2009) Word sense disambiguation: A survey. *ACM Comput. Surv.* 41(2): 10:1-10:69.

Navigli R and Ponzetto SP (2010) BabelNet: Building a Very Large Multilingual Semantic Network. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, PA, USA, 2010, pp. 216–225. ACL '10. Association for Computational Linguistics. Available at: http://dl.acm.org/citation.cfm?id=1858681.1858704 (accessed 15 May 2014).

Newman MEJ (2003) The structure and function of complex networks. *SIAM REVIEW* 45: 167–256.

Ortiz Suárez PJ, Sagot B and Romary L (2019) Asynchronous pipelines for processing huge corpora on medium to low resource infrastructures. In: *Proceedings of the Workshop on Challenges in the Management of Large Corpora (CMLC-7) 2019. Cardiff, 22nd July 2019*, 4 July 2019, pp. 9–16. Leibniz-Institut für Deutsche Sprache. Available at: https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/9021 (accessed 4 June 2025).

Ortiz Suárez PJ, Romary L and Sagot B (2020) A Monolingual Approach to Contextualized Word Embeddings for Mid-Resource Languages. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (eds D Jurafsky, J Chai, N Schluter, et al.), Online, July 2020, pp. 1703–1714. Association for Computational Linguistics. Available at: https://aclanthology.org/2020.acl-main.156/ (accessed 4 June 2025).

Ploux S and Victorri B (1998) Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. In: *Traitement Automatique des Langues, Traitement automatique des langues*, 1998, pp. 161–182. Available at: http://halshs.archives-ouvertes.fr/halshs-00009433 (accessed 19 November 2013).

Polguère A (2009) Lexical Systems: Graph Models of Natural Language Lexicons. *Language Resources and Evaluation* 43(1). Springer: 41–55.

Polguère A (2014) From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography* 27(4): 396–418.

Segond F (2000) Framework and Results for French. *Computers and the Humanities* 34(1/2). Springer: 49–60.

Veronis J (2001) Sense tagging: does It make sense? In: *Corpus Linguistics'2001 Conference*, 2001.

Watts DJ and Strogatz SH (1998) Collective dynamics of /`small-world/' networks. *Nature* 393(6684): 440–442.

Aurélie Picton / Pascaline Duri
Université de Genève / Université Lyon 2

« Mon loup m'a attaquée… » : Regard sur l'appropriation et la mise en discours de la terminologie médicale dans les récits de patients atteints de maladies chroniques

La notion d'*expertise patient* désigne l'ensemble des connaissances que les personnes atteintes de maladies chroniques acquièrent sur leur maladie au fil des années, par l'expérience vécue, le rapport au système de santé et les interactions avec les équipes médicales (par ex. Carretier et coll., 2009 ; Flora, 2015 ; Derradji, 2017). Si ce concept est notamment questionné par l'anthropologie et les sciences de la communication (par exemple Akrich & Rabeharisoa, 2012 ; Bellander & Landqvist, 2020 ; Flora, 2020), il suscite depuis plusieurs années également l'intérêt de la linguistique et de la terminologie, en particulier à travers l'analyse des échanges dans les forums de patients (par ex. Delavigne, 2020 ; Humbert-Droz et Picton, 2024). Dans ces travaux, la question de l'appropriation des termes occupe une place centrale (par ex. Vogh & Courbon, 2018 ; Delavigne et coll., 2022). En effet, comprendre la manière dont les termes sont utilisés et compris par les patients alimente non seulement les réflexions sur leur communication avec le personnel soignant et les patients, mais aussi la légitimité de leur expertise en tant que patients (par ex. Flora, 2015 ; Garric & Herbland, 2020).

Ce phénomène s'inscrit dans une transformation plus large du paradigme de l'expertise médicale : depuis plusieurs décennies, le savoir « profane » des patients se juxtapose au savoir « scientifique » des soignants pour constituer une dynamique inédite de circulation des connaissances (Akrich & Rabeharisoa, 2012). L'émergence de la notion de *patient expert* transforme ainsi, dans une certaine mesure, la répartition des rôles dans le contexte médical, passant d'un modèle principalement paternaliste à un modèle plus ouvertement participatif (par ex. Pomey et coll., 2015, 2024). Dans ce contexte, l'interaction entre ces deux formes d'expertise favorise la création d'espaces partagés entre les patients où les termes médicaux circulent, entre appropriation d'un savoir scientifique et vécu quotidien (Carretier et coll., 2009).

Ces dernières années, certains courants, tels que la médecine narrative (conceptualisée par Charon, 2006, 2015), soulignent l'importance d'autres formes de discours où les patients partagent leur expertise et décrivent leur expérience : l'écriture dite réflexive, notamment dans le cade de récits personnels, permet de saisir leur singularité et leur vécu dans la prise en charge personnalisée de la maladie, et offre un moyen d'enrichir les pratiques médicales. Les récits de patients deviennent ainsi aujourd'hui des pratiques pleinement reconnues, tant pour leur impact sur la prise en charge que sur la qualité de la relation soignant-soigné (par ex. Delavigne, 2017 ; Galichon, 2023 ; Charon, 2015).

En terminologie, cette perspective ouvre un champ de recherche original et nouveau : l'analyse de la terminologie des maladies chroniques telle qu'elle apparaît dans ces récits personnels de patients. Plus précisément, dans cette communication, nous nous interrogeons sur la manière dont les patients s'approprient la terminologie médicale, la manière dont elle est narrée, mise en discours, modifiée ou reformulée dans leurs récits personnels, et sur ce que ces usages terminologiques peuvent révéler de leur rapport à la maladie.

Pour ce faire, nous proposons d'observer ces récits personnels à travers une analyse en corpus et par le biais d'une approche linguistique outillée. Le corpus constitué pour cette étude est composé de récits de patientes et patients atteints de maladies chroniques, organisé en deux parties :

- *Récits_maladies_chroniques* : un premier sous-corpus (70 000 occurrences) est composé de 62 récits de patients atteints de différentes maladies chroniques (VIH, lupus, sclérose en plaques, etc.), issus de différents sites d'associations (tels que l'Association Boulimie Anorexie (ABA), l'Association suisse du Lupus, Draw Your Fight), et de centres de recherche (comme le Centre Hépato-Biliaire Paul Brousse).

- *Récits_diabète* : un second corpus (107 000 occurrences) est centré sur le recueil de 62 récits de patients atteints de diabète, extraits du site de la Fédération française des diabétiques (FFD).

Après avoir extrait les termes présents dans notre corpus à l'aide de Sketch Engine (Kilgarriff et coll., 2014), nous proposons de nous appuyer sur les fonctionnalités du logiciel d'analyse textométrique TXM (Heiden et coll., 2010), afin d'observer les spécificités de nos sous-corpus et explorer les fonctionnements des termes sélectionnés dans ce type de discours (en particulier concordances et collocations). Nous proposons plus particulièrement d'examiner la terminologie médicale utilisée par les patients atteints de maladies chroniques sous deux angles :

- Une analyse de termes partagés par les patients, quelle que soit la maladie chronique concernée (sous-corpus Récits_maladies_chroniques). Nous cherchons ainsi à décrire la mise en discours des termes « transversaux » (ou transdisciplinaires (Humbert-Droz et Picton, 2024)) utilisés par les patients, en particulier les termes relatifs à la description des symptômes (par exemple *douleur*, *fatigue*) et à celle des traitements (*scan, IRM*, etc.).

- Une analyse de termes propres à une maladie, ici le diabète, (sous-corpus Récits_diabète), tels que *insuline, glycémie* ou *injection*. Nous cherchons ici à observer la manière dont une terminologie plus spécifique et moins transversale, et donc aux modalités d'appropriation potentiellement différente, est mise en discours.

En croisant les perspectives de la terminologie et la narration, nous faisons l'hypothèse que la construction narrative du récit d'un vécu avec une maladie chronique s'appuie sur une mise en discours particulière des termes médicaux. Cette étude vise ainsi à alimenter les réflexions sur les modalités d'appropriation et de partage des termes médicaux par les patients dans un contexte discursif spécifique, celui des récits et de l'écriture réflexive, ainsi que sur les dynamiques de l'expertise patient que ce contexte révèle, entre savoir scientifique et expérience vécue.

Nos résultats pourront également contribuer à enrichir les pistes d'analyse et de réflexion relatives à l'amélioration de la relation entre patients et professionnels de santé, en posant les jalons de ce qui pourrait, à terme, être nommé une terminologie narrative.

Références citées

Akrich, M., & Rabeharisoa, V. (2012). Patients et experts profanes : La dynamique des savoirs participatifs dans le domaine de la santé. *Anthropologie & Santé*, 5. https://journals.openedition.org/anthropologiesante/1342

Carretier J., Delavigne V. et Fervers B. (2009). Du langage expert au langage patient: Vers une prise en compte des préférences des patients dans la démarche informationnelle entre les professionnels de santé et les patients. *Sciences-Croisées*, 6.

Charon, R. (2006). *Narrative Medicine. Honoring the Stories of Illness*. Oxford University Press.

Charon, R. 2006. *Narrative Medicine: Honoring the Stories of Illness*, New York, Oxford University Press, trad. fr. R. Charon, *Médecine narrative. Rendre hommage aux histoires de maladie*, Paris, Sipayat, 2015.

Delavigne, V. (2017). L'écriture pour les patients : une écriture singulière ? *Pratiques*, (173-174). https://doi.org/10.4000/pratiques.3268

Flora L. (2015). Du patient « passif » au patient expert. In Déchamp-Le Roux C. et Rafael F. (Eds), *Santé mentale: Guérison et rétablissement*. Montrouge: John Libbey Eurotext, 109-119.

Flora, L. (2020). L'expertise du patient formateur et la reconnaissance du savoir expérientiel dans l'éducation thérapeutique. *Médecine/Sciences*, 36(11), 1023-1025.
https://www.medecinesciences.org/en/articles/medsci/full_html/2020/12/medsci200206s/medsci200206s.html

Galichon, I. (2023). Les récits multiples de la médecine narrative. *Essais*, 21.

Garric, N., & Herbland, A. (2020). [Référence complète à compléter selon source précise, non fournie ici].

Heiden, S., Magué, J.-P., & Pincemin, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In S. Bolasco, I. Chiari, & L. Giuliano (Eds.). In: Actes des 10e Journées internationales d'Analyse statistique des Données Textuelles/International Conference on the Statistical Analysis of Textual Data.

Humbert-Droz, L., & Picton, A. (2024). Révéler l'expertise partagée par les patientes atteintes de diabète et d'endométriose: une analyse exploratoire de forums médicaux. In: Actes des 17es Journées internationales d'Analyse statistique des Données Textuelles/International Conference on the Statistical Analysis of Textual Data.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine : Ten years on. *Lexicography*, 1(1), 7-36.

Pomey, M.-P., Flora, L., Karazivan, P., Dumez, V., Lebel, P., Vanier, M.-C., Débarges, B., Clavel, N., & Jouet, E. (2015). The Montreal Model : The Challenges of a Partnership Relationship between Patients and Healthcare Professionals. *Santé Publique*, HS, 41-50.

Pomey, M.-P., Schaad, B., Lasserre-Moutet, A., Böhme, P., & Jackson, M. (2024). Towards a New Integrated Model for Taking Into Account the Experiential Knowledge of People With Chronic Diseases, Integrating Mediation, Therapeutic Education and Partnership : The Expanded Chronic Care Patient–Professional Partnership Model. *Health Expectations*, 27(5), e70054. https://doi.org/10.1111/hex.70054

Vogh, E., & Courbon, C. (2018). « "For me, it was very genetic" : usage de termes issus du domaine de la génétique dans le discours en ligne de diabétiques anglophones », *in* Christophe Roche (éd.), *Actes du colloque TOTh 2016*, Chambéry, Éditions de l'Université de Savoie, 2018, 83-103.

Nesrine Raissi
Universités de Montpellier III et Bourgogne Europe

Dire le traumatisme par les témoins du camp de Rivesaltes

La mémoire des témoins de Rivesaltes est une mémoire traumatique telle qu'elle a été défini par Cyrulnick (2012b). Cette mémoire contient toute la souffrance du passé liée à la guerre, à l'exil et à l'internement. Lorsque le témoin prend la parole pour raconter ce passé pesant, il mobilise le langage afin de « restaurer une mémoire abîmée dans son encodage même pour en faire une mémoire vive, active, dicible et possiblement curative » (Mayaffre et Ben Hamed, 2014, p. 2-3). Dès lors, quel est le lexique mobilisé par les témoins historiques pour dire le traumatisme et justifier cette mémoire souffrante ? Cette contribution cherche à montrer comment le témoin historique exprime, justifie et construit sa mémoire traumatique à travers les mots.

Une série d'entretiens, réalisés et transcrits dans le cadre de l'équipex Matrice, porté par Denis Peschanski, a permis de recueillir entre 2007 et 2014 les témoignages de personnes internées au camp de Rivesaltes. On propose d'analyser un corpus de vingt-huit témoignages de réfugiés espagnols et de déportés juifs, détenus entre 1941 et 1942. Ce corpus, de 883 095 mots, sera traité à l'aide du logiciel textométrique TXM. En l'interrogera en se basant notamment sur les travaux de Plantin, Doury, et Traverso (2000), de Doucet (2014) ou encore de Novakova et Tutin (2019) qui se sont interrogés sur la mise en mots de l'expérience subjective (trauma, mémoire et émotion) et la manière dont le langage permet de dire et de partager l'émotion liée à cette expérience. Nous nous appuierons également sur les travaux de Mayaffre et de Ben Hamed (2014) et de Mayaffre et *al* (2016) qui, par le biais de la textométrie, abordent les traces lexicales engendrées par le traumatisme de la Shoah.

Une première approche exploratoire a montré que, dans les témoignages de Rivesaltes, la mémoire traumatique se construit et se légitime par l'emploi du **lexique émotionnel** (« Et euh, on a bien souffert après, quand nous nous sommes trouvés dans le camp et que nous n'avions presque rien, ni souliers, ni habits convenables pour l'hiver. On a souffert du froid terriblement », Hilda, 2009), souvent appuyé par le **lexique de l'intensification** (« Mais nous, on croyait que c'était des Allemands, alors on avait terriblement peur alors le, le, le passeur nous a dit … et nous, on était encore sur le côté français, on est resté », Norbert, 2008), par le **lexique de l'impuissance et de la victimisation** (« On ne pouvait pas se laver, on n'avait pas d'eau à boire, on était obligé de se tenir entre les repas qui n'étaient pas des repas, qui nous laissaient très très affamés », Amira, 2012), par le **lexique de la mémoire** qui souligne la persistance du souvenir (« Bon, quand on est parti de « Rivasaltas » avec le, le train, on est arrivé à « Rivasaltas » … Enfin, je l'ai déjà dit ça, mais il y a un passage que je voudrais dire et que je n'oublie jamais et c'est terrible ! C'est un passage quand on était là, il y avait la mer, il n'y avait rien, il n'y avait pas de latrine, il n'y avait rien », George, 2014) ou encore par **le lexique des contrastes** (« Vous savez, on était bien quand on était jeune, mon père …. On avait … Bah, on n'était pas des riches à millions, mais on était bien, on était heureux et, du jour au lendemain, on a tout perdu. On était sur les routes, quand même. On était réfugié, on était sur les routes, sous les bombardements », Josefa, 2008).

Cette catégorisation sera élargie et approfondie par une étude minutieuse du lexique relevé. Cette démarche montrera comment le lexique devient un outil de transmission et de justification de la mémoire traumatique dans le témoignage historique.

Références bibliographiques

Cyrulnik, B. (2012b). Mémoire traumatique et résilience, D. Peschanski (dir.), *Mémoire et mémorialisation*, (Paris : Hermann), p.133-143

Doucet, P. (2014). Lire l'indicible comme expérience aliénante dans *Insensatez* de Horacio Castellanos Moya. *ILCEA* [En ligne], 54. URL : http://journals.openedition.org/ilcea/19271 ; DOI : https://doi.org/10.4000/ilcea.19271

Mayaffre, D., & Ben Hamed. M. (2014). Récits de mort et souvenir traumatique. Trames et traces lexicales des témoignages sur la Shoah. *Argumentation et Analyse du Discours*, Varia, 13, p.1-20. ff10.4000/aad.1836ff. ffhal-01253245f

Pincemin, B., Mayaffre, D., Heiden, S., & Weyl, P. (2016). Génétique mémorielle. Shoah, mémoire et ADT. Mayaffre, D., Poudat, C., Vanni, L., Magri, V., & Follette, P. *JADT 2016 - Statistical Analysis of Textual Data*, volume 2, p.495- 506, Nice, France : Presses de FacImprimeur. hal-01361988

Novakova, I. et Tutin, A. (2009). *Le lexique des émotions*. Ellug : Université Stendhal.

Plantin, C., Doury M., et Traverso, V. (2000). *Les émotions dans les interactions*. Presses universitaires de Lyon.

Elise Schürgers
Université de Liège

Quand l'autonymie vient à la formule. L'activité métadiscursive autour de la locution « fake news » et sa portée argumentative

La présente proposition s'inscrit prioritairement dans la sous-thématique « Argumentation » renseignée par l'appel, mais entend également contribuer, quoiqu'indirectement, aux questionnements liés aux notions de « Polysémie et Ambiguïté ».

Notre cas d'étude est celui de la circulation – dans la presse généraliste francophone contemporaine, en France et en Belgique – de l'emprunt lexical « *fake news* ». Dans le courant de l'année 2017, ce dernier a connu un fonctionnement discursif faisant de lui une formule[9[1]] (Krieg-Planque 2006 ; 2009). Pétris d'une ambiguïté qui facilite ses reprises et ses appropriations, la locution peine à s'accrocher à un sens circonscrit ou stable, que l'on pourrait considérer comme partagé par les locuteurs francophones d'une société donnée, et ce malgré les initiatives lexicographiques à son endroit (*cf.* notamment l'avis de la CELF). Cela étant, le discours de presse atteste du caractère majoritairement incontournable de la formule : à l'abord de plusieurs thématiques et face à une série d'événements d'actualité, l'anglicisme est devenu une unité lexicale dont on se passe difficilement, tout comme il est une désignation, un descripteur, dont on sait qu'il sera reconnu et retiendra l'attention. Malgré cela, « *fake news* » est un mot que l'on rechigne souvent à employer sans guillemets tant les locuteurs y décèlent, entre autres, différentes lignes d'opinion et d'instrumentalisation. Notre recherche s'emploie ainsi conjointement à documenter et analyser les formes de réflexivité langagière qui marquent la trajectoire de la formule « *fake news* » et à interroger les enjeux argumentatifs dont cette activité métadiscursive est porteuse.

Les passages que nous traitons, appartenant à un vaste corpus d'énoncés attestés[10[2]], sont ceux qui font de la formule un « *objet signe* » (Authier-Revuz 2003, p. 31) ; ces usages s'arrêtent sur la locution, signalent qu'elle fait problème et nous renseignent sur les enjeux que les locuteurs font peser sur ses emplois. L'activité métalangagière observée prend la forme d'énoncés définitoires, de commentaires métadiscursifs opacifiants portant sur des faits de langue ou de discours (Authier-Revuz 1995), de modalisations autonymiques (*ibid.*) et d'îlots textuels (*ibid.* 1996). En privilégiant l'entrée par les formes du métalangage ordinaire (« l'épilinguistique » d'Antoine Culioli), nous emboîtons le pas à l'approche adoptée par Alice Krieg-Planque dans sa recherche sur la formule *purification ethnique* (2003), mais aussi à de nombreux travaux intéressés par ce que l'on nomme tantôt les « affrontements sur les mots » (Micheli 2011), tantôt les « polémiques métalangagières » (Garric 2015), tantôt les « débats lexico-sémantiques » (Calabrese 2018). Se prêtant particulièrement bien à la combinaison d'approches discursives et argumentatives autour d'unités lexicales clés (Micheli 2013), cet objet de recherche permet d'approcher le processus au fil duquel les locuteurs élaborent autant un objet de discussion publique et son sens social que les moyens linguistiques qui leur permettront de s'exprimer à son sujet.

Le résultat de ce travail s'organise principalement autour de deux grands « registres de problématisation » (Krieg-Planque 2003), relevant de ce que Jacqueline Authier-Revuz (1996 ; 2003) a appelé la « non-coïncidence du discours à lui-même », créé par dialogisme interdiscursif (*e.g.* : *X, selon le mot de…*), et « la non-coïncidence entre les mots et les choses » (*e.g.* : *ce que l'on pourrait peut-être appeler X*). Nous montrerons que cette activité métadiscursive est

9[1] Nous illustrerons par une analyse des formes langagières ce constat déjà posé par des chercheurs en sciences de l'information et de la communication (Allard-Huver 2017 ; Froissart 2020 ; Giry 2021).

10[2] Constitué à partir des productions de six quotidiens (*Le Figaro, Libération, Le Monde* ; *La Libre Belgique, Le Soir, La Dernière Heure*), le corpus comptabilise un total de 1019 articles pour 1520 occurrences de la locution, lesquels s'échelonnent sur les seize mois qui, répartis entre octobre 2004 et mai 2019, attestent de la forte médiatisation de la locution « *fake news* ». Nous analysons donc des énoncés en contexte, relevant du commentaire métadiscursif sur la locution en question.

productrice d'une tension entre deux orientations : l'une œuvre à la trivialisation de l'anglicisme comme élément de langage polémique, l'autre travaille à en faire un objet de débat, soulevant les implications liées à l'acte définitoire. Entre « danger » et « hochet », « *fake news* » se révèle être un mot dont l'emploi est conjointement lesté de significations politiques alarmantes et décrédibilisé comme un artifice communicationnel sans grande profondeur. Sur cette base, nous mettrons d'abord en évidence que la propriété rétorsive[11[3]] de « *fake news* » est un moteur de cette activité métadiscursive et de son efficace argumentative, et, ensuite, que la préoccupation des commentateurs médiatiques de la formule est moins sa signification en langue que la détermination des réalités qui tomberaient ou non sous son égide, ainsi que ce que représente le geste de définition lui-même. En cela, les locuteurs activent la fonction métapragmatique du langage (Silverstein 1993) : en s'arrêtant sur ce que « *fake news* » pourrait ou devrait désigner et sur ce que les critères de détermination de son empan dénotatif impliquent dans un certain cadre (législatif, surtout), les commentateurs de la formule supposent en effet que le fonctionnement sémantico-référentiel de la locution a des effets intersubjectifs en contexte.

Si l'activité métadiscursive autour de l'unité lexicale « *fake news* » est le terrain de logiques de distinctions, au cours desquelles l'énonciateur « pose en s'opposant » (Amossy 2015), elle est également un moyen de s'arroger un espace de valeurs commun avec l'énonciataire et de désamorcer par avance les soupçons qui pourraient peser sur le point de vue défendu et de construire l'image d'un locuteur ancré dans le discours social, au fait des conflictualités (langagières) qui agitent la thématique sur laquelle il se prononce, et ce afin de recentrer ces tensions autour de ce qui, à ses yeux, doit être le sujet des préoccupations collectives.

Références citées :

Allard-Huver François, 2017, « Fake news », dans *Publictionnaire*, 2017. URL : https://publictionnaire.huma-num.fr/notice/fake-news/.

Amossy Ruth, 2015, « La polémicité des définitions en discours. La formule "délégitimation d'Israël dans la presse française" », dans Johannes Angermuller et Gilles Philippe (éds.), *Analyse du discours et dispositifs d'énonciation. Autour des travaux de Dominique Maingueneau*, Limoges, Lambert-Lucas, pp. 248-259.

_____2018, *Une formule dans la guerre des mots. « La délégitimation d'Israël »*, Paris, Classiques Garnier.

Authier-revuz Jacqueline, 1992, « Repères dans le champ du discours rapporté », *L'information grammaticale*, n° 55, p. 91-115.

_____1995, *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris, Larousse, coll. « Sciences du langage ».

_____1996, « Remarques sur la catégorie de "l'îlot textuel" », *Cahiers du français contemporain*, vol. 3, Hétérogénéités en discours, 1996.*op. cit.*

_____2003, « Le fait autonymique : langage, langue, discours. Quelques repères », dans Marianne Doury et Sandrine Reboul-Touré (éds.), *Parler des mots : Le fait autonymique en discours*, Paris, Presses Sorbonne Nouvelle, coll. « Sciences du langage », p. 67-96.

Calabrese Laura, 2018, « Faut-il dire migrant ou réfugié ? Débat lexico-sémantique autour d'un problème public », *Langages*, vol. 210, n° 2, 2018.

Froissart Pascal, 2020, « Vrai et faux en contexte médiatique. "Fake news", infox, rumeur », dans Salomé Kintz (éd.), *Décoder les fausses nouvelles et construire son information avec la bibliothèque*, Villeurbanne, Presses de l'enssib, coll. « La Boîte à outils », pp. 22-37.

Garric Nathalie, 2015 « Polémique métalangagière sur une (dé)nomination : le terme "confrontation" », *Langue française*, vol. 188, n° 4, pp. 61-76.

11[3] À la suite de Bernard Dupriez et de Marc Angenot, Ruth Amossy décrit la rétorsion comme une procédure rhétorique qui « retourne contre l'adversaire les armes qu'il a choisi d'employer afin de le contre-attaquer efficacement » (Amossy 2018, p. 169).

Giry Julien, « Les fake news comme concept de sciences sociales », *Questions de communication*, n° 1, 2021, pp. 371-394.

Krig-Planque Alice, 2009, *La notion de « formule » en analyse du discours. Cadre théorique et méthodologique*, Paris, Presses Universitaires de Franche-Comté, 2009.

_____2003, *« Purification ethnique ». Une formule et son histoire*, Paris, CNRS éditions, coll. « CNRS Communication ».

Micheli Raphaël, 2011, « Quand l'affrontement porte sur les mots en tant que mots : polémique et réflexivité langagière », *Semen. Revue de sémio-linguistique des textes et discours*, n° 31, p. 97-112.

_____2013, « Les querelles de mots dans le discours politique : modèle d'analyse et étude de cas à partir d'une polémique sur le mot "rigueur" » [en ligne], *Argumentation et Analyse du Discours*, vol. 10, 2013, URL : https://journals.openedition.org/aad/1446#tocto1n2.

Silverstein Michael, 1993, « Metapragmatic discourse and metapragmatic function », dans John A. Lucy (éd.), *Reflexive Language: Reported Speech and Metapragmatics*, Cambridge, Cambridge University Press.

Athina Sioupi
Université de Thessalonique

Towards a Verb-Lexicon Model: Greek and German in a contrastive constructional perspective

Introduction. The paper presents ongoing research on modeling the intricate interplay between morphology, syntax, and semantics and on advancing contrastive studies across languages focusing on verbs. To this end, we employ constructicons for German and Greek - two languages representing typologically and genetically distant branches of Indo-European that differ substantially in their morphological and syntactic profiles. Our research seeks to model German and Greek verb valence patterns or constructions as form-meaning pairings organized in language-specific structured repositories of reconstructions.

Objectives and scope. The aim of this work is to present an analysis of comparative concepts for specific verb classes, such as *change of state verbs* (CoS) (i.e., *open*, *close*), *break verbs* (i.e., *break*, *crack*) (Levin, 2013, Fillmore, 1970 inter alia) in Greek in contrast to German. In this respect, a key objective is the development of the Greek constructicon, and the enrichment of the German one where needed. Particular attention is paid to cross-linguistic correspondences through the alignment of the two constructicons, enabling systematic comparison of argument structure realizations and constructional patterns.

Previous work. Grounded in the theoretical framework of Construction Grammar (Goldberg 1995, 2006, 2013; Croft 2022), constructicons, that is, structured repositories of constructions, originated with the English Constructicon (Filmore 2008; Fillmore, Lee-Goldman & Rhomieux 2012; Lee-Goldman & Petruck 2018), which established the foundations for systematic constructional description. Since then, constructicons have been created for several other languages, including Swedish (Lyngfelt et al. 2018), Japanese (Ohara 2013), Russian (Endresen et al. 2022), Brazilian Portuguese (Laviola et al 2017), and German (*https://framenet-constructicon.hhu.de,* last access 14.10.2025), each reflecting language-specific constructional inventories and usage patterns. Approaches to constructicon creation involve the identification and clustering of constructions and their interrelations. Constructicon alignment (Lyngfelt et al 2018) has been pursued mainly through FrameNet-based semantic anchoring and feature-based correspondences across languages.

Methodology. We have adopted a combined top-down and bottom-up methodology for constructicon creation and population. From a top-down perspective, we begin with theoretical descriptions of verb valences and constructions identified in previous work on Greek and in the German Constructicon, selecting representative verbs and verb patterns that instantiate core constructional types (e.g. transitive, reflexive, causative, change of state or aspectually marked patterns). From a bottom-up perspective, we conduct a corpus-based study aimed at identifying recurrent form–meaning pairings and constructional variants attested in usage.

Construction annotation is performed manually using a dedicated annotation platform and on top of dependency-annotated corpora of German and Greek, which allow the extraction of syntactic and morphological information relevant to constructional patterning, including argument structure, valence, and morphosyntactic alternations.

Analysis. Greek and German differ in how morphological marking is realized. In Greek, anticausative Change of State verbs (CoS) appear with two morphologically distinct forms: (i) an active morphology, e.g. *lerose* 'dirty' (2a) and/or (ii) a non-active morphology, e.g. *lerothike* (2b), while the causative CoS is expressed by a one morphological form, *lerono* 'dirty' (2c):

(2) a. To trapezomantilo lerose.[12[1]] (anticausative)

the tablecloth.$_{NOM}$ dirtied.$_{3SG}$

b. To trapezomantilo lerothike.

the tablecloth.$_{NOM}$ dirtied.$_{NACT.3SG}$ (anticausative)

c. O Jannis lerose to trapezomantilo.

the-John dirtied.$_{3SG}$ the tablecloth.$_{NOM}$

'John dirtied the tablecloth'. (causative)

(Alexiadou et al. 64 examples 4a,b).

In German, verbs participating in the causative alternation, can appear in two main morphological patterns: (i) as morphologically unmarked, known as *unmarked anticausatives* (3a,b), and (ii) as (reflexively) *marked anticausatives*, appearing with the reflexive pronoun 'sich' (4a,b) (Markantonatou, 1992, Alexiadou, Anagostopoulou & Schäfer, 2015, inter alia).

(3) a. Hans/Die Sonne schmilzt die Butter. (causative)

John/the sun melts the butter.$_{ACC}$

b. Die Butter schmilzt. (anticausative)

The butter.$_{NOM}$ melts.$_{3SG}$

(4) a. Hans öffnet die Tür. (causative)

John.$_{NOM}$ opens.$_{3SG}$ the door

b. Die Tür öffnet sich. (marked anticausative)

The door.$_{NOM}$ opens.$_{3SG}$ REFL (Schäfer 2008, Haspelmath 2019, inter alia)

Verbs expressing a change-of-state take agents and/or causers as external arguments in their causative use, but may also occur as lacking an external syntactic and semantic argument in their anticausative (or inchoative) use (Levin & Rappaport Hovav, 1998, Reinhart, 2002, Schäfer, 2008, Alexiadou et al., 2015: 64) as shown in (5a,b):

(5) a. John/the wind opens the door. (transitive / causative CoS)

b. The door opens. (intransitive CoS / inchoativ / anticausative)

The traditional syntactic analysis for transitive / causative CoS and intransitive CoS verbs (intrCoS) / inchoatives / anticausatives / unaccusatives is illustrated in (6a,b), respectively:

(6) a.[TP … [VP $DP_{AGENT/CAUSER}$ [verb DP_{THEME}]]] (transitive / causative CoS)

b. [TP … [VP verb DP_{THEME}]] (intransitive CoS / inchoative / anticausative)

(e.g. Levin & Rappaport Hovav, 1995, Reinhart 2002)

12[1] Abbreviations: 3SG= 3rd singular, NACT=non-active, NOM=nominative, ACC= accusative.

The corresponding semantic analysis can be captured as in (7a,b), respectively:

(7) a.[x CAUSE [BECOME [y <STATE>]]] (transitive / causative CoS)

b. [BECOME [y <STATE>]] (intransitive CoS / inchoative / anticausative)

(e.g. Dowty 1979, Levin & Rappaport Hovav, 1995)

The analysis serves as the base for building the repository of Greek verbs and enriching an already existing database of German verbs; the two databases model the morpho-syntactic and semantic properties of verbs (Herbst, 1992, 2011, Herbst & Schüller, 2008) and provide mappings to their semantic representations.

Conclusion. We aim to present our ongoing research on modeling the lexicon of German and Greek via the development and alignment of subset of constructicons in the two languages. We will present the methodological approach adopted, the analyses conducted, and the results obtained so far. Beyond its immediate descriptive value, our work is intended to advance comparative, construction-based linguistic research by providing empirically grounded, interoperable representations of the interaction between morphology, syntax, and semantics across languages. At the same time, the outcomes of our work have significant potential for Natural Language Processing (NLP), supporting applications that require linguistically informed modeling of grammatical constructions - including, but not limited to, language learning and educational technologies, multilingual parsing, and cross-linguistic syntactic processing. Ultimately, the project seeks to bridge theoretical linguistics and computational implementation, fostering synergy between linguistic constructive studies and practical NLP applications.

Selected References

Alexiadou, A., Anagnostopoulou E. & Schäfer, F. (2015). *External arguments in transitivity alternations. A layering approach*. Oxford University Press.

Boas, H. C. (2014b). Zur Architektur einer konstruktionsbasierten Grammatik des Deutschen. In A. Lasch & A. Ziem (Eds.), *Grammatik als Netzwerk von Konstruktionen. Sprachwissen im Fokus der Konstruktionsgrammatik* (= Sprache und Wissen 15) (pp. 37-63). Mouton de Gruyter.

Croft, W. (2022). *Morphosyntax: Constructions of the World's Languages*. Cambridge: Cambridge University Press.

Fillmore, C. (1970). The grammar of hitting and breaking. In R. A. Jacobs, & P. A. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 120-133). Ginn.

Fillmore, C. J., Lee-Goldman, R. & Rhomieux, R. (2012). The FrameNet Constructicon. In: Boas, Hans C. & Ivan A. Sag (eds.), Sign-Based Construction Grammar. Stanford: CSLI Publications, 309–372.

Laviola, A., Lage, L., Marçâo, N. D., Tavares, T., Almeida, V., Matos, E. E., & Torrent, T. (2017). The Brazilian Portuguese Constructicon: Modeling Constructional Inheritance, Frame Evocation and Constraints in FrameNet Brasil. In AAAI Spring Symposia.

Lee-Goldman, R. & Petruck, M. R. L. (2018). The FrameNet constructicon in action. In: Lyngfelt, Benjamin et al. (eds.) 2018b, Constructicography: Constructicon development across languages. Amsterdam: John Benjamins, 19–39.

Lyngfelt, B., Borin, L., Ohara, K., & Torrent, T. T. (Eds.). (2018). *Constructicography: Constructicon development across languages* (Vol. 22). John Benjamins Publishing Company.

Ohara, K.H. (2013). Toward Constructicon Building forJapanese in Japanese FrameNet. Special Issue on Frame Semantics and Its Technological Applications. Veredas, vol. 17, no.1, pp.11-27.

Agnès Tutin / Francis Grossmann
Université Grenoble Alpes

Phrases préfabriquées interrogatives à fonction métalinguistique hétéroadressées : quelles constructions ?

La fonction métalinguistique peut être définie comme renvoyant aux différentes manifestations de cette propriété qu'est la réflexivité des langues naturelles, qu'il s'agisse d'expressions portant sur l'emploi de la langue (métalinguistiques, dans un sens restreint), sur l'acte de dire ou le « dire en train de se faire » (méta-énonciatives) ou bien sur le discours lui-même (métadiscursives). Dans les interactions orales et écrites, on relève de nombreuses phrases relevant de cette fonction, souvent intégrées dans la catégorie des « marqueurs de discours » propositionnels (Andersen 1998 ; Dostie 2004; Rouanne 2016) comme *on va dire, on l'a dit …* Certaines d'entre elles s'en distinguent toutefois car elles n'ont pas un fonctionnement parenthétique (*tu sais quoi ? qu'est-ce que ça veut dire ?*). Un premier inventaire de ces phrases préfabriquées à fonction métalinguistique montre qu'elles sont assez nombreuses, puisqu'on en dénombre à peu près 200 pour le français, intégrant majoritairement le verbe *dire*.

Un ensemble de fonctions « classiques » associé à ces différentes expressions a déjà été bien été identifié pour le français (par exemple, Steuckardt 2005, 2018). Parmi ces éléments, on pourra citer : la reformulation (*je veux dire*), la précaution oratoire (*si j'ose dire*) à, la recherche de formulation (*comment dirais-je ?, comment dire ?*), l'insistance (*je veux dire*), l'auto-correction (*qu'est-ce que je raconte*). D'autres fonctions propres à la structure des interactions ont toutefois été moins étudiées : l'annonce/amorce (*il faut que je te dise; tu sais quoi ?*), la demande de précision (*c'est-à-dire ?*), la confirmation de la compréhension auprès de l'allocutaire (*tu vois ce que je veux dire ?*), la reprise du discours d'autrui (*tu disais ?*). Nous souhaitons dans cette communication nous centrer en particulier sur ce deuxième type d'expressions, hétéroadressées, qui interpellent l'allocutaire.

Nous nous restreindrons aux expressions qui constituent des « phrases préfabriquées », et qui présenteront les propriétés suivantes : ce sont des énoncés autonomes (dont la tête n'est pas régie par un autre élément), verbaux (*tu disais ?*) ou averbaux (*c'est-à-dire ?*) , dont la réalisation lexicale est relativement contrainte par la fonction langagière (*tu sais quoi ?* et non *tu sais ce que je veux dire ?*). Les phrases préfabriquées qui nous intéressent n'intègrent donc pas des expressions comme *autrement dit, soi-disant, à vrai dire* qui ne peuvent pas être des énoncés autonomes. Cette communication sera centrée sur les phrases préfabriquées métalinguistiques spécifiquement hétéro-adressées, en particulier celles qui apparaissent avec la modalité interrogative, et ont la structure <*tu/vous* V [X] > : *tu me suis ? tu vois ce que je veux dire ? tu comprends ? tu sais quoi/tu connais la dernière ? tu sais quoi ?* Les exemples suivants illustrent quelques-uns des emplois de ces phrases.

1. [L1] alors je te propose en parfum
   [L2] vas -y
   [L1] euh abricot romarin / c'est de la tisane hein
   [L2] ouais ouais
   [L1] vanille
   [L2] vanille / c'est du thé nan
   [L1] ouais / euh infusion verveine menthe
   [L2] tu sais quoi
   [L2] je crois que je vais prendre abricot romarin pour goûter / ça me tente bien
   (Corpus oral TCOF, conversation entre amies)

2. [A] K alors il y a quelque chose qu' il faut que je vous explique les enseignements sont regroupés dans une unité
[EE] hum hum
[A] il y a plusieurs enseignements dans l'unité / et en fait pour avoir les E C T S il faut prendre l'ensemble de l'unité
[EE] d'accord
[A] vous voyez ce que je veux dire
[A] bon et les unités vous pouvez les repérer à travers des numéros
[EE] hum hum
(Corpus FLEURON, Conversation entre étudiants)

Notre objectif est à la fois théorique et descriptif. Il vise d'une part à proposer un modèle rendant compte du schéma constructionnel des phrases préfabriquées concernées en intégrant à la fois des paramètres sémantiques, syntaxiques et interactionnels. Cette approche intégrative (voir par exemple, Dobrovol'skij, 2020 ; Mellado Blanco, 2021) nous parait particulièrement adaptée à notre objet d'études, parce qu'elle permet de modéliser les phrases qui nous intéressent sous forme de phrasèmes constructionnels, définis comme des associations structures-sens comportant des éléments fixes et des éléments variables. Pour ce faire, nous observerons les données à partir d'un corpus oral authentique diversifié, intégrant de nombreux corpus contemporains du français (Corpus ESLO2, ORFEO-CEFC et MPF), totalisant 5,4 millions de tokens et mis en ligne sur le Lexicoscope (Kraif, 2019). Cela nous permettra d'identifier les spécificités de fonction liées aux genres conversationnels. On peut ainsi faire l'hypothèse que le discours didactique ou les réunions de travail feront davantage usage de formules de « vérification de la compréhension » (*vous me suivez ?*), alors que la conversation informelle comportera davantage de phrases d'amorce/annonce (*il faut que je te dise. tu sais quoi ?*). Ces relevés seront complétés par l'observation d'un corpus de dialogues romanesques récents (>= 1980), qui permettra de déterminer si l' « oral représenté » prend bien en compte la dimension métalinguistique dans les dialogues.

Références bibliographiques :

Andersen, H. L. (2007). Marqueurs discursifs propositionnels. *Langue française*, *154*(2), 13-28.

Dobrovol'skij, D. O. (2020). German Constructional Phrasemes with Deictic Elements in Bilingual Dictionaries. In P. Cotta Ramusino and F. Mollica (Eds.). *Contrastive Phraseology: Languages and Cultures in Comparison* (pp- 115-128). Newcastle upon Tyne: Cambridge Scholars Publishing.

Guryev, A. (2021). « Car je suis un… tu sais quoi ? Rappeur »: Étude pragma-syntaxique de la séquence 'Tu sais quoi ?' *Lexique*, 29, 97-115

Kraif, O. (2019). Explorer la combinatoire lexico-syntaxique des mots et expressions avec le LEXICOSCOPE. *Langue française*, *203*(3), 67-82.

Mellado Blanco, C. (ed.) (2021). *Productive Patterns in Phraseology and Construction Grammar. A Multilingual Approach*. Berlín: de Gruyter.

Rouanne, L., & Anscombre, J. C. (2016). *Histoires de dire: petit glossaire des marqueurs formés sur le verbe DIRE*. G. Kleiber (Ed.). Berne: Peter Lang.

Steuckardt A. (2005) : « Les marqueurs formés sur dire », in A. Steuckardt et A. Niklas-Salminen (dir.), Les marqueurs de glose. Aix en Provence, Publications de l'Université de Provence, coll. « Langues et langage », p. 51-65.

Steuckardt, A. (2018). Les marqueurs de reformulation formés sur dire: exploration outillée. *Langages*, *212*(4), 17-34.

Tutin, A. (2019). Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI. *Cahiers de Lexicologie*, 114, 63-91.

Daniel Vojtek
Université P. J. Šafárik de Košice

Élaboration d'un corpus de terminologie grammaticale multilingue (slovaque-français-anglais-latin) : quelle méthode ?

La nomenclature grammaticale souffre d'une variabilité et d'une instabilité marquantes, malgré l'objectif des grammairiens et des terminologues de concilier qualité du discours grammatical et efficacité didactique. La raison en est que les choix terminologiques des grammairiens sont en partie guidés par la théorie linguistique qui préside à l'élaboration de la grammaire et la terminologie grammaticale connaît des évolutions, des tendances et parfois aussi des antagonismes (Leduc-Adine, 1980 ; Buyssens, 1985). L'étude de la terminologie grammaticale (TG ci-après) sur le plan comparé dans l'espace linguistique centre-européen connaît un historique solide et au cours des dernières années elle a évolué en touchant plusieurs aspects, en passant des analyses multiparamétriques, notamment lexico-sémantiques (Vojtek, 2020) vers un sondage discursif profond, recouvrant une période représentative (Puchovská, 2021). Dans ce sens, la comparaison systémique interlinguale de la TG ouvre toujours des pistes nouvelles. La présente communication représente un continuum des travaux réalisés et vise à étudier les modalités et les possibilités méthodologiques de l'élaboration d'un corpus terminologique multilingue, tout en restant inspirée et fortement fondée sur la nécessité et les besoins objectifs tels que décrits dans les travaux récents (Swiggers, 2010 ; Křečková, 2012). Les langues incluses sont le slovaque, le français, l'anglais et le latin. Objectif demeurant de nature tantôt terminologique tantôt traductologique et équivalentiel, il n'est en aucun cas moins lexicographique (Swiggers, 1989). L'apport prépondérant du projet en cours se lit dans l'approche aussi synchronique que diachronique, qui retrace l'équivalence terminologique des langues étudiées tissée sur l'origine latine d'une grande partie des termes de grammaire.

La méthode de l'élaboration du corpus terminologique devrait se baser sur trois points fondamentaux :

a. Le choix des ressources lexicographiques actuelles (datant d'après 2000), tant unilingues que bilingues ou multilingues (morphologie et syntaxe),

b. Traitement multiparamétrique du matériel lexical au sein d'un fichier commun incluant les quatre langues touchées,

c. Adoption et application du seul critère équivalentiel qui est la correspondance conceptuelle découlant des caractéristiques définitionnelles de chaque terme.

S'encadrant fortement dans la perspective métalexicographique ou métaterminographique, le projet en cours représente une tentative de systémiser l'ensemble de la terminologie grammaticale sur un plan quadrilingue. Dans ce sens, le choix du cadre méthodologique et sa description restent un préalable nécessaire à toute étape ultérieure des travaux lexicographiques et/ou ceux analysant et décrivant les mécanismes régissant tout le procédé de l'élaboration d'un dictionnaire sui generis.

Références bibliographiques :

Buyssens, É. (1985). Terminologie grammaticale et linguistique fonctionnelle. *Enjeux : Revue de didactique du français*, 71, 86-97.

Křečková, V. (2012). *Terminologie & linguistique : construction des ensembles terminologiques bilingues (slovaque-français)*. Fakulta humanitných vied, Univerzita Mateja Bela.

Leduc-Adine, J.-P. (1980). De la terminologie grammaticale : quelques problèmes théoriques et pratiques. *Langue française*, 47, 6-24. https://doi.org/10.3406/lfr.1980.5058

Puchovská, Z. (2021). Le discours grammatical contextualisé du français de conception slovaque. Analyse de la période entre 1993 et 2018. Dans Z. Puchovská (Éd.), *Le discours grammatical contextualisé slovaque dans la description du français. (1918-2018)* (pp. 143-233). Éditions des archives contemporaines.

Swiggers, P. (1989). Les dictionnaires de terminologie linguistique : bibliographie systématique. *L'Information grammaticale*, 40, 12-15.

Swiggers, P. (2010). Terminologie, terminographie et métalangage linguistiques : quelques réflexions et propositions. *Revue roumaine de linguistique*, *55*(3), 209-222.

Vojtek, D. (2020). *Preklad gramatickej terminológie (na materiáli francúzštiny a slovenčiny)*. Prešovská univerzita v Prešove.



POSTERS

Lian Chen / Huy-Linh Dao / Alexander Delaporte / Damien Nouvel
Université d'Orléans / INALCO / CRLAO-CNRS / INALCO

Étude et détection automatique des néologismes en chinois et en vietnamien (2015–2025)

Cette recherche s'inscrit dans le cadre du projet NeoLex, consacré à l'étude et à la détection automatique des néologismes en chinois et en vietnamien entre 2015 et 2025. L'objectif principal est de recenser, analyser et modéliser les unités lexicales nouvelles apparues dans ces deux langues, en articulant les approches linguistique, computationnelle et lexicographique. Ce travail répond à une double lacune : d'une part, l'absence de ressources comparables entre langues asiatiques dans les projets internationaux de néologie, souvent centrés sur l'Europe (par exemple, le projet ANR EDyLex de Sagot & Nouvel, 2009, Néoveille, ou encore ENEOLI – *European Network on Lexical Innovation*, 2024-2027, https://eneoli.eu) ; d'autre part, le manque d'outils numériques adaptés au traitement automatique du chinois et du vietnamien, deux langues typologiquement différentes mais historiquement connectées. L'objectif de cette étude n'est pas de proposer une analyse contrastive approfondie entre le chinois et le vietnamien, mais d'évaluer la robustesse d'une même méthode de détection automatique de néologismes appliquée à deux langues typologiquement différentes.

Sur le plan théorique, en chinois comme en vietnamien, les néologismes (Sablayrolles, Boulanger, Cabré, Sagot & Nouvel) sont définis comme des unités nouvelles par leur forme, leur sens ou leur usage. Ils peuvent être décrits selon plusieurs processus communs : emprunts, calques sémantiques, compositions internes et créations socio-numériques. En chinois, les xīn cí (新词) résultent principalement de la composition lexicale (如 网红 *wanghong*, « influenceur ») et de calques liés aux innovations technologiques (脑机, « interface cerveau-ordinateur »). En vietnamien, les từ mới reflètent davantage une hybridation linguistique, incluant des emprunts phonétiques à l'anglais (thần kiu), des calques du chinois (trà xanh) et des créations numériques (thả thính).

Sur le plan méthodologique, le projet combine des approches issues du traitement automatique des langues (TAL). Deux corpus comparables ont été constitués :

(i) un lexique de référence, compilé à partir de dictionnaires chinois et vietnamiens publiés avant 2015 (environ 230 000 entrées pour le chinois et 80 000 pour le vietnamien) ;

(ii) un corpus médiatique en ligne (articles de presse numérique), composé de 1 286 articles chinois et 759 articles vietnamiens publiés entre 2015 et 2025, recueilli par web scraping sur six sites d'actualité (人民日报, 澎湃新闻, 新华网 pour le chinois ; Kenh14, Cafebiz et Genk pour le vietnamien), puis prétraité par normalisation, tokenisation (THULAC pour le chinois, PyVi pour le vietnamien) et harmonisation dans des formats CSV/JSON unifiés.

L'identification automatique des néologismes repose sur une approche en deux temps, fondée sur le principe de l'out-of-vocabulary (OOV), consistant à comparer les formes du corpus récent aux lexiques antérieurs. Inspirée des travaux de Sagot & Nouvel (2009, 2013), cette méthode a été adaptée aux langues analytiques et isolantes : elle permet de détecter des séquences inédites, d'en estimer la fréquence et la première année d'apparition, tout en intégrant un premier filtrage du bruit (nombres, entités nommées, symboles). Dans un second temps, les candidats ainsi identifiés sont affinés à l'aide de méthodes statistiques complémentaires, notamment l'analyse des n-grammes (1 à 5 unités), le calcul du PMI (Pointwise Mutual Information) et de l'entropie contextuelle, afin d'évaluer leur cohésion lexicale. Cette double approche, à la fois statistique et contextuelle, permet d'identifier aussi bien des néologismes monosyllabiques que des unités polylexicales émergentes. Les premiers résultats mettent en évidence plusieurs tendances, notamment l'émergence de termes sociopolitiques, technologiques et culturels tels que IA, 新冠 (Covid-19), 防控 (« prévention et contrôle »), 脑机 (« interface cerveau-ordinateur ») en chinois, et IA, smartphone, VinFast, sống ảo (« vie virtuelle ») en vietnamien.

Cependant, l'évaluation manuelle met en évidence du bruit (chiffres, noms, symboles), démontrant la nécessité d'un filtrage linguistique (normalisation, reconnaissance d'entités nommées, détection d'expressions multi-mots).

L'analyse linguistique menée à l'aide de l'outil de textométrie TXM a permis d'examiner l'évolution et les usages de néologismes chinois et vietnamiens dans quatre grands domaines — technologie, santé, politique et économie — à partir d'un corpus médiatique (2015–2025). L'analyse TXM met en évidence la diversité des processus néologiques, l'influence croissante des technologies et des médias, ainsi que la manière dont le chinois et le vietnamien intègrent, chacun à leur façon, les transformations linguistiques et culturelles de l'ère numérique.

Bibliographie :

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American, 284(5)*, 34–43.

Cartier, E. (2011). Néologie et description linguistique pour le TAL. *Langages*, 183, 105–117.

Dahlberg, C., Liu, T. Q., & Chen, C. F. (s.d.). Vocabulary Entry of Neologism: A Lexicographical Project aided with NLP Application. *Nanjing University, School of Foreign Studies*.

Hồ Tú Bảo, & Lương Chi Mai. (s.d.). *Về xử lý tiếng Việt trong công nghệ thông tin.* Viện Công nghệ Thông tin & Viện Khoa học và Công nghệ Tiên tiến Nhật Bản.

Humbert-Droz, J. (2023). When Terms Become Neologisms: A Contribution to the Study of Neology from the Perspective of Determinologisation. *CLINA: An Interdisciplinary Journal of Translation, Interpreting and Intercultural Communication, 9(1)*, 135–158. https://doi.org/10.14201/clina202391135158

Le-Hong, P., Roussanaly, A., & Nguyen, T. M. H. (2015). A syntactic component for Vietnamese language processing. *Journal of Language Modelling, 3(1)*, 145–184.

McCrae, J., Bosque-Gil, J., Gracia, J., Buitelaar, P., & Cimiano, P. (2017). The OntoLex-Lemon model: Development and applications. In I. Kosem et al. (Eds.), *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 Conference* (pp. 587–597). Lexical Computing CZ.

Mejri, S., & Sablayrolles, J.-F. (s.d.). *Présentation : Néologie, nouveaux modèles théoriques et NTIC.* Laboratoire « Lexiques, Dictionnaires, Informatique » (CNRS UMR 7187), Université Paris 13.

Pan, W., Liu, T., Sun, Y., Gong, B., Zhang, Y., & Yang, P. (2021). **融合自**编码器和对抗训练的中文新词发现方法 [Finding Chinese New Word by Combining Self-encoder and Adversarial Training]. **第二十届中国**计算语言学大会论文集, 736–746 页, **呼和浩特**, **中国**, 2021**年**8**月**13–15**日**.

SABLAYROLLES, J. (1997). Néologismes : Une typologie des typologies. *Cahier du CIEL, 1996–1997*, 11–48.

Sagot, B., Nouvel, D., Mouilleron, V., & Baranes, M. (2013). Extension dynamique de lexiques morphologiques pour le français à partir d'un flux textuel. *TALN-RÉCITAL 2013, Les Sables d'Olonne, 17–21 juin.*

Sablayrolles, J.-F. (2006). La néologie aujourd'hui. In C. Gruaz (Ed.), *À la recherche du mot : De la langue au discours* (pp. 141–157). Lambert-Lucas.

Tan, Q. (2024). Linguistic analysis of Chinese neologisms from 2017 to 2021. *International Journal of Language and Literary Studies, 6(2)*, 260–275. https://doi.org/10.36892/ijlls.v6i2.1449

Christine Da Silva /Marine Le Mené / Caroline Masson /Christophe Benzitoun / Loïc Liegeois
Université de Lorraine / Université de Québec / Université Sorbonne nouvelle / Université de Lorraine, ATILF, CNRS / Université Clermont Auvergne

Prise en compte de la variation dans l'évaluation des compétences lexico-sémantiques des enfants : réflexions théoriques et méthodologiques

L'évaluation du lexique joue un rôle essentiel pour apprécier les compétences langagières des enfants, comme des adultes. D'un point de vue développemental, cela s'explique en partie par le lien étroit qui existe entre le développement lexical et celui des apprentissages ultérieurs. En effet, le niveau de vocabulaire est un bon prédicteur du développement morphosyntaxique (Tat Cheung et al., 2024) comme de la réussite scolaire (Kurdek & Sinclair, 2001). En orthophonie, ce niveau est mesuré le plus souvent à l'aide d'épreuves standardisées et étalonnées. Deux types d'épreuves sont le plus couramment utilisés : une épreuve de dénomination, pour mesurer le versant de la production, et une épreuve de désignation d'images, pour évaluer le versant de la compréhension (da Silva-Genest et al., 2024). Bien que pertinentes pour situer le développement d'un enfant par rapport à un groupe de référence, ces épreuves ont fait l'objet de nombreuses critiques relatives à la qualité psychométrique de certaines épreuves (Betz et al., 2013 ; Macchi et al., 2023), ou aux choix linguistiques réalisés pour élaborer le matériel d'évaluation (Rondal, 2003 ; da Silva-Genest et al., 2024). Parmi les limites constatées figurent le nombre généralement limité d'items testés, le manque de transparence dans la sélection des items ou encore le problème de la polysémie des images. De surcroît, il apparaît que la variation linguistique, quelle qu'elle soit (diatopique, diastratique ou diaphasique), n'est que très peu prise en compte dans les épreuves utilisées pour tester les compétences lexicales. Ces outils sont donc peu propices à considérer la langue autrement que de façon normative et homogènes. Ainsi, bien que la richesse des manipulations langagières et les choix lexicaux effectués par les locuteurs puissent être des critères pertinents pour évaluer les compétences lexico-sémantiques et communicationnelles d'un individu, les mobiliser présupposerait de passer par des mesures fines du langage de l'enfant dans des situations plus spontanées, ce qui peut sembler plus difficile à mettre en place dans une pratique clinique (Klatte et al., 2022).

À la lumière de nos premières réflexions sur les outils d'évaluation en orthophonie et sur la nécessité de considérer le caractère dynamique du système linguistique (da Silva-Genest, 2020; da Silva-Genest, 2024), notre objectif est de déterminer de façon plus exhaustive la façon dont la variation est traitée dans le cadre de l'évaluation du langage des enfants – et plus précisément de leurs compétences lexico-sémantiques – et d'identifier si la place faite à la variation a évolué avec l'apparition de nouveaux outils d'évaluation. L'étude s'appuie sur deux types de données : d'une part, les outils d'évaluation et en particulier les épreuves standardisées ciblant l'évaluation des compétences lexico-sémantiques.

Nous avons analysé des épreuves lexicales de dénomination et de désignation d'images issues de 8 tests de langage francophones publiés entre 2001 et 2023 : ELO (Khomsi, 2001), EVALO 2-6 (Coquet, Roustit et Ferrand, 2009), EXALANG 3-6 (Helloin et Thibault, 2006), EXALANG 5-8 (Croteau et al., 2010), N-EEL (Chevrie-Muller et Plaza, 2001), CLÉA (Pasquet et al., 2014), L2MA-2 (Chevrie-Muller et al., 2010), PPVT-5 (Douglas & Dunn, 2023), EVT-3 (Williams, 2023). Pour chaque épreuve lexico-sémantique, ont été considérées les caractéristiques méthodologiques d'élaboration des épreuves ainsi qu'un inventaire des items proposés et des variantes acceptées. Ces données ont ensuite été analysées en tenant compte de la date de publication des outils afin d'apprécier l'évolution des mesures d'évaluation et des critères associés.

L'analyse de ces épreuves standardisées met en évidence des éléments partagés par la plupart des tests (e.g. items *cou* et *table* relevés dans presque tous les tests pour le versant de la compréhension), mais aussi des différences, liées selon nous en grande partie à la date de publication des outils. En effet, les versions les plus récentes se distinguent par une plus grande quantité d'items testés et une plus grande diversité des catégories grammaticales représentées. Par exemple, pour les tâches de désignation d'images, les outils les plus anciens testent entre 20 et 121

items et ne proposent généralement pas de verbes (4 tests sur 5) ; le test de compréhension lexicale le plus récent (PPTV-5) évalue quant à lui 239 items, dont 58 verbes et 23 adjectifs. De même, les outils les plus récents donnent la possibilité de dénommer de différentes façons une image, alors que cette possibilité de variation était presque inexistante dans les manuels des tests les plus anciens.

De manière générale, les résultats obtenus montrent que la variation demeure, en fonction des tests, un facteur peu pris en compte. L'évaluation s'inscrit nécessairement dans une tension entre la prise en compte de normes établies et l'attention portée à la variabilité des pratiques et des expériences individuelles. Notre communication visera à mettre en évidence l'intérêt de constituer une base de données des variantes lexicales observées dans les discours des locuteurs francophones et à discuter du rôle de la variation dans un double objectif : mieux définir les compétences langagières des enfants et comprendre son importance et son impact dans le développement langagier.

Références bibliographiques :

Betz, S., Eickhoff, J. & Sullivan, S. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment, *Language, Speech, and Hearing Services in Schools,* 44, 133-146.

Chevrie-Muller, C., Maillart, C., Simon, A.-M. & Fournier, S. (2010). *L2MA2 Batterie de langage oral, langage écrit, mémoire et attention*. ECPA.

Coquet, F., Roustit, J. & Ferrand, P. (2009). *EVALO 2-6*. Ortho Édition.

Croteau, B., Helloin, M.-C. & Thibault, M.-P. (2010). *EXALANG 5-8 : Batterie informatisée d'examen du langage oral et écrit*, Mont-Saint-Aignan, Orthomotus.

da Silva-Genest, C., Liégeois, L., Masson, C., Benzitoun, C. & Le Mené, M., Rabot, sécateur, téléphérique : interroger la pertinence des choix lexicaux dans les outils d'évaluation du langage en orthophonie, *Lidil* [En ligne], 62 | 2020, mis en ligne le 03 novembre 2020, URL : http://journals.openedition.org/lidil/8422 ; DOI : https://doi.org/10.4000/lidil.8422

da Silva-Genest, C., Masson, C., Liégeois, L., Benzitoun, C. & Le Mené Guigourès, M. (2024). Les problèmes liés à l'évaluation des compétences lexico-sémantiques orales des enfants : si j'avais un marteau, un rabot, un râteau, un tournevis… *Travaux de linguistique*, 89(2), 7-28. https://doi.org/10.3917/tl.089.0007.

Dunn, D. (2023). PPVT-5 Test de vocabulaire passif en images Peabody, 5ème édition. ECPA.

Helloin, M.-C. & Thibault, M.-P. (2006). *EXALANG 3-6 : Batterie informatisée pour l'évaluation du langage oral et langage écrit.* Orthomotus.

Klatte, I. S., van Heugten, V., Zwitserlood, R. & Gerrits, E. (2022). Language Sample Analysis in Clinical Practice: Speech-Language Pathologists' Barriers, Facilitators, and Needs. *Language, speech, and hearing services in schools*, *53*(1), 1–16. https://doi.org/10.1044/2021_LSHSS-21-00026

Khomsi, A. (2001). *Évaluation du langage oral*. ECPA.

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*(3), 451–455. https://doi.org/10.1037/0022-0663.93.3.451

Macchi, L., Herman, F., Colli-Vaast, L., Merle, A. & Danchin, P. (2023). Propriétés psychométriques des tests francophones de langage oral chez l'enfant. *Éla. Études de linguistique appliquée*, 210(2), 145-160. https://doi.org/10.3917/ela.210.0019.

Pasquet, F., Parbeau-Guéno, A. & Bourg, E. (2014). *CLéA : communiquer, lire et écrire pour apprendre. Batterie d'évaluation du fonctionnement et des dysfonctionnements du langage, oral et écrit*. ECPA Pearson.

Rondal, J.-A. (2003). *L'évaluation du langage*. Mardaga.

Tat Cheung, H., Hsu, C. L., & Ts'ou, B. (2024). The role of vocabulary and grammar in the listening text comprehension of school-age Cantonese-speaking children with developmental language disorder. *Brain and language*, 248, 105360. https://doi.org/10.1016/j.bandl.2023.105360

Williams, K. (2023). *EVT-3 Test de vocabulaire actif en images, 3ème édition*. ECPA.

Henriette Huber
Université de Düsseldorf

Word(s) in progress: A usage-based analysis of univerbation tendencies in New High German

During the composition of handwritten texts, e.g. in an exam setting, the brain has to make many quick decisions not only about the content that writers want to produce, but also how to write in such a way that the text is coherent and comprehensible and that (in languages with codified language norms) the individual words correspond to orthographic conventions. For that to work, information needs to be recalled from the mental lexicon, worked into the sentence structure that the writer aims for, and be converted into the motoric information needed to actually write the individual letters by hand – a process that overall does not seem to happen sequentially (cf. Kandel, 2023). It is no surprise then that such a split-second decision under the influence of sometimes inconclusive orthographic norms based on case-by-case judgements – as is the case for writing some linguistic units as words or phrases in German – can lead to variation in how these units are recalled depending on the writer as well as the specific context. Even when the norm is clear and known to the writer, there is no guarantee in the first place that the retrieval of those norms from memory goes off without a hitch every time, never mind during a stressful situation such as a final exam.

This leads us to the following questions: From a usage-based point of view, what factors play an influence on language users in regard to considering certain linguistic units words or phrases while producing written text, using German as an example? What makes a phrase progress into a word in people's minds? Donalies (2018, p. 12) for example argues that writers use solid spelling to signal the semantic cohesiveness of the concepts that are combined. So how much of a role do orthographic norms play for the language user when they are not set in stone and have changed due to a shift in wide-spread usage before?

To investigate these circumstances, we are going to focus on univerbation, which, according to Lehmann (2020, pp. 206ff.), is the diachronic process of at least two syntagmatically adjacent word forms (a chunk) transitioning into one graphemic word over time due to frequent usage with the syntactic structure still visible at the end of the process. Graphemic variation, he writes, constitutes as synchronic indication of the process taking place. Take these examples of spelling variants in German from the WebXL metacorpora (DWDS, 2025), for instance, which at first glance seem to stem from the same construction pattern:

(1) a. <in Folge dessen>
b. <infolge dessen>
c. <infolgedessen>
‚in consequence of'

(2) a. <auf Grund dessen>
b. <aufgrund dessen>
c. <aufgrunddessen>
‚on the basis of'

Both of these units were originally composed of three separate graphemic word forms, but converged over time – presumably due to frequent collocation. Both of them contain a preposition (*in/auf*), a noun (*Folge, Grund*) and an inflected form of the definite article (*dessen* – GEN.m/n), and both can be used synonymously for each other according to the Duden dictionary (cf. Dudenredaktion, 2025). It would be easy to assume that the norm is the same for both. However, in the case of (1), the only orthographically correct variant is (1c), while for (2), it is (2b) that is recommended and (2a) that is also deemed acceptable. What makes (1c) different from (2c)? Is frequency of use the only reason why one is considered acceptable and one is not, even though all three variants are attested? Likewise, we have frequent occurrences such as <gar nicht> vs. <garnicht> 'not at all' or <vor allem> vs. <vorallem> 'above all', for which to this day, only the syntagmatic variant is codified, but we can find multiple instances of both variants dating back over a century. If we follow Lehmann's definition, might these be prime candidates for future widely used and accepted univerbations?

As a pilot study to investigate these kinds of occurrences, all instances of tokens which were erroneously written either as one word or separated in GraphVar (Romstadt et al., 2024), a corpus consisting of German school-exit exams in the subjects of German, history and biology spanning from 1948-2018, will be accumulated as a first dataset. This corpus constitutes an excellent basis for an in-depth analysis of univerbation from a usage-based, cognitive point of view, as its spontaneously written texts with little time for revision potentially represent a less polished version of one's own thinking and writing process as well as not allowing any outside interference or editing. Furthermore, the exams having been written by hand means that any revisions that did occur are more traceable than is the case with most typewritten texts (especially from the past). When Fiehler (2011) conducted his study on univerbation using newspaper corpora, he noted this exact fact: That those newspaper articles are generally well revised and had at least one other person reading over them before publishing, making it a less ideal data basis. Likewise, Donalies (2018) tried to avoid this issue by including additional texts from Google, in hope of including data from less professional writers. However, one cannot discount the possibility of auto-correct having been used in at least some of those cases.

To analyse the data, the elicited tokens will be then annotated to account for factors such as word frequency, analogy, idiomaticity, word length and morphosyntactic context, factors that have been found to influence someone's decision in the past (cf. Fiehler, 2011; Donalies, 2018; Schäfer/Sayatz, 2024 for noun+verb units), as well as part-of-speech and syntactic flexibility. Afterwards, we will utilise *distinctive* and *covarying collexeme analyses* (cf. Stefanowitsch/Gries, 2009) to test the statistic validity of said factors, and in a second approach, a distributional semantic investigation using *fastText* (based on Stupak/Baayen, 2022) to see whether employing semantic vectors is able to provide any insight into the matter of morphosyntactic context and idiomaticity specifically. The results will gain us insight into why people univerbate as well as the current state of ongoing univerbation in German.

References

Donalies, E. (2018). *Wetterbeobachter, Zeitlang, wahrsagen, zartfühlend, kurzerhand, dergestalt: Handbuch zur Univerbierung*. Universitätsverlag Winter.

Dudenredaktion. (2025). infolgedessen. In *Duden online.* Retrieved October 6, 2025, from https://www.duden.de/rechtschreibung/infolgedessen

DWDS. (2025). *WebXL*. Retrieved July 7, 2025, from https://www.dwds.de/d/korpora/webxl

Fiehler, R. (2011). Korpusbasierte Analyse von Univerbierungsprozessen. In M. Konopka, J. Kubczak, C. Mair, F. Šticha, & U. H. Waßner (Eds.), *Grammatik und Korpora 2009. Dritte Internationale Konferenz. Mannheim, 22.-24.9.2009* (pp. 141-155). Narr.

Kandel, S. (2023). The APOMI Model of Word Writing: Anticipatory Processing of Orthographic and Motor Information. In R. Hartsuiker, & K. Strijkers (Eds.): *Language Production* (pp. 209-232). Routledge.

Lehmann, C. (2020). Univerbation. *Folia Linguistica Historica, 41(1)*, 205-252. https://doi.org/10.1515/flih-2020-0007

Romstadt, J., Strombach, T., & Berg, K. (2024). GraphVar – Ein Korpus für graphematische Variation (und mehr). In S. Krome, M. Habermann, H. Lobin, & A. Wöllstein (Eds.): *Orthographie in Wissenschaft und Gesellschaft* (pp. 425-435). De Gruyter.

Schäfer, R., & Sayatz, U. (2024). Between syntax and morphology: German noun+verb units. *Glossa: a journal of general linguistics, 9(1)*, 1-36. https://doi.org/10.16995/glossa.10069

Stefanowitsch, A., & Gries, S. Th. (2009). Corpora and grammar. In A. Lüdeling, & M. Kytö (Eds.): *Corpus Linguistics: An International Handbook*, Vol. 2 (pp. 933-952). De Gruyter Mouton.

Stupak, I. V., & Baayen, R. H. (2022). An inquiry into the semantic transparency and productivity of German particle verbs and derivational affixation. *The Mental Lexicon 17(3)*, 422-457.

Evelyne Jacquey / Laurence Kister / Sandrine Ollinger
Université de Lorraine, ATILF, CNRS

Annotation de contextes en référence et en facettes sémantiques pour faciliter la désambiguïsation lexicale.

La communication affichée proposée a pour fondement les résultats d'une campagne d'annotation manuelle des occurrences d'un lexème ambigu. Elle consiste en une expérience destinée à repérer les éléments linguistiques qui pourraient guider ou confirmer le processus de désambiguïsation lexicale des occurrences du lexème ambigu. Pour cela, nous nous intéressons à des relations de coréférences de deux types : les relations anaphoriques du lexème ambigu et les segments linguistiques (désormais indices) qui évoquent un ou plusieurs sens du lexème ambigu.

Les contextes analysés sont une sélection de contextes étendus issus d'un corpus préalablement annoté manuellement en lexies. Le corpus constitué pour l'expérience se compose de trente contextes : dix extraits qui ont soulevé des questions lors de leur annotation et vingt extraits choisis avec pour objectif de proposer un maximum de sens différents du lexème ambigu. Pour identifier les différentes formes de référence et les indices linguistiques susceptibles de guider la sélection du sens, les contextes sélectionnés sont constitués de la manière suivante :

- sept phrases avant l'occurrence du lexème ambigu
- sept phrases après l'occurrence du lexème ambigu quand il n'y a pas de contexte avant
- une ou plusieurs phrase(s) avant et après l'occurrence du lexème ambigu quand le contexte avant ou après lexème ambigu est court

Pour cette expérience nous utilisons l'interface d'annotation INCEpTION[13[1]] pour identifier et annoter les relations de coréférence et les indices linguistiques avec la méthodologie suivante :

- identification des éléments qui entretiennent des relations de référence avec le lexème ambigu et identification des indices ;
- annotation des relations anaphoriques et des indices (voir les jeux d'étiquettes ci-après) ;
- croisement des annotations réalisées avec celles de l'annotation en lexies des occurrences du lexème ambigu dans le corpus ;
- réunions de consensus entre les annotateurs.

Deux jeux d'étiquettes ont été utilisés pour cette expérience. Le premier permet d'annoter les relations de coréférence : anaphores fidèles, anaphores pronominales, anaphores possessives et anaphores infidèles. L'annotation des anaphores associatives en lien avec le lexème ambigu a été écartée car elle est fortement sous représentée dans l'échantillon de contexte utilisé. Le second concerne les indices qui évoquent les sens du lexème ambigu. Le jeu d'étiquettes comporte cinq facettes sémantiques qui correspondent aux sens du lexème ambigu dans la ressource lexicale de référence choisie :

- facette 1 - astre dans le ciel ;
- facette 2 - étoile, foyer du système planétaire de la terre ;
- facette 3 - lumière, chaleur ou temps qu'il fait ;
- facette 4 - zone de X (avoir un bien au X, aller en vacances au X, aller vivre au X) ;
- facette 5 - situation, fait, objet ou personne qui illumine la vie de quelqu'un et le rend heureux.

13[1] INCEpTION est une plateforme d'annotation sémantique qui offre une assistance intelligente et une gestion des connaissances tout en intégrant des tâches connexes dans une plateforme web commune. INCEpTION permet la compilation des tâches successives et la création ou l'extension de bases de connaissances spécifiques aux corpus. La plateforme est développée par laboratoire UKP de l'Université technique de Darmstadt.

Les premières observations à l'issue de cette campagne d'annotation montrent que le lexème ambigu entretient très peu de relations anaphoriques dans le corpus observé. Même si les anaphores fidèles et les anaphores pronominales sont présentes, elles sont assez rares et elles n'interviennent pas directement dans le processus de « prédiction » des sens des occurrences du lexème ambigu. En revanche, les relations de coréférence entre les indices semblent intéressantes à creuser.

Une analyse quantitative des résultats a été réalisée dans l'objectif d'évaluer les contextes pour lesquels il y a convergence entre l'annotation en lexies et l'annotation en facettes sémantiques. Il apparaît que le taux de convergence n'est pas systématiquement conditionné par le nombre de facettes et d'indices.

En ce qui concerne l'analyse qualitative des résultats, on observe trois cas de figures : une divergence totale des facettes, une convergence totale des facettes et des convergences/divergences partielles des facettes. Lorsque nous croisons ces observations avec l'annotation en lexies, nous relevons une convergence entre facettes et lexies en cas de convergence totale ou partielle de l'annotation en facettes.

En conclusion, il apparaît que les relations de coréférence autour du lexème ambigu n'apportent pas d'information permettant de faciliter la désambiguïsation. Les indices quant à eux semblent pouvoir guider ou confirmer le choix du sens du lexème ambigu retenu par l'annotateur quand :

- on est en présence d'un grand nombre d'indices en faveur d'une même facette ;
- on repère la présence simultanée de certaines facettes qui semblent s'attirer.

Le rôle des relations de coréférences entre les indices reste à analyser de manière systématique.

Parmi les perspectives ouvertes par cette première expérience, nous pouvons envisager d'étendre l'identification et l'annotation des relations de référence présentes dans les contextes et appliquer la méthodologie à d'autres lexèmes et d'autres corpus. En nous appuyant sur une autre campagne d'annotation en lexies, nous envisageons, en particulier, d'utiliser un jeu de vingt et un lexèmes dont les sens sont équi-répartis dans un corpus de quatre cent vingt emplois (vingt emplois par lexèmes). Si ce corpus plus conséquent se révèle insuffisant nous utiliserons le FrenchTenTen pour l'étendre et en diversifier les emplois. Nous prévoyons aussi d'examiner l'incidence de la distance entre les sens des lexèmes de l'expérience : métaphore vs métonymie.

Références bibliographiques :

BARQUE Lucie ; HAAS, Pauline, 2022. La polysémie régulière. In : Lexique n°31.

DELABORDE, Marine, 2020. Analyse en corpus de chaînes de référence : la coréférence non-stricte à l'épreuve de la linguistique outillée. Thèse de Doctorant. Université de la Sorbonne Nouvelle – Paris III.

KLEIBER, Georges, 1999. Problèmes de sémantique. La polysémie en question. Villeneuve d'Asq : Presses du Septentrion.

KLEIBER, Georges, 2008. Petit essai pour montrer que la polysémie n'est pas un sens interdit. In : DURAND, Jacques ; HABERT, Benoit ; LAKS, Bernard (éd.). Congrès Mondial de Linguistique Française - CMLF'08. Paris, pp. 338-349.

LANDRAGIN, Frédéric, 2021. Etudes de chaîne s de références en français : liens entre modélisation linguistique et analyse quantitative. In : Bulletin de la Société de Linguistique de Paris. CXVI (1), pp.41-75.

LANDRAGIN, Frédéric, 2022. Expressions référentielles et chaînes de référence en français : le projet Democrat et exploration des rapports entre linguistiques textuelle et linguistique de corpus. In Echo des études romanes, 18(1), pp. 49-65.

LEHMANN, Alise ; MARTIN-BERTHET, Françoise, 2018. La polysémie. In : Sémantique, morphologie et lexicographie. 5e éd. Paris : Armand Colin. Chap. 5, pp. 95-142.

POLGUERE, Alain, 2020. Apports de la lexicographie réticulaire : le cas de la polysémie relationnelle. In : Langue(s) et Parole(s), n°5, 135-156.

VICTORRI, Bernard ; FUCHS, Catherine, 1996. La polysémie – construction dynamique du sens. Montréal : Hermès.

KLIE, Jan-Christoph ; BUGERT, Michael ; BOULLOSA, Beto ; ECKART DE CASTILHO, Richard et GUREVYCH, Iryna, 2018. The Platform : Machine-Assisted and Knowledge-Oriented Interactive Annotation. In Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics (COLING 2018), Santa Fe, New Mexico, USA.

Filip Kalaš / Pavol Lipták
Bratislava University of Economics and Business, Faculty of Applied Languages /
Bratislava University of Economics and Business, Faculty of Commerce

Quantifying Bilingual Terminology Extraction for Slovak Public Administration: GPT Prompting vs. RAG

This paper presents a quantitative evaluation of bilingual terminology extraction for Slovak–English public-administration discourse. The study compares two GPT-5.2 configurations on the same ~35,000-token corpus of official texts and translations: a prompt-only baseline and a Retrieval-Augmented Generation (RAG) pipeline that retrieves domain passages prior to generation. The research follows the project's remit (automated extraction + human validation of bilingual terminology; comparison of direct prompting vs. RAG-enhanced extraction) and uses the same task framing and corpus scope as the original design.

Our design is grounded in three current findings. First, long-context LLMs exhibit position sensitivity ("lost in the middle"): performance peaks when evidence is at the beginning or end of the context and degrades when evidence lies mid-context (Liu et al., 2024). Second, RAG variants, especially those enriched with knowledge graphs, improve retrieval specificity and factuality over prompt-only systems (Xu et al., 2024). Third, terminology control in MT/LLM pipelines benefits from explicit glossary and constraint mechanisms, showing measurable gains from term extraction + integration procedures (Kim et al., 2024; Moslem et al., 2023).

The paper set out to quantify whether RAG > Prompt on bilingual term extraction under controlled conditions and identify where gains originate (retrieval vs. generation).

We use a corpus of Slovak public-administration documents with authoritative English counterparts (≈35,000 tokens), spanning subdomains such as regional governance, urban policy, and finance.

We compare a Prompt baseline—templated zero/few-shot prompts with a fixed schema producing bilingual output—and a RAG system that uses a hybrid retriever (dense with lexical fallback) over sentence/paragraph chunks with an optional knowledge-graph (KG) overlay for key entities/relations (institution→competence; procedure→document), following evidence that KG-aware RAG improves retrieval specificity (Xu et al. 2024)

We report micro- and macro-precision, recall, and F1 with 95% bootstrap confidence intervals; we also compute Average Precision (AP) and AUC-PR for ranked candidates, assess calibration via the Brier score and reliability curves, quantify lexical diversity using the Type–Token Ratio and Shannon entropy, and estimate PMI-based co-occurrence density to capture multiword cohesion. Two expert terminologists validate all candidates, and we quantify annotator reliability and model–human concordance using Cohen's κ and Krippendorff's α. For significance testing, we compare F1 and AP with document-paired t-tests (or Wilcoxon signed-rank tests when normality is violated), apply Holm corrections for multiple comparisons, report effect sizes (Cohen's d, Cliff's δ).

Prompts follow up-to-date, task-explicit, schema-first guidelines to ensure reproducibility across runs and model variants.

The quantitative design offers clear advantages: it provides objectivity and replicability through confidence intervals and effect sizes; threshold-free ranking metrics such as AP and AUC-PR reduce the sensitivity of conclusions to a single F1 threshold; calibration analysis makes over- or under-confidence visible—an established issue in generative information extraction; agreement statistics separate model error from human variability.

We hypothesise that RAG will outperform Prompt on F1 and AP, with larger gains for multiword terms and when evidence is located mid-context; improvements may diminish for highly formulaic terms and ultra-rare items. This

expectation aligns with reported RAG benefits in domain retrieval (Xu et al., 2024) and with documented position sensitivity in long-context models (Liu et al., 2024).

References

Fleischmann, K., & Lang, C. (2025). *Terminologie in der KI: Wie mit Terminologie der Output von LLMs und GenAI optimiert werden kann*. In Deutscher Terminologie-Tag: Akten des Symposiums, Worms.

Kim, S., Sung, M., Lee, J., Lim, H., & Gimenez Perez, J. F. (2024). *Efficient terminology integration for LLM-based translation in specialized domains*. In B. Haddow, T. Kocmi, P. Koehn, & C. Monz (Eds.), Proceedings of the Ninth Conference on Machine Translation (WMT 2024). Association for Computational Linguistics. https://aclanthology.org/2024.wmt-1.51/

Larsen, M., & Weßels, D. (2022). *Chain-of-thought prompting. KI-Transfer Hub SH*. https://kuenstliche-intelligenz.sh/de/archiv/chain-of-thought-prompting kuenstliche-intelligenz.sh

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). *Lost in the middle: How language models use long contexts.* Transactions of the Association for Computational Linguistics, 12, 157–173. https://doi.org/10.1162/tacl_a_00638

Moslem, Y., Romani, G., Molaei, M., Kelleher, J. D., Haque, R., & Way, A. (2023). *Domain terminology integration into machine translation: Leveraging large language models*. In P. Koehn, B. Haddow, T. Kocmi, & C. Monz (Eds.), Proceedings of the Eighth Conference on Machine Translation (pp. 902–911). Association for Computational Linguistics. https://aclanthology.org/2023.wmt-1.82/

Muegge, U. (2023). *Terminology extraction for translation and interpretation made easy*.

OpenAI. (2025). *Best practices for prompt engineering with the OpenAI API*. https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api OpenAI Help Center

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). *Chain-of-thought prompting elicits reasoning in large language models*. arXiv. https://arxiv.org/abs/2201.11903

Xu, Z., Cruz, M. J., Guevara, M., Wang, T., Deshpande, M., Wang, X., & Li, Z. (2024). *Retrieval-Augmented Generation with Knowledge Graphs for customer service question answering*. arXiv. https://arxiv.org/abs/2404.17723

Acknowledgments

The study has been elaborated within the projects of young teachers, scholars and postgraduate students *A-25-106/3020-18 Artificial intelligence: challenges for linguistics and marketing*, I-26-101-00 *Language as a Sales Tool: Modelling Trust and Effectiveness in AI-Driven Brand Communication* and APVV 24-0302 *Harmonisation of Slovak and English terminology of local government in the Slovak Republic*

Marie-Laurence Knittel
Université de Lorraine, ATILF, CNRS

L'ambiguïté des adjectifs en -ant : structure argumentale et propriétés du suffixe -ant

Les adjectifs en -ant n'ont fait l'objet que de quelques travaux (Halmøy 1984, Arnavielle 2003), surtout en relation avec les formes verbales en -ant (participe présent et gérondif). Cependant, aucun d'entre eux ne mentionne l'ambigüité régulière de certaines de ces formes, illustrée en (1-3).

1. *blanchissant*
   a. *des cheveux blanchissants* : des cheveux qui blanchissent
   b. *une lessive blanchissante* : une lessive qui blanchit

2. *vieillissant*
   a. *un homme vieillissant* : un homme qui vieillit
   b. *une coiffure vieillissante* : une coiffure qui vieillit

3. *cicatrisant*
   a. *une plaie cicatrisante :* une plaie qui cicatrise
   b. *une lotion cicatrisante :* une lotion qui cicatrise

L'ambigüité des adjectifs ci-dessus n'est pas toujours enregistrée dans les dictionnaires (1, 2 vs 3), mais des exemples peuvent aisément être repérés sur Internet, ce qui témoigne de la régularité de cette alternance sémantique.
Dans cette présentation, nous montrons que cette ambigüité est héritée de celle des verbes de base de ces adjectifs, et que les propriétés du suffixe -ant varient également selon la valeur de la base verbale.

Héritage de la structure argumentale du verbe

Les sens disponibles pour les adjectifs en (1-3) peuvent être décrits à partir des observations de Rainer (1999) et de Fábregas (2016, 2020), portant sur les adjectifs déverbaux de l'espagnol.

Dans les exemples (a) de (1-3), les adjectifs prédiquent sur le Patient du verbe de base, i.e. l'entité qui subit l'action dénotée par le verbe.

4. a. *des cheveux blanchissants* : des cheveux qui blanchissent, i.e. qui deviennent blancs
   b. *un homme vieillissant* : un homme qui vieillit, i.e. qui devient (plus) vieux
   c. *une plaie cicatrisante* : une plaie qui cicatrise

Dans les exemples (b), ils prédiquent sur l'agument Cause, i.e. l'entité non-volitionnelle à l'origine de l'événement décrit par le verbe. Ils sont paraphrasables par 'qui peut causer l'action dénotée par V'.

5. a. *une lessive blanchissante* : une lessive qui blanchit, i.e. qui rend (plus) blanc
   b. *un vêtement vieillissant* : un vêtement qui vieillit, i.e. qui fait paraître plus vieux
   c. *une lotion cicatrisante* : une lotion qui cicatrise, qui favorise la cicatrisation

Les exemples ci-dessus montrent que l'ambiguité des adjectifs provient de celle des verbes de base, qui présentent l'alternance anticausative (Levin 1993, Alexiadou et al. 2006; Pylkkänen 2008, Zribi-Hertz 1987), schématisée en (6) :

6. a. $\text{Sujet}_{\text{Cause}}$ V $\text{Objet}_{\text{Patient}}$
   b. $\text{Sujet}_{\text{Patient}}$ V

Les verbes anticausatifs admettent deux emplois ; en emploi transitif, ils sélectionnent un sujet interprété comme une Cause et un Objet interprété comme un Patient (6a). En emploi intransitif, ils conservent l'argument Patient, réalisé comme sujet (6b).
L'exemple (7) illustre le fonctionnement anticausatif de *vieillir,* base de *vieillissant* (2).

7. a. *Ces lunettes*$_{Cause}$ *le*$_{Patient}$ *vieillissent beaucoup.*
   b. *Ses parents*$_{Patient}$ *ont beaucoup vieilli.*

L'ambiguïté des adjectifs en -ant est donc héritée de celle de leurs verbes de base. La structure verbale (7a) donne lieu aux adjectifs dispositionnels, qui en conservent l'argument Cause (2b) ; la structure (7b) produit les adjectifs potentiels, qui maintiennent l'argument Patient (2a). Dans les deux cas, l'argument maintenu correspond au sujet de la structure verbale. Il est ainsi possible de former un adjectif à partir de chacune des structures argumentales.

Valeur des adjectifs

Rainer (1999) et Fábregas (2016, 2020) remarquent que les adjectifs déverbaux, au contraire des verbes à l'indicatif, ne peuvent pas décrire des événements effectifs. Ainsi, les adjectifs en (5) décrivent des capacités des entités sur lesquels ils prédiquent à participer à des événements, mais n'impliquent pas la réalisation des événements en question. L'exemple (8) n'est donc pas contradictoire, alors même que la séquence Nom + Adjectif est suivie par une continuation niant l'occurrence de l'événement décrit par le verbe de base de l'adjectif.

8. *Cette lotion cicatrisante n'est pas efficace, la peau n'a pas encore cicatrisé.*

De tels adjectifs sont qualifiés de 'dispositionnels' par ces auteurs, du fait qu'ils décrivent des dispositions des entités sur lesquels prédiquent ces adjectifs.
Il en est de même dans le cas des adjectifs en *-able*, qui prédiquent sur des Patients, et sont appelés 'potentiels'.

9. a. *un tissu lavable* : un tissu qu'on peut laver
   b. *Ce manteau est lavable, mais il n'a encore jamais été lavé.*

Cependant, les adjectifs en -ant construits sur des bases anticausatives et prédiquant sur des Patients ne présentent pas toujours ce comportement. Si certains ont bien une valeur potentielle (10), la plupart ne peuvent pas entrer dans des structures de type (9b) sans provoquer une contradiction (11).

10. a. *un matériau cassant* : qui peut se casser facilement
    b. *Ce matériau cassant ne s'est pas cassé, car il a été manipulé avec précaution.*

11. a. *# Ses cheveux blanchissants ne blanchissaient pas.*
    b. *# Ses parents vieillissants ne vieillissent pas.*
    c. *# La plaie cicatrisante ne cicatrisait pas.*

On observe au contraire qu'ils sont paraphrasables par leur verbe de base employé au présent ou au progressif *être en train de V* :

12. a. *des cheveux blanchissants* : des cheveux qui {blanchissent / sont en train de blanchir}
    b. *les personnes vieillissantes* : les personnes qui {vieillissent / sont en train de vieillir}
    c. *une plaie cicatrisante* : une plaie qui {cicatrise / est en train de cicatriser}

Ces adjectifs sont donc bien aptes à décrire la participation à un événement en cours, et non la simple capacité ou tendance à participer à un certain type d'événement. Ils accèdent donc à la lecture épisodique, typique des verbes conjugués à l'indicatif.

Rôle du suffixe

Nous supposons que la valeur épisodique des adjectifs dérivés de verbes anticausatifs et prédiquant sur des Patients résulte d'une réanalyse du suffixe -ant de participe présent / gérondif, à valeur progressive (cf. *J'ai trouvé Paul dormant*$_{Part.Prst}$ *tranquillement = J'ai trouvé Paul en train de dormir tranquillement*) en un marqueur adjectival à valeur épisodique.

C'est donc ce suffixe -ant à valeur épisodique qui apparaît sur les verbes anticausatifs en emploi intransitif, mais également sur d'autres verbes qui partagent avec les précédents l'intransitivité, l'atélicité et la sélection de sujets non-agentifs (Knittel 2023) comme le montrent les exemples en (13).

13. a. *des étoiles scintillantes* : des étoiles qui sont en train de scintiller
    b. *des mains tremblantes* : des mains qui sont en train de trembler
    c. *une flamme vacillante* : une flamme qui est en train de vaciller

Références

Alexiadou, A., Anagnostopoulou, E. & F. Schäfer. 2006. The properties of anticausatives crosslinguistically. In: M. Frascarelli (ed.), *Phases of Interpretation.* Berlin: Mouton.

Arnavielle, Teddy. 2003. Le participe, les formes en -ant: positions et propositions. *Langages 149. 37–54.*

Fábregas, A.. 2016. Deconstructing the non-episodic readings of Spanish deverbal adjectives. *Word Structure* 9 (1), 1-41.

Fábregas, A.. 2020. Morphologically Derived Adjectives in Spanish, Amsterdam, John Benjamins.

Halmøy, O.. 1984. À propos de l'adjectif en –ant, dit "verbal". *Revue Romane* 19(1), 48-64.

Knittel, M.L. 2023. Deverbal adjectives with episodic value in French: evidence for mixed categories. *Probus* 35-2, 277-325

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation.* Chicago: University of Chicago Press.

Pylkkänen, L.. 2008. *Introducing arguments.* Linguistic Inquiry Monographs 49, Cambridge MA: MIT Press

Rainer, F.. 1999. La derivación adjetival. In I. Bosque & V. Demonte (eds), *Gramática descriptiva de la lengua española,* Madrid, Espasa Calpe, 4595-4645.

Zribi-Hertz, A.. 1987. La réflexivité ergative en français moderne. *Français moderne, 55(1-2), 23-54.*

Matthieu Quignard / Audrey Mazur
ENS Lyon

L'impact de la dyslexie-dysorthographie sur les pratiques lexicales : un outil de recherche et d'aide à la remédiation

1. Contextualisation

Dans le cadre des projets *ETUDYS* et *FLEXIDYS*[14[1]], des productions écrites et orales d'étudiants présentant une dyslexie et contrôles appariés ont été collectées. Les étudiants ont été invités à produire quatre productions textuelles : narratif écrit, narratif oral, expositif écrit et expositif oral (cf. pour une description du protocole, par exemple, Mazur-Palandre Quignard et Witko, 2021, 2022). Le corpus ainsi collecté permet d'avoir pour chaque individu quatre textes sur une même thématique, permettant des comparaisons linguistiques qualitatives et quantitatives. Ce corpus reste un échantillon de ce que nous pourrions observer par ailleurs mais dimensionné pour permettre de répondre à des questions de recherche en psycholinguistique. Il est ainsi composé de 172 textes, 86 oraux et 86 écrits. La distribution (moyenne et écart-type) des nombres de mots est donnée dans le tableau 1. L'ensemble du corpus comporte 43 662 mots (ponctuation exclue).

	Expo. oraux	Expo. écrits	Narr. oraux	Narr. écrits
Étudiants Dyslexiques (21)	301 (266)	198 (102)	368 (235)	211 (133)
Étudiants Contrôles (22)	281 (203)	183 (139)	305 (205)	186 (113)

Tableau 1- Nombre moyen de mots par texte selon les groupes, genres et modalités

Une des questions de recherche de ces projets portait sur le lexique et les potentiels « mots à risque » pour les étudiants présentant de la dyslexie-dysorthographie. Pour ce faire, les techniques de la linguistique de corpus ont été mobilisées pour développer une chaine de traitement linguistique, ALEX, permettant, pour chaque mot produit par les individus, d'en obtenir des propriétés lexicales. Cette chaine de traitement linguistique a permis de produire : 1- un corpus annoté permettant une exploitation psycholinguistique des données ; 2- une base de données lexicale enrichie et enfin 3- un outil aidant à la remédiation à destination des orthophonistes. Ce présent papier a comme objectif de présenter le développement de la chaine de traitement lexicale ALEX et d'en montrer les intérêts à la fois en recherche et en clinique (orthophonie).

2. La chaine de traitement lexicale ALEX

2.1 Présentation d'ALEX

La chaine de traitement lexicale ALEX a été développée en s'appuyant sur les bases lexicales Lexiques (Lexique.org) et Manulex (Lété et al., 2004). Elle indique pour chaque mot donné les propriétés suivantes :

- Longueur : nombre de syllabes écrites, de lettres et de phonèmes ;
- Propriétés graphémiques et phonétiques : l'indicateur de cohérence phonème/graphème, l'indicateur de cohérence graphème/phonème (Lété et al., 2004) et la structure phonologique (succession consonne/voyelle) ;
- La fréquence des lemmes et des formes, selon Lexique.org ;

14[1] Accueil, intégration et accompagnement des ETUdiants DYSlexiques à l'Université ; Financements PEPS CNRS, ASLAN, ENS de Lyon

- La classe grammaticale : adjectif, adverbe, etc., incluant la ponctuation.

2.2 Étapes d'intervention

Plusieurs étapes ont été nécessaires. La première a consisté à obtenir les productions textuelles des sujets à partir des transcriptions annotées dans le logiciel CLAN. Nous avons utilisé deux scripts de conversion de format (CLAN vers XML/TEI, puis XML/TEI vers texte brut) pour filtrer les annotations CLAN préexistantes.

La deuxième étape a consisté à tokeniser le corpus à l'aide du logiciel TreeTagger (Schmid, 1994) et du modèle de base du français (french-utf8.par). Les 67 mots des textes qui n'étaient pas inclus dans le modèle ont été ajoutés manuellement afin que les 86 textes écrits puissent être analysés. Les ajouts étaient des noms propres et des mots très spécifiques dans le sens où ces mots, par exemple, ne sont pas encore entrés dans les bases de données ou peu utilisés (*transgenre, powerpoint*). Ce processus permet d'obtenir, pour chaque mot, son lemme et sa catégorie grammaticale, selon le modèle TreeTagger (32 catégories). À l'issue de cette étape, nous avons extrait le lexique complet du corpus écrit, qui comprend 3 031 formes différentes et 16 702 formes au total.

Au cours de la troisième étape, nous avons injecté des informations lexicales en plus des lemmes et des parties de discours (POS) produits par TreeTagger. Les informations lexicales ont été fournies par deux bases de données lexicales pour le français : Lexique.org (de Boris New et Christophe Pallier) et Manulex-infra (Lété, Sprenger-Charolles et Colé, 2004). Lexique contient un grand nombre d'informations sur la phonétisation et la fréquence d'utilisation des mots dans les textes littéraires français et les films sous-titrés. Manulex a été développée par Lété et collaborateur (2004) pour la description des caractéristiques infralexicales des formes (45 000 entrées ; Ortega et al., 2010 ; Peereman et al., 2013). Cette base de données fournit pour chaque mot des indicateurs de difficulté orthographique (cohérence orthographique), calculés pour le début, le milieu, la fin du mot et pour le mot entier. Comme dans le cas de TreeTagger, nous avons dû compléter à la main les informations manquantes correspondant aux mots de notre corpus, inconnues pour ces bases de données. Il convient de noter que la langue française contient un grand nombre de formes ambiguës (par exemple, *le* peut être à la fois un déterminant et un pronom). Pour résoudre ces ambiguïtés et attribuer l'information correcte à la forme ciblée, nous avons basé notre appariement sur la forme, le lemme et le POS donnés par TreeTagger. Comme ces catégories POS ne coïncident pas strictement avec celles de Lexique et Manulex, nous avons dû reconfigurer toutes les catégories POS pour établir un jeu de balises partagé entre les trois bases de données. Si par ce processus nous y perdons en précision, nous y gagnons en cohérence entre les bases lexicales et dans la désambiguïsation des formes. La plus grande partie de l'enrichissement manuel a consisté à calculer la cohérence orthographique, selon l'algorithme Manulex[15][2]. La base Manulex a en effet une couverture lexicale limitée puisqu'elle a été développée à partir de manuels scolaires primaires français. Cette tâche manuelle a concerné 500 formes.

Cette chaîne de traitement ALEX permet de travailler sur les formes correctes du français. Les formes graphiques erronées peuvent être prises en compte d'une façon indirecte, une fois leur forme cible identifiée.

3. Utilité de cette chaine de traitement en recherche et en clinique

Cette chaine de traitement a été précieuse pour plusieurs études scientifiques permettant de mieux comprendre les impacts de la dyslexie-dysorthographie sur la production écrite textuelles et en particulier sur le lexique. Ces travaux ont révélé que les étudiants présentant une dyslexie : 1- utilisent, à l'écrit, les mêmes types de mots en termes de longueur, de fréquence, de classe grammaticale ou de cohérence orthographique, que les étudiants contrôles ; 2- mais font beaucoup plus d'erreurs d'orthographe et 3-produisent un sous-ensemble inattendu d'erreurs (mots à la fois courts et homophones).

15[2] Le manuel de Manulex (https://www.manulex.org/docs/fr/Manulex-infra.pdf) indique que la valeur de la cohérence d'un mot est la moyenne des valeurs de cohérence de chaque paire graphème-phonème qui le compose.

Cette chaine de traitement et les résultats lui étant associés ont suscité l'intérêt des praticiens en orthophonie. D'ALEX, nous avons donc développé l'application AnaliDys[16[3]] (Mazur, Quignard et Witko, 2025) afin de leur permettre de caractériser les propriétés lexicales des mots utilisés et erronés de leurs patients. AnaliDys peut en effet orienter la remédiation des patients présentant des troubles du langage écrit en spécifiant les mots corrects et erronés produits, en délivrant des analyses et cartographies des mots utilisés et erronés, pour un meilleur suivi de l'évolution des pratiques lexicales.

Références bibliographiques

Lété, B., Sprenger-Charolles, L. et Colé, P. (2004). Manulex: A grade-level lexical database from French elementary-school readers. *Behav. Res. Methods Instrum. Comput*., 36, 156–166.

Lété, B., Peereman, R. et Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French chil-dren: A regression-based study. *J. Mem. Lang*., 58:952–977.

Mazur-Palandre A., Quignard, M. et Witko, A. (2021). Confronting Lexical Choice and Error Distribution in Written French: New Insights into the Linguistic Insecurity of Students with Dyslexia. Brain Sciences, 11, 922.

Mazur, A., Quignard, M. et Witko, A. (2022). Persistance des difficultés en production écrite chez l'étudiant dyslexique : implémentation d'une chaîne de traitement lexical en pratique clinique. In Bois-Parriaud, F., Joly-Froment, S. et Witko, A. (Eds.), Actualités du langage écrit, 277-304. Isbergues, Ortho Édition.

Mazur, A., Quignard, M. et Witko, A. (2025). L'importance de l'échange entre recherche et clinique : l'application Analidys. *ELA, Études de Linguistique Appliquée*, 218(2), 163-174.

Ortega, É. et Lété, B. eManulex: Electronic Version of Manulex and Manulex-Infra Databases. 2010. Available online: http://www.manulex.org (accessed on 9 July 2021).

Peereman, R., Lété, B. et Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behav. Res. Methods*, 39:593–603.

Peereman, R., Sprenger-Charolles, L. et Messaoud-Galusi, S. (2013). The contribution of morphology to the consistency of spelling-to-sound relations: A quantitative analysis based on French elementary school readers. *Année Psychol. Top. Cogn. Psychol*., 113:3–33.

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Lexique : http://www.lexique.org/

MANULEX : http://www.manulex.org/fr/home.html

Tree Tagger : https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

16[3] https://aslan.universite-lyon.fr/projet-dysortho-l-application-analidys-253846.kjsp?RH=416260547274113

Jozef Štefčík
Université de Bratislava

Measuring quality of Monolingual Terminology Extraction: A Replicable Comparative Study of Prompt-Based vs Retrieval-Augmented approach

This study addresses an ongoing national project at the Faculty of Applied Languages, University of Economics in Bratislava, related to the harmonisation of English-Slovak terminology used by regional governments. It focuses on applied research to develop a consistent terminological framework to support local government administration, interdisciplinary collaboration, and international communication.

The resulting framework is designed for integration into digital environments, including interactive platforms, digitised glossaries, and multilingual online databases. It builds on fundamental research in concept-oriented terminology, the categorisation of conceptual systems, and systematic terminological management.

The study proposes a controlled comparison of two OpenAI large language model configurations for terminology extraction and termbase population in a specialised domain: (1) prompt-based extraction using ChatGPT Pro as a desktop application, and (2) context-based extraction using an API-based Retrieval-Augmented Generation (RAG) pipeline. The dataset consists of parallel English and Slovak documents in the domains of environmental policy and forestry, totalling approximately 53,000 words per language. Two monolingual gold-standard glossaries, one in English and one in Slovak, validated by domain experts, serve as reference resources for evaluation.

The study is motivated by recent research showing that RAG combines parametric model knowledge with external, updateable non-parametric memory, while its effectiveness depends strongly on retrieval quality, retrieval design, and the integration of structured lexical resources. The overall aim is to qualitatively assess precision and the signal-to-noise ratio in terminology extraction using the same corpus and extraction schema, and to determine whether retrieval-augmented prompting improves bilingual terminology extraction compared with direct prompting. The extracted term sets are validated by a subject-matter expert using termhood and unithood criteria, and the sample includes both single-word and multi-word expressions relevant to the domain.

To ensure methodological validity and replicability, both experimental conditions use the same corpus, glossary resources, model family, and fixed generation parameters. In the prompt-based condition, terminology extraction is guided by a standardised instruction template. In the RAG condition, glossary entries and document chunks are indexed as separate retrieval units and dynamically injected during extraction. The output is normalised and evaluated against the gold-standard glossaries using precision and recall, supplemented by the glossary-alignment rate and consistency across repeated term occurrences. Ambiguous cases are double-annotated and adjudicated, with inter-annotator agreement reported.

The study aims to determine whether glossary-centred retrieval augmentation offers a more robust and reproducible methodology for terminology extraction than prompt-only interaction with large language models in specialised language-processing settings. Its main objectives are to compare prompt-based terminology extraction in ChatGPT with context-based extraction in an API-based RAG pipeline, and to test whether retrieval-based contextualisation improves the identification of domain terms that should be Slovak-English aligned in subsequent stages of terminology work.

Two research questions are formulated as follows:

1. Does a RAG-based configuration improve terminology extraction performance compared with prompt-only extraction?

2. Does retrieval-based injection of glossary and document context improve alignment with monolingual gold-standard glossaries?

Looking ahead, the study will investigate the extent to which the accuracy of terminology extracted via RAG can be formalised using deterministic criteria derived from curated databases, and under what conditions such criteria remain scalable, valid, and empirically preferable to probabilistic or human-dependent methods.

References

Fleischmann, K., & Lang, Ch. (2025). Terminologie in der KI. Wie mit Terminologie der Output von LLMs und GenAI optimiert werden kann. *Deutscher Terminologie-Tag. Akten des Symposiums Worms*.

Larsen, Moritz; Weßels, Doris (2022): Chain of Thought Prompting – KI Transfer Hub SH https://kuenstliche-intelligenz.sh/de/archiv/chain-of-thought-prompting (retrieved on 22.09.2025)

Liu, Nelson F. et al. (2023): *Lost in the Middle: How Language Models Use Long Contexts https://arxiv.org/abs/2307.03172 (retrieved on 6.10.2025)*

Li, Tianle et.al (2024): Long-Context LLMs Struggle with Long In-context Learning https://arxiv.org/abs/2404.02060 *(retrieved on 6.10.2025)*

Muegge, U. (2023). *Terminology extraction for translation and interpretation made easy*. Amazon UK.

Xu, Zhentao, et al. (2024): Retrieval Augmented Generation with Knowledge Graphs for Customer Service Question Answering. In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, S.2905-2909.*

OpenAI: Best practices for prompt engineering with the OpenAI API. https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api (retrieved on 1.10.2025)

ISO 704:2009, *Terminology work — Principles and methods*

Informations pratiques

Lieux Nancy | Campus Lettres et Sciences Humaines | ATILF | Bâtiment CNRS | Salle Paul Imbs & Bâtiment G | Salle G04

Organisation ATILF (CNRS – Université de Lorraine), axe de recherche *Lexique*

Soutiens ATILF | CNRS | Université de Lorraine, UFR ALL, UFR SHS, Programme ORION, Projet IMPACT LUE INSIGHT | Linguist List | Ortho Editions | Nouveaux Cahiers d'Allemand

Contact lexiques@sciencesconf.org

Web https://lexiques.sciencesconf.org
https://www.atilf.fr/recherche/manifestations/colloques/20261216-colloque-lexiques



: ENTRÉE CAMPUS
: POINT D'ACCUEIL / LOGE
BU : BIBLIOTHÈQUE UNIVERSITAIRE
: POINT DE RESTAURATION
: SOIP (Service Orientation Insertion Professionnelle)
: MAUL (Musée Archéologique de l'Université de Lorraine)
: AMPHITHÉÂTRE
MDE : MAISON DE L'ÉTUDIANT
ATILF : UNIVERSITÉ DE LORRAINE / CNRS
Laboratoire d'Analyse et Traitement Informatique de la Langue Française
Centre de documentation M. Dinet ❶ Site linguistique ❷ Site didactique des langues

: SERVICE DE SANTÉ
: SUAPS (Activités Physiques et Sportives)
WC : TOILETTES
: PORTE D'ACCÈS AUX BÂTIMENTS
: POINT DE RASSEMBLEMENT
: ASCENSEUR
: ENTRÉE POUR PERSONNES À MOBILITÉ RÉDUITE

