

Beyond single words: the Interplay of phraseology, Multi-Word Units (MWU) and Complex Noun Phrases (CNP) in specialised discourses and AI-assisted translation

720, Olympe de Gouges Building
Grands Moulins Campus, Paris Cité University

December 11, 2025

Acronym: **BeyondSingleWords2025**

URL: bsw2025.sciencesconf.org

Research on phraseology, multi-word units (MWU, or more broadly, multi-word expressions), and complex noun phrases (CNP), which is conducted by linguists in French and in English, but also in Spanish, Italian, and other languages, converges on the fact that constrained linguistic forms are likely to contribute to an overall perception of human language, and that they can be systematised, to which Sinclair (1991) refers as “semi-preconstructed phrases”. From the very beginning, corpus linguistics has provided tools for identifying recurrent patterns perceived as units of meaning, including collocational phenomena (Sinclair, 1991; Gledhill, 2000), multi-word units (Granger & Meunier, 2008; Tutin, 2018), and complex noun phrases (Maniez, 2020; Kübler et al., 2022). Although these phenomena are often studied separately due to their inherent complexity, they form a continuum. The “BeyondSingleWords2025” workshop aims to examine them together in order to discuss their shared characteristics, particularly those standing out in specialised discourse studies.

In phraseology, key research areas include disciplinarisation (Burger, 2007; Mieder, 2009; Legallois & Tutin, 2013), systematisation (Mel’čuk, 2012, 2023), formal description (Gross, 1996; Moon, 1998), the theorisation of fixed expressions (Mejri, 1997; Gries, 2008), and the role of phraseology in language teaching (Berlage, 2014; Cavalla, 2020; Nation, 2022).

Regarding complex lexical units or multi-word units (also referred to as “séquences polylexicales” in French), research mainly focuses on their decoding and formalisation. This notion fosters interdisciplinarity (Granger & Meunier, 2008), which leverages new technologies to characterise and identify these forms by using corpora (Benetti, 1995; Sag et al., 2002; Savary et al., 2017; Pastor & Colson (eds.), 2020; Wahl & Gries, 2020). However, the boundaries between complex lexical units or MWU and CNP remain ambiguous (Mejri (ed.), 2004; Gries, 2022).

Complex noun phrases, extensively studied in English grammar (Quirk, 1985; Biber, 1999; Keizer, 2007; Schlücker, 2019), raise issues in translation and reveal structural and variational complexity in the use of language (Maniez, 2020; Kübler et al., 2022).

Specialised discourse serves as a prime field of observation for these linguistic forms, at the intersection of phraseology, terminology, and discourse analysis, while interdisciplinary research efforts highlight the interest and complexity of this phenomenon. The “BeyondSingleWords2025” workshop seeks to address these issues within the framework of research on specialised languages and AI-assisted specialised translation.

Research in phraseology and related fields is indeed a highly dynamic area, as a large number of conferences and workshops dedicated to these units of language and meaning demonstrate. Some recent and upcoming events include: “EUROPHRAS 2025”; “Interdisciplinary Approaches to Phraseological Units in World Languages: Linguistics – NLP & AI – Translation – Literature” (2025); “PHRASEOPRAG 2024: Pragmatic Phraseology in Oral and Mediated Interactions: Cross-Perspectives and Interdisciplinary Insights”; “Study Days on Constructional Approaches and Phraseology” (2024); “Phraseology Extraction: From Corpora to Deep Learning” (2022); International Conference “PhraséoTerme”, held at the University of Verona (2020), etc.

Researchers from the ALTAE laboratory (formerly CLILLAC-ARP) wish to propose a joint discussion about these units in specialised languages, following their workshop on generic collocations held in 2018: [Journée d'étude consacrée aux collocations génériques](#). The upcoming workshop will highlight the need to question these units, particularly in view of technological evolutions and the emergence of LLM, which are currently reshaping both linguistic research and language users' behaviors.

Proposals for communications on phraseology and its relations with multi-word units and/or complex noun phrases namely in specialised language and translation may concern the following themes or areas:

- contribution of the descriptions in specialised languages to global understanding and to theorisation of collocational phenomena
- descriptive and/or formal analysis in different fields
- contrastive analysis
- terminological and phraseological databases
- automatic recognition and extraction
- evaluation and processing in machine translation (NMT or generative AI)



Les travaux sur la phraséologie, les unités polylexicales (en anglais plus souvent désignées par Multi-Word Units, ou de manière plus large, Multi-Word Expressions) et les groupes nominaux complexes (en anglais Complex Noun Phrases), chez les linguistes français comme anglais, mais aussi espagnols, italiens, etc., convergent vers le constat que les formes à caractère contraint participent d'une perception globale de la langue et sont systématisables, ce que Sinclair (1991) appelle les préfabriqués de la langue. En effet, la linguistique de corpus a permis, dès le départ, de reconnaître des « schémas récurrents » (*recurrent patterns*) perçus comme des unités de sens, entre autres les phénomènes collocationnels (Sinclair, 1991 ; Gledhill, 2000), les unités polylexicales (Mejri, 1999 ; Granger & Meunier, 2008, Tutin, 2018) ou encore les groupes nominaux complexes (Maniez, 2020 ; Kübler et al., 2022). Ces phénomènes, souvent étudiés de manière isolée, en raison de leur complexité inhérente, constituent néanmoins un continuum que la journée d'étude *BeyondSingleWords2025* se propose d'aborder ensemble, afin de discuter de leurs caractéristiques communes, notamment celles qui ressortent des études des discours spécialisés.

En phraséologie, la disciplinarisation (Burger, 2007 ; Mieder, 2009 ; Legallois et Tutin 2013), la systématisation (Mel'čuk, 2012, 2023), la description formelle (Gross, 1996 ; Moon, 1998), le processus du figement et sa théorisation (Mejri, 1997 ; Gries, 2008) ainsi que la phraséologie dans l'enseignement (Berlage, 2014 ; Cavalla, 2020 ; Nation, 2022) constituent les axes majeurs de recherche.

Concernant les unités lexicales complexes (« séquences polylexicales » en français), les travaux se concentrent principalement sur leur décodage et leur formalisation. Cette notion ouvre une interdisciplinarité (Granger & Meunier, 2008) qui mobilise de nouvelles technologies pour caractériser et identifier ces formes à partir de corpus (Benetti, 1995 ; Sag et al., 2002 ; Savary et al., 2017 ; Pastor & Colson (éd.), 2020 ; Wahl & Gries, 2020). Toutefois, les frontières entre les unités lexicales complexes, d'un côté, et les groupes nominaux complexes, de l'autre, restent floues (Mejri (éd.), 2004 ; Gries, 2022).

Les groupes nominaux complexes, largement étudiés en grammaire anglaise (Quirk, 1985 ; Biber, 1999 ; Keizer, 2007 ; Schlücker, 2019), posent des difficultés en traduction et révèlent une complexité structurelle et variationnelle de l'emploi du langage humain (Maniez, 2020 ; Kübler et al., 2022).

Le discours spécialisé constitue un terrain d'observation privilégié pour ces formes linguistiques, à l'intersection de la phraséologie, de la terminologie et de l'analyse du discours. Des tentatives de recherche, souvent pluridisciplinaires, révèlent l'intérêt et la complexité de ce phénomène. La journée d'étude *BeyondSingleWords2025* a pour objectif d'aborder ces

phénomènes dans le cadre des recherches en langues de spécialité et la traduction spécialisée assistée par l'IA.

Les nombreuses journées d'études et conférences consacrées à ces unités de langue et de sens témoignent de la vivacité et de la richesse des recherches en phraséologie et dans les domaines connexes, comme en témoignent les manifestations récentes ci-dessous, dont certaines ont lieu cette année :

EUROPHRAS 2025, Approches interdisciplinaires des unités phraséologiques (UP) dans les langues du monde : Linguistique - TAL & IA - Traduction - Littérature, en 2025, *PHRASEOPRAG - La Phraséologie pragmatique des interactions orales et médiées : Regards croisés et perspectives interdisciplinaires*, journées d'étude *Approches constructionnelles et phraséologie* en 2024, *Phraseology Extraction: From Corpora to Deep Learning* en 2022 ou encore le colloque international *PhraséoTerme* tenu à l'Université de Vérone en 2020.

Les chercheurs du laboratoire ALTAE (ex-CLILLAC-ARP) souhaitent ainsi revenir sur les problématiques soulevées par ce type d'unités en langues de spécialité, après la [journée d'étude consacrée aux collocations génériques](#) qu'ils avaient organisée en 2018, pour mettre en lumière le besoin de continuer à interroger ce niveau d'analyse en langue, notamment au vu des évolutions technologiques et de l'apparition des LLM, qui bouleversent tant la recherche que le comportement des usagers de la langue.

Les propositions de communication sur la phraséologie et ses relations avec les unités polylexicales et/ou les groupes nominaux complexes spécifiquement en langues de spécialité et traduction spécialisée pourront porter sur les thématiques ou les axes suivants :

- apport de la description en langues de spécialité à la compréhension globale et à la théorisation des phénomènes
- analyse descriptive et/ou formelle dans différents domaines
- analyse contrastive
- bases de données terminologiques et phraséologiques
- reconnaissance et extraction automatique
- évaluation et traitement en traduction automatique (TAN ou IA générative)

Modalités de soumission

Les propositions, rédigées en français ou en anglais, comprendront un résumé de 300 à 400 mots ainsi qu'une bibliographie.

Elles devront être déposées en docx ou en pdf sur : <https://bsw2025.sciencesconf.org/>.

Langues de la journée d'étude : le français et l'anglais

Bibliographie

- Adamson, Sylvia & Victorina González-Díaz (2009). "History and structure in the English noun phrase : introduction". In : Transactions of the Philological Society 107, 3, pp. 255-261.
- Benetti, Laurence. (1995). « La structure des syntagmes nominaux complexes : problèmes de reconnaissance ». In: Scolia : Sciences Cognitives, Linguistiques et Intelligence Artificielle, 5, actes des rencontres linguistiques de BEBEFRI-Strasbourg, Neuchâtel, pp. 13-30.
- Berlage, Eva. (2014). *Noun Phrase Complexity in English. Studies in English Language*. United Kingdom: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Burger, Harald, Dmitrij Dobrovolskij, Peter Kühn & Neal R. Norrick. (2007). *Phraseology: an international handbook of contemporary research*, De Gruyter Mouton.
- Cavalla, Cristelle & Dominique Legallois. (2020). « Caractériser et identifier les unités phraséologiques pour leur enseignement ». Action Didactique, Enseignement des expressions préfabriquées, 6, pp. 12-30.
- Colson, Jean-Pierre. (à paraître). « Le traitement informatique du défigement : des corpus à l'intelligence artificielle ». Le français moderne, 93, 1.
- Condamines, Anne. (2005). « Linguistique de corpus et terminologie ». In: Langages, 39^e année, 157. La terminologie : nature et enjeux, sous la direction de Loïc Depecker. pp. 36-47.
- Corpas Pastor, Gloria. and Jean-Pierre Colson (eds.). (2020). *Computational Phraseology*, John Benjamins Publishing Company.
- Corpas Pastor, Gloria & Ruslan Mitkov (eds.). (2022). *Computational and Corpus-Based Phraseology*, Springer, Europhras.
- Frérot, Cécile & Mojca Pecman. (2021). *Des corpus numériques à l'analyse linguistique en langues de spécialité*. UGA Éditions.
- Gledhill, Christopher. (2000). *Collocations in science writing*. Gunter Narr Verlag Tübingen, 22, 270 p.
- Granger, Sylviane & Meunier, Fanny (eds.). (2008). *Phraseology: an interdisciplinary perspective*, Amsterdam & Philadelphia: John Benjamins.
- Gries, Stefan Th. (2022). "Multi-word units (and tokenization more generally): a multi-dimensional and largely information-theoretic approach", Lexis [Online], 19. <https://journals.openedition.org/lexis/623>.
- Gross, Gaston. (1996). *Les expressions figées en français : noms composés et autres locutions*, Ophrys.
- Gross, Maurice. (1982). « Une classification des phrases 'figées' du français ». Revue québécoise de linguistique, 11, 2, pp. 151-185.
- Keizer, Evelien. (2007). *The English Noun Phrase. The Nature of Linguistic Categorization*. Cambridge University Press.
- Kübler, Natalie, Alexandra Mestivier & Mojca Pecman. (2022). "Using Comparable Corpora for Translating and Post-Editing Complex Noun Phrases in Specialised Texts: Insights from English-to-French Specialised

- Translation". In: *Extending the Scope of Corpus-Based Translation Studies*. Bloomsbury Academic.
- Legallois, Dominique & Agnès Tutin. (2013). « Présentation : Vers une extension du domaine de la phraséologie ». *Langages, Vers une extension du domaine de la phraséologie*, 189, 1, pp. 3-25.
- Maniez, François. (2020). « Chapter 18: The identification of potentially ambiguous noun phrases in scientific English: a crucial aspect of translator ». In: *Strategies and Analyses of Language and Communication in Multilingual and International Contexts*. Cambridge Scholars Publishing, pp. 187-195.
- Mejri, Salah. (1997). *Le figement lexical : descriptions linguistiques et structuration sémantique*, Publications de la Faculté des lettres de la Manouba, série linguistique, volume X, 632 p.
- Mejri, Salah. (1999). « Unité lexicale et polylexicalité », *Linx*, 40, pp. 79-93.
- Mejri, Salah (ed.). (2004). *Polysémie et Polylexicalité*. *Syntaxe & Sémantique*, 5, Presses universitaires de Caen.
- Mel'čuk, Igor A. (2023). *General Phraseology: Theory and Practice*, *Linguisticæ Investigationes, Supplementa 36*, Amsterdam / Philadelphie, John Benjamins Publishing Company, 281 p.
- Mel'čuk, Igor A. (2012). "Phraseology in the language, in the dictionary, and in the computer", *Yearbook of Phraseology*, 3, pp. 31-56.
- Mieder, Wolfgang. (2009). *International Bibliography of Paremiology and Phraseology*. 2 vols. De Gruyter.
- Moon, Rosamund. (1998). *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Clarendon Press.
- Nation, Paul. (2022). "Finding and learning multiword units," in *Learning Vocabulary in Another Language*, Cambridge: Cambridge University Press, pp. 433-468.
- Pecman, Mojca. (2005). « De la phraséologie à la traductologie proactive : essai de synthèse des fondements théoriques sous-tendant la recherche en phraséologie ». *Meta*, 50, 4, <https://doi.org/10.7202/019853ar>.
- Quirk, Randolph. Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. (1985). *A Comprehensive Grammar of The English Language*, Longman.
- Sag, Ivan A, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. (2002). "Multiword Expressions: A Pain in the Neck for NLP". *Computational Linguistics and Intelligent Text Processing, Third International Conference, CICLing 2002, Mexico City, Mexico*, pp. 1-15.

Comité d'organisation

Lichao Zhu (Université Paris Cité, France)
Natalie Kübler (Université Paris Cité, France)
Mojca Pecman (Université Paris Cité, France)

Comité d'accueil

Anastasia Buturlakina
José Cornejo Carcamo
Agnès Ganet
Virginia Laconi
Xiaolin Ma
Joachim Minder
Anastasia Toropko

Comité scientifique

Gloria Corpas Pastor (Université de Grenade, Espagne)
Patrick Drouin (Université de Montréal, Canada)
Pascaline Dury (Université Lyon Lumière 2, France)
Pamela Faber (Université de Grenade, Espagne)
Paolo Frassi (Université de Vérone, Italie)
Cécile Frérot (Université Grenoble Alpes, France)
Christopher Gledhill (Université Paris Cité, France)
Clive Hamilton (Université Paris Cité, France)
John Humbley (Université Paris Cité, France)
Olivier Kraif (Université Grenoble Alpes, France)
Natalie Kübler (Université Paris Cité, France)
Marie-Claude L'Homme (Université de Montréal, Canada)
François Maniez (Université Lyon Lumière 2, France)
Salah Mejri (Université Sorbonne Paris Nord, France)
Luis Meneses-Lerín (Université d'Artois, France)
Alexandra Mestivier (Université Paris Cité, France)
Stéphane Patin (Université Paris Cité, France)
Mojca Pecman (Université Paris Cité, France)
Beatriz Sanchez Cardenas (Université de Grenade, Espagne)
Agnès Tutin (Université Grenoble Alpes, France)
Lichao Zhu (Université Paris Cité, France)

Calendrier

Date limite de soumissions : 15 juin 2025
Notification aux auteurs : début septembre 2025