Translating Variation in Multi-Word Terms: Is the human going to stay in the loop?

Natalie Kübler
(invited speaker)

ABSTRACT

This study examines terminological variation in specialised translation, with a focus on multi-word terms (MWTs), complex noun phrases (CNPs), and the specific challenges they pose in Earth and Planetary Sciences (EPS). Drawing on established typologies, the distinction between denominative and conceptual variation is explored, along with their linguistic, cognitive, and discursive causes. The analysis highlights the limitations of existing terminological resources, particularly regarding MWTs, which are often underrepresented in term bases. Empirical data show that both NMT and generative language models frequently introduce inappropriate variants and errors when processing MWTs and CNPs. Classroom-based translation and post-editing experiments confirm a high frequency of terminological errors, especially those involving misidentification of semantic heads or literal translations. Term extraction experiments using ChatGPT-4o and Termostat© demonstrate inconsistencies in automatic outputs, necessitating manual validation and corpus-based verification. These findings underscore the need for corpus-informed translation practices and terminological training that integrates both human expertise and computational tools. The study contributes to the development of structured pedagogical approaches for terminology work in translation training and addresses current limitations in the automated handling of scientific terminology in multilingual contexts.

0 Introduction

This paper offers reflections, illustrated with various examples, on how Generative AI (GenAI) can support translators in handling specialised translation, with a focus on the complex issue of translating multi-word term (MWT) variation. After outlining the concept of term variation and its causes and types, we will discuss the question of MWT variation in specialised translation. The advantages and limitations of Neural Machine Translation (NMT) and GenAI for specialised translation are then examined. Finally, the paper presents two experiments on how GenAI, specifically ChatGPT 4o addresses specialised translation and the processing of MWTs.

1 Types of Term Variation

Over the past two decades, interest in terminological variation has grown significantly, with considerable advances in the field (Humbley & Picton, 2017). The advent of digital corpora and corpus-linguistic methodologies enabled the systematic observation of recurring variation in texts. Terminologists, once domain experts with linguistic inclinations, have increasingly become linguists or translators—language specialists rather than subject-matter

experts. This shift has led to the rise of textual terminology (Bourrigault & Slodzian) and has elevated the role of corpora in extracting specialised, context-sensitive data (Pecman & Kübler, 2022), particularly through Knowledge-Rich Contexts (Meyer, 2001). As a result, research has shifted from suppressing variation to analysing it as a legitimate and meaningful phenomenon.

Term variation can be classified according to form (denominative) and meaning (conceptual). Denominative variation, defined by Sager (1990) as the use of multiple terms for a single concept, often reflects synonymy and is linked to neonymy (Gledhill & Pecman, 2018). Cabre (1999, p. 250) describes it as formal instability—morphological, graphical, or phonetic—occurring in lexical structures. This phenomenon is now well-documented, particularly in neology, where term creation and disappearance frequently result in terminological proliferation and synonymic variation (Dury et Picton, 2009; Pantazara & Tziafa, 2019).

Daille (2017) identifies several causes of denominative variation, including communicative needs, conceptual evolution, and user preferences. Freixa (2006) categorises dialectal causes into diatopic, diastratic, and diachronic variation. Diachronic variation, closely tied to neology, includes primary/secondary (Sager, 1997) and origin/transfer neology (Rondeau, 1984 in Humbley, 2012). Functional, discursive, interlinguistic, and cognitive factors further account for variation (Freiwa, 2006). Conceptual variation does not always entail denominative change, as shown by Pecman (2025) with *data pollution*, where the meaning shifted despite lexical stability.

Additional types include rhetorical (Pecman, 2012), emotional (Condamines, 2017), and ideological variation (Rossi, 2017). These diverse forms underscore the multifaceted, multifunctional nature of terminological variation, which has been enriched by broader linguistic research.

2 Variation in specialised translation

In linguistic practice, term variation creates uncertainty for individuals (researchers, translators, non-specialist writers) and organisations, particularly during company mergers (Gaudin, 1994; Alarcón-Navío et al., 2016; De Vecchi, 2012). Several studies in specialised translation—such as Jimenez-Crespo & Tercedor-Sanchez (2017), Sabela-Fernandez & Kerremans (2011), and Li (2023)—highlight that translating term variation is a key issue, often amplified in the translation process.

To manage term variation, translators typically rely on term bases. However, these

resources are rarely exhaustive or up to date, especially for multi-word terms (MWT), which are more variable and harder to standardise. For instance, the established EPS term *true polar wander* was missing from several major databases. According to Cabezas-García & León Araúz (2020), MWTs are less lexicalised and behave like phraseological units. Their link to neology may explain their underrepresentation in standard terminological resources, pointing to a need for enrichment. Translation challenges also stem from contrasting views on equivalence: while term bases assume direct term-to-term matches, translation practice favours functional, context-based solutions. For example, the English term *creep* has several French equivalents in EPS—*fluage*, *reptation*, *cryptosolifluxion*—each context-dependent. Corpus analysis confirms additional variants like *glissement* or *déformation asismique lente*, revealing the complexity of establishing reliable equivalence.

Finally, a single-word term with multiple equivalents can generate compound terms, but not all equivalents are usable in new formations. For instance, *dislocation creep* becomes *fluage par dislocation*, not *reptation par dislocation*. Studies (e.g. Bowker 1998; Fernandez-Silva & Kerremans 2011; Giacomini, 2018) confirm that MWTs are more prone to variation than single-word terms, as each component can shift or expand in form.

As previously noted, the translation process further amplifies the challenges associated with term variation. Our experience in annotating error in Learner Translation Corpora (LTC) indicates (Kübler et al. 2016, 2018, 2022, 2024; Minder et al., 2025) that, in the context of specialised translation, terminological errors represent the most frequent error type. Several factors contribute to the high incidence of terminological errors. Firstly, students often struggle to identify MWTs, as well as to understand the internal semantic relations among their constituents. In some cases, they fail to correctly identify the head of the term, leading to meaning distortions. When no established French equivalent exists for an English term, students frequently resort to literal translation, which further contributes to terminological errors.

An additional difficulty lies in the fact that MWT are often part of complex noun phrases (CNPs, Kübler et al. 2018, 2022), which makes it difficult for translators to find the correct equivalent and to process term variation. The increasing use of machine translation tools, whether NMT or GenAI, brings back these long-standing questions that the translation community has faced in relation to machines (Kübler et al. 2022, Cabezas-García & León-Araúz, 2023). Over the last decades, much has been done to specialise machine translation systems with terminological resources (Semenov et al. 2023). However, MWT terms processing has received little attention and machines are not consistent while translating

terms, especially in CNPs, thus introducing inappropriate variation.

3 The advent of AI in specialised translation

NMT employs deep learning models to produce translations by predicting word sequences based on large bilingual corpora. NMT typically yields fluent results but often struggles with domain-specific terminology. The emergence of genAI marks a significant evolution in this field. Unlike NMT, generative models use contextual “reasoning” to “interpret” meaning, rather than aligning words directly, enabling them to handle complex instructions and multi-step tasks. This expanded capacity introduces both new possibilities and limitations in the context of specialised translation. GenAI for example, can extract terms from a source text and obtain their equivalents in another language, thereby processing bilingual terminology upstream of the translation process. Thanks to prompts and Chain-of Thought Prompting, it can produce translations that appear natural and fluent, yet may still contain inaccuracies or errors. It can also be employed to evaluate translated texts, although its outputs may contain errors and inconsistencies (Minder et al. 2025). The comparison of translations produced by NMT and GenAI systems reveals that, while both can generate accurate output, human intervention remains necessary to ensure precision and consistency. For instance, Farrell (2023) compares three systems: two NMT tools (DeepL and Google Translate) and one GenAI model (ChatGPT-4). The results indicate that DeepL outperforms Google Translate, which in turn performs better than ChatGPT. However, the evaluation relies solely on automatic metrics and does not provide insight into the specific translation issues encountered. There are few studies evaluating specialised translations produced by GenAI, and even fewer that specifically address terminological challenges. While Lyu et al. (2023) suggest that fine-tuning such systems can help resolve terminology-related issues, Siu (2024) demonstrates that these problems persist.

We conducted an experiment to compare DeepL and ChatGPT-4o in translating an excerpt from a research article in Earth and Planetary Sciences. One of the sentences included the term *moonquake*, a neologism in English formed by analogy with *earthquake*. The correct French equivalent is *séisme lunaire*. In French scientific discourse, the standard term for *earthquake* is *séisme*, whereas *tremblement de terre* is typically used in general language. Unlike English, which uses *earthquake* across both registers, French maintains this distinction for different registers. The following example illustrates that DeepL produced a more accurate translation of *moonquake* than ChatGPT-4o.

Table 1: Translation of *moonquake* by DeepL and ChatGPT

Source text:	*a limiting distance beyond which a **moonquake** and a fault are considered unrelated must be established*.
MT Chat GPT:	*il est nécessaire d'établir une distance limite au-delà de laquelle un ***tremblement de lune** et une faille sont considérés comme non liés.*
MT DeepL:	*il faut établir une distance limite au-delà de laquelle un **séisme lunaire** et une faille sont considérés comme n'étant pas liés.*

As evocative as the translation *tremblement de lune* may be in French, it is not an established term. Unlike *tremblement de terre*, which is the conventional expression for *earthquake*, no analogous construction exists in French to form *tremblement de lune*. A second experiment was carried out in which ChatGPT was asked to evaluate the translations provided by DeepL. ChatGPT was not asked to evaluate its own translations, as a previous study (Minder et al., 2025) demonstrated that self-evaluation presents significant challenges. The following response was obtained for the translation by DeepL:

Table 2: ChatGPT’s comment on the translation of moonquake by DeepL

ChatGPT (about DeepL translation):	"Moonquake" Translation: The term "**moonquake**" is translated as "**séisme lunaire**", which is correct **but loses the specific connotation of 'moonquake'.** A more direct translation like "**tremblement de lune**" could have been considered, though "séisme lunaire" is not incorrect

In this instance, the system offers no meaningful assistance and risks misleading the specialised translator seeking support.

After a reminder of term, and especially, MWT variation, we will tackle the issue of translating term variation with examples in English and French. We will then address the issue of evaluating how TAN and GenAI translate terminological variation and MWTs, particularly within complex noun phrases. Finally, we will show some possible solutions, especially on a human-in-the loop approach of GenAI.

4 Translating Multi-Word terms and Complex Noun Phrases

Research on the machine translation of complex noun phrases (CNPs) remains limited, especially when these structures include specialised terminology. Similarly, the automatic translation of MWTs has not been widely explored. In a classroom experiment (Kübler et al., 2022), we examined the challenges faced by students translating CNPs in Earth and Planetary Science texts, producing an error typology based on annotated student

translations. From a different angle, Cabezas-García and León-Araúz (2023) analysed the translation of English MWTs into Spanish by machine translation systems and developed an error typology specific to MWTs. Although the two studies adopt different perspectives — Kübler et al. (2022) analysing student translations and Cabezas-García and León-Araúz (2023) examining machine-generated output—their error typologies show striking similarities. This overlap suggests that both human translators and machine translation systems tend to produce comparable errors with terminology-rich structures.

In another classroom-based study on student post-editing of Earth and Planetary Science research articles (Kübler, Mestivier, Pecman, in preparation), the general error typology we developed (Kübler, 2008; Castagnoli et al. 2011; Kübler et al. 2024) was applied to both the machine-translated texts and the post-edited versions. The translations were produced using DeepL and ChatGPT-4o. As is typical in specialised domains, most errors were terminological in nature. Both systems struggled with complex noun phrases containing specialised terms, although the difficulties did not always occur in the same cases. The students’ post-edited texts were also annotated, revealing the same type of terminological and CNP errors; instances of both overconfidence and underconfidence in the output of MT systems were observed, depending on their level of language, findings that align with those reported in Kübler et al. (2022) & Schuhmacher (2020, 2023).

Qualitative analysis shows that single-word terminological variants are a frequent source of error in the translation of MWTs. Table 3 presents a complex noun phrase containing two MWTs, where individual components were mistranslated by both DeepL and ChatGPT. However, the machine-translated output was correctly revised during post-editing.

Table 3: erroneous literal term in FR, correctly modified in the post-editing process

Source text:	*The resulting structure is bounded by **low-angle thrusting** along the external faults*
MT DeepL:	*La structure résultante est délimitée par des ***poussées à faible *angle** le long des failles externes*
MT ChatGPT:	(…) **chevauchements à faible *angle**
PE:	*la structure résultante est délimitée par le **chevauchement à faible pendage** le long des failles externes*

Table 4 illustrates a case of a graphical denominative variant of a MWT that can alternate with an acronym in English.

Table 4: Incorrect MT, erroneously corrected

Source text:	*We also develop an algorithm to invert for the* ***[centroïd MT]*** *of the [VLPs]*
MT DeepL:	*Nous développons également un algorithme pour inverser *le* ***[centroïde MT]*** *des [****VLP]***
MT ChatGPT**:**	(…)[****moment *tensoriel *centroidal] (MT)*** *des [*VLP]*
Incorrect PE:	*Nous développons également un algorithme pour inverser les* ***[moments sismiques centroïdes]*** *des *[****signaux longue durée[***
Correct PE:	(…)[***moments sismiques centroïdes]*** *des* [***signaux à longue période]***

The MWT *centroid MT* appears as the head of a CNP that includes the prepositional phrase *des VLP* containing an acronym. Both acronyms have no French equivalent, as French employs the expanded forms: *moment sismique centroïde* for English *MT* and *signaux longue période* for English *VLP*. Neither the MT systems nor the post-editing student were able to produce a fully accurate translation for the English CNP. However, the student did identify the correct French expanded equivalent for *centroid moment tensor*, though they did not provide the appropriate equivalent for the acronym *VLP*.

As previously noted, terminological databases typically include single-word terms or, at most, two-terms, while entries exceeding two words are rare. However, the examples discussed above show that specialised texts frequently contain CNPs incorporating MWTs, in which individual lexical components may vary erroneously.

For two decades, our university has developed a corpus-based approach to specialised translation (Kübler 2003), drawing on earlier frameworks, such as Zanettin (1998) and Bowker (1998). As part of this approach, students engage in a year-long project centered on the final translation of a research article in EPS. During the first term, they compile a comparable English-French corpus related to the article’s subject area and conduct a detailed terminological analysis. By exploring the corpus, students are able to define domain-specific terms, identify appropriate term equivalents, and document usage contexts -- among other things (Kübler et al 2018). These findings are used to create term records, which are then entered into the ARTES database, a terminological and phraseological resource developed at our institution (Pecman & Kübler 2011, Kübler & Pecman 2012). Emphasis is placed on documenting MWTs to address gaps in official databases or emerging terms.

In the second term, students translate and post-edit the selected articles. They also participate in numerous classroom activities focusing on using corpora to resolve specific specialised translation challenges, including the translation of complex noun phrase. In Kübler et al. (2016), we analysed translation produced with and without corpus use; our results indicated that corpus use significantly improved the students’ translation. The various experiment we conducted later have shown (2018, 2022, 2024) that corpus use can help students improve their translations or their post-edited texts. These findings are important as they demonstrate how the development of corpus competence enables translators to solve problems that NMT or GenAI cannot resolve effectively, thereby reinforcing the essential role of human involvement in the translation process.

Taking this into account, we are currently working on providing term variation information to LLMs in order to better translate term variation. This is part of an ANR-funded project, the MaTOS[1] project, whose aim consists in improving the translation by LLMs of complete research papers in Natural Language Processing, Earth and Planetary Sciences, and Medicine, in order to better disseminate scientific information in languages other than English (cf. The Helsinki Initiative on Multilinguism in Scholarly Education[2]). Preliminary results show that term variation is very common in EPS for example, and generates difficulties for machine translation by GenAI.

We have demonstrated that translations generated by genAI cannot be considered more reliable than those produced by NMT. This applies to the overall translation of the text, as well as to the translation of MWTs, complex noun phrases, and term variation—three interrelated phenomena. However, it is possible to develop equivalent terminology between English and French through an iterative process involving constant interaction between the human translator and the system. In a pilot study aimed at developing a teaching protocol for using such systems in terminological validation and definition building, Pecman (2025) shows how they can support the understanding of the evolving meaning of an existing term and assist in formulating comprehensive definitions.

This work on understanding terms is part of the initial phase of translating a text, which involves understanding the subject area by analysing the terminology and identifying equivalents in the target language, including collocations and variants (Pecman & Kübler

1 MaTOS (Machine Translation for Open Science) is a project funded by the French National Research Agency : https://anr-matos.github.io/

2 https://www.helsinki-initiative.org/en

2011). Siu (2023) shows how ChatGPT and ChatGPT-4 can be used to extract terms and their definitions. The demonstration has two main limitations. First, the examples are taken from very short medical texts; extracting terminology from a full research article or a larger corpus remains a challenge (Giacomini, 2025). Second, medicine is a field that has been widely studied in terminology and specialised translation, and the system may be drawing on existing term banks and bilingual texts.
Using ChatGPT-4o, we prepared a bilingual English–French glossary with definitions to support the translation or post-editing of a research article in Earth and Planetary Sciences. The article contains 13 170 words.

Relying on White et al. (2023) prompt patterns classification, we provided the context in which the task was taking place (provided a definition of term with examples (few-shots) and explained the task in several steps:

Table 5: example of prompt for the term extraction task

You are a translator in specialised translation; you have to translate a research article in earth and planetary science. The article is highly specialised and contains many terms. A term is a unit of meaning in a particular domain, that correspond to a concept and has a specific definition. A term can be composed of one single word or of several words. Here are a few examples in earth and planetary science, in English: volcano, mantle wedge, active source refraction seismology; in French: croute, volcan effusif, élément réfractaire lithophile, fluide issu de la déstabilisation. Extract all the terms from the uploaded article, find their equivalent in French by searching in academic documents. Create a three-column table: First column: English term; second column: French equivalent; third column: definition in English.

The results show that the system has difficulty extracting terms from an entire article. This is a common issue when translating full documents (Peng et al., 2025). To address this limitation, the article was divided into five parts, which were submitted one after the other. Follow-up questions were used to obtain more terms. A sample of the first 39 terms, corresponding to the first section (2300 words), was analysed. Table 6 presents the results.

Table 6: results with examples

Term candidates in English	39	
Erroneous French equivalent	9	*seismic moment release> *liberation de moment > moment sismique relâché*
Not a term, but collocations	5	*triggered SSE > *évènement de glissement lent déclenché > glissement lent a été déclenché*
Not a term, not a collocation	1	**fluid pressure change : change* is found several lines later
The term exists, but is not present in the text	2	*fault mechanics*: the term in the text is *mechanic of faults*

The sample is limited, but it highlights the challenges a translator would face when using the results without further processing.

We compared ChatGPT’s results with Termostat© (Drouin, 2003[3]), a tool designed for term extraction, 1,443 term candidates (both single-word and multi-word units) were identified from the full article, and 331 from the first section. In both cases, the results require manual validation. All of the suggested terms were checked against the article and a comparable corpus in Earth and Planetary Sciences, compiled at Université Paris Cité (Kübler et al. 2016, 2018, 2022), to confirm or reject the term candidates and their French equivalents. Corpus analysis confirmed that the system was not entirely reliable and helped identify suitable translations.

5 Conclusion

This study has examined the challenges of translating multi-word term (MWT) variation in specialised domains such as Earth and Planetary Sciences. The findings show that neither Neural Machine Translation (NMT) nor generative models consistently manage terminological variation, particularly within complex noun phrases. While such tools can support term extraction and glossary building, their output often lacks precision and requires manual validation. Effective specialised translation still depends on corpus-informed, human-in-the-loop approaches to ensure accuracy and consistency.

References

Gledhill, C., & Pecman, M. (2018). On alternating pre-modified and post-modified nominals such as aspirin synthesis vs. synthesis of aspirin: Rhetorical and cognitive packing in English science writing. *Fachsprache-Journal of Professional and Scientific Communication*, *40*(1+ 2), 24-46.

Cabre Castellví, M. T. (1999). *Terminology: Theory, methods and applications* (Vol. 1). John Benjamins Publishing.

Pantazara, M., & Tziafa, E. (2018). Les termes de la crise économique grecque dans les corpus. *Meta*, *63*(3), 739-765.

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

Daille, B. (2017). Term variation in specialised corpora. Amsterdam: John Benjamins

Gaudin, F. (1994). L'insécurité linguistique des scientifiques. A propos d'une enquête socioterminologique. *Cahiers de lexicologie*, (65), p-47.

Freixa, J. (2006). Causes of denominative variation in terminology: A typology proposal. *Terminology. International Journal of Theoretical and Applied Issues in Specialised Communication*, *12*(1), 51-77.

[3] https://termostat.ling.umontreal.ca/index.php

Sager, Juan Carlos. 1997. Term formation. In Sue Ellen Wright and Gerhard Budin (eds.), Handbook of Terminology Management, vol 1, 25–41. Philadelphia: John Benjamins.

Humbley, J. (2012). Retour aux origines de la terminologie: l’acte de dénomination. *Langue française*, (2), 111-125.

Pecman, Mojca (in press) Drafting definitions for emerging concepts and terms undergoing semantic shift within the ARTES knowledge base: A protocol for integrating LLMs into terminological analysis by experimental approach. Terminologija 32 (2025).

Condamines, A. (2017). The emotional dimension in terminological variation. Multiple perspectives on terminological variation. John Benjamins Publishing Company. Pp. 11-18

Rossi, M. (2017). Chapter 8. Terminological metaphors and the nomadism of specialised terms: Some observations on intralinguistic and interlinguistic variation. In *Multiple Perspectives on Terminological Variation*. John Benjamins Publishing Company. Pp. 181-211

Alarcón-Navío, E., López-Rodríguez, C., & Tercedor-Sánchez, M. 2016. Variation dénominative et familiarité en tant que source d’incertitude en traduction médicale. *Meta*, *61*(1), 117-144.

De Vecchi, D. (2012). What do “They” Mean by That? *LSP Journal-Language for special purposes, professional communication, knowledge management and cognition*, *3*(2).

Jiménez-Crespo, Miguel A. & Maribel Tercedor. (2017) “Lexical Variation, Register and Explicitation in Medical Translation: A Comparable Corpus Study of Medical Terminology in US Websites Translated into Spanish.” TIS: Translation and Interpreting Studies 12: 3, pp. 405-426.

Fernández-Silva, S., & Kerremans, K. (2011). Terminological variation in source texts and translations: A pilot study. *Meta*, *56*(2), 318-335.

Li, B. (2023). Conceptual deviation in terminology translation: A case study on translating COVID-19 terminology in multilingual news media. *Terminology*, *29*(2), 351-382.

Cabezas-García, M., & León-Araúz, P. (2023). Machine versus corpus-based translation of multiword terms. *Digital Scholarship in the Humanities*, *38*(Supplement_1), i6-i16.

León-Araúz, Pilar & Cabezas-García M. (2020) “Term and translation variation of multi-word terms.” In: Mogorron Huerta, Pedro (ed.) 2020. Análisis multidisciplinar del fenómeno de la variación fraseológica en traducción e interpretación / Multidisciplinary Analysis of the Phenomenon of Phraseological Variation in Translation and Interpreting. MonTI Special Issue 6, pp. 210-247.

Bowker, L. (1998). Using specialised monolingual native-language corpora as a translation resource: A pilot study. *Meta*, *43*(4), 631-651.

Gledhill, C., & Pecman, M. (2018). On alternating pre-modified and post-modified nominals such as aspirin synthesis vs. synthesis of aspirin: Rhetorical and cognitive packing in English science writing. *Fachsprache-Journal of Professional and Scientific Communication*, *40*(1+ 2), 24-46.

Kübler, N., Mestivier, A., Pecman, M., & Zimina, M. (2016). Exploitation quantitative de corpus de traductions annotés selon la typologie d’erreurs pour améliorer les méthodes d’enseignement de la traduction spécialisée. *Actes des 13èmes Journées internationales d’analyse statistique des données textuelles (JADT 2016)*, 7-10.

Kübler, N., Mestivier, A., & Pecman, M. (2018). Teaching specialised translation through corpus linguistics: translation quality assessment and methodology evaluation and enhancement by experimental approach. *Meta*, *63*(3), 807-825.

Kübler, N., Mestivier, A., & Pecman, M. (2022). Using comparable corpora for translating and post-editing complex noun phrases in specialised texts. Syviane Granger & Marie-Aude Lefer. *Extending the Scope of Corpus-Based translation Studies*, Bloomsbury Publishing, p. 237-266

Kübler, N., Martikainen, H., Mestivier, A., & Pecman, M. (2024). Post-editing neural machine translation in specialised languages: The role of corpora in the translation of phraseological structures. In *Recent Advances in Multiword Units in Machine Translation and Translation Technology* (pp. 57-78). John Benjamins Publishing Company.

Semenov, K., Zouhar, V., Kocmi, T., Zhang, D., Zhou, W., & Jiang, Y. E. (2023, December). Findings of the WMT 2023 Shared Task on Machine Translation with Terminologies. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 663-671).

Minder, J., Wisniewski, G., & Kübler, N. (2025). Testing LLMs' Capabilities in Annotating Translations Based on an Error Typology Designed for LSP Translation: First Experiments with ChatGPT. *arXiv preprint arXiv:2504.15052*.

Farrell, M. (2023, July). Preliminary evaluation of ChatGPT as a machine translation engine and as an automatic post-editor of raw machine translation output from other machine translation engines. In *Proceedings of the International Conference HiT-IT 2023* (pp. 108-113). Incoma Ltd..

Lyu, C., Du, Z., Xu, J., Duan, Y., Wu, M., Lynn, T., ... & Wang, L. (2023). A Paradigm Shift: The Future of Machine Translation Lies with Large Language Models. *arXiv preprint arXiv:2305.01181*.

Siu, S. C. (2023). ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. *Available at SSRN 4448091*

Kübler, N. (2008), ‘A comparable Learner Translator Corpus: creation and use’, in P. Zweigenbaum (ed) *Proceedings of the Comparable Corpora Workshop of the LREC Conference*, 73-78, May 28-30 2008, Marrakech, Morocco.

Castagnoli, S., Ciobanu, D., Kunz, K., Kübler, N., & Volanschi, A. (2011). Designing a learner translator corpus for training purposes. *Corpora, language, teaching, and resources: From theory to practice*, *12*, 221-248.

Schumacher, P. (2020). Post-édition et traduction humaine en contexte académique : une étude empirique.Transletters : International Journal of Translation and Interpreting, (4), 239-274.

Schumacher, P. (2023). Traduction humaine et postédition : contrôle qualité en contexte académique. *Meta*, *68*(3), 510-536.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.

Peng, Z., Bawden, R., & Yvon, F. (2024). Investigating Length Issues in Document-level Machine Translation. *arXiv preprint arXiv:2412.17592*.

Natalie Kübler

Professeur UFR EILA, linguistique et traduction

Université Paris Cité

E-mail: natalie.kubler@u-paris.fr