DOI: 10.23817/strans.17-6

Received: 23.06.2025
Accepted: 10.02.2026

STUDIA TRANSLATORICA
2026 • vol. 17 105–121
ISSN 2084–3321 e-ISSN 2657-4802

Barbara Heinisch
Institute for Applied Linguistics, Eurac Research/ Italy

Leveraging large language models in lexicography, terminology and specialized translation

Abstract

Leveraging large language models in lexicography, terminology and specialized translation

The advent of large-language models (LLMs) such as GPT-5 has opened new avenues for research and practice in lexicography, terminology work and specialized translation. Terminology connects these three fields, serving as a backbone for precision and clarity in professional (multilingual) communication. This contribution explores how professionals can use LLMs to transform the management and application of terminology while addressing ethical and sustainability challenges.

In lexicography, LLMs can facilitate the creation and maintenance of specialized dictionaries by extracting and defining lemmas, identifying usage patterns and providing contextual examples. The contribution also examines how text-generative LLMs might reshape corpus design and challenge traditional notions of what constitutes 'authentic text' as more content is generated by means of LLMs.

In specialized translation, LLMs can support terminology management by suggesting equivalents and usage examples across languages, potentially improving efficiency and accuracy.

In terminology work, LLMs can process large datasets, identify patterns provide definitions, find variants, visualize concept systems and assist in building and updating (multilingual) terminology databases.

Overall, LLMs offer professionals new ways to approach their work, but professional expertise, critical judgment and established tools remain essential.

Keywords: dictionaries, terminology resources, terminology work, domain-specific terminology, language technology

© The author(s). This is an open-access article distributed under the CC BY-NC-ND 4.0 license.

1. Introduction

Large language models (LLMs), such as GPT-5, and particularly their chatbot interfaces, such as ChatGPT or DeepSeek, have revolutionized human-computer interaction and access to language-based AI, as they enable users to engage with complex models through natural language in an intuitive and user-friendly manner (Rudolph et al. 2023). Due to their capability of generating human-like texts (in addition to images, videos and sound, among others), they have impacted several fields and professions, including lexicography, terminology management and specialized translation.

Although the recent emergence of LLMs offers numerous new possibilities and modes of work, the integration of technology is by no means a novel phenomenon for lexicographers, terminologists or specialized translators. Technology was, and is, crucial in lexicography, specialized translation and terminology work. Terminologists rely on a set of tools, including corpus analysis tools (Picht 2001) for finding occurrences or frequencies of terms or for finding definitions or contexts, terminology extraction tools to extract candidate terms from texts (Zielinski/ Ramírez Safar 2005) or terminology management systems (Drewer/ Schmitz 2017) to compile, edit, store and disseminate terminology. Translators rely on a variety of tools as well (Bowker/ Corpas Pastor 2015), including computer-assisted translation tools, which are interoperable with a wide range of components, including terminology recognition, machine translation systems, alignment tools, but also quality assurance/checkers, project management and workflow tools (Bowker/ Corpas Pastor 2015: 872). Similar to translators and terminologists, also lexicographers use a wide range of tools in their work, and automation is (and was) an important aspect in (corpus-based) lexicography (Jakubíček/ Rundell 2023: 518; Klosa 2007: 106–115). This encompasses the creation of large corpora and the development of tools and methods for the automatic generation of word lists, example sentences or collocations. These tools were typically designed for specific tasks and were employed individually or in combination to support the drafting of complete dictionary entries, thereby enhancing the efficiency of the dictionary-making process (Jakubíček/ Rundell 2023: 518). Compared to LLMs, whose output is usually unpredictable (Herrera-Poyatos et al. 2025: 2), each of these systems, due to their "deterministic approach guarantees consistent and predictable outputs" (Massion 2024: 86).

Lexicography, specialized translation and terminology work are inherently interconnected through their shared focus on the analysis, organization and mediation of language (and domain-specific knowledge). While lexicography documents and structures lexical information (usually in general language), terminology work focusses on the language of a specialized domain, thereby working with terms in a concept-oriented manner. Specialized translators, in

turn, rely heavily on lexicographical and terminological resources to ensure accurate and contextually appropriate transfer of texts across languages. In addition, they also create these resources for themselves and take on the role of lexicographers or terminologists in individual steps throughout the translation process, e.g. when researching the meaning of (general) language words or trying to locate terms in the concept system of a domain.

The boundaries between (specialized) lexicography and terminology are blurred and debated (Bowker 2017: 138). While lexicography (mainly) deals with 'words', terminology focusses on 'terms'. Whereas lexicography (usually) targets lay people as users, terminology (often) addresses domain experts (Bowker 2017: 139). Another major difference between lexicography and terminology is their approach: Although terminology work (mainly) has an onomasiological approach ("starting from the analysis of concepts and then assign terms as designations of these concepts," Kageura 2015: 57), lexicography follows a semasiological one ("starting from the analysis of words and look for the senses or meanings", Kageura 2015: 57). The outputs provided also differ; while terminology work provides terminology databases (among others) in one or several specialized fields, the products of lexicography are dictionaries that focus on general language (Bowker 2017: 139).

As both address words (or terms) in context, corpus-based approaches are applied in both disciplines in order to obtain linguistic information from a body of text, which also means that both use the same tools (to a certain extent), e.g. corpus analysis tools (Bowker 2017: 143). According to Costa et al. (2023), lexicography and terminology are closely related disciplines, both concerned with the study of 'terms' and their meanings to support the understanding and dissemination of knowledge. While they share the same objects of study, they differ in methodology and underlying frameworks, as each responds to distinct social needs and produces lexical resources tailored to specific users and purposes (Costa et al. 2023: i18). Especially the user needs have led to "the rapprochement between lexicography and terminology" (Bowker 2017: 144).

Specialized translators must not only engage with the products of lexicography and terminology but are often required to undertake lexicographic and terminology work themselves to produce a translation. Moreover, they must possess knowledge of the relevant institutional and cultural systems, as well as an awareness of differences between countries and the historical evolution of terms. As Edelmann (2014: 200) argues, this complex interplay of linguistic and conceptual expertise necessitates that specialized translators also assume the role of terminologists, with a thorough understanding of terms and concept systems in both the source and target languages.

More broadly, the integration of LLMs offers new opportunities for professionals in lexicography, specialized translation and terminology work to enrich

their work by supporting tasks such as domain-specific research, terminology extraction and knowledge organization. When used critically, only when really necessary and in combination with established methods and tools, these technologies can enhance the quality, efficiency and scope of professional practice when supported by expert knowledge (of lexicography, terminology and specialized translation). At the same time, their use raises ethical considerations, including questions of sustainability and responsible use. Addressing these concerns requires practitioners to engage with LLMs thoughtfully and reflexively, ensuring that technological assistance remains aligned with professional standards and values.

The advent of LLMs has prompted a wave of experimental applications across various disciplines, including lexicography, specialized translation and terminology work. This study provides an overview of key insights emerging from these initial explorations and assesses the implications for current practices in these three fields. It should be noted, however, that the scope of documented studies likely represents only a fraction of the experiments currently being conducted in practice, many of which remain unpublished or are ongoing.

2. LLMs for lexicography, specialized translation and terminology work

Central to the effective use of LLMs is the design of prompts, which shape the model's output and determine its alignment with linguistic, domain-specific and communicative requirements. Apart from the LLM and the specific version used, the quality of the prompt is decisive for the quality of the LLM output. Effective prompt selection requires a nuanced understanding of user intent, model capabilities and domain specificity (Ekin 2023: 5). Crafting clear and precise prompts, tailored to the task and incorporating any necessary constraints, is essential for guiding the model toward producing more accurate and relevant outputs (Ekin 2023: 5). This gave rise to the development of prompting principles (Bozkurt 2024; Chen et al. 2023; Saleem 2024), ranging from general guidelines to prompt design frameworks (that might change from one model to the next). In the following, the results of studies focusing on the use of LLMs for purposes of lexicography, terminology work or specialized translation are discussed.

2.1. LLMs and lexicography

With the advent of LLMs, lexicographers have been evaluating the usefulness of ChatGPT and other LLMs for lexicographic purposes, leading to a wide range of results. According to Rundell (2024: 7–12), LLMs offer promising

applications for lexicography, particularly in generating definitions for straightforward or technical terms and in lexical profiling. Thus, writing dictionary definitions works well for simple (and common) terms but not for polysemous ones. However, the usefulness of LLMs is currently limited by issues such as inaccurate examples, poor handling of polysemous words, grammatical inconsistencies and hallucinated sources. While LLMs may assist in drafting dictionary content, their lack of transparency and unreliable output make them unsuitable as standalone tools for lexicographic work without rigorous post-editing and access to verifiable data sources.

Jakubíček and Rundell (2023) also used ChatGPT (GPT3.5 and GPT4) for a range of lexicographic tasks including generating definitions, identifying syntax patterns, handling grammatical information, determining word class and transitivity, providing example sentences and distinguishing marked items (e.g. formal or archaic words). ChatGPT demonstrated strong performance in distinguishing between antonyms and synonyms (Jakubíček/ Rundell 2023: 515). It performed relatively well in producing definitions (particularly for technical terms, which supports Rundell's (2024: 7–8) and Lew's (2023: 1) findings) but struggled significantly with example generation (since it produced examples following monotonous structures and subject areas), syntactic pattern identification (with mixed results) and word-sense disambiguation (which is one of the hardest activities in lexicography). Regarding word-sense disambiguation, the authors found "a similar tendency to split one sense unnecessarily while completely missing another common meaning" (Jakubíček/ Rundell 2023: 516). They also used ChatGPT for grammatical categorization (where some words were categorized in the wrong word class) and handling of nuanced or intransitive uses. Overall, while ChatGPT shows potential in certain areas, especially definition drafting (despite using obsolete styles), it lacks the precision and contextual sensitivity required for more complex lexicographic tasks (Jakubíček/ Rundell 2023: 517).

Among lexicographers, concerns regarding the use of LLMs in lexicography are related to the 'hallucinations' (fabrications) produced by LLMs and the fact that LLMs are a black box because "with the current tools for dictionary compilation, [lexicographers] can always go back to the corpus, which is not possible with ChatGPT" (Rundell in Schryver 2023: 366). Other concerns are related to the LLMs' non-deterministic output (since the LLM output can change for the same prompt), the English bias (in the training data) and the prompting strategies (Schryver 2023: 377). The major disadvantages of LLMs for lexicography as outlined by Jakubíček/Rundell (2023) revolve around the issues of using Google or GPTs as sole sources of references. *Googleology* and *GPTology* (Jakubíček/ Rundell 2023: 520) are problematic because tools like ChatGPT use unknown, non-deterministic data sources with limited stability

and reproducibility, making them unsuitable for academic research as a general-purpose search system. Furthermore, the data used by GPT models is often outdated and may reflect outdated lexicographic styles and methods, as the system has learned from publicly available, not necessarily high-quality or current dictionaries. Moreover, relying entirely on GPT for dictionary creation eliminates critical human roles, as GPT models do not allow for the verification of evidence, undermining the lexicographic process of checking and analyzing corpus data for accuracy. In contrast, advantages of using LLMs in lexicography are that they are user-friendly, cost-effective, continuously evolving and easily integrable into dictionary writing systems via programming interfaces (Jakubíček/ Rundell 2023: 518–519).

In conclusion, with appropriately crafted prompts, an LLM like ChatGPT can either independently generate a dictionary or, more reliably, assist in the process by producing high-quality draft entries that lexicographers can review and refine (Schryver 2023: 377).

2.2. LLMs and terminology

LLMs offer a broad range of possibilities for terminology work, for both terminologists and specialized translators supporting various stages from initial preparation to finalization. This includes domain identification, which can be particularly helpful when specialized translators are uncertain about the field of specialization of a source text and need to determine the appropriate specialized language (Łukasik 2023). Beyond extraction, LLMs can support terminologists in drafting definitions, selecting preferred terms, identifying synonyms or developing concept systems (Reineke 2023: 25).

Unlike traditional tools though, LLMs can provide additional information such as definitions and usage contexts, particularly when integrated with real data sources and functioning as both search engine and chatbot (Heinisch 2026: 109). However, caution is needed: if the LLM-generated definitions and contexts are not drawn from verifiable sources, they may be unreliable or inappropriate for the domain in question. Even when real sources are used, there is no guarantee that they are relevant in the domain or context, making expert consultation advisable. Consequently, LLM outputs should be supplemented with traditional terminology resources, as domain knowledge is essential for validating the reliability of definitions. Some LLMs can also integrate client-provided or external terminology databases (e.g. via Model Context Protocol), enabling more controlled and accurate translation workflows (Heinisch 2026: 107–112), including retrieval augmented generation or terminology augmented generation (Fleischmann/ Lang 2025: 90–94), which allow to generate texts in accordance with the preferred terms in a terminology database. Additional tasks for which terminologists may use LLMs include suggesting target-

language equivalents, outlining concept relations, recommending sources, generating data categories for terminology databases and exporting structured formats such as TBX files (Reineke 2023: 25; Heinisch 2026: 110–111) as well as terminology verification and validation in addition to suggestions for sources (Heinisch 2025: 16–20).

In conclusion, LLMs can support both terminologists and specialized translators in researching, preparing or managing terminology throughout the terminology and translation workflow. However, competences in terminology work, domain knowledge and LLM literacy are required to prepare useful prompts, but also to assess and revise the LLM output for the task at hand.

LLMs can also process and generate files in formats needed by lexicographers, terminologists or specialized translators, including structured XML, JSON, CSV, TEI XML used in the production of dictionaries, as well as TBX files relevant for terminologists and TMX and XLIFF files used in translation (Schryver 2023: 360). While there are also LLMs that are either fine-tuned to the fields mentioned above (including translation), users may also develop their own LLMs (or rather small or very small language models) and train them with their data. Nevertheless, LLMs are not always the best solution, which will be discussed later.

2.3. LLMs and specialized translation

While (specialized) translators have long incorporated machine translation engines into their workflows, LLMs represent a novel and transformative development within the field of translation technologies. Although both neural machine translation (NMT) systems and LLMs are typically based on transformer architectures, a key distinction lies in their training data: While NMT systems mainly use curated parallel corpora, LLMs are often trained on monolingual data with a strong bias towards English data (Hendy et al. 2023: 1). Another major difference is that NMT systems can only be used for translation, while LLMs "can be prompted to solve arbitrary tasks" (Guerreiro et al. 2023: 1501).

Several studies have compared the performance of NMT systems and LLMs for translation purposes, and often, LLMs are capable of generating translations that are comparable in quality to those produced by supervised machine translation models (Hendy et al. 2023: 14; Peng et al. 2023: 3–7). Basically, the results for high-resource languages, such as English or German are better than for low-resource ones, even with zero-shot prompting (Hendy et al. 2023: 1). Moreover, hallucinations remain a persistent issue in LLM-based translation (Guerreiro et al. 2023: 1501), particularly for low-resource language pairs and when translating from English, often reflecting biases in the training data. Smaller distilled models may hallucinate less than larger ones, possibly due to architectural features that constrain output generation. Unlike traditional

systems, LLMs exhibit distinct hallucination patterns (such as off-target translations and overgeneration) that are difficult to mitigate within similarly trained models. As the occurrence of hallucinations cannot be predicted, they might not be present in the next LLM output for the same task (Guerreiro et al. 2023: 1501). Moreover, the performance of LLMs can be improved through few-shot situations allowing for in-context learning (Hendy et al. 2023: 4).

The linguistic features (Sizov et al. 2024: 1183) employed by NMT systems and LLMs, as well as the 'errors' they make in translation tasks, are therefore also different (Guerreiro et al. 2023: 1505; Hendy et al. 2023: 4). This means that the translation output by an LLM requires more than traditional machine translation post-editing (MTPE). Although both the outputs of LLMs and NMT systems might be of good quality for high-resource languages, they "still differ from texts originally written by native speakers" (Sizov et al. 2024: 1183).

At present, no genuine multilingual models exist that can deliver equally strong performance across the full range of the world's languages (or even across a small portion of them). Evaluations reveal inconsistencies in performance across tasks, even within the same language. While the goal of training such models is broad and includes multilingual and multitasking competences, achieving this requires extensive and diverse data, raising challenges related to scalability, inclusivity and fairness (Hendy et al. 2023: 23).

Apart from the general translation capabilities of LLMs, recent studies assessed different prompting strategies for translation tasks in different language directions. A study exploring prompt design for ChatGPT's translation performance from English into Chinese Mandarin for an article in the field of popular science (He 2024: 7) evaluated four prompt types: a basic prompt (translation command, target language and purpose), a prompt based on a translation brief (including contextual factors such as function, target audience and medium) and prompts assigning the persona of either an *author* or a *translator*. Results showed that assigning ChatGPT the *translator* persona yielded the highest-quality translations, outperforming both the basic prompt and the *author* persona. Surprisingly, the prompt based on the translation brief (typically central to professional translation practice) resulted in the lowest performance. These findings suggest that, contrary to expectations, supplying detailed translation brief information does not necessarily enhance LLM translation output, whereas role assignment (specifically the translator persona) significantly improves results. However, due to the vast development of LLMs, this might already be outdated. Related work by Yamada (2023) investigated the use of ChatGPT for translating marketing texts from English into Japanese, focusing on how prompting strategies based on dynamic equivalence and user-centered translation affect output quality. The findings by Yamada (2023: 9) indicated that explicitly incorporating communicative intent

and audience considerations into prompts notably improved the overall quality of translations, especially in handling culturally nuanced and idiomatic expressions, aligning the results more closely with professional industry standards.

In the context of Chinese–English translation, Siu (2023) evaluated ChatGPT's effectiveness across various translation-related activities, including the clarification of expressions within specific contexts, the interpretation of culturally embedded elements and the explanation of terminology. Furthermore, ChatGPT was assessed for its utility in simplifying complex texts, generating multiple raw translation outputs and serving as a source of inspiration for translators' own translations. In the realm of quality assurance, the model was employed for error detection, grammatical correction, overall quality assessment and stylistic editing. Siu (2023) also explored ChatGPT's potential in interactive translation workflows, where iterative user interaction leads to progressively refined outputs.

In contrast to conventional machine translation tools, such as DeepL or Google Translate, which offer limited opportunities for users to input additional information regarding the translation task, LLMs enable the incorporation of extensive context information (Timothy 2023). Users can specify critical parameters such as the source and target languages (including language varieties), the intended perspective, the distinction between general-language and specialized texts, the text type (e.g. annual report, medical certificate, blog article), the target audience (e.g. domain experts or the general public) and the author's communicative intent. Furthermore, users can request alternative forms of output, such as a summarized translation rather than a full rendering of the source text. Providing translation examples within the prompt can further align the model's output with the stylistic and textual expectations of the translator. Although some machine translation systems now offer limited customization options (such as selecting a formal, informal or academic style), the capacity of LLMs to incorporate a comprehensive translation brief within the prompt represents a significant advancement, offering enhanced flexibility and a closer alignment with client-specific requirements.

In conclusion, the major advantage of LLMs compared to NMT systems is that they allow users to provide additional context, including a translation brief or a persona, which can improve the translation output of the LLM. The adaptation to the purpose and target audience as well as the consideration of the language variety or corporate language can increase the efficiency of translators. Also, the summarization or simplification of the source text, brainstorming while translating or revision of the translation can support translators in the translation process.

The growing integration of LLMs into lexicography, specialized translation and terminology work not only creates new methodological possibilities but

also raises ethical questions. While this contribution explores how these technologies can support professionals in managing and applying (specialized) knowledge, their adoption must also be considered in relation to broader societal, environmental and professional implications. Issues such as their negative environmental impact, the potential influence of LLMs on language use and norms and concerns about job displacement highlight the need for a critical and responsible approach to their use. The following section therefore situates the practical use of LLMs within an ethical framework, examining how technological innovation in these fields can be balanced with professional responsibility and expertise as well as sustainability considerations.

3. Ethical aspects of LLMs

An aspect too often overlooked and ignored is the ethics underlying the use of LLMs (not only for lexicography, terminology work or translation but also their ecological, social and economic implications), which can be broadly subsumed under sustainability. Since the integration of LLMs into professional language practices affects not only how these tasks are performed but also the conditions under which they are performed, their broader implications need to be considered alongside their practical benefits. This perspective is particularly relevant when assessing the role of professionals and their responsibility in deciding when, how and to what extent LLMs should be used.

3.1. Sustainability

As LLMs require vast resources, including raw material for producing hardware, the energy for training, deploying and prompting the systems and water to cool down the server farms, concerns about the use of LLMs (in lexicography, terminology work and specialized translation) span multiple dimensions. Environmentally, their operation significantly increases energy and water consumption (McKean/ Fitzgerald 2024: 19) and requires vast amounts of raw materials (Crawford/ Joler 2018). The energy usage throughout the LLM lifecycle is coupled with enormous carbon emissions arising from the training and operation of LLMs (Singh et al. 2025: 11). Thus, the rapid proliferation of AI technologies is expected to impose considerable environmental burdens, with estimates indicating that AI may contribute as much as 5% of global greenhouse gas emissions by 2030 (Andrae 2020: 55).

Ethically, LLMs risk amplifying bias, as their training data underrepresent marginalized groups, potentially reinforcing stereotypes in dictionary content (McKean/ Fitzgerald 2024: 20) or translations. Technically, issues like confabulation, hallucination, lack of transparency and non-reproducibility undermine trust and reliability (McKean/ Fitzgerald 2024: 19–23). Professionally,

the rise of LLMs threatens to devalue human linguistic expertise, potentially endangering the profession itself (McKean/ Fitzgerald 2024: 21). Finally, unresolved intellectual property and copyright questions raise serious legal and ethical challenges, particularly around data memorization and authorship of AI-generated content (McKean/ Fitzgerald 2024: 21–23).

Singh et al. (2025: 11–12) propose several solutions to these issues. These include implementing sustainable practices such as energy-efficient training, sustainable hardware, eco-friendly data centers and lifecycle assessment. Nevertheless, it is worth critically examining whether LLMs represent the most appropriate solution in the first place. For the tasks outlined above, the use of small or very small language models (capable of running locally on personal computers) may offer a more efficient and sustainable alternative (Schick/ Schütze 2020: 9) in addition to traditional tools, such as corpus analysis tools (Jakubíček/ Rundell 2023: 518).

3.2. Job displacement

The rise of LLMs brings significant concerns regarding job displacement and the devaluation of human expertise, as their increasing use in fields like lexicography and translation risks automating roles traditionally conducted by skilled professionals. Beyond economic implications, this shift also reflects broader challenges, including the opacity of model outputs, diminished human oversight and growing inequality in access to advanced technologies, issues that demand critical reflection and informed governance to ensure ethical and equitable integration (Tokayev 2023: 17).

3.3. Language change

Language, both general and specialized, is constantly changing. With the emergence of new domains, new terms are also evolving. The question arises what influence LLMs will have on (specialized) language in the future. While Massion (2024: 86) argues that an "[o]ver-reliance on AI-generated content can lead to a loss of originality, diversity and authenticity in language and communication," others observe that some phrases (such as "meticulous" and "commendable") have become more popular in academic (PubMed records) publications since the advent of LLMs (Matsui 2024: 2). It has to be noted that the usage frequency of these words and phrases had already been rising for several years prior to the release of ChatGPT, indicating that while ChatGPT may have accelerated their popularity, these expressions were already gaining momentum (Matsui 2024: 2). Another observation in the medical field is that human-written medical texts tend to be more concrete, diverse and context-specific, whereas ChatGPT-generated texts prioritize fluency and coherence but often rely on general terminology with less context-relevant detail (Liao

et al. 2023: 1). While the frequency of certain terms (in academic literature) changes according to research trends (some terms being more popular at certain times), the change in the frequency of adjectives, as well as the decreased use of 'are' and 'is' are likely to be related to LLM-generated texts (Geng/ Trotta 2024: 3). Therefore, the proliferation of ChatGPT may subtly lead to the decreased use of certain words or phrases without necessarily recognizing it (Matsui 2024: 12).

As mentioned above, using LLMs for the creation of dictionaries, terminology resources or specialized translation raises several concerns, including that LLMs are a black box and do not provide real sources (Schryver 2023: 366) (unless prompted to do so; and even then, also not reliably). For language professionals, this is an issue because they need to access the source, i.e. the authentic text. However, since more and more texts are produced (and post-edited) by means of LLMs, the corpora that these language professionals will (be able to) use in the future will change. The boundaries between 'synthetic'/ 'artificial' data and 'authentic' data might become blurred.

Furthermore, due to the English-centered training of LLMs "prompts formulated in other languages than English are answered using also (and mainly) English data" (Jakubíček/ Rundell 2023: 512). This also means that outputs in languages other than English are "thus patterned onto the (linguistic) characteristics and structures of the English language" (Schryver 2023: 364). For languages or language varieties that are not well-represented in LLMs, there is the risk of linguistic (and cultural) "homogenization as these technologies may promote dominant cultures at the expense of local or minority perspectives" (Tokayev 2023: 17). However, the positive impact of LLMs on academic writing (in fields where English is the lingua franca) is that they 'level the playing field' (to some extent) because "they help non-English native writers to improve the quality and flow of their text, as well as to translate into English from their mother tongue or vice versa" (Geng/ Trotta 2024: 10).

3.4. Dictionaries and terminology resources of the future

The question arises whether traditional dictionaries and terminology resources will continue to be (relevant and) needed as the reference works they are today. Therefore, the above-mentioned studies primarily focused on evaluations of the ability of LLMs to support their production. Nonetheless, Schryver raises the question that future dictionaries might be subsumed into (or entirely overtaken by) other digital tools: "we should have asked ChatGPT to produce the type of lexical analysis that will be needed in the future, rather than mimic the past" (Schryver 2023: 380). Furthermore, the question emerges as to who the primary users of dictionaries and terminology resources will be in the future: humans or machines (Schryver 2023: 380), such as LLMs.

4. Conclusion

LLMs like ChatGPT can be promising tools in lexicography, terminology work and specialized translation by accelerating (or augmenting) certain steps in the workflow and assisting with the detection of patterns and structures across languages, language varieties, text types and domains. When used critically, they can support professionals in various steps of their work and in generating outputs that can then be refined through expert analysis. However, critical limitations remain: LLMs lack transparency in data sourcing, cannot reliably reproduce results and currently do not support evidence-based validation through corpus back-linking, which is essential in both lexicography and terminology work. For this reason, the expertise of professionals, such as lexicographers, terminologists and specialized translators, remains central to creating useful prompts, as well as to assessing, contextualizing and revising LLM-generated output in the fields of lexicography, terminology work and specialized translation. Overreliance on these models risks perpetuating certain (outdated) language norms, homogenizing language and marginalizing professional expertise, making it essential that LLMs be used as supportive tools rather than standalone systems. Terminologists, lexicographers and specialized translators therefore play a key role in shaping how these technologies are applied in practice, combining them with established tools and methods. It is crucial that terminologists, lexicographers, specialized translators and even non-experts avoid relying solely on LLMs, as traditional tools like corpus analysis software and well-maintained terminology databases remain more reliable, less resource-intensive and better suited for many tasks.

References

Andrae, Anders S. G. (2020). "Hypotheses for Primary Energy Use, Electricity Use and CO_2 Emissions of Global Computing and Its Shares of the Total Between 2020 and 2030". In: *WSEAS Transcations of Power Systems* 15. Pp. 50–59.

Bowker, Lynne (2017). "Lexicography and terminology". In: Fuertes-Olivera, P.A. (ed.) *The Routledge Handbook of Lexicography*. Abingdon/New York. Pp. 138–151.

Bowker, Lynne/ Corpas Pastor, Gloria (2015). "Translation Technology". In: Mitkov, R. (ed.) *The Oxford Handbook of Computational Linguistics*. Oxford. Pp. 871–905.

Bozkurt, Aras (2024). "Tell Me Your Prompts and I Will Make Them True: The Alchemy of Prompt Engineering and Generative AI". In: *Open Praxis* 16(2). Pp. 111–118.

Chen, Banghao/ Zhang, Zhaofen/ Langrené, Nicolas/ Zhu, Shengxin (2023). "Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review". In: *arXiv e-prints*. (https://arxiv.org/html/2310.14735v4, accessed: 23.10.2023).

Costa, Rute/ Salgado, Ana/ Ramos, Margarida/ Almeida, Bruno/ Silva, Raquel/ Carvalho, Sara/ Khan, Fahad/ Tasovac, Toma/ Khemakhem, Mohamed/ Romary, Laurent (2023). "A crossroad between lexicography and terminology work: Knowledge organization and domain labelling". In: *Digital Scholarship in the Humanities* 38. Pp. i17–i29.

Crawford, Kate/ Joler, Vladan (2018). *Anatomy of an AI system. The Amazon Echo as an anatomical map of human labor, data and planetary resources*. (https://anatomyof.ai/index.html, accessed: 31.07.2024).

Drewer, Petra/ Schmitz, Klaus-Dirk (2017). *Terminologiemanagement. Grundlagen – Methoden – Werkzeuge*. Berlin/Heidelberg.

Edelmann, Gerhard (2014). "Rechungslegungsnormen und Übersetzung: Der Fachübersetzer als Terminologe". In: *trans-kom* 7(2). Pp. 184–203.

Ekin, Sabit (2023). *Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices*. [Preprint]. TechRxiv. DOI: 10.36227/techrxiv.22683919.v2.

Fleischmann, Klaus/ Lang, Christian (2025). "Terminologie in der KI. Wie mit Terminologie der Output von LLMs und GenAI optimiert werden kann". In: Drewer, P./ Felix, M/ Pulitano, D. (eds.) Terminologie in der KI – KI in der Terminologie. Akten des Symposions Worms, 27.-29. März 2025. München/Karlsruhe/Bern. Pp. 83–95.

Geng, Mingmeng/ Trotta, Roberto (2024). "Is ChatGPT Transforming Academics' Writing Style?". In: *arXiv e-prints*. (https://arxiv.org/abs/2404.08627, accessed: 31.10.2024).

Guerreiro, Nuno M./ Alves, Duarte M./ Waldendorf, Jonas/ Haddow, Barr / Birch, Alexandra/ Colombo, Pierr / Martins, André F. T. (2023). "Hallucinations in Large Multilingual Translation Models". In: *Transactions of the Association for Computational Linguistics* 11. Pp. 1500–1517.

He, Sui (2024). "Prompting ChatGPT for Translation: A Comparative Analysis of Translation Brief and Persona Prompts", in: *arXiv e-prints*. (https://arxiv.org/abs/2403.00127, accessed: 31.11.2024).

Heinisch, Barbara (2025). "Large language models for terminology work: A question of the right prompt?". In: *Journal for Language Technology and Computational Linguistics* 38(2). S. 13–30.

Heinisch, Barbara (2026). "Text-generative Large Language Models in Specialized Translation: Not Only Suitable for Translation". In: Walter, K./ Agnetta, M. (eds.) *Applying artificial intelligence in translation. Possibilities, processes and phenomena,* New York/London. Pp. 103–116.

Hendy, Amr/ Abdelrehim, Mohamed/ Sharaf, Amr/ Raunak, Vikas/ Gabr, Mohamed/ Matsushita, Hitokazu/ Kim, Young Jin/ Afify, Mohamed/ Awadalla, Hany Hassan (2023). "How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation". In: *arXiv e-prints*. (https://arxiv.org/abs/2302.09210, accessed: 5.11.2024).

Herrera-Poyatos, David/ Peláez-González, Carlos/ Zuheros, Cristina/ Herrera-Poyatos, Andrés/ Tejedor, Virilo/ Herrera, Francisco/ Montes, Rosana (2025). "An overview of model uncertainty and variability in LLM-based sentiment analysis. Challenges, mitigation strategies and the role of explainability". In: *arXiv e-prints*. (https://arxiv.org/abs/2504.04462, accessed: 5.01.2026).

Jakubíček, Miloš/ Rundell, Michael (2023). "The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography?". In: *Electronic lexicography in the 21st century (eLex 2023). Invisible Lexicography. Proceedings of the eLex 2023 conference*. Pp. 508–523.

Kageura, Kyo (2015). "Terminology and lexicography". In: Kockaert, H.J./ Steurs, F. (eds.) *Handbook of Terminology*. Amsterdam. Pp. 45–59.

Klosa, Annette (2007). "Korpusgestützte Lexikographie: besser, schneller, umfangreicher?". In: Kallmeyer, W./ Zifonun G. (eds.) *Sprachkorpora – Datenmengen und Erkenntnisfortschritt*. Berlin/New York. Pp. 105–122.

Lew, Robert (2023). "ChatGPT as a COBUILD lexicographer". In: *Humanities and Social Sciences Communications* 10(1). Pp. 1–10.

Liao, Wenxiong/ Liu, Zhengliang/ Dai, Haixing/ Xu, Shaochen/ Wu, Ziha / Zhang, Yiyang/ Huang, Xiaoke/ Zhu, Dajiang/ Cai, Hongmin/ Li, Quanzheng/ Liu, Tianming/ Li, Xiang (2023). "Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study". In: *JMIR medical education* 9. Pp. 1–15.

Łukasik, Marek (2023). *Terminology Work in the Era of AI-generated Content. A Case of ChatGPT,* presentation at *5th International Conference on Terminology: Scientific, Administrative and Educational Dimensions of Terminology. 19–20 Oct. 2023*. Vilnius.

Massion, François (2024). "Terminology in the Age of AI: The Transformation of Terminology Theory and Practice". In: *Journal of Translation Studies* 1. Pp. 67–94.

Matsui, Kentaro (2024). "Delving into PubMed Records: Some Terms in Medical Writing Have Drastically Changed after the Arrival of ChatGPT". Preprint. DOI: 10.1101/2024.05.14.24307373.

McKean, Erin/ Fitzgerald, Will (2024). "The ROI of AI in lexicography". In: *Lexicography* 11(1). Pp. 7–27.

Peng, Keqin/ Ding, Liang/ Zhong, Qihuang/ Shen, Li/ Liu, Xuebo/ Zhang, Min/ Ouyang, Yuanxin/ Tao, Dacheng (2023). "Towards Making the Most

of ChatGPT for Machine Translation". In: *arXiv e-prints.* (https://arxiv.org/abs/2303.13780, accessed: 5.01.2024).

Picht, Heribert (2001). "Korpora als Ausgangspunkt für die Extraktion von terminologischen Daten". In: *Fagspråk, kommunikasjon, kulturkunnskap Festskrift for Harald Frønsdal og Einar Hansen*. Bergen. Pp. 38–48.

Reineke, Detlef (2023). "Terminologiearbeit mit ChatGPT & Co.". In: *Fachzeitschrift für Terminologie*. 19(1). Pp. 25–28.

Rudolph, Jürgen/ Tan, Samson/ Tan, Shannon (2023). "ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?". In: *Journal of Applied Learning & Teaching* 6(1). Pp. 242–263.

Rundell, Michael (2024). "Automating the Creation of Dictionaries: Are We Nearly There?". In: *Humanising Language Teaching* 26(1). Pp. 1–14.

Saleem, Maab (2024). *11 ChatGPT Prompt Frameworks Every Marketer Should Know.* (https://buttercms.com/blog/chatgpt-prompt-frameworks, accessed: 30.09.2024).

Schick, Timo/ Schütze, Hinrich (2020). "It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners". In: *arXiv e-prints.* (https://arxiv.org/abs/2009.07118, accessed: 31.10.2024).

Schryver, Gilles-Maurice de (2023). "Generative AI and Lexicography: The Current State of the Art Using ChatGPT". In: *International Journal of Lexicography* 36(4). Pp. 355–387.

Singh, Aditi/ Patel, Nirmal Prakashbhai/ Ehtesham, Abul/ Kumar, Saket/ Khoei, Tala Talaei (2025). "A Survey of Sustainability in Large Language Models: Applications, Economics, and Challenges". In: *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*. Pp. 8–14.

Siu, Sai Cheong (2023). "ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation". In: *SSRN Electronic Journal*. Pp. 1–36.

Sizov, Fedor/ España-Bonet, Cristina/ van Genabith, Josef/ Xie, Roy/ Dutta Chowdhury, Koel (2024). "Analysing Translation Artifacts: A Comparative Study of LLMs, NMTs, and Human Translations". In: *Proceedings of the Ninth Conference on Machine Translation*. Stroudsburg. Pp. 1183–1199.

Timothy, Maxwell (2023). *How to Use ChatGPT as a Language Translation Tool.* (https://www.makeuseof.com/how-to-translate-with-chatgpt, accessed: 31.10.2024).

Tokayev, Kassym-Jomart (2023). "Ethical implications of large language models a multidimensional exploration of societal, economic, and technical concerns". In: *International Journal of Social Analytics* 8(9). Pp. 17–33.

Yamada, Masaru (2023). "Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability". In: *arXiv e-prints*. (https://arxiv.org/abs/2308.01391, accessed: 31.10.2024).

Zielinski, Daniel/ Ramírez Safar, Yamile (2005). "Research meets practice: t-survey 2005: An online survey on terminology extraction and terminology management". In: *Translating and the Computer* 27. Pp. 2–27.

Barbara Heinisch
Eurac Research
Institute for Applied Linguistics
Viale Druso 1
39100 Bolzano
Italy
barbara.heinisch@eurac.edu
ORCID: 0000-0002-1362-4088