Barbara Heinisch

Large language models for terminology work: A question of the right prompt?

Abstract

Text-generative large language models (LLMs) offer promising possibilities for terminology work, including term extraction, definition creation and assessment of concept relations. This study examines the performance of ChatGPT, Perplexity and Microsoft CoPilot for conducting terminology work in the field of the Austrian and British higher education systems using strategic prompting frameworks. Despite efforts to refine prompts by specifying language variety and system context, the LLM outputs failed to reliably differentiate between the Austrian and German systems and fabricated terms. Factors such as the distribution of German-language training data, potential pivot translation via English and the lack of transparency in LLM training further complicated evaluation. Additionally, output variability across identical prompts highlights the unpredictability of LLM-generated terminology. The study underscores the importance of human expertise in evaluating LLM outputs, as inconsistencies may undermine the reliability of terminology derived from such models. Without domain-specific knowledge (encompassing both subject-matter expertise and familiarity with terminology principles) as well as LLM literacy, users are unable to critically assess the quality of LLM outputs in terminological contexts. Rather than indiscriminately applying LLMs to all aspects of terminology work, it is crucial to assess their suitability for specific tasks.

1 Introduction

Large Language Models (LLMs), capable of processing and generating human-like text, are transforming numerous professions (Eloundou, Manning, Mishkin, & Rock, 2023), including specialized translation and terminology management. Since LLMs utilize distinct approaches to generate and comprehend language, they fundamentally change the function of terminology (Massion, 2024) and the way how terminologists and specialized translators approach terminology work. Nevertheless, technology has long supported translation, from Computer-Assisted Translation (CAT) tools to terminology extraction software, corpus analysis and alignment tools (Rothwell, Moorkens, Fernández-Parra, Drugan, & Austermühl, 2023). While machine translation tools have long been a staple for translators, LLMs bring a new level of versatility, as they cannot only be used for translation per se but also for translation-related tasks, such as clarifying meaning, editing style, detecting errors or assuring quality (Siu, 2023).

Also terminologists use a wide range of specialized software for terminology work, including term extraction (Steurs, de Wachter, & de Malsche, 2015) or the management of terminology in terminological databases (Drewer & Schmitz, 2017). LLMs can enhance terminology work by efficiently extracting relevant terms (Hamm, 2025), generating definitions (Reineke, 2023) in context and by assessing language-variety-specific terminology (Heinisch, 2020). They can be used for finding equivalents across languages and terminology validation. Additionally, LLMs help terminologists establish relationships between concepts and verify proper terminology use, streamlining the overall process (Massion, 2024).

Terminology, defined as "set of designations [...] and concepts [...] belonging to one domain [...] or subject [...]" (*ISO 1087:2019 Terminology work and terminology science — Vocabulary*, 2019) is crucial in specialized communication because it ensures precision, consistency and clarity in communication. In specialized translation, effective terminology management improves translation efficiency and ensures the quality of the target text. Therefore, terminology management is pivotal, including the identification of terms and concepts, the extraction of candidate terms as well as the organization and validation of terms before storing and maintaining them in terminological databases (Steurs et al., 2015). One major challenge in multilingual terminology work is determining the equivalence between concepts (Hohnhold, 1990). To accurately interpret and use terminology, terminologists and translators must consider its domain, the system it belongs to and the specific context in which it appears. Thus, terminology is domain-specific, system-bound and context-dependent. Therefore, the question arises how LLMs perform in (selected) tasks aimed at (multilingual) terminology work.

2 Method

The theoretical framework of this paper is grounded in Wüster's General Theory of Terminology (Wüster, 1974), whose foundational principles continue to inform contemporary terminological practice as codified in the ISO standard (*ISO 704:2022 Terminology work — Principles and methods*, 2022) on terminology work — Principles and methods. This study forms part of the larger project *UniTermGPT: University Terminology in German in the Age of ChatGPT*, which explores how ChatGPT handles university terminology across selected German language varieties. Therefore, this pilot study also addresses prompt engineering in the context of terminology injection into LLMs with a focus on system-bound terminology. The objective was to evaluate the potential of multilingual LLM terminology work in the field of university terminology in both German and English.

Since the larger project focusses on ChatGPT only, this study aimed to assess the general suitability of LLMs for translation-oriented terminology tasks by using three different LLMs: ChatGPT (GPT-4o-mini), Perplexity ('default model') and Microsoft CoPilot. The selection of ChatGPT, Perplexity and Microsoft Copilot for this study was partly informed by their prominence in contemporary academic and professional

contexts. ChatGPT was chosen due to its widespread adoption, Perplexity, as an early model integrating LLM outputs with real-time web search, is particularly suited to terminology work involving emergent terms, a phenomenon often underrepresented in static models. Microsoft Copilot, integrated into Microsoft 365, is widely used in organizations through its presence in tools like Word and Excel.

Given the practical orientation of this study, the LLMs were prompted (in German and partly in English) to address four terminology-related tasks: (1) the identification of key terms within a domain (distinct from term extraction, which typically presupposes the existence of a corpus), (2) term extraction from web-based sources (without having to compile a corpus beforehand), (3) the generation and extraction of definitions, and (4) the establishment of concept relations. These tasks simulate realistic scenarios faced by terminologists or specialized translators who require a foundational terminological database under time constraints.

The performance of the LLM also depends on the prompt being used. Selecting the right (user) prompt involves understanding user intent, model understanding and the specificity of the domain. Clear, specific prompts tailored to the task and any necessary constraints help guide the model to produce better results (Ekin, 2023). Prompt engineering principles (Bozkurt, 2024; Chen, Zhang, Langrené, & Zhu, 2013; Saleem, 2024) are a dime a dozen, ranging from general guidelines to prompt engineering frameworks. In this study, the CARE and RACE frameworks were used. CARE (context, action, result, example) consists of context, i.e. background information, action as the definition of the tasks to be completed, result to state the expected outcome and example to provide the LLM with concrete examples of what the output should be. RACE (role, action, context, expectation), on the other hand, focusses on the role, also sometimes referred to as persona the LLM should assume, action and context (which are similar to CARE) and expectation to specify the expected result. Additionally, these prompts were improved by LLMs for optimizing prompts. Moreover, prompt chaining, whereby a large task is broken into a sequence of smaller subtasks, each handled by its own prompt, was used and the domain specificity, system boundness and context dependence of terminology was considered in the prompt (Table 1).

Table 1: Consideration of domain specificity, system boundness and context dependence of terminology in the prompts

Terminology characteristics	**Aspect**	**Considered in the prompt**
Domain specificity	University terminology; studies (subdomain); admission (focus)	"in the university context", "in the field of university admission"
System boundness (system)	(Austrian) university system	"Austrian higher education system", "universities in Austria"
System boundness (language variety)	Austrian	"Austrian German", "the Austrian variety of the German language"
Context dependence	University vs. university of applied sciences; terminological variation; certain universities	By including source hierarchy, domains (ac.at) from which terms and definitions should be extracted and varying the terminology used in the prompt, e.g. "university" or "higher education system"; "Benennungen" or "Termini" (in German)

To determine whether the issue was simply finding the 'right prompt' or if it was influenced by factors beyond a single model, three different LLMs were tested. The goal was not to compare these models (or to compare them with traditional corpus analysis tools) but to gain a broader understanding of how useful LLMs are for terminology-related tasks. By using multiple models, the analysis was not limited to just one LLM, allowing for more comprehensive conclusions. This study adopts a qualitative approach to analyzing both the design of prompts and the outputs generated by the language model, with a focus on understanding LLM capabilities in specific terminological contexts, including domain-specific and system-bound (language-variety-specific) terminology. Therefore, the analysis focusses on the following aspects: 1) If the term actually exists (or is hallucinated); 2) If the term is bound to the correct system (e.g. Austrian university terminology or corporate language, if prompted) and 3) if the term is specific to the domain (and not from any other domain), i.e. the university or higher education domain (Table 2).

Table 2: Criteria and aspects considered in analyzing the LLM output

Criterion	Subcriterion	Guiding questions
Term existence	Real term (vs hallucination, pseudo-terminology)	Is the term real and used in recognized sources (e.g. termbases, glossaries)? Has the LLM generated a non-existent or fabricated term?
Domain specificity	University terminology; studies (subdomain); admission (focus)	Does the term belong to the relevant specialized (sub-)field?
System boundness	Correct system	Is the term from the (Austrian or British) university system? Does the term cover the relevant language variety?
Context dependence	Corporate language (if requested)	Is the term used by the relevant university? Is it the preferred term (at the university)?

3 Results

LLMs face challenges in several key areas of bilingual terminology work. The three LLMs analyzed in this study struggle with completing multiple steps or sequences, even when given step-by-step instructions within a single prompt. Additionally, they often fail to provide accurate terminological definitions, especially when requested in structured formats like tables. Moreover, they often do not provide original terms and definitions but translations of (German) terms and definitions (in English). LLMs also tend to mix Austrian university terminology with terms from the German higher education system, making it difficult to focus exclusively on the desired system, language variety and context. Lastly, when extracting definitions from websites, the models can produce inconsistent results unless the task is confined to a single, focused source.
Despite prompts that took the domain specificity, the system boundness and the context dependence of terminology into account, it was not possible to achieve the desired result. In some cases, the LLM outputs did not differentiate between the Austrian and German higher education systems. As a result, the outputs, which should have been related to the Austrian university system, contained terms from both contexts. For example, *Numerus Clausus (NC)* is not a restriction for university admission in Austria, whereas in Germany it is. The analyzed LLMs generally do not distinguish between Austrian and German university terminology despite prompts specifying the 'Austrian university system' or

'Austrian German'. However, they occasionally generate university-specific terms, even without being prompted, including proprietary system names like *u:space*, a platform of the University of Vienna. Also, when prompted for relations between concepts, terms from the German university system were included. While the concept relations generated by the LLMs are generally usable, they should be approached with caution. The terminologist must possess the ability to differentiate between 'German' and 'Austrian' university terminology in order to develop an accurate concept system based on the LLM output. For example, ChatGPT defined the following subordinate concepts for admission to studies (*Zulassung zum Studium*): „allgemeine Universitätsreife, fachgebundene Universitätsreife, Studienberechtigungsprüfung, Berufsreifeprüfung, Quotenregelung". While the results may seem plausible to non-experts, they contain pseudo-terminology (or hallucinations) as *fachgebundene Universitätsreife* is not a term used within the Austrian higher education system. The correct, albeit not commonly used, term would be *fachgebundene Hochschulreife*.

In the case of bilingual terminology work, in which a RACE prompt (illustrated in the Appendix) was used, the LLMs did not provide original definitions for the English or German terms provided as, in some cases, the LLMs just translated the definitions into the other language, thereby even inventing terms. For example, *conditional offer* and *unconditional offer* refer to university admissions with or without certain conditions in the English higher education system. The German terms *unbedingte Zulassung* and *bedingte Zulassung* as provided by ChatGPT are, however, not used in Austria at all. This means that the LLM outputs were often not useful to create or assess relationships between concepts and to prepare concept systems for the Austrian and (British) English higher education systems. Even by varying the specification of the respective language variety (e.g. 'Austrian German' or 'Austrian variety of German'), as well as the further specification of the system (e.g. 'Austrian higher education system', sometimes also 'at the University of Vienna') and the context, the LLM outputs could not be significantly improved.

Although the quality of the prompt has a significant impact on the quality of the LLM output, the characteristics of each LLM also play a role. The user's knowledge of these characteristics is termed 'model understanding' (Ekin, 2023). These characteristics include, for example, how up-to-date the training data are and how the training data for German are distributed across the German, Austrian, Swiss (and other) varieties of German. This is aggravated by the fact that the providers of the LLMs are often non-transparent with regard to such information. It is equally opaque whether the studied LLMs use English as a pivot language (i.e. translate the prompt into English), when the user enters German prompts, before returning the output to German. Some outputs allude to that, for example, the term *Notendurchschnitt (GPA)* was output by Microsoft CoPilot. However, the English abbreviation *GPA (grade point average)* is not a common abbreviation for *Notendurchschnitt* in German. This may lead to biases in multilingual terminology work and specialized translation in general.

4 Discussion

The variability of LLM outputs, even when using the same prompts, presents significant challenges in terminology-related tasks. Problems such as the creation of pseudo-terminology or inconsistencies highlight the critical importance of domain-specific expertise when using LLMs for terminology work. Without sufficient knowledge in the relevant field, such as Austrian or British university terminology, users may struggle to evaluate the quality and relevance of the generated output. This underscores the need for LLM literacy and a critical approach to LLM use in such specialized tasks. Human expertise remains crucial to meet the specific demands of terminology work, even when LLMs are involved. One key factor influencing the quality of LLM output is prompt engineering. The iterative process of refining prompts to better suit the task at hand is essential (Ekin, 2023), particularly in terminology work where domain specificity is vital. LLMs, unlike traditional tools such as corpus analysis software, require well-crafted prompts to produce accurate and relevant terminology outputs. In contrast, traditional tools can provide frequent terms within a domain without requiring extensive domain-specific input.

Despite their usefulness, LLMs do not necessarily enhance the productivity of terminologists. This study suggests that applying prompt chaining (breaking down tasks into smaller, sequential prompts) yields better results than attempting to address everything in a single prompt. This method is particularly important in multilingual terminology work, where equivalences between languages need to be established. However, LLMs as tools in terminology work can also be time-consuming, requiring terminologists to carefully scrutinize the output, verify sources and ensure the accuracy of definitions and web sources provided by the model. LLMs pose several challenges for terminology work, including biases and multilingual limitations. Their reliance on predominantly English training data (Wang et al., 2024) affects multilingual terminology since terms may be inherently altered by being filtered through English (Heinisch, in print), which makes it difficult to find equivalents in languages, other than English. Biases in, or a lack of training data can skew terminology work, particularly in emerging or niche domains (Heinisch, in print), and LLMs' tendency to hallucinate (terms) further complicates their reliability. Moreover, ecological concerns related to the energy consumption and carbon footprint of LLMs (Rojas, 2024) should be considered when choosing appropriate tools for terminology tasks. The limitations of this study lie in the selection of the LLMs and the prompting frameworks used: the analyzed models are not representative of all commercially available LLMs. Since terminology work is often multilingual (as demonstrated in this study), future research could include models with a stronger multilingual focus, such as EuroLLM (Martins et al., 2025). Furthermore, future research could employ more advanced prompting strategies and frameworks, as those used in this study were intentionally kept simple and concise. Moreover, the larger UniTermGPT project intends to include more language varieties as well as additional annotators. Given the study's focus on the practical application of LLMs by terminologists and specialized translators, no German

fine-tuned models were used. The aim was to reflect realistic workflows using a single general-purpose model, particularly for high-resource languages like German and English, which are typically well-supported by such models. While Retrieval-Augmented Generation (RAG) enables LLMs to access supplementary information, Terminology-Augmented Generation (TAG) (Fleischmann & Lang, 2025) represents a complementary approach tailored to domain-specific language use. TAG integrates several components: deterministic retrieval from structured terminological databases (as opposed to probabilistic retrieval from vectorized data), the generation of precise and processable outputs in standardized or prose-like terminology formats and real-time access to terminology resources via APIs (Fleischmann & Lang, 2025). This methodology is particularly beneficial for content generation tasks that require adherence to domain-specific or corporate language norms, such as specialized translation or technical communication, where consistent use of (validated) terminology is essential.

5 Conclusion

While large language models offer promising possibilities for terminology work, it is evident that not all domains, languages and language varieties are equally supported by these systems. The presence of biases and hallucinations poses significant challenges in multilingual and domain-specific terminology tasks, underscoring the importance of human expertise in mitigating these issues. To ensure reliable outputs, LLM literacy is essential for users engaging with these tools in terminology work. Furthermore, prompt engineering plays a crucial role in shaping the quality of the LLM's responses, though the inherent characteristics of the model, such as its training data and its capabilities, also influence the results. In some cases, smaller, fine-tuned models focused on terminology tasks or even traditional tools like corpus analysis tools may be more suitable. The balance between LLM usage and traditional terminology tools, as well as the expertise required to navigate LLM outputs, is crucial for effective and high-quality terminology work.

Acknowledgement

This paper was funded by the EC-MCSA Seal of Excellence Programme of the Autonomous Province of Bozen/Bolzano – Department for Innovation, Research and University, project *University terminology in German in the age of ChatGPT (UniTermGPT)*.

References

Bozkurt, A. (2024). Tell Me Your Prompts and I Will Make Them True: The Alchemy of Prompt Engineering and Generative AI. *Open Praxis*, *16*(2), 111–118. Retrieved from https://search.informit.org/doi/pdf/10.3316/informit.T202404100001439007354109O

Chen, B., Zhang, Z., Langrené, N., & Zhu, S. (2013). *Unleashing the potential of prompt engineering in Large Language Models: a comprehensive review.* Retrieved from http://arxiv.org/pdf/2310.14735

Drewer, P., & Schmitz, K.-D. (2017). *Terminologiemanagement: Grundlagen - Methoden - Werkzeuge.* Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-53315-4

Ekin, S. (2023). *Prompt Engineering for ChatGPT: A Quick Guide to Techniques, Tips, and Best Practices.* Institute of Electrical and Electronics Engineers (IEEE). doi: 10.36227/techrxiv.22683919.v1

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.* Retrieved from https://doi.org/10.48550/arXiv.2303.10130

Fleischmann, K., & Lang, C. (2025). Terminologie in der KI. Wie mit Terminologie der Output von LLMs und GenAI optimiert werden kann. In P. Drewer, F. Mayer, & D. Pulitano (Eds.), *Terminologie in der KI – KI in der Terminologie. Akten des Symposions Worms, 27.–29. März 2025* (pp. 83–95). München / Karlsruhe / Bern: Deutscher Terminologie-Tag e.V.

Hamm, J. (2025). Terminologische Konsistenz und generative KI – ein Perfect Match? Produktiver Einsatz von Sprachmodellen im Terminologiemanagement und beim Post-Editing. In P. Drewer, F. Mayer, & D. Pulitano (Eds.), *Terminologie in der KI – KI in der Terminologie. Akten des Symposions Worms, 27.–29. März 2025* (pp. 151–163). München / Karlsruhe / Bern: Deutscher Terminologie-Tag e.V.

Heinisch, B. (2020). Sprachvarietätenabhängige Terminologie in der neuronalen maschinellen Übersetzung: Eine Analyse in der Sprachrichtung Englisch-Deutsch mit Schwerpunkt auf der österreichischen Varietät der deutschen Sprache. In C. Schöch (Ed.), *DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts* (pp. 211–214). doi: 10.5281/zenodo.4621962

Heinisch, B. (in print). Next-Gen Terminology: Transforming Terminology Work with Large Language Models. *Across Languages and Cultures.*

Hohnhold, I. (1990). *Übersetzungsorientierte Terminologiearbeit: Eine Grundlegung für Praktiker.* Stuttgart: InTra, 1. Fachübersetzergenossenschaft. Retrieved from http://media.obvsg.at/AC00246114-1001

ISO 1087:2019 Terminology work and terminology science — Vocabulary (2nd ed.). (2019). https://www.iso.org/standard/73906.html. (Published by International Organization for Standardization (ISO))

ISO 704:2022 Terminology work — Principles and methods (4th ed.). (2022). https://

www.iso.org/standard/81503.html. (Published by International Organization for Standardization (ISO))

Martins, P. H., Fernandes, P., Alves, J., Guerreiro, N. M., Rei, R., Alves, D. M., ... Martins, A. F. (2025). EuroLLM: Multilingual Language Models for Europe. *Procedia Computer Science*, *255*, 53-62. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877050925006210 (Proceedings of the Second EuroHPC user day) doi: https://doi.org/10.1016/j.procs.2025.02.260

Massion, F. (2024). Terminology in the Age of AI: The Transformation of Terminology Theory and Practice. *Journal of Translation Studies*, *4*(1). Retrieved from https://www.peterlang.com/document/1495905 doi: 10.3726/JTS012024.04

Reineke, D. (2023). Terminologiearbeit mit ChatGPT & Co. *Fachzeitschrift für Terminologie*, *19*(1), 25–28. Retrieved from http://dttev.org/images/edition/ausgaben/edition-2023-1-e-version.pdf

Rojas, S. (2024). Evaluating the Environmental Impact of Large Language Models: Sustainable Approaches and Practices. *Innovative Computer Sciences Journal*, *10*(1), 1–6. Retrieved from https://innovatesci-publishers.com/index.php/ICSJ/article/view/153

Rothwell, A., Moorkens, J., Fernández-Parra, M., Drugan, J., & Austermühl, F. (2023). *Translation tools and technologies* (1st edition ed.). London New York: Routledge, Taylor & Francis Group. doi: 10.4324/9781003160793

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

Siu, S. C. (2023). ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4448091

Steurs, F., de Wachter, K., & de Malsche, E. (2015). Terminology tools. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology* (pp. 222–249). Amsterdam: John Benjamins Publishing Company.

Wang, W., Tu, Z., Chen, C., Yuan, Y., Huang, J.-t., Jiao, W., & Lyu, M. (2024). All Languages Matter: On the Multilingual Safety of LLMs. In *Findings of the Association for Computational Linguistics ACL 2024* (pp. 5865–5877). Stroudsburg, PA, USA: Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.349

Wüster, E. (1974). Die Allgemeine Terminologielehre - Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften. *Linguistics*, *12*(119), 61–106. Retrieved 2025-05-21, from https://doi.org/10.1515/ling.1974.12.119.61 doi: doi:10.1515/ling.1974.12.119.61

Appendix

The appendix contains examples of prompts (according to the RACE and CARE frameworks) used in this study. The LLMs were prompted in German or English depending on the task at hand: For example, if the task was related to Austrian university terminology, the prompt was written in German, and if the task was solely related to the British university terminology, the prompt was in English. As all the German prompts will be made available as part of the larger UniTermGPT project, the following section only gives examples of the prompts in English (including translations of the German prompts). First, the different aspects of the selected prompting frameworks are illustrated. Second, examples of full prompts are provided. Third, an example of a prompt improved by *Prompt Maker* by Ruben Hassid is shown.

Selected prompting frameworks applied to terminology tasks

The two selected prompting frameworks (RACE and CARE) applied to terminology tasks:

Table 3: Prompting frameworks applied to terminology tasks

Task	CARE prompt	RACE prompt
(1) Term identification	**Context:** Identify university admission terminology in Austria (and the UK). **Action:** Create list of key terms. **Result:** A list of admission-related terms (with brief descriptions). **Example:** transcript of records = official document summarizing a university student's academic performance and progress to date.	**Role:** Terminologist for higher education. **Action:** Generate list of admission terms. **Context:** Austrian and British university system. **Expectation:** 10–15 terms per system (with brief definitions).

Task	CARE prompt	RACE prompt
(2) Term extraction	**Context:** Analyze university admission terminology in Austria (and the UK). **Action:** Extract relevant terms from websites or authoritative sources. **Result:** A glossary of admission-related terms (with source). **Example:** Sammelzeugnis = Zeugnis über alle absolvierten Prüfungen eines Studierenden an einer Universität.	**Role:** Terminologist for higher education. **Action:** Generate list of admission terms from university websites. **Context:** Austrian and British university system. **Expectation:** 10–15 terms per system (with source).
(3) Definition generation or extraction	**Context:** List of terms collected from domain of university admission. **Action:** Define each term using ISO 704/1087 principles. **Result:** Structured definitions in German/English. **Example:** "Zulassungsbescheid" – Verwaltungsakt, mit dem eine Universität Studienwerber*innen formal mitteilt, dass sie für ein Studium in einer bestimmten Studienrichtung unter den angegebenen Bedingungen aufgenommen sind.	**Role:** Expert in ISO-compliant terminology work. **Action:** Write precise definitions. **Context:** Multilingual terminology work and management. **Expectation:** 10-15 definitions of concepts in German (and English).
(4) Concept relations	**Context:** In terminology, concepts are related to each other. **Action:** Identify hierarchical (generic and partitive) (and associative) relations. **Result:** Concept system or structure with relation types. **Example:** "University" → "Faculty" (partitive).	**Role:** Terminologist comparing university systems. **Action:** Map concept relations. **Context:** Prepare concept system in German / English. **Expectation:** Hierarchical or associative concept relations.

Examples of full prompts according to the selected prompting frameworks

Examples of full prompts (mainly translated from German) according to the RACE and CARE prompting frameworks for the four selected terminology tasks are listed in the following:

Table 4: Prompting frameworks for terminology tasks using prose-style prompts

Task	CARE prompt	RACE prompt
(1) Term identification	In the context of university admission in Austria and the UK, generate a list of the most frequently used terms in this domain. Your task is to identify and list core terminology used in university admission processes. The result should be a bilingual list (Austrian German–British English) of 10 terms including definitions according to terminological principles. For example: Studienwerber*in: Person, die an einer Universität die Zulassung zu einem bestimmten Studium beantragt. – Applicant: person who has submitted an application for admission to a university.	As a terminologist specialising in higher education, identify key terms related to university admissions in both Austria and the UK. Consider how admission is structured in each country. Your output should be a list of 10 admission-related terms per system, each with a short definition in context.

Task	CARE prompt	RACE prompt
(2) Term extraction	For terminology management in the Austrian university system, using official Austrian websites or normative documents, extract key terms related to the study admission process in Austria. This includes identifying domain-specific terminology from legal texts or university materials. Provide the term, its source (and a short contextual definition) in a three-column table. For example: 1) Studieneingangs- und Orientierungsphase, 2) Angebot von Lehrveranstaltungen aus den das jeweilige Diplom- oder Bachelorstudium besonders kennzeichnenden Fächern, das der Information und der Orientierung der Studienanfängerinnen und Studienanfänger dient, 3) `URL`.	Act as a terminology expert conducting web-based term extraction in the field of Austrian university terminology. Focus on extracting university admission terminology from Austrian institutional websites (e.g. ministry or university portals). Present 10–15 relevant terms from the Austrian higher education system and include their source (URL or document title).

Task	CARE prompt	RACE prompt
(3) Definition generation or extraction	For comparative terminology work in the university sector in German and English, compile a list of the most common terms in the field of university admissions. Add definitions for these terms from Austrian and British websites and preferably from normative or official sources, such as laws or documents from authorities or organisations. Compare terms from the Austrian and British higher education systems. Here is an example: Sammelzeugnis = 'Zeugnis über alle absolvierten Prüfungen eines Studierenden an einer Universität', the English equivalent is: transcript of records = 'official document summarising a university student's academic performance and progress to date'. Add the URL of the website where you found the definitions.	You are a terminologist and need definitions of terms in the university context in Austrian German and British English in the field of university admissions. Given a preselected list of terms related to university admission, define each term using principles from ISO 704 and ISO 1087. Where possible, extract official definitions from trusted sources (e.g. laws or university guidelines). Create a table with 6 columns: 1st column: German term. 2nd column: German definition for this term. 3rd column: Source of the definition from the Internet by specifying a real website from Austria (domain: `ac.at`). 4th column: English term. 5th column: English definition. 6th column: Source of the definition from the Internet by specifying a real website from Great Britain (domain: `ac.uk`).
(4) Concept relations	In the context of terminological concept systems, analyze the relationships between the concepts (and terms) identified in the Austrian university context. Your task is to construct a concept structure for the field of university admission in Austria showing hierarchical (generic and partitive) or associative relationships. The result should support the creation of (comparative) concept systems in multilingual terminology work, e.g. University → Faculty (partitive).	As an ISO-compliant terminologist comparing the Austrian and British higher education systems, identify concept relationships among university admission terminology. Develop a concept system that includes hierarchy (e.g. admission document → Zulassungsbescheid) and differences between the Austrian and British university admission systems. Document the structure clearly for bilingual comparison.

Example of prompt improved by Prompt Maker LLM

Example of prompt (in German) improved by *Prompt Maker* by Ruben Hassid (although the terminology of terminology is not correct in the improved prompt, e.g. *Fachbegriff* instead of *Benennung*, the LLM output with the improved prompt was not changed, except for the headings in the sample table):

Promptbeschreibung: Terminologiearbeit im Hochschulbereich (AT/UK)
Rolle & Kontext: Agiere als Experte für vergleichende Terminologiearbeit im Hochschulbereich mit Spezialisierung auf Studienzulassungsverfahren in Österreich und Großbritannien. Deine Aufgabe ist es, eine präzise, systematische Liste der häufigsten Fachbegriffe in diesem Bereich zu erstellen und die entsprechenden Begriffe auf Deutsch und Englisch zu vergleichen.
Ziel & Anforderungen:

1. Erstelle eine umfassende Liste relevanter Begriffe aus dem Bereich der Studienzulassung.
2. Für jeden Begriff liefere folgende Informationen:
   - **Deutsche Bezeichnung**
   - **Englische Entsprechung (britisches Hochschulsystem)**
   - **Definition in beiden Sprachen**, basierend auf offiziellen Quellen (z. B. gesetzliche Regelungen, Dokumente von Hochschulen oder Behörden)
   - **Quelle der Definition** (falls verfügbar)
3. Beachte spezifische Unterschiede zwischen dem österreichischen und britischen Hochschulsystem. Falls es keine exakte Entsprechung gibt, erläutere die Unterschiede.

Beispieltabelle

Table 5: Vergleich von Fachbegriffen der Studienzulassung (AT/UK)

Deutsche Benennung	Englische Benennung	Deutsche Definition	Quelle Def. DE	Englische Definition	Quelle Def. EN
Sammel-zeugnis	transcript of records	Zeugnis über alle absolvierten Prüfungen eines Studierenden an einer Universität.	[Name der Quelle]	Official document summarizing a university student's academic performance and progress to date.	[Name der Quelle]

Detaillierter Arbeitsablauf

1. **Recherche:** Identifiziere die wichtigsten Begriffe im Bereich der Studienzulassung in Österreich und Großbritannien.
2. **Kategorisierung:** Ordne die Begriffe in sinnvolle Gruppen ein (z. B. Bewerbungsunterlagen, Zulassungsbedingungen, Prüfungsanforderungen).
3. **Definitionserstellung:** Nutze vertrauenswürdige Quellen, um exakte Definitionen bereitzustellen. Falls es keine direkte Entsprechung gibt, erkläre die Unterschiede.
4. **Formatierung:** Präsentiere die Ergebnisse in einer strukturierten Tabelle für bessere Übersichtlichkeit.

Wichtige Hinweise:

- Bevorzuge offizielle Quellen wie Hochschulgesetze, Webseiten von Universitäten oder Ministerien.
- Falls Begriffe Mehrdeutigkeiten aufweisen, gib eine kurze Kontext-Erklärung dazu.
- Falls es im britischen System kein direktes Äquivalent gibt, erläutere den Unterschied.

*Nimm dir Zeit und arbeite die Aufgabe **Schritt für Schritt** durch.*

Correspondence

Barbara Heinisch

Eurac Research
Institute for Applied Linguistics
Bolzano, Italy
barbara.heinisch@eurac.edu