SKASE Journal of Translation and Interpretation, 2025; 18(2): 117–48
doi: 10.33542/JTI2025-S-7

AI and digital literacy in translation competence development

Athanasios Breskas*, Silvia Hansen-Schirra*, Dimitrios Kapnas*

*Johannes Gutenberg University

Abstract

Numerous studies in translation process research (TPR) show that both professional and student translation processes become more efficient, and the outputs become of even higher quality and more consistent using AI as machine pre-processing. The aim of this empirical keylogging study was to explore how school pupils and university students use conventional translation tools, e.g., machine translation (MT), and innovative developments, such as ChatGPT, in comparison. Specifically, we looked at how (often) the participants used the tools, the effects on language processing and translation quality, and which translation strategies were employed. We expected students who, in comparison to the pupils, display greater translation competence, to deliver better final translations and to harness the power of the available tools more effectively to their advantage. The results of this study confirm this assumption and are in line with TPR studies. Students used a greater variety of tools to solve different problems and also prompted more when only allowed to use ChatGPT. This study shows that tools alone, in their current state, do not make up for lacking language skills. On the contrary: Language skills are necessary to evaluate the tool output and make informed decisions. Finally, this study suggests an expansion of the existing translation and post-editing competence models, as future translation students have already come in contact with prevalent innovative language technologies and thus have different prerequisites.

Keywords: *Translation Competence; Artificial Intelligence; Machine Translation; Keylogging; Prompting*

1. Introduction

The motivation for this study lies in the technological advancements in the field of generative AI (GenAI), which have surged dramatically in recent years. On the one hand, AI-based programs are increasingly used in everyday life (e.g. speech recognition systems or MT), while on the other hand, they are also making inroads into professional contexts. In our study, we thus focus on the question of how the use of GenAI impacts translation tasks. How do laypeople use such tools for translation, in our case for school tasks, and how do (semi-) professional translators integrate such tools into translation processes?

This study also seeks to explore to what extent lay translators are already influenced by their daily interaction with AI systems, including MT on mobile phones, and whether they are already familiar with potential translation technologies (such as DeepL, Google Translate, or other online resources). In the past, translation students typically had little to no experience with translation tools and online resources before beginning their studies. Such skills were part of a clearly defined set of translational competencies developed during their studies. However, nowadays, students often bring prior experience, as they use translation tools on their mobile phones for personal tasks and increasingly interact with GenAI to complete school

assignments, including tasks in foreign language classes. They, therefore, possess knowledge that laypeople or beginners in translation studies did not have in the past.

Our study seeks to merge these lines of research questions by comparing school pupils, i.e., lay translators, with university students, i.e., novice or intermediate translators, regarding their use of innovative AI tools in contrast to conventional tools, i.e., MT and online resources. The study employs methods from TPR, such as keylogging and screen recording, to measure translation efficiency and strategies. The quality of translations is evaluated, and a questionnaire is used to gather metadata on participants and their usage behavior regarding translation resources. The collected data are triangulated, and quantitative as well as qualitative results are discussed. This study sheds light on the question to which extent pre-academic knowledge regarding tool usage exists today. And if so, this knowledge should be addressed and integrated in academic programs, which makes an adaptation of didactic concepts pertaining to translation competence models necessary. First, we will describe the state-of-the-art of competence models, refer to TPR studies and outline our hypotheses, present our methods, and discuss our results.

2. Competence models and competence development

2.1. Translation competence models

AI and digital literacy have not been in the focus of translation competence models. Established models such as by PACTE (2003) and Göpferich (2009) define translation competence as a multi-component skill set that enables translators to produce accurate and appropriate translations. Key components of the models include:

- Bilingual sub-competence: proficiency and communicative competence in both the source and target languages, the ability to understand and produce appropriate texts in both languages.
- Extra-linguistic sub-competence: cultural, domain, and subject-specific knowledge relevant to the translation task.
- Instrumental sub-competence: proficiency in using tools, online resources and interfaces, such as dictionaries, databases and CAT tools.
- Strategic sub-competence: problem-solving and decision-making skills and the capacity to develop effective strategies to address challenges during the translation process.
- Psycho-physiological components: cognitive and behavioral abilities like memory, attention, and perseverance.
- Knowledge about translation: awareness of translation theories, processes, and practices.
- Translation routine activation: automatic recall of routines and standards for handling recurring translation scenarios.

The PACTE (2003) model emphasizes the dynamic interaction between these components, with a strong focus on empirical research and experimental validation. Göpferich (2009) emphasizes the importance of monitoring mechanisms to ensure quality and consistency, highlighting the interplay of cognitive processes. The importance of digital

literacy is only addressed within the instrumental sub-competence, which seems to be too short-sighted against the background of the most recent digital and big data developments. Considering the AI hype, these models are not state-of-the-art anymore.

Concerning digital and AI literacy, more practice-oriented frameworks come into play, e.g., the EMT (EMT Board and Competence Task-Force 2022) competence model or the definition of translation competence in the DIN ISO 17100 (2015). The EMT (European Master's in Translation) model links translation education with professional market needs, ensuring that graduates are prepared for the challenges of the translation industry. It is designed to standardize translation training across Europe and identifies the following key competence areas, which are similar to those mentioned above: language and culture, translation, personal and interpersonal skills, service provision and domain-specific knowledge. In addition to the models above, they further define the technological skills as follows: mastery of translation technologies, such as Computer-Assisted Translation (CAT) tools, MT Post-Editing (PE), and terminology management. In the area of neural MT, the interaction with AI-driven Large Language Models (LLMs) is implicitly included in the competence definition. The DIN ISO 17100 (2015) standard does not explicitly address PE of MT, but it has a strong focus on research and tools competencies since it demands the familiarity with and the ability to use tools and technology necessary for the translation process, including CAT tools, terminology management systems, file formats and software related to translation projects, and knowledge of quality control procedures. Additional notes from the DIN ISO 17100 (2015) suggest that translators are expected to continuously develop their competencies through training and practice, which implicitly covers the continuous development in digital literacy.

2.2. Post-editing competence models

In addition to the translation competence models introduced above, Nitzke & Hansen-Schirra (2021) introduce a competence model focusing on PE competencies. This model emphasizes the interaction between human and machine, highlighting the importance of understanding MT-specific errors and workflows. The model considers both cognitive demands (e.g. recognizing MT limitations) and technological proficiency, setting it apart from traditional translation competence models. It reflects the growing role of MT in professional workflows, offering a framework for the training of post-editors. The PE competence model is based on various translation competence models discussed above and the revision competence model by Robert et al. (2017). In addition to the translation competencies introduced above, the PE model covers the following PE-specific competencies:

- Error recognition and correction: the ability to identify, categorize and correct errors in machine-translated (MT) texts quickly and precisely.
- Knowledge of MT systems: understanding how MT systems are developed, trained, and function. Awareness of the typical advantages and limitations of MT systems, not only linguistically but also regarding aspects like training processes, security, or sustainability.
- Advisory competence: the ability to advise clients or stakeholders, which is increasingly important as translation and PE workflows grow more complex. This competence helps clients make informed decisions and weigh the opportunities and risks associated with using MT.

The model is rounded off by PE-specific soft skills, some of which overlap with those required for traditional translation:

- Psycho-physiological traits: focus, stress resistance, and analytical thinking are crucial for PE.
- Adherence to PE guidelines: the ability to follow PE instructions and quality standards precisely and efficiently.
- Interest in technology: a curiosity about technical developments and tools.
- Positive self-image: confidence in one's abilities is essential for effective PE.

This model can be regarded as one of the first frameworks specifically for PE competencies and enhances the awareness of the skills gap between traditional translators and post-editors, emphasizing the need for specialized competencies in the evolving translation landscape. The same holds true for the DIN ISO 18587 (2017), which also deals with PE competencies and qualifications. The competencies outlined there align closely with those in the PE competence model. The standard identifies several skills for post-editors that overlap with general translation competencies: translation competence, linguistic and textual competence, research and information management as well as cultural, technical, and subject-specific competence. Additionally, for professional PE, the standard includes:

- Knowledge of MT systems and CAT tools.
- The ability to assess the feasibility of editing MT output concerning time and effort.
- Adherence to PE guidelines.

These MT and PE competence models provide a structured framework for understanding the skills required in post-editing. They emphasize the importance of foundational translation skills, PE-specific technical and cognitive abilities as well as soft skills tailored to the unique challenges of post-editing MT. They highlight the growing complexity of translation workflows and the need for specialized competencies to ensure quality and efficiency in PE processes.

2.3. Translation competence development models and MT literacy

Having defined the diverse competencies that a translator and a post-editor should have, it is important to consider how these competencies can be developed and trained. The translation competence models by PACTE (2000) and Risku (1998) aim to explain how translators acquire, refine, and apply their competencies over time, often incorporating insights from cognitive science, pedagogy, and translation studies. These models suggest various stages of competence development: there is a staged progression of translation competence acquisition, typically moving from novice to expert (Risku 1998). The stages are characterized by:

- Lay translators: pre-translational competence.
- Novice translators: focus on basic linguistic equivalence, limited ability to navigate complex texts or cultural nuances.
- Intermediate translators: increased ability to handle challenges through training and experience, with growing use of translation tools.
- Expert translators: mastery of translation processes, strategic competence, and seamless integration of cultural, linguistic, and domain-specific knowledge.

Risku (1998) describes competence development from lay to expert translator according to the following requirements: macro strategic development, information integration, planning and decision making, self-organization. PACTE (2000) describes the learning process from pre-translational to translational competence as the development and integration of the above-described sub-competencies. Concerning AI and digital literacy, there is little research on which basic knowledge potential students may already have and which competencies they still need to learn. Bridging this gap from a didactic perspective, Krüger & Hackenbuchner (2024) suggest a matrix for MT-oriented teaching of data literacy, which defines competence descriptors for data planning, collection and production, data evaluation as well as data use within the context of MT data literacy. However, the new generation of lay and novice translators dramatically differ from lay and novice translators generations ago. In contrast to the past, when various digital translation tools that came to use in the language industry were unknown to most of the students before beginning their studies, nowadays, many young people have already come in contact with LLMs and/or GenAI (Sengar et al. 2024) and even schools employ these models as teaching assistants in the classroom (Khanmigo n.d.). With the onset of GenAI, the language industry has already begun adapting to this change (Lionbridge 2025; Carr 2023). Accordingly, language and translation studies curricula should follow suit to better prepare their students for the demands of the job market. Universities need not only adapt their curricula to a shifting job market but also to a generation that is already sensitized to the way LLMs work and what they can provide. Additionally, this generation has a basic understanding of what MT is and actively uses it on a frequent basis. In the coming sections, we will try to explain what defines this young generation and what implications this might have for the motivation of this study.

3. Translation process research and consultation of external resources

Research into the use of external resources while translating is not new in TPR (Carl et al. 2021). As consulting external resources is defined as a constituting element of various competence models (see previous sections), it is no wonder that TPR has investigated its effect on translation quality and efficiency and to the translator. Research into this field also focuses on improving translator education and better preparing students for the job market.

For example, Chodkiewicz (2015) conducted a study with 34 B.A. students in Applied Linguistics, whose native language was Polish. She qualitatively studied their behavior while translating from English to Polish and vice versa. The author recorded the participants' sessions using screen recording software and later analyzed the videos. She annotated the search queries and tools used, relying on an annotation scheme derived from the literature. In conclusion, almost all participants used external sources during translation. Participants "who performed L2 translation relied more heavily on a divers [sic!] types of searches" (Chodkiewicz 2015: 137), which can be interpreted as a means of dealing with inferior L2 language competence compared to their L1. Furthermore, the most common type of websites queried were bilingual dictionaries, online encyclopedias, online forums, and search engines. The author also notes that the participants did not just research "equivalents of particular terms, but also [...] word meaning, extralinguistic knowledge, and TL correctness" (Chodkiewicz 2015: 137). This points to the participants' translation competence, as they have

been sensitized about typical problems when transferring meaning from one language to another and the corresponding translation strategies. Finally, the author seems to question the participants' technical computer skills, as many of them were neither familiar with the "Control Find command" (Chodkiewicz 2015: 138), a basic computer shortcut used to find and process information in a text quickly, nor the usage of "quotation marks to search for an entire phrase" (Chodkiewicz 2015: 138), which enables the user to utilize search engine algorithms to more precisely narrow down the results.

TPR has suggested a quantitative analysis of the usage of external sources as well. Researchers have developed computer tools that allow them to conduct quantitative analyses by meticulously logging every action during the translation process. These tools include but are not limited to Translog-II (Jakobsen & Schou 1999) and InputLog (Leijten & VanWaes 2013). They enable the recording, replaying, and analysis of the entire translation process. More importantly, they do not interfere with the reading, writing, and/or translation processes in any way, as they do not require the participant to allocate cognitive resources into something not relevant to the task. Additionally, they run silently in the background thus not distracting the participant, like Think Aloud Protocols would (e.g. Jakobsen 2003). Furthermore, Translog can be paired with eye trackers (Carl 2012), which log every input. Simultaneously, it records eye movement data and maps that to the same timeline as the mouse and keyboard inputs, thus shedding light further into the translator's mind while translating. InputLog on the other hand logs keyboard and mouse inputs, its advantage being that it logs the corresponding window in which these inputs were made. Carl et al. (2016: 42) sum up that "in a browser-based application, Inputlog knows which window is on focus. Successive keystrokes can accordingly be associated with the web page in focus. In this way web searches can be tracked and reconstructed".

Daems et al. (2016) were one of the first to demonstrate the advantages of these tools in their study. The authors examined the effectiveness of consulting external resources during translation and PE using InputLog and CASMACAT, an interactive PE suite that tried to guess the next word that was supposed to be typed, which was quite innovative for the time. The authors confirm that "whereas most previous studies were limited to screen capture software to analyze the usage of external resources, [they] present a more convenient way to capture this data, by combining the functionalities of CASMACAT with those of InputLog, two state-of-the-art logging tools" (Daems et al. 2016: 111). In total, they analyzed 80 translation sessions from English to Dutch of ten M.A. students, 40 of which were completed from scratch and 40 while post-editing MT output. The authors chose eight newspaper articles on a variety of subjects and manipulated them to keep them comparable with regards to complexity, sentence length, number of words and sentences, etc.

After the experiment, the authors analyzed the focus events provided by InputLog and manually grouped them into distinct categories dependent on the external resource used, e.g., search engines, concordancers, dictionaries, encyclopedias, and others. After conducting several statistical tests, the authors formulate the following conclusions: Overall, the participants needed significantly more time to complete the tasks when translating from scratch than when post-editing, which is in line with other empirical studies (e.g. Plitt & Masselot 2010). The types of consulted resources do not differ between the two conditions, however "significantly less time is spent in encyclopedias and other types of resources compared to dictionaries, concordancers and search engines, for both types of translation" (Daems et al. 2016: 130). The distribution of allotted time to the individual resources is similar

between human translation and PE. Nonetheless, participants spend less time in research in the PE condition than in the human translation condition. This is indicative of more efficient information processing, because in the PE condition, the participants managed to complete the tasks quicker and needed less time to find the correct information and solution to their translation problems than in the human translation condition. This is mirrored in the quality analysis: "We can therefore conclude that there is no significant difference in overall quality between both types of translation (PE and human translation)" (Daems et al. 2016: 125). Achieving the same quality in a shorter amount of time means better overall efficiency. The authors made a final interesting finding when triangulating time and quality:

> When looking at post-editing, longer consultation of external resources was accompanied by higher overall error scores, whereas the opposite was true for human translation, where longer consultation of external resources was accompanied by lower overall error scores. This leads us to believe that participants are more successful in problem solving by consulting different resources when translating than when post-editing. This finding is in line with the suggestion by Yamada that post-editing requires different skills from human translation (2015). (Daems et al. 2016: 131)

For the present study, the research gap can be defined as follows: First, we want to look deeper into the translation strategies, especially the research strategies of lay translators (pupils) in comparison to novice or intermediate translators (students). Secondly, we wish to investigate how GenAI is used as an external resource in contrast to other translation tools or online resources within the translation process. The following experiment will shed light on these research questions.

4. Method and experimental design

The experiment was conducted at the Rudi-Stephan-Gymnasium in Worms, Germany, and at Faculty 06 of the Johannes Gutenberg University in Germersheim, Germany. Both the Johannes Gutenberg University and the Rudi-Stephan-Gymnasium are partners in the ForThem Alliance (Forthem n.d.). In accordance with German legislation, all academic studies conducted in educational establishments must undergo a comprehensive examination concerning ethics, data protection, and data security at the state level. It should be noted that the present study complied with all relevant legal and ethical requirements of the school advisory authority of the federal state of Rhineland-Palatinate.

First, the participants were informed about the scope of the study, the data collected, the anonymous analysis of the data, their right to withdraw from the study and to have their data deleted, in accordance with university policy. This privacy statement was signed by the participants or, when underage, their guardians.

Second, the participants were required to complete an anonymous five-minute English test, namely the Cambridge University English Assessment Test for Schools. The test comprises 25 multiple-choice items, with a single correct response for each. Subsequently, the participants were automatically assigned a score between 0 and 25, indicating the number of correct responses they provided. Subsequently, the scores of each participant were documented. The English test was employed as a quantitative measure of the participants' English proficiency, facilitating the process of data triangulation.

Third, participants were required to complete a 26-point questionnaire (see Appendix A), adapted from Oster (2019). The primary function was to assess usage behavior regarding tools and AI and to permit further data triangulation. In addition to providing basic biographical data (age, gender, class, native language, and the age at which they began studying English), participants were asked to indicate the frequency with which they utilize various digital tools, including search engines, online dictionaries, online encyclopedias, MT tools, and chatbots. They were also asked to rate the helpfulness of these tools on a three-point Likert scale.

Additionally, participants were requested to indicate all the tools they utilize for academic tasks and the rationale behind their selection of these tools (time savings, technological proficiency, lack of knowledge, complex assignments, lack of motivation). Additionally, participants were queried as to whether they would endorse the integration of chatbots, search engines, online dictionaries, online encyclopedias, and MT tools within the academic setting. They were also asked whether they would be interested in learning more about these tools in an educational context and whether they would prefer to see these tools utilized more frequently in the classroom (yes, no, undecided).

Fourth, to ensure that all participants had a uniform understanding of the capabilities and applications of LLMs such as ChatGPT, a brief video (Bucher & Humpa 2023) was presented to address these topics. The participants were now prepared to start the experiment.

The experiment comprised two tasks: an English reading task and a translation task from German to English. In this paper, we only shed light on the translation task. Nonetheless, each task was presented in two conditions. In the first condition, the participants were permitted to utilize all available online tools, apart from LLMs/chatbots. In the second condition, the participants were only permitted to employ LLMs/chatbots. For each task, there were two texts, one for each condition.

In conclusion, each participant was tasked with working on four texts, two of which were to be translated into English (see Appendix C) and two of which were intended for reading comprehension. The objective was to ascertain the influence of conventional and innovative AI tools on the participants' output. To reiterate, the analysis of the reading task is beyond the scope of this paper. The texts were presented in a pseudo-randomized order (see Appendix B) to prevent any participant from working on the same text in both conditions. Furthermore, this method ensured the elimination of fatigue and task order effects, as no participant was presented with the same sequence of texts, tasks, and conditions.

To test for interactivity, the program InputLog, a keylogging software (Leijten & van Waes 2013), was employed. InputLog enables researchers to precisely record and reconstruct the writing processes of individuals engaged in text composition on a computer. For the purposes of this experiment, InputLog Version 8.0.0.17 (beta) was utilized.

To enhance the reconstruction of the writing process, the screen of the participants was additionally recorded using OBS Studio (OBS). This approach facilitated data triangulation, monitoring of data, and a chronological understanding of the participants' search queries and prompting behavior, in addition to the exported Excel sheets provided by InputLog.

The quality of the translation task was evaluated using an MQM-based approach (MQM Council). Two independent reviewers, both faculty researchers, were instructed to mark any errors identified according to the custom MQM error typology (American Translators Association) (see Figures 2a & 2b). They were then asked to compare their notes and finally submit a single file containing all the agreed upon annotated errors.

The initial cohort consisted of 13 students from the Rudi-Stephan-Gymnasium in Worms, Germany (here: pupil participants, 'P'). Of these, five were in their penultimate year of secondary education (11th grade), while the remaining eight were in their final year (12th grade). The age range of the participants was between 16 and 18 years. The following biographical information about the participants was obtained directly from the questionnaire. Of the participants, nine identified German as their native language (69.23%, as indicated in Table 1), while one indicated English, two selected other languages (Albanian, Urdu), while one did not specify a native language. In this section of the questionnaire, multiple options were permitted. Seven of the participants are male and six are female. Five of the participants have attended English classes in school since the first grade, five since the third grade, and three since the fourth grade. Twelve of the participants plan to choose English as a subject in their final exams; one did not. All the pupil participants consume English media frequently or very frequently (100%, see Table 1).

The following table provides a visual representation of the biographical data of the participant groups. While we do acknowledge the imbalance of the number of participants and gender, we agreed on a pragmatical approach. The gender distribution amongst the student group is representative of all students in our faculty. Regarding the number of pupil participants, a small percentage of the initially planned pupils showed up for the study.

Table 1: Biographical and meta data of participant groups

Participants	Pupils	Students
N Participants	13	55
Mean Age	16.77	25.75
Male	7	9
Female	6	46
Other	0	0
German as native language %	69.23%	45.45%
Consume media in English frequently or very frequently %	100.00%	83.64%
Use online dictionaries frequently or very frequently %	15.38%	89.09%
Use machine translation tools frequently or very frequently %	69.23%	69.09%
Use chatbots frequently or very frequently %	38.46%	36.36%
Have been using chatbots for at least 6 months or longer %	46.15%	61.82%

Additionally, all the respondents indicated that they utilize chatbots. Five use them frequently or very frequently (38.46%, see Table 1), one uses them sometimes, and seven use them rarely. Seven pupil participants indicated that they found chatbots to be very helpful, while six indicated that they found them somewhat helpful. No respondents indicated that they found chatbots not helpful. Eight of the 13 pupil participants expressed uncertainty about the use of chatbots for schoolwork, while four were in favor and one was opposed.

Twelve of the 13 participants favored the use of other, conventional, tools, such as search engines, online dictionaries, online encyclopedias, and MT tools for schoolwork, with one participant remaining undecided. In addition, nine pupil participants indicated a desire to gain further insight into the potential applications of chatbots in school classes, while three remained undecided and one expressed opposition. Ten pupil participants utilize search engines

very frequently, two use them frequently and one sometimes. Eleven pupil participants found search engines to be very helpful, while two found them somewhat helpful.

No pupil participant reported using online dictionaries very frequently; two do so frequently, seven do so sometimes, two rarely, and two never do. This indicates that a total of 15.38% of the pupil participants use online dictionaries very frequently or frequently (see Table 1). Four respondents indicated that they find online dictionaries to be very helpful, eight rated them as somewhat helpful, and one rated them as not helpful. It is notable that online encyclopedias are not utilized very frequently by any of the pupils. Eight respondents indicated that they frequently use online encyclopedias, while five respondents indicated that they sometimes do so. Seven respondents stated that they find online encyclopedias to be very helpful, while six respondents indicated that they find them to be somewhat helpful.

Regarding MT tools, such as Google Translate or DeepL, four respondents indicated that they use them very frequently, five respondents indicated that they use them frequently, three respondents indicated that they use them sometimes, and one respondent indicated that they use them rarely. This indicates that a total of 69.23% of pupil participants use MT tools very frequently or frequently (see Table 1). Ten respondents indicated that they found MT to be very helpful, while three respondents indicated that they found it to be somewhat helpful. In this section of the questionnaire, participants were permitted to select multiple options. Twelve participants utilize search engines for their schoolwork and homework, five also employ online dictionaries, five utilize MT, seven utilize online encyclopedias, and only four utilize chatbots. The primary reasons that the pupil participants utilize digital tools for their schoolwork were time savings (11), technological proficiency (2), a lack of knowledge (8), complex tasks (7), a lack of motivation (2), and other factors (1).

The student cohort from Faculty 06 of the Johannes Gutenberg University in Germersheim comprised 55 individuals (here: student participants, ‘SP’), with an age range between 19 and 48 years and an average age of 25 years. Of the total number of this group of participants, 46 were female and nine were male. 27 of the students were pursuing a bachelor’s degree, while 25 were pursuing their master’s. Three students did not respond to this inquiry. Of the total student sample, 25 specified German as their native language (45.45%, see Table 1), while the remainder identified Spanish, Italian, Greek, Arabic, Portuguese, Russian, Turkish, Chinese, Polish, and Tamil as their native language.

Eight of the student participants have been attending English classes since the first grade, one since the second grade, eight since the third grade, four since the fourth grade, eight since the fifth grade and one since the sixth grade. Four respondents indicated that they had attended English classes since primary school, which could be any of the first four grades. Additionally, 21 of the 55 participants did not respond to this question. Of the 55 respondents, 44 indicated that they had taken advanced language classes, 37 of whom had also taken advanced English classes.

A total of 36 of the 55 student participants indicated that they consume English-language media very frequently, with ten participants reporting that they do so frequently, four participants reporting that they do so sometimes, and two participants reporting that they do so rarely. This shows that a total of 83.64% of the student participants engage with media in English frequently or very frequently (see Table 1). Additionally, four of the student participants utilize chatbots very frequently, while 16 students employ them frequently, 17 use them sometimes, ten utilize them rarely, and eight never do so. This suggests that a total of 36.36% of the student participants interact with chatbots at a high frequency (see Table 1).

21 students indicated that they find chatbots to be very helpful, while 27 students indicated that they find them to be somewhat helpful. Five students indicated that they find chatbots to be not helpful, while two did not respond to the inquiry. 25 students indicated that they believe chatbots should be utilized for academic purposes, while twelve expressed opposition, 17 remained undecided, and one did not respond to the inquiry. Nevertheless, 53 of the respondents expressed a preference for utilizing alternative conventional tools, including search engines, online dictionaries, online encyclopedias, and MT tools, for academic purposes, while two respondents were against this. Additionally, 48 students expressed interest in learning more about integrating chatbots into the classroom, while one student expressed no interest, and six students remained undecided.

45 of the student participants utilize search engines very frequently, nine do so frequently, and one does so sometimes. 48 of the students find search engines to be very helpful, seven find them somewhat helpful. 27 of the students employ online dictionaries very frequently, 22 frequently, four sometimes, and two rarely. This indicates that 89.09% of the student group utilize online dictionaries very frequently or frequently (see Table 1). 43 of the students surveyed indicated that they find online dictionaries to be very helpful. A further twelve respondents indicated that they find online dictionaries somewhat helpful. Eleven of the participating students utilize online encyclopedias very frequently, 23 do so frequently, 17 sometimes, and four rarely. A total of 30 students indicated that they find online encyclopedias to be very helpful, while 25 students indicated that they find them to be somewhat helpful.

Regarding MT tools, such as Google Translate or DeepL, 18 students indicated that they use these tools very frequently, 20 students indicated that they use them frequently, 15 students indicated that they use them sometimes, one student indicated that they use them rarely, and one student indicated that they never use them. This means that a total of 69.09% of the participating students use MT tools very frequently or frequently (see Table 1). A total of 33 students participating in the study indicated that they find MT to be very helpful, while 21 students find it helpful. In response to the question of which tools they regularly employ for the completion of homework and assignments, 52 students indicated the use of search engines, 49 the use of online dictionaries, 41 the use of MT, 34 the use of online encyclopedias, and 25 the use of chatbots. The primary motivations for utilizing digital tools for academic endeavors were identified as time savings (45), technological proficiency (18), a lack of knowledge (41), complex tasks (27), a lack of motivation (10), and other factors (5).

5. Results

5.1. Errors in translation

In this chapter we examine the results of the experiment. We will focus on translation competence and how it is reflected in research strategies and final output quality. Figure 1 shows the quality analysis of the translation task.

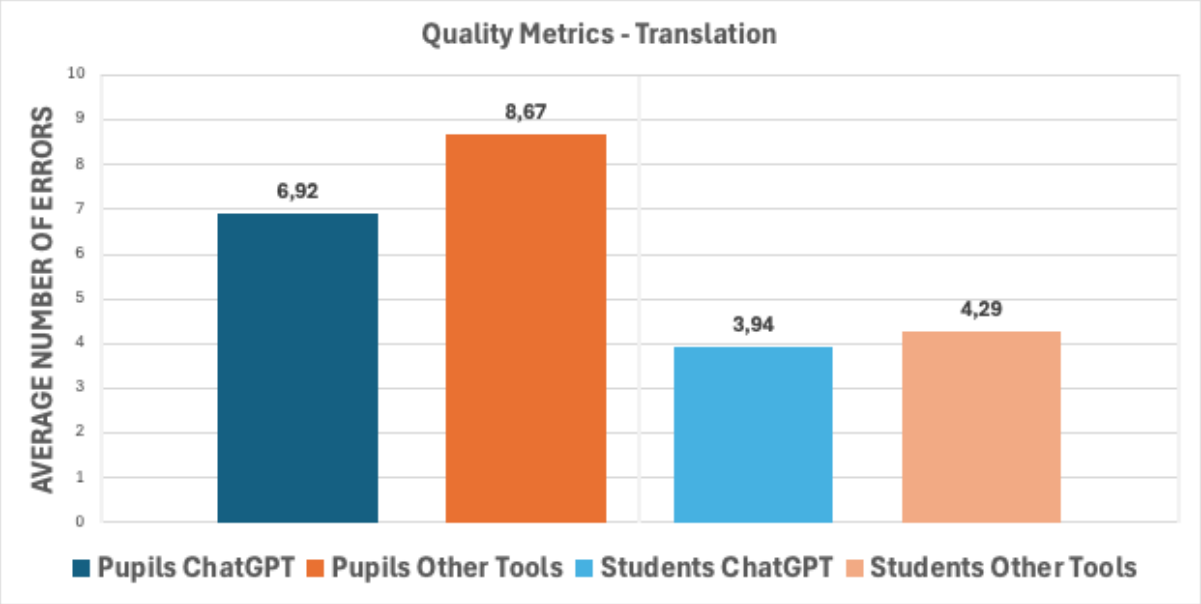

Figure 1: Average number of errors in both conditions (Quality). Both participant groups

The average number of errors is shown on the Y-axis. The dark blue bar represents the number of errors made by pupils in the ChatGPT condition, while the dark orange bar represents the number of errors made by pupils in the Other Tools condition. The light blue bar represents the number of errors made by students in the ChatGPT condition, while the light orange bar represents the number of errors made by students in the Other Tools condition. In the ChatGPT condition, pupils made an average of 6.92 errors. In the Other Tools condition, pupils made 8.67 errors on average. In the ChatGPT condition, students made an average of 3.94 errors whilst making 4.29 errors on average in the Other Tools condition. Figures 2a and 2b show the average detailed error annotation scores for both groups per participant in both conditions. The dark blue bars represent the average number of errors in each annotation category for the pupil participant group in the ChatGPT condition, while the dark orange bars represent the average number of errors in each annotation category for the pupil participant group in the Other Tools condition.

The light blue bars represent the average number of errors in the respective annotation category of the student participant group in the ChatGPT condition, while the light orange bars represent the average number of errors in the respective annotation category of the student participant group in the Other Tools condition.

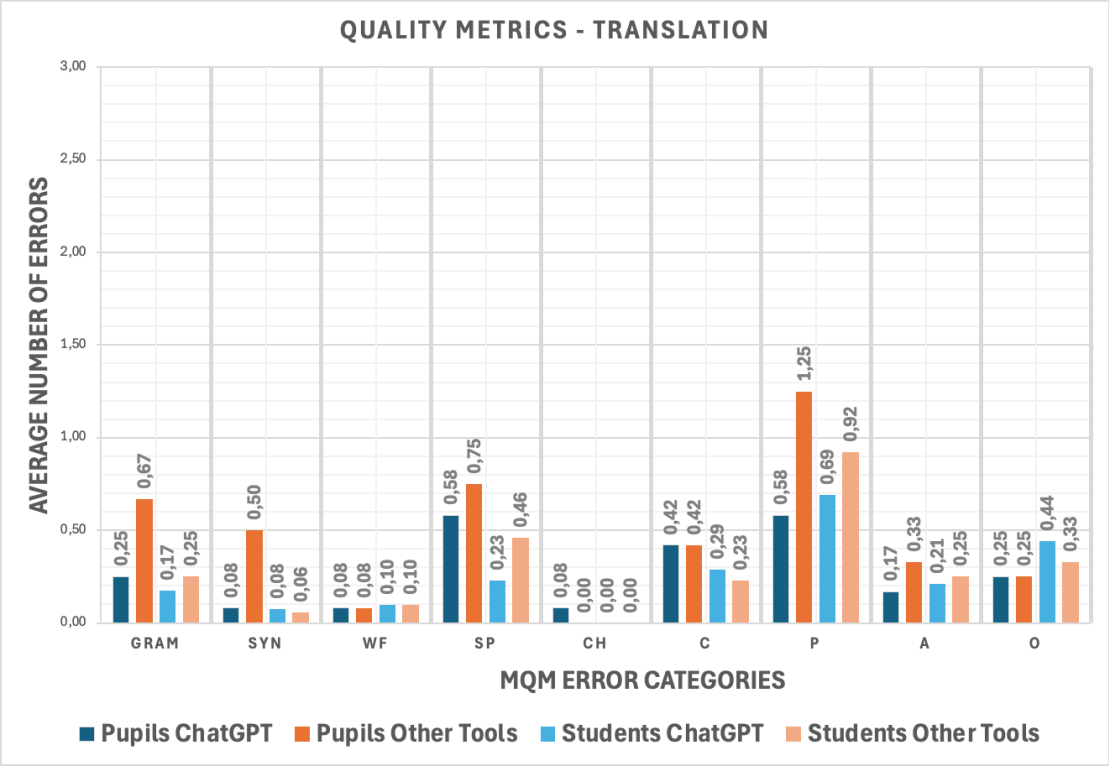

Figure 2a: Average number of errors in the respective annotation category in both conditions (Quality) per participant

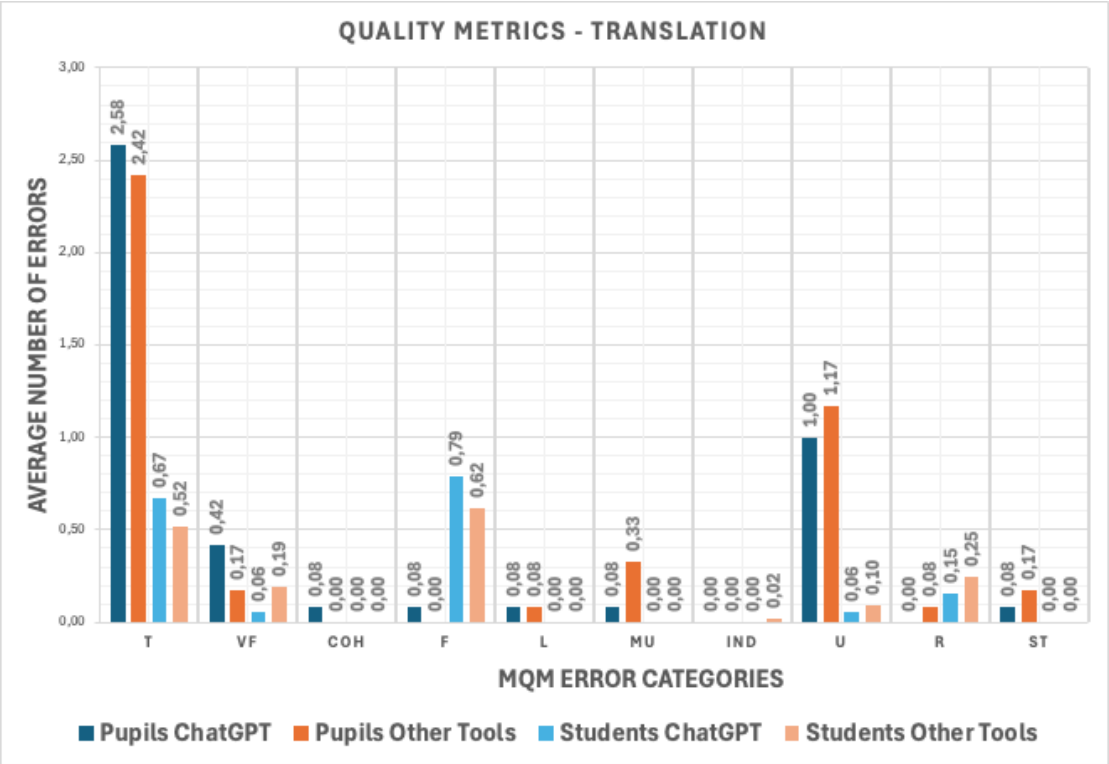

Figure 2b: Average number of errors in the respective annotation category in both conditions (Quality) per participant

The Y-axis shows the average number of errors, on the X-axis from left to right: Grammar (GRAM), Syntax (SYN), Word Form (WF), Spelling (SP), Character (CH), Capitalization (C), Punctuation (P), Addition (A), Omission (O), Terminology (T), Verb Form (VF), Cohesion (COH), Faithfulness (F), Literalness (L), Misunderstanding (MU), Indecision (IND), Usage (U), Register (R), and Style (ST).

In the category of terminology, it is found that pupils in the ChatGPT condition made 2.58 errors on average compared to the students' average of 0.67 errors. In the Other Tools condition, pupils made 2.42 errors on average compared to the students' average of 0.52 errors. In the category of faithfulness, it is found that pupils in the ChatGPT condition made 0.08 errors on average compared to the students' average of 0.79 errors. In the Other Tools condition, pupils made no errors, while the students' average was 0.62.

5.2. Research strategies

Next, we analyzed the target groups' research strategies. Students use a much greater variety of different tools than pupils do. The 13 pupils used seven different tools in total. 55 students used 23 tools in total. On average, students used twice as many tools as pupils: While pupils used on average only 1.77 different tools per translation session, students used 3.15 different tools. The sheer number of different tools used by students is indicative of their translation competence. They search and choose an appropriate tool to solve a given translation problem. In contrast, pupil participants resorted to one-size-fits-all solutions, the MT tools. And as the average number of tools used is smaller than 2, it means that many pupils just used the MT output as their only problem-solving strategy.

These findings are summed up in Figures 3 and 4. Figure 4 is analogous to Figure 3 but focuses on the student participants. These figures do not take into consideration how many times one specific tool was accessed by a participant but rather show the number of participants who utilized a tool at least once as part of their problem-solving strategy to create a final translation. In addition, of course one participant can use more than just one tool, i.e., the basic quantity of Figures 3 and 4 can be higher than the actual number of unique participants. This enables us to see the relation of the types of tools used rather than the individual tools, which is analogous to the methodology of Daems et al. (2016). Thus, we can compare the types of tools the two target groups access to translate the source texts. For example, Bing and Google are both search engines; Google Translate, DeepL Translate, and Youdao Translate are all neural MT tools etc. Pupil participants only used MT tools, search engines and online dictionaries. 13 individual pupils accessed seven different tools. In sum 23 pupil participants accessed all tools combined at least once. Roughly 69% of all tools accessed at least once by the pupil participants were MT tools, 22% were search engines and 9% were online dictionaries. Moreover, it is interesting to note that MT tools were the most accessed across both target groups (16 in total for pupils and 53 for students). Almost every student used DeepL to complete the translation task, showing that DeepL is the preferred NMT tool for problem solving. The pupils did not use any other NMT tool, so we can conclude that they are aware of the existence of MT tools, and the overwhelming majority choose DeepL to translate the source text.

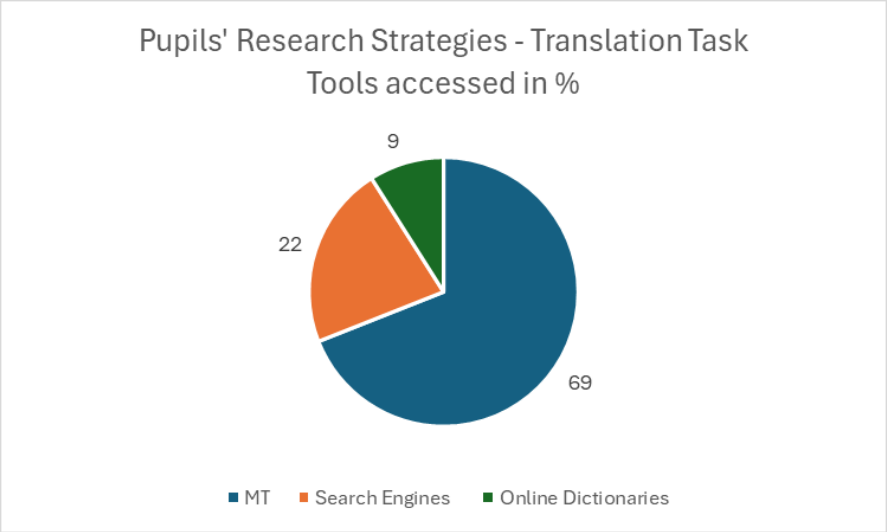

Figure 3: Tools used by the pupil group in the Other Tools condition. In percentages

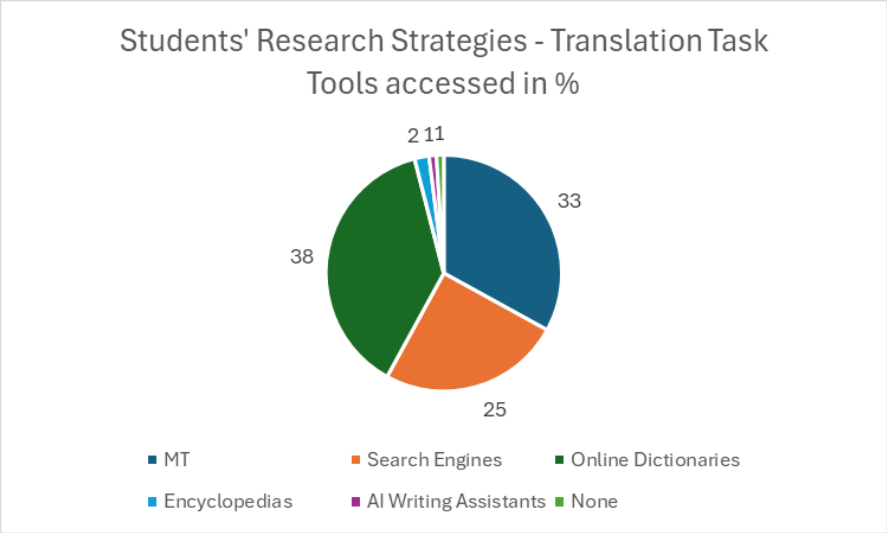

Figure 4: Tools used by the student group in the Other Tools condition. In percentages

Roughly 33% of the student participants used MT, 25% search engines, 33% online dictionaries, 2% (online) encyclopedias, 1% AI writing assistants and 1% did not use any tool at all. As elaborated upon before, students did not just use more tools on average, but also a greater variety of tools (see Figure 4 legend). This is indicative of translation competence, as solving different translation problems calls for different tools. Also, depending on the translation stage (drafting, translation, revision etc.; see EMT Board and Competence Task-Force 2022) different tools may be considered.

Furthermore, we also looked at the translation strategies in the ChatGPT condition, where we also expected to find indicators for translation competence development. Figures 5 and 6 refer to the prompting behavior in the translation task when using ChatGPT only. Figure 5 shows the average number of prompts pupils and students wrote when translating the texts in the ChatGPT condition.

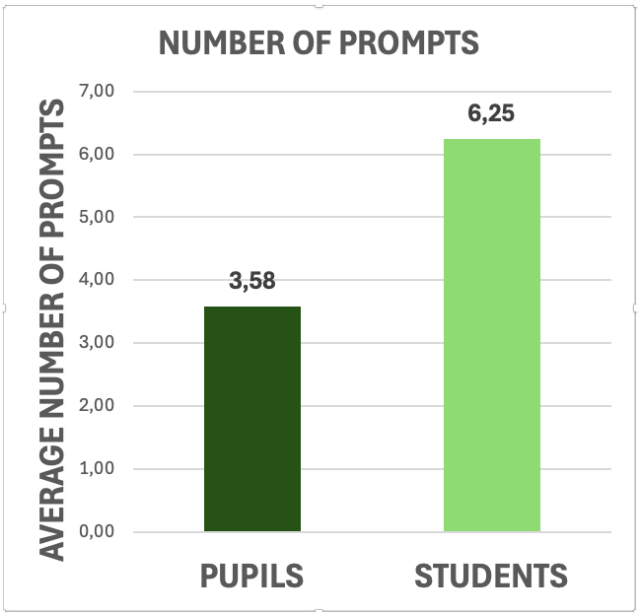

Figure 5: Number of prompts written on average. Both groups, ChatGPT condition

On average, pupils wrote 3.58 prompts, while students wrote 6.25 prompts. This suggests that students engaged in significantly more back-and-forth interactions with the tool compared to pupils. Next, we examined the average number of words per prompt.

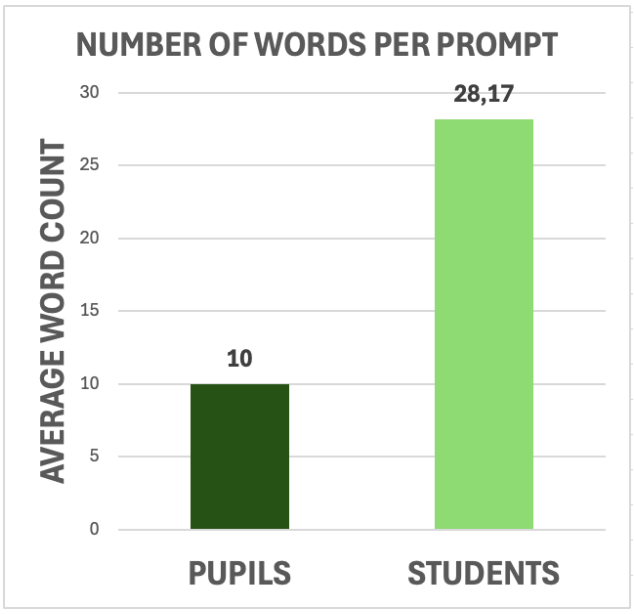

Figure 6: Number of words per prompt written. Both groups, ChatGPT condition

Figure 6 shows that, on average, each of the pupils' prompts was 10 words long, whereas the students' prompts were almost 3 times longer at 28.17. These results are indicative of the students' increased interactivity. Not only did they prompt more, but their prompts were

more elaborate. Providing more (relevant) context to the LLM yields more favorable results, a technique referred to as prompt engineering:

> By employing prompt engineering techniques, academic writers and researchers can unlock the full potential of language models, harnessing their capabilities across various domains. This discipline opens up new avenues for improving AI systems and enhancing their performance in a range of applications, from text generation to image synthesis and beyond. (Giray 2023: 2629)

Thus, we observe that students' translation competence is adaptable to innovative tools like ChatGPT.

6. Discussion

These results clearly show the effect of translation competence development. In all our analyses, university students performed better than school pupils.

One interesting source of errors in both target groups was terminology. On average, approximately 28% of all errors made by pupils in the Other Tools condition were terminology errors in contrast to the students' average of 12%. In sum, the majority of the pupils' and students' mistakes were terminology errors. Moreover, both target groups made more terminology errors on average when using ChatGPT than when not using ChatGPT (see Figures 2a and 2b). On average, roughly 37% of all errors pupils made in the ChatGPT condition were terminology errors. In contrast, on average only 17% of all errors made by students in the ChatGPT condition were terminology errors.

For example, P10 translated "Das Gefühl hoffnungsloser Traurigkeit entsteht in den unzähligen **mittleren** [...] Städten" as "The feeling of hopeless sadness arises in countless **medium sized** [...] citys[sic!]" (boldface by authors) in the ChatGPT condition. The terminologically sound translation would be 'middle-sized'. Vardaro et al. (2019) argue that terminology errors are amongst the most prevalent error categories in NMT output. It is not possible to comment on the prevalence of terminology errors in the raw ChatGPT output. However, it can be assumed that the participants in this study exhibited a level of trust comparable to that reported by Huschens et al. (2023). In their study, they presented participants with the same texts in different UI conditions and found that "participants tend to attribute similar levels of credibility" to either UI condition (Huschens et al. 2023: 1). Furthermore, participants rated "AI-generated content as being clear and more engaging" (Huschens et al. 2023: 1). We therefore speculate that our participants possibly overlooked terminological issues due to the false sense of credibility attributed to the ChatGPT output.

An interesting observation in Figure 2b is that student participants made more faithfulness errors than pupils. For example, SP03 translated this source sentence "Die moderne Zivilisation auf einen nachhaltigen Weg zu bringen, **gleicht mehr und mehr** dem Versuch, einen Deich zu halten, gegen den die Flut drückt." (boldface by authors) as "Trying to make our modern world more sustainable **is a bit like** trying to patch up a leaking dam against an oncoming flood" (boldface by authors). The annotators marked the bold phrase in the target text as a faithfulness error, as it should have been something like 'more and more' or 'continuously more' and not just 'a bit', albeit not being completely wrong as is. Also, for

example SP07 translated the source sentence "Vielleicht entsteht dieses Gefühl nicht in den inneren Zirkeln der Metropolen, deren Zentren **immer noch begeistern können**." (boldface by authors) as "This feeling might not arise in the inner circles of the metropolises, whose city centers are **still cause for wonder**" (boldface by authors). There is no causal relationship in the source text. The target text is not completely wrong either, it simply adds a nuance that is not present in the source text. Pupil participants did not make such mistakes, which leads us to assume that the students who display a greater translation competence in comparison to the pupils tend to choose formulations that are farther away from the source material and therefore not entirely accurate. During their studies, they learn about equivalence (Catford 1965) and Skopos theory (Vermeer 2006), which may lead them to take greater risks to translate the *meaning* of the text and not *just* the words. Put differently, pupils resorted to more literal translation solutions, while students tended to allow themselves to deviate from the literal and textual constraints of the source text. We believe that this has implications concerning the literal translation hypothesis theorized by Malmkjaer (2005) and Chesterman (2011). Schaeffer et al. (2016) define it as follows: "A literal translation is the first or default solution a translator applies to the source text, often only as an interim solution before a less literal translation is considered or produced" (Schaeffer et al. 2016: 189). We believe that literality interacts with translation competence development and plan to conduct further research into this matter.

University students display a more sophisticated use of tools which is indicative of better translation strategies. As has been shown in Figures 3 and 4, students not only use a wider variety of tools but also use, on average, more tools per task than pupils. To better illustrate these findings, we will present examples of the translation sessions.

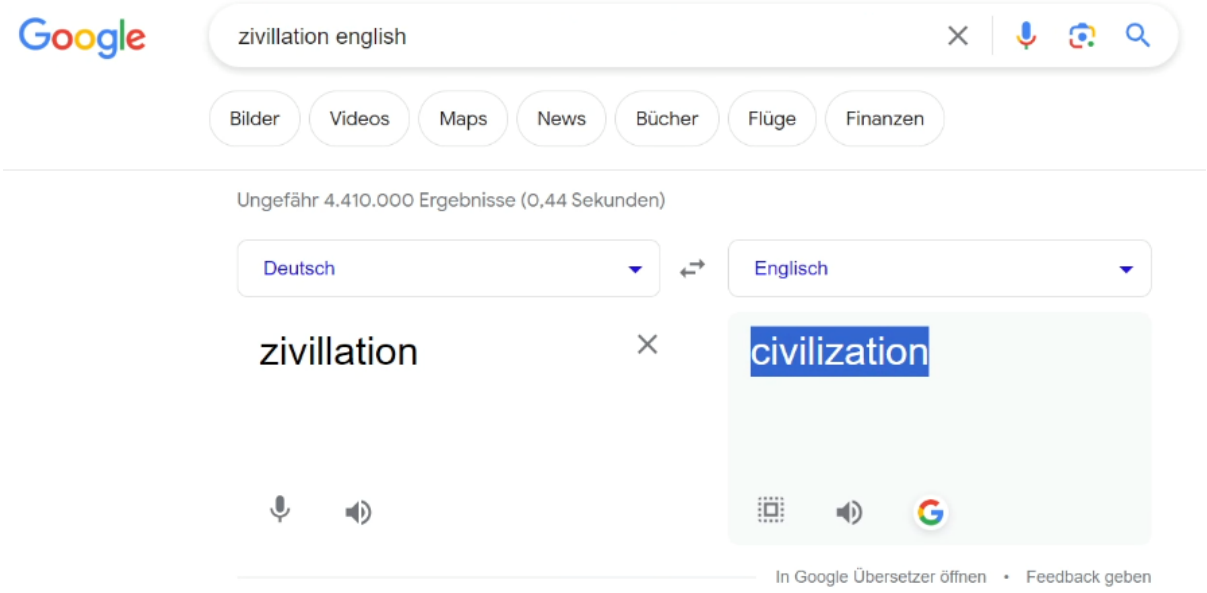

Figure 7: P01, Other Tools condition

In Figure 7, P01 typed *zivillation english* in the Google search bar. Google's first result was its own translation program, Google Translate. We observed that the pupil did not specifically look for an MT tool, they just wanted to solve the translation problem. This goes

to say, that the participant did not think about which tool would best suit their needs and just chose an arbitrary solution provided by Google. This is critical, because in this example the pupil used an NMT-program as a dictionary and did not provide any context for the tool, which can lead to mistranslations. In addition, the participant did not copy the word directly from the source document into the search bar, but chose to type it in manually, resulting in a spelling error. The spelling error did not affect the performance of the tool, as Google Translate correctly identified the word and provided a correct translation of the otherwise misspelled word. This performance is attributed to the NMT-architecture, as a statistical MT tool would most likely not be able to correctly identify the word (Bertoldi et al. 2010). The lack of translation competence also becomes evident in the ChatGPT condition in Figure 8.

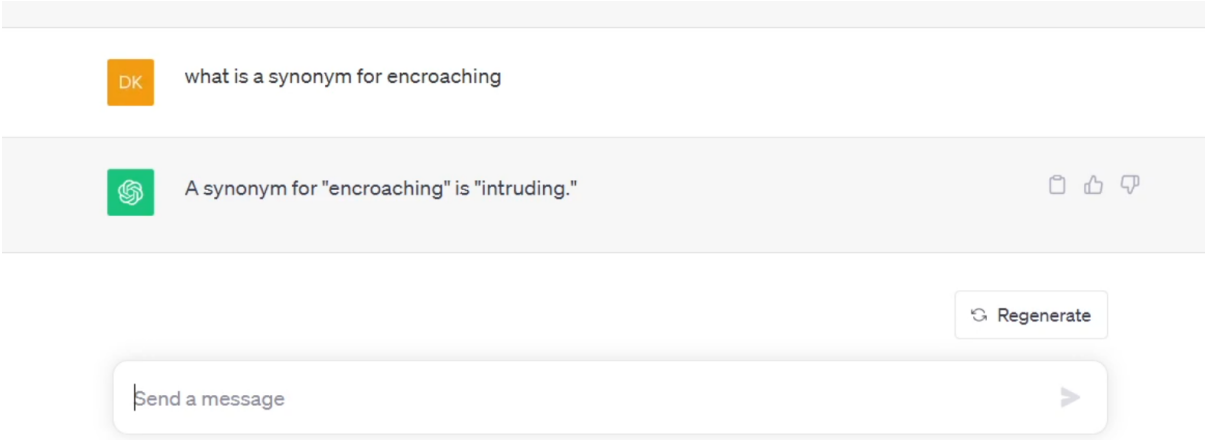

Figure 8: P03, ChatGPT condition

P03 asked ChatGPT to provide a synonym for the word *encroaching*. While it is spelled correctly, the context or register in which the target word should be placed was not specified. Not every synonym for a word can replace any other in every context and situation. Furthermore, the participant did not further research/prompt to assess whether *intruding* could be used synonymously in the same context as *encroaching* can, which resulted in a terminology error. The German source sentence "[d]ie moderne Zivilisation auf einen nachhaltigen Weg zu bringen, gleicht mehr und mehr dem Versuch, einen Deich zu halten, gegen den die Flut drückt." was translated as "Trying to steer modern civilization onto a sustainable path increasingly resembles attempting to maintain a levee against an intruding flood." While there are other errors in this translation, we will focus on the word *intruding*. Surprisingly, the participant had already copied the entire text into ChatGPT and kindly asked the tool to translate the text. After ChatGPT provided the translation, the participant wrote the prompt shown above. We cannot know why the participant decided to ask whether *intruding* was a good fit or not. Something motivated the participant to look for an alternative. What is also interesting is that ChatGPT had context in a previous prompt. ChatGPT did not make a connection to the previous prompt, and the participant did not verify the result. Instead, they blindly trusted the tool's output, which is in line with previous studies (e.g. Huschens et al. 2023).

Of course, student participants also had to deal with the problem of finding the correct word for the given context. SP06 demonstrated their translation competence by asking ChatGPT to verify that the machine output was applicable to the specific context, as shown in Figure 9.

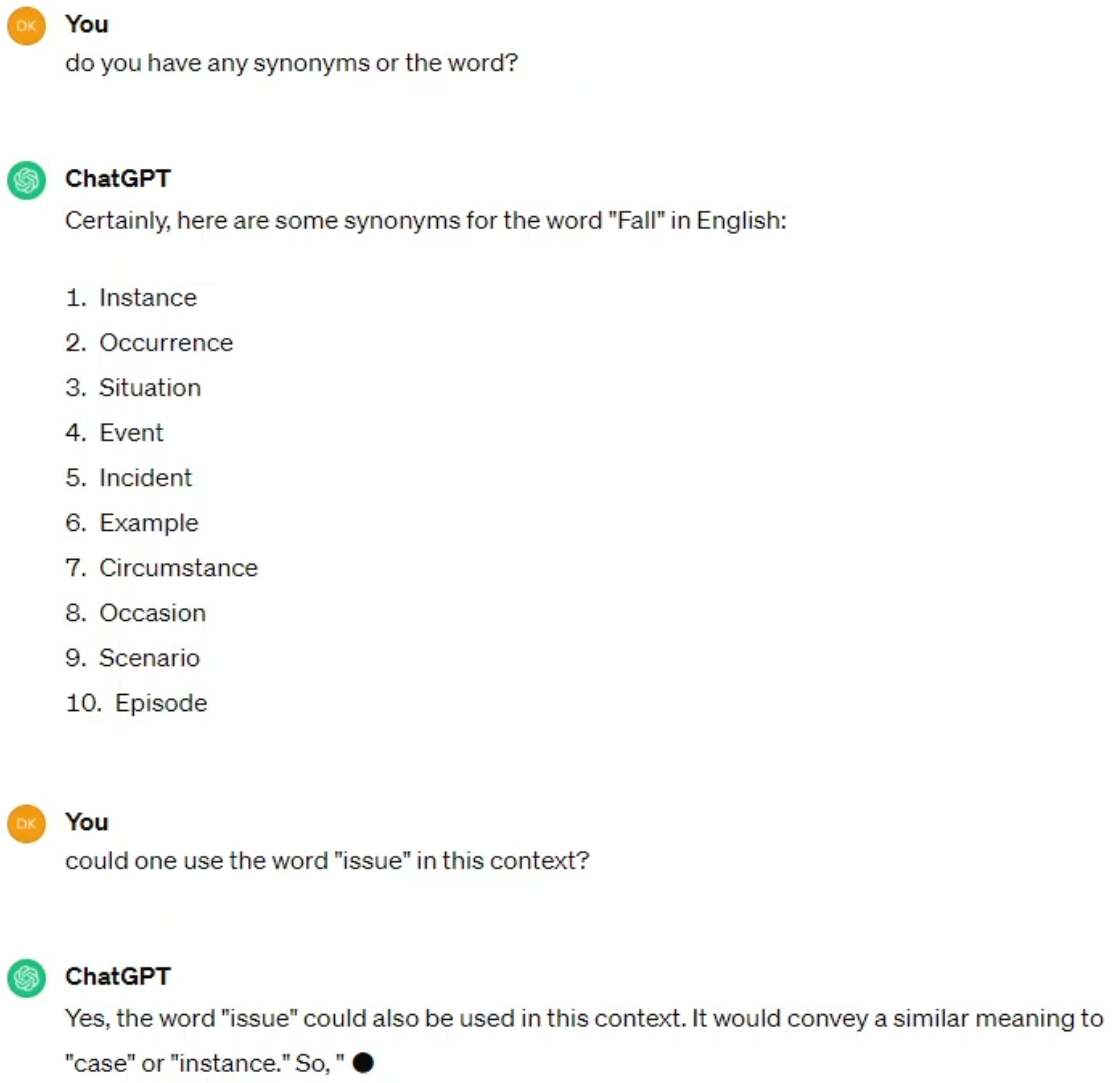

Figure 9: SP06, ChatGPT condition

Not only did they request multiple synonyms, but they also checked whether a new word, that was not included in ChatGPT's reply, would still work in that context. We also observe that SP06 asks questions. These questions are addressed to ChatGPT as if ChatGPT were a real person. Notice the difference: P03 wrote "what is a synonym for encroaching" vs. "do **you** have any synonyms or [sic!] the word?" (boldface by authors). This *humanization* of the tool is a novel form of interactivity, which conventional AI tools like NMT cannot achieve. Furthermore, this is in line with what Kußmaul refers to as *learning buddy* (see Kußmaul 2015).

This *learning buddy* phenomenon can be observed in Figure 10 as well. For example, SP10 primarily translated *from scratch* and used ChatGPT very sparingly only to verify individual words or phrases. However, when they finished their translation, they pasted their entire translation into ChatGPT and asked whether it would improve anything ("würdest du an meiner Übersetzung bzw. der Grammatik was verbessern?", 'would you improve anything regarding my translation or grammar?'; translated by authors). ChatGPT took the additional context into consideration and then replied with a revised version and details on what had been changed. The student then copied that text into the Word document and made final adjustments before submitting their final translation. Of course, this did not lead to a flawless translation, as there are still many mistakes that ChatGPT did not correct, e.g., terminology, style, and register. The tool does not constitute a panacea for correcting deficiencies. It is, however, possible, that if the participant had more critically examined the tool output and prompted more concisely, more errors could have been avoided.

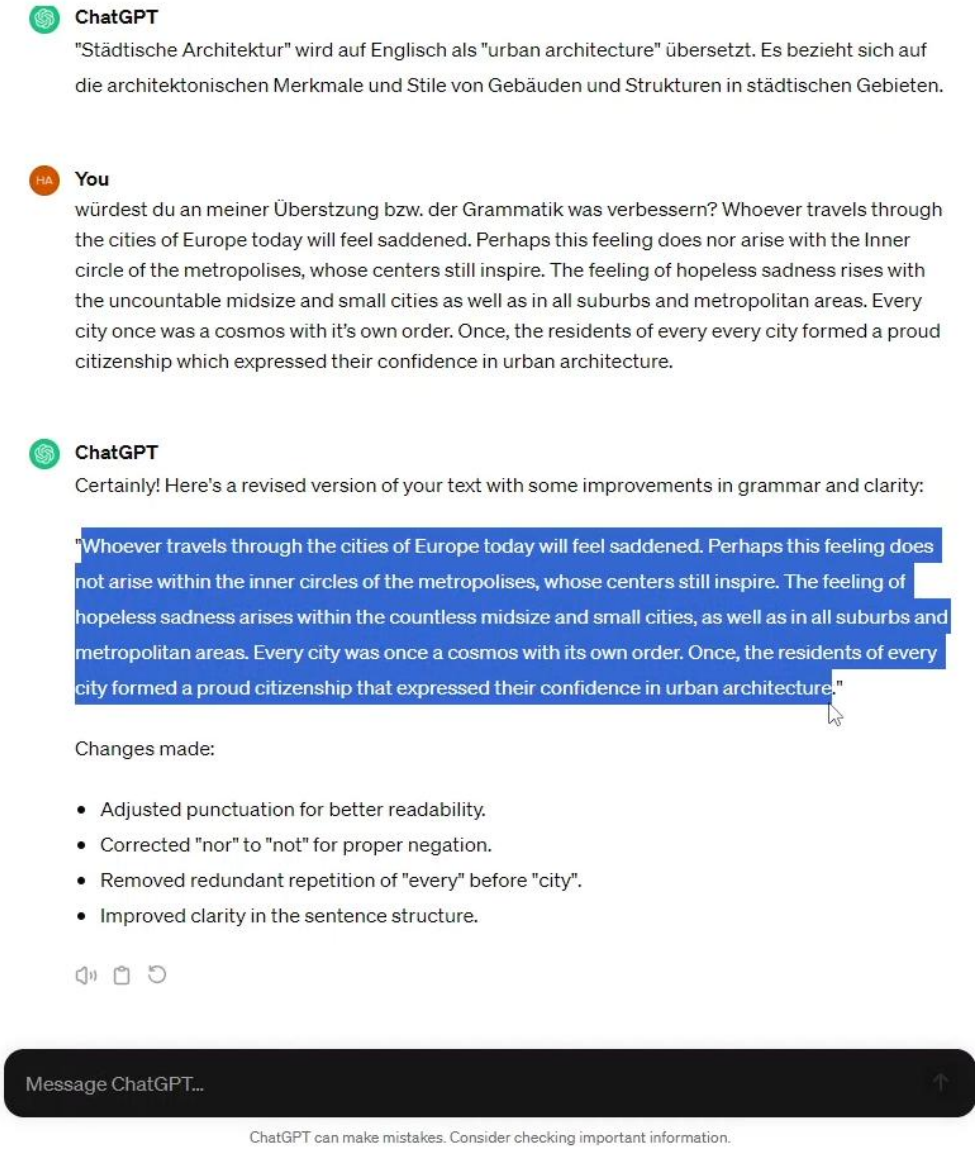

Figure 9: SP10, ChatGPT condition

In Figure 11, SP26 effectively interacted with ChatGPT as a MT tool and as a learning buddy to find a solution to a lexical problem. First, the participant pasted the entire source text into ChatGPT and requested an English translation and ChatGPT complied. Next, the participant demanded a more idiomatic translation. The participant compared the two outputs and post-edited in the Word window accordingly. Finally, the participant noticed that in the first machine output, ChatGPT had translated *Selbstbewusstsein* with 'self-confidence' but in the second output as 'self-assurance'. The participant then challenged the tool's inconsistent choice of words by demanding to know the difference between self-confidence and self-assurance. This strategy is called *chain prompting* and is used to improve the quality of ChatGPT's output (Sawalha et al. 2024).

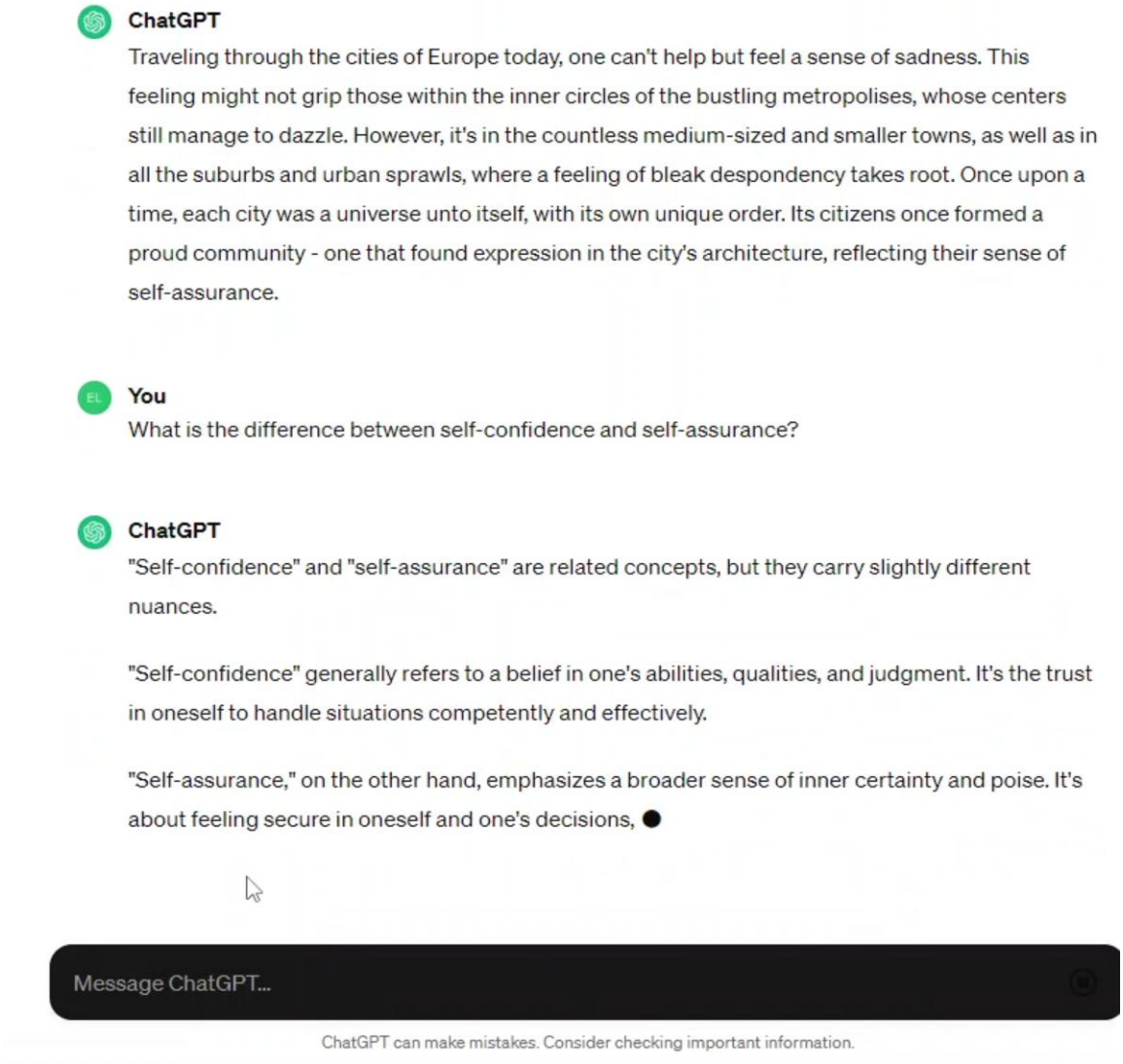

Figure 101: SP26, ChatGPT condition

In sum, SP26 first used ChatGPT as an MT tool by providing the entire source text. Then they leveraged the advantages of LLMs by requesting a second variant of the same source text, something simple NMT systems are not capable of. By cross-referencing the two machine outputs and choosing the more suitable variant, this participant avoided some mistakes. Then they used ChatGPT as a learning buddy to better understand the meaning of a certain word and make an informed decision. This troubleshooting process would have required the usage of many different tools in the non-ChatGPT condition.

7. Conclusion

The study presented in this paper compares the use of GenAI vs. traditional translation tools in foreign language learning in two groups, a group of pupils with little to no translation competence, and a group of translation students, who are developing their translation competence by studying translation at the Johannes Gutenberg University Mainz. From a methodological point of view, a scientifically valid and replicable experimental environment was successfully created, wherein all groups were subjected to identical conditions and comparable settings. However, internet connectivity issues did arise, though they did not impact the output of the translation process and thus the scope of the study. The number of pupil participants could be considered a limitation to the study.

The results show that GenAI has a positive effect on translation tasks. In terms of quality, we can gather two things from the results: 1) that the usage of ChatGPT boosted both groups' final translation quality, and 2) the more sophisticated translation strategies employed by students improve quality in both conditions (i.e. all online tools allowed except ChatGPT and only ChatGPT allowed). Students generally make fewer mistakes across both conditions, except in some categories, e.g., Faithfulness. The greater the translation competence, the less literal the translations are, which might be explained with the literal translation hypothesis. This is grounds for further research.

We also gather that students employ more sophisticated research strategies in both conditions: in the ChatGPT condition by prompting more and prompting more detailed. We can observe strategies of prompt engineering since they use chain prompting to refine their prompts or to let ChatGPT revise their texts. In the Other Tools condition, the students show their research competence by using a wider and more suitable array of tools.

The results show that the existing competence models and standards need to be extended in such a way that AI literacy is included alongside competencies that involve research, translation, and revision. Furthermore, the previous knowledge of the new generations of students must be considered. In addition, the PE competence models and standards may have to be updated, since effective interaction with GenAI-tools might make (conventional) PE of MT obsolete. Our results might indicate a paradigm shift away from MT as a tool producing draft translations – whose quality must be checked – towards collaborating with GenAI as a translation and revision buddy. There is still room for research concerning whether these learning buddy scenarios result in more efficient processes and products of higher quality. There is also still room for technological developments that integrate or combine GenAI with traditional CAT environments.

Nevertheless, the research underscores the transformative potential of AI technologies, such as ChatGPT, in the field of education. It highlights the necessity to equip educators with targeted information regarding the potential applications and constraints of these technologies to guarantee their meaningful integration into the everyday teaching process.

Regarding the next steps of our study, it is necessary to evaluate the data gathered from keylogging to ascertain the extent of the technical and temporal effort involved. This may be achieved by analyzing the number of keystrokes per working window, the number of window switches, and the time taken to complete each task. The results of our study demonstrated that the student participants exhibited superior quality output. If we were to verify that this was achieved with fewer keystrokes, it would serve as a definitive indicator of translation competence.

Another crucial element of the study is the acknowledgment that AI-backed tools, such as ChatGPT, have the capacity to enhance the learning process by offering tailored assistance for challenging tasks. The research results indicate that by involving a larger and more diverse group of subjects, the identified limitations can be addressed and a more comprehensive understanding of the potential benefits and challenges of using these technologies can be gained. A comparison with other AI tools could also be carried out to achieve these goals. Furthermore, the long-term effects of the integration of AI tools into educational contexts should be investigated. In addition, specific error categories and the effects of targeted training measures or model adaptations on the performance of ChatGPT in the classroom should be researched.

Appendix A: Fragebogen zur Studie „Einsatz von (KI-gestützten) Tools im Fremdsprachenunterricht“

Vom Versuchsleiter auszufüllen

Datum:	
VP-Nr.:	
Score Sprachtest	

Bitte füllen Sie den nachfolgenden Teil aus.

Alter	
Studiengang	□ BA □ MA □ Sonstiges:
Fachsemester	
Geschlecht	□ m □ w □ divers
Muttersprache	
Englisch als Studiensprache	□ ja □ nein
Konsumieren Sie Medien (z.B. Filme, Bücher, soziale Medien, Videospiele, Musik) in englischer Sprache?	□ sehr oft □ oft □manchmal □ selten □ nie

Nutzen Sie Suchmaschinen (z.B. Google, Bing, Ecosia)?	□ sehr oft □ oft □manchmal □ selten □ nie
Wie hilfreich finden Sie Suchmaschinen?	□ sehr □ etwas □ gar nicht
Nutzen Sie Online-Wörterbücher (z.B. Leo, Pons, Linguee)?	□ sehr oft □ oft □manchmal □ selten □ nie
Wie hilfreich finden Sie Online-Wörterbücher?	□ sehr □ etwas □ gar nicht
Nutzen Sie Online-Enzyklopädien (z.B. Wikipedia)?	□ sehr oft □ oft □manchmal □ selten □ nie
Wie hilfreich finden Sie Online-Enzyklopädien?	□ sehr □ etwas □ gar nicht
Nutzen Sie Online-Übersetzungstools (z.B. Google Translate oder DeepL)?	□ sehr oft □ oft □manchmal □ selten □ nie
Wie hilfreich finden Sie Online-Übersetzungstools?	□ sehr □ etwas □ gar nicht
Nutzen Sie Chatbots (z.B. ChatGPT)?	□ sehr oft □ oft □manchmal □ selten □ nie
Wie hilfreich finden Sie Chatbots?	□ sehr □ etwas □ gar nicht
Wie lange nutzen Sie schon Chatbots?	□ Heute zum 1. Mal davon gehört □ Seit 3 Monaten □ Seit 6 Monaten □ Länger als 6 Monate

Welche dieser Tools nutzen Sie regelmäßig für die Schule und/oder Hausaufgaben?	□ Suchmaschinen □ Online-Wörterbücher □ Online-Enzyklopädien □ Online-Übersetzungstools □ Chatbots □ Andere:
Kreuzen Sie die Hauptgründe an, warum Sie Tools nutzen.	□ Zeitersparnis □ Technologie-Interesse □ Fehlendes eigenes Wissen □ Schwierige Aufgaben □ Mangel an Motivation □ Andere:

Befürworten Sie die Nutzung von Chatbots im Studium?	□ ja	□ nein	□ weiß nicht
Befürworten Sie die Nutzung von Suchmaschinen, Online-Wörterbüchern, -Enzyklopädien und -Übersetzungstools im Studium?	□ ja	□ nein	□ weiß nicht
Möchten Sie mehr über den Umgang mit Chatbots im Unterricht lernen?	□ ja	□ nein	□ weiß nicht
Möchten Sie mehr über den Umgang mit Suchmaschinen, Online-Wörterbüchern, -Enzyklopädien und -Übersetzungstools im Unterricht lernen?	□ ja	□ nein	□ weiß nicht
Möchten Sie, dass Chatbots häufiger im Unterricht verwendet werden?	□ ja	□ nein	□ weiß nicht
Möchten Sie, dass Suchmaschinen, Online-Wörterbüchern, -Enzyklopädien und -Übersetzungs-tools häufiger im Unterricht verwendet werden?	□ ja	□ nein	□ weiß nicht
Denken Sie, dass Chatbots in Ihrer Berufspraxis eine Rolle spielen werden?	□ ja	□ nein	□ weiß nicht
Denken Sie, dass Suchmaschinen, Online-Wörterbüchern, -Enzyklopädien und -Übersetzungs-tools in Ihrer Berufspraxis eine Rolle spielen werden?	□ ja	□ nein	□ weiß nicht

Vielen Dank für die Teilnahme an unserer Studie!

Appendix B: Pseudo-Randomization Scheme

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Texte	4							1	2	3	4	
2	Reading	2						mt	1mt_R	2mt_R	3mt_T	4mt_T	
3	Translating	2	Tools	GPT				ot	1ot_R	2ot_R	3ot_T	4ot_T	
4	Conditions	2	mit tools = alles außer ChatGPT	ohne tools = Nur ChatGPT					Presidents	Devon	Zündstoff	Leuchttürme	
5													
6									mit tools = alles außer ChatGPT				
7									ohne tools = Nur ChatGPT				
8													
9	P01	Task 1	1mt_R	2ot_R	3mt_T	4ot_T		P49	Task 17	2ot_R	1mt_R	3mt_T	4ot_T
10	P02	Task 2	4ot_T	1mt_R	2ot_R	3mt_T		P50	Task 18	1mt_R	4ot_T	2ot_R	3mt_T
11	P03	Task 3	3ot_T	4mt_T	1mt_R	2ot_R		P51	Task 19	4mt_T	3ot_T	1mt_R	2ot_R
12	P04	Task 4	2ot_R	3ot_T	4mt_T	1mt_R		P52	Task 20	3ot_T	2ot_R	4mt_T	1mt_R
13	P05	Task 5	2mt_R	1ot_R	3ot_T	4mt_T		P53	Task 21	1ot_R	2mt_R	3ot_T	4mt_T
14	P06	Task 6	4mt_T	2mt_R	1ot_R	3ot_T		P54	Task 22	2mt_R	4mt_T	1ot_R	3ot_T
15	P07	Task 7	3mt_T	4ot_T	2mt_R	1ot_R		P55	Task 23	4ot_T	3mt_T	2mt_R	1ot_R
16	P08	Task 8	1ot_R	3mt_T	4ot_T	2mt_R		P56	Task 24	3mt_T	1ot_R	4ot_T	2mt_R
17	P09	Task 9	1mt_R	2ot_R	4ot_T	3mt_T		P57	Task 25	4ot_T	2ot_R	3mt_T	1mt_R
18	P10	Task 10	4ot_T	1mt_R	3mt_T	2ot_R		P58	Task 26	3mt_T	1mt_R	2ot_R	4ot_T
19	P11	Task 11	3ot_T	4mt_T	2ot_R	1mt_R		P59	Task 27	2ot_R	4mt_T	1mt_R	3ot_T
20	P12	Task 12	2ot_R	3ot_T	1mt_R	4mt_T		P60	Task 28	1mt_R	3ot_T	4mt_T	2ot_R
21	P13	Task 13	2mt_R	1ot_R	4mt_T	3ot_T		P61	Task 29	4mt_T	1ot_R	3ot_T	2mt_R
22	P14	Task 14	4mt_T	2mt_R	3ot_T	1ot_R		P62	Task 30	3ot_T	2mt_R	1ot_R	4mt_T
23	P15	Task 15	3mt_T	4ot_T	1ot_R	2mt_R		P63	Task 31	1ot_R	4ot_T	2mt_R	3mt_T
24	P16	Task 16	1ot_R	3mt_T	2mt_R	4ot_T		P64	Task 32	2mt_R	3mt_T	4ot_T	1ot_R
25	P17	Task 17	2ot_R	1mt_R	3mt_T	4ot_T		P65	Task 1	1mt_R	2ot_R	3mt_T	4ot_T
26	P18	Task 18	1mt_R	4ot_T	2ot_R	3mt_T		P66	Task 2	4ot_T	1mt_R	2ot_R	3mt_T
27	P19	Task 19	4mt_T	3ot_T	1mt_R	2ot_R		P67	Task 3	3ot_T	4mt_T	1mt_R	2ot_R
28	P20	Task 20	3ot_T	2ot_R	4mt_T	1mt_R		P68	Task 4	2ot_R	3ot_T	4mt_T	1mt_R
29	P21	Task 21	1ot_R	2mt_R	3ot_T	4mt_T		P69	Task 5	2mt_R	1ot_R	3ot_T	4mt_T
30	P22	Task 22	2mt_R	4mt_T	1ot_R	3ot_T		P70	Task 6	4mt_T	2mt_R	1ot_R	3ot_T
31	P23	Task 23	4ot_T	3mt_T	2mt_R	1ot_R		P71	Task 7	3mt_T	4ot_T	2mt_R	1ot_R
32	P24	Task 24	3mt_T	1ot_R	4ot_T	2mt_R		P72	Task 8	1ot_R	3mt_T	4ot_T	2mt_R
33	P25	Task 25	4ot_T	2ot_R	3mt_T	1mt_R		P73	Task 9	1mt_R	2ot_R	4ot_T	3mt_T
34	P26	Task 26	3mt_T	1mt_R	2ot_R	4ot_T		P74	Task 10	4ot_T	1mt_R	3mt_T	2ot_R
35	P27	Task 27	2ot_R	4mt_T	1mt_R	3ot_T		P75	Task 11	3ot_T	4mt_T	2ot_R	1mt_R
36	P28	Task 28	1mt_R	3ot_T	4mt_T	2ot_R		P76	Task 12	2ot_R	3ot_T	1mt_R	4mt_T
37	P29	Task 29	4mt_T	1ot_R	3ot_T	2mt_R		P77	Task 13	2mt_R	1ot_R	4mt_T	3ot_T
38	P30	Task 30	3ot_T	2mt_R	1ot_R	4mt_T		P78	Task 14	4mt_T	2mt_R	3ot_T	1ot_R
39	P31	Task 31	1ot_R	4ot_T	2mt_R	3mt_T		P79	Task 15	3mt_T	4ot_T	1ot_R	2mt_R
40	P32	Task 32	2mt_R	3mt_T	4ot_T	1ot_R		P80	Task 16	1ot_R	3mt_T	2mt_R	4ot_T
41	P33	Task 1	1mt_R	2ot_R	3mt_T	4ot_T		P81	Task 17	2ot_R	1mt_R	3mt_T	4ot_T
42	P34	Task 2	4ot_T	1mt_R	2ot_R	3mt_T		P82	Task 18	1mt_R	4ot_T	2ot_R	3mt_T
43	P35	Task 3	3ot_T	4mt_T	1mt_R	2ot_R		P83	Task 19	4mt_T	3ot_T	1mt_R	2ot_R
44	P36	Task 4	2ot_R	3ot_T	4mt_T	1mt_R		P84	Task 20	3ot_T	2ot_R	4mt_T	1mt_R
45	P37	Task 5	2mt_R	1ot_R	3ot_T	4mt_T		P85	Task 21	1ot_R	2mt_R	3ot_T	4mt_T
46	P38	Task 6	4mt_T	2mt_R	1ot_R	3ot_T		P86	Task 22	2mt_R	4mt_T	1ot_R	3ot_T
47	P39	Task 7	3mt_T	4ot_T	2mt_R	1ot_R		P87	Task 23	4ot_T	3mt_T	2mt_R	1ot_R
48	P40	Task 8	1ot_R	3mt_T	4ot_T	2mt_R		P88	Task 24	3mt_T	1ot_R	4ot_T	2mt_R
49	P41	Task 9	1mt_R	2ot_R	4ot_T	3mt_T		P89	Task 25	4ot_T	2ot_R	3mt_T	1mt_R
50	P42	Task 10	4ot_T	1mt_R	3mt_T	2ot_R		P90	Task 26	3mt_T	1mt_R	2ot_R	4ot_T
51	P43	Task 11	3ot_T	4mt_T	2ot_R	1mt_R		P91	Task 27	2ot_R	4mt_T	1mt_R	3ot_T
52	P44	Task 12	2ot_R	3ot_T	1mt_R	4mt_T		P92	Task 28	1mt_R	3ot_T	4mt_T	2ot_R
53	P45	Task 13	2mt_R	1ot_R	4mt_T	3ot_T		P93	Task 29	4mt_T	1ot_R	3ot_T	2mt_R
54	P46	Task 14	4mt_T	2mt_R	3ot_T	1ot_R		P94	Task 30	3ot_T	2mt_R	1ot_R	4mt_T
55	P47	Task 15	3mt_T	4ot_T	1ot_R	2mt_R		P95	Task 31	1ot_R	4ot_T	2mt_R	3mt_T
56	P48	Task 16	1ot_R	3mt_T	2mt_R	4ot_T		P96	Task 32	2mt_R	3mt_T	4ot_T	1ot_R
57													

Appendix C: Source texts for translation

Translation Brief 1 – Other Tools Condition
Aufgabe: Übersetzen Sie diesen Text ins Englische. Der Text ist sehr schwer – versuchen Sie trotzdem, Ihr Bestes zu geben. Die Übersetzung soll in der Schülerzeitung veröffentlicht werden. Sie dürfen alle Tools nutzen, die Sie kennen **AUßER** Chatbots (z.B. ChatGPT). Erlaubt sind zum Beispiel Suchmaschinen (Google etc.), Online-Wörterbücher (Leo, Pons, Linguee etc.), Online-Enzyklopädien (Wikipedia etc.), Online-Übersetzungstools (Google Translate etc.). Es bleibt Ihnen überlassen, wie Sie mit diesen Tools interagieren, ob Sie Wörter, Fragen, Sätze oder ganze Textpassagen in die Tools eingeben. Sie dürfen selbst entscheiden, wie oft und wie lange Sie welche Tools nutzen. Sie haben maximal 20 Minuten Zeit. Bitte nutzen den von uns vorgegebenen Browser. Minimieren Sie bitte keine Fenster; springen Sie zwischen den Fenstern hin und her.

Translation Brief 2 – ChatGPT Condition
Aufgabe: Übersetzen Sie diesen Text ins Englische. Der Text ist sehr schwer – versuchen Sie trotzdem, Ihr Bestes zu geben. Die Übersetzung soll in der Schülerzeitung veröffentlicht werden. Sie dürfen **NUR** Chatbots (z.B. ChatGPT) verwenden. Es sind KEINE anderen Tools erlaubt. Es bleibt Ihnen überlassen, wie Sie mit dem Chatbot interagieren, ob Sie Wörter, Fragen, Sätze oder ganze Textpassagen in das Tool eingeben. Sie dürfen selbst entscheiden, wie oft und wie lange Sie den Chatbot nutzen. Sie haben maximal 20 Minuten Zeit. Bitte nutzen den von uns vorgegebenen Browser. Minimieren Sie bitte keine Fenster; springen Sie zwischen den Fenstern hin und her.

Text 1
Die moderne Zivilisation auf einen nachhaltigen Weg zu bringen, gleicht mehr und mehr dem Versuch, einen Deich zu halten, gegen den die Flut drückt. Hat man gerade noch mit bloßen Händen den einen Riss gestopft, tun sich daneben schon die nächsten auf. Der jüngste Fall: Pflanzen als Energiequelle der Zukunft. Vor zwei Jahren noch gepriesen, vergeht nun kaum ein Monat, in dem nicht Umwelt- und Entwicklungsorganisationen vor dramatischen Konsequenzen für Klima, Umwelt und Ernährungssicherheit warnen. (Goethe Institut 2017)

Text 2
Wer heute durch die Städte Europas fährt, der wird traurig. Vielleicht entsteht dieses Gefühl nicht in den inneren Zirkeln der Metropolen, deren Zentren immer noch begeistern können. Das Gefühl hoffnungsloser Traurigkeit entsteht in den unzähligen mittleren und kleineren Städten sowie in allen Vorstädten und Agglomerationen. Einst war jede Stadt ein eigener Kosmos mit eigener Ordnung. Einst formten die Bewohner einer jeden Stadt eine stolze Bürgerschaft – die ihr Selbstbewusstsein in städtischer Architektur zum Ausdruck brachte. (Goethe Institut 2011)

Please type in your translation here:

References

American Translators Association. Explanation of error categories, Version 2022. (https://www.atanet.org/certification/how-the-exam-is-graded/error-categories/) (Accessed 2025-09-10).

Bertoldi, Nicola & Cettolo, Mauro & Federico, Marcello. 2010. Statistical machine translation of texts with misspelled words. In *Human Language Technologies. The 2010 Annual Conference of the North American Chapter of the ACL*. (https://aclanthology.org/N10-1064.pdf) (Accessed 2024-12-19).

Bucher, Benedikt & Humpa, Michael. ChatGPT Anleitung. KI auf Deutsch nutzen. *Chip*, 14 July 2023. (https://www.chip.de/news/ChatGPT-Anleitung-KI-auf-Deutsch-nutzen_184795765.html) (Accessed 2025-09-10).

Cambridge University Press & Assessment. 2024. (https://www.cambridgeenglish.org/de/test-your-english/for-schools/) (Accessed 2024-12-09).

Carr, Paul. 2023. Generative AI. Friend or foe for the translation industry? *Forbes*, 11 Aug. 2023. (https://www.forbes.com/councils/forbestechcouncil/2023/08/11/generative-ai-friend-or-foe-for-the-translation-industry/) (Accessed 2025-09-10).

Carl, Michael. 2012. Translog-II. A program for recording user activity data for empirical reading and writing research. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*. Istanbul: European Language Resources Association (ELRA). 4108–12.

Carl, Michael & Bangalore, Srinivas & Schaeffer, Moritz. 2016. *New directions in empirical translation process research. Exploring the CRITT TPR-DB.* Cham: Springer. doi: 10.1007/978-3-319-20358-4.

Carl, Michael. 2021. *Explorations in empirical translation process research*. Cham: Springer. doi: 10.1007/978-3-030-69777-8.

Catford, John C. 1965. *A linguistic theory of translation. An essay on Applied Linguistics*. London: Oxford University Press.

Chesterman, Andrew. 2011. Reflections on the literal translation hypothesis. In Alvstad, Cecilia & Hild, Adelina & Tiselius, Elisabet (eds.), *Methods and strategies of process research: Integrative approaches in translation studies*. Amsterdam: John Benjamins. doi: 10.1075/btl.94.05che.

Chodkiewicz, Marta. 2015. Undergraduate students' use of external sources in revising and justifying their translation decisions based on instructor feedback. *Lublin Studies in Modern Languages and Literature* 39(2): 124–41. (https://journals.umcs.pl/lsmll/article/viewFile/2713/1943) (Accessed 2024-11-26).

Daems, Joke & Carl, Michael & Vandepitte, Sonia & Hartsuiker, Robert & Macken, Lieve. 2016. The effectiveness of consulting external resources during translation and post-editing of general text types. In Carl, Michael & Bangalore, Srinivas & Schaeffer, Moritz (eds.), *New directions in empirical translation process research. Exploring the CRITT TPR-DB*. Cham: Springer. 111–33. doi: 10.1007/978-3-319-20358-4.

DIN ISO 18587. 2017. Translation services – post-editing of machine translation output – requirements. (https://www.iso.org/standard/62970.html) (Accessed 2025-09-10).

DIN ISO 17100. 2015. Übersetzungsdienstleistungen. Anforderungen an Übersetzungsdienstleistungen. (https://www.din.de/de/mitwirken/normenausschuesse/nat/veroeffentlichungen/wdc-beuth:din21:246682286) (Accessed 2024-12-20).

EMT Board and Competence Task-Force. European Master's in Translation – Competence Framework, Oct. 2022. (https://commission.europa.eu/system/files/2022-11/emt_competence_fwk_2022_en.pdf) (Accessed 2025-09-10).

Forthem. n.d. (https://www.forthem-alliance.eu) (Accessed 2024-12-09).

Giray, Louie. 2023. Prompt engineering with ChatGPT. A guide for academic writers. *Annals of Biomedical Engineering* 51: 2629–33. doi: 10.1007/s10439-023-03272-4.

Göpferich, Susanne. 2009. Towards a model of translation competence and its acquisition. The longitudinal study TransComp. In Göpferich, Susanne & Jakobsen, Arndt Lykke & Mees, Inger (eds.), *Behind the mind. Methods, models and results in translation process research*. Copenhagen: Samfundslitteratur.

Goethe Institut. 2011. C2 Modellsatz. (https://www.goethe.de/pro/relaunch/prf/materialien/C2/c2_modellsatz.pdf) (Accessed 2025-09-10).

Goethe Institut. 2017. C2 Übungssatz. (https://www.goethe.de/pro/relaunch/prf/materialien/C2/c2-uebungssatz.pdf) (Accessed 2025-09-10).

Huschens, Martin & Briesch, Martin & Sobania, Dominik & Rothlauf, Franz. 2023. Do you trust ChatGPT? – Perceived credibility of human and AI-generated content. ArXiv. doi: 10.48550/arXiv.2309.02524.

Jakobsen, Arnt Lykke. 2003. Effects of think aloud on translation speed, revision and segmentation. In Alves, Fabio (ed.), *Triangulating translation. Perspectives in process-oriented research*. Amsterdam: Benjamins. 69–95. doi: 10.1075/btl.45.08jak.

Jakobsen, Arnt Lykke & Schou, Lasse. 1999. Translog documentation. In Hansen, Gyde (ed.), *Probing the process in translation methods and results*. Copenhagen: Samfundslitteratur. 1–36.

Khanmigo. n.d. (https://www.khanmigo.ai/) (Accessed 2024-11-28).

Krüger, Ralph & Hackenbuchner, Janiça. 2024. A competence matrix for machine translation-oriented data literacy teaching. *Target* 36(2): 245–275. doi: 10.1075/target.22127.kru.

Kussmaul, Paul. 2015. *Verstehen und Übersetzen. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr Francke Attempto.

Leijten, Mariëlle & Van Waes, Luuk. Keystroke logging in writing research. Using Inputlog to analyze and visualize writing processes, 29 June 2013. *Written Communication* 30(3): 358–392. (https://journals.sagepub.com/doi/10.1177/0741088313491692) (Accessed 2025-09-10). doi: 10.1177/0741088313491692.

Lionbridge. 2025. (https://www.lionbridge.com/generative-ai/) (Accessed 2024-11-28).

Malmkjaer, Kirsten. 2005. Norms and nature in translation studies. *Synaps* 16: 13–19.

MQM Council. 2023. (https://themqm.org/error-types-2/typology/) (Accessed 2025-09-10).

Nitzke, Jean & Hansen-Schirra, Silvia. 2021. *A short guide to post-editing*. Berlin: Language Science Press. doi: 10.5281/zenodo.5646896.

Oster, Katharina. 2019. Lexical activation and inhibition of cognates among translation students. Mainz: Johannes Gutenberg-Universität. (Doctoral dissertation.) doi: 10.25358/openscience-3748.

PACTE. 2000. Acquiring translation competence. Hypotheses and methodological problems of a research project. In Beeby, Allison & Ensinger, Doris & Presas, Marisa (eds.), *Investigating translation. Selected papers from the 4th International Congress on Translation, Barcelona, 1998*. Amsterdam: John Benjamins. 99–106. doi: 10.1075/btl.32.13pac.

PACTE. 2003. Building a translation competence model. In Alves, Fabio (ed.), *Triangulating translation. Perspectives in process-oriented research*. Amsterdam: John Benjamins. 43–66. doi: 10.1075/btl.45.06pac.

Plitt, Mirko & Masselot, François. 2010. A productivity test of statistical machine translation post-editing in a typical localization context. *Prague Bulletin of Mathematical Linguistics* 93: 7–16. doi: 10.2478/v10108-010-0010-x.

Risku, Hanna. 1998. *Translatorische Kompetenz. Kognitive Grundlagen des Übersetzens als Expertentätigkeit*. Tübingen: Stauffenburg.

Robert, Isabelle & Remael, Aline & Ureel, Jim JJ. 2017. Towards a model of translation revision competence. *The Interpreter and Translator Trainer* 11(1): 1–19. doi: 10.1080/1750399X.2016.1198183.

Sawalha, Ghadeer & Taj, Imran & Shoufan, Abdulhadi. 2024. Analyzing student prompts and their effect on ChatGPT's performance. *Cogent Education* 11(1): 2397200. doi: 10.1080/2331186X.2024.2397200.

Sengar, Sandeep Singh & Hasan, Affan Bin & Kumar, Sanjay & Carroll, Fiona. 2024. Generative artificial intelligence. A systematic review and applications. *Multimedia Tools and Applications* 83: 23661–700. doi: 10.1007/s11042-024-20016-1.

Schaeffer, Moritz & Dragsted, Barbara & Hvelplund, Kristian Tangsgaard & Balling, Laura Winther & Carl, Michael. 2016. Word translation entropy. Evidence of early target language activation during reading for translation. In Carl, Michael & Bangalore, Srinivas & Schaeffer, Moritz (eds.), *New directions in empirical translation process research. Exploring the CRITT TPR-DB*. Cham: Springer International Publishing. 183–210. doi: 10.1007/978-3-319-20358-4_9.

Vardaro, Jennifer & Schaeffer, Moritz & Hansen-Schirra, Silvia. 2019. Comparing the quality of neural machine translation and professional post-editing. In *Eleventh International Conference of Quality of Multimedia Experience (QoMEX), 2019*. Berlin, Germany. 1–3. (https://ieeexplore.ieee.org/abstract/document/8743218) (Accessed 2025-09-10). doi: 10.1109/QoMEX.2019.8743218.

Vermeer, Hans J. 2006. *Versuch einer Intertheorie der Translation*. Berlin: Frank & Timme.

Xu, Ziwei & Jain, Sanjay & Kankanhalli, Mohan. 2024. Hallucination is inevitable. An innate limitation of large language models. ArXiv. (https://arxiv.org/abs/2401.11817) (Accessed 2024-11-29). doi: 10.48550/arXiv.2401.11817.

Yamada, Masaru. 2015. Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings. *Machine Translation* 29: 49–67. doi: 10.1007/s10590-014-9167-7.

Athanasios Breskas
An der Hochschule 2
76726 Germersheim
Germany
e-mail: atbreska@uni-mainz.de

Silvia Hansen-Schirra
An der Hochschule 2
76726 Germersheim
Germany
e-mail: hansenss@uni-mainz.de

Dimitrios Kapnas
An der Hochschule 2
76726 Germersheim
Germany
e-mail: dikapnas@uni-mainz.de