

CERSTIN MAHLOW

Large Language Models and Artificial Intelligence as Tools for Teaching and Learning Writing

Abstract

Rapid advances in artificial intelligence (AI), specifically large language models (LLMs), have recently generated significant debate. This article explores the impact of these developments on learning to write in a first or second/foreign language, specifically German. We examine the technology behind AI-based tools and the natural language processing (NLP) tasks for which they were originally designed. This will help us identify the possibilities and limitations of their use in the context of language learning. We then examine how this technology can be used effectively in language teaching and learning. In conclusion: the availability of these tools will allow language teaching to focus on the non-mechanical aspects of writing; automatically generated personalized teaching and learning materials will make room for and support human-human interaction.

Keywords: artificial intelligence, large language models, natural language processing, written language acquisition

Entwicklungen im Bereich große Sprachmodelle (Large Language Models – LLM) und künstliche Intelligenz (KI) sind ein wichtiges und zugleich umstrittenes Thema im aktuellen Diskurs. Dieser Artikel untersucht den Einfluss dieser Entwicklungen auf das Sprachenlernen. Er diskutiert die Technologie, die KI-basierten Werkzeugen zugrunde liegt, sowie Aufgaben aus dem Bereich Sprachtechnologie, für die sie ursprünglich entwickelt wurden. Dies erlaubt es, Chancen und Grenzen für die Verwendung im Kontext des Sprachenlernens zu identifizieren. Anschließend wird gezeigt, wie KI-basierte Sprachtechnologie effizient und effektiv für den Schriftspracherwerb eingesetzt werden kann. Die Verfügbarkeit solcher Werkzeuge ermöglicht die Automatisierung repetitiver und standardisierter Prozesse und erlaubt die Konzentration der Lehrpersonen auf menschliche Interaktion.

Schlüsselwörter: Künstliche Intelligenz, große Sprachmodelle, automatische Sprachverarbeitung, Schriftspracherwerb

1 Introduction

Language technology and Natural Language Processing (NLP) have advanced rapidly and now seem to have achieved a degree of maturity to be used in most daily communication situations. Machine translation (MT), as one example, has been proven to not replace human-led translation; but it has changed established processes (Bowker/Ciro 2019) and raised questions about the need to learn a foreign language, and how to incorporate MT into the Common European Framework of Reference for Languages (CEFR) (Delorme Benites/Lehr 2021). Human-written translation still yields the most satisfactory results and writing in one's own language is an important skill. Writing is also a fundamental part of learning and this paper will focus on German, its use as a first and second/foreign language. Language technology in writing tools can support writers when struggling with certain aspects like spelling and vocabulary, or inputting text. It thus is also an important instrument in inclusive language acquisition.

Applications using language technology are often called Artificial Intelligence (AI) tools or AI-based tools. The sudden availability of such applications to the general public in late November 2022, with ChatGPT from OpenAI as the most prominent example, seems to indicate that indeed automatic production of texts in human-like quality is now possible. Millions of users tried out ChatGPT in December 2022, the underlying resources got integrated into existing or new applications. Social and traditional media were flooded with astonishing examples of texts produced by ChatGPT, but slowly skeptical voices appeared, pointing to flaws or even errors. Previous to ChatGPT, Meta had released Galactica, which swiftly had to be deactivated as users discovered and showed serious issues that had not been properly communicated or addressed by Meta. ChatGPT in contrast appeared to "have learnt" from this and made clear that not everything in the texts produced should be taken too seriously; the application also refuses to produce clearly offensive or malicious texts.

Technology of any kind is created to serve certain purposes. Language technology in general is designed and implemented to solve specific language-related tasks. ChatGPT and similar applications produce texts according to prompts in an interactive way while the user is in the lead. Purposeful uses of automated text production in dialog-like situations involving chatbots and integrated as functionality in word processors of any kind are still to be discovered. First studies on *prompt engineering* (e.g., Chang et al. 2023; Zamfirescu-Pereira et al. 2023) are published, which aim to determine how to best communicate specific intents to an LLM-based AI system. While peer-reviewed studies addressing AI in education focus on classifying writing (e.g., Wulff et al. 2022) or reducing workload in answer grading (e.g., Weegar/Idestam-Almquist 2023) involve writing, only little has been published considering AI-based support for learning how to write or writing to learn.

What is shared among teachers privately or at larger scale via newsletters are syllabi and best practices.¹ Systematic studies are yet to be conducted and published.

In this paper I focus on how to use current language technology for language learning with an emphasis on writing – i.e., learning how to write in German. Writing is both an every-day activity and a significant competence learners acquire to actively and successfully participate in communicative situations. Writing and learning how to write in a first as well as in a foreign or second language is thus an important skill. Writing competencies are specifically addressed as written production in the CEFR (Council of Europe 2020, 2001). Producing written texts involves cognitive and linguistic competencies. Cognitive competencies address knowledge about genres and organizing relevant information or facts to achieve a certain communicative goal. Linguistic competencies on the other hand cover general morphological and syntactical knowledge, but also refer to a repertoire of vocabulary, phrasal structures, and discourse devices language learners acquire. Both aspects are interconnected: What is usually covered by “style” might be interpreted and described as appropriate genre-specific linguistic features.

Learning how to write thus should address all aspects and offer opportunities to explore and train cognitive, pragmatic, as well as linguistic competencies. Written texts produced by language learners are often used as evidence of learners’ performance and serve assessments purposes. They supposedly show whether or not learners are able to correctly recognize specific communicative situations and apply appropriate linguistic features to master this scenario and thus prove competence at a certain level.

If we agree that similar to earlier experiences in the field of translation (Bowker/Ciro 2019), language technology will change established writing processes but never replace humans completely, learning how to write will remain to be part of curricula. Available applications as well as potential new ones – given the general capabilities of the underlying resources – will become an integral part of language learning. Focusing on opportunities that come with almost disruptive changes to learning and teaching will allow to focus on specific pedagogical aspects (again) and support learners in new ways.

The rest of this paper is structured as follows: First, I look at the underlying technology for AI-based applications like ChatGPT and show what they are designed for in section 2. This will help to understand general chances and limits. Section 3 looks at opportunities for automation in learning and teaching writing German for which language technology in fact can be used efficiently and effectively.

2 LLMs as core of language-processing and language-producing AI applications

Specific applications of language technology are the focus and the center of enthusiastic optimism as well as of fierce debates about dangers of all kinds. Single applications are used

¹ For example the newsletter by Lance Cumming: <https://world.hey.com/lance.cummings> (last access 2023-05-20).

as proxy in discourses – at the time of writing, in late spring 2023, “ChatGPT” is used when actually discussing AI-based language technology at large – where we rather should look at the underlying resources and technology to be able to evaluate and assess general possibilities and limits. This certainly requires good understanding of the nature of these resources, general scenarios they could be used in, and assumptions of what would be needed to make it useful for an intended specific task (e.g., percentage of correct results or performance compared to humans).

The core of today’s AI-based applications in language learning and teaching are Large Language Models (LLMs). If we understand how language models are constructed and what they can be used for in general with reasonable confidence, we will be able to determine the role of existing and future applications based on these resources. A crucial aspect when talking about language technology are general approaches and tasks in Natural Language Processing (NLP), which clearly profit from improving LLMs and which can be used as resources and tools for activities in learning and teaching. The currently most impressive use of LLMs is the production of texts in reaction to user prompts. From a technical point of view, ChatGPT is just a chatbot using a specific LLM with a very convincing and easy to use interface.

2.1 Large Language Models

Large language models (LLMs) are resources for language technology applications that use deep neural networks with a huge number of features and parameters trained on massive amounts of text data. The potential of LLMs became evident with Google’s BERT (Devlin et al. 2019) and PaLM (Chowdhery et al. 2022) – 340 million and 540 billion parameters, respectively. OpenAI has published several versions of their Generative Pretrained Transformer (GPT) models. GPT-3, introduced in 2020, has 175 billion parameters (Brown et al. 2020). For the most recent version, GPT-4, available since Spring 2023, OpenAI has not released technical information (OpenAI 2023). GPT-3.5 has been used for the chatbot ChatGPT with additional fine-tuning and the capability to store and use previous utterances both from the bot and the user. Users can ask the chatbot for specific output by drafting a verbose request in their language, they can reconfigure their original intent in a dialog with the system.

For training – i.e., for creating the models in the first place –, these models use *unlabeled* text. The texts have not been preprocessed or annotated for any information: they use so-called unsupervised learning. Which also means that there is no significant control or knowledge of the sources and original intentions of these texts: one would perhaps like to exclude machine-produced texts (either generated or automatically translated texts), to verify the language of texts and probably treat multilingual texts separately, use weighting to address and level out biases, etc. Modeling only takes into account what in NLP are usually called left and right contexts of words, i.e., what is before and after each single wordform.

In a broad linguistic abstraction: language models entail only *co-text* of language units but have no access to *context*. The power of a model is influenced both by the amount of training data and by the number of parameters the model uses to first classify co-text and later retrieve possible or plausible co-text by using a new architecture called transformers (see Tay et al. 2022). *Large* language models are simply language models trained on large amounts of texts, using large amounts of parameters, and large sizes of co-text taken into account – GPT-4 has a co-text window of 32 thousand words.

Language models and their use could be understood as an actual implementation of Firth’s well-known statement: “You shall know a word by the company it keeps” (Firth 1957, 11), which influenced lexical semantics and was used to allow for drawing conclusions not only on the meaning of words but also on the concepts behind words or sequences of words. However, words and their co-texts do not allow to deduce all facts that readers as well as writers are aware of and use. Obvious examples include the need for providing additional context when reading and processing historical texts (Piotrowski 2012) as well as the still hard task of unambiguously determining specific persons and places which needs additional resources in named entity recognition (Wang et al. 2021).

However, LLMs – and language models in general – are used in various specific tasks in Natural Language Processing (NLP) and are one reason for significant performance gains in these tasks.

2.2 Automated Text Production vs. Natural Language Generation

The output of applications using LLMs is often called “generated text,” which points towards the field of Natural Language Generation (NLG). NLG is the production of “understandable texts in [...] human languages from some underlying non-linguistic representation of information” (Reiter/Dale 2000) and has been an active research field for several decades. In analogy to text-to-speech, it has also been named *data-to-text* (Schneider et al. 2022).

NLG systems rely heavily on data, from measure points and facts to entire databases. This data provides the context for the text being generated, which is provided in a structured form to the generator. Further, the intended audience, genre, length, and other features, are made explicit and used as parameters. Dale (2020) and Schneider et al. (2022) provide an overview of commercial NLG. The latter authors emphasize that all necessary features for generation must be made explicit. They emphasize that

data-to-text systems in real-world applications still require such a share of human configuration and control and the creative contribution share of the software [...] is still so limited that it would not be adequate to claim creative autonomy of the software in the process. (Schneider et al. 2022, 1)

In contrast, LLM-based automated text production is an advanced version of predictive texting found on mobile phones (Ganslandt et al. 2009). Rather than relying on facts and information to determine the next words and sentences, it uses already existing words and sentences – i.e., the existing co-text – as a basis to statistically determine the next

word(s). The larger the co-text window, the better the text produced. The produced text is grammatically correct, cohesive, and consistent, but is only linguistically acceptable and believable, disregarding its content.

All information that seems to be included in the text is currently merely due to frequent cooccurrences of words, not based on facts and thus not trustworthy. No references to any context can be made, no conclusions on underlying knowledge, understanding, or intention of an assumed author can be drawn. Depending on the genre the LLM is asked to produce, a text may contain what looks like facts – e.g., dates, places, or even correctly formatted bibliographical references. But also these are only *plausible*, they are not true; they could, but most often do not exist. The LLM just produces sequences of words that follow general patterns for bibliographical entries consisting of strings that are names, followed by an arbitrary four-digit-number, a string looking like a title, followed by a string mimicking a publishing house, etc. While references produced by a LLM are thus most probably non-existing, one can very confidently use them to check whether the *format* of a list of bibliography entries adheres to a certain citation style or is compatible with a list of example entries.

Language technology can only provide automated writing assistance, not creative autonomy. There is no *interaction* between a human writer and a text-producing machine, only very sophisticated *interactivity*. Language technology is used for writing support in settings of automated text production as defined by Mahlow/Dale (2014). Dale/Viethen (2021) provide an overview on writing assistance based on state-of-the-art NLP approaches and resources – i.e., LLMs. These tools, while aiding beyond (copy-)editing, are mainly used for specific genres and tasks, such as blogs, poetry, expanding, rewriting, and shortening texts. In 2021, they started to be integrated into experimental editors (e.g., Dang et al. 2022; Yuan et al. 2022), but did not yet have widespread use, and had not been used for writing instruction.

LLM-based systems are not creative: they do not invent anything as they have no agency. They just react to arbitrary input – be it a prompt from a human or previously produced text. All creativity is only in the mind of human readers interpreting these texts, while ignoring authorship and not being aware of the circumstances – i.e., the process – of their production, as Sejnowski (2023) shows.

The more powerful LLMs get, the more users forget that texts automatically produced by any application using these resources are just plausible extensions of existing co-text – a variant of the ELIZA effect (Hofstadter 1996, 155-169). These co-texts are either existing (parts of) texts, which are then expanded, or prompts the system seems to “react” to: “GPT-3 is still very capable of generating nonsense, but on the whole it’s more plausible nonsense; and with appropriate fine tuning and prompting, the texts it generates can be eerily convincing.” (Dale/Viethen 2021, 516)

3 Automation in language learning and teaching

Section 2 covered some general characteristics of LLMs. This section focuses on language technology that specifically addresses the educational needs of language learners and teachers as well as language technology originally intended to solve certain NLP tasks, but whose affordances allow their use in language learning.

In general, major technological developments, as can be described in waves of “industrial revolutions,” enable and support the consistent, almost automatic execution of tasks that previously required significant human effort. Automation of any kind frees up capacity that humans can use for other activities.

In learning to write, two such examples are mastering handwriting and using a keyboard, i.e., some of the “mechanics” of writing. In the beginning, learners have to do everything very carefully and consciously (Dowling 1994). Once these processes have been internalized, writers are usually no longer aware of them (Fayol 1999; Kellogg 2008). Consequently, they have more cognitive capacity (Piolat et al. 2004) available to focus, for example, on spelling – another skill that can be mastered almost automatically (Fayol 1999) – coherence, and in particular the development of creative ideas (Torrance/Galbraith 1999).

The following section looks at specific NLP tasks that already are or can be used in language learning settings with a focus on writing in a first or second/foreign language.

3.1 Natural Language Processing Tasks for Language Learning and Teaching

The field of Computer-Assisted Language Learning (CALL) has progressed a lot since Levy (1997) established the term, which later changed to Technology-Enhanced Language Learning (TELL). Zhang/Zou (2022) provide a comprehensive overview on the current state of the art focusing on five major types of technology for second and foreign language learning: mobile-assisted, multimedia, socialized, speech-to-text and text-to-speech assistance, and gamification. Here we take a different perspective and show which NLP tasks can be incorporated in TELL, focusing on generally available technology for non-educational purposes whose affordances allow for use in educational applications and settings (Mahlow/Hediger 2021).

Answer extraction (e.g., Schwitter et al. 2000; Seonwoo et al. 2020) and question answering (e.g., Adlakha et al. 2022; Kwiatkowski et al. 2019) are used for the construction of reading comprehension tests. Extracted keywords (e.g., Hulth 2003) and concepts can be provided as seeds for higher level systems to look up suitable references or to point readers to more elaborate explanations of these concepts considering the individual learner’s level of competence (Chinkina et al. 2016; Meurers et al. 2010).

Summarization of texts (e.g., Jia et al. 2020; Wang et al. 2020) shortens texts by keeping the original message intact. This can be used by learners to make their writing fit the formal constraints of a writing task, they will have to check for consistency and coherence, though. Summarization can be used by teachers to provide learners with an abbreviated version of a longer text from a newspaper or a lengthy Wikipedia entry.

Checkers for spelling, grammar, and style have been around since the 1970s (Heidorn et al. 1982; Macdonald et al. 1982) and are now an every-day tool teachers recommend to use. Originally designed to support expert writers (Fontenelle 2005; Heidorn 2000), they have been adapted and advanced to address learners' needs (e.g., Gamon 2010; Tschichold et al. 1997); a development which is still ongoing (e.g., Sjöblom/Creutz/Vahtola 2021; Yuan/Bryant 2021). As for spelling and punctuation, it could be argued that writers do no longer need to practice these skills, since today's checking programs are capable of correcting almost all errors, and the use of dictation would prevent spelling errors from the start. Although dictation software is becoming more and more tolerant of background noise, accents, pronunciation, stutter, pauses and variation in intonation, they still work best for L1 speakers of standard varieties in quiet situations. Thus, the use of dictation in inclusive written language acquisition is possible but must be carefully planned and accompanied. Additionally, the output of those applications is not completely trustworthy, the writer has to be competent to check them and thus has to learn how to properly spell and place punctuation.

Speech-to-text (automatic transformation of spoken utterances into written texts) and text-to-speech (automatic rendering of written texts into audio) is used for dictation to help writers struggling with spelling or writer's block, and for having written texts read aloud to the writer to help them detect issues for revising. Dictation would also theoretically make it possible to avoid learning how to enter text altogether – either with a pen or a keyboard. This can be understood as the delegation of certain tasks to a trusted entity that is able to perform these tasks with high quality, so that the user can focus on other aspects. In the cases of dictation, text expansion (typing only a few letters and have the editor expand to the complete word), and predictive texting (suggestion of possible next words the writer can choose from), the writer – i. e. the learner – has the leading role: They decide *what* to write, technology is just helping to reduce some effort or is helping to overcome obstacles like dyslexia and permanent or temporary physical inability to type or to use a pen. If the focus of a specific exercise or task is not on spelling and grammar but on coherence and the mastering of certain aspects of genre and register, technology supporting inputting text do not interfere with the main goal but actually free mental capacity to focus on the task at hand.

Machine translation (MT) is applying deep neural learning models for some years now, making it possible to get good results even for low-resource languages. There are valid arguments that MT literacy should be included as competence into the CEFR (Delorme Benites/Lehr 2021), making it clear that MT is part of language learning. As Lee (2020) shows, using MT helps students to improve their lexico-grammatical competencies as well as revision strategies for writing texts.

Sentence similarity measures how similar sentences are with respect to syntactical and morphosyntactical structures (Das/Smith 2009), or to vocabulary and semantic (Fernando/Stevenson 2009). It can thus be used for formative assessment situations and to check how similar texts produced from learners are to provided examples.

One crucial aspect is the quality or correctness of results achieved by respective applications. Teachers will have to evaluate whether current state-of-the-art results are *good enough* to be used in real-world settings; it is nearly impossible to determine this without considering concrete context and tasks as shown by Mahlow/Piotrowski (2009) for automatic lemmatization and morphosyntactic analysis. The combination of resources and tasks need thorough evaluation to determine the overall quality and decide on suitability for language learning purposes.

Considering these general possibilities of current digital technology, we can identify some areas in the learning and teaching of writing that require consistent execution and that, on the other hand, would benefit from standardization and personalization while being agnostic to the content. Such areas are the provision of feedback, providing writing and learning companions for students, and the creation of assignments and prompts. What is particularly intriguing is the fact that LLMs can be used for all of those tasks without further intervention or programming. Additionally, applications having access to LLMs can execute all of those tasks simultaneously within seconds.

3.2 Feedback

Providing immediate and consistent feedback supports learning. With respect to writing, automatic essay scoring (AES) (Shermis/Burstein 2002) generated heated debate, while automatic writing evaluation (AWE), with its focus on providing feedback to improve learners' writing skills, was much more positively received (Fu et al. 2022).

As mentioned earlier, LLMs can recognize and produce plausible writing, which is error-free in terms of spelling and grammar, consistent, and cohesive. It is also possible to access text output that is specific to a particular language level and for a particular purpose or genre. This, in turn, means that LLMs can be confidently used to immediately and consistently evaluate learner input in terms of linguistic features and provide feedback on how texts can be improved to better meet these criteria. For learners, this represents a more acceptable or comfortable experience: Feedback will be consistent but adaptive to the learner – which has been shown to be effective (Leontjev 2014) – and the machine is not emotionally involved as teachers might be (or as learners might assume their teachers would be).

These scenarios and their corresponding applications already exist in Intelligent Tutoring Systems (ITS) (e.g., Wang et al. 2023). Their performance and capabilities will only benefit from the use of increasingly better underlying language technology, in our case LLMs.

As mentioned above, one task in NLP is summarization. Applications that use LLMs are able to provide a summary of a paragraph, section, or even an entire text within milliseconds as an additional feature (e.g., Dang et al. 2022). It can be used as part of a feedback system: the writer is provided with a summary of the text they themselves produced. They then have to decide whether or not this summary fits their original

communicative intent and identify passages in the text to be revised. Discrepancies can be due to both a lack of linguistic competence (the writer does not know how to express something appropriately in a particular language) and the fact that the text does not contain the intended information and needs to be expanded. Even when the quality of the summarization is not good enough – which depends on the language as well as the topic and genre – to be submitted without revisions, for immediate, somewhat “objective” feedback to the writer, it proves useful and gives explicit agency to the writer who decides how to react.

These scenarios are the more empowering for learners, the more consistent and replicable suggestions provided by the machine are. Teachers can focus on specific aspects, make connections to previous writing tasks, etc., while being relieved of the time-consuming and repetitive task of finding plausible alternatives to show variations and patterns in writing.

3.3 Learning and Writing Companion

Machine translation can be used for round-trip-translations (Somers 2005), a suitable approach to automatically produce a variant of a sentence or paragraph. It allows writers to get suggestions on alternative formulations of sentences and paragraphs of their texts: a draft written in German can be automatically translated into English or French, and without having the writer look at this translation, it is immediately translated back into German. Due to the properties of LLMs, the resulting text often uses more idiomatic syntax, some words and phrases may be replaced by synonyms or plausible alternatives respecting the co-text. This way, the writer receives a revised version as a proposal for alternative phrasing for their original text which teachers can also use as starting point to make learners aware of and discuss vocabulary choice, grammar, style, etc.

The NLP resources involved can be combined and integrated into editors, but some programming skills are needed.² Meanwhile, various add-ons and plug-ins are available to provide rewriting functionality for sentences or paragraphs. For example, the functionality “Umformulieren”³ as additional feature in “LanguageTool” that can be used in the LanguageTool editor as well as in all other writing applications LanguageTool can be integrated into. German is one of the main languages LanguageTool supports, also acknowledging main variants for German in Austria, Germany, and Switzerland. Another example is the online editor Lex,⁴ developed to offer a “Google docs-style editing experience” with limited formatting support but access to several OpenAI GPT variants. Here, the quality of rewriting or shortening sentences and paragraphs depends on the language, due to the assumed training data it is usually better for English than for German and other languages.

2 See for example TXL by Titus van der Malsburg <https://github.com/tmalsburg/txl.el> (last access 2023-05-21) to use DeepL directly from Emacs.

3 <https://languagetool.org/insights/de/beitrag/umformulieren/> (last access 2023-05-21).

4 <https://lex.page> (last access 2023-05-21).

Sentence similarity measures can be used to detect topical relevance in learner essays (Rei/Cummins 2016) as starting point for feedback that can be given either fully automatically – focusing on whether or not some expected topics are already present in the learner’s writing – or serves as resource for teachers or writing counselors.

Sentence similarity is also used to create phrasebooks from authentic learner input, to produce textbook examples to be presented to writers as acceptable instances of rhetoric patterns, and as ad hoc created variants of sentences or phrases the writer produced themselves. Note that here authentic material – either from own personal writing, from writing of peers, or from reference corpora – is extracted and presented to learners and teachers. In non-interactive settings, this material can be produced by teachers running queries in corpus linguistics tools. However, one has to be familiar with a specific corpus query language and additionally one has to “translate” all vague information from the prompt into various specific formal abstractions to be able to construct a set of queries.

As an example situation: Learners of German as a second/foreign language are asked to practice argumentative writing. These texts should contain several aspects of or views on a situation or concept; one specific prompt could be “Schreiben Sie zu Vor- und Nachteilen von E-Autos verglichen mit Diesel-Autos.”⁵ The focus is not on the content but on mastering discourse devices to express different views, in this case pros and cons. What learners need are phrases to start a section or paragraph on cons after having written about the pros. Teachers might provide a handful of useful phrases either from textbooks or extracted from corpora of argumentative essays. However, to extract relevant phrases, one would have to create appropriate corpus queries by making explicit what constitutes transition sentences expressing alternatives on a syntactic level and maybe considering specific words – which is not at all trivial.

LLMs on the other hand can output similar material in terms of phrasebooks or example sentences within seconds. For our example, the prompt: “Ich schreibe einen argumentativen Text über Vor- und Nachteile von E-Autos verglichen mit Diesel-Autos. Generiere mir 10 Möglichkeiten, den Abschnitt zu Nachteilen einzuleiten” has ChatGPT output 10 useful sentences like the following:⁶

Obwohl E-Autos viele Vorteile bieten, gibt es auch einige Nachteile, die bei einem Vergleich mit Diesel-Autos berücksichtigt werden sollten.

Trotz ihrer umweltfreundlichen Natur haben E-Autos einige Nachteile gegenüber Diesel-Fahrzeugen, die nicht übersehen werden dürfen.

Neben den offensichtlichen Vorteilen weisen E-Autos auch einige Nachteile auf, die bei der Entscheidung zwischen ihnen und Diesel-Autos berücksichtigt werden müssen.

5 Engl: “Write about pros and cons of electric cars compared to diesel cars.”

6 Session with ChatGPT on 2023-05-2. Translation of the prompt: “I write an argumentative text on pros and cons of electric cars compared to Diesel cars. Generate 10 possibilities to start the paragraph on cons.”

Es ist wichtig, die potenziellen Nachteile von E-Autos zu betrachten, um eine fundierte Wahl zwischen ihnen und Diesel-Fahrzeugen zu treffen.

Trotz ihres steigenden Beliebtheitsgrads sind E-Autos nicht frei von Nachteilen, die es zu bedenken gilt, insbesondere im Vergleich zu Diesel-Autos.

Während E-Autos zweifellos ihre Vorzüge haben, sollten wir auch die möglichen Nachteile in Betracht ziehen, die sie im Vergleich zu Diesel-Fahrzeugen haben.

Es gibt bestimmte Aspekte von E-Autos, die als Nachteile betrachtet werden können, wenn man sie mit Diesel-Fahrzeugen vergleicht.

Es ist wichtig, sich der möglichen Nachteile von E-Autos bewusst zu sein, um eine umfassende Entscheidung darüber treffen zu können, ob sie Diesel-Fahrzeugen überlegen sind.

E-Autos haben zweifellos ihre Vorteile, aber es gibt auch einige Nachteile, die es zu beachten gilt, insbesondere im Vergleich zu Diesel-Autos.

Trotz ihrer Umweltfreundlichkeit weisen E-Autos einige Nachteile auf, die bei der Abwägung zwischen ihnen und Diesel-Fahrzeugen berücksichtigt werden sollten.

These sentences produced by an LLM are *synthetic* sentences, they cannot be found in actual texts produced by writers. However, they are very similar to authentic sentences and might thus be used as additional learning material constituting examples how one could formulate a sentence or phrase. Using the “regenerate” command, one can get even more also suitable sentences again and again. This automatic production can now be used in two settings: (1) allowing the writer to choose one to include in their own writing as they would from a traditional phrasebook and (2) looking at a large number of these sentences to discover the generic rhetorical features and patterns, syntactical patterns, and a repertoire on vocabulary to use. The second setting provides learning opportunities on several levels: syntactic and rhetoric patterns, vocabulary, register, etc.

3.4 Create tasks and check solutions

The use of actual language use from corpora for teaching and learning languages has been advocated for for a long time now. One obstacle for language teachers has always been operating corpus linguistics tools and having to learn specific corpus query languages. Additionally, the availability of appropriate corpora has been an issue. Having an LLM produce typical phrases, sentences, and paragraphs for specific genres and communicative situations allows teachers to provide a large number of such examples that *could* be found in actual texts and are acceptable for the writing task at hand by just formulating in natural language what these examples should express or look like. At the same time, learners too can get immediate examples even including specific words and phrases from their writing task.

The creation of appropriate tasks and exercises for learners is an important aspect of language teaching, particularly to support internal differentiation, which requires personalized instruction. Automatic exercise generation can help to ensure comparability of individual tasks with respect to topic, difficulty, language level, etc. Taking into account that LLMs are able to produce plausible texts meeting specific criteria, they can be used to produce writing prompts, fill-in-the-blank exercises, and even more sophisticated types of tests, which have been shown to be helpful for learners, but very challenging to create (Mahlow/Hess 2004; Mahlow/Piotrowski/Fenske 2010).

It is important to note that the automatic generation of exercises still requires manual specification, as demonstrated by Heck et al. (2022). The use of language models in a co-creative way can make this process easier. For instance, Zou et al. (2022) showed that language models can be used to automatically generate true/false reading comprehension questions from texts. In both cases, further progress on the power of LLMs and hybrid neuro-symbolic approaches – i. e., the combination of LLMs with knowledge bases – will improve availability and quality in the coming months. We will also see combinations with templates that allow for seamless integration into learning management systems (LMS) and the like – one already existing example is QuestionAid⁷.

LLMs can produce examples of specific phenomena in grammar or vocabulary even in ad hoc situations. To emphasize again: phrases and sentences automatically produced are linguistically acceptable and plausible, but they are not real-world examples, they are not authentic. These sentences *could* appear in texts and would be considered valid, but they cannot necessarily be found in any actually *existing* text. In contrast, examples extracted from corpora are instances of actual language use as it has been observed, and one can access the co-text of these phrase and sentences. But as argued above: mastering corpus linguistics tools is more challenging than formulating a natural language prompt. Here, language teachers have to carefully weigh advantages and disadvantages and decide which approach to apply in which situation.

LLMs can provide teachers and learners with detailed instructions for communicative situations in role-plays or writing prompts. Most often, role-plays are used in language learning to practice the use of discourse devices, where it does not matter whether the “facts” used in weather reports or tourist recommendations about museums are actually true. Similarly, writing prompts ask learners to use appropriate linguistic devices to write a text in a particular genre. Chatbots based on LLMs can be used as partners for interactive text-adventure games, they can take over the role of chat partners who also point out challenges to writers and correct grammar and spelling.

Taking into account generative AI beyond texts, some typical exercises in writing can be turned into more interactive self-regulated activities. Describing images and pictures challenges learner to use specific vocabulary as well as grammatical constructions to describe

⁷ <https://www.question-aid.com> (last access 2023-05-20).

which elements in an image are positioned where. There usually is no one single solution, various descriptions are possible. Careful examination of such description is needed to decide if it is in fact appropriate and to point to aspects still missing or even wrong. Generative AI systems like Dall-E⁸ or Midjourney⁹ produce images based on descriptions in natural language. Thus, if a description of a given image is appropriate, those tools should output a similar image when prompted with this description.

4 Conclusion

The rapid advancement of language technology, exemplified by applications using large language models like ChatGPT, has opened up new possibilities for teaching and learning writing. AI-based language technology has the potential to enhance language learning experiences, in particular learning how to write; but its integration into education requires careful consideration of its capabilities and limitations.

Writing in German, as both a first and second language, requires the development of cognitive, pragmatic, and linguistic competencies. Learners must understand genre conventions, organize information effectively, and employ appropriate linguistic features. Language technology can facilitate and enhance the learning process by providing opportunities for practice, feedback, and assessment. However, it is important to acknowledge that language technology is not a substitute for human guidance and expertise. Rather, it should be viewed as a tool to support and augment the teaching and learning of writing.

As language technology continues to evolve, it will inevitably impact writing processes and respective pedagogical approaches. However, it is crucial to recognize that technology alone cannot replace the role of teachers in guiding and shaping students' writing skills. Learning how to write will remain an integral part of curricula, with language technology serving as a valuable resource for learners and teachers alike. By embracing the opportunities presented by language technology and incorporating it thoughtfully into language learning environments, we can enhance writing instruction and empower learners to become effective communicators in German and other languages.

In conclusion, the integration of large language models and AI-based tools into language learning requires a balanced approach that combines the capabilities of technology with the expertise of language teachers. By harnessing the power of language technology while maintaining the centrality of human guidance, we can create dynamic and effective learning environments that help learners acquire the skills they need to succeed in written communication in a self-regulatory manner.

8 <https://labs.openai.com/> (last access 2023-05-21).

9 <https://www.midjourney.com/> (last access 2023-05-21).

References

- Adlakha, Vaibhav/Dhuliawala, Shehzaad/Suleman, Kaheer/Vries, Harm de/Reddy, Siva (2022) TopiOCQA: Open-domain conversational question answering with topic switching. In: *Transactions of the Association for Computational Linguistics* 10, 468-483. https://doi.org/10.1162/tacl_a_00471.
- Agrawala, Maneesh/Wobbrock, Jacob O./Adar, Eytan/Setlur, Vidya (2022) *UIST '22: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. Bend, OR, USA, 29 October 2022 – 2 November 2022. <https://dblp.org/db/conf/uist/index.html> [last access 2023-06-02].
- Allal, Linda/Chanquoy, Lucile/Largy, Pierre (Eds.) *Revision. Cognitive and Instructional Processes, Studies in Writing*. Boston, Dordrecht, London: Kluwer.
- Bowker, Lynne/Ciro, Jairo Buitrago (2019) *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Bingley: Emerald Group Publishing.
- Brown, Tom/Mann, Benjamin/Ryder, Nick/Subbiah, Melanie/Kaplan, Jared D./Dharawal, Prafulla/Neelakantan, Arvind/Shyam, Pranav/Sastry, Girish/Askell, Amanda/Agarwal, Sandhini/Herbert-Voss, Ariel/Krueger, Gretchen/Henighan, Tom/Child, Rewon/Ramesh, Aditya/Ziegler, Daniel/Wu, Jeffrey/Winter, Clemens/Hesse, Chris/Chen, Mark/Sigler, Eric/Litwin, Mateusz/Gray, Scott/Chess, Benjamin/Clark, Jack/Berner, Christopher/McCandlish, Sam/Radford, Alec/Sutskever, Ilya/Amodei, Dario (2020) Language models are few-shot learners. In: Larochelle, Hugo/Ranzato, Marc'Aurelio/Hadsell, Raia/Balcan, Maria-Florina/Lin, Hsuan-Tien (Eds.), 1877-1901.
- Chang, Minsuk/Druga, Stefania/Fiannaca, Alex/Vergani, Pedro/Kulkarni, Chinmay/Cai, Carrie/Terry, Michael (2023) *The prompt artists*. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2303.12253> [last access 2023-06-02].
- Chinkina, Maria/Kannan, Madeeswaran/Meurers, Detmar (2016) Online information retrieval for language learning. In: *Proceedings of ACL-2016 System Demonstrations*. Stroudsburg, PA, USA: Association for Computational Linguistics, 7-12.
- Chowdhery, Aakanksha/Narang, Sharan/Devlin, Jacob/Bosma, Maarten/Mishra, Gaurav/Roberts, Adam/Barham, Paul/Chung, Hyung Won/Sutton, Charles/Gehrmann, Sebastian/Schuh, Parker/Shi, Kensen/Tsvyashchenko, Sasha/Mayne, Joshua/Rao, Abhishek/Barnes, Parker/Tay, Yi/Shazeer, Noam/Prabhakaran, Vinodkumar/Reif, Emily/Du, Nan/Hutchinson, Ben/Pope, Reiner/Bradbury, James/Austin, Jacob/Isard, Michael/Gur-Ari, Guy/Yin, Pengcheng/Duke, Toju/Levskaya, Anselm/Ghemawat, Sanjay/Dev, Sunipa/Michalewski, Henryk/Garcia, Xavier/Misra, Vedant/Robinson, Kevin/Fedus, Liam/Zhou, Denny/Ippolito, Daphne/Luan, David/Lim,

- Hyeontaek/Zoph, Barret/Spiridonov, Alexander/Sepassi, Ryan/Dohan, David/Agrawal, Shivani/Omernick, Mark/Dai, Andrew M./Pillai, Thanumalayan Sankaranarayana/Pellat, Marie/Lewkowycz, Aitor/Moreira, Erica/Child, Rewon/Polozov, Oleksandr/Lee, Katherine/Zhou, Zongwei/Wang, Xuezhi/Saeta, Brennan/Diaz, Mark/Firat, Orhan/Catasta, Michele/Wei, Jason/Meier-Hellstern, Kathy/Eck, Douglas/Dean, Jeff/Petrov, Slav/Fiedel, Noah (2022) *PaLM: Scaling language modeling with pathways*. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2204.02311> [last access 2023-06-02].
- Clematide, Simon/Klenner, Manfred/Volk, Martin (Eds.) (2014) *Searching Answers – Festschrift in Honour of Michael Hess on the Occasion of His 60th Birthday*. Münster: MV-Verlag.
- Council of Europe (2020) *Common european framework of reference for languages: Learning, teaching, assessment – companion volume*. Strasbourg, France: Council of Europe Publishing.
- Council of Europe (2001) *Common european framework of reference for languages*. Cambridge, UK: Cambridge University Press.
- Dale, Robert (2020) Natural language generation: The commercial state of the art in 2020. In: *Natural Language Engineering* 26(4), 481-487. <https://doi.org/10.1017/S135132492000025X>.
- Dale, Robert/Moisl, Herman/Somers, Harold (Eds.) (2000) *Handbook of Natural Language Processing*. New York, NY, USA: Marcel Dekker.
- Dale, Robert/Viethen, Jette (2021) The automated writing assistance landscape in 2021. In: *Natural Language Engineering* 27(4), 511-518. <https://doi.org/10.1017/S1351324921000164>.
- Dang, Hai/Benharrak, Karim/Lehmann, Florian/Buschek, Daniel (2022) Beyond text generation: Supporting writers with continuous automatic text summaries. In: Agrawala et al. (eds.), article No. 98. <https://doi.org/10.1145/3526113.3545672>.
- Das, Dipanjan/Smith, Noah A. (2009) Paraphrase identification as probabilistic quasi-synchronous recognition. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. Stroudsburg, PA, USA: Association for Computational Linguistics, 468-476.
- Delorme Benites, Alice/Lehr, Caroline (2021) Neural machine translation and language teaching: Possible implications for the CEFR. In: *Bulletin suisse de linguistique appliquée* 114, 47-66. <https://doi.org/10.21256/zhaw-24716>.

- Devlin, Jacob/Chang, Ming-Wei/Lee, Kenton/Toutanova, Kristina (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 4171-4186.
- Dowling, Carolyn (1994) Word processing and the ongoing difficulty of writing. In: *Computers and Composition* 11(3), 227-235. [https://doi.org/10.1016/8755-4615\(94\)90015-9](https://doi.org/10.1016/8755-4615(94)90015-9).
- Fayol, Michel (1999) From on-line management problems to strategies in written composition. In: Torrance, Mark/Jeffery, Gaynor (Eds.), 13-23.
- Fernando, Samuel/Stevenson, Mark (2009) A semantic similarity approach to paraphrase detection. In: *Proceedings of the 11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics*. Oxford: UK Special Interest Group for Computational Linguistics.
- Firth, John Rupert (1957) *Studies in linguistic analysis*. Oxford: Blackwell.
- Fontenelle, Thierry (2005) Dictionnaires et outils de correction linguistique. In: *Revue française de linguistique appliquée* X(2), 119-128.
- Fu, Qing-Ke/Zou, Di/Xie, Haoran/Cheng, Gary (2022) A review of AWE feedback: Types, learning outcomes, and implications. In: *Computer Assisted Language Learning* 0(0), 1-43. <https://doi.org/10.1080/09588221.2022.2033787>.
- Gamon, Michael (2010) Using Mostly Native Data to Correct Errors in Learners' Writing. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 163-171.
- Ganslandt, Sebastian/Jörwall, Jakob/Nugues, Pierre (2009) Predictive text entry using syntax and semantics. In: *Proceedings of the 11th International Conference on Parsing Technologies, IWPT '09*. Stroudsburg, PA, USA: Association for Computational Linguistics, 37-48.
- Heck, Tanja/Meurers, Detmar/Nuxoll, Florian (2022) Automatic exercise generation to support macro-adaptivity in intelligent language tutoring systems. In: *Intelligent CALL, Granular Systems, and Learner Data: Short Papers from EUROCALL 2022*, 162-167.
- Heidorn, George E. (2000) Intelligent writing assistance: Techniques and applications for the processing of language as text. In: Dale, Robert/Moisl, Herman/Somers, Harold (Eds.), 181-207.

- Heidorn, George E./Jensen, K./Miller, L.A./Byrd, R.J./Chodorow, M.S. (1982) The EPISTLE text-critiquing system. In: *IBM Systems Journal* 21(3), 305-326. <https://doi.org/10.1147/sj.213.0305>.
- Hofstadter, Douglas R. (1996) *Fluid Concepts and Creative Analogies*. London: Basic Books.
- Hulth, Anette (2003) Improved automatic keyword extraction given more linguistic knowledge. In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA: Association for Computational Linguistics, 216-223.
- Jakobs, Eva-Maria/Perrin, Daniel (Eds.) (2014) *Handbook of Writing and Text Production, Handbooks of Applied Linguistics*. Berlin, Germany: de Gruyter Mouton.
- Jia, Ruipeng/Cao, Yanan/Tang, Hengzhu/Fang, Fang/Cao, Cong/Wang, Shi (2020) Neural extractive summarization with hierarchical attentive heterogeneous graph network. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 3622-3631.
- Kellogg, Ronald T. (2008) Training writing skills: A cognitive developmental perspective. In: *Journal of Writing Research* 1(1), 1-26.
- Kwiatkowski, Tom/Palomaki, Jennimaria/Redfield, Olivia/Collins, Michael/Parikh, Ankur/Alberti, Chris/Epstein, Danielle/Polosukhin, Illia/Devlin, Jacob/Lee, Kenton/Toutanova, Kristina/Jones, Llion/Kelcey, Matthew/Chang, Ming-Wei/Dai, Andrew M./Uszkoreit, Jakob/Le, Quoc/Petrov, Slav (2019) Natural questions: A benchmark for question answering research. In: *Transactions of the Association for Computational Linguistics* 7, 452-466. https://doi.org/10.1162/tacl_a_00276.
- Larochelle, Hugo/Ranzato, Marc'Aurelio/Hadsell, Raia/Balcan, Maria-Florina/Lin, Hsuan-Tien (Eds., 2020) *Advances in Neural Information Processing Systems*. Red Hook: Curran Associates.
- Lee, Sangmin-Michelle (2020) The impact of using machine translation on EFL students' writing. In: *Computer Assisted Language Learning* 33(3), 157-175. <https://doi.org/10.1080/09588221.2018.1553186>.
- Leontjev, Dmitri (2014) The effect of automated adaptive corrective feedback: L2 english questions. In: *APPLES: Journal of applied language studies* 8(2), 43-66.
- Levy, Michael (1997) *Computer-assisted language learning: Context and conceptualization*. Oxford: Oxford University Press.

- Macdonald, Nina H./Fraser, Lawrence T./Gingrich, Patricia S./Keenan, Stacey A. (1982) The Writer's Workbench: Computer aids for text analysis. In: *IEEE Transactions on Communication* 30(1), 105-110.
- Mahlow, Cerstin/Dale, Robert (2014) Production Media: Writing as Using Tools in Media Convergent Environments. In: Jakobs, Eva-Maria/Perrin, Daniel (Eds.), 209-230.
- Mahlow, Cerstin/Hediger, Andreas (2021) Education as loosely coupled system of technology and pedagogy. In: *On Education. Journal for Research and Debate* 4(12), ohne Seitenzahlen. https://doi.org/10.17899/on_ed.2021.12.7.
- Mahlow, Cerstin/Hess, Michael (2004) Sentence Completion Tests for Training and Assessment in a Computational Linguistics Curriculum. In: *COLING-2004 Workshop on eLearning for Computational Linguistics and Computational Linguistics for eLearning*. Stroudsburg, PA, USA: Association for Computational Linguistics, 61-70.
- Mahlow, Cerstin/Piotrowski, Michael (2009) A target-driven evaluation of morphological components for German. In: Clematide, Simon/Klenner, Manfred/Volk, Martin (Eds.), 85-99.
- Mahlow, Cerstin/Piotrowski, Michael/Fenske, Wolfram (2010) A solid foundation: Why e-assessment should be based on a systematic typology of test items. In: Nunes, Miguel B./McPherson, Maggie (Eds.), 47-51.
- Meurers, Detmar/Ziai, Ramon/Amaral, Luiz/Boyd, Adriane/Dimitrov, Aleksandar/Metcalf, Vanessa/Ott, Niels (2010) Enhancing Authentic Web Pages for Language Learners. In: *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*. Stroudsburg, PA, USA: Association for Computational Linguistics, 10-18.
- Nunes, Miguel B./McPherson, Maggie (Eds.) (2010) *Proceedings of the IADIS International Conference e-Learning 2010, Freiburg, Germany, July 26-29, 2010*. Freiburg i. Br.: IADIS Press.
- OpenAI (2023) *GPT-4 technical report*, Preprint at arXiv. <https://10.48550/arXiv.2303.08774> [last access 2023-06-02].
- Piolat, Annie/Roussey, Jean-Yves/Olive, Thierry/Amada, Murielle (2004) Processing time and cognitive effort in revision: Effects of error type and of working memory capacity. In: Allal, Linda/Chanquoy, Lucile/Largy, Pierre (Eds.), 21-38.
- Piotrowski, Michael (2012) *Natural Language Processing for Historical Texts*. San Rafael, CA, USA: Morgan & Claypool.
- Rei, Marek/Cummins, Ronan (2016) Sentence similarity measures for fine-grained estimation of topical relevance in learner essays. In: *Proceedings of the 11th Workshop on*

- Innovative Use of NLP for Building Educational Applications*. Stroudsburg, PA, USA: Association for Computational Linguistics, 283-288.
- Reiter, Ehud/Dale, Robert (2000) *Building Natural Language Generation Systems*. *Studies in natural language processing*. Cambridge, UK: Cambridge University Press.
- Schneider, Adela/Madsack, Andreas/Heininger, Johanna/Chen, Ching-Yi/Weißgraber, Robert (2022) Data-to-text systems as writing environment. In: *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)*. Dublin, Ireland: Association for Computational Linguistics, 1-10.
- Schwitler, Rolf/Molla, Diego/Fournier, Rachel/Hess, Michael (2000) Answer extraction towards better evaluations of NLP systems. In: *ANLP-NAACL 2000 Workshop: Reading Comprehension Tests as Evaluation for Computer-Based Language Understanding Systems*, Stroudsburg, PA, USA: Association for Computational Linguistics, 20-27.
- Sejnowski, Terrence J. (2023) Large Language Models and the Reverse Turing Test. In: *Neural Computation* 35(3), 309-342. https://doi.org/10.1162/neco_a_01563.
- Seonwoo, Yeon/Kim, Ji-Hoon/Ha, Jung-Woo/Oh, Alice (2020) Context-aware answer extraction in question answering. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2418-2428.
- Shermis, Mark D./Burststein, Jill C. (Eds.) (2002) *Automated essay scoring: A cross-disciplinary perspective*. New York, NY, USA: Routledge.
- Sjöblom, Eetu/Creutz, Mathias/Vahtola, Teemu (2021) Grammatical error generation based on translated fragments. In: *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. Reykjavik, Iceland (Online): Linköping University Electronic Press, Sweden, 398-403.
- Somers, Harold (2005) Round-trip translation: What is it good for? In: *Proceedings of the Australasian Language Technology Workshop 2005*. Sydney, Australia, 127-133.
- Tay, Yi/Dehghani, Mostafa/Bahri, Dara/Metzler, Donald (2022) Efficient transformers: A survey. In: *ACM Computing Surveys* 55(6). <https://doi.org/10.1145/3530811>.
- Torrance, Mark/Galbraith, David (Eds.) (1999) *Knowing What to Write: Conceptual Processes in Text Production*. 1st ed. Studies in writing. Amsterdam: Amsterdam University Press.
- Torrance, Mark/Jeffery, Gaynor (Eds.) (1999) *The Cognitive Demands of Writing: Processing Capacity and Working Memory Effects in Text Production*, *Studies in Writing*. Amsterdam: Amsterdam University Press.

- Tschichold, Cornelia/Bodmer, Franck/Cornu, Etienne/Grosjean, Francois/Grosjean, Lysiane/Kubler, Natalie/Lewy, Nicolas/Tschumi, Corinne (1997) Developing a new grammar checker for English as a second language. In: *From Research to Commercial Applications: Making NLP Work in Practice*. Stroudsburg, PA, USA: Association for Computational Linguistics, 7-12.
- Wang, Huanhuan/Tlili, Ahmed/Huang, Ronghuai/Cai, Zhenyu/Li, Min/Cheng, Zui/Yang, Dong/Li, Mengti/Zhu, Xixian/Fei, Cheng (2023) Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. In: *Education and Information Technologies*. <https://doi.org/10.1007/s10639-022-11555-x>.
- Wang, Xinyu/Jiang, Yong/Bach, Nguyen/Wang, Tao/Huang, Zhongqiang/Huang, Fei/Tu, Kewei (2021) Improving named entity recognition by external context retrieving and cooperative learning. In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 1800-1812.
- Wang, Zhengjue/Duan, Zhibin/Zhang, Hao/Wang, Chaojie/Tian, Long/Chen, Bo/Zhou, Mingyuan (2020) Friendly topic assistant for transformer based abstractive summarization. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 485-497.
- Weegar, Rebecka/Idestam-Almquist, Peter (2023) Reducing Workload in Short Answer Grading Using Machine Learning. In: *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-022-00322-1>.
- Wulff, Peter/Mientus, Lukas/Nowak, Anna/Borowski, Andreas (2022) Utilizing a Pre-trained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections. In: *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-022-00290-6>.
- Yuan, Ann/Coenen, Andy/Reif, Emily/Ippolito, Daphne (2022) Wordcraft: Story writing with large language models. In: *27th International Conference on Intelligent User Interfaces, IUI '22*. New York, NY, USA: Association for Computing Machinery, 841-852.
- Yuan, Zheng/Bryant, Christopher (2021) Document-level grammatical error correction. In: *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*. Stroudsburg, PA, USA: Association for Computational Linguistics, 75-84.

- Zamfirescu-Pereira, J.D./Wong, Richmond Y./Hartmann, Bjoern/Yang, Qian (2023) Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*. New York, NY, USA: Association for Computing Machinery.
- Zhang, Ruofei/Zou, Di (2022) Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. In: *Computer Assisted Language Learning* 35(4), 696-742. <https://doi.org/10.1080/09588221.2020.1744666>.
- Zou, Bowei/Li, Pengfei/Pan, Liangming/Aw, Ai Ti (2022) Automatic true/false question generation for educational purpose. In: *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 61-70.