

Artificial Intelligence vs. Corpus Literacy

Ansätze zur Vermittlung reflektierter Schreibkompetenz in der Wissenschaftssprache Englisch

Julia Schlüter, Katharina Deckert

Abstract: Dieser Beitrag evaluiert die Anwendung der KI-Tools ChatGPT (Version 3.5 und 4), DeepL Write und Grammarly bei der Korrektur und Endredaktion eines Textes in der Wissenschafts(fremd)sprache Englisch aus sprachwissenschaftlicher Sicht (basierend auf linguistischen Korpora). Dabei wird systematisch unterschieden zwischen Fällen, in denen KI-generierte Änderungen zielführend sind, in denen überflüssige Änderungen vorgenommen werden und in denen Fehler unkorrigiert bleiben. Anhand dieses exemplarischen use case wird die Relevanz eines souveränen, kritisch-informierten Umgangs mit KI-Tools durch Anwender:innen demonstriert.

Adopting a linguistic perspective (based on language corpora), this contribution evaluates the application of the AI tools ChatGPT (versions 3.5 and 4), DeepL Write and Grammarly for correcting and revising a text in English as an academic lingua franca. A systematic distinction is made between cases in which AI-generated changes are conducive to improvements, in which superfluous corrections are suggested, and in which real errors remain uncorrected. In this exemplary use case, the study thus demonstrates the relevance of users' expertise and critical awareness in the application of AI tools.

Keywords: ChatGPT; DeepL Write; Grammarly; Korpuslinguistik; Fehlerkorrektur; wissenschaftliches Schreiben auf Englisch; corpus linguistics; error correction; academic writing in English

1. Einleitung

Durch das Erscheinen frei zugänglicher und leicht handhabbarer KI-Tools haben sich in zahlreichen informationsverarbeitenden und -generierenden Disziplinen neue Chancen aufgetan, so auch in der universitären Forschung und Lehre. Aus diesem Grund ist es wichtig, Studierende und Forschende zu einem verantwor-

tungsbewussten und reflektierten Umgang mit diesen innovativen Werkzeugen zu befähigen. Während viele verlockend erscheinende Anwendungsmöglichkeiten im Zusammenhang mit dem Verfassen wissenschaftlicher Texte äußerst kritisch zu sehen sind (Limburg et al., 2023), soll in diesem Beitrag die vielversprechende Einsatzmöglichkeit von KI-Tools wie *ChatGPT*, *DeepL Write* und *Grammarly* als Korrekturhilfen aus sprachwissenschaftlicher Perspektive betrachtet und evaluiert werden.

In der modernen empirisch-deskriptiven Sprachwissenschaft gehört die Nutzung großer digitaler Textdatenbanken, sogenannter Korpora, seit Ende des 20. Jahrhunderts zum unabdingbaren Handwerkszeug; ohne sie wären Grundlagenforschung und Angewandte Linguistik heute kaum denkbar. Fast ebenso lange werden diese Methoden auch zur Professionalisierung von fortgeschrittenen Lernenden und Lehrkräften der Fremdsprache Englisch nachdrücklich empfohlen (Mukherjee, 2002; Mair, 2002; Timmis, 2015), unter anderem zur selbstständigen Klärung sprachlicher Zweifelsfälle beim Redigieren und Korrigieren von Texten. In der Praxis finden diese Bemühungen allerdings nur begrenzten Anklang, wohl auch, weil der Erwerb von *corpus literacy*, also von Kompetenzen in der effizienten Informationsgewinnung aus Korpora, hohe kognitive und methodische Anforderungen an Lernende stellt. Das Feedback zu einem semesterfüllenden Seminar zur gezielten Vermittlung von Korpusanalysetechniken, das regelmäßig an der Universität Bamberg angeboten wird, zeigt das immer wieder (Großmann & Schlüter, 2024).

Ist diese anspruchsvolle Kompetenz aber angesichts der Verfügbarkeit von KI-Sprachtools überhaupt noch notwendig und zeitgemäß? Wie in vielen anderen Tätigkeitsfeldern stellt sich in dieser und anderen universitären Lehrveranstaltungen zum wissenschaftlichen Schreiben aktuell die Frage, ob die Künstliche Intelligenz das Training der ›natürlichen Intelligenz‹ mit dem Ziel der Erlangung fachspezifischer methodischer Fähigkeiten überflüssig macht.

Large Language Models (LLMs) wie *ChatGPT*, *DeepL Write* und *Grammarly* basieren ebenfalls auf Korpora (wenn auch um ein Vielfaches größeren als den üblichen linguistischen Referenzkorpora) und lassen sich noch dazu leicht dialogisch in natürlicher Sprache steuern. Beim wissenschaftlichen Schreiben bieten sie vielseitige hilfreiche (jedoch noch zu evaluierende) Anwendungen, vom Brainstorming über die Strukturierung von Argumenten bis hin zur Formatierung von Publikationen (Buck & Limburg, 2023, S. 71). Eine weitere Funktion, der das besondere Augenmerk dieses Beitrags gilt, ist die sprachliche Korrektur von selbst verfassten Texten, hier exemplarisch für die Wissenschafts(fremd)sprache Englisch vorgestellt.

Englisch ist die *lingua franca* der Wissenschaft; jedoch sind Nicht-Muttersprachler:innen hinsichtlich der Rezeption und Produktion englischer Texte naturgemäß häufig im Nachteil. KI-Tools können hier helfen, das wissenschaftliche Schreiben zu verbessern, mangelnde Sprachkompetenz auszugleichen und mit der internatio-

nen Konkurrenz mitzuhalten. So kann den Tools zugutegehalten werden, dass sie Bildungs- und Karriereschränken abbauen und die Chancengleichheit in der Wissenschaft fördern (Amano et al., 2023; Lavista Ferres, 2023).

Das Ziel dieses Beitrags ist, die Vorteile und Grenzen von LLMs bei der Textkorrektur (unseres Wissens erstmalig) systematisch und sprachwissenschaftlich fundiert zu evaluieren.

2. Vorgehensweise

Um das Korrekturverhalten von KI-Tools zu beurteilen, wurde ein geeigneter Beispieltext erstellt, der einen Auszug aus einer von Studierenden verfassten wissenschaftlichen Arbeit (wie sie beispielsweise in den Geistes- und Sozialwissenschaften häufig als Nachweis erreichter Lernziele gefordert wird) darstellen soll. Gezielt wurden 15 typische Lernerfehler aus verschiedenen Problemfeldern des Englischen im Text platziert. Als Vergleichsmaßstab für die Unterscheidung zwischen ›richtig‹ und ›falsch‹ – besser: ›normkonform‹ und ›normabweichend‹ – dient der Sprachgebrauch, der bei Muttersprachler:innen des Englischen in wissenschaftlichen Arbeiten üblich ist. Empirische Evidenz hierfür kann mit einem beträchtlichen Maß an linguistischem Know-how aus den eben schon genannten Korpora gewonnen werden. Hierzu wurde die frei im Netz zugängliche Korpusplattform www.english-corpora.org genutzt (Davies, 2008).

Aus Platzgründen werden hier nur zwei der im Text verwendeten Fehler exemplarisch aus linguistischer Perspektive diskutiert. Der Satzkontext, mit durch Unterstreichung kenntlich gemachten Fehlern, lautet:

There's a *stark* tendency of *judging* grammar as either right or wrong, while usage data highlights the diversity of linguistic expression and gradual frequency distinctions in a scale of 0 to 100 percent.

Beispielsweise wird die Verwendung von Kontraktionen wie *there's* in einem wissenschaftlichen Aufsatz im Allgemeinen (noch) nicht toleriert. Dies kann mit einer Korpusabfrage gezeigt werden, die die Häufigkeit der Kontraktion in verschiedenen Textgenres¹, standardisiert auf der Basis von einer Million Wörtern, visualisiert. Dabei wird augenfällig, dass *there's* in gesprochener Sprache sehr häufig und in fiktionalen Texten, Zeitschriften und Zeitungen gelegentlich vorkommt, aber in wissenschaftlichen Texten deutlich vermieden wird.

Weiterhin ist die Formulierung *a stark tendency* eher unüblich und könnte durch Interferenz des deutschen Ausdrucks *eine starke Tendenz* bedingt sein. Eine ge-

¹ Link zur Korpusuche: <https://www.english-corpora.org/coca/?c=coca&q=120242260>

bräuchlichere Kollokation, die in der Liste der häufig mit *tendency* kombinierten Adjektive² (nach Ausschluss der Adjektive mit andersartigen Bedeutungen) ganz oben zu finden ist, wäre *a strong tendency*. Das Adjektiv *stark* kommt dagegen unter den Top 1000 der adjektivischen Attribute von *tendency* nicht vor.

Um *ChatGPT* als Korrekturhilfe zu verwenden, muss zunächst ein zielführender Prompt formuliert werden, der das Tool möglichst detailliert instruiert, ihm Normen vorgibt und Grenzen setzt. Die im Test verwendete Anweisung lautete:

Can you help me correct the following text written by a non-native writer who is a university student of English? It should be written in British English for an informed academic audience. Could you focus on making corrections only, without rephrasing too much of the text? And could you highlight the corrections you made in your text output?

Der zu korrigierende Text wurde anschließend in den Chat eingegeben. Unter anderem generierte dies den folgenden Korrekturvorschlag:

There's a **strong tendency to judge** grammar as either right or wrong, while usage data highlights the diversity of linguistic expression and gradual frequency distinctions **on** a scale **from 0 to 100 per cent**. (*ChatGPT*)

In diesem Beispiel wurde keine Korrektur von *There's* zu *There is* vorgenommen. Dafür wurde *stark tendency* adäquat zu *strong tendency* korrigiert. Die drei weiteren Fehler wurden ebenfalls verbessert. Darüber hinaus wurde die Formulierung *a scale of 0 to 100* zu *a scale from 0 to 100* geändert, was durchaus plausibel erscheint. Eine anschließende Korpusuche nach den Präpositionen, die den Zahlenbereich einer Skala definieren können,³ zeigt allerdings, dass die ursprüngliche Version mit *of* deutlich häufiger ist als die von *ChatGPT* vorgeschlagene Alternative *from*. Derartige nicht zwingend notwendige Änderungen wurden konsequent protokolliert. Praktisch nie wurden dagegen falsche Formulierungsvorschläge gefunden.

Bei der Verwendung von *DeepL Write* und *Grammarly*, die beide in ihren freien Online-Versionen verwendet wurden, ist das Formulieren eines Prompts nicht notwendig. Stattdessen können in beiden Apps verschiedene Einstellungen vorgenommen werden, in diesem Fall also wiederum britisches Englisch sowie ein formeller, wissenschaftlicher Stil. Die von *DeepL Write* vorgeschlagenen Modifikationen lauteten:

2 Link zur Korpusuche: <https://www.english-corpora.org/coca/?c=coca&q=120274866>

3 Link zur Korpusuche: <https://www.english-corpora.org/coca/?c=coca&q=120244302>

There's is a stark tendency of judging → to judge grammar as either right or wrong, while → but usage data highlights → shows the diversity of linguistic expression and gradual frequency distinctions in → on a scale of 0 to 100 per cent. (*DeepL Write*)

Hier wurde *There's* registerkonform zu *There is* verändert; *stark* wurde ersatzlos gestrichen; *a scale of 0 to 100* blieb unverändert. Jedoch fallen (neben zwei der drei hier nicht weiter betrachteten, aber notwendigen Korrekturen) in der Satzmitte zusätzliche Änderungen ins Auge, die allenfalls stilistische Varianten, doch keinen echten Korrekturbedarf darstellen.

In *Grammarly* werden Korrekturvorschläge durch Unterstreichung markiert und können per Mausklick angezeigt und akzeptiert werden:

There's a stark tendency of judging grammar as either right or wrong, while usage data highlights the diversity of linguistic expression and gradual frequency distinctions in a scale of 0 to 100 percent. (*Grammarly*)

Wie bereits bei *ChatGPT* wurde *There's* in *Grammarly* nicht zu *There is* korrigiert. Auch *stark tendency* wurde übersehen. *A scale of 0 to 100* blieb unangetastet, was der Korpus-evidenz entspricht. Drei weitere Fehler wurden bemerkt, während keine zusätzlichen Varianten vorgeschlagen wurden.

3. Ergebnisse

Die Tests zeigen, dass von *ChatGPT*, *DeepL Write* und *Grammarly* viele der bewusst eingebauten echten Fehler entdeckt werden. Jedoch unterscheiden die Tools nicht zwischen solchen Korrekturen und anderen alternativen Ausdrucksweisen, die aus linguistischer Sicht lediglich mögliche Formulierungsvarianten sind. Daher sollten Nutzer:innen bei der Verwendung dieser Tools berücksichtigen, dass trotz einer klar auf Fehlerkorrektur abzielenden Aufgabenstellung nicht selten mehr als nötig geändert wird.

Dieses erratisch erscheinende Korrekturverhalten entsteht dadurch, dass LLMs auf statistischen Wahrscheinlichkeiten beruhen. Ihr Output ist probabilistisch; er entspricht also den Wahrscheinlichkeiten, die die Modelle in ihrer Trainingsphase aus den in sie hineingegebenen Textmengen abgeleitet haben (Stöffelbauer, 2023). Während *Grammarly* und *DeepL Write* bei der oben beschriebenen Vorgehensweise reproduzierbar stets dieselben Alternativen (und im Fall von *DeepL Write* weitere Ersetzungsvorschläge als kontextabhängige Drop-Down-Listen) unterbreiten, kommt bei *ChatGPT* ein Zufallsfaktor hinzu: Das generative Sprachmodell ist so programmiert, dass es nicht immer die Option mit der höchsten Wahrscheinlichkeit wählt, so dass es bei jedem Durchlauf etwas andere Ausgaben liefert (Stöffelbauer,

2023). Mithilfe der Möglichkeit, per Knopfdruck einen neuen Output zu einem vorgegebenen Prompt zu generieren, kann die Treffsicherheit der Korrekturvorschläge statistisch evaluiert werden. Anhand von *ChatGPT*s frei verfügbarer Version 3.5 und der kostenpflichtigen Version 4 wurden für dieses Vorhaben jeweils 100 Durchläufe ausgeführt. Die so erhaltenen Korrekturen und Varianten werden in Tab. 1 einzelnen Durchgängen von *Grammarly* und *DeepL Write* gegenübergestellt.

Tab. 1: Ergebnisüberblick für alle verwendeten Tools

	Fehler n = 15		Weitere Änderungen n = 10
	davon korrigiert	davon unentdeckt	
<i>Grammarly</i>	53 %	47 %	10 %
Korrekturdurchläufe: 1	(8/15)	(7/15)	(1/10)
<i>ChatGPT</i> 3.5	64 %	36 %	26 %
Korrekturdurchläufe: 100	(967/1500)	(533/1500)	(264/1000)
<i>ChatGPT</i> 4	70 %	30 %	32 %
Korrekturdurchläufe: 100	(1015/1500)	(450/1500)	(318/1000)
<i>DeepL Write</i>	87 %	13 %	50 %
Korrekturdurchläufe: 1	(13/15)	(2/15)	(5/10)

Alle vier Tools finden zumindest die Hälfte der im Beispieltex eingebaute echten Fehler. Das Sprachmodell von *ChatGPT* 4 erweist sich im direkten Vergleich als zuverlässiger als dasjenige von Version 3.5. Mit 87 Prozent schneidet *DeepL Write* bezüglich adäquater Korrekturen von Fehlern am besten ab, nimmt sich aber gleichzeitig auch die größten Freiheiten, nicht notwendige Änderungen vorzunehmen: Fünf der insgesamt zehn verzeichneten Varianten gehen allein auf *DeepL Write* zurück. Eine genauere Untersuchung der statistischen Häufigkeit dieser Varianten in linguistischen Korpora ist für eine Folgestudie geplant und wird eine präzisere Bewertung ermöglichen.

4. Diskussion

In Abwägung dieser Ergebnisse, gegenübergestellt in Tab. 2, ist von der Verwendung von KI-Tools zu Zwecken der redaktionellen Überarbeitung und Korrektur von Texten durch einen großen Nutzerkreis keineswegs abzuraten. Vielmehr lässt sich eine souveräne, kritisch-informierte Nutzung für viele Einsatzbereiche grundsätz-

lich befürworten und wird im eingangs beschriebenen Seminar sowie diesem Beitrag exemplarisch veranschaulicht. Wie die Textbeispiele und Statistiken gezeigt haben, korrigieren KI-Tools bei der vorgeschlagenen zielgerichteten und kontrollierten Verwendungsweise viele Fehler, übersehen zwar auch einige, aber produzieren äußerst selten zusätzliche Normabweichungen. Wenn eine überschaubare Restunsicherheit akzeptabel ist, können sie die Textqualität signifikant verbessern, Nutzende im Sinne eines *computational offloading* partiell entlasten und dadurch zeitliche Ressourcen für höherrangige Aufgaben bereitstellen (KI-Campus, 2023). Wie einleitend bereits erwähnt, besteht ein unbestreitbarer Vorteil in ihrer niederschweligen Nutzung durch linguistische Laien auch ohne fachspezifische methodische Kompetenzen. Unerwünschte KI-Effekte wie Halluzinationen sind bei den gezeigten engen Prompts praktisch ausgeschlossen.

Tab. 2: Stärken von KI- und korpusbezogenen Kompetenzen im Vergleich

AI literacy	Corpus literacy
<ul style="list-style-type: none">· Adäquate Korrektur vieler Fehler bei gleichzeitigem Übersehen anderer Fehler· Beträchtliche Restunsicherheit· Signifikante Textverbesserung· Chance des <i>computational offloading</i>· Einfache Bedienung auch durch linguistischeLaien	<ul style="list-style-type: none">· Souveräne Unterscheidung notwendiger, überflüssiger und fehlender Korrekturen· Unmittelbarer Zugang zu authentischen Textdaten· Belastbare, reproduzierbare Evidenz· Professionelle Urteilsfähigkeit· Intellektuelle Herausforderung

Die *Artificial Intelligence* sollte jedoch, wenn es um die zuverlässige Unterscheidung von ›richtig‹ und ›falsch‹ sowie dazwischenliegender Abstufungen geht, stets von einer geschulten *natural intelligence* geleitet werden, beispielsweise wenn davon eine Notengebung abhängt. Mithilfe korpuslinguistischer Methoden haben die berichteten Tests gezeigt, dass KI-Tools oft weitergehende Änderungen vorschlagen, die Nutzende unnötig verunsichern könnten. Ohne kritisches Bewusstsein bezüglich der Grenzen der KI könnten diese zu dem falschen Schluss kommen, ein überflüssigerweise korrigierter Ausdruck sei tatsächlich korrekturbedürftig, und ihn fortan zu vermeiden suchen. Insofern will der vorliegende Beitrag sowohl ein gesundes Misstrauen gegenüber KI-Tools als auch ein gesundes Vertrauen in die eigene sprachliche Intuition vermitteln.

Studierenden der Anglistik/Amerikanistik, und damit angehenden Expert:innen für die englische Sprache, wird so die Schlüsselfunktion der *corpus literacy*

vor Augen geführt, also einer auf selbstständiger (nicht durch LLMs vermittelter) Sprachdatenanalyse beruhenden professionellen Urteilsfähigkeit: Korpora liefern eindeutig reproduzierbare Evidenz, die belastbare Entscheidungen erlaubt. Jedoch ist eine routinierte und souveräne *corpus literacy* nicht ohne intensives Training zu erreichen. Stärken und Schwächen der natürlichen und Künstlichen Intelligenz stehen in einem Spannungsverhältnis, das die Sprachwissenschaft – wie auch andere Disziplinen – vor inhaltliche und hochschuldidaktische Herausforderungen stellt. Dieser Beitrag hat exemplarisch für die Englische Linguistik gezeigt, wie eine sinnvolle Aufgabenteilung und gewinnbringende Kosten-Nutzen-Abwägung zwischen beiden aussehen kann.

Disclaimer: Der Fokus dieses Beitrags liegt auf linguistischen Aspekten der Überarbeitung und Korrektur von Texten durch KI-Tools. Die problematischen und z.T. noch klärungsbedürftigen datenschutz- und urheberrechtlichen Aspekte der Eingabe von Texten in die Tools werden hierbei ausdrücklich ausgespart.

Förderhinweis: Die Entwicklung des KorPLUS-Selbstlernpakets zum Erwerb von *corpus literacy* wurde gefördert durch die Stiftung Innovation in der Hochschullehre (Projekt »Digitale Kulturen der Lehre entwickeln – DiKuLe«, Universität Bamberg).

Literaturverzeichnis

- Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J.D., Montaña-Centellas, F., Paudel, K., White, R.L. & Verissimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biol* 21(7), e3002184, <https://doi.org/10.1371/journal.pbio.3002184>
- Buck, I. & Limburg, A. (2023). Hochschulbildung vor dem Hintergrund von Natural Language Processing (KI-Schreibtools): Ein Framework für eine zukunftsfähige Lehr- und Prüfungspraxis. *die hochschullehre* 9(6). 70–84. <https://doi.org/10.3278/HSL2306W>
- Davies, M. (2008–) *The Corpus of Contemporary American English (COCA)*. Abgerufen am 18.01.2024 von <https://www.english-corpora.org/coca/>
- DeepL Write. Beta-Version. Abgerufen am 18.01.2024 von <https://www.deepl.com/de>
- Grammarly. Abgerufen am 18.01.2024 von app.grammarly.com
- Großmann, C. & Schlüter, J. (2024). Corpus literacy in der Lehrer*innenbildung: Englisch (lernen) lehren mit Korpora. In A. Rosen & K. Beuter (Hg.), *Englische Sprachwissenschaft und Fachdidaktik im Dialog: Chancen zur Stärkung der Lehrkräfte-*

- bildung* (S. 185–201). Narr Francke Attempto. <http://doi.org/10.24053/9783381112524>
- KI-Campus — Die Lernplattform für Künstliche Intelligenz. (2023). Sprachassistenten als Chance für die Hochschullehre: Modul 2: Einfluss von Sprachtechnologien auf die Hochschuldidaktik. Abgerufen am 03.05.2024 von <https://moodle.ki-campus.org/course/view.php?id=19§ion=2>
- Lavista Ferres, J. M. (2023). Lasst künstliche Intelligenz für euch schreiben! *Zeit Online*. Abgerufen am 16.01.2024 von <https://www.zeit.de/2023/48/chatgpt-wissen-schaft-englisch-chancengleichheit-sprachkenntnisse>
- Limburg, A., Bohle-Jurok, U., Buck, I., Grieshammer, E., Gröpler, J., Knorr, D., Mundorf, M., Schindler, K., Wilder, N. (2023). Zehn Thesen zur Zukunft des wissenschaftlichen Schreibens. Diskussionspapier Nr. 23. Berlin: *Hochschulforum Digitalisierung*. Abgerufen am 05.05.2024 von https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_DP_23_Zukunft_Schreiben_Wissenschaft.pdf
- Mair, C. (2002). Empowering non-native speakers: The hidden surplus value of corpora in continental English departments. In B. Kettemann & G. Marko (Hg.), *Teaching and learning by doing corpus analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora*, Graz 19–24 July, 2000 (S. 119–130). Brill. <http://dx.doi.org/10.1163/9789004334236>
- Mukherjee, J. (2002). *Korpuslinguistik und Englischunterricht: Eine Einführung* (Sprache im Kontext, Bd. 14.). Lang.
- OpenAI. *ChatGPT*. Abgerufen am 18.01.2024 von chat.openai.com
- Stöffelbauer, A. (2023). *How Large Language Models work: From zero to ChatGPT*. Abgerufen am 03.05.2024 von <https://medium.com/data-science-at-microsoft/how-large-language-models-work-91c362f5b78f>
- Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. London/New York: Routledge.

