

Journal für Schreibwissenschaft

Ausgabe 28 (3/2024), 15. Jahrgang

JoSch



Textfeedback in Praxis und Forschung – Teil II

Herausgebende
Franziska Liebetanz, Leonardo Dalesandro, Nicole Mackus, Özlem Alagöz-Bakan

Gastherausgebende
Ulrike Bohle-Jurok, Ella Grieshammer, Alina Lira Lorca

schreibwissenschaft

KI-generierte Übersetzung (MT für Machine Translation) spielt in mehrsprachigen Schreibprozessen eine große Rolle. Der Beitrag stellt Ergebnisse aus zwei Gruppendiskussionen vor, in denen mehrsprachige Studierende über Nutzungsmöglichkeiten und Limitationen von MT im Schreibprozess sprachen. In den in der Auswertung gebildeten Kategorien wird MT-Nutzung unter dem Aspekt von Rückmeldung betrachtet: Die Teilnehmenden nutzen MT-Vorschläge auf unterschiedliche Weise als Rückmeldung zu Textentwürfen, geben aber auch selbst Rückmeldung zu diesen MT-Vorschlägen. Nach Klärung des Begriffs Rückmeldung und Beschreibung des methodischen Vorgehens werden die Kategorien mit Bezug zu weiteren Forschungsergebnissen sowie didaktischen Implikationen vorgestellt.

Schlagworte: KI-Tools; KI-generierte Übersetzung; MT; Rückmeldung
Zitievorschlag: Brinkschulte, Melanie; Grieshammer, Ella (2024): Rückmeldung durch bzw. zu KI-generierte(r) Übersetzung. Eine explorative Studie. JoSch 3(24), 63-76. <https://doi.org/10.3278/JOS2403W006>

E-Journal Einzelbeitrag
von: Melanie Brinkschulte, Ella Grieshammer

Rückmeldung durch bzw. zu KI-generierte(r) Übersetzung Eine explorative Studie

aus: Ausgabe 28: Textfeedback in Praxis und Forschung - Teil II
(JOS2403W)
Erscheinungsjahr: 2025
Seiten: 63 - 76
DOI: 10.3278/JOS2403W006

Rückmeldung durch bzw. zu KI-generierte(r) Übersetzung

Eine explorative Studie

Melanie Brinkschulte & Ella Grieshammer

Abstract

KI-generierte Übersetzung (MT für Machine Translation) spielt in mehrsprachigen Schreibprozessen eine große Rolle. Der Beitrag stellt Ergebnisse aus zwei Gruppendiskussionen vor, in denen mehrsprachige Studierende über Nutzungsmöglichkeiten und Limitationen von MT im Schreibprozess sprachen. In den in der Auswertung gebildeten Kategorien wird MT-Nutzung unter dem Aspekt von Rückmeldung betrachtet: Die Teilnehmenden nutzen MT-Vorschläge auf unterschiedliche Weise als Rückmeldung zu Textentwürfen, geben aber auch selbst Rückmeldung zu diesen MT-Vorschlägen. Nach Klärung des Begriffs Rückmeldung und Beschreibung des methodischen Vorgehens werden die Kategorien mit Bezug zu weiteren Forschungsergebnissen sowie didaktischen Implikationen vorgestellt.

Einleitung

Nach der Veröffentlichung der kostenfreien Version von ChatGPT-3 im November 2022 wurde bekanntermaßen viel darüber diskutiert, was diese disruptive Entwicklung für das studentische Schreiben bedeutet (siehe z. B. Buck/Limburg 2023). Um nicht nur *über* die Studierenden zu sprechen, sondern auch *mit* ihnen, führten wir Autorinnen zwei parallele Gruppendiskussionen mit Studierenden der Interkulturellen Germanistik zum Thema KI-Nutzung beim Schreiben durch. Die am 26. Januar 2023 realisierten Gruppendiskussionen orientierten sich an einem zuvor entwickelten Leitfaden, waren aber, wie es die Methode vorsieht, thematisch relativ offen.

Gleich zu Beginn beider Diskussionsrunden wurde durch Erfahrungsberichte deutlich, dass die teilnehmenden Studierenden bis auf wenige Ausnahmen zu diesem Zeitpunkt noch keinerlei Erfahrung mit ChatGPT oder anderer textgenerierender KI hatten. Alle Teilnehmenden berichteten jedoch von umfangreichen Erfahrungen mit KI-basierter Übersetzung, im Folgenden als MT (für Machine Translation) bezeichnet. Wie sich herausstellte, nutzten alle Teilnehmenden maschinelle Übersetzungstools wie DeepL sowohl im Studium als auch in Nebenjobs regelmäßig. Dies deckt sich mit den Ergebnissen zahlreicher Studien (siehe hierzu z. B. Alm/Watanabe (2022), Lee (2023) oder Jolley/Maimone (2022)), die feststellen,

Melanie Brinkschulte & Ella Grieshammer

dass mit zunehmender Qualität der durch MT erzeugten Outputs ein Großteil aller Studierenden in mehrsprachigen Schreibkontexten regelmäßig MT nutzt.

Die Teilnehmenden an unseren Gruppendiskussionen äußerten größtenteils differenzierte Ansichten zu Potenzialen und Limitationen der Nutzung von MT-Tools beim Schreiben. Während der Zeitpunkt anscheinend zu früh war, um verwertbare Ergebnisse zur Nutzung textgenerierender KI zu erhalten, bilden die Gruppendiskussionen eine gute Grundlage, um die studentische MT-Nutzung zu explorieren. Die offen gestellte Forschungsfrage nach der Nutzung von KI-basierten Tools im Schreibprozess wurde dementsprechend modifiziert zu den Fragen: Wie nutzen Studierende MT in ihren mehrsprachigen Schreibprozessen? Inwiefern spielen hierbei Rückmeldungen – durch und zu maschinell generierte(n) Übersetzungen – eine Rolle? Die Ergebnisse werden in diesem Beitrag vorgestellt. Damit wird auch das Ziel verfolgt, die Bedeutung von MT in mehrsprachigen Schreibprozessen stärker in den Fokus der Schreibdidaktik zu rücken – wie es im Kontext von EFL Writing bereits vielfach getan wird. Da wir MT-Übersetzung im Zusammenhang mit Rückmeldung betrachten, klären wir zunächst, was wir unter dem Begriff der Rückmeldung verstehen, bevor wir unsere Methodik näher beschreiben. Auf bisherige Forschungsergebnisse zu MT in mehrsprachigen Schreibprozessen gehen wir bei der Darstellung unserer Ergebnisse ein.

Zum Verständnis von Rückmeldung im Kontext von MT

Lee (2022: 2) merkt an, dass MT für Lernende ein individualisiertes Feedback darstellen kann. Dieser Sichtweise schließen wir uns in Teilen an und betrachten an dieser Stelle das Verhältnis von MT zu Rückmeldung und Feedback in Schreibprozessen:

MT-Tools wie DeepL oder Google Translate generieren Vorschläge für zielsprachliche Formulierungen auf Basis von Eingaben durch Schreibende. Insofern geben MT-Tools eine direkt verfügbare Rückmeldung zu Entwürfen, die Schreibende in ihren Inhalten selbst erdacht und aufgeschrieben haben. Diese Rückmeldung kann sich je nach Eingabe auf Formulierungen auf der Wort-, Satz- und Textebene des eingegebenen Textentwurfs beziehen. Die Übersetzungsvorschläge können Schreibende dann in ihren Textentwurf integrieren.

Rückmeldungen durch MT sind also dadurch charakterisiert, dass sie direkt und ohne Zeitverzug verfügbar sind und zielsprachliche Vorschläge für bereits geschriebene Textteile bieten. MT-Rückmeldungen von DeepL oder Google Translate enthalten derzeit (Stand der Gruppendiskussion, aber auch Stand der Veröffentlichung) keine Erläuterungen zu den Übersetzungsvorschlägen, wie z. B. Erklärungen zu Grammatik, Pragmatik oder Registerwahl. Auch werden keine Erläuterungen zu Textsortenkonventionen oder zu Adressat*innenorientierung angefügt. Die Schreibenden müssen eigenständig – nach selbst gewählten Kriterien – entscheiden, wie sie mit den generierten Vorschlägen umgehen. Dialoge, in denen Schreibende Begründungen oder Erläuterungen zu Übersetzungsvorschlägen erhalten könnten, sind in klassischen, kostenfreien MT-Tools wie DeepL aktuell nicht vorgesehen.

In pädagogischen Verständnissen wird hingegen ein dialogischer Prozess zwischen Schreiber*in und Leser*in als wesentliches Merkmal von Feedback angesehen, um eine interaktive Auseinandersetzung mit einem Textentwurf zu ermöglichen (Hattie/Timperley 2007: 81). Mubarak (2013: 54) betont in seinem fremdsprachendidaktischen Konzept von Feedback, dass in fremdsprachigen Schreibprozessen das Feedback der Lehrkraft Vorschläge und Schritte für die weitere Textbearbeitung liefert. Insofern fehlt MT-Rückmeldungen dieser aushandelnde Diskurs zwischen eingebender (Schreiber*in) und vorschlagender (MT) Instanz. Die Interaktion zwischen Schreiber*in und MT kann eher als eine Gegenüberstellung sprachlicher Produkte ohne weitere Kommentierung oder Bewertung seitens der vorschlagenden MT charakterisiert werden. Schreibende, die MT nutzen, müssen auf Grundlage ihrer Sprach- und Schreibkompetenzen selbstständig zu einer Einschätzung der MT-Vorschläge kommen und entscheiden, ob sie die Vorschläge übernehmen, überarbeiten oder verwerfen. Somit sind (Fach-)Sprach- und Schreibkompetenzen sowie die Fähigkeit zu autonomen Entscheidungsprozessen ausschlaggebend für die weitere Verarbeitung eines Textentwurfs.

Aus lernpsychologischer Sicht unterscheiden Müller/Ditton (2014: 15) zwischen Feedback und Rückmeldung. Unter Rückmeldungen verstehen Müller/Ditton inhaltlich-sprachliche Anmerkungen, Korrekturen und Ähnliches, ohne dass die Feedbacknehmenden direkt angesprochen werden. Feedback umfasst hingegen nach diesem Verständnis, dass die Feedbacknehmenden direkt adressiert werden, Feedback also eine Reaktion auf eine Performance darstellt. Dabei sehen Müller/Ditton Rückmeldung und Feedback nicht als dissoziierte Konzepte an, sondern vielmehr als zwei Extrempunkte auf einem dynamischen Kontinuum: Auf der einen Seite steht die rein inhaltlich-sprachliche Rückmeldung auf eine Performance, auf der anderen die direkte Adressierung des Feedbacknehmenden. Entsprechend ergeben sich Übergänge und Mischformen von Rückmeldung zu Feedback.

Dem Verständnis von Rückmeldung schließen wir uns in diesem Beitrag an, um die Interaktion zwischen Schreiber*in und MT begrifflich zu fassen: MT-Tools liefern Übersetzungsvorschläge und liefern damit Hinweise zur sprachlichen Gestaltung, sprechen die Schreibenden jedoch nicht direkt an und geben keine erläuternden Hinweise.

Methodik, Datengrundlage, Teilnehmende

Wie eingangs erwähnt, war das ursprüngliche Ziel unserer Studie im Januar 2023, Studierende des Masterstudiengangs Interkulturelle Germanistik zu Einstellungen, Haltungen und Nutzungspraxis hinsichtlich textgenerierender KI wie ChatGPT zu befragen. Daher wählten wir die Erhebungsmethode der Gruppendiskussion, da diese sowohl Teilnehmenden die Gelegenheit bietet, über individuelle Erfahrungen und Einstellungen zu sprechen, als auch Einblicke in sozial verankerte Meinungen und Haltungen von Gruppen gibt (Lamnek 2010: 376 f.). Das besondere Potential von Gruppendiskussionen liegt darin, dass Gelegenheiten geschaffen werden, um Wahrnehmungsweisen, Einstellungen, Konventionen

und Muster aufzudecken, sodass auch implizite Wissensbestände und Haltungen zutage treten. Somit werden gerade durch Gruppendiskussionen kollektive Orientierungen, die das Handeln von Individuen – teilweise implizit – beeinflussen, expliziert und präsentiert. Diese individuellen Einstellungen und kollektiven Orientierungen werden nicht genuin in der Diskussionssituation gebildet, vielmehr durch die Methode der Gruppendiskussion sichtbar und somit einer Analyse zugänglich (Lamnek 2010: 385 ff.; Scherf 2018: 87). Die Gruppendiskussion ist als explorative Methode einzuordnen und entsprechend offen konzipiert (Lamnek 2005: 40 ff.). Zudem werden die Gütekriterien explorativer Forschung eingehalten. Dies bedeutet auch, dass die Teilnehmenden die Gesprächsinhalte bestimmen. In unserem Fall verschob sich der Schwerpunkt von textgenerierender KI auf Erfahrungen mit MT-Tools und den Haltungen gegenüber diesen, da dieses Thema von den Teilnehmenden als relevant gesetzt wurde. Damit einhergehend verschob sich auch der Fokus unserer Forschungsfrage.

An den zwei parallelen, jeweils ca. 50-minütigen Gruppendiskussionen am 26.01.2023 nahmen insgesamt 18 Studierende teil. Dabei handelt es sich um Studierende des ersten Fachsemesters im Masterstudiengang Interkulturelle Germanistik/Deutsch als Fremdsprache an der Universität Göttingen. Diese Studienkohorte stammt aus dem deutschen, europäischen und außereuropäischen Bildungskontext. Die Teilnehmenden zeichnen sich dadurch aus, dass sie sprachenaffin sind, mehrere Sprachen gelernt haben und diese zu einem großen Anteil auf einem höheren Sprachniveau beherrschen (mind. B1 GeR). 16 der 18 Teilnehmenden studieren auf Deutsch als weiterer Wissenschaftssprache (ca. C1 GeR), da sie ihren BA in einer anderen Sprache als Deutsch erreicht haben. Alle verfügen durch ein zuvor abgeschlossenes Studium über akademische Schreiberfahrungen, haben diese jedoch in unterschiedlichen universitären Bildungskontexten gesammelt. Einige sind oder waren als Übersetzer*innen, einige als Fremdsprachenlehrer*innen tätig. Alle Teilnehmenden haben Kenntnisse in Fremdsprachendidaktik und interkultureller Kommunikation. Insgesamt lässt sich die Gruppe als mehrsprachig und an Sprache(n) interessiert charakterisieren.

Die Gruppendiskussionen fanden im Rahmen eines Seminars zur Einführung in empirische Forschungsmethoden statt, das eine der Autorinnen dieses Beitrags leitet. Im Seminar hatten die Studierenden die Gruppendiskussion als Methode kennengelernt und wurden im Vorfeld angeregt, KI-Tools für das eigene Schreiben zu erproben und/oder über ihre Erfahrungen mit KI-Tools nachzudenken. Sie nahmen freiwillig an der Studie teil und unterzeichneten eine Einverständniserklärung. Den Auftakt zu den Gruppendiskussionen bildete eine kurze Einführung in das Thema durch die beiden Moderatorinnen (Autorinnen dieses Beitrags). Im Anschluss wurden die Teilnehmenden anhand eines offen gehaltenen Leitfadens zu ihren Erfahrungen mit KI beim Schreiben befragt, wobei die Ausgestaltung der Diskussion weitgehend bei den Teilnehmenden lag.

Die Audioaufnahmen der beiden Gruppendiskussionen wurden orientiert an der Schriftsprache transkribiert; stark von (syntaktischen oder anderen) Normen abweichende Äußerungen wurden geglättet, Namen und Bezüge pseudonymisiert. Die Transkripte wurden mittels qualitativer Inhaltsanalyse (Mayring 2022) analysiert, da diese Methode der Datenauswertung ein umfassendes und zugleich offenes Analyseinstrument für Inhalte bie-

tet, die in den Gruppendiskussionen ausgeführt wurden. Die Kategorienbildung erfolgte induktiv (Mayring 2022: 68 ff.), um die Aussagen der Teilnehmenden umfänglich erfassen und schließlich inhaltlich strukturiert bündeln zu können. Eine Interrater-Reliabilität wurde erzielt, indem wir Autorinnen zusammen mit zwei Kolleginnen unabhängig voneinander die Daten analysierten.

Ergebnisse und Diskussion

Bereits während der Gruppendiskussionen wurde deutlich, dass sich die MT-Nutzung der Teilnehmenden nicht auf ein bestimmtes Sprachenpaar reduzieren ließ. Stattdessen gaben die Teilnehmenden an, MT auf vielfältige Weise für Übersetzungen zwischen Erstsprache(n), Zielsprache(n) und weiteren verfügbaren Sprachen zu nutzen. Um welche Sprachen es sich hierbei konkret handelte, unterschied sich innerhalb der sprachlich sehr heterogenen Gruppen je nach Teilnehmenden. Teilweise wurden in den Äußerungen keine konkreten Sprachen benannt, teilweise wurde deutlich, dass es um Übersetzungen zwischen einzelnen Sprachenpaaren ging wie Chinesisch-Deutsch, Deutsch-Spanisch, Chinesisch-Englisch, Deutsch-Russisch, Arabisch-Deutsch.

Bei der Auswertung der Transkripte unter der Frage, welche Arten von Rückmeldung die Studierenden bei der MT-Nutzung thematisieren, konnten wir induktiv drei Oberkategorien bilden, zu sehen in Tabelle 1: Zum einen gaben die Teilnehmenden an, eine Übersetzung ihres Textentwurfs (hierbei handelte es sich um einzelne Wörter, Sätze und oder Abschnitte) von MT in der Zielsprache generieren zu lassen (Kategorie 1). Im Sinne des oben dargelegten Verständnisses handelt es sich bei dieser automatisch generierten Übersetzung um eine Rückmeldung, d. h. um einen inhaltlich-sprachlichen Vorschlag in der ausgewählten Zielsprache. Zu dieser Kategorie ließen sich drei Unterkategorien bilden, die jeweils fokussieren, welche weitergehenden Ziele die Teilnehmenden mit den Rückmeldungen verfolgen.

Zum anderen berichteten die Teilnehmenden, wie sie diese generierten Übersetzungsvorschläge bewerten, also von eigenen Rückmeldungen zu den Vorschlägen im Sinne von Evaluationen, wie geeignet ihnen der Vorschlag erscheint (Kategorie 2). Hier interessierte uns, welche Arten von Rückmeldung die Teilnehmenden zu den MT-Vorschlägen geben und wie sie ihre Einschätzungen begründen. Aus diesem Interesse heraus entstanden vier weitere Unterkategorien.

Einige Teilnehmende berichteten zudem von Rückmeldungen, die dem Textfeedback zuzuordnen sind: Hier ging es um Reaktionen von externen Lesenden zu mit Hilfe von MT generierten Übersetzungen (Kategorie 3). Da diese Kategorie in den Gruppendiskussionen seltener und weniger facettenreich besprochen wurde, gab es hier keine weitere Bildung von Unterkategorien.

Die folgende Tabelle stellt diese Kategorien im Detail dar:

Tabelle 2

Kategorien zur Nutzung von Machine Translation (MT) in zwei Gruppendiskussionen

Kategorien	Unterkategorien
1. MT liefert Übersetzung eines Textentwurfs als Rückmeldung. Analysefokus: Wozu nutzen die Schreibenden diese Rückmeldungen?	1a. Auswahl von Synonymen 1b. Vorschläge für zielsprachliche Fachterminologie (Ausgleich eines Wissensdefizits) 1c. Erweiterung eigenen Wissens über idiomatische Wendungen
2. Schreiber*in gibt Rückmeldung zu MT-Übersetzungsvorschlag. Analysefokus: Welche Rückmeldungen geben die Schreibenden und worauf basieren diese Rückmeldungen?	2a. Sprachliche Wissensgrundlage für Bewertung von MT-Vorschlägen 2b. Qualitätsunterschiede bei verschiedenen Sprachpaaren 2c. Qualitätsverlust bei komplexeren Sätzen und Auslassen von Informationen 2d. Fehlende Registerunterschiede
3. Weitere*er Leser*in gibt Rückmeldung zu Text. Analysefokus: Welche Rückmeldungen geben weitere Lesende?	

Im Folgenden werden die einzelnen Kategorien anhand von jeweils einem Ankerbeispiel aus den Transkripten vorgestellt, um die jeweilige Form der Rückmeldung durch bzw. zu MT zu charakterisieren. Zu jedem Einzelergebnis wird der Bezug hergestellt zu Studien, die zu ähnlichen Ergebnissen kommen bzw. zu Studienergebnissen, die für unser Ergebnis relevant sind. Außerdem wird die Aussage zu jeder Kategorie, wenn nötig, problematisiert und es werden erste schreibdidaktische Schlussfolgerungen abgeleitet.

Kategorie 1a) Auswahl von Synonymen

Ein Aspekt MT-generierter Rückmeldung, den mehrere Teilnehmende hervorheben, ist die Möglichkeit, sich für einzelne Ausdrücke mehrere Übersetzungsvorschläge anzeigen zu lassen und daraus wählen zu können. Ein Zitat der Teilnehmerin Amira verdeutlicht dies:

„Und das Gute an dieser Funktion finde ich, dass, wenn man da drauf klickt, kommen ja Synonyme. Und da suche ich mir immer das Relevanteste aus oder das, wo ich denke, ok, das ist gut. Und das ist eigentlich ganz toll, finde ich, dass man nicht immer switchen muss für Synonyme.“ (Amira, 10:30, GDE)

In Amiras Aussage wird deutlich, dass sie die Auswahlmöglichkeit, die MT liefert, besonders schätzt, gerade auch die Tatsache, dass die Auswahl mit der Übersetzung kombiniert ist, sodass sie nicht zwischen verschiedenen Nutzungsoberflächen wechseln muss. Gleichzeitig

zeigt sie an, über die Kompetenz zu verfügen, das für sie „Relevanteste“ bzw. Passendste auswählen zu können. Dass MT dabei unterstützen kann, geeigneten Wortschatz auszuwählen, berichten auch Teilnehmende in anderen Studien (siehe z. B. Lee 2022). Aus schreibdidaktischer Sicht ist hierbei relevant, dass das schnelle Nachschlagen von Synonymen in der Zielsprache das Schreiben entlasten kann. Das Suchen nach geeigneten zielsprachlichen Ausdrücken, die für Schreibende zunächst nur in der Erstsprache oder anderen Sprachen verfügbar sind, wird somit verkürzt. Kritisch daran zu sehen wiederum ist zum einen die daraus resultierende Unterbrechung des Schreibflusses durch die Möglichkeit, schnell Synonyme anzeigen lassen zu können. Diese könnte u. U. dazu führen, dass Schreibende den Wahlmöglichkeiten bezüglich einzelner Ausdrücke viel Aufmerksamkeit widmen und ranghöhere Aspekte des Textes (HOCs) aus den Augen verlieren. Zum anderen ist fraglich, ob Schreibende tatsächlich immer in der Lage sind, die adäquateste Übersetzung auszuwählen oder ob hier nicht eventuell *falsche Freunde* oder für den entsprechenden Kontext weniger geeignete Ausdrücke gewählt werden.

Da MT in der in dieser Kategorie geäußerten Verwendung wie ein klassisches Wörterbuch genutzt wird, bieten sich in der didaktischen Umsetzung Strategien der Wörterbucharbeit an. D. h. hier könnte im Rahmen von Schreibworkshops mit den Teilnehmenden erarbeitet werden, welche Zeitpunkte im individuellen Schreibprozess sich für das Nachschlagen bzw. MT-Generieren von Ausdrücken eignen, sodass inhaltliche Aspekte des Textes nicht aus dem Blickfeld geraten und der Schreibfluss möglichst wenig unterbrochen wird. Auch Strategien zur Überprüfung der Angemessenheit von Vorschlägen können hier thematisiert werden, z. B. mit Hilfe von Datenbanken, Suchmaschinen, Rückübersetzungen etc. (siehe auch Barczaitis et al. 2022).

Kategorie 1b) Vorschläge für zielsprachliche Fachterminologie (Ausgleich eines Wissensdefizits)

Eine weitere Funktion, für die MT genutzt wird, ist das Vorschlagen von Ausdrücken mit dem Fokus auf Fachterminologie. Diese Funktion nennt die Teilnehmerin Shiyan, als sie sagt:

„Damals musste ich ein Dokument vom Deutschen ins Chinesische übersetzen und in diesem Dokument geht es um medizinische Geräte. Und ehrlich gesagt habe ich keine Ahnung oder kein Hintergrundwissen in diesem Bereich. Und deshalb benutzte ich DeepL sehr, sehr oft.“ (Shiyan, 14:00, GDE)

Sie gibt also an, MT zum Ausgleich eines Wissensdefizits zu nutzen: Begriffe einer ihr fremden Fachsprache (Medizin) sind ihr unbekannt, daher nutzt sie MT. Die MT-Nutzung für fachsprachliche Terminologie kann als potentiell problematisch angesehen werden, denn zumindest 2021 war Ergebnis einer Studie, die maschinell generierte Übersetzungen deutscher Abstracts ins Englische analysierte, dass kostenfreie MT-Tools Fachbegriffe häufig nicht optimal übersetzen (Delorme Benites 2021). Da MT-Systeme aber beständig weiterlernen und ihre Entwicklung daher sehr dynamisch ist, ist zu vermuten, dass auch die Überset-

zung von Fachbegriffen kontinuierlich optimiert wird. Im Fall von Shiyan ist zudem anzumerken, dass sie im Weiteren angibt, die vorgeschlagenen chinesischen Fachbegriffe mit anderen Quellen zu überprüfen. Denn sie fährt fort: „(...), weil ich diese chinesische Terminologie nie gehört habe, und dann musste ich noch auf einer chinesischen Webseite herausfinden, welche Wörter am meisten benutzt werden.“ (Shiyan, 14:00, GDE) Insofern zeigt sie ein Problembewusstsein dafür, dass sie von MT vorgeschlagene Ausdrücke, die sie selbst noch nie gehört hat, auf ihre Verwendungshäufigkeit hin analysieren muss. Für die Schreibdidaktik bedeutet dies, dass die automatisierte Übersetzung von Fachbegriffen thematisiert werden sollte. Hier wäre zu überlegen, ob sich mit (Rück-)Übersetzungen von Fachbegriffen im Unterricht experimentieren lässt, um die Tauglichkeit von MT zu testen und eigenen Übersetzungen bzw. der Arbeit mit klassischen Wörterbüchern oder anderen Quellen gegenüberzustellen.

Kategorie 1c) Erweiterung eigenen Wissens über idiomatische Wendungen

Mehrere Teilnehmende berichten davon, dass sie MT nicht nur für die konkrete Produktion oder Überarbeitung eines akademischen Textes nutzen, sondern durch die MT-Nutzung auch ihr Wissen über idiomatische Wendungen (hier in der Regel auf das Deutsche als Zielsprache bezogen) erweitern. So sagt z. B. Tian: „Mit dieser Methode kann ich etwas lernen und einige Redewendungen oder Sprichwörter (sic!). Ich finde, es ist ganz nützlich für mich.“ (Tian, 41:00, GDE) Mit „Methode“ bezieht er sich auf die von ihm zuvor beschriebene MT-Nutzung, zunächst einen Text auf Deutsch zu schreiben, den gleichen dann auf Chinesisch zu schreiben und diesen chinesischen Text dann durch MT übersetzen zu lassen, sodass er den MT-generierten mit dem selbst geschriebenen deutschen Text vergleichen kann. Ähnlich berichten auch andere Teilnehmende, die das Deutsche als Fremdsprache gelernt haben, dass sie mit Hilfe von MT neue Ausdrücke lernen oder sogar lernen, (Grammatik-)Fehler zu vermeiden. Auf das Potential von MT für den Fremdsprachenerwerb wird auch in der Literatur häufig hingewiesen (siehe Goodwin-Jones 2022, Jolley/Marmoine 2022). In Tians Äußerung wird ein komplexes Vorgehen deutlich, das den MT-Vorschlag zur Überprüfung der eigenen zielsprachlichen Textproduktion nutzt. In dieser Strategie zeigt sich besonders, dass MT hier als Rückmeldung genutzt wird: Indem das eigene zielsprachliche Textprodukt verglichen wird mit dem MT-Text, erfährt Tian als Schreiber eine Rückmeldung zu seinem Text, die veranschaulicht, wie sein Text in einer anderen Version in der Zielsprache aussehen könnte. Ähnliche Vorgehensweisen werden auch von Tsai (2020) und Xu (2020) beschrieben. Mit Blick auf die Schreibdidaktik lässt sich für solche MT-bezogenen Strategien vermutlich analog zum Umgang mit Schreibstrategien festhalten: Strategien, die Studierende für sich entwickelt und erprobt haben, sollten zunächst einmal im Sinne eines kompetenz- und ressourcenorientierten Ansatzes gewürdigt werden. Im nächsten Schritt könnte überprüft werden, in welchen Settings oder Kontexten diese Strategien angemessen sind und ob sie sich durch zusätzliche Strategien erweitern lassen, z. B. wenn unterschiedliche Textsorten oder Texte für bestimmte Leser*innen zu verfassen sind.

Kategorie 2a) Sprachliche Wissensgrundlage für Bewertung von MT-Vorschlägen

Im Rahmen der Gruppendiskussionen wurde auch danach gefragt bzw. von den Teilnehmenden unaufgefordert diskutiert, auf welcher Grundlage sie die MT-generierten Vorschläge bewerten, wie also ihre eigene Rückmeldung im Sinne einer Evaluation des Vorschlags zustande kommt. In ihren Äußerungen beschreiben die Teilnehmenden, mit Hilfe welcher sprachlichen Kompetenzen sie die Angemessenheit einer MT-Übersetzung beurteilen. Jessie (31:40, GDM) sagt dazu: „Zum Beispiel hier brauche ich einfach ein Funktionsverb, aber bei dem ist ein langer Satz. Mit meinem höheren Sprachniveau kann ich die Fehler dieser KI erkennen und korrigieren.“ Was Jessie hier äußert, sagen in ähnlicher Form auch weitere Teilnehmende: dass die KI durchaus Fehler bei der Übersetzung macht, dass die Studierenden ihr Kompetenzniveau jedoch so hoch einschätzen, dass sie diese Fehler erkennen und beheben können. Dass diese Einschätzung zutreffen könnte, scheint eine Studie von Kol/Schcolnik/Spector-Cohen (2018) nahezulegen: Die Autor*innen fanden heraus, dass Studierende bereits mit B2-Sprachniveau viele Fehler maschineller Übersetzung erkennen können. Didaktisch ließe sich dies nutzen, indem Studierende in mehrsprachigen Schreibprozessen aufgefordert werden könnten, MT zu verwenden und Übersetzungsvorschläge sowie ihre Entscheidungen über die Vorschläge zu notieren, etwa im Sinne eines Integrated Problem Decision Reports (Heine 2019). Hierdurch könnten die Studierenden sich bewusster machen, wie sie mit MT-Vorschlägen arbeiten und welche Kriterien sie für Entscheidungen bezüglich deren Annahme, Veränderung oder Ablehnung heranziehen.

Kategorie 2b) Qualitätsunterschiede bei verschiedenen Sprachenpaaren

Durch die sprachlich heterogene Zusammensetzung der beiden Diskussionsgruppen kam es zu zahlreichen Vergleichen der Übersetzungsqualität bei verschiedenen Sprachenpaaren. Da die Teilnehmenden hauptsächlich über DeepL sprachen, stellten sie fest, dass es sich hierbei um ein in Deutschland entwickeltes Tool handelt, das vor allem auf die Übersetzung zwischen Deutsch und anderen Sprachen fokussiert sei. Jinjin sagt z. B.: „Ich habe auch mit DeepL übersetzt von Deutsch auf Chinesisch. Das funktioniert gut. Aber als ich von Japanisch auf Chinesisch übersetzt habe, funktionierte das nicht so gut.“ (10:25, GDE) Dass MT bei verschiedenen Sprachenpaaren – auch unabhängig vom sprachlich-kulturellen Sitz des Software-Herstellers – unterschiedlich gut funktioniert, gilt als relativ belegt: Verschiedene Sprachenpaare erzeugen unterschiedliche Übersetzungsqualität und unterschiedliche Typen von Fehlern (Lee 2022). Die Lebendigkeit, mit der auch sonst im Seminarkontext eher zurückhaltende Teilnehmende über ihre Erfahrungen mit MT bei verschiedenen Sprachenpaaren berichteten, lässt darauf schließen, dass dieses Thema zumindest bei sprachaffinen Studierenden auf Interesse stößt und ein Erinnern an eigene Erfahrungen auslöst. Didaktisch ließe sich dies nutzen, um in Lehrveranstaltungen oder Beratungen zum mehrsprachigen Schreiben die Leistungsfähigkeit von MT bei verschiedenen Sprachenpaaren zu diskutieren und in Experimenten mit den Schreibenden vertrauten Sprachen zu testen. Bei weniger sprachaffinen

Studierenden wäre vermutlich eher das Ziel, ein gesundes Misstrauen gegenüber den Fähigkeiten von MT-Tools zu entwickeln und Kriterien zu erarbeiten, mit denen sich MT-Vorschläge evaluieren lassen.

Kategorie 2c) Qualitätsverlust bei komplexeren Sätzen und Auslassen von Informationen

Bei den Rückmeldungen zu MT-Generaten fällt auf, dass einzelne Teilnehmende eine Beobachtung äußern, die auch in der Literatur beschrieben wird: Bei längeren Sätzen nimmt die Übersetzungsqualität ab (Bowker/Ciro 2019, Koehn/Knowles 2017, zit. n. Lee/Briggs 2021 sowie Park 2017, Kim 2019, zit. n. Lee 2022). Feng (16:40, GDE) sagt hierzu: „(...) weil es viele logische Fehler gibt und sie können nur einfache Sätze, kurze Sätze erkennen.“ Interessanterweise verweist Feng an anderer Stelle auf ein Phänomen, das auch Delorme Benites (2021) in ihrer Analyse MT-generierter Übersetzung deutscher Abstracts ins Englische feststellt: Bisweilen werden in MT-Übersetzungen ganze Sätze ausgelassen (zumindest war dies 2021 noch der Fall, für 2024 liegen unseres Wissens noch keine validen Untersuchungen vor), wenn auch die Ursache hierfür unklar ist. Feng stellt fest: „Aber wenn dieser Satz ein langer Satz ist oder wenn es ein Abschnitt ist, dann werden einige Sätze darin übersehen und ignoriert.“ (Feng, 16:40, GDE) Hierin offenbaren sich also eindeutig Limitationen von MT – auch wenn diese möglicherweise in Zukunft behoben werden können. Diese Limitation spricht dafür, MT ausschließlich zur Übersetzung einzelner Sätze oder Ausdrücke zu nutzen, was auch der in den Gruppendiskussionen angesprochenen sowie in weiterer Forschung erhobenen MT-Nutzung entspricht (siehe Überblick von Jolley/Maimone 2022). Als schreibdidaktische Implikation wäre hier zu nennen, dass aufgrund der dynamischen Entwicklung Studierende wie auch Lehrende über aktuelle Potenziale und Limitationen KI-basierter Tools informiert sein, diese immer wieder selbst testen und ggf. an Fortbildungen zu KI und Schreiben teilnehmen sollten. In Workshops zum mehrsprachigen Schreiben könnte die aktuelle Leistungsfähigkeit diverser Tools bei verschiedenen Sprachenpaaren und Fachthemen erprobt werden, sodass die Teilnehmenden Schlussfolgerungen für den Einsatz in ihren Schreibprozessen ziehen.

Kategorie 2d) Fehlende Registerunterscheidung

Eine Limitation, die mehrere Diskussionsteilnehmende nennen, ist die teilweise fehlende Registerunterscheidung in MT-Generaten: MT-Tools wie DeepL verwenden bisweilen lexikalisch und syntaktisch innerhalb eines Vorschlags keine einheitliche Register- bzw. Stilebene, wie diese Aussage von Darja (33:00, GDE) verdeutlicht: „Manchmal klingt es so komisch. Manchmal ist es die Verwendung der Wörter, z. B. wenn diese formellen und informellen Verben z. B. zusammengestellt sind und es ist so ein bisschen- es klingt nicht so gut.“ Diese Beobachtung, dass MT bisweilen nicht lexikalisch angemessen ist und Kontext sowie kulturelle Hintergründe und damit verbundene Konnotationen nicht einbezieht, findet sich auch in der Literatur zu MT wieder (Lee 2022, bezugnehmend auf Briggs 2018, Ducar/Schocket 2018). Mehr noch als bei anderen Limitationen wird hier deutlich, dass insbesondere eine

umfassende pragmatische Kompetenz in der Zielsprache nötig ist, um MT-Vorschläge beurteilen zu können (siehe auch Kategorie 2a). Gleichzeitig wird diese Limitation vermutlich in Zukunft reduziert oder sogar eliminiert: Kostenlos verfügbare KI-basierte Tools wie ChatGPT, die unter anderem übersetzen können, weisen durch die dialogische Interaktion zwischen Mensch und KI bereits die Möglichkeit auf, die Ausgabe in einem bestimmten Stil bzw. in einem bestimmten Kontext oder einer bestimmten Rolle zu verlangen. Allerdings bleibt die Notwendigkeit bestehen, dass die Schreibenden überprüfen, ob ein vorgeschlagenes Register wirklich in den Kontext des geschriebenen Textes passt. Dieses leitet über zu einer weiteren didaktischen Implikation: Schreibende sollten lernen, kritisch mit Registervorschlägen umzugehen und diese analytisch zu prüfen, indem sie beispielsweise deren Verwendung in anderen authentischen Zusammenhängen recherchieren.

Eine andere Möglichkeit besteht darin, dass Studierende dazu angeleitet werden könnten, Texte in unterschiedlichen Stilen, wie sie Tools wie z. B. DeepL Write ausgeben, auf die syntaktische und lexikalische Umsetzung des Stils hin zu analysieren. Somit ließe sich KI nutzen, um ein stärkeres Bewusstsein für verschiedene Register und Stile zu formen.

Kategorie 3) Weitere*r Leser*in gibt Rückmeldung zu Text

Einige Diskussionsteilnehmende thematisieren die Rückmeldung externer Leser*innen zu MT-erzeugten Texten. So berichtet Jessie (10:00, GDM):

„Vor zwei Jahren habe ich als Übersetzerin gearbeitet, nämlich ich musste chinesische Texte in Englische übersetzen oder umgekehrt. Und wenn ich DeepL benutzen konnte, bekam ich meine Arbeit sehr schnell fertig. Aber mein Boss war nicht immer zufrieden mit meiner Arbeit und ich habe mich erinnert, dass mein Boss einmal sagte: Du musst diese App nicht so häufig benutzen, weil die Texte nicht so authentisch sind.“

Anders als in den bisherigen Kategorien wird in diesem Fall erstmals angesprochen, dass ein MT-Text Feedback von einem adressierten Leser („mein Boss“) erhält. Das Feedback ist hier negativ und betont die größere Authentizität einer genuin menschlich bearbeiteten Übersetzung. In Jessies Äußerung schwingt zudem das Argument der Arbeitsersparnis durch den Einsatz von MT mit, das für professionell arbeitende Übersetzer*innen eine große Rolle spielt, indem zeitlicher Aufwand und Qualität einer Übersetzung zueinander in Beziehung gesetzt werden. Auch diese Erfahrung von Jessie, die aus einem zurückliegenden Jahr stammt, ist innerhalb ihres zeitlichen Rahmens zu sehen und vor dem Hintergrund der Tatsache, dass sich MT stetig weiterentwickelt und auch kostenfreie Tools immer besser werden. Im Rahmen der Gruppendiskussionen scheinen diese Anekdoten, die die Outputqualität der MT in Frage stellen, vor allem die Funktion zu haben, die in beiden Gruppen vorherrschende Meinung zu bestätigen: MT kann zwar hilfreich sein, jedoch sind kompetente Sprachnutzer*innen wie die Teilnehmenden ihr überlegen.

Fazit

Betrachtet man die Eigenschaft von Gruppendiskussionen, Meinungsbildungsprozesse abilden zu können, dann können wir als Konsens beider Diskussionsgruppen festhalten: Die Teilnehmenden sind davon überzeugt, dass MT – in Abhängigkeit vom Sprachenpaar – für fortgeschrittene Sprachnutzer*innen eine hilfreiche Unterstützung beim Schreiben darstellt, weil MT unmittelbare Rückmeldungen gibt. Gleichzeitig betonen unsere Teilnehmenden, dass es eine recht hohe Sprachkompetenz braucht, um selbst wiederum kompetente Entscheidungen zu den Vorschlägen geben zu können. Insgesamt herrscht die Meinung vor, den MT-Tools bzgl. der Sprachverwendung, wie z. B. bei der Registerwahl, überlegen zu sein.

Die Gruppendiskussionen haben zum einen bestätigt, was auch die Literatur zu MT zeigt: dass MT aus mehrsprachigen Schreibprozessen Studierender nicht mehr wegzudenken ist. Die lebhaften Diskussionen der Teilnehmenden über das Thema ließen zum anderen darauf schließen, dass gerade für an Sprache(n) interessierte Studierende das Thema KI und ihr Einsatz beim mehrsprachigen Schreiben ein relevantes Thema zu sein scheint, dass sie aber eher selten zum Sprechen darüber aufgefordert werden. Die Teilnehmenden verdeutlichen zudem durch ihre Aussagen über die weitergehende Verwendung und Modifizierung der durch MT generierten Vorschläge, dass sie über vielfältige Strategien verfügen, die sie zielorientiert einsetzen. Interessant wäre es, auch aus Sicht der Translationswissenschaften den Einsatz von MT zu analysieren, um mehr über aktivierte Übersetzungsstrategien der MT-Nutzenden zu erfahren.

Das Thema MT-Nutzung in mehrsprachigen Schreibprozessen ist dafür, dass es diese Tools in kostenlosen, online zugänglichen Varianten schon seit langem gibt, im deutschsprachigen schreibdidaktischen Diskurs eher unterrepräsentiert. Hingegen findet sich im internationalen Raum ein breit gefächerter Diskurs zum Thema mehrsprachiges Schreiben und MT-Nutzung, der hier nur in knappen Auszügen wiedergegeben werden konnte. Unserer Erfahrung nach findet die Frage nach MT-Übersetzung in mehrsprachigen Schreibprozessen auch noch eher selten Eingang in schreibdidaktische Formate wie Schreibworkshops und Schreibberatung.

Insgesamt sprechen die Ergebnisse dieser explorativen Studie dafür, MT-Nutzung im Rahmen einer translingualen Schreibdidaktik (siehe auch Barczaitis et al. 2022) ernst zu nehmen, indem MT mit ihrem Potential für die Ausgestaltung von Schreibprozessen expliziert wird. Hierzu gehören sowohl ein Sichtbarmachen und ein Austausch über bereits verwendete Strategien beim Einsatz von MT als auch die Anleitung zur Nutzung weiterer verfügbarer Sprache(n) im Schreibprozess (siehe auch Zhou et al. 2022).

Literatur

- Alm, Antonie/Watanabe, Yuki (2022): Online Machine Translation for L2 Writing Across Languages and Proficiency Levels. In: *Australian Journal of Applied Linguistics*. Jg. 5. Nr. 3. 135–157. DOI: <https://doi.org/10.29140/ajal.v5n3.53si3>
- Barczaitis, Irina/Brinkschulte, Melanie/Grieshammer, Ella/Stoian, Monica (2022): *Mehr-sprachiges Schreiben im akademischen Kontext unterstützen, anleiten, begleiten*. Bielefeld: wbv media und UTB. <https://doi.org/10.36198/9783838558011>
- Briggs, Neil (2018): Neural Machine Translation Tools in the Language Learning Classroom: Students' Use, Perceptions, and Analyses. In: *The JALT CALL Journal*. Jg. 14. Nr. 1. 3–24. DOI: <https://doi.org/10.29140/jaltcall.v14n1.221>
- Buck, Isabella/Limburg, Anika (2023): Hochschulbildung vor dem Hintergrund von Natural Language Processing (KI-Schreibtools). Ein Framework für eine zukunftsfähige Lehr- und Prüfungspraxis. In: *die hochschullehre*. Jg. 9. Nr. 6. 70–84. <https://doi.org/10.3278/HSL2306W>
- Delorme Benites, Alice (2021): *Maschinelle Übersetzung für akademische Texte. Hinweise und Empfehlungen*. DOI: <https://doi.org/10.21256/zhaw-22301>. URL: <https://digital-collection.zhaw.ch/handle/11475/22301> (Zugriff: 08.04.2024)
- Goodwin-Jones, Robert (2022): Partnering with AI: Intelligent Writing Assistance and Instructed Language Learning. In: *Language Learning & Technology*. Jg. 26. Nr. 2. 5–24.
- Hattie, John/Timperley, Helen (2007): The Power of Feedback. In: *Review of Educational Research*. Jg. 77. Nr. 1. 81–112. <https://doi.org/10.3102/003465430298487>
- Heine, Carmen (2019): Integrated Problem Decision Reports in Schreibdidaktik und Beratung. In: *Journal für Schreibwissenschaft*. Jg. 17. Nr. 1. 48–55. <https://doi.org/10.3278/JOSI90IW048>
- Jolley, Jason R./Maimone, Luciane (2022): Thirty Years of Machine Translation in Language Teaching and Learning: A Review of the Literature. In: *L2 Journal*. Jg. 14. Nr. 1. 26–44. <https://doi.org/10.4995/eurocall.2018.10140>
- Kol, Sara/Schcolnik, Miriam/Spector-Cohen, Elana (2018): Google Translate in Academic Writing Courses? In: *The EUROCALL Review*. Jg. 26. Nr. 2.
- Lamnek, Siegfried (2010): *Qualitative Sozialforschung*. Weinheim: Beltz.
- Lamnek, Siegfried (2005): *Gruppendiskussion: Theorie und Praxis*. Weinheim: Beltz.
- Lee, Sangmin-Michelle (2022): An investigation of machine translation output quality and the influencing factors of source texts. In: *ReCALL*, Jg. 34. Nr. 1. 81–94. DOI: <https://doi.org/10.1017/S0958344021000124>
- Lee, Sangmin-Michelle/Briggs, Neil (2021): Effects of Using Machine Translation to Mediate the Revision Process of Korean University Students' Academic Writing. In: *ReCALL*. Jg. 33. Nr. 1. 18–33. <https://doi.org/10.1017/S0958344020000191>
- Mayring, Philipp (2022): *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim: Beltz. https://doi.org/10.1007/978-3-658-37985-8_43
- Mubarak, Mohamed (2013): *Corrective feedback in L2 writing: A study of practices and effectiveness in the Bahrain context*. University of Sheffield.

- Müller, Andreas/Ditton, Hartmut (2014): Feedback: Begriff, Formen und Funktionen. In: Ditton, Hartmut/Müller, Andreas (Hrsg.): *Feedback und Rückmeldungen. Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. Münster: Waxmann, II–28.
- Scherf, Daniel (2018): Gruppendiskussionen. Ein Verfahren zur Erhebung kollektiver Orientierungen. In: Boelmann, Jan (Hrsg.): *Empirische Forschung in der Deutschdidaktik. Band 2: Erhebungs- und Auswertungsverfahren*. Baltmannsweiler: Schneider Hohengehren. 81–98.
- Tsai, Shu-Chiao (2020): Chinese Students' Perceptions of Using Google Translate as a Translingual CALL Tool in EFL Writing. In: *Computer Assisted Language Learning*. Jg. 35. Nr. 5–6. 1–23. <https://doi.org/10.1080/09588221.2020.1799412>
- Xu, Jun (2020): Machine Translation for Editing Compositions in a Chinese Language Class: Task Design and Student Beliefs. *Journal of Technology and Chinese Language Teaching*. Jg. II. Nr. 1. 1–18.
- Zhou, Shu/Zhao, Shuo/Groves, Michael (2022): Towards a Digital Bilingualism? Students' Use of Machine Translation in International Higher Education. In: *Journal of English for Academic Purposes*. Jg. 60. 1–13. <https://doi.org/10.1016/j.jeap.2022.101193>

Autor*innen

Melanie Brinkschulte, Dr. phil., ist Schreibwissenschaftlerin, und leitet den Schlüsselkompetenzbereich Interkulturelle Interaktionen der Georg-August-Universität Göttingen.

Ella Grieshamer, Dr. phil., leitet das Internationale Schreiblabor der Universität Göttingen und ist zurzeit 1. Vorsitzende der gefsus. Zu ihren (Forschungs-)Interessen zählen das Sprechen über Schreiben und Texte, mehrsprachige Schreibprozesse, theoretische Grundlagen der (Schreib-)Beratung sowie Schreiben mit KI-Unterstützung.