universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Post-Editing durch Studierende: Eine Analyse von temporalem, technischem und kognitivem Aufwand

verfasst von / submitted by

Paul Weilguny, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 070 331 342
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Translation Deutsch Englisch
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Dagmar Gromann, BSc

Danksagung

Nach den vergangenen sieben Jahren, die ich am Zentrum für Translationswissenschaft im Rahmen des Bachelor- und Masterstudiums verbracht habe, kann ich nun nachvollziehen, warum Studieren als die schönste Zeit im Leben eines jungen Menschen bezeichnet wird. Während dieser Jahre war das Zentrum für mich nicht nur eine Ausbildungsstätte, sondern auch ein Ort an dem unzählige Freundschaften geschlossen wurden, an dem sich Studierende für ein Miteinander einsetzten und an dem auch nie Konkurrenzdenken herrschte. Aus all diesen Gründen wird mir dieser Lebensabschnitt stets in angenehmer Erinnerung bleiben.

Ein besonderer Dank gilt auch meiner Familie, die mir das Studium durch ihre Unterstützung ermöglicht hat, sowie meinem langjährigen Studienkollegen Mario, der selbst in der stressigsten Prüfungswoche nie um einen Scherz verlegen war.

Weiters möchte ich mich bei Mag. Trisha Kovacic-Young bedanken, die es mir im Rahmen eines Praktikums unter anderem ermöglicht hat, ein unternehmensinternes, maschinelles Übersetzungssystem selbstständig zu trainieren und zu betreuen. Durch ihren Enthusiasmus für diese Technologie und meine Erfahrungen aus erster Hand entstand letztendlich die Idee zum Thema für diese Masterarbeit.

Ebenso möchte ich mich bei Univ.-Prof. Dragos Ioan Ciobanu, PhD und der HAITrans Forschungsgruppe bedanken, die mir das Eyetracking-Labor am ZTW unmittelbar nach dessen Fertigstellung für meine wissenschaftliche Studie zu Verfügung gestellt haben.

Einen besonderen Dank möchte ich auch an die drei Teilnehmer*innen meiner Studie aussprechen, die mir trotz Verzögerungen und Unvorhersehbarkeiten ihre wertvolle Zeit geschenkt haben, um einen kleinen, aber wichtigen Beitrag zur Forschung zu leisten.

Zuletzt möchte ich mich bei Ass.-Prof. Mag. Dr. Dagmar Gromann, BSc für die Unterstützung beim Ausformulieren meines Wunschthemas bedanken sowie für ihre stets gründliche und fachkundige Betreuung, wodurch sich das Verfassen dieser Arbeit nicht nur unkompliziert gestaltete, sondern auch mit Freude verbunden war.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Als in den 1990er Jahren Firmen wie Microsoft und IBM ihre eigenen maschinellen Übersetzungssysteme entwickelten, erkannte man erstmals den Bedarf an Personen, die diese maschinellen Übersetzungen editieren müssen, um einen Zieltext von zufriedenstellender Qualität zu erhalten. Über das Post-Editing als eigene Disziplin war zu diesem Zeitpunkt wenig bekannt, weshalb die Forschung sich intensiver mit diesem aufsteigenden Aspekt der Sprachindustrie zu beschäftigen begann. Parallel mit der Forschung zur maschinellen Übersetzung (MÜ) (z.B. Canfora & Ottmann 2020; Moorkens 2017; Sin-Wai 2019) wurde nun auch Post-Editing (PE) in der Translationswissenschaft erforscht (z.B. Alves et al. 2016; Guerberof Arenas 2020; O'Brien 2011). Vor allem in den 2010er Jahren schien das Forschungsinteresse erstmals besonders groß zu sein, und folglich wurden Arbeiten zu verschiedensten Aspekten des PE publiziert. Guerberof Arenas (2014a) beschäftigte sich etwa mit dem Produktivitätszuwachs, den PE selbst unerfahrenen Anwender*innen beschaffen kann. Herbig et al. (2019) führten eine multivariate Analyse durch, um den Aufwand eines PE-Auftrags schon im Vorhinein abschätzen zu können, und de Faria Pires (2020) erforschte die Leistung von Studierenden beim PE und befragte sie anschließend zu ihrer Einstellung. Dieses intensive, aktuelle Forschungsinteresse kann als Indiz für die Wichtigkeit der Beschäftigung mit Post-Editing gesehen werden, die rasante Entwicklung stellt die Forschung jedoch auch vor folgende Frage: Wie eruiert man, ob Post-Editing für einen vorliegenden, maschinell übersetzten Text eingesetzt werden soll? Diese Frage bezieht sich einerseits auf die Qualität und Brauchbarkeit der maschinellen Übersetzung, und andererseits auf den Aufwand, der für Post-Editor*innen mit der Aufgabe verbunden sein wird. Ein Modell, dass Einschätzungen zum temporalen, technischen und kognitiven Aufwand einer PE-Aufgabe liefern kann, würde es ermöglichen, Post-Editing noch effizienter einsetzen zu können. Zusätzlich würden diese Erkenntnisse beim Erstellen von Post-Editing-spezifischen Curricula helfen, mithilfe welcher eine neue Generation an Post-Editor*innen ausgebildet werden könnte.

Im theoretischen Teil dieser Arbeit sollen also zuerst Entwicklung, Funktionsweise und aktuelle Stellung der maschinellen Übersetzung in Sprachindustrie und Translationswissenschaft erläutert werden sowie ein Überblick über Arten, Ablauf und das Kompetenzprofil von Post-Editing gegeben werden. In diesem Kapitel soll auch gesondert auf Post-Editing im professionellen Kontext, also durch Übersetzer*innen sowie im nicht-professionellen Kontext, also durch Studierende beziehungsweise angehende Übersetzer*innen, eingegangen werden und die unterschiedlichen Einstellungen der zwei

Gruppen beschrieben werden. In zwei weiteren Theoriekapiteln sollen zum einen der translatorische Qualitätsbegriff und diverse Metriken zur Evaluierung der Qualität von maschinellen Übersetzungssystemen behandelt werden, und zum anderen die Funktionsweise von Eyetrackern, da diese Kapitel auch die für den empirischen Teil gewählten Erhebungstechniken theoretisch erklären. In diesem sollen abschließend die Ergebnisse einer Post-Editing Studie mit drei Teilnehmer*innen, Masterstudierenden am Zentrum für Translationswissenschaft, präsentiert werden. Die Leistung der Teilnehmer*innen beim Post-Editieren einer maschinellen Übersetzung wurde einerseits hinsichtlich des temporalen, technischen und kognitiven Aufwands durch Eyetracking und weitere Methoden gemessen, zum anderen wurden die Studierenden auch interviewt, um einen genaueren Einblick in ihre Einstellungen gegenüber Post-Editing zu erhalten und Auffälligkeiten in den Messergebnissen besser nachvollziehen zu können.

1.1 Motivation und Zielsetzung der Arbeit

In der Sprachindustrie lässt sich beobachten, dass professionelle Übersetzer*innen Post-Editing skeptisch gegenüberstehen und derartige Aufträge nur zögerlich annehmen. Diesbezüglich wurden bereits einige Studien durchgeführt, etwa von Aranberri et al. (2014) oder Gaspari et al. (2015), die zu den Schlüssen kamen, dass professionelle Übersetzer*innen PE und MÜ einerseits als Konkurrenz betrachten und lieber selbst übersetzen, da die damit verbundene Bezahlung höher ist. Außerdem stellten sie fest, dass Übersetzer*innen Post-Editing trotz des gemessenen Produktivitätsgewinns als langsamer und störender empfinden und das Humanübersetzen bevorzugen. Nachdem die Bereitschaft der etablierten Übersetzer*innen mit PE zu arbeiten gering zu sein scheint, fehlt es vermehrt an Post-Editor*innen. Somit stellt sich die Frage, ob eine neue Generation an Übersetzer*innen, etwa Studierende der Translationswissenschaft, bereit und offen für Post-Editing sind. Hinzu kommt, dass die Qualitätsanforderungen von Kund*innen in manchen Fachbereichen niedriger geworden sind. Es gibt daher eine steigende Anzahl an Übersetzungsaufträgen, die mit maschineller Übersetzung und anschließendem PE durch Studierende in einem zufriedenstellenden Maß erledigt werden könnten (Yamada 2015).

Jedoch verfügen viele Studierende nicht über die nötigen Post-Editing-Kompetenzen und -Strategien. Die Notwendigkeit von PE-Kursen an Hochschulen wurde bereits von Guerberof Arenas und Moorkens (2019) thematisiert, dennoch ist auf europäischer Ebene nur eine sehr langsame Entwicklung und Anpassung der Curricula zu beobachten. Auch die EMT-Expert*innengruppe weist darauf hin, dass Studierende die Grundlagen der MÜ kennen sollen, um maschinell übersetzte Texte angemessen post-editieren zu können (EMT 2017: 8).

Die Ergebnisse dieser Arbeit und der damit verbundene Erkenntnisgewinn über die Defizite, die bei Studierenden während des Post-Editings bestehen, können in Verbindung mit bereits publizierten Forschungsarbeiten einen wichtigen Beitrag zur Erstellung eines praxisorientierten Curriculums leisten. Falls die Zukunft des Post-Editings in den Händen der Studierenden liegen sollte, wäre es von Bedeutung die Forschung auch auf diese Gruppe zu fokussieren und für sie eine praxisnahe Ausbildung zu entwickeln.

1.2 Forschungsfrage und Annahmen

Die Forschungsfrage dieser Arbeit soll primär auf das Erfassen des mit Post-Editing verbundenen Aufwands abzielen. Da Post-Editing und Übersetzen als zwei unterschiedliche Arbeitsweisen einzustufen sind, die sich durch ihre Abläufe und kognitiven Anforderungen an die arbeitende Person unterscheiden, soll erforscht werden, welcher Aspekt bei PE mit dem meisten Aufwand verbunden ist. Anders als beim Übersetzen, bei dem ein Zieltext durch mehrmaliges Lesen und Dekodieren des Ausgangstextes erzeugt wird, ist der Zieltext beim Post-Editing bereits vorhanden und kann mittels diverser Bearbeitungsstrategien auf das angestrebte Qualitätsniveau gebracht werden. Wie dieser Prozess des Editierens bei Studierenden ohne Post-Editing-Ausbildung abläuft und welche temporalen, technischen und kognitiven Aufwände dabei entstehen, soll in dieser Arbeit erforscht werden. Die Forschungsfrage lautet also: *Welcher Aspekt von Post-Editing ist für Studierende nachweislich mit dem höchsten temporalen, technischen und kognitiven Aufwand verbunden?*

Zudem sollen im Rahmen dieser Arbeit folgende Unterfragen beantwortet werden:

1. *Welche Einstellung haben Studierende zu Post-Editing?*
2. *Welche Strategien wenden Studierende beim Durchführen von Post-Editing an?*
3. *Welche Wichtigkeit schreiben Studierende einer Post-Editing-Ausbildung an Hochschulen zu?*

Die mit den Fragen verbundenen Annahmen leiten sich teilweise aus den Ergebnissen anderer Forschungsarbeiten mit ähnlichen Schwerpunkten ab, auf denen die vorliegende Arbeit aufbauen soll.

Annahme 1: *Die durch den Eyetracker gemessenen Fixationen der Studierenden werden häufiger auf den Zieltext als auf den Ausgangstext gerichtet sein.*

Annahme 2: *Die Studierenden werden trotz ihrer fehlenden Post-Editing-Ausbildung ein relativ hohes Maß an Produktivität erzielen können.*

Annahme 3: *Die Studierenden werden dazu tendieren, übermäßig viele Änderungen an der maschinellen Übersetzung vorzunehmen.*

2 Theoretische Grundlagen

In diesem ersten Kapitel soll die theoretische Grundlage für den darauffolgenden empirischen Teil geschaffen werden, indem die beiden, für diese Arbeit zentralen Themenbereiche der maschinellen Übersetzung und des Post-Editings ausführlich erläutert werden. Dabei soll neben einer geschichtlichen Abhandlung vor allem der aktuelle Stand dieser Disziplinen beschrieben werden. Gewissermaßen herrscht eine Interdependenz zwischen der maschinellen Übersetzung und dem Post-Editing, da beide Disziplinen in manchen Bereichen nicht ohneeinander auskommen. Durch das vorherrschende Misstrauen und die Ablehnung von Übersetzer*innen gegenüber der maschinellen Übersetzung und dem Post-Editing (do Carmo & Moorkens 2020; Vidal et al. 2020) soll in diesem Kapitel auch auf die Einstellung von angehenden Übersetzer*innen in der Ausbildung eingegangen werden und die Frage behandelt werden, ob diese die kommende Generation der Post-Editor*innen darstellen könnten. Darüber hinaus sollen in diesem Kapitel die Funktionsweise von Eyetrackern und die Messbarkeit von Produktivität und temporalem, technischem und kognitivem Aufwand in Bezug auf Post-Editing beschrieben werden.

2.1 Maschinelle Übersetzung

Maschinelle Übersetzung, folgend als MÜ bezeichnet, soll in dieser Arbeit als die automatische Übersetzung eines Textes aus einer menschenlesbaren Sprache in eine andere verstanden werden (Kenny 2019: 428). Seit ihren Anfängen in der zweiten Hälfte des 20. Jahrhunderts hat sich die MÜ mittlerweile zu einem mächtigen Werkzeug entwickelt, das vor allem in der Translationswissenschaft einen der zentralsten Forschungsgegenstände darstellt. Zum einen sind Arbeitsprozesse, die MÜ beinhalten, bereits so stark an Technologie gebunden, dass auch die Personen, die daran arbeiten kaum noch als davon unabhängig gesehen werden können. Die *Human-Computer-Interaction*, die in Bezug auf die Translationswissenschaft u.a. von O'Brien (2012) beschrieben wird, ist bereits Arbeitsalltag der modernen Übersetzer*innen geworden, und so verschwimmen auch die Grenzen zwischen maschineller Übersetzung und computergestützter Humanübersetzung immer weiter. Allerdings ist hier eine Unterscheidung wichtig, da nach wie vor lediglich 7% aller Übersetzungsaufträge in der Sprachindustrie rein maschinell erledigt werden und diese Arbeitsweise auf Grund ihrer rezenten Entwicklung in den Köpfen vieler Personen als nicht vertrauenswürdig gilt (Poibeau 2017: 97). Im Rahmen dieser Arbeit soll also zwischen Humanübersetzung und maschineller Übersetzung unterschieden werden. Erstere ist ein Prozess, in dem ein Mensch unter Benutzung jeglicher ihm verfügbarer technischer

Hilfsmittel und Ressourcen jene Schritte vornimmt, die für die Umwandlung eines ausgangssprachlichen Textes in einen zielsprachlichen Text notwendig sind. Bei der maschinellen Übersetzung wird diese Umwandlung von einer Maschine durchgeführt, auch wenn die Funktionsweise dieser Maschine durch menschliche Arbeit erzielt wurde (Kenny 2019: 429). In diesem Kapitel soll also auf die Entwicklung der maschinellen Übersetzung im 20. und 21. Jahrhundert eingegangen werden sowie ein Überblick über die historisch und aktuell relevanten MÜ-Architekturen präsentiert werden. Weiters sollen die aktuellen Anwendungsbereiche sowie die Grenzen von gängigen MÜ-Systemen beschrieben werden und wie diese sich auf den Markt der Sprachindustrie und den Übersetzerberuf auswirken. Zuletzt soll ein Ausblick über mögliche Entwicklungen und die Zukunft der MÜ in den nächsten Jahren gegeben werden und geklärt werden, welche Folgen dies für Übersetzer*innen haben könnte.

2.1.1 Geschichte

Angelehnt an Poibeau (2017: 24f) lässt sich die Entwicklung der MÜ im 20. und 21. Jahrhundert in sechs Phasen unterteilen, deren ausschlaggebendste Errungenschaften und Durchbrüche in diesem Kapitel grob umrissen werden sollen.

Phase 1: 1930er bis 1949

Auch wenn sich die ersten Ideen automatisierter Übersetzungssysteme bis ins 17. Jahrhundert zurückverfolgen lassen, so findet sich der erste konkrete Vorschlag einer MÜ in Form zweier Patente aus den 1930er Jahren. Der Franzose Georges Artsrouni und der Russe Petr Trojanskij entwickelten unabhängig voneinander das Konzept für ein elektromechanisches Gerät, das in einem dreistufigen Prozess als Übersetzungswörterbuch fungiert (Kenny 2019: 429). Laut Trojanskijs Konzept sollte ein/e einsprachige/r, menschliche/r Bearbeiter*in demnach den Ausgangstext mittels eines universalen Schemas in das System einarbeiten, wodurch alle grammatischen Funktionen der Wörter erfasst werden. Der/die Bearbeiter*in sollte in einem zweiten Schritt alle Wörter des Ausgangstextes in dem Übersetzungswörterbuch finden und zuordnen, bevor die Maschine anschließend die äquivalenten Wörter in der Zielsprache finden und diese anhand der Grammatikinformationen korrekt übertragen würde. In einem zusätzlichen Schritt, der jedoch nicht offizieller Teil des Patents war, sollte ein/e zweisprachige/r Bearbeiter*in im Text inhaltliche Bedeutung und Stil verbessern. Auch wenn Trojanskijs Idee weit vor der Erfindung der ersten kommerziellen Computer in den Nachkriegsjahren geboren wurde, so stellt sie eine der ersten (theoretischen) nicht-numerischen Anwendungen eines elektronischen Computers dar. Darüber hinaus entspricht

der von ihm vorgeschlagene dreistufige Prozess mit bilingualem/r Nachbearbeiter*in gewissermaßen dem, was heute als Post-Editing maschineller Übersetzung bekannt ist (Kenny 2019: 431).

Phase 2: 1949 bis 1959

Die ersten Bestrebungen um MÜ-Systeme waren in den Nachkriegsjahren noch stark durch das Konzept der Dekodierung geprägt, welches vor allem im Zweiten Weltkrieg an Bedeutung gewann, da der britische Informatiker Alan Turing so erfolgreich den *Enigma-Code* entschlüsseln konnte (Koehn 2010: 15). Die Mathematiker Claude E. Shannon und Warren Weaver gelten als Pioniere der MÜ-Forschung in den Nachkriegsjahren und zogen in ihrer Forschung Parallelen zu Turing. In ihrem *Sender-Empfänger Modell (*Shannon & Weaver 1949), dass sich auf jede Form der Kommunikation, auch Translation, übertragen lässt, kodiert ein/e Sender*in eine Mitteilung und überträgt sie über einen Kanal an eine/n Empfänger*in, der/die die Nachricht dekodiert. Das Übersetzen ist nach ihrer Auffassung ein Prozess, der auf das Erfassen und Umwandeln der kleinsten semantischen Einheiten, also Wörtern, bezieht (Poibeau 2017: 31). Sie vertraten die heute als problematisch geltende Annahme, dass jeder fremdsprachige Text eigentlich in der eigenen Sprache verfasst ist und lediglich in unverständliche, fremde Symbole kodiert wurde. Die Aufgabe der Übersetzer*innen ist es folglich, diese fremden Symbole durch verständliche auszutauschen, also zu dekodieren, und somit einen Text in der eigenen Sprache zu schaffen (Koehn 2010:16). Unzufrieden mit der Leistung der ersten auf wörtlicher Übertragung funktionierenden Systeme lieferte Weaver (1949) mit Veröffentlichung seines Memorandums Vorschläge zur Überwindung der Barriere, die durch die unzähligen verschiedenen Sprachen entsteht. Als Lösung schlug Weaver (1949: 8ff) mehrere Konzepte vor, etwa dass der Kontext bei der Übersetzung berücksichtigt werden soll und auch, dass die Verwendung einer Interlingua Abhilfe schaffen könnte. Weaver erwähnt explizit, dass es sich hier um rein theoretische Konzepte handelt, die seinen eigenen Überlegungen als Nicht-Linguist entspringen und auf Grund des damaligen Forschungsstandes und der fehlenden technologischen Kapazitäten nicht umsetzbar waren. Rückblickend handelt es sich bei dem Dokument jedoch um eine äußerst zukunftsweisende Beschreibung, die später zur Entwicklung statistischer Methoden führten (Poibeau 2017: 32f).

Bereits einige Jahre nach Veröffentlichung von Weavers Memorandum war das Interesse an der Disziplin der MÜ stark gewachsen. Der israelische Mathematiker und Linguist Jehoschua Bar-Hillel galt Anfang der 1950er Jahre als eine der wichtigsten Persönlichkeiten in der MÜ-Forschung und organisierte 1952 am Massachusetts Institute of

Technology die erste MÜ-Konferenz. Um dafür Förderungen zu erhalten, wurde am 7. Januar 1954 an der amerikanischen Universität von Georgetown unter der Leitung von Léon Dostert und in Zusammenarbeit mit *IBM* ein Projekt um ein Übersetzungssystem gestartet, das vom Russischen ins Englische übersetzen konnte. Bei dem Georgetown-IBM-Experiment wurden mittels dieses Systems, das auf einem *IBM 701* Großrechner lief, 49 Sätze, die zuvor jedoch bewusst ausgewählt wurden, aus dem Russischen in das Englische übersetzt, wobei auch Grammatikregeln und 250 Wörter aus einem Glossar verwendet wurden (Kurz 2014: 403). Trotz der subjektiven Steuerung des Experiments lösten die erzielten Resultate Optimismus aus und sorgten weltweit für die Entstehung von Forschungsgruppen, die alle unterschiedliche Aspekte der maschinellen Übersetzung untersuchten. Außerdem wurde durch die mediale Berichterstattung Interesse geweckt und amerikanische und britische Regierungsinstitutionen subventionierten eigene Forschungsprojekte (Poibeau 2017: 34). Unter anderem durch den Kalten Krieg und den Wettkampf mit der Sowjetunion motiviert, investierte die amerikanische Regierung zwischen 1956 und 1966 rund 20 Millionen US-Dollar in die Entwicklung von MÜ-Systemen (Kurz 2014: 402).

Phase 3: 1959 bis 1966

Nach den ersten Erfolgen der MÜ beim Georgetown-IBM-Experiment stellte der Mathematiker und Linguist Jehoschua Bar-Hillel 1959 die Realisierbarkeit einer *Fully Automatic High Quality Translation* (FAHQT), also einer hochqualitativen maschinellen Übersetzung, die keine menschliche Vor- oder Nachbearbeitung benötigt, erstmals in Frage. Ebenfalls kritisierte er, dass die gängigen Systeme selbst unter kontrollierten Bedingungen keine zufriedenstellenden Ergebnisse liefern können (Kurz 2014: 403). In seinem Bericht über den aktuellen Stand der MÜ, bewertete Bar-Hillel (1959: 18ff) die laufenden Forschungsbemühungen um die MÜ negativ, da er der Auffassung war, dass die existierenden Systeme keine Übersetzungen mit zufriedenstellender Qualität erzeugen konnten, und stellte in Aussicht, dass FAHQT nicht erzielbar sei. Er schlug eine Umorientierung der Forschung in Richtung Entwicklung computergestützter Übersetzungssysteme für Humanübersetzer*innen vor, mit denen beispielsweise auch ein Post-Editing der maschinell übersetzten Texte möglich wäre (Poibeau 2017: 37). Bar-Hillels Bericht hatte auch zur Folge, dass jene Agenturen, die in den 50er Jahren die Finanzierung der Forschungsprojekte ermöglicht haben, 1964 einen Bericht in Auftrag gaben, um eine Einschätzung der Fortschritte der MÜ zu erhalten. Nach seiner Gründung durch die *National Science Foundation* 1965 wurde das *Automatic Language Processing Advisory Committee* (ALPAC) damit beauftragt, jenen Bericht über den aktuellen Stand und die Aussichten der MÜ anzufertigen. Der mittlerweile berühmte ALPAC-

Bericht wurde von ALPAC (1966) veröffentlicht und beschäftigte sich mit dem Stand der Übersetzung im weiteren Sinne, die maschinelle Übersetzung wird tatsächlich nur in einem fünfseitigen Kapitel beschrieben. Dem beauftragten Gremium von sechs Experten, bestehend aus Linguisten, Experten für künstliche Intelligenz und Psychologen, wurde außerdem Voreingenommenheit und die Verwendung veralteter Zahlen und Daten vorgeworfen, wodurch auch nicht alle relevanten Bereiche beurteilt wurden (Kurz 2014: 403). Zudem war keiner der Experten selbst im Bereich der MÜ tätig, jedoch wurden Forscher aus dem Bereich interviewt. In dem Bericht wurde außerdem nur auf MÜ vom Russischen ins Englische eingegangen (Poibeau 2017: 41). Laut Bericht war die MÜ doppelt so teuer und zeitaufwändiger als vergleichbare Humanübersetzungen. Durch die ausreichende Verfügbarkeit an qualifizierten Übersetzer*innen, die schneller arbeiteten und qualitativere Übersetzungen produzierten, bestehe somit in absehbarer Zukunft kein Bedarf an MÜ (Kenny 2019: 431). Trotz der bedenklichen Aussagekraft des Berichts stellte die US-Regierung 1966 die Subventionierung der maschinellen Übersetzungsforschung ein und kam auf Grund der steigenden Kosten und sinkenden Qualitätsniveaus der Texte zu dem Fazit, es gäbe „no immediate or predictable prospect of useful machine translation“ (ALPAC 1966: 32). Im ALPAC-Bericht wurde ebenfalls aufgezeigt, dass auch ein Post-Editing der maschinell übersetzten Texte weder kostengünstiger noch schneller als Humanübersetzung sei. Mit einem jährlichen Budget von 20 Millionen US-Dollar waren die Ausgaben für Übersetzungen in den USA relativ niedrig und es wurde beschlossen, diese in die Verbesserung der Humanübersetzung durch Ausbildungen zu investieren. Diese Entscheidung führte in den folgenden Jahren zu einer Abwendung von der MÜ in den USA, aber zu einer Zuwendung zu der *Machine Assisted Human Translation* (MAHT), die unter anderem in darauffolgenden Jahrzehnten *Computer Assisted Translation Tools* (CAT-Tools) und Translation Memories hervorbrachten (Kurz 2014: 403).

Phase 4: 1966 bis Ende 1980er

Außerhalb der USA bestand jedoch weiterhin Forschungsinteresse an der MÜ, vor allem in zwei- oder mehrsprachigen geographischen Regionen wie Kanada oder Europa, die einen praktischen Nutzen für und Bedarf an MÜ hatten. So wurde in Kanada 1977 das regelbasierte maschinelle Übersetzungssystem *MÉTÉO* entwickelt, das bis zu seiner Auflassung 2002 täglich 45.000 Wörter aus Wetterberichten vom Englischen ins Französische übersetzte. Ein weiteres Beispiel ist das regelbasierte *System Translation* (*SYSTRAN)*, das 1968 von Peter Toma entwickelt wurde und zur Übersetzung von Texten aus der Luftfahrt zwischen Russisch und Englisch diente. 1976 wurde das System von der *NASA* verwendet und 1975 von der EU-

Kommission gekauft, wo es seither im Einsatz ist (Kurz 2014: 404f). Im Laufe der Jahre kamen immer mehr Sprachpaare und Verbesserungen der Funktionsweise hinzu, wodurch das *SYSTRAN* nach wie vor in Verwendung ist, auch wenn die Europäische Kommission 2013 mit *MT@EC* ihr hauseigenes maschinelles Übersetzungssystem entwickelte, welches 2017 wiederum durch das System *eTranslation* als Nachfolger abgelöst wurde (Kenny 2019: 432). Auch in Nordamerika kamen in den 1970er und 80er Jahren regelbasierte MÜ-Systeme erneut verstärkt zum Einsatz, um interne Dokumentation in multilingualen Organisationen wie der *NATO* zu übersetzen (Kenny 2019: 432f). Diese internationalen Beispiele zeigen, dass zwar ein begrenzter, aber dennoch realer Bedarf an MÜ-Systemen bestand, und dass maßgeschneiderte Lösungen für administrative oder technische Anwendungsbereiche schon damals adäquate Texte produzieren konnten. 1988 kam es durch die Forschungsgruppe um Peter F. Brown von *IBM* zu einem Paradigmenwechsel. Auf Grund der rapiden technologischen Fortschritte, der stärkeren Rechen- und Speicherleistung von Computern und der großen Verfügbarkeit maschinenlesbarer Textkorpora wurde vorgeschlagen, dass die Zeit der statistikbasierten MÜ-Systeme nun gekommen war. Innerhalb von etwa einem Jahr konnten durch Verwendung großer Korpora funktionale Prototypen angefertigt werden, die ganz ohne die zeitintensive Unterstützung von Linguist*innen brauchbare Übersetzungen erzeugen konnten. (Stein 2018: 7). Diese Systeme der statistischen maschinellen Übersetzung gewannen an Beliebtheit, jedoch beschränkte sich die Forschung auf wenige Projekte. Der Schwerpunkt lag nach wie vor auf regelbasierten Systemen, statistische Systeme wurden erst ab der Jahrtausendwende im großen Rahmen untersucht und entwickelt. (Kenny 2019: 433).

Phase 5: 1990er bis 2016

Mit der erhöhten Verfügbarkeit von leistungsstarken Computern und immer mehr Internetanschlüssen kam es um die Jahrtausendwende zu einer Kommerzialisierung von MÜ-Systemen, die gewissermaßen parallel mit den ersten CAT-Tools erschienen. 1997 entwickelte die Suchmaschine *AltaVista*, später *Yahoo*, das erste Online-Übersetzungssystem *Babel Fish*, welches zwischen 36 Sprachkombinationen übersetzen konnte (Schmalz 2019: 194). Das regelbasierte System nutzte vordefinierte Regeln zur Verwendung von Syntax und Grammatik und arbeitete auf Wortebene. Trotz der rudimentären Funktionsweise und der teilweise erheiternden Übersetzungsfehler wurden mit *Babel Fish* 2001 täglich rund 1,3 Millionen Übersetzungen durchgeführt (Schmalz 2019: 195). Neben diesen Online-Systemen erschienen auch kommerzielle Systeme von *IBM* oder *Siemens* sowie Systeme zum privaten Gebrauch, wie etwa *T1* oder *SYSTRAN PC*. Im Kontext von Übersetzungsagenturen fanden rechnerbasierte Systeme anfangs nur wenig Anklang, jedoch wurden sie mit der Zeit immer

mehr in Arbeitsprozesse und neue Tätigkeitsbereiche, wie die Lokalisierung, eingebunden, um Übersetzer*innen zu unterstützen (Kurz 2014: 407). Die Vielsprachigkeit, die im Internet bereits 2005 durch eine Milliarde Nutzer*innen vertreten war, stellte die Möglichkeit dar, maschinelle Übersetzungssysteme auch Privatpersonen zugänglich zu machen, wie *Babel Fish* bereits erfolgreich gezeigt hatte. Mit der Einführung des Dienstes *Google Translate* im Jahr 2006 wurden Nutzer*innen neue Möglichkeiten im Umgang mit den mehrsprachigen Inhalten des Internets geboten, und es entwickelten sich erste Gemeinschaften an Amateurübersetzer*innen, die auch Rückmeldungen zu den von *Google* übersetzten Texten lieferten (Kenny 2019: 433f). Die neuen leistungsstärkeren und flexibleren statistischen Systeme, die korpusbasiert, also anhand von Parallelkorpora arbeiten, stellten bis Mitte der 2010er Jahre den Schwerpunkt des Forschungsinteresses dar. Für das Erstellen rechnergestützter statistischer MÜ-Systeme für den privaten oder kommerziellen Gebrauch wurde 2006 an der Universität von Edinburgh das Open-Source SMÜ-Toolkit *Moses* entwickelt. Das Toolkit wurde von Programmierern der John Hopkins Universität noch verbessert und dann der Öffentlichkeit als freier Download zur Verfügung gestellt (Koehn 2020: 37). Zusammen mit den wachsenden frei zugänglichen Beständen an Parallelkorpora war es erstmals also allen Nutzer*innen möglich, ein SMÜ-System auf dem eigenen Rechner zu erstellen (Koehn 2020: 38).

Phase 6: ab 2016

Als sich ab 2014 eine Stagnation der Fortschritte der statistischen maschinellen Übersetzung abzeichnete, begann die Forschung andere Wege einzuschlagen und erzielte zwischen 2015 und 2016 den Durchbruch mit einem neuartigen System, das Übersetzungen anhand neuronaler Netze, die mit Parallelkorpora trainiert werden, anfertigte. Bereits in den 1990er Jahren gab es unter anderem durch Forcada und Neco (1997) erste Forschungsprojekte zur Anwendung rekurrenter neuronaler Netze für maschinelle Übersetzung, die jedoch auf Grund der begrenzten Parallelkorpora nicht aussagekräftig waren. Nach intensiver Forschung im Bereich der statistischen Systeme wurde um 2010 erkannt, dass die Fortschritte stagnierten, wodurch man sich erneut vermehrt den neuronalen Systemen zuwandte. Eine der ersten Forschungsarbeiten zu *Convolutional Neural Networks* stammt von Kalchbrenner und Blunsom (2013), die mit ihrem Modell gute Übersetzungen für kurze, jedoch nicht für längere Sätze erzielten. Die neuronale maschinelle Übersetzung (NMÜ) erzielte bereits in ersten Versuchen bessere Resultate als ihr Vorgänger, weshalb sich auch die Forschung innerhalb von zwei Jahren auf neuronale Netze fokussierte. Bereits 2016 veröffentliche *Google* die neue Version seines Onlinedienstes *Google Translate*, die nun neuronal arbeitet und mit Stand

2021 102 Sprachen unterstützt (Koehn 2020: 39f). Auch wenn die MÜ bis zum heutigen Tag viele Entwicklungsschritte durchlaufen hat und in ihrer Kapazität gewachsen ist, so bestehen nach wie vor einige der ursprünglichen Probleme, vor allem das fehlende Welt- und Kulturwissen der Systeme sowie deren Fähigkeit den Kontext über Satzgrenzen hinaus zu erkennen (Schmalz 2019: 195).

2.1.2 Ansätze

Angelehnt an Kurz (2014: 410) wurde zur besseren Übersichtlichkeit über die verschiedenen Ansätze der MÜ folgendes, in Abb.1 gezeigtes Modell erstellt, welches auch die 2015/2016 entwickelte, neuronale maschinelle Übersetzung (NMÜ) als neue Variante des korpusbasierten Ansatzes beinhaltet. Analog zur Geschichte der MÜ aus dem vorherigen Kapitel sollen die Ansätze in diesem Kapitel in chronologischer Abfolge beschrieben werden.

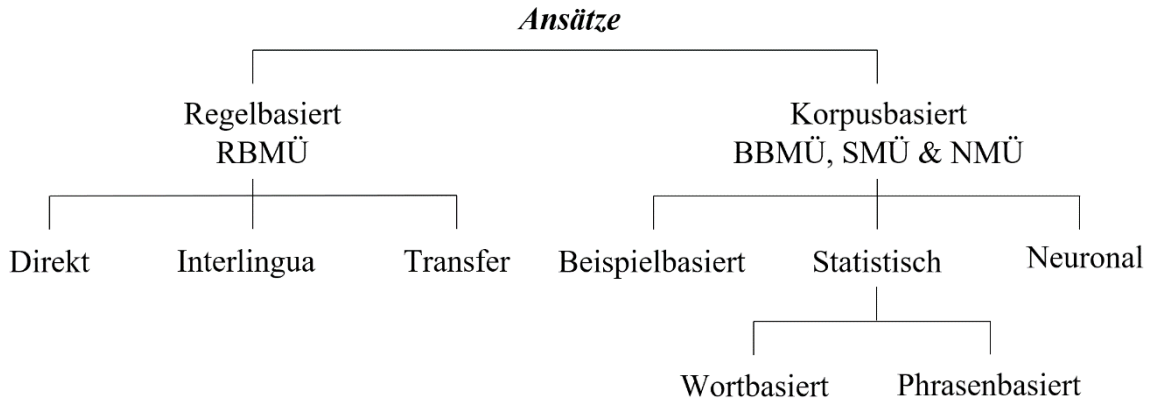


Abbildung 1: Ansätze der maschinellen Übersetzung

2.1.2.1 Regelbasierter Ansatz

Die regelbasierte maschinelle Übersetzung (RBMÜ) umfasst den direkten, den Transfer- und den Interlingua-Ansatz. Bei den Ansätzen der RBMÜ werden die zu übersetzenden Texte nach syntaktischen und semantischen Regeln bearbeitet, da das Erfassen und Übertragen der Bedeutung ein besonderes Ziel in dieser ersten Entwicklungsstufe der MÜ war (Kurz 2014: 413). Beim direkten Ansatz kommen nur wenige Regeln zum Einsatz und der Zieltext wird ohne intermediäre Repräsentation erzeugt (Kurz 2014: 412). Regelbasierte Systeme wurden durch die Zusammenarbeit von Computer- und Sprachwissenschaftler*innen in der Nachkriegszeit entwickelt, da einerseits Grammatikregeln und spezielle Wörterbücher benötigt wurden, diese andererseits aber auch in maschinenlesbare Form gebracht werden mussten (Forcada 2010: 218f).

Der direkte Ansatz der MÜ trifft auf die allerersten MÜ-Systeme der 1940er und 50er Jahre zu, bei denen von einer Sprache A in eine Sprache B ohne Erzeugen einer intermediären Repräsentation der Bedeutung übersetzt wird (Kenny 2019: 434). Der direkte Ansatz basiert auf dem Prinzip einer wortwörtlichen Übersetzung, bei der neben einer morphologischen Analyse des Ausgangstextes und einer Syntaxbereinigung des Zieltextes die Wörter als kleinste Einheit der Übersetzung übertragen werden (Kurz 2014: 411f). Direkte MÜ-Systeme basieren auf Wörterbüchern, in denen Paare an Ausgangs- und Zielwörtern gespeichert sind, die als Wort-für-Wort Übersetzungen übernommen werden können. Daher können diese Systeme allerdings nur begrenzt mit semantischen Merkmalen oder metaphorischen Ausdrücken umgehen. Zum Erzeugen der Übersetzung kommen auch einfache syntaktische Regeln zur Anwendung, damit Wortreihenfolgen und Satzstrukturen in der Zielsprache korrekt wiedergegeben werden. Diese Systeme funktionieren gut für verwandte Sprachen und können helfen, die Aussage eines Textes zu verstehen, auch wenn dieser grammatikalisch und formal inkorrekt ist (Poibeau 2017: 21f). Allerdings wurde dieser Ansatz von Mathematiker*innen in den Nachkriegsjahren entwickelt, die das Übersetzen als einen Prozess der Dekodierung sahen, wodurch das Wissen von Übersetzungswissenschaftler*innen und Linguist*innen nur bedingt in die Entwicklung einfloss. Dieser Ansatz gilt heutzutage als überholt und ist nicht mehr im Einsatz, jedoch erwies er sich für die ersten Systeme in den 1940er und 50er Jahren als bahnbrechend, wie etwa beim *IBM*-Übersetzer, der beim Georgetown-IBM Experiment erfolgreich zur Anwendung kam (Kurz 2014: 411f).

Transfersysteme sind komplexer als direkte Systeme, da sie eine syntaktische und grammatikalische Analyse des ausgangssprachlichen Satzes durchführen und dessen Struktur nutzen können, wodurch die Beschränkungen der wortbasierten Vorgehensweise umgangen werden und auch die Syntax der Zieltexte fehlerfreier ist (Kenny 2019: 435). Der Transferansatz läuft in den drei Schritten von Analyse, Transfer und Synthese ab und nutzt eine Reihe von morphologischen, syntaktischen, semantischen und kontextbezogenen Regeln, die in beliebigem Umfang programmiert und kombiniert werden können (Stein 2018:10). Im ersten Schritt der Analyse wird der Ausgangstext analysiert und eine ausgangssprachliche Repräsentation der ausgangssprachlichen Bedeutung erstellt. Im zweiten Schritt wird unter Anwendung des Wörterbuchs ein lexikalischer Transfer, und unter Anwendung der Regeln ein struktureller Transfer durchgeführt. Dabei wird der Ausgangstext mittels eines einsprachigen Wörterbuches der Ausgangssprache analysiert und die Bedeutung mittels eines zweisprachigen Wörterbuches in eine intermediäre Repräsentation verwandelt (Kurz 2014: 412f). Anschließend wird die intermediäre Repräsentation der Ausgangssprache in eine

intermediäre Repräsentation in der Zielsprache übertragen. In einem dritten und letzten Schritt wird der konkrete zielsprachliche Text aus der Repräsentation mittels eines einsprachigen Wörterbuches in der Zielsprache erzeugt (Forcada 2010: 218f). Durch die zusätzlichen Analyseschritte kann die RBMÜ im Regelfall eine bessere Übersetzung als die direkte MÜ anfertigen, jedoch ist der Ansatz anfälliger für Fehler, beispielsweise wenn die Analyse des Ausgangstextes keine Ergebnisse liefert und der Übersetzungsprozess nicht weitergeführt werden kann (Kurz 2014: 412f). Ab einer gewissen Komplexität an Regeln erreichen diese Systeme außerdem ihr maximales Potential und ihre Funktion kann durch Konflikte von Regeln eingeschränkt sein (Stein 2018: 10).

Eine weitere Form der RBMÜ, die in den 1970er Jahren relativ erfolglos erforscht wurde, ist der interlinguale Ansatz. Die Interlingua-Methode setzt eine universale Sprache voraus, die im Übersetzungsprozess als Zwischenebene fungiert. Ähnlich wie beim Relais-Übersetzen würde ein Text in diese Sprache übersetzt werden und der Text anschließend in der Zielsprache anhand dieser neutralen Version produziert werden. Im Gegensatz zum Relais-Übersetzen und auch zum normalen Übersetzungsprozess würde bei dieser Methode jedoch keinerlei Information verloren gehen (Stein 2018: 10). Idealerweise könnte ein solches System Inhalt und Bedeutung jedes Textes in einer Ausgangssprache analysieren, in eine sprachunabhängige Repräsentation (Interlingua) umwandeln, wobei sämtliche Spuren der Ausgangssprache verwischt werden. Die tatsächliche Umsetzung solcher Systeme konnte allerdings nie realisiert werden (Kurz 2014: 413f).

2.1.2.2 Korpusbasierter Ansatz

Der korpusbasierte Ansatz umfasst die beispielbasierte maschinelle Übersetzung (BBMÜ), die statistische maschinelle Übersetzung (SMÜ) und die neuronale maschinelle Übersetzung (NMÜ). Im Gegensatz zur RBMÜ basiert dieser Ansatz auf der Annahme, dass Übersetzungswissen nicht unbedingt manuell in Form von Regeln festgelegt werden muss, sondern auch aus bilingualen Parallelkorpora automatisch von Systemen erlernt und folglich auf andere Texte angewandt werden kann (Kenny 2019: 435). Ein wichtiger Bestandteil des korpusbasierten Ansatzes, der für die BBMÜ und SMÜ benötigt wird, ist das sogenannte Übersetzungsmodell. Darunter versteht man eine in der Trainingsphase erstelle Einheit innerhalb eines MÜ-Systems, die aus den Parallelkorpora linguistische Informationen extrahiert, aus diesen Wahrscheinlichkeitsmodelle erstellt und diese Gesamtheit an Informationen anschließend nutzen kann, um Texte maschinell zu übersetzen (Kenny 2019: 433). Ein Parallelkorpus ist eine Sammlung von Texten, die mit der übersetzten Version desselben Textes in einer anderen Sprache aligniert sind. Unter Alignieren versteht man das

Zuordnen eines ausgangssprachlichen Satzes oder Segments mit seinem zielsprachlichen Äquivalent (Koehn 2010: 55). Ein Satz ist dabei als eine sprachliche Einheit zu verstehen, die syntaktisch und semantisch autonom ist. Bei der Alignierung wird davon ausgegangen, dass der übersetzte Text dieselbe Struktur wie der Ausgangstext hat und somit die Sätze auch in derselben Reihenfolge auftreten (Poibeau 2017: 49f). Die alignierten Texte, die zu einem Parallelkorpus zusammengefasst werden können, werden auch als Bitexte bezeichnen. Bitexte gelten als eine der wertvollsten Sprachressourcen für MÜ, da sie im Gegensatz zu Wörterbüchern auch auf den Kontext und andere Merkmale menschlicher Sprache schließen lassen (Poibeau 2017: 48f). Bitexte können beispielsweise in der Form von Translation Memory Dateien (TMX) aus CAT-Tools verfügbar sein oder manuell oder automatisch anhand von multilingualen Textsammlungen aligniert werden (Poibeau 2017: 49).

Die beispielbasierte maschinelle Übersetzung wurde Mitte der 1980er entwickelt. Da regelbasierte Systeme in ihrer Wartung immer komplexer wurden und nur gesamte Sätze analysieren konnten, was oft zu fehlerhaften Übersetzungen führte, sollten die beispielbasierten Systeme auch mit Satzteilen oder Wörtern aus den Parallelkorpora umgehen können. Der beispielbasierte Ansatz ähnelt in seiner Funktionsweise einem Translation-Memory-System. Anstatt der Verwendung von statistischen Sprach- und Übersetzungsmodellen, die Aufschluss über Wortreihenfolgen geben, wird der Ausgangstext mit dem Korpus verglichen und das Segment mit der statistisch höchsten Übereinstimmung ausgewählt. Der Zieltext wird also durch die Neuanordnung von Segmenten auf Satzebene oder Teilsegmenten auf Wortebene erzeugt (Kurz 2014: 417). Der große Nachteil dieses Ansatzes ist, dass nicht immer ein Äquivalent im Parallelkorpus gefunden wird, das gänzlich mit dem zu übersetzenden Text übereinstimmt. In diesen Fällen produziert das System entweder gar keine oder nur eine wortwörtliche Übersetzung (Poibeau 2017: 57).

Ab der Jahrtausendwende stellte der statistische Ansatz die gängigste Art der MÜ dar. SMÜ-Systeme treffen Übersetzungsentscheidungen anhand von bedingten Wahrscheinlichkeiten, also Ereignissen, die als Reaktion auf ein bereits aufgetretenes Ereignis stattfinden. Diese Wahrscheinlichkeiten werden dabei in der Trainingsphase aus Parallelkorpora berechnet und in einem Wahrscheinlichkeitsmodell (*probability model*) abgespeichert (Stein 2018: 10). Ziel der SMÜ ist es, aus dem Parallelkorpus jene zielsprachliche Einheit zu finden, die die höchste Wahrscheinlichkeit hat, mit der ausgangssprachlichen Einheit übereinzustimmen. Idealerweise könnte ein solches System auf alle möglichen Sätze einer Sprache zugreifen. Da Korpora jedoch immer nur gewisse Größen haben, arbeiten diese Systeme mit Annäherungen, also möglichst universalen Regeln zum

Umgang mit Texten, die im Wahrscheinlichkeitsmodell gespeichert sind (Stein 2018: 11). Das Wahrscheinlichkeitsmodell enthält mehrere kleinere Modelle. Einerseits gibt es das bilinguale Übersetzungsmodell, welches alignierte Sätze, Wörter und Wortreihenfolge umfasst und die lexikalischen Wahrscheinlichkeiten enthält, dass Wort *B* die wahrscheinlichste Übersetzung für Wort *A* ist. Das monolinguale Sprachmodell, welches für die Zielsprache erstellt wird, gibt an, wie wahrscheinlich es ist, dass ein zielsprachlicher Satz mit einer genauen Abfolge von Worten von einem Menschen so verfasst werden würde (Forcada 2010: 219f). Die Analyse von Abfolgen soll nicht nur helfen die Bedeutung zu übertragen, sondern auch natürlich klingende Sätze zu erzeugen. Ein Sprachmodell kann jedoch auch Entscheidungen bezüglich der besten Übersetzungskandidaten, der Wortreihenfolge und der Syntax treffen. Voraussetzung für das Funktionieren solcher Systeme sind wie bei den meisten korpusbasierten Methoden jedoch die Qualität und Größe des Parallelkorpus, damit das System ausreichend Trainingsmaterial zum Erstellen des Sprachmodells hat. Ein zu kleiner Korpus wäre für die Sprache nicht repräsentativ und würde zu unnatürlichen Formulierungen und falscher Wortwahl führen (Koehn 2010: 181f). Außerdem ist das Analysieren von ganzen Sätzen bei SMÜ-Systemen auf Grund der geringen Wahrscheinlichkeit, komplett identische Sätze im Parallelkorpus zu finden, keine sinnvolle Vorgehensweise. Die analysierten sprachlichen Einheiten werden daher kleiner gefasst und man unterscheidet zwischen wortbasierten und phrasenbasierten SMÜ-Systemen (Stein 2018: 11).

Wortbasierte Systeme kamen beispielsweise in den frühen 1990er Jahren im *Candide* Projekt von *IBM* zum Einsatz (Koehn 2010: 82). Beim wortbasierten Ansatz wird davon ausgegangen, dass jedes ausgangssprachliche Wort ein zielsprachliches Äquivalent hat und dadurch zu ersetzen ist. Jedes Auftreten des Wortes im Parallelkorpus wird erfasst und dazu die verschiedenen Übersetzungsmöglichkeiten im Zieltext. Dann kann die Wahrscheinlichkeit jeder möglichen Übersetzung für dieses Wort eruiert werden, wobei ein höherer Wert eine höhere Wahrscheinlichkeit darstellt, dass die Übersetzung die Richtige ist (Koehn 2010: 83). Bei der Übersetzung können jedoch Probleme auftreten, wenn Mehrworteinheiten übersetzt werden müssen, weshalb der phrasenbasierte Ansatz entwickelt wurde (Stein 2018: 12). Als Phrase wird eine Sequenz aufeinanderfolgender Wörter bezeichnet, etwa Nominalphrasen oder andere Mehrworteinheiten. Phrasen sind zwar nicht zwingendermaßen syntaktische Einheiten, eignen sich jedoch besser als wortbasierte Ansätze, um größere Teile des Satzes zu analysieren und übersetzen (Koehn 2010: 127f). Durch das Evaluieren der Häufigkeit von Wortkombinationen und Phrasen können Rückschlüsse auf den Kontext gezogen werden, wenn davon ausgegangen wird, dass Wörter mit ähnlichem Sinn auch gemeinsam auftreten.

Bei nicht verwandten Sprachen, die sich in ihrer Syntax oder Wortfolge unterscheiden, können durch das Einbinden spezifischer Sprachmodelle grobe Fehler vermieden werden (Schmalz 2019: 198). Moderne SMÜ-Modelle arbeiten überwiegend phrasenbasiert, in der Trainingsphase werden also nicht einzelne Wörter, sondern Wortfolgen, sogenannte Phrasen, aus dem Trainingskorpus für das Erstellen der Modelle herangezogen. Sämtliche sinnvolle oder nicht sinnvolle Kombinationen dieser Wortfolgen werden auch *n*-grams genannt und bezeichnen die Abfolgen von ein, zwei, drei oder *n* Wörtern im Trainingskorpus (Kenny 2019: 436). Da diese *n*-grams jedoch unabhängig voneinander behandelt und übersetzt werden, haben SMÜ-Systeme unter anderem Probleme mit mehrteiligen Prädikaten, reflexiven Verben und flektierenden Sprachen (Kenny 2019: 436f). Gleichzeitig ermöglichen SMÜ-Systeme jedoch eine konsistente Verwendung von Terminologie in der Übersetzung, vorausgesetzt diese Terminologie ist im Trainingskorpus einheitlich (Schmalz 2019: 198).

Um die Nachteile der *n*-gram-basierten SMÜ zu umgehen, benötigt man ein System, das Sätze in ihrer Gesamtheit analysieren und übersetzen kann. Dieser Durchbruch gelang 2015/2016 mit Entwicklung der neuronalen maschinellen Übersetzung (NMÜ). Auch dieses System lernt anhand von Parallelkorpora, jedoch verwendet es neuronale Netze, in denen Tausende von einzelnen Einheiten oder künstlichen Neuronen, analog zu den Neuronen im menschlichen Gehirn, mit anderen Neuronen verbunden werden. Der Aktivierungszustand jedes Neurons hängt von den Reizen ab, die es von anderen Neuronen erhält, sowie von der Stärke der Verbindungen zwischen diesen Neuronen. Erst wenn mehrere verbundene Neuronen gemeinsam aktiviert werden, kann dies als Repräsentation eines Wortes samt Kontext interpretiert werden (Kenny 2019: 436f). Durch die Verwendung von neuronalen Netzen entfällt die Nutzung von Sprach- und Übersetzungsmodellen und die Relationen einzelner Wörter im Satz können beachtet werden. NMÜ-Systeme arbeiten mittels einer „nicht linearen Abbildung [...] basierend auf einer Vektordarstellung des Ein- und Ausgangssatzes über mehrere Zwischenstufen und mehrere Abstraktionsgrade“ (Schmalz 2019: 199). Diese nicht linearen Abbildungen werden in der Trainingsphase in mehreren Durchläufen als aufeinanderfolgende Schichten erzeugt. Während die äußeren Schichten von menschlichen Bearbeitern eingesehen und zwecks der Gewichtung manipuliert werden können, sind die tieferen Schichten nicht zugänglich (Kenny 2019: 436f). NMÜ-Systeme verfügen über einen *Encoder*, oder Verschlüsseler, der Repräsentationen jedes Wortes oder jeder Zeichenfolge des Ausgangstextes inklusive des Kontextes erstellt, und einen *Decoder*, oder Entschlüsseler, der für jede Position im Zieltext das Wort eruiert, welches die wahrscheinlichste Fortsetzung eines bereits gewählten Wortes davor ist. NMÜ-Systeme

basieren also auf monolithischen neuronalen Netzen, die aus Parallelkorpora eine semantische Repräsentation von Zielwörtern anhand der Ausgangswörter und der umliegenden Zielwörter berechnen und erlernen. Die durch den Encoder erzeugten Repräsentationen von jedem Wort werden vom Decoder herangezogen, um das wahrscheinlichste Wort für jede Position im Zielsatz zu eruieren (Forcada 2017: 296). Durch diesen Prozess ist das System auch in der Lage, Strukturen, also Relationen von Wörtern oder Wortgruppen, zu erkennen und somit eine Hierarchie in der Syntax und Wortreihenfolge zu lernen. Das System arbeitet allerdings nicht ausschließlich hierarchisch, sondern auch multidimensional, indem jedes Wort oder jede Gruppe von Worten in ihrem Kontext und ihrer Umgebung mit anderen Worten analysiert wird. Außerdem sind neuronale Systeme in der Lage, Kontinuitäten in Texten erkennen zu können, beispielsweise wenn ein bereits erwähnter Inhalt paraphrasiert wird, ohne dass dabei dieselben Wörter oder Synonyme verwendet werden (Poibeau 2017: 83).

2.1.3 Anwendungsbereiche

Wie in Kapitel 2.1.1 besprochen, begann Mitte der 2000er Jahre die Entwicklung einer Reihe von kommerziellen und frei zugänglichen MÜ-Systemen für diverse Nutzer*innengruppen. Die MÜ war also nicht mehr Forscher*innen in Laboren oder aufwändigen Rechenzentren vorbehalten, sondern konnte von allen Menschen mit einem Endgerät und Internetzugang genutzt werden. Die Anwendungsbereiche sind teilweise so divers wie die MÜ-Systeme an sich und lassen sich in die drei Kategorien Assimilation, Dissemination und Kommunikation unterteilen (Forcada 2010: 217).

Bei der Assimilation, oder Aufnahme, geht es darum die Systeme zu nutzen, um durch die Übersetzung eine Idee des Inhalts zu bekommen. Für private, teils aber auch kommerzielle Zwecke, wird die MÜ für das sogenannte *gisting* genutzt, also das Verstehen des groben Inhalts in einer Sprache, die man selbst nicht spricht. Fehler und Detailgrad der Übersetzung sind dabei sekundär, solange den Nutzer*innen die Kernaussage des fremdsprachigen Textes zum Großteil korrekt übermittelt wird (Forcada 2010: 216). Durch das Aufeinandertreffen verschiedenster Sprachen im Internet und der diversen Bedürfnisse von Nutzer*innen kann MÜ hier sinnvoll eingesetzt werden, um Texte jeder Art in ihrer Kernaussage in eine Sprache zu übersetzen. Dies kann entweder von einer Website automatisch vorgenommen werden oder durch die Nutzer*innen mittels Klicken auf eine Schaltfläche erreicht werden. Durch dieses bewusste Einleiten des Prozesses der Übersetzung sind sich Nutzer*innen auch im Klaren darüber, dass diese Übersetzung von einer Maschine stammt, und passen ihre Erwartungen dementsprechend an. Solange die Nutzer*innen einen ungefähren Einblick in den fremdsprachlichen Text erhalten haben, hat die MÜ ihren Zweck erfüllt (Koehn 2020: 19f).

Neben den freien Online-MÜ-Systemen, die von privaten Nutzer*innen für das *gisting* verwendet werden, setzen einige Unternehmen spezialisierte, hauseigene MÜ-Systeme ein, die auf Rechnern in einem lokalen Netzwerk laufen und für das automatische Übersetzen interner Kommunikation verwendet werden können, oder auch von Mitarbeiter*innen, um sich einen Überblick über vertrauliche Dokumente in einer Fremdsprache zu verschaffen (Koehn 2020: 21).

Die Dissemination, oder Verbreitung, zielt darauf ab, durch maschinelle Übersetzung publikationsreife Texte in anderen Sprachen zu erzeugen. Wie bereits erwähnt, sind gängige MÜ-Systeme für die meisten Fachbereiche und Textsorten nicht in der Lage, FAHQT zu erzielen. Bei der Dissemination geht es also um das Erzeugen eines publikationsreifen Textes mit hoher Qualität, der von Menschen vor- oder nachbearbeitet werden muss. Um die Qualität der rohen maschinellen Übersetzung zu verbessern, kann ein Post-Editing durchgeführt werden, welches im aktuellen Entwicklungsstand von MÜ-Systemen einen der größten Beschäftigungsbereiche für Übersetzer*innen darstellt (Forcada 2010: 217). Die Wichtigkeit von menschlicher Nachbearbeitung der Rohübersetzungen wurde bereits bei der Erforschung der ersten MÜ-Systeme erkannt und immer wieder aufgegriffen und scheint nun eine wichtige Rolle in der weiteren Entwicklung des Übersetzungsberufs darstellen (Koehn 2020: 23ff). Die Anschaffung und Betreuung des Systems sowie die anfallenden Kosten für Post-Editing der Texte sind ebenfalls wichtige Faktoren, die sich eventuell erst nach einer gewissen Anwendungsdauer rentieren (Kurz 2014: 421).

Den dritten großen Anwendungsbereich der MÜ stellt die Kommunikation dar, die das maschinelle Übersetzen von digitaler Kommunikation wie E-Mails, Diskussionen oder Beiträgen in Sozialen Medien umfasst. Kommunikation innerhalb einer Sprache oder über Sprachgrenzen hinweg, egal ob mündlich oder schriftlich, ist ein großer Bestandteil des Internets, auf den viele Nutzer*innen angewiesen sind. Auf rein schriftlicher Ebene können neben dem Übersetzen der *Google* Suchergebnisse auch Chaträume, Forumsbeiträge, Support- und Hilfsportale und Beiträge auf sozialen Medien durch das Einbinden von MÜ-Systemen auf diese Seiten scheinbar per Mausklick übersetzt werden (Koehn 2020: 23f). Mit der Einführung cloudbasierter MÜ-Systeme haben sich auch neue multimodale Anwendungsbereiche ergeben, etwa das maschinelle Übersetzen von automatisch erzeugten oder professionell erstellten Untertitel für Videos oder Nachrichtenbeiträge (Koehn 2020: 28). Mit dem wachsenden Interesse von Forscher*innen und Konsument*innen an Spracherkennung und Spracheingabe werden auch MÜ-Systeme um diese Funktionen erweitert. In Kombination mit anderen Technologien, wie etwa der Spracherkennung, *Text-to-*

Speech oder *Speech-to-Text* Programmen, kann die maschinelle Übersetzung mehrsprachige Kommunikation ohne merkbare Zeitverzögerung ermöglichen. *Skype* bietet mit dem *Skype Translator* beispielsweise die Möglichkeit, Sprachanrufe mittels Spracherkennung, maschineller Übersetzung und Sprachsynthese in quasi Echtzeit zu dolmetschen. Auch *SYSTRAN Translate* und *Google Translate* bieten mittlerweile Spracheingabe an und auch Chatdienste wie *WeChat* haben bereits integrierte Dolmetschfunktionen für Sprachanrufe (Poibeau 2017: 104).

Im Rahmen dieser Arbeit soll auch die Nutzer*innengruppe der Studierenden an translationswissenschaftlichen Hochschulen, beziehungsweise angehende Übersetzer*innen, genannt werden. Ihr Zugang zu MÜ-Systemen ist durch alle drei Anwendungsbereiche geprägt, wie aus einer Studie von Heinisch und Lušicky (2019: 45f) hervorgeht, die sich mit den Erwartungen von angehenden Übersetzer*innen an MÜ beschäftigt. Dabei gaben 93% der Befragten an, dass sie MÜ im Laufe ihrer Ausbildung nutzten, 41% taten dies wöchentlich, 15% monatlich und 19% mehrmals pro Jahr (Heinisch & Lušicky 2019: 44). Die zwei beliebtesten Systeme bei den angehenden Übersetzer*innen waren *DeepL* (69%) und *Google Translate* (59%). Zu den Hauptgründen für die Nutzung von MÜ zählen Zeitersparnis (69%), *gisting* (66%), Heranziehen einer Referenzübersetzung (55%) und das Vermeiden repetitiver Arbeit (31%). Frei verfügbare MÜ-Systeme sind somit bereits ein großer Bestandteil von Übersetzungsausbildungen, und auch wenn sie nicht immer Teil der offiziellen Curricula sind, so herrscht bei angehenden Übersetzer*innen scheinbar bereits eine große Bereitschaft, diese Systeme in ihre Arbeitsprozesse zu integrieren.

2.1.3.1 Maschinelle Übersetzung in der Sprachindustrie

Auf dem Markt der Sprachindustrie war der Bereich der MÜ laut *TAUS*-Bericht (van der Meer & Ruopp 2014) im Jahr 2014, also bereits vor Einführung neuronaler Systeme, rund 250 Mio. US-Dollar wert. Durch intensives Forschungsinteresse und rapide Fortschritte ist dieser Bereich in den letzten sieben Jahren stark gewachsen und ist laut *Global Market Insights*-Bericht (Wadhwani & Gankar 2021) mit Stand 2021 650 Mio. US-Dollar wert. Den größten Marktanteil haben dabei Nordamerika mit 21,5% und Europa mit 30%. In Nordamerika entfallen etwa 30,9 Mio. US-Dollar des Marktanteils auf NMÜ, 61,4 Mio. auf SMÜ und 72,2 Mio. auf RMBÜ. In den nächsten sechs Jahren wird von einer Verdreifachung dieser Summen ausgegangen. Bis zum Jahr 2027 wird weltweit für die MÜ mit einem Marktwachstum von 25% gerechnet, was zu einem Marktwert von rund 3 Mrd. US-Dollar führen wird (Wadhwani & Gankar 2021). Zu den größten internationalen Unternehmen in der MÜ-Entwicklung zählen unter anderem *Alibaba Cloud, Cloudwords Inc., Omniscien Technologies, Tencent*

Cloud TMT und *Yandex.Cloud LLC* (Wadhwani & Gankar 2021). Auch die Ergebnisse von *Nimdzi 100 (*Hickey & Agulló García 2021), einer jährlichen Befragung der 100 größten Sprachdienstleister*innen weltweit, deuten auf ein starkes Wachstum von MÜ und PE im Bereich der Sprachdienstleistungen hin. Zu den weltweit größten Sprachdienstleister*innen zählen *TransPerfect, Lionbridge, LanguageLine Solutions* und *SDL*, die einen Jahresumsatz zwischen 500 und 850 Mio. US-Dollar und ein jährliches Wachstum von 6 bis 10% verzeichnen. Die globale Sprachindustrie wächst jährlich um etwa 6% und nahm 2020 55 Mrd. US-Dollar ein, für 2021 wurden 58,3 und für 2025 73 Mrd. US-Dollar prognostiziert. Die häufigsten Dienstleistungen von Sprachdienstleister*innen waren 2020 Übersetzen und Lokalisieren mit 97,5%, MÜ und PE mit 71,5% und Untertitelung mit 68,4% (Hickey & Agulló García 2021). Verglichen mit ähnlichen Befragungen von vor fünf Jahren hat die MÜ als Disziplin stark zugelegt und wird auch in kommenden Jahren voraussichtlich an Wichtigkeit gewinnen (Hickey & Agulló García 2021). Abgesehen von solchen Befragungen sind die genauen Zahlen und Umsätze der Sprachindustrie und somit der MÜ nur schwer berechenbar. Die *Generaldirektion Übersetzung* (DGT) der Europäischen Kommission hat beispielsweise ihr internes Übersetzungsbudget 2013 mit 330 Mio. Euro angegeben und jährlich über zwei Millionen übersetzte Seiten verzeichnet, davon 93% mittels Humanübersetzung (Poibeau 2017: 97).

2.1.3.2 Gängige maschinelle Übersetzungssysteme

Frei verfügbare, beziehungsweise kostenlose Online-MÜ-Systeme traten erstmals in den 1990er Jahren mit Veröffentlichung des Dienstes *Babel Fish* auf, der auf der Technologie von *SYSTRAN* basierte und 2012 durch den *Bing Translator* ersetzt wurde. Derartige Online-Systeme haben in den letzten 15 Jahren ein starkes Wachstum erlebt und der wohl erfolgreichste Dienst wurde 2006 mit *Google Translate* eingeführt. *Google Translate* basierte bei seiner Einführung auf einem von *IBM* entwickelten statistischen Ansatz, wurde im Laufe der Jahre jedoch deutlich verbessert und profitiert auch von *Googles* stets wachsendem Suchmaschinenverzeichnis (Johnson et al. 2017: 345f). *Google Translate* verzeichnete im Mai 2016, wohlgemerkt kurz vor Umstellung auf die neuronalen Architektur, 500 Millionen Nutzer*innen und 143 Milliarden übersetzte Wörter pro Tag (Way 2018: 3f). Als Suchmaschinenbetreiber hat *Google* Zugriff auf die größte Menge an Sprachdaten und Bitexten, jedoch sind diese meist nicht verifiziert und die Qualität daher durchmischt. Der Betreiber des NMÜ-Systems *DeepL* verfügt hingegen zusätzlich über den Online-Übersetzungsspeicher *Linguee* und kann somit große Mengen an qualitativen, institutionsnahen Bitexten für das Training seines NMÜ-Systems nutzen (Schmalz 2019:

203). Andere freie Online-MÜ-Systeme sind der bereits erwähnte *Bing Translator*, *SYSTRAN Translate*, *PROMPT.One Translator oder Yandex.Translate* (Poibeau 2017: 99).

Parallel zu den freien Systemen existieren auch kommerzielle Systeme in Form von Abonnements, oftmals gibt es hier beschränkte aber kostenlose Testversionen sowie monatlich lizenzierbare kostenpflichtige Versionen mit verschiedenen Umfängen. Viele dieser Produkte werden als *Software as a Service* angeboten und sind cloudbasiert, wodurch Benutzer*innen nicht durch die Kapazität ihres Rechners beschränkt sind (Poibeau 2017: 98). Mit der Einführung von NMÜ-Systemen ab 2016 wurde das Online-Angebot vergrößert und Firmen wie *Globalese* konnten sich mit ihren für große Unternehmen maßgeschneiderten NMÜ-Systemen in diesem neuen Feld etablieren (Poibeau 2017: 101). Zu den Nutzer*innen dieser kommerziellen Systeme zählen Firmen, Unternehmen, politische sowie staatliche Institutionen, die maßgeschneiderte Systeme nutzen und diese mit gepflegten Datenbeständen an Altübersetzungen und Terminologien zur Entlastung der Übersetzer*innen in großen, fortlaufenden Projekten mit engen Zeitplänen einsetzen (Schmalz 2019: 196).

2.1.3.3 Limitationen der maschinellen Übersetzung

SMÜ und NMÜ-Systeme weisen ähnliche Fehlertypen auf, wenn es um Inkonsistenzen bei Zahlen oder Terminologie und um fehlerhafte Übersetzungen von Eigennamen geht. Einige NMÜ-spezifische Fehler sind oft semantischen Ursprungs und können dazu führen, dass kontextrelevante Wörter falsch übersetzt werden, was eine Verfälschung der Aussage des gesamten Satzes mit sich bringen kann. Auch im Umgang mit unbekannten oder seltenen Wörtern, also Wörtern, die das System in der Trainingsphase nicht erlernt hat, da sie nicht im Parallelkorpus waren, können Fehler auftreten. Systeme die mit *sub-word units*, also Zeichenfolgen, arbeiten, können durch das Zusammensetzen dieser Zeichenfolgen sogenannte Nichtwörter erzeugen (Forcada 2017: 303). Diese Wörter orientieren sich an den lexikalischen Konventionen der Zielsprache und bestehen beispielsweise aus zwei Worthälften, die separat gesehen auch existieren und Sinn ergeben, in Kombination jedoch nicht. Dieser Fehlertyp trat bei SMÜ-Systemen nicht auf, da ihre Funktionsweise eine solche Erzeugung von Wörtern nicht erlaubt, stattdessen wären die unbekannten Wörter ausgelassen worden. Dadurch waren diese Fehler beim Post-Editing auch leicht erkennbar, bei NMÜ-Systemen setzt die teilweise hohe Plausibilität der neu erstellten Wörter von den Nachbearbeitern eine hohe sprachliche Kompetenz voraus, um diese zu identifizieren (Forcada 2017: 304). Auch wenn durch die NMÜ bereits beachtliche Fortschritte gemacht wurden, sind MÜ-Systeme generell nicht für alle Textsorten geeignet. Technische Texte mit strukturhafter Syntax, festgelegtem Vokabular und instruktivem Charakter wie etwa

Gesetzestexte, Verträge oder Finanzberichte eignen sich gut und können mit geringem Post-Editing-Aufwand publiziert werden. Vor allem Texte, die unter Einsatz einer kontrollierten Sprache, also einer Sprache, die in ihrem lexikalischen und syntaktischen Umfang begrenzt ist, können von MÜ-Systemen gut übersetzt werden. Literarische Text oder Marketingtexte, die durch kreativen und metaphorischen Sprachgebrauch, eine variierte Syntax und andere Stilmittel wie Wortspiele bestechen, können selbst von neuronalen Systemen nicht mit zufriedenstellender Qualität übersetzt werden (Kurz 2014: 418f). Um jedoch überhaupt übersetzen zu können, müssen korpusbasierte MÜ-Systeme zuerst trainiert werden, was eine hohe Rechenleistung mit eigener Hardware, vor allem Grafikprozessoren, voraussetzt. Einige frei verfügbare NMÜ-Toolkits, die mit eigenen Korpora trainiert werden können, wären *OpenNMT, Nematus, AmuNMT* oder *Neural Monkey*. Im Gegensatz zu vergleichbaren freien SMÜ-Toolkits sind diese Kits jedoch wegen der Anforderungen an Hardware und Wissen der Anwender*innen nicht für Übersetzer*innen geeignet (Forcada 2017: 302). Selbst ein fertig trainiertes, neuronales System könnte von den meisten Desktop-PCs nicht in Echtzeit ausgeführt werden. Daher bieten einige Firmen wie *Globalese* cloudbasierte Lösungen an, die vom Endgerät des Nutzer*innen unabhängig sind, jedoch gesichert auf internen Servern des Anbieters laufen.

Neben den technischen Limitationen von MÜ-Systemen im Laufe ihrer Entwicklung besteht für die gängigen korpusbasierten, statistischen oder neuronalen Systeme auch noch die Problematik bezüglich der Parallelkorpora. Größe und Qualität der Parallelkorpora, die für das Training der korpusbasierten MÜ-Systeme verwendet werden, sind der wichtigste Faktor für ein Funktionieren der Systeme. Dabei gilt, dass ein MÜ-System nur so gut wie sein Trainingsmaterial sein kann. Werden die Systeme anhand von Korpora mit schlechter Qualität trainiert, so wird dies als *garbage in-garbage out,* oder *GIGO*, bezeichnet, da die schlechte Qualität der Korpora auch auf die Qualität des Translats übergeht (Clark 1972: 97). Die Bitexte der Parallelkorpora sollen daher aus seriösen und offiziellen Quellen stammen, wie etwa aus dem Umfeld von Institutionen wie der *EU* oder *UNO* (Schmalz 2019: 203). Ein Großteil der verfügbaren Parallelkorpora zu Trainingszwecken enthält Englisch als eine der beiden Sprachen, da viele zwei- oder mehrsprachige Quellen im Internet meist in Kombination mit dieser Sprache verfügbar sind. Korpora für Sprachen mit kleinen Sprechergemeinschaften und geringer Verfügbarkeit an Textmaterial sind jedoch eine Hürde für die korpusbasierte MÜ (Stein 2018: 14). Manche Firmen und Übersetzungsbüros können zum Erstellen von Parallelkorpora auf alte Übersetzungen ihrer Kund*innen zurückgreifen und somit kunden- oder domänenspezifische MÜ-Systeme trainieren. Auch das Internet kann

als größte Ressource für mehrsprachige Texte herangezogen werden, jedoch sind diese Texte meist nicht explizit als Korpora oder Bitexte angegeben oder formatiert (Poibeau 2017: 48f).

Mit der rapiden Entwicklung und hohen Leistungsfähigkeit von NMÜ-Systemen kam es zu großem Enthusiasmus bezüglich des zukünftigen Potentials dieser Systeme. Vor allem Online-MÜ-Systeme sind schnell, gratis und benutzerfreundlich, wodurch es leicht ist, die damit verbundenen Risiken zu übersehen. Zum einen gibt es Cyber-Risiken, die vor allem cloudbasierte Dienste wie *DeepL* betreffen können, da persönliche und vertrauliche textuelle Informationen so ins Internet oder in falsche Hände gelangen können. Die oben besprochene Problematik der Verfügbarkeit von Korpora hat mit Einführung von frei verfügbaren Online-MÜ-Systemen eine neue Dimension angenommen, da durch jede Eingabe eines Textes oder Satzes von Nutzer*innen quasi unbewusst Trainingsmaterial für die Systeme zugesteuert wird. Dienste wie *Google Translate* schreiben sogar explizit in ihre Geschäftsbedingungen, dass man mit Verwenden des Dienstes jedes ausgangssprachliche Element (Wort, Satz, Text) als Lizenz an *Google* übergibt, der über diesen frei verfügen kann und ihn auch zur Entwicklung nutzen oder öffentlich publizieren darf. Bei der Benutzung von Online-MÜ-Systemen sollte also immer Folgendes bedacht werden: „If the service is free, you are the product" (Serra & Weyergraf 1980: 104). Besonders kritisch ist es, wenn professionelle Übersetzer*innen vertrauliche Dokumente in derartige Online-Systeme kopieren, ohne die Kund*innen darüber zu informieren (Canfora & Ottmann 2020: 63f). Andere Risiken beinhalten finanzielle oder physische Schäden, die durch Fehlübersetzungen von MÜ-Systemen ausgelöst wurden, sowie die Frage nach der Verantwortlichkeit in solchen Fällen, da Maschinen rechtlich nicht verantwortlich gemacht werden können (Canfora & Ottmann 2020: 59). NMÜ-Anbieter wie *DeepL* oder *Google* schließen eine Haftung im Schadensfall komplett aus. Laut aktueller EU-Gesetzgebung bezieht sich Haftung immer auf menschliches Verhalten und setzt im Falle eines fehlerhaften Textes einen Autor oder eine Autorin voraus. Da MÜ-Systeme nicht als juristische Person gelten, sind sie auch von dieser Regelung ausgenommen. Alternativ gibt es in der EU-Kommission Bestrebungen, dass jegliche Produzent*innen, Besitzer*innen oder Nutzer*innen dieser Systeme im Schadensfall die Haftung übernehmen (Canfora & Ottmann 2020: 62f).

2.1.4 Ausblick über die weitere Entwicklung

Die Entwicklung von NMÜ-Systemen ab 2016 löste großen Enthusiasmus in manchen Kreisen aus und wurde als Paradigmenwechsel bezeichnet. Ähnlich wie bei den Reaktionen auf das Georgetown-IBM-Experiment der 1950er Jahre warnten einige Expert*innen mit der Einführung von NMÜ jedoch vor einem „peak of inflated expectations“ (Canfora & Ottmann 2020: 63). Die anfänglichen Erfolge dürfen nicht über die Risiken und Nachteile dieser neuartigen Systeme hinwegtäuschen und auch Aussagen bezüglich der Fortschrittlichkeit der Systeme sind mit Vorsicht zu genießen. Mit Veröffentlichung seiner NMÜ-Architektur behauptete der Suchmaschinenbetreiber *Google* etwa, dass in einigen Sprachkombinationen *human parity* erzielt wurde, also „der (qualitative) Gleichstand zwischen dem Ergebnis einer durch ein Stück Software generierten Übersetzung und dem durch einen menschlichen Profi-Übersetzer erzeugten“ (Porsiel 2020: 2). Auf diese Behauptungen folgte harsche Kritik von Übersetzer*innen, die die Messbarkeit von *human parity* anzweifelten und die Firmen dazu brachten, ihre Aussagen zu relativieren. Zeitgleich änderte sich durch solche Aussagen jedoch auch die Einstellung vieler Auftraggeber*innen zum Übersetzen, sie erwarteten „dass sie ab sofort (viel) mehr Text in (viel) mehr Sprachen, in (sehr) viel kürzerer Zeit, zu (sehr) viel geringeren Kosten bei (mindestens) gleichbleibender (Human-)Qualität bekommen können“ (Porsiel 2020: 5). Wie in Abb. 2 zu sehen ist, sinkt durch den wirtschaftlichen Druck auch das Qualitätsniveau, das Auftraggeber*innen bereit sind zu akzeptieren (Moorkens 2017: 470).

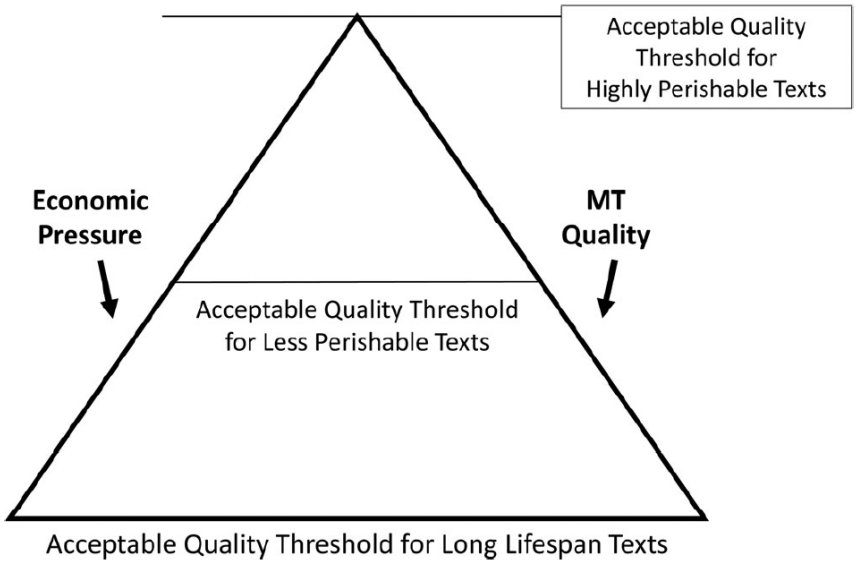


Abbildung 2: Akzeptable Qualitätsniveaus der MÜ (Moorkens 2017: 469)

Dem Enthusiasmus von MÜ-Entwickler*innen und Auftraggeber*innen steht die Skepsis vieler Übersetzer*innen gegenüber, die eine Bedrohung ihres Berufes fürchten (Porsiel 2020: 4). Ein weiterer Faktor, der oft übersehen wird, ist die Abhängigkeit der Maschine vom Menschen, da auch NMÜ-Systeme anhand von Parallelkorpora trainiert werden müssen, deren Bitexte zuerst von menschlichen Übersetzer*innen anzufertigen sind. Dies wirft nicht nur die Frage auf, wie mit Übersetzungsdaten in Bezug auf Urheberrecht umgegangen werden soll, sondern auch, welche Rolle der menschliche Übersetzer*innen in der weiteren Entwicklung der NMÜ spielen wird (Moorkens 2017: 472). Einen Mittelweg stellt zuletzt noch das Post-Editing dar, bei dem sowohl technisches Wissen über die Funktionsweise von MÜ-Systemen und deren typische Fehler als auch die sprachliche Expertise beim Editieren der Rohübersetzung wichtig sind (Moorkens 2017: 473). Das Berufsfeld des Post-Editings erlebt aktuell ein starkes Wachstum und stellt viele Übersetzer*innen vor die Entscheidung sich umzuschulen oder an der Entwicklung der NMÜ mitzuarbeiten (Moorkens 2017: 471).

Wie bei jeder neuen Technologie ist auch bezüglich der MÜ ein gewisses Maß an Skepsis und Kritik angebracht, jedoch liegt es an den Übersetzer*innen sich mit diesen neuen Technologien auseinanderzusetzen und sie in den eigenen Arbeitsprozess zu integrieren, um im Idealfall Kosten und Zeit zu sparen (Porsiel 2020: 3). Freiberufliche Übersetzer*innen und große Sprachdienstleister*innen sehen sich mit den hohen Erwartungen von Auftraggeber*innen bezüglich der neuen Technologien konfrontiert und müssen ausreichend Wissen und passende Argumente parat haben, um diese Erwartungen zu dämpfen und realisierbare Alternativen vorzuschlagen, mit deren Ergebnisse beide Parteien schlussendlich zufrieden sind (Porsiel 2020: 6). Die kommenden Jahren werden vor allem durch Anpassung bestehender oder Entwicklung neuer Übersetzungsworkflows geprägt sein sowie den Aufbau von Fachpersonal in Form von Post-Editor*innen und Computerlinguist*innen, die die computergestützten Abläufe und MÜ-Systeme betreuen und sinnvoll nutzen können (Porsiel 2020: 8f). Mit dem momentanen Entwicklungsstand der NMÜ ist also davon auszugehen, dass bilinguale Sprachprofis und Übersetzer*innen vor, während und nach Prozessen der maschinellen Übersetzung nach wie vor gebraucht werden und durch ihre Beteiligung das Funktionieren solcher Systeme überhaupt erst ermöglichen. „Maschinelle Übersetzungssysteme können nur von Anwendern sinnvoll eingesetzt werden, die mehr wissen als das System. Ihr größter Markt liegt jedoch im Bereich von Anwendern, die ein System wünschen, das mehr weiß als sie selbst.“ (Schubert 1999: 429) Um das volle Potenzial von MÜ-Systemen auszuschöpfen, ist also eine fachkundige Betreuung sowie eine korrekte

Nutzung aller verfügbarer Ressourcen notwendig. Auch wenn der Übersetzerberuf in den kommenden Jahrzehnten durch Aufkommen neuer Berufsfelder wie dem Post-Editing große Veränderungen durchlaufen wird, so wird diese Zeit nicht durch eine Vormachtstellung der MÜ, sondern durch eine Koexistenz und Kollaboration von Übersetzer*innen und MÜ-Systemen geprägt sein (Kurz 2014: 399).

2.2 Der translatorische Qualitätsbegriff

In Hinblick auf die Verwendung mehrerer Metriken zur Evaluierung der Qualität maschineller Übersetzungssysteme im empirischen Teil dieser Arbeit soll in diesem Kapitel auf die unterschiedliche Auffassung von Qualität in den Bereichen der maschinellen Übersetzung und Humanübersetzung eingegangen werden.

2.2.1 Auffassungen von Übersetzungsqualität

Der Qualitätsbegriff wird in den Bereichen der Sprachindustrie und der Translationswissenschaft nach wie vor intensiv diskutiert, zumal in beiden Bereichen unterschiedliche Vorstellungen darüber herrschen (Castilho et al. 2018: 16). Während in der Translationswissenschaft nach objektiv mess- und quantifizierbaren Merkmalen für Qualität gesucht wird, wird der Qualitätsbegriff im praktischen Alltag der Sprachindustrie viel flexibler gehandhabt, da er hier sehr stark durch das subjektive Qualitätsempfinden von Auftraggeber*innen definiert wird (Bruhn 2019: 37). Diesem wertorientierten Qualitätsbegriff, der nach dem Preis-Leistungs-Verhältnis eines Produktes, also einer Übersetzung, beurteilt wird, steht der textorientierte Qualitätsbegriff gegenüber, der das Ziel der wissenschaftlichen Forschung verfolgt. Somit wird deutlich, dass in Bezug auf Übersetzungsqualität nicht mit universalen Definitionen, wie jener aus der Norm ISO 9000 (2015a) gearbeitet werden kann, die Qualität als „realisierte Beschaffenheit einer Einheit bezüglich der Qualitätsforderung“ (Bruhn 2019: 33) definiert. Um beide Bereiche abzudecken, bieten Koby et al. (2014) eine breit- und eine enggefasste Definition der Übersetzungsqualität an, die den Auffassungen der Sprachindustrie und Translationswissenschaft entsprechen sollen.

Die breite Definition der Übersetzungsqualität bezieht sich auf die Dienstleistungsebene, vertritt das ausgangstextorientierte Konzept der Adäquatheit und geht davon aus, dass die Vorgaben für eine Übersetzung vor deren Beginn von Auftraggeber*in und Sprachdienstleister*in explizit besprochen werden sollen (Castilho et al. 2018: 16). „A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider,

taking into account end-user needs." (Koby et al. 2014: 416) Die enge Definition der Übersetzungsqualität bezieht sich auf die Textebene, vertritt das zieltextorientierte Konzept der *Fluency*, des Sprachflusses, und geht davon aus, dass das Festlegen von Vorgaben überflüssig ist, da Auftraggeber*innen oft nicht in der Lage sind richtig einzuschätzen, welche Schritte seitens der Sprachdienstleister*innen nötig sind (Castilho et al. 2018: 16).

> A high-quality translation is one in which the message embodied in the source text is transferred completely into the target text, including denotation, connotation, nuance, and style, and the target text is written in the target language using correct grammar and word order, to produce a culturally appropriate text that, in most cases, reads as if originally written by a native speaker of the target language for readers in the target culture. (Koby et al. 2014: 416f)

Als Begründer der funktionsorientierten Übersetzungstheorie entwickelten Reiß und Vermeer (1984) die kultur- und sprachenpaarunabhängige Skopostheorie, aus der sich auch einige Qualitätsmerkmale für eine Übersetzung ableiten lassen. Zentral für diese Theorie ist das bewusste und von einer Motivation getriebene translatorische Handeln einer Person, um einen bestimmten Zweck zu erfüllen. Anders formuliert ist „die Dominante aller Translation [...] deren Zweck" (Reiß & Vermeer 1984: 96), also die Funktion und Wirkung, die eine Übersetzung auf Empfänger*innen in der Zielsprache und -kultur hat. Die Übersetzungsqualität kann dabei anhand von Adäquatheit, also dem Grad an inhaltlicher Äquivalenz zwischen Ausgangs- und Zieltext (Nord 2006:22) sowie anhand von zieltextlicher *Fluency* gemessen werden (Reiß & Vermeer 1984: 140). Übersetzungsqualität soll dabei nicht auf die absolute, perfekte, sondern auf die im Rahmen des Auftrags maximal mögliche Adäquatheit und *Fluency* abzielen (Koby et al. 2014: 415). Sprachdienstleister*in sollten von Auftraggeber*innen auch Zielgruppe und Zweck der Übersetzung erfragen, da diese Faktoren maßgeblich in den Arbeitsprozess und das Qualitätsverständnis einfließen (Koby et al. 2014: 415). Die Wichtigkeit des Zwecks für Übersetzungsqualität findet sich auch in der Norm ISO 17100 (2015b: 16): „Der Übersetzer muss in Übereinstimmung mit dem Zweck des Übersetzungsprojekts, einschließlich der sprachlichen Konventionen der Zielsprache und der betreffenden Projektspezifikationen, übersetzen."

Somit lässt sich aus den diversen Definitionen zusammenfassen, dass Übersetzungsqualität sich zumindest anhand folgender drei Merkmale festmachen lässt: einem zielsprachlich guten Stil sowie grammatikalische Korrektheit, Adäquatheit bezüglich der pragmatischen und semantischen Äquivalenz zwischen Ausgangs- und Zieltext sowie Einhaltung der Projektvorgaben, also dem Zweck der Übersetzung (Chatzikoumi 2019: 138).

2.2.2 Qualitätsmetriken der maschinellen Übersetzung

Im Falle der Qualität von maschinell übersetzten Texten wird meist mit denselben Definitionen gearbeitet, da der Zweck der Übersetzung theoretisch derselbe ist, unabhängig davon mit welchen Mitteln sie erzeugt wurde. Jedoch ist eine solche Gleichstellung von maschineller Übersetzung und Humanübersetzung nicht unproblematisch, da die Endnutzer*innen maschineller Übersetzungen und deren Anforderungen stark variieren können. Von *gisting* eines Textabschnitts im Internet bis zur Übersetzung eines zu publizierenden Artikels kann die maschinelle Übersetzung alles abdecken, wodurch auch die Qualitätsanforderungen der Endnutzer*innen stark variieren (Chatzikoumi 2019: 138). Daher wurden für MÜ-Systeme eigene Techniken der Qualitätsevaluierung, sogenannte Metriken, entwickelt, die entweder automatisiert oder manuell funktionieren. Automatisierte Metriken nutzen Humanübersetzungen als Referenz für eine automatische Berechnung der Qualität, wohingegen manuelle Metriken die rechnergestützte Evaluierung der Qualität durch Bearbeiter*innen vorsehen (Chatzikoumi 2019: 139). Im Rahmen dieser Arbeit soll in Hinblick auf die Gestaltung des empirischen Teils nur auf die automatisierten Metriken eingegangen werden. Die Anforderungen an diese Metriken sind unter anderem die Einheitlichkeit und die intuitive Interpretation ihrer Ergebnisse, ihre Geschwindigkeit und Verlässlichkeit, ein großer Anwendungsbereich und, am allerwichtigsten, ihre Korrelation mit menschlichen Beurteilungen (Chatzikoumi 2019: 140).

Zum einen sind hier auf Referenzübersetzungen basierende Metriken zu nennen, die die Qualität der maschinellen Übersetzung anhand ihrer Ähnlichkeit zu einer Humanübersetzung messen. Eine der ältesten dieser Metriken ist die Editierdistanz nach Levenshtein (1966), welche die Mindestanzahl an Änderungen misst, die benötigt wird, um eine Buchstabenfolge *a* in eine Buchstabenfolge *b* zu verwandeln. Als Änderungen werden hier Löschen, Einfügen und Ersetzen verstanden (Chatzikoumi 2019: 140). Im Bereich der MÜ-Evaluierung sind hier die Metriken *Word Error Rate* (WER) (Nießen et al. 2000) und die davon abgeleitete *Translation Edit Rate* (TER) (Dorr & Snover 2006) zu nennen. Die WER stammt ursprünglich aus dem Bereich der automatischen Spracherkennung und misst das Verhältnis der Anzahl aller Änderungen im MÜ-Output zur Wortanzahl der Referenzübersetzung (Castilho et al. 2018: 25). Die TER wird anhand der Anzahl an Änderungen in der maschinellen Übersetzung gemessen, die nötig ist, um diese mit der Referenzübersetzung übereinstimmen zu lassen und dann durch die durchschnittlichen Wortanzahl aus der Referenzübersetzung geteilt. Als Änderung gelten hier Löschen,

Einfügen, Verschieben und Ersetzen. TER bietet gute Korrelationen mit menschlichen Beurteilungen (Dorr & Snover 2006: 225).

Neben der Editierdistanz gibt es auch auf *Precision und Recall* basierende Metriken, die anhand von *n*-grams funktionieren und in zwei Stufen berechnet werden. Unter Precision versteht man das Verhältnis an korrekten *n*-grams, die in mindestens einer Referenzübersetzung auftreten, innerhalb des MÜ-Outputs. Für einen übersetzten Satz mit 10 *n*-grams, wovon 6 als akzeptabel gelten, wäre die Precision also 6/10. Unter Recall versteht man das Verhältnis an korrekten *n*-grams, die in mindestens einer Referenzübersetzung auftreten. Ist ein Satz der Referenzübersetzung also 15 *n*-grams lang, so ist der Recall 6/15 (Chatzikoumi 2019: 141). Die verbreitetste Precision und Recall Metrik ist die 2001 entwickelte *Bilingual Evaluation Understudy* (BLEU) (Papineni et al. 2002), welche sprachpaar- und domänenunabhängig ist. Maschinelle Übersetzungen werden von BLEU durch den Vergleich mit einer Human-Referenzübersetzung anhand ihrer Adäquatheit und *Fluency* bewertet. Dabei wird der maschinellen Übersetzung eine höhere Qualität zugeschrieben, je näher sie an der Referenz ist und dies mittels eines Werts zwischen 0 und 1 ausgedrückt, wobei 1 die höchstmögliche Qualität darstellt (Chatzikoumi 2019: 141). BLEU liefert im Vergleich zu anderen Metriken jedoch weniger verlässliche Werte auf Satzebene (Agarwal & Lavie 2008: 115), benötigt zudem eine große Menge an Referenzübersetzungen (Dorr & Snover 2006: 224) und wird dafür kritisiert, dass die Recall-Komponente bei den Berechnungen nicht ausreichend berücksichtigt wird. Weitere nennenswerte Precision und Recall Metriken sind *General Text Matcher* (GTM) (Melamed et al. 2003) und *Metric for Evaluation of Translation with Explicit Ordering* (METEOR) (Banerjee & Lavie 2005). METEOR wurde mit dem Ziel entwickelt, Korrelationen mit menschlichen Beurteilungen auf Satzebene zu verbessern. Es berechnet einen Wert anhand expliziter Wort-zu-Wort Übereinstimmungen zwischen der maschinellen Übersetzung und einer oder mehreren Referenzübersetzungen (Agarwal & Lavie 2008: 115). Für Wörter ohne Übereinstimmungen und Änderungen der Syntax gibt es Abzüge in der Bewertung, wodurch diese Metrik so gut mit menschlichen Beurteilungen übereinstimmt (Dorr & Snover 2006: 224).

Auf Referenzübersetzungen basierende Metriken sind kostengünstig, schnell und lassen sich stets wiederverwenden, was vor allem in der Entwicklungsphase von MÜ-Systemen zur Feinabstimmung hilfreich ist (Castilho et al. 2018: 26). Gleichzeitig entsteht durch die Verwendung von Referenzübersetzungen die Problematik, dass selbst gute maschinelle Übersetzungen schlecht bewertet werden, wenn sie stärker von der Referenz abweichen (Dorr & Snover 2006: 230). Diese Metriken überschätzen somit die tatsächliche

Fehlerrate in der Übersetzung und sind durch eine fehlende Flexibilität charakterisiert, da sie auf Ebene von Zeichenketten arbeiten und daher darauf angewiesen sind, wortwörtliche Übereinstimmung zwischen den zwei Übersetzungen zu finden (Agarwal & Lavie 2008: 116). Auch die Interpretation der Evaluierungsergebnisse gestaltet sich als schwierig und gibt wenig Aufschluss über die exakten Stärken und Schwächen eines MÜ-Systems (Chatzikoumi 2019: 144). Im Bereich des Post-Editing eignen sich referenzbasierte Metriken wie BLEU, METEOR und TER auch dafür, um Vorhersagen bezüglich der Qualität des MÜ-Systems sowie dem mit PE verbundenen temporalen Aufwand zu treffen. Diese Vorhersagen können Sprachdienstleister*innen dabei helfen, sich für oder gegen die maschinelle Übersetzung eines Textes zu entscheiden, wenn dieser im Vergleich zur Humanübersetzung besser oder schlechter abschneiden würde (Castilho et al. 2018: 28).

Neben den auf Referenzübersetzungen basierenden Metriken existieren auch noch *Quality Estimation* (QE) Metriken die dazu dienen, die Qualität eines MÜ-Systems anhand eines Inputs abzuschätzen, ohne dass dabei ein Output vorhanden ist. Auch wenn QE-Metriken streng genommen keine Qualitätsmetriken sind, so können sie zum Einsatz kommen, wenn keine Referenzübersetzungen vorhanden sind, oder auch bei der Auswahl des besten MÜ-Systems aus mehreren Kandidaten, wofür referenzbasierte Metriken ebenfalls ungeeignet sind. Moderne QE-Metriken arbeiten teils neuronal und sind daher auch in der Lage auf Satz- anstatt auf Wortebene zu arbeiten sowie eine automatische Alignierung von Übersetzungen vorzunehmen (Chatzikoumi 2019: 145).

2.3 Eyetracking - Funktionsweise und Anwendungen

In den vergangenen Jahren haben sich Eyetracker vor allem im Forschungsbereich der *Human-Computer-Interaction* als wertvolles Mittel zur Datenerfassung erwiesen. Bei der Verwendung von Computern ist neben der physischen Interaktion mittels Tastatur und Maus vor allem die visuelle Interaktion über den Bildschirm ein Bereich, der viel Aufschluss über Ablauf und Ergonomie dieser modernen Arbeitsweise bietet. Zudem ist das Sehen der wichtigste Sinn, da der Mensch im Schnitt 87% aller Informationen mit den Augen wahrnimmt und dies mit einem Informationsfluss von 2×108 bit/s geschieht (Nauth 2012: 18). Das Gerät für die Messung von solchen Augenbewegungen wird als *Eyetracker* bezeichnet. Die damit verbundene Disziplin wird folglich als *Eyetracking*, zu Deutsch auch Blickerfassung oder Okulographie, bezeichnet. (Nauth 2012: 17). In seiner modernen Form dient der Eyetracker dazu, die Augenbewegungen eines/einer Betrachter*in auf einem Bildschirm zu messen. Hierzu wird am häufigsten ein videobasierter Cornea-Reflektion Eyetracker verwendet (Duchowski 2017: 49).

Solche Eyetracker bestehen aus einer Kombination aus Infrarot- und Video-Okulographie (Nauth 2012: 25f) und messen Augenbewegungen primär anhand der Cornea-Reflektion-Methode. Dafür wird eine fixierte Lichtquelle, etwa Infrarotlicht, auf das Auge gerichtet, welches auf der Hornhaut nur ein gewisses Maß an Licht reflektieren kann, da der Rest vom Glaskörper absorbiert wird, um die Netzhaut zu schützen. Diese Reflektion geschieht 3,5 mm unter der Augenoberfläche und lässt die Pupille auf der Kamera des Eyetrackers als helle Fläche aufscheinen. Um die Pupillenmitte entstehen durch das Auftreffen der Infrarotstrahlen vier Reflektionen, die genaue Messungen der Position der Retina je nach Winkel der Reflektion ermöglichen, wobei jede Veränderungen der Reflektion als Augenbewegung aufgefasst wird (Duchowski 2017: 52). Zusätzlich zu den Koordinaten beider Pupillen auf einer x- und y-Achse wird auch der Abstand der Pupillen zueinander erfasst (Rajs et al. 2016: 116). Zusätzlich zu der Lichtquelle befindet sich im Eyetracker eine Kamera mit einer durchschnittlichen Frequenz von 60Hz oder höher, die das Gesicht des/der Betrachter*in erfasst und Abstand und Höhe der Augen im Gesicht misst (Rajs et al. 2016: 115). Manche Eyetracker benötigen dafür einen daumengroßen Sticker mit Zielscheibenmuster, der auf die Stirn des/der Betrachter*in etwa zwei bis drei Zentimeter oberhalb der Augenbrauen mittig angebracht werden muss und als Referenzpunkt dient (SR Research 2021). Wurde die Kamera korrekt kalibriert, zeichnet sie die rohen Augenbewegungen auf, welche dann digital verarbeitet und mit den gemessenen Cornea-Reflektionen kombiniert werden, um die Augenbewegungen relativ zur Bewegung des Kopfes darzustellen (Duchowski 2017: 57). Jegliche Form der Bewegung oder Nicht-Bewegung des Auges wird folglich als *Point Of Regard* (POR), zu Deutsch Augenposition, bezeichnet und stellt die Position des Auges auf dem Bildschirm zu jeder beliebigen Zeit dar (Duchowski 2017: 85).

Durch den Eyetracker können verschiedenste Augenbewegungen aufgezeichnet werden, die vier wichtigsten sollen hier kurz erläutert werden. *Fixationen* sind Vorgänge mit einer Dauer von 150-600 ms (Millisekunden), bei denen die Retina stabilisiert und auf ein Objekt im visuellen Umfeld gerichtet wird (Nauth 2012: 22). Hierbei wird gemessen, wie lange der/die Betrachter*in mit seinem/ihrem Blick auf diesem Objekt verweilt (Duchowski 2017: 44). Rund 90% aller visuellen Wahrnehmungen sind Fixationen. *Glatte Augenfolgebewegungen* sind Vorgänge, bei denen die Retina den Pfad eines bewegten Gegenstandes mitverfolgt, sofern dieser eine gewisse Geschwindigkeit nicht überschreitet (Duchowski 2017: 43). *Regressionen* sind hingegen Rücksprünge zu bereits fixierten Punkten beim Betrachtungsvorgang (Nauth 2012: 22) Als *Sakkade* wird die Umpositionierung der

Retina auf ein neues Element im visuellen Umfeld definiert. Dieser Vorgang kann bewusst vorgenommen werden oder reflexiv als Reaktion auf einen visuellen Reiz geschehen. Sakkaden sind notwendig, um von einer Fixation, einer Augenfolgebewegung oder einer Regression zur nächsten zu gelangen. Die Dauer der Umpositionierung ist mit zwischen 10 und 100 ms so kurz, dass der/die Betrachter*in während dieses Vorgangs quasi blind ist und keine visuellen Reize aufnehmen kann, bis die Retina auf das neue Element fixiert ist (Duchowski 2017: 40). Dadurch wird eine Reizüberflutung durch unfixierte Elemente verhindert. Aus all diesen visuellen Informationen setzt das Gehirn anschließend ein Bild der Umgebung zusammen, welches aus fixierten Punkten, aber auch dem peripheren Sichtfeld besteht (Nauth 2012: 22). Die oben genannte Augenbewegungen sind für Messungen mit Eyetrackern am wichtigsten, da sie auf freiwillige, offenkundige, visuelle Aufmerksamkeit schließen lassen, da Fixationen, Regressionen und Augenfolgebewegungen dem Verlangen entsprechen, den Blick auf einen bestimmten Bereich zu richten und dort zu halten, um visuelle Informationen zu erfassen (Duchowski 2017: 45).

Eyetracker können entweder als festinstalliertes System auf, unter oder vor dem Monitor, vor dem der/die Betrachter*in sitzt, befestigt werden, oder als kopfgetragenes System mittels einer Vorrichtung am Kopf des/der Betrachter*in. In beiden Fällen muss das System kalibriert werden, indem dem/der Betrachter*in auf dem Bildschirm Punkte an verschiedenen Positionen angezeigt werden, auf die er/sie blicken sollen. Hiermit wird das gesamte Blickfeld ausgelotet und sichergestellt, dass die physischen Abstände zwischen Betrachter*in, Eyetracker und Bildschirm korrekt eingestellt sind. Auch die Schwellenwerte für Pupillenmitte und die Cornea-Reflektionen werden so überprüft, damit es nicht zu Interferenzen durch lange Wimpern oder Reflektionen in Brillengläsern kommt (Duchowski 2017: 85). Kopfgetragene Eyetracker garantieren zwar eine weniger fehleranfällige Messung, da sich das Gerät immer in derselben Position relativ zu den Augen befindet, und sie ermöglichen dem/der Träger*in auch einen höheren Grad an Mobilität, etwa für die Verwendung außerhalb von Laboren, jedoch sind sie durch ihre teils unhandliche Konstruktion für Teilnehmer*innen von Experimenten störender als ein festinstallierter Eyetracker (Nauth 2012: 23). Festinstallierte Systeme erlauben ebenfalls einen gewissen Grad an Bewegungsfreiheit und können dabei je nach Positionierung und Design komplett unauffällig in die Forschungsumgebung, etwa unter einen Bildschirm, integriert werden (Duchowski 2017: 97). Darüber hinaus senden diese Eyetracker die aufgezeichneten Daten in Echtzeit an einen verbundenen Computer, wodurch eindeutige, parameterbasierte Daten zur Auswertung gewonnen werden können (Nauth 2012: 24). Für den privaten Gebrauch von

modernen festinstallierten Eyetrackern sind der *Tobii Eye Tracker 5* der *Tobii AB* zu nennen, für den Gebrauch in der Forschung etwa der *EyeLink Portable Duo* von *SR Research*, der auch im Rahmen der empirischen Studie dieser Arbeit zur Anwendung kam.

Ein Eyetracker zeichnet prinzipiell nur Rohdaten wie Koordinaten und Dauer von Fixationen, Sakkaden, Regressionen oder Augenfolgebewegungen auf, die anschließend mit Hilfe einer Software ausgewertet und zu Informationen gemacht werden müssen (Nauth 2012: 28). Die Analyse dieser Rohdaten dient dazu, Einblicke in das Aufmerksamkeitsverhalten des/der Betrachter*in zu erhalten. Dafür wird jede Stelle, an der sich die durchschnittliche temporale Frequenz der Augenposition ändert und somit beispielsweise das Ende einer Fixation darstellt, gesucht (Rajs et al. 2016: 119). Auf eine Fixation folgt stets eine Sakkade, die den Übergang bis zur nächsten Fixation darstellt. In erster Linie werden diese Übergänge mittels Summierung und Mittelwert berechnet. Es wird eine temporale Mindestdauer festgelegt, die eine Augenposition erfüllen muss, um als Fixation zu gelten, was auch als *Verweildauer* bezeichnet wird, und dann beobachtet, ob das Signal in seiner Frequenz variiert oder nicht (Duchowski 2017: 142). Solche Aufzeichnungen sind in ihrer Grundform jedoch nur rein numerische Daten, die Nutzer*innen wenig Auskunft über die Messung geben, weshalb zur Analyse eine Softwareanwendung nötig ist, die diese Daten zu Informationen machen kann. Im Falle des oben genannten *EyeLink Portable* Duo bietet *SR Research* (2021) die Anwendung *EyeLink Data Viewer* an, welche es ermöglicht statistische Analysen durchzuführen oder die Fixationen und Sakkaden der gesamten Aufzeichnung anzuzeigen. Dafür kann ein *Gaze Plot* erstellt werden, welcher Reihenfolge und Position aller Fixationen auf dem ganzen Bildbereich anzeigt, oder auch eine *Heat-* oder *Fixation-Map*, bei der sämtliche Fixationspunkte je nach Dauer und Verteilung farblich von grün, was eine geringe Anzahl und Dauer von Fixationen anzeigt, bis rot, was eine hohe Anzahl und Dauer von Fixationen anzeigt, auf dem betrachteten Inhalt visualisiert werden (Nauth 2012: 29).

Eyetracker können in die zwei Bereiche der diagnostischen und der interaktiven Anwendung unterteilt werden. Bei der interaktiven Anwendung stellt der Eyetracker ein Eingabegerät dar, welches die Augenbewegungen eines/r Nutzer*in als Zeigegerät, ähnlich einer Computermaus, nutzt und es ermöglicht, Elemente auf einem Bildschirm auszuwählen (Duchowski 2017: 248). Für Wissenschaft und Forschung ist jedoch die diagnostische Anwendung wichtiger, bei der der Eyetracker zur Aufzeichnung der Augenbewegung dient, um Aufmerksamkeitsmuster bezüglich eines visuellen Reizes erforschen zu können. Im Gegensatz zur interaktiven Anwendung haben die Augenbewegungen hier keinen Einfluss auf die betrachteten Elemente und der/die Betrachter*in kann diese auch nicht beeinflussen. In

diesem Anwendungsbereich ist vor allem im Rahmen von Studien und Experimenten eine gewisse Diskretheit wünschenswert, da Betrachter*innen die Präsenz des Eyetrackers im Idealfall nicht wahrnehmen und durch ihn von ihrer eigentlichen Aufgabe nicht abgelenkt werden sollen (Duchowski 2017: 247). Diagnostische Anwendungsbereiche für Eyetracker finden sich in der Medizin, etwa bei der Funktionellen Magnetresonanztomographie, in der Optometrie beim Durchführen von Lese- oder Sehtests, in der Kunst bei der Wahrnehmung von Kunstwerken oder Filmen und Videos, in der Ausbildung von Flugzeugpilot*innen, bei der Erforschung der visuellen Aufmerksamkeit beim Autofahren, im Bereich des Marketing zur Erforschung des Blickverhaltens bei Print- oder Fernsehwerbungen oder auch bei der Rezeption von Inhalten auf Websites (Duchowski 2017: 300ff). In der Forschung werden Eyetracker primär für die *Human-Computer-Interaction* (HCI) verwendet, die in Kapitel 2.4.2 genauer beschrieben werden soll. Sie kommen zum einen bei Studien zur Benutzerfreundlichkeit und Barrierefreiheit von Softwareanwendungen zum Einsatz, verstärkt aber auch im Bereich der Translationswissenschaft zum Erforschen der *Translator-Computer-Interaction* (TCI) (Duchowski 2017: 315). Auf die konkrete Anwendung von Eyetrackern bei Übersetzungs- oder Post-Editing Studien soll in Kapitel 2.4.5.2 eingegangen werden.

2.4 Post-Editing

Wie in Kapitel 2.1.1 bereits erwähnt, trat PE bereits mit der Entwicklung der allerersten MÜ-Systeme als deren Begleiterscheinung auf, da in gewisser Form schon immer eine menschliche Bearbeitung der maschinellen Rohübersetzung nötig war. Auch mit dem aktuellen Entwicklungsstand von NMÜ-Systemen ist eine FAHQT nur in wenigen Anwendungsbereichen erzielbar, was die Wichtigkeit von PE erneut hervorhebt. Da diese Zusammenarbeit von Mensch und Maschine auch in kommenden Jahrzehnten noch unabdingbar sein wird, stellt sich jedoch auch die Frage, ob es sich bei dem Post-Editing um maschinengestützte Humanübersetzung, oder um humangestützte maschinelle Übersetzung handelt (Koponen 2016: 131). Veale und Way (1997:5) definieren Post-Editing als "The correction of machine translation output by human linguists/editors". O'Brien (2011: 197) greift diese Definition auf und erweitert sie um einige Aspekte: "Post-editing is the correction of raw machine translated output by a human translator according to specific guidelines and quality criteria." Beide Definitionen zeichnen sich durch das Wort *correction*, also Korrektur aus, welches im Falle von Post-Editing etwas irreführend sein kann. Es gilt, das PE maschinell übersetzter Texte von der Revision human-übersetzter Texte abzugrenzen, da es sich um unterschiedliche Aufgaben mit verschiedenen Anforderungen handelt. Wo sich die Revision auf das Überprüfen eines von Menschen übersetzten Textes nach dem Vier-Augen-

Prinzip bezieht, ist Post-Editing auf maschinelle Übersetzung ausgerichtet (Nitzke 2019: 12). Der Unterschied wird dabei in der Art und Häufigkeit der Fehler deutlich, denn während Revisor*innen bei der Humanübersetzung lediglich auf einige wenige Rechtschreib- oder Tippfehler stoßen können, sehen sich Post-Editor*innen mit häufig auftretenden Fehlern in der Rohübersetzung konfrontiert, die in einigen Fällen durch die Komplexität von NMÜ-Systemen gar nicht nachvollziehbar sind, und in dieser Form nie in einem humanübersetzten Text auftreten würden (Nitzke 2019: 13). Weiters ist der Arbeitsprozess beim PE langwieriger und kognitiv anspruchsvoller, da Post-Editor*innen simultan drei Prozesse bewältigen müssen. Laut Somers (2003) Typologie zum PE umfassen diese das Lesen des Ausgangstextes (AT), das vergleichende Lesen der rohen maschinellen Übersetzung und durch das Vornehmen von Änderungen an derselben die Produktion des Zieltextes (ZT). „Dabei besteht die Aufgabe des Post-Editings nicht in der vollständigen Neuübersetzung eines Segments, sondern in der punktuellen Verbesserung des Ergebnisses der Maschinellen Übersetzung." (Kurz 2014: 422) Von der Rohübersetzung soll also so viel wie möglich beibehalten werden und so viel wie nötig an den relevanten Stellen editiert werden. Dieses Editieren bezieht sich auf das Vornehmen einer oder mehrerer der vier Änderungen *Löschen*, *Einfügen*, *Verschieben* und *Ersetzen* (do Carmo & Moorkens 2020: 37).

Das Post-Editing maschinell übersetzter Texte bietet einen potenziellen Produktivitätszuwachs von bis zu 70% in der Übersetzungstätigkeit (Kurz 2014: 422) und ermöglicht es, große Textmengen zu verarbeiten. Für gewisse Fachbereiche, Textsorten und Sprachpaare können besonders gute maschinelle Übersetzungen erzeugt werden, wodurch auch der PE Aufwand geringer ausfällt. Auf Basis einer guten MÜ können Übersetzer*innen laut Deparaetere et al. (2014) mit PE ihre Produktivität, gemessen an Wörtern pro Tag, um 30 bis 70% steigern. Gleichzeitig schneiden post-editierte Texte bei der Evaluierung durch Qualitätsmetriken, wie *General Text Matcher* (GTM) und *Translation Edit Rate* (TER), in Bezug auf Flüssigkeit und Fehler ebenfalls konsistent besser ab. Auch wenn das PE hochqualitativer MÜ produktiver ist, muss bedacht werden, dass diese Qualität von vielen Faktoren abhängig ist. Zu bedenken ist die Art der MÜ Engine (statistisch oder neuronal), das Sprachpaar, der Fachbereich, die häufigsten Fehler und Probleme von MÜ-Systemen sowie die Vorerfahrung und Einstellung des/der Post-Editor*in zur eigenen Aufgabe. Variiert nur einer dieser Faktoren, kann auch die Qualität und Produktivität des PE um einiges geringer ausfallen, und in einigen Fällen kann sich ein PE auch aufwändiger als eine Humanübersetzung desselben Textes gestalten. Hinzu kommen kürzere Abgabefristen als bei normalen Übersetzungsaufträgen sowie geringere Wort- oder Zeilenpreise, die damit

gerechtfertigt werden, dass der Text bereits übersetzt und daher theoretisch mit weniger Aufwand verbunden ist (Guerberof Arenas 2020: 333). Eine der wichtigsten ungeklärten Fragen der Disziplin, mit der sich auch diese Arbeit beschäftigen soll, ist also wie sich eruieren lässt, ob PE für einen vorliegenden maschinell übersetzten Text sinnvoll und produktiv ist, oder ob der damit verbundene Aufwand zu hoch wäre (Koponen 2016: 132).

2.4.1 Geschichte

Bei der Entwicklung der MÜ war PE stets als notwendige Begleiterscheinung vorhanden und wurde von den Entwickler*innen der Disziplin auch explizit als Bestandteil der frühen MÜ-Systeme genannt, wie etwa im Patent von Petr Trojanskij aus den 1930er Jahren, das zweisprachige Bearbeiter*innen beschreibt, die im maschinell übersetzten Text Bedeutung und Stil ergänzen. Edmundson und Hays (1958) entwickelten erstmals ein konkretes Konzept für ein MÜ-System inklusive PE, das wissenschaftliche Texte aus dem Russischen ins Englische übersetzen sollte und für die *RAND Corporation*, einen Think Tank des US-Militärs gedacht war. Der/die Post-Editor*in sollte dabei die maschinelle Rohübersetzung mittels eines Codes bearbeiten, der die grammatikalischen Informationen jedes Wortes erfasst, wodurch auch eine gute zielsprachliche Kompetenz sowie Fachwissen benötigt wurden (Koponen 2016: 133). Die Ausgangssprache Russisch musste für diesen Prozess jedoch nicht beherrscht werden. Dieses Konzept ließ sich jedoch nicht erfolgreich umsetzen, wodurch das Programm nach wenigen Jahren eingestellt wurde.

In der Forschung der 1950er Jahre galt PE als Schwachstelle der MÜ, da der letzte Schritt im Dekodierungsprozess dem/der menschlichen Bearbeiter*in überlassen war. Ziel war zu dieser Zeit jedoch das Erreichen einer FAHQT ohne menschliche Beihilfe. Die Notwendigkeit von Menschen im PE-Prozess wurde jedoch schnell klar, wodurch diese Zusammenarbeit auch als gemischte maschinelle Übersetzung (*mixed machine translation*) bezeichnet wurde. Dies warf erstmals die Frage auf, welche Prozesse dabei der Mensch übernehmen soll. Anfänglich standen menschliche Bearbeiter*innen am Rand des Übersetzungsprozesses und galten als Partner*innen der Maschine, die nur die nötigsten Änderungen vornehmen sollten (Vieira et al. 2019: 2). Mitte der 1950er Jahre wurden zur Betreuung von MÜ-Systemen erstmals Freiberufler*innen als Pre-Editor*innen eingesetzt, die die Ausgangssprache gut beherrschten, sowie als Post-Editor*innen, die Fachgebiet und Zielsprache gut beherrschten, jedoch nicht die Ausgangssprache. Mitte der 1960er Jahre waren beispielsweise 43 Personen in der MÜ-Abteilung der amerikanischen *Air Force* beschäftigt, darunter einige Pre- und Post-Editor*innen, die täglich dabei halfen 100.000 Wörter maschinell aus dem Russischen zu übersetzen. Zur selben Zeit beschäftigte sich die

Forschung mit der Bewertung von MÜ-Output und der Dauer, die PE in Anspruch nimmt (Nitzke 2019: 14). Durch den ALPAC-Bericht, in dem das PE in sämtlichen Kriterien schlechter als die Humanübersetzung eingestuft wurde, kam jedoch auch diese Forschung zum Stillstand (Nitzke 2019: 15). Orr und Small (1967) publizierten im selben Jahrzehnt dennoch die erste Forschungsarbeit zu PE, in der sie das Leseverständnis von Teilnehmer*innen verglichen, die eine maschinelle Rohübersetzung, eine post-editierte Version der Rohübersetzung und eine humanübersetze Version desselben wissenschaftlichen Textes lesen mussten. Die Teilnehmer*innen mussten anschließend Fragen beantworten, die auf die Verständlichkeit der Texte schließen ließen, was aufzeigte, dass zwischen dem post-editierten und humanübersetzten Text nur geringe Unterschiede bestanden, aber viel größere zwischen der rohen MÜ und der Humanübersetzung (Koponen 2016: 135).

Im europäischen Raum bestand durch die Verwendung von *SYSTRAN* beispielsweise Bedarf an Post-Editor*innen, die auch gut ausgebildet und effektiv sein sollten, obwohl sie damals nur geringe Änderungen an den Texten vornehmen mussten (Nitzke 2019: 15). Auch andere Institutionen und Firmen führten PE in die Arbeitsabläufe ihrer Übersetzungsabteilungen ein, jedoch entwickelten die meisten auf Grund der fehlenden Forschung eigene Lösungen. Die *Pan American Health Organization (PAHO),* deren Ziel es ist in ihren Mitgliedsstaaten Krankheiten zu bekämpfen und Gesundheitssysteme zu verbessern, führte 1980 eines der ersten, dedizierten PE-Systeme für die interne Verwendung ein. In Kombination mit einem eigenen, regelbasierten MÜ-System namens *PAHOMTS*, werden 90% aller Dokumente maschinell übersetzt und dann post-editiert, Humanübersetzer*innen kommen kaum zum Einsatz. Das System verwendet weder eine kontrollierte Sprache, noch werden Texte pre-editiert, durch die regelbasierte Architektur und die Spezialisierung auf medizinische Texte ist das System eigenständig in der Lage, hochqualitative Rohübersetzungen zu erzeugen. In Kombination mit Post-Editing gibt die *PAHO* eine Produktivitätssteigerung von 30-50% gegenüber Humanübersetzungen an (Nitzke 2019: 35). 1994 führte die Europäische Kommission zur Ergänzung des MÜ-Systems *SYSTRAN* den *PER-Service (post-édition rapide)* ein, der von Freiberufler*innen ohne PE-Vorerfahrung betreut wurde und nur zum Einsatz kam, wenn durch Humanübersetzer*innen für gewisse Aufträge Abgabefristen überschritten wurden. Von den Kund*innen der post-editierten Texte wurden Rückmeldungen eingeholt, die wiederum in *SYSTRAN* eingearbeitet wurden, um das System zu verbessern. Da eine post-editierte Normseite 50% günstiger als eine humanübersetzte war, die Produktivität von 30.000 Seiten auf 180.000 Seiten gesteigert wurde und der *PER-Service* bis 1997 doppelt so häufig genutzt wurde, schrieb die

Europäische Kommission 1998 erstmals eine Stellenanzeige für eine/n Post-Editor*in in den Sprachen Deutsch-Englisch-Französisch aus (Nitzke 2019: 35f). Allerdings war diese Person nicht für die Korrektur von maschinellen Rohübersetzungen zuständig, die laut Europäischer Kommission darauf abzielt einen publikationsreifen Text zu erzeugen, sondern für das Post-Editing von maschinellen Rohübersetzungen, wobei der Zieltext nicht publiziert wird und der Stil unwichtig ist, solange Informationsgehalt und Inhalt korrekt übertragen werden (Nitzke 2019: 36).

In den letzten zehn Jahren wuchs auch das Forschungsinteresse in der Translationswissenschaft und es wurden Richtlinien, wie 2010 durch *TAUS (Translation Automation User Society)*, oder Normen, wie 2017 durch ISO eingeführt (Koponen 2016: 135). In den letzten Jahren wurden auch open-source CAT-Tools, wie etwa *CASMACAT* oder *MateCat*, entwickelt, die Schnittstellen zu MÜ-Systemen haben und über einige spezifische PE Funktionen verfügen (Nitzke 2019: 15). Zwar gab es vereinzelte Bestrebungen zur Entwicklung eigenständiger Post-Editing-Tools, wie das *ACCEPT Post-Editing Environment* durch Routier et al. (2013) oder das *PET: Post-Editing Tool* durch Aziz und Specia (2012), diese konnten sich jedoch nicht in einem mit CAT-Tools vergleichbaren Ausmaß durchsetzen. Heutzutage hat sich das PE also zu einer eigenen Dienstleistung entwickelt, die von speziell ausgebildeten Übersetzer*innen unter Einhaltung von Normen wie der ISO durchgeführt wird, und in der der menschliche Bearbeiter nicht mehr eine untergeordnete Rolle spielt, sondern diese professionelle Aufgabe mit seiner Expertise bewältigt und steuert (Vieira et al. 2019: 3).

2.4.2 Post-Editing als Human-Computer-Interaction

In der Einleitung dieses Kapitels wurde bereits die Problematik angesprochen, dass die Grenzen zwischen maschinengestützter Humanübersetzung und humangestützter maschineller Übersetzung immer stärker verschwimmen und auch das Post-Editing nicht eindeutig zuordenbar ist (Läubli & Green 2019: 370). Die gegenseitige Abhängigkeit von Mensch und Maschine, genauer von Übersetzer*in und Maschine, soll in diesem Kapitel durch das Konzept der *Human-Computer-Interaction* (HCI), beziehungsweise der *Translator-Computer-Interaction* (TCI) beleuchtet werden.

Die Entwicklungen und Durchbrüche im Bereich der Übersetzung der letzten vierzig Jahre waren stets stark an die Entwicklung und Optimierung von Computern gebunden und daher auch interdisziplinärer Natur. Durch den großen Übersetzungsbedarf ist der moderne Übersetzerberuf von Wiederholungen und umfangreichen Textmengen geprägt, die sich durch maschinelle Unterstützung von Übersetzer*innen (M*achine Assisted Human Translation,*

MAHT) gut bewältigen lässt. Gleichzeitig ist die Expertise von Übersetzer*innen in der Betreuung von MÜ-Systemen gefragt (*Human Assisted Machine Translation*, HAMT), da diese eigenständig nur selten zur Produktion einer FAHQT in der Lage sind und PE benötigt wird (O'Brien 2012: 101f). Das Interesse der Translationswissenschaft an diesen neuen Forschungsgebieten ist groß, da vor allem die Weiterentwicklung des Übersetzerberufs angesichts neuer Technologien von Bedeutung ist. Hier setzt auch die HCI an, die als Erforschung der Interaktion von Menschen, Computern und Aufgaben definiert wird und sich in ihren Anfängen in den 1990er Jahren vor allem mit der Einstellung von Menschen zu Softwareprogrammen und deren Benutzerfreundlichkeit beschäftigte (O'Brien 2012: 103). Da auch das Übersetzen mittlerweile oft rein computergestützt abläuft, handelt es sich eindeutig um eine Form der HCI, die folglich als *Translator-Computer-Interaction* (TCI) bezeichnet werden soll. Mit Einführung der ersten elektrischen Schreibmaschinen mit Textspeicher in den späten 1970er Jahren wurde das Übersetzen erstmals zu einer computergestützten Aufgabe. Als zehn Jahre später die ersten Textverarbeitungsprogramme und CAT-Tools auf den Markt kamen, wurden die Vorteile dieser Anwendungen schnell bekannt. Vor allem die Lokalisierungs- und EDV-Industrie setzten auf diese neuen Technologien, die erstmals die Wiederverwendung von übersetzten Segmenten ermöglichten und zu einem enormen Produktivitätszuwachs führten. Die CAT-Tools wurden um Module zur Einbindung von Terminologiedatenbanken erweitert und die Oberfläche der Systeme wurde nutzerfreundlicher gestaltet, wodurch diese auch Übersetzer*innen ohne technische Vorkenntnisse effektiv nutzen konnten. Mit dem Zuwachs an Internetanschlüssen Mitte der 1990er Jahre wurden gedruckte Wörterbücher überflüssig, da vermehrt Korpora und digitale Sprachdaten verfügbar waren (O'Brien 2012: 104). Der Grad an TCI hat durch die Verfügbarkeit an freien MÜ-Systemen ab Mitte der 2000er Jahre noch mehr zugenommen. Neben privaten Anwender*innen zählen mittlerweile auch Übersetzungsbüros, Sprachdienstleister*innen sowie freiberufliche Übersetzer*innen zu den Hauptnutzer*innen von MÜ-Systemen. Der Grad an Interaktion beim Einfügen des Ausgangstextes ist gering, die Systeme arbeiten in diesem Schritt quasi autonom, jedoch ist der Grad an Interaktion nach Erzeugen der Rohübersetzung umso höher, wenn diese von Übersetzer*innen, oder in diesem Falle Post-Editor*innen, bearbeitet wird (O'Brien 2012: 106).

Die Einstellung von Übersetzer*innen zur Technologie ist somit wichtiger als die Technologie selbst, denn solange sie nicht bereit sind sie zu nutzen, kann diese auch nicht die Produktivität steigern (Bundgaard 2017: 126). In einer Studie mit acht professionellen Teilnehmer*innen beschäftigte sich Bundgaard (2017) mit der Einstellung von

Übersetzer*innen, deren Arbeit eine Woche lang durch MÜ-Vorschläge unterstützt wurde. Während die Einstellung der Teilnehmer*innen eher negativ war und die MÜ in vieler Hinsicht als Einschränkung und unerwünschte Komponente im Übersetzungsprozess gesehen wurde, erkannten dennoch einige Teilnehmer*innen auch den möglichen Produktivitätszuwachs und die Sinnhaftigkeit dieser Technologie. Ähnlich wie bei der Einführung von CAT-Tools vor über 20 Jahren haben Übersetzer*innen eine flexible und pragmatische Einstellung gegenüber solcher neuartigen Technologien und sind in der Lage als Akteur*innen geeignete Handlungen zu setzen, um die Defizite dieser Programme auszugleichen. Die Unaufhaltbarkeit dieser Entwicklung wird von den Teilnehmer*innen jedoch auch kritisch betrachtet, sie fürchten um den Verlust ihres Berufs und somit auch um einen Verlust ihrer Handlungsfähigkeit, wie ein*e Teilnehmer*in es formulierte: „It is me who makes this translation, it is me who decides what to write“ (Bundgaard 2017: 142).

2.4.3 Prozesse und Abläufe

Bevor auf die verschiedenen Arten des PE eingegangen wird, sollen zuerst die Begrifflichkeiten geklärt und Gründe genannt werden, warum PE und Revision im Kontext der MÜ nicht als Synonyme zu verstehen sind.

Wie bereits in Kapitel 2.4 erwähnt, ist das Post-Editing maschinell übersetzter Texte von der Revision humanübersetzter Texte zu unterscheiden. Laut do Carmo und Moorkens (2020: 36) ist Übersetzen als Schreibprozess zu verstehen, bei dem Übersetzer*innen den Ausgangstext lesen, verstehen und einen Zieltext produzieren. In diesem Prozess wird das Schreiben nur pausiert, wenn Probleme auftreten, also Passagen, die mehrmals gelesen werden müssen, um sie zu verstehen und eine Übersetzung zu finden. Die Revision ist hingegen als Leseprozess zu verstehen, bei dem Revisor*innen den Zieltext lesen und ihn auf seine Korrektheit überprüfen. Hier wird nur pausiert, wenn Fehler auftreten, also Passagen, die redigiert werden müssen, um die Aussage des Textes korrekt wiederzugeben. Parallel zu Humanübersetzen und Revision als Schreib- und Leseprozess treten mit Fortschritt der Technologie nun MÜ und PE auf, die aber nicht als Synonyme der ersten beiden Begriffe zu verstehen sind.

Auch wenn mehrere Studien, wie etwa jene von Koehn (2010), gezeigt haben, dass beim PE mehr Zeit mit Lesen als mit Schreiben verbracht wird, so ist dies nicht Grund genug für eine Gleichsetzung mit der Revision. Zum einen ist die Erwartungshaltung eines/einer Übersetzer*in zu beachten, der/die einen der beiden Aufträge annimmt. Bei der Revision eines humanübersetzten Textes weiß der/die Übersetzer*in, dass es sich bei dem vorliegenden Text um eine finalisierte Übersetzung handelt, in der nur noch kleine Fehler ausgebessert

werden müssen. Dieselbe Erwartungshaltung ist bei einem PE Auftrag nicht möglich, denn es handelt sich hier um keine finalisierte Übersetzung, sondern um einen Vorschlag oder eine Hypothese eines möglichen Zieltextes, der durch den/die Übersetzer*in in einen finalisierten Text verwandelt werden muss. Hierbei sind deutlich mehr Bearbeitungen nötig als bei der Revision, der Prozess ist kognitiv weitaus aufwändiger und der/die Bearbeiter*in muss mit den häufigsten Fehlertypen vertraut sein und gleichzeitig Zweck und Zielpublikum der Übersetzung bedenken (do Carmo & Moorkens 2020: 40). Bei der Humanübersetzung läuft der gesamte Schreib- beziehungsweise Erzeugungsprozess der Übersetzung im Gehirn ab, da Bedeutung und Sinn des Ausgangstextes rezipiert, umgewandelt und in einen Zieltext generiert werden. Auch wenn PE theoretisch gesehen ähnlich der Revision nur auf das Bearbeiten einiger weniger Fehler abzielen sollte, so ist der/die Post-Editor*in in der Praxis meist mit gröberen Fehlern und Unstimmigkeiten in der Rohübersetzung konfrontiert. Dadurch verwandelt sich auch das PE von einem Lese- in einen Schreibprozess, da der/die Bearbeiter*in einen Teil der Erzeugung des Textes selbst übernehmen muss (do Carmo & Moorkens 2020: 47).

Weiters besteht nach der Durchführung von PE die Problematik, dass ein post-editierter Text nicht automatisch als redigiert gilt. In vielen kommerziellen Bereichen werden post-editierte Texte nach dem Vier-Augen-Prinzip einer Revision zur Qualitätssicherung unterzogen. Die Wichtigkeit, die dem post-editierten Text aktuell zugeschrieben wird und die Tatsache, dass die Revision in diesem Prozess als eigenständiger Schritt durchgeführt wird, lässt darauf schließen, dass PE nicht mit der Revision gleichgesetzt werden kann. Vielmehr bedeutet dies aber, dass das PE im Kontext moderner Übersetzungstechnologien als Weiterentwicklung der Humanübersetzung gesehen werden sollte. Wo früher sowohl Übersetzen als auch Revision durch einen Menschen vorgenommen wurden, so wird der Mensch durch die MÜ nun in einem dieser Bereiche nicht mehr gebraucht (do Carmo & Moorkens 2020: 42). Allerdings erfüllt PE denselben Zweck wie die Humanübersetzung, die Produktion eines qualitativen Zieltextes auf effiziente und effektive Weise, und setzt von dem/der Bearbeiter*in bilinguale Lese-, Schreib- sowie Kulturkompetenz voraus. Während Humanübersetzung und Post-Editing maschineller Übersetzungen momentan noch koexistieren, so ist es denkbar, dass zweiteres im Laufe der Jahre an Bedeutung gewinnen wird und die Humanübersetzung immer mehr verdrängen könnte.

Angesichts dieser Entwicklungen ist es wichtig zu verstehen, welche grundlegenden Abläufe mit dem PE verbunden sind. Post-Editing umfasst das Vornehmen einer oder mehrerer der vier Änderungen *Löschen*, *Hinzufügen*, *Verschieben* und *Ersetzen* (do Carmo &

Moorkens 2020: 44). Diese Änderungen können in zwei Kategorien unterteilt werden: jene, die sich auf den Inhalt der sprachlichen Einheiten, also Wörter und Wortgruppen, beziehen und jene, die sich auf die Position dieser Einheiten in einem Segment beziehen. Löschen und Hinzufügen sind primäre Änderungen, durch sie wird die Anzahl der sprachlichen Einheiten in einem Segment erhöht oder verringert und auch der Inhalt kann sich ändern. Verschieben und Ersetzen sind hingegen sekundäre Änderungen, da bei ihnen die Anzahl der sprachlichen Einheiten gleichbleibt und sich lediglich ihre Position im Segment ändert. Sekundäre Änderungen setzen sich aus primären Änderungen zusammen. Verschieben umfasst das Löschen einer Einheit an einer Position und das Einsetzen derselben Einheit an einer anderen Stelle. Ersetzen hingegen umfasst das Löschen einer Einheit an einer Position und das Einsetzen einer Einheit von einer anderen Stelle an dieser Position (do Carmo & Moorkens 2020: 44f). PE darf jedoch nicht als einfache Aneinanderreihung verschiedener Änderungen wie oben beschrieben verstanden werden, denn Post-Editor*innen verändern die Segmente auf komplexere Weise. Sätze werden oft in nicht-linearer Abfolge gelesen und bearbeitet, es können Wörter durch Phrasen ersetzt werden, das Ersetzen kann weitere Änderungen im selben Satz nötig machen (z.B. Deklination) und manche Segmente können sich auf andere beziehen, wodurch eine Änderung in einem Segment eine Änderung in allen anderen Segmenten nach sich zieht (do Carmo & Moorkens 2020: 45).

Der Prozess des PE ist jedoch weitaus umfangreicher und besteht nicht nur aus dem reinen Vornehmen von Änderungen. Laut der Typologie von Somers (2003) lässt sich PE in die drei Prozesse Lesen des Ausgangstextes, Lesen der maschinellen Übersetzung und Vornehmen von Änderungen an zweiterer unterteilen. Torrejón und Rico (2013: 168f) schlagen zur genaueren Beschreibung des Prozesses eine erweiterte Typologie vor, die hier in chronologischer Reihenfolge kurz beschrieben werden soll. Zum einen umfasst sie ausgangstextbezogene Prozesse, wie das Lesen des Ausgangstextes zur Sinnerfassung, dem Erkennen von Mustern und Problemstellen sowie dem späteren Vergleich mit der maschinellen Übersetzung. Referenzbezogene Prozesse können nach einer ersten Lektüre des Ausgangstextes optional gewählt werden und umfassen eine Recherche von Terminologie, Paralleltexten und andere, für die Aufgabe relevante Quellen. MÜ-bezogene Prozesse umfassen das Lesen der maschinellen Rohübersetzung, das Erkennen von Problemstellen oder Segmente, die mit dem Ausgangstext abgeglichen werden müssen und erste Überlegungen, wie Segmente umformuliert werden können. Die Prozesse der Zieltextproduktion machen den größten Teil eines Post-Editings aus und umfassen das oben beschriebene Vornehmen der vier Änderungsarten unter der Vorgabe bestimmter Richtlinien und unter Beachtung des Zwecks,

also ob ein leichtes oder vollständiges PE notwendig ist. In diesem Schritt sollte pro Segment nicht länger als zwei Sekunden überlegt werden, ob es effizienter ist den MÜ-Vorschlag zu editieren, oder zu löschen und selbst neu zu übersetzen (Zaretskaya 2017: 122). Anschließend folgen die Prozesse zur Zieltextevaluierung, in denen der produzierte Zieltext sowohl mit der maschinellen Rohübersetzung als auch dem Ausgangstext abgeglichen wird. Durchzogen wird das gesamte Post-Editing von globalen aufgabenbezogenen Prozessen: der/die Post-Editor*in ist bei seiner/ihrer Arbeit für die Aufgabensteuerung verantwortlich und muss die oben beschriebenen Prozesse bewusst einteilen und durchführen können. Darüber hinaus kann es nötig sein, dass der/die Post-Editor*in auch Rückmeldungen bezüglich der Qualität der MÜ an den/die Auftraggeber*in gibt und Verbesserungsvorschläge mitteilt. Die hohe Komplexität und der große Umfang eines Post-Editings verstärken somit die oben geäußerte Annahme, dass Revision keinesfalls mit PE gleichgesetzt werden kann, sondern dass es sich dabei um die Weiterentwicklung beziehungsweise Digitalisierung des Übersetzungsprozesses handelt (Torrejón & Rico 2013: 169).

2.4.4 Arten von Post-Editing

Bevor in diesem Kapitel genauer auf die verschiedenen Arten des Post-Editing eingegangen wird, soll zuerst die verwandte Disziplin des Pre-Editings kurz erläutert werden. Durch Pre-Editing, also die Vorbearbeitung eines ausgangssprachlichen Textes, sollen aus diesem Fehler und Textpassagen, die bei MÜ-Systemen bekannterweise zu Fehlern führen, entfernt werden. Dadurch kann der Text besser maschinell übersetzt werden und das Post-Editing gestaltet sich im Idealfall einfacher. Beim Pre-Editing werden auf den Text entweder manuell oder automatisch bestimmte Regeln angewandt, die Fehler entfernen sollen. *Acrolinx* ist beispielsweise eine Anwendung zur Content-Gestaltung, die es Nutzer*innen ermöglicht Texte automatisch zu pre-editieren, indem grammatikalische und syntaktische Regeln in eine Datenbank eingegeben und dann automatisch auf einen Text angewandt werden. Auch Mischformen sind möglich, wie etwa bei der *ACCEPT Post-Editing Environment* (Routier et al. 2013), bei dem das System den Text analysiert und Vorschläge zu problematischen Stellen macht, die von Nutzer*innen aber akzeptiert oder abgelehnt werden müssen. Diese Regeln können darauf abzielen Rechtschreibfehler zu entfernen, die Syntax mehr an die Zielsprache anzupassen, Eigennamen hervorzuheben, die nicht übersetzt werden sollen, oder komplexe, lange Sätze in kürzere zu unterteilen (Nitzke 2019: 17f).

Eine besonders effektive Methode des Pre-Editings ist das Verwenden einer kontrollierten Sprache, die in ihrem lexikalischen und syntaktischen Umfang begrenzt ist. Ziel einer kontrollierten Sprache ist es, lexikalische Mehrdeutigkeiten und komplexe

grammatikalische Strukturen bei der Textproduktion zu vermeiden, damit der Text sowohl von Leser*innen, als auch von MÜ-Systemen einfacher verstanden werden kann. Diese Regeln, auch terminologische und grammatikalische Richtlinien, sollen helfen, die Qualität der maschinellen Rohübersetzung zu verbessern (Guerberof Arenas 2020: 334). Das Pre-Editing geschieht quasi bereits beim Verfassen des Textes, denn für diese dürfen beispielsweise nur eine bestimmte Auswahl an Vokabeln verwendet werden, bei denen einer Bedeutung nur ein Begriff zugeordnet wird und umgekehrt. Auch ist es möglich, dass nur bestimmte syntaktische Konstruktionen und Zeitformen erlaubt sind. Vor allem im Bereich der technischen Dokumentation kommen kontrollierte Sprachen schon länger zum Einsatz, und auch in der Lokalisierungsindustrie kann dadurch eine höhere Produktivität erzielt werden (Nitzke 2019: 17f).

Aus Zeit- und Kostengründen wird meist jedoch auf ein Pre-Editing verzichtet und vielmehr darauf geachtet, die für den Auftrag richtige Art von Post-Editing auszuwählen. Beim monolingualen PE hat der/die Post-Editor*in beispielsweise nur Zugang zum maschinell übersetzten, zielsprachlichen Text und kann diesen folglich nicht mit dem Ausgangstext vergleichen. Für diese Art des PE sind allerdings auch nur monolinguale Post-Editor*innen vorgesehen, die die Ausgangssprache nicht beherrschen, dafür Fachwissen über den Inhalt des Textes besitzen. Das monolinguale PE ist für die Forschung von Interesse, da der Prozess vereinfacht werden kann und nicht zwingendermaßen von Sprachexpert*innen oder Übersetzer*innen übernommen werden müsste (Vieira 2019: 320). Wie in Kapitel 2.4.1 erwähnt, kamen beispielsweise in der Europäischen Kommission monolinguale Lai*innen als Post-Editor*innen zum Einsatz, es können jedoch auch Fachexpert*innen ein PE durchführen. Ähnlich wie bei der fachlichen Prüfung zeigten Studien bereits, dass monolinguale Fachexpert*innen beim PE eines ihnen vertrauten Themas über 90% korrekte Sätze erzeugen konnten, die außerdem in ihrem Ausdruck besser klangen. Die gängigste Variante ist jedoch das bilinguale PE, bei dem die maschinelle Rohübersetzung durch Sprachexpert*innen oder Post-Editor*innen mit dem Ausgangstext abgeglichen wird und anschließend Änderungen vorgenommen werden. Im Vergleich zum monolingualen Ansatz produziert diese Vorgehensweise Texte, bei denen der Inhalt adäquat übertragen wurde und auch Fehler der MÜ beseitigt wurden (Vieira 2019: 321).

Die meisten bilingualen PE-Systeme, die für kommerzielle Anwendungszwecke und die Forschung genutzt werden, funktionieren statisch, da der/die Post-Editor*in bereits bei Arbeitsbeginn den Ausgangstext und vollständig übersetzten Zieltext zur Verfügung hat. Den Text in seiner Gesamtheit zu editieren, bringt den Vorteil, dass auf Einheitlichkeit von

Terminologie und Inhalt geachtet werden kann, und auch, dass häufige Fehler in einem Durchgang entfernt werden können. Eine neuere Entwicklung stellt das interaktive Post-Editing dar, bei dem Post-Editoren*innen, oder in diesem Fall eher Übersetzer*innen, mit einem Programm arbeiten, das über ein integriertes MÜ-System verfügt und bei dem der Text Segment für Segment maschinell übersetzt und anschließend gleich post-editiert wird. Dadurch, dass der Text post-editiert wird, während er entsteht, können diese Systeme den Anwender*innen beispielsweise Übersetzungsvorschläge machen, die maschinell erzeugt wurden oder aus einem eingebundenen Translation Memory stammen, wie das interaktive CAT-Tool *Lilt Translate*. Andere Systeme können aus den manuellen Änderungen eines/einer Post-Editor*in lernen und diese bereits in die maschinelle Übersetzung des darauffolgenden Segments einfließen lassen, wie es etwa das open source CAT-Tool *CASMACAT* ermöglicht (Vieira 2019: 322). Als dritte Art ist noch das automatische Post-Editing zu erwähnen. Hierzu werden PE-Modelle erstellt, die anhand von maschinellen Rohübersetzungen und deren alignierten, von Menschen post-editierten Zieltexten, trainiert werden und folglich ähnliche Rohübersetzungen automatisch post-editieren können. Derartige Systeme werden vor allem vom *ADAPT-Centre* der Dublin City University erforscht (Vieira 2019: 328).

Um die Qualitätsansprüche von Kund*innen an einen post-editierten Text den Post-Editor*innen zu vermitteln und es diesen zu ermöglichen, nicht zu viele oder zu wenige Änderungen vorzunehmen, wurden einige Richtlinien erstellt, wie etwa die *TAUS Machine Translation Post-Editing Guidelines* (Massardo et al. 2016) oder die *ISO 18587 Post-editing of machine translation output* (ISO 2017). Die TAUS Richtlinie kommt vor allem in der Lokalisierungsindustrie zum Einsatz und soll Kund*innen und Sprachdienstleister*innen Vorschläge für ein erfolgreiches PE liefern (Koponen 2016: 135). Sie unterscheidet dabei zwischen zwei Qualitätsniveaus: Das erste Niveau *good enough* zielt auf die Produktion eines Textes ab, der verständlich und semantisch korrekt ist, bei dem keine Informationen entfernt oder hinzugefügt wurden, der so viel maschinelle Rohübersetzung wie möglich wiederverwendet, der grammatikalisch nicht fehlerfrei sein und bei dem keine stilistischen Verbesserungen vorgenommen werden (Massardo et al. 2016: 17). Das zweite Niveau *publishable quality* hingegen zielt auf die Produktion eines publikationsreifen Textes mit vergleichbarer Qualität zu einer Humanübersetzung ab, der grammatikalisch, syntaktisch und semantisch korrekt ist, bei dem die korrekte Terminologie verwendet wird, der so viel maschinelle Rohübersetzung wie möglich wiederverwendet und der korrekt formatiert ist (Massardo et al. 2016: 16). Die TAUS Richtlinie sieht vor, dass die Kriterien *good enough* und *publishable quality* immer anhand der Qualität des Zieltextes definiert werden, damit

beispielsweise vermieden wird, dass eine maschinelle Rohübersetzung mit schlechter Qualität nur ein leichtes Post-Editing erhält (Guerberof Arenas 2020: 336). Idealerweise erhalten Post-Editor*innen auch Informationen bezüglich des PE-Auftrags, etwa welches MÜ-System verwendet wurde, welchen Verwendungszweck der/die Kund*in anstrebt, sprachpaarspezifische Fehler die zu erwarten sind und welche Elemente nicht bearbeitet werden sollen (Guerberof Arenas 2020: 337). Die Norm *ISO 18587* wurde nach der TAUS Richtlinie veröffentlicht und orientiert sich stark daran, denn sie unterscheidet ebenfalls zwischen denselben zwei Qualitätsniveaus, bezeichnet sie jedoch anders. *Light post-editing,* also leichtes Post-Editing, ist nach ISO 18587 (2017) ein „process of post-editing […] to obtain a merely comprehensible text without any attempt to produce a product comparable to a product obtained by human translation […]". *Full post-editing*, also vollständiges Post-Editing, ist nach ISO 18587 (2017) ein „process of post-editing […] to obtain a product comparable to a product obtained by human translation […]".

Basierend auf der TAUS Richtlinie und der ISO 18587 Norm haben sich das leichte und das vollständige Post-Editing in der Sprachindustrie zu den gängigsten zwei Arten des PE etabliert. Die Wahl der PE-Methode hängt vom Verwendungszweck ab, also ob der Zieltext nur intern verwendet wird und eine kurze Nutzungsdauer hat, oder ob der Text publiziert wird und eine lange Nutzungsdauer hat (Nitzke 2019: 13f). Gewissermaßen hängt daher auch die Klassifizierung der drei großen Anwendungsbereiche der MÜ aus Kapitel 2.1.3 mit PE zusammen. So setzen maschinell übersetzte Texte, die für die Assimilation oder *gisting* bestimmt sind, nach ISO 18587 ein leichtes Post-Editing (nach TAUS *good enough*), und Texte, die für die Dissemination oder Publikation bestimmt sind, nach ISO 18587 ein vollständiges Post-Editing (nach TAUS *publishable quality)* voraus (Guerberof Arenas 2020: 335). Leichtes Post-Editing zielt auf einen Zieltext ab, der verständlich sein muss und nicht für die Publikation bestimmt ist. Syntax, Grammatik und Stil können Fehler enthalten und sind von untergeordneter Wichtigkeit, solange sie das Verständnis nicht erschweren und den Inhalt nicht verfälschen. Vollständiges Post-Editing zielt auf einen publikationsreifen Zieltext ab, der verständlich und inhaltlich genau sein muss, dessen Syntax und Grammatik fehlerfrei sind. Bezüglich des Stils wird akzeptiert, dass der Text nicht mit einer Humanübersetzung vergleichbar ist, aber dennoch einen guten Stil haben muss (Nitzke 2019: 13f). Auch innerhalb dieser zwei Arten des PE gibt es Abstufungsgrade, die sich nach dem Verwendungszweck richten. Dies kann Übersetzer*innen vor eine neue Herausforderung stellen, denn durch die Revision sind sie es gewohnt, eine Übersetzung in ihrer Gesamtheit zu redigieren, beim leichten PE müssen sie jedoch wissen, welche Fehler ausgebessert werden

und welche nicht (Guerberof Arenas 2020: 335). Vieira (2019: 325) weist außerdem darauf hin, dass sich die Umsetzung dieser zwei Qualitätsniveaus bei PE selbst in Unternehmen mit erfahrenen Übersetzer*innen als problematisch gestalten kann, da die Abgrenzungen eventuell nicht klar genug sind und viele Übersetzer*innen auch bei einem leichten PE dazu tendieren, mehr Fehler zu korrigieren als vorgesehen. Durch die steigende Leistungsfähigkeit von MÜ-Systemen könnten in Zukunft immer bessere Rohübersetzungen erzeugt werden, die folglich auch weniger Post-Editing benötigen würden, wodurch auch die Unterscheidung von leichtem und vollständigem PE überflüssig werden könnte.

Die Vielzahl an Arten und Verwendungsmöglichkeiten von PE, die in diesem und dem vorherigen Kapitel beschrieben wurden, verdeutlichen erneut, dass sich bereits ein Wandel von humangestützter maschineller Übersetzung zu maschinell-gestützter Humanübersetzung vollzieht. Dies wirft die Frage auf, welches Maß an Handlungsfähigkeit den Post-Editor*innen bei ihrer Tätigkeit zukommt und in welchem Umfang Änderungen an den maschinellen Rohübersetzungen erwünscht sind. Das Spektrum an Handlungsfähigkeit ist in der nachstehenden Abb. 3 dargestellt.

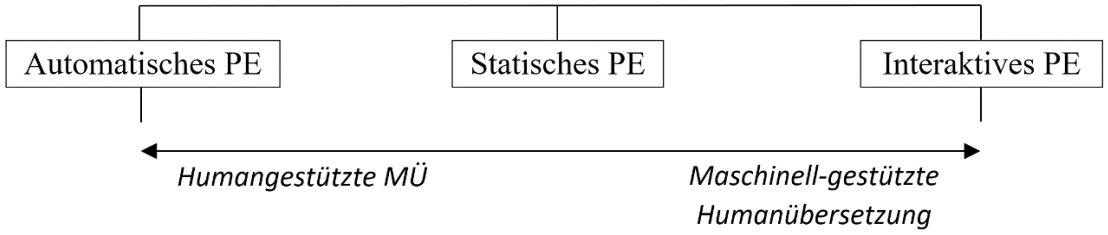


Abbildung 3: Menschliche Handlungsfähigkeit beim Post-Editing

Am einen Ende des Spektrums der Handlungsfähigkeit steht das automatische PE, welches ohne jegliche menschliche Bearbeitung funktioniert und eigenständig abläuft. Am anderen Ende des Spektrums befindet sich das interaktive PE, welches auch als interaktives Übersetzen bezeichnet wird, um den teils negativ konnotierten Begriff des Post-Editings zu umgehen. Hierbei ist die Handlungsfähigkeit des/der Post-Editor*in am wenigstens eingeschränkt, denn die maschinelle Rohübersetzung muss nicht zwingend übernommen werden, sondern kann als Vorschlag gesehen werden. Hierbei handelt es sich um eine maschinelle Unterstützung der Humanübersetzung. In der Mitte des Spektrums steht das statische PE, die gängigste Art der Disziplin, die auch mehrere Abstufungen umfasst. Je nach Zweck und Anwendungsbereich kann diese zentrale Position Eigenschaften der beiden Enden, also eher maschinengestützt oder humangestützt, annehmen. Beispielsweise tendieren die frühen Anwendungen von statischem PE durch nicht-qualifizierte, einsprachige

Bearbeiter*innen eher in Richtung des maschinenzentrierten automatischen PE, wohingegen moderne Einbindungen von MÜ-Vorschlägen in CAT-Tools in Richtung interaktives PE tendieren (Vieira 20219: 328). Dies wirft die Frage auf, ob interaktives PE überhaupt als Post-Editing gesehen werden sollte, da es zum einen viel häufiger als interaktives Übersetzen bezeichnet wird. Da Post-Editor*innen, oder eher Übersetzer*innen, die das Post-Editing maschineller Übersetzungsvorschläge zur Produktivitätssteigerung nutzen, diese Vorschläge nicht zwingendermaßen übernehmen müssen, ist dieser Anwendungsbereich wohl eher als Übersetzen und nicht als PE zu sehen. Diese Mischform ist aus keinem modernen Übersetzungsprozess wegzudenken, da jede/r Übersetzer*in früher oder später MÜ als Quelle für Vorschläge nutzt, wodurch diese interaktive Form auch häufiger als das eigentliche, statische PE ist. Gleichzeitig erschwert diese Flexibilität der Grenzen eine Zuordnung von PE als Form humangestützter maschineller Übersetzung oder maschinell-gestützter Humanübersetzung (Vieira 20219: 329).

2.4.5 Messbarkeit der Post-Editing Leistung

Eines der wichtigsten Argumente für den Einsatz von Post-Editing sind die möglichen Produktivitätsgewinne im Vergleich zur Humanübersetzung sowie der verminderte Aufwand, der etwa auf die geringere Anzahl an Tastatureingaben zurückzuführen ist. Beim Post-Editing stehen Produktivität und Aufwand jedoch in einer Wechselbeziehung zueinander. Je größer der Aufwand wird, desto stärker sinkt auch die Produktivität ab und wirkt somit dem eigentlichen Ziel des Post-Editing entgegen. In diesem Kapitel soll auf die Eigenschaften dieser beiden Begriffe eingegangen werden, die auch als häufigste Messinstrumente für die Evaluierung von PE-Leistung gelten.

2.4.5.1 Produktivität

O'Brien (2011: 198) definiert Produktivität als das Verhältnis von Qualität und Quantität der produzierten Einheiten im Verhältnis zum vorausgesetzten Aufwand pro Zeiteinheit. Produktivität wird in der Sprachindustrie üblicherweise anhand von temporalen Metriken wie Wörtern pro Minute (WPM) oder Wörtern pro Stunde (WPS) gemessen (Herbig et al. 2019: 91). PE stellt für die Sprachindustrie eine notwendige Strategie zur Bewältigung des steigenden Volumens an zu übersetzenden Texten dar, welches mit reiner Humanübersetzung und ohne maschineller Beihilfe in vielen Bereichen nicht mehr möglich wäre. Wichtigster Faktor ist dabei die Produktivität von Post-Editor*innen, die im Vergleich zu Übersetzer*innen Zieltexte mit zumindest gleichbleibender Qualität, aber in deutlich kürzerer Zeit liefern sollen (do Carmo & Moorkens 2020: 45). Die großen Akteur*innen in der

Sprachindustrie sehen Produktivität als einen der wichtigsten zeitbasierten Leistungsindikatoren, der auch herangezogen wird, um die eigene Leistung mit der Konkurrenz zu vergleichen und bei Bedarf den stündlichen Durchsatz an Wörtern zu erhöhen (do Carmo & Moorkens 2020: 49).

Nach Robert (2013: 32) ist bei der Humanübersetzung eine tägliche Produktivität von 2.000 Wörtern, beim PE jedoch von bis zu 3.500 Wörtern zu erwarten. Die tatsächliche Produktivität hängt dabei stark von externen Faktoren, wie der Qualität des MÜ-Systems ab, wonach ein Produktivitätszuwachs in einer Spanne von 500 bis 5.000 Wörtern pro Tag erreichbar sein kann (Koponen 2016: 134). Auch Guerberof Arenas (2014a: 168) belegt derart große Schwankungen, die auch sprachpaarabhängig sind und von 42% bis 131% reichen können. Auch wenn in manchen Studien, wie jener von Carl et al. (2011) mit unerfahrenen Post-Editor*innen, keine signifikanten Produktivitätszuwächse verzeichnet wurden, so ist dies nicht primär auf die fehlende Erfahrung, sondern auf die Ungeeignetheit des gewählten Textes oder MÜ-Systems zurückzuführen, da Guerberof Arenas (2014b) in ihrer Studie mit erfahrenen und unerfahrenen Post-Editor*innen keine großen Unterschiede im Produktivitätszuwachs der beiden Gruppen verzeichnen konnte. Ob eine Person durch PE tatsächlich ihre Produktivität erhöhen kann, hängt jedoch stark davon ab, ob sich ihre bevorzugte Arbeitsweise gut mit den Vorschlägen des MÜ-Systems kombinieren lässt. Dabei unterscheiden sich Post-Editor*innen weniger durch die Anzahl der Änderungen, die im Zieltext vorgenommen wurden, sondern durch ihre Geschwindigkeit und die Anzahl ihrer Tastatur- und Mauseingaben. Jene, die hier zu einem *over-editing*, einer Überbearbeitung der maschinellen Übersetzung tendieren, eliminieren die eigentliche Zeiteinsparung, die Post-Editing mit sich bringen kann (Koponen 2016: 136).

Ebenso lässt sich keine Korrelation zwischen der Dauer, die für eine PE-Aufgabe benötigt wird, und der Anzahl an vorgenommenen Änderungen feststellen. Übersetzer*innen, die in ihrer Arbeitsweise eher langsamer sind, können zwar mehr von PE profitieren als schnellere Übersetzer, die Qualität der Zieltexte ist bei beiden Gruppen jedoch vergleichbar (Guerberof Arenas 2014a: 168). Auch Post-Editor*innen, die an Texten arbeiten, die nicht ihrer Erstsprache entsprechen, können mit einem durchschnittlichen Produktivitätszuwachs von 90% deutlich stärker von PE profitieren als Personen, die in ihrer Erstsprache arbeiten. Hier muss jedoch in Relation gesetzt werden, dass Post-Editor*innen, die in ihrer Erstsprache arbeiten, ohnehin weniger Zeit für denselben Auftrag benötigen als nicht-erstsprachliche Personen. In Hinblick auf die Zeitdauer, die Erst- und Zweitsprachler*innen für dieselbe PE-Aufgabe benötigt haben, lässt sich jedoch ein ausgleichender Effekt feststellen, wonach beide

Gruppen in etwa dieselbe Dauer für das Fertigstellen benötigen (Depraetere et al. 2014: 85). Fehlende Vorerfahrung in Kombination mit einer tendenziellen Ablehnung gegenüber PE kann hingegen auch dazu führen, dass PE weniger produktiv als eine Humanübersetzung desselben Textes ist, wie eine Studie von Gaspari et al. (2014: 67) mit professionellen Übersetzer*innen zeigt. Aus den nach der Aufgabe ausgefüllten Fragebögen geht hervor, dass die Teilnehmer*innen es nicht nur bevorzugten selbst zu übersetzen, sondern diese Arbeitsweise auch als schneller einschätzten.

2.4.5.2 Aufwand

Krings (2001) unterscheidet in seiner Post-Editing-Typologie drei Arten von Aufwand: temporaler Aufwand, technischer Aufwand und kognitiver Aufwand. Die Relation dieser drei Elemente ist folgendermaßen festgelegt: *technischer A. + kognitiver A = temporaler A*. Der technische Aufwand umfasst dabei die Anzahl aller Änderungen, die mit Maus und Keyboard während des PE am Text vorgenommen werden, und ist mit Hilfe von Keyloggern und Metriken wie TER oder BLEU direkt messbar (Koponen 2016: 140), ebenso wie der temporale Aufwand, der anhand der Wörter pro Minute gemessen werden kann. Der kognitive Aufwand kann hingegen nur indirekt gemessen werden und wird als das Ausmaß von Anforderungen definiert, das von einer Aufgabe ausgeht und sich auf die mentalen Ressourcen einer Person zur Informationsbearbeitung auswirkt (Herbig et al. 2019: 92). Zur Bewältigung dieses Aufwands setzt eine Person ihre kognitive Kapazität ein, also den Gesamtbetrag an mentalen Ressourcen, die ein Individuum für einen Auftrag anwenden kann. Diese Kapazität variiert je nach Vorerfahrung einer Person, wodurch beispielsweise erfahrene Post-Editor*innen den Aufwand derselben Aufgabe niedriger als unerfahrene einschätzen werden (Vieira 2016: 42). Außerdem steuert der kognitive Aufwand die Zeit, die Post-Editor*innen aufwenden müssen, und beeinflusst die Änderung, die sie vornehmen.

Textbezogene Merkmale, die mit einem erhöhten Aufwand in Verbindung stehen, sind unter anderem Satzlänge und -struktur sowie Fehler in der Rohübersetzung bezüglich Syntax, Wortreihenfolgen oder inkorrekt übersetzten Wortspielen oder Redewendungen (Koponen 2016: 142). In der Forschung ist man dennoch dazu übergegangen, den kognitiven Aufwand nicht am zu bearbeitenden Text, sondern direkt an den Post-Editor*innen zu messen. Zwar wurde gezeigt, dass längere Sätze beim PE mit einem höheren kognitiven Aufwand verbunden sein können, jedoch lässt dies nur schlecht auf den temporalen Aufwand schließen (Vieira 2019: 323). Stattdessen wird beispielsweise das Pausenverhalten beim Tippen herangezogen, da man davon ausgeht, dass eine Ansammlung kurzer Pausen (< 1s) innerhalb eines Segments ein besseres Indiz für kognitiv aufwändige Stellen sein kann, da der/die Post-

Editor*in diese Pausen zum Nachdenken benötigt (Lacruz & Shreve 2014: 266). Eine weitere Möglichkeit ist die Verwendung von vier- oder fünfstufigen Skalen, auf denen die Post-Editor*innen ankreuzen sollen, in welchem Umfang jedes Segment eines Textes editiert werden musste, um einen qualitativen Zieltext zu erhalten. Diese Methode ist jedoch stark subjektiv und kann je nach Qualität der maschinellen Rohübersetzung und Vorerfahrung der Post-Editor*innen inkonsistente Einschätzungen des Aufwands liefern (Koponen 2016: 139f). In der Pionierstudie von O'Brien (2006) über das Editieren von Fuzzy Matches wurde erstmals ein Eyetracker verwendet, um den mit PE verbundenen kognitiven Aufwand zu messen. Eyetracking bewies sich hier als relativ kostengünstige, umfangreiche und unaufdringliche Technik für die Erhebung des Aufwands und kommt seitdem immer stärker in Studien zum Einsatz (Moorkens 2018: 56). Dabei wird primär die Anzahl, Dauer und Positionierung von Fixationen herangezogen, da davon ausgegangen wird, dass diese Aufschluss über die Dauer geben, die der/die Post-Editor*in benötigt, um ein Wort oder eine Phrase kognitiv zu verarbeiten (Jakobsen 2019: 402). Zusätzlich ist es hier üblich, bei der Auswertung der Eyetracking-Daten *Areas of Interest* (AOI), also Interessensbereiche auf der Benutzeroberfläche, in der post-editiert wird, zu definieren, etwa in den Bereichen des Ausgangs- und Zieltextes. Somit können Anzahl und Dauer von Fixationen, die in einen dieser Bereiche fallen, separat ausgewertet werden und Aufschluss über das PE-Verhalten geben, beispielsweise ob der/die Post-Editor*in sich mehr auf Ausgangs- oder Zieltext fokussiert (Moorkens 2018: 59). Während die visuelle Aufmerksamkeit beim Übersetzen beinahe gleichmäßig zwischen Ausgangs- und Zieltext aufgeteilt ist, so kann sie beim Post-Editing bis zu sechsmal länger im Bereich des Zieltextes sein. Dies lässt darauf schließen, dass der Ausgangstext herangezogen wird, um die inhaltliche Korrektheit der maschinellen Rohübersetzung zu prüfen, der Fokus der Post-Editor*innen danach allerdings auf dem Editieren des Zieltextes liegt, unter vereinzelter Betrachtung des Ausgangstextes (Mesa-Lao 2014: 231).

Auch die durchschnittliche Fixationsdauer innerhalb eines Interessensbereichs kann ein Indiz für kognitiven Aufwand sein, da beispielsweise die durchschnittliche Fixation beim Editieren des Zieltextes bis zu doppelt so lange wie die durchschnittliche Fixation beim Lesen des Ausgangstextes sein kann (Alves et al. 2016: 131). Auch wenn die durchschnittliche Fixationsdauer beim Schreiben bzw. Tippen nicht pauschal definiert werden kann, so wird jene beim Lesen in den meisten Fällen mit 200-250 ms angegeben (Jakobsen 2019: 401). Die Fixationen im Zieltext sind nicht nur auf Grund der kognitiven Prozesse, die für das Editieren und Formulieren nötig sind, um einiges länger, sondern auch, weil der simultan ablaufende

technische Aufwand des Tippens per Tastatur mehr Zeit erfordert und die Augen somit länger an derselben Stelle verharren müssen, bis die Eingabe abgeschlossen ist (Alves et al. 2016: 132). Ebenso aussagekräftig kann das Berechnen der durchschnittlichen Fixationsdauer pro Segment sein, welche sich aus der durchschnittlichen Fixationsdauer geteilt durch die durchschnittliche Anzahl an Änderungen pro Segment ergibt (Alves et al. 2016: 129). Eine hohe Fixationsdauer hat dabei mit der kognitiven Bereitschaft des/der Post-Editor*in zu tun, denn je weniger er/sie das Auftreten eines Ereignisses, etwa eines Wortes, erwartet, umso länger wir er/sie sich darauf im Falle des Auftretens fixieren. Das Analysieren und Verarbeiten eines unerwarteten Ereignisses ist somit mit einem höheren kognitiven Aufwand verbunden und Interessensbereiche mit einer höheren durchschnittlichen Fixationsdauer können auf eine Vielzahl an solchen unerwarteten Ereignissen schließen lassen (Jakobsen 2019: 402). Vereinzelt wurde in der Forschung auch die Annahme aufgegriffen, dass der Durchmesser der Pupillen ein Indiz für den kognitiven Aufwand sein kann, so zuletzt von Fernández et al. (2016). Auch wenn teils belegt werden konnte, dass der Pupillendurchmesser bei höherem kognitiven Aufwand größer wird, so konnte die Kausalität dieser Beobachtung nie eindeutig geklärt werden. Dennoch kann es von Vorteil sein, die Pupillendurchmesser im Kontext mit anderen Messergebnissen zu betrachten, da es sich dabei um unbewusst gesteuerte Vorgänge handelt, die diverse kognitive Verarbeitungsprozesse nach außen hin sichtbar machen (Jakobsen 2019: 402).

2.4.6 Post-Editing im professionellen Kontext

Sakamoto und Yamada (2020: 89) definieren im Zusammenhang mit dem PE maschinell übersetzter Texte im professionellen Kontext vier soziale Gruppen, nämlich Projektmanager*innen, Auftraggeber*innen, das Management von Sprachdienstleister*innen und Übersetzer*innen. Die Interaktion dieser Gruppen miteinander wurde durch die Entwicklung dieser neuen Technologien stark beeinflusst und hat nicht zuletzt zu Meinungsverschiedenheiten bezüglich Qualität, Bezahlung und Abgabefristen geführt. Vor allem professionelle Übersetzer*innen, die durch ihre Qualifikationen immer mehr in die Rolle der Post-Editor*innen gedrängt werden, stellen in der aktuellen Forschung die interessanteste Gruppe dar. Dennoch gestaltet sich ihre Erforschung als schwierig, da in der professionellen Übersetzungsgemeinschaft mitunter Misstrauen gegenüber dem PE und der MÜ herrscht. Vereinfacht gesagt kann man zwischen jenen Übersetzer*innen unterschieden, die PE-Aufträge annehmen, und jenen, die sie ablehnen. In ihrer Studie mit 28 Teilnehmer*innen aus den oben erwähnten vier sozialen Gruppen fanden Sakamoto und Yamada (2020: 88), dass sprachliche und fachliche Kompetenz sowie

Übersetzungskompetenz wichtige Kriterien eines/r guten Post-Editor*in sind. Die Teilnehmer*innen der Studie gaben jedoch auch an, dass Revisor*innen, Terminolog*innen oder Projektmanager*innen genauso gut für die Aufgabe geeignet sein können, vorausgesetzt sie haben eine PE-Ausbildung erhalten. Flexibilität, Genauigkeit und die Bereitschaft Vorgaben zu befolgen, wurden als wichtigste Eigenschaften von Post-Editor*innen genannt, wodurch auch klar wird, warum viele Übersetzer*innen, die die Kreativität, Freiheit und Vielfältigkeit ihrer Tätigkeit schätzen, dem PE abgeneigt sind. Unter den befragten Übersetzer*innen der Studie herrschte der Konsens, dass PE und Humanübersetzen zwei verschiedene Tätigkeiten mit unterschiedlichen Anforderungen und Abläufen sind (2020: 90). Jedoch ist eine gewisse Wertung zu erkennen, da einige Übersetzer*innen angaben, PE wäre für angehende, unerfahrene Übersetzer*innen ein guter Berufseinstieg und würde es ihnen ermöglichen, ihre sprachlichen Kompetenzen zu erweitern, um eventuell zu „vollwertigen" Übersetzer*innen zu werden.

Trotz ihres Misstrauens sehen sich viele Übersetzer*innen in ihrem Arbeitsalltag dennoch mit PE konfrontiert, da auch immer mehr Sprachdienstleister*innen PE in ihre Arbeitsprozesse integrieren und Aufträge vergeben, die PE benötigen (Silva 2014). In den Leistungsspektren großer Sprachdienstleister*innen reiht sich PE neben Korrektorat, Untertitelung, Copywriting und Transkreation ein und wird vor allem von freiberuflichen Übersetzer*innen angenommen, die daran interessiert sind, ihr Wissen im Bereich der MÜ zu erweitern. Erfahrene In-house Übersetzer*innen sind diesen Entwicklungen meist weniger aufgeschlossen, was oft auf fehlendes Wissen zurückzuführen ist. Sie befürchten, dass sie mit maschinellen Rohübersetzungen von schlechter Qualität arbeiten müssen und für diese Aufgabe auch nicht ausreichend bezahlt werden. Wie bereits erläutert, verstehen viele Übersetzer*innen PE und Revision als Synonyme, auch wenn die beiden Arbeiten gänzlich andere Prozesse umfassen. In diesem Zusammenhang wäre Aufklärungsarbeit notwendig, um skeptischen Übersetzer*innen zu kommunizieren, dass PE als Sonderform des Übersetzens gesehen werden kann, bei dem menschliche Bearbeiter*innen die Übersetzung anhand von Übersetzungsvorschlägen, jedoch durch ihre eigenen Entscheidungen anfertigen. Die Handlungsfähigkeit von Übersetzer*innen ist zwar nicht im selben Maß wie bei der Humanübersetzung gegeben, aber dennoch vorhanden. Diese Einschränkung der Handlungsfähigkeit wird jedoch durch den möglichen Produktivitätszuwachs und die höhere Effizienz der Arbeitsweise ausgeglichen (Zaretskaya 2017: 117).

2.4.6.1 Berufsprofil von Post-Editor*innen

Das Berufsprofil eines/r Post-Editor*in zu definieren, bringt die Herausforderung mit sich, dass es sich bei den meisten Post-Editor*innen um professionelle Übersetzer*innen handelt, die sich in diesem Bereich spezialisiert haben und bereits viele Qualifikationen und Erfahrungen mitbringen. Sowohl in der Forschung als auch in der Sprachindustrie scheint der Konsens zu herrschen, dass Übersetzer*innen tendenziell am besten für diese Aufgabe geeignet sind, da nur sie die Genauigkeit einer Übersetzung einschätzen und die Fehler der MÜ erkennen können sowie ein fundiertes Wissen über sprachlichen Transfer besitzen (Torrejón & Rico 2013: 167). Um statisches, bilinguales PE erfolgreich in Übersetzungsprozesse einbinden zu können, ist es also auch wichtig, die Aufgabenbereiche und das Kompetenzprofil eines/r Post-Editor*in zu definieren, die sich in vielerlei Hinsicht von jenen des/r Übersetzer*in unterscheiden. Zusätzlich ist ein/e Post-Editor*in auch für sämtliche mit dem PE zusammenhängende Aufgaben verantwortlich. Als zentrale Handlungsfigur koordiniert er/sie diverse Prozesse und steuert Abläufe, die in Zusammenhang mit dem Computer stehen, da mit aktuellem Entwicklungsstand kein System in der Lage ist, unüberwacht und eigenständig zu arbeiten (Guerrero Romeo 2018: 89f).

In der Post-Editing Abteilung von Sprachdienstleister*innen könnten Vollzeit Post-Editor*innen nach Guerrero Romeo (2018) folgende drei Aufgabenbereiche übernehmen müssen. Als erster Schritt jedes PE-Prozesses muss ein maschinell übersetzter Text erzeugt werden. Es kann die Aufgabe von Post-Editor*innen sein, Projektmanager*innen beim Training einer MÜ-Engine zu unterstützen, indem sie ihre Expertise nutzen, um die korrekten Bitexte für Parallelkorpora auszuwählen, Terminologiedatenbanken einzubinden und Gewichtungen zu vergeben. Nach der Trainingsphase gilt es die Qualität der Engine zu evaluieren, indem beispielsweise Qualitätsmetriken wie *BLEU* oder *TER* angewandt werden, oder indem eine menschliche Evaluation anhand von Fehlerkriterien durchgeführt wird. Es können auch Rückmeldungen von Übersetzer*innen eingeholt werden, die durch die Engine übersetzte Texte zur Bearbeitung erhalten haben. Nach vollständiger Evaluierung kann der/die Post-Editor*in sein/ihr gesammeltes Wissen in die Verbesserung der gewählten Engine einfließen lassen, sich für eine andere Architektur an MÜ-System (heutzutage primär statistisch oder neuronal) entscheiden, oder auch beschließen, dass PE für den vorliegenden Auftrag nicht sinnvoll ist. Nach Auswahl der Engine werden in diesem Schritt auch die zu befolgenden PE Richtlinien ausgewählt (Guerrero Romeo 2018: 92). In einem zweiten Schritt wird das Post-Editing durchgeführt, wobei hier als erstes auch ein Pre-Editing des Ausgangstextes notwendig sein kann. Das Wissen des/der Post-Editor*in über das

ausgewählte und trainierte MÜ-System kann hier helfen, effizient jene Stellen auszubessern, die dem System Probleme bereiten könnten.

Anschließend wird das Post-Editing mit all seinen in Kapitel 2.4.3 beschriebenen Schritten durch einen oder mehrere Post-Editor*innen durchgeführt. Weiters ist der/die Post-Editor*in für die Qualitätskontrolle des Zieltextes zuständig, was auch eine Revision durch andere Post-Editor*innen oder Übersetzer*innen beinhalten kann. Im dritten und letzten Schritt werden Rückmeldungen von allen beteiligten Post-Editor*innen eingeholt, die sich auf häufige Fehlertypen in der MÜ und auf fehlende oder verbesserungsbedürftige Angaben in den Richtlinien beziehen. Nach Lieferung des Textes an den/die Auftraggeber*in kann auch von diesem/r Rückmeldung eingeholt werden und beispielsweise Wünsche bezüglich der Terminologie abgespeichert und für künftige Aufträge übernommen werden (Guerrero Romeo 2018: 93f). Rückmeldungen bezüglich MÜ-Fehlern sollen den manuellen Arbeitsaufwand bei künftigen Projekten verringern und können sich auf Fehler in Zeichensetzung, Terminologie oder Inhalt beziehen. Dabei sollten Rückmeldungen zu einem Segment so spezifisch wie möglich sein, also sich auf das exakte, betroffene Wort beziehen, und auch so objektiv wie möglich sein, also nicht auf stilistische Präferenzen der Post-Editor*innen (Zaretskaya 2017: 121).

Abbildung 4: Kompetenzprofil für Post-Editing

Das in Abb. 4 dargestellte Kompetenzprofil für Post-Editing lässt sich durch die Verbindung der Kompetenzprofile von Torrejón und Rico (2013: 169f) und Nitzke et al. (2019: 250) in die vier Bereiche Grundkompetenzen, sprachliche Kompetenzen, translatorische Kompetenzen und technische Kompetenzen unterteilen. Die darin enthaltenen Kompetenzen sollen anschließend kurz beschrieben werden und spiegeln auch die aktuellen Anforderungen wider, die in der Sprachindustrie an Post-Editor*innen gestellt werden und von Cid (2020a) in einer Studie identifiziert wurden. Viel wichtiger jedoch ist, dass in einer Folgestudie von Cid (2020b) dieselben Anforderungen auch von Sprachdienstleistern*innen und professionellen Übersetzer*innen akzeptiert und als wichtig empfunden wurden, was für die Anwendbarkeit dieses Modells spricht.

Die Grundkompetenzen eines/r Post-Editor*in umfassen zum einen die *Kompetenz zur Abschätzung von Risiken*, die es dem/der Post-Editor*in erlaubt unter Berücksichtigung von Anwendungsbereich und Lebensdauer des Textes die mit ihm verbundene Risikostufe einzuschätzen. Dies ermöglicht weiters den Umgang mit subjektiven Qualitäts- und Anforderungskriterien sowie Unklarheiten seitens der Auftraggeber*innen (Nitzke et al. 2019: 248). Die *strategische Kompetenz* hilft dabei, fundierte Handlungen zu setzen, etwa bei der Entscheidung für ein leichtes oder vollständiges PE, beim Auswählen des besten MÜ-Vorschlags, beim Entscheiden nach der *2-Sekunden-Regel*, ob ein Segment bearbeitet oder neu übersetzt wird, und vor allem beim Einhalten der PE-Richtlinien, die mit dem Auftrag verbunden sind (Torrejón & Rico 2013: 170). Die *Beratungskompetenz* ermöglicht es dem/der Post-Editor*in, das abgeschätzte Risiko und die strategisch gewählte Vorgangsweise an den/die Projektleiter*in oder Kund*innen zu kommunizieren und diese über mögliche Konsequenzen zu informieren. Die *Dienstleistungskompetenz* bedeutet, dass Post-Editor*innen in der Lage sind, die Kosten für einen PE-Auftrag angemessen anhand von Aufwand und Qualität der MÜ zu berechnen und dem/der Projektleiter*in oder Auftraggeber*in diesen Preisvorschlag transparent zu präsentieren. Neben Verhandlungskompetenz hat der/die Post-Editor*in zum/r Auftraggeber*in eine vermittelnde oder informierende Rolle bezüglich der angebotenen Dienstleistung, weswegen der/die Post-Editor*in mit der Funktionsweise des Übersetzungsmarktes vertraut sein soll (Nitzke et al. 2019: 248).

Die sprachlichen Kompetenzen umfassen sämtliche Kompetenzen, die auch von Übersetzer*innen gefordert werden. Durch die *bilinguale Kompetenz* beherrscht der/die Post-Editor*in Ausgangs- und Zielsprache auf vergleichbarem Niveau, durch die *extralinguistische Kompetenz* verfügt er/sie über Textsorten- und Fachbereichswissen und kann mit Hilfe

seiner/ihrer transkulturellen Kompetenz Unterschiede und Gemeinsamkeiten dieser Bereiche in beiden Sprachen identifizieren. Durch seine/ihre sprachliche Expertise verfügt ein/e Post-Editor*in auch über *Revisionskompetenz* und hat Strategien für die Korrektur von Texten anderer Autor*innen, ohne dabei zu viele oder zu wenige Änderungen vorzunehmen (Nitzke et al. 2019: 249).

Die translatorischen Kompetenzen umfassen die *Übersetzungskompetenz*, die sich primär auf die Vertrautheit des/r Post-Editor*in mit Abläufen im Übersetzungsprozess bezieht. Er/sie ist daher mit Auftragsmodellen und Gestaltungsrichtlinien vertraut, verfügt über translationswissenschaftlich relevantes, theoretisches Wissen, wie etwa Skopos, Textsortenkonventionen und Zielgruppengerechtheit, und verfügt über Lösungsstrategien für sprachlich oder kulturell sensible Textstellen. Die *MÜ-Kompetenz* hilft dem/r Post-Editor*in, Funktionsweise, Arten und häufige Fehlertypen von MÜ-Systemen zu verstehen und somit potentielle Probleme frühzeitig zu erkennen und die Qualität zu verbessern. Weiters verfügt er/sie noch über *PE-Kompetenz*, weiß also über den Ablauf des PE-Prozesses Bescheid, kann Änderungen gemäß der vorgeschriebenen Richtlinien vornehmen und ist mit den gängigen Fehlertypen vertraut, die beispielsweise durch NMÜ-Systeme erzeugt werden, und kann diese gezielt suchen und bearbeiten (Nitzke et al. 2019: 250).

Die technischen Kompetenzen umfassen die *Recherchekompetenz*, die es dem/r Post-Editor*in ermöglicht, Fachbereich spezifische Glossare, Terminologiedatenbanken und Parallelkorpora zu finden und diese in den Arbeitsprozess einzubinden. Bei der manuellen Evaluierung der Übersetzungsqualität eines MÜ-Systems kann auch Recherche nötig sein. Besonders wichtig ist die *instrumentelle Kompetenz*, denn ein/e Post-Editor*in muss stets über die gängigsten, am Markt verbreiteten Anwendungen und Programme Bescheid wissen und diese anwenden können (Nitzke et al. 2019: 249). Diese umfassen das Verständnis von und Wissen über die Funktionsweisen von desktop- und cloudbasierten MÜ-Systemen, CAT-Tools, Terminologiemanagementsystemen, Programmen zur Alignierung oder QA-Tools. Der/die Post-Editor*in soll im Umgang mit diesen Anwendungen möglichst geschult und effizient sein, um seine/ihre Produktivität zu maximieren. Dies kann auch grundlegende Programmierkompetenzen voraussetzen, um sich beispielsweise Tastenkürzel oder Makros für häufig genutzte Eingaben und Aktionen zu erstellen, beispielsweise das Löschen ganzer Segmente, die Funktion *Suchen und Ersetzen* oder das Übernehmen von Vorschlägen (Zaretskaya 2017: 122).

2.4.6.2 Einfluss von Berufserfahrung auf Post-Editing

Der Beruf der Post-Editor*innen ist gewissermaßen als Weiterentwicklung des Übersetzerberufes zu sehen, da er sich in seinem Umfang und den benötigten Kompetenzen mit diesem teilweise deckt. Nach wie vor gibt es nur eine sehr begrenzte Anzahl an Ausbildungsmöglichkeiten für PE, wodurch diese Arbeit primär von professionellen Übersetzer*innen ausgeführt wird. Da ihre Eignung dafür immer wieder erwähnt wird, soll in diesem Kapitel kurz ausgeführt werden, welchen Einfluss die Erfahrung aus dem Übersetzer*innenberuf auf das Post-Editing hat.

In ihrer Studie mit 24 professionellen Übersetzer*innen stellt Guerberof Arenas (2014b) die Hypothese auf, dass Übersetzer*innen, die technische Erfahrung haben und mit der Lokalisierungsindustrie vertraut sind, auch beim PE eine höhere Produktivität, gemessen anhand der Wörter pro Minute, haben werden. Obwohl die Teilnehmer*innen im Durchschnitt mindestens zwei, in manchen Fällen auch mehr als acht Jahre Berufserfahrung als Übersetzer*innen hatten, in unterschiedlichen Bereichen tätig waren und mit CAT-Tools vertraut waren, so hatten 65% keine oder kaum Arbeitserfahrung mit PE. 66,7% der Teilnehmer*innen gaben an, dass innerhalb der letzten drei Jahre auch weniger als 25% ihrer Arbeit auf PE entfiel. Trotz dieser Streuung unter den Teilnehmer*innen zeigen die Ergebnisse der Studie, dass unerfahrenere Übersetzer*innen zwar eine langsamere Bearbeitungsgeschwindigkeit aufwiesen, diese im Vergleich zu den erfahrenen Übersetzer*innen jedoch nur wenige Prozent betrug. Im Rahmen des Experiments mussten abwechselnd MÜ-Vorschläge und Fuzzy Matches zwischen 85-94% post-editiert werden. Die unerfahrenen Übersetzer*innen machten bei der Bearbeitung der Fuzzy Matches zwar mehr Fehler, jedoch nicht bei MÜ-Vorschlägen. Dies deutet darauf hin, dass die Verwendung von MÜ-Vorschlägen im PE-Prozess einen ausgleichenden Effekt haben kann, wodurch die Berufserfahrung professioneller Übersetzer*innen keine allzu große Rolle spielt und die Tätigkeit des PE schon bereits zwei Jahre nach absolvierter Ausbildung effektiv durchgeführt werden kann (Guerberof Arenas 2014b: 72).

Auch wenn die Ergebnisse dieser Studie in sich einheitlich waren, so lassen sie nicht darauf schließen, dass jede/r gute Übersetzer*in auch automatisch ein/e gute/r Post-Editor*in ist. Auch wenn sich viele der Kompetenzen der beiden Disziplinen überschneiden, so unterscheiden sie sich primär durch die Geschwindigkeit, mit der Entscheidungen getroffen werden müssen, und die Anzahl an Änderungen, die nötig sind, um einen qualitativen Zieltext zu produzieren. Während Übersetzer*innen auf Basis von Fuzzy Matches arbeiten, die von Menschen übersetzt und auf ihre Qualität geprüft sind, arbeiten Post-Editor*innen mit

ungeprüften, maschinell erzeugten Vorschlägen von manchmal mangelhafter Qualität, die mit dem Ausgangstext abgeglichen werden müssen, bevor sie bearbeitet werden können. Übersetzer*innen sind es zudem gewohnt, stets eine publikationsreife Übersetzung von höchster Qualität zu erzeugen und dafür auch mehr Zeit aufzuwenden, wohingegen Post-Editor*innen an die vorgegebenen Qualitätsrichtlinien gebunden sind und bei einem leichten PE etwa nur die gröbsten Fehler beheben sollen (Aranberri 2017: 90f). Die Eignung zum/r Post-Editor*in hängt daher zum einen von der Übersetzungserfahrung und einer spezifischen PE-Ausbildung ab, zum anderen aber auch von der persönlichen Präferenz oder Bereitschaft für die Arbeitsweise des PE, die normierter und restriktiver als das klassische Übersetzen abläuft.

2.4.6.3 Einstellung professioneller Übersetzer*innen zu Post-Editing

Heutzutage müssen Übersetzer*innen stets größere Volumina an Texten in den verschiedensten Domänen bearbeiten, was bereits durch die Einführung von CAT-Tools in den 1990er Jahren vereinfacht wurde. MÜ und PE ermöglichen nun unter Voraussetzung der korrekten Kombination an den Parametern MÜ-Qualität, Abgabefrist, Textsorte, Erfahrung und Einstellung der Post-Editor*innen auch einen Zuwachs an Produktivität. Gleichzeitig werden PE-Aufträge nur mit 60-65% des Wortpreises einer Übersetzung bezahlt, die Auftraggeber*innen erwarten jedoch dieselbe, wenn nicht sogar eine höhere Produktivität, da die Leistung des Menschen nicht schlechter als die der Maschine sein sollte (do Carmo & Moorkens 2020: 38). Dabei soll die Qualität identisch mit der einer Humanübersetzung sein und dennoch so viel wie möglich von der maschinellen Rohübersetzung beibehalten werden. Der Vormarsch von PE am Übersetzungsmarkt wird von Übersetzer*innen teils kritisch gesehen, da diese ein gänzliches Verschwinden der Humanübersetzung und somit auch ihrer Handlungsfähigkeit befürchten (Sakamoto 2019: 201). Mit Aufkommen der neuronalen maschinellen Übersetzung sind die Einstellungen professioneller Übersetzer*innen zu PE außerdem zu einer zentralen Frage der Forschung geworden. Dabei wird die Disziplin selbst in der Forschung oft mit den Begriffen Misstrauen, Widerstand oder Zukunftsangst in Verbindung gebracht, was durch mehrere Forschungsarbeiten mittels Fragebögen und Interviews belegt wurde.

Vidal et al. (2020) und Bundgaard (2017) führten Studien mit vier bzw. acht angestellten Übersetzer*innen durch, die zusätzlich zur Befragung auch kurze Post-Editing Aufträge absolvieren mussten. Pérez Macías (2020) und Guerberof Arenas (2013) führten reine Fragebogenstudien mit 104 bzw. 24 freiberuflichen und angestellten Übersetzer*innen durch, die alle zu ähnlichen Aspekten befragt wurden. Die Teilnehmer*innen der Studie von

Vidal et al. (2020) gaben vor dem Auftrag an, sich nicht auf PE zu freuen und auch Angst davor zu haben, dass MÜ Übersetzer*innen in Zukunft ersetzen könnte oder sie möglicherweise zu Post-Editor*innen machen würde. Sie gaben jedoch an, dass sie sich vom MÜ-System eine gute Qualität erwarten, die wiederum den Arbeitsschritt des PE angenehmer machen würde (Vidal et al. 2020: 54f). Nach dem Auftrag gab die Hälfte der Teilnehmer*innen an, sie könnte sich vorstellen mehr PE-Aufträge anzunehmen, solange die Bezahlung stimmt. Die andere Hälfte bewertete die Erfahrung negativ und fühlte sich in ihrer Kreativität eingeschränkt, sie bestätigten jedoch einen wahrgenommenen Produktivitätszuwachs. Alle Teilnehmer*innen waren mit ihrem Endprodukt zufrieden und gaben an, dass sie produktiver als bei der Humanübersetzung waren, die Qualität jedoch dieselbe sei, auch wenn sie das Übersetzen dem PE vorziehen würden (Vidal et al. 2020: 56). Auch die Teilnehmer*innen der Studie von Bundgaard (2017) gaben an, dass einige Segmente fast ohne Bearbeitung übernommen werden konnten oder alternativ eine gute Inspiration für die eigene Übersetzung boten, wodurch eine Zeitersparnis möglich war (Bundgaard 2017: 132). Gleichzeitig lieferte die MÜ auch Vorschläge, die nicht brauchbar waren, da manche Elemente gar nicht oder terminologisch falsch übersetzt worden waren (Bundgaard 2017: 133). Unter den Teilnehmer*innen herrschte der Konsens, dass PE und Humanübersetzen zwei verschiedene Disziplinen sind, wobei bei ersterer die kreative Komponente gänzlich fehlt und das ständige Vergleichen der beiden Texte kognitiv anstrengender ist. Ebenso fühlten sie sich durch die MÜ-Vorschläge gefangen und zurückgehalten, da die Lösung scheinbar schon vorgegeben war und der Ablauf nicht jenem des sonst vertrauten Übersetzungsprozesses entsprach (138f).

In Pérez Macías (2020) Studie sahen 61% der 122 Befragten PE als eine Möglichkeit das eigene Portfolio auf dem Übersetzungsmarkt zu erweitern, 26% sahen es hingegen als Gefahr für den eigenen Beruf. Unter den Teilnehmer*innen herrschte jedoch der Konsens, dass PE keine Modeerscheinung ist und Übersetzer*innen es in ihren Arbeitsprozess integrieren müssen, um die wachsenden Textvolumina zu bewältigen (Pérez Macías 2020: 26). 73% bevorzugten die Arbeit mit CAT-Tools, 16% hingegen das reine Post-Editing. Dies hat jedoch mit der Vertrautheit mit den diversen Softwareanwendungen zu tun. Selbst wenn Übersetzer*innen dem PE gegenüber aufgeschlossen sind, kann das Arbeiten in einer unausgereiften, kontraintuitiven Softwareumgebung ausreichen, um sie umzustimmen. Werden Post-Editing Funktionen jedoch in vertraute CAT-Tools integriert, kann auch bei den Übersetzer*innen eine größere Akzeptanz geschaffen werden (Pérez Macías 2020: 29). In der Studie von Guerberof Arenas (2013) ist bei der Frage nach der Einstellung zu PE eine breite

Streuung zu beobachten. Zwar herrscht eine Tendenz zur Mitte, die Ablehnung des PE überwiegt jedoch leicht. Einige genannte Gründe sind fehlende Vorerfahrung mit und Ausbildung für Post-Editing sowie die fehlende Freude an und Schwierigkeit mit der Tätigkeit, auch wenn die Produktivität merkbar gesteigert wird. Für das PE sprechen hingegen die bessere Ergonomie, beispielsweise durch weniger Tastatureingaben und eine bessere terminologische und inhaltliche Konsistenz, die für repetitive Texte hilfreich ist (Guerberof Arenas 2013: 83). Was in all diesen Studien bereits anklingt, ist die Wichtigkeit der Ergonomie beim Übersetzungsprozess, die sich auf die Effizienz und Produktivität auswirkt, mit denen eine Aufgabe erledigt werden kann.

In einer Studie mit 70 Übersetzer*innen aller 24 Amtssprachen der DGT (Generaldirektion Übersetzung) der Europäischen Kommission erforschten Cadwell et al. (2016) deren Einstellungen zur Nutzung und Bearbeitung von Vorschlägen aus dem hauseigenen hybriden statistischen MÜ-System MT@EC. Diese Ergebnisse wurden von Cadwell et al. (2018) in einer Folgestudie mit den Einstellungen von 20 Übersetzer*innen des britischen Sprachdienstleisters ALPHA CRC verglichen, um die Unterschiede zwischen diesen zwei Gruppen an Übersetzer*innen aufzuzeigen, die unterschiedliche Zugänge zu MÜ und PE haben. Die häufigsten Gründe, die von beiden Gruppen für eine Verwendung von MÜ genannt wurden, waren die Steigerung von Produktivität, die ausreichende Qualität der rohen MÜ für manche Sprachpaare und Textsorten, die Nutzung des Vorschlags als Inspiration für eigene Übersetzungen, die Zeitersparnis durch das Kopieren von Vorschlägen und die Nützlichkeit in Fällen, wo das hauseigene TM nur wenige Treffer für einen Text liefert (Cadwell et al. 2016: 235). Die ALPHA-Gruppe erwähnte zusätzlich, dass die Implementierung von MÜ unaufhaltbar sei und dass auch vermehrt vorausgesetzt wird, dass Übersetzer*innen MÜ in ihren Arbeitsprozess aufnehmen und zu nutzen lernen (Cadwell et al. 2018: 310). Gegen die Verwendung von MÜ sprach hingegen bei beiden Gruppen die als schlecht wahrgenommene Qualität der MÜ, die neuartigen Fehler von NMÜ-Systemen, die beim PE übersehen werden können, die erhöhte Konzentration, die PE voraussetzt, sowie die Angst, durch die Maschine ersetzt zu werden (Cadwell et al. 2016: 237). Die wichtigsten Gründe, die nur von der ALPHA-Gruppe erwähnt wurden, waren, dass PE langsamer als Humanübersetzen ist, dass es keinen Spaß mache, und dass ein gutes Translation Memory dieselben Vorteile wie MÜ bringen würde (Cadwell et al. 2018: 311).

Die meisten der genannten Gründe haben mit der Ergonomie der Tätigkeit, also dem Wohlbefinden der Übersetzer*innen zu tun, die an der Optimierung ihres eigenen Arbeitsprozesses interessiert sind und wirtschaftliche oder finanzielle Faktoren auf größerer

Ebene weniger bewusst bedenken (Cadwell et al. 2016: 238). In beiden Gruppen erwähnten nämlich nur 25% der Teilnehmer*innen, dass MÜ die Arbeit der Übersetzer*innen entwerten würde und, dass Übersetzer*innen der MÜ nicht vertrauen würden. Auch hier geht es also primär um Ergonomie und Produktivität anstatt um Ansehen oder Prestige des Berufes (Cadwell et al. 2018: 311). Die DGT-Gruppe war gegenüber MÜ positiver eingestellt und an der Verwendung der Technologie interessiert. Durch die Implementierung eines einheitlichen Systems in ihrer Organisation konnten sie dessen Vor- und Nachteile an eigener Hand erleben. Die ALPHA-Gruppe hatte diese Möglichkeit auf Grund des Fehlen eines solchen Systems bei ihrem/r Sprachdienstleister*in nicht. Sie äußerte sich zufrieden mit ihrer gängigen Arbeitsweise der computergestützten Humanübersetzung, blickte den zukünftigen Entwicklungen der MÜ jedoch kritisch entgegen. Diese Studie zeigt somit die Wichtigkeit der Vertrautheit mit MÜ-Systemen auf, die helfen kann, Vorurteile zu reduzieren und die Akzeptanz gegenüber der Technologie zu erhöhen (Cadwell et al. 2018: 312). Auch die oben genannten Studien decken sich mit diesen Erkenntnissen, da die befragten Übersetzer*innen primär Ergonomie, Produktivität und Optimierung des eigenen Arbeitsprozesses als wichtigste Faktoren nannten, die für oder gegen die Akzeptanz von PE sprechen. Im modernen Übersetzungsprozess findet PE wegen fehlender dedizierter Software meist in CAT-Tools statt, wobei Übersetzer*innen neben den gewohnten TM-Treffern auch MÜ-Vorschläge angezeigt bekommen und so mehr Möglichkeiten haben, diese zu übernehmen oder als Inspiration zu nutzen. Nach einer Studie von Guerberof Arenas (2014a) unterscheidet sich das Post-Editieren von 85-94%en TM-Treffern und von MÜ-Vorschlägen kaum in Bezug auf Fehleranzahl, inhaltlicher Korrektheit und Stil. Darüber hinaus wurde PE in vergangenen Jahren nicht nur in die Curricula translatorischer Ausbildungen aufgenommen, sondern auch in die Arbeitsprozesse großer Sprachdienstleister*innen. So war PE 2019 bereits fester Bestandteil in 10% der weltweiten Sprachdienstleister*innen, die Dunkelziffer an PE, das von freiberuflichen Übersetzern oder kleinen Übersetzungsbüros durchgeführt wird, dürfte jedoch viel höher sein und kippt das in der Forschung vorherrschende Narrativ der Übersetzer*innen, die sich strikt gegen PE wehren, ins Positive. Gleichzeitig zeigt es auch die Wichtigkeit einer eigenständigen PE-Ausbildung und der regelmäßigen Auseinandersetzung mit den neuen Technologien durch Übersetzer*innen auf (do Carmo & Moorkens 2020: 39).

2.4.7 Post-Editing im nicht-professionellen Kontext

Wie im Laufe dieser Arbeit bisher mehrmals erwähnt, ist PE zu einer wichtigen Kompetenz im Portfolio von Übersetzer*innen geworden, die trotz ihrer teils gemischten Einstellung gegenüber der Digitalisierung und Automatisierung des Übersetzungsprozesses die

Produktivitätsgewinne durch PE anerkennen und durch ihre Übersetzungsausbildung einen Teil der für PE erforderlichen Kompetenzen bereits mitbringen. Durch den rasanten Zuwachs an Nachfrage und Wichtigkeit, den PE in den letzten Jahren erlebt hat, haben nur die wenigsten all jener Übersetzer*innen, die momentan als Post-Editor*innen tätig sind, diese Disziplin im Rahmen ihrer Ausbildung erlernen können, da sie selbst mit momentanem Stand kein Bestandteil der Curricula mancher translationswissenschaftlicher Hochschulen ist (Daems et al. 2017: 248). Folglich haben sich viele ihre PE-Kompetenzen autodidaktisch oder durch Fortbildungen angeeignet, wodurch in der Praxis die Auffassung und der Umfang der Tätigkeit von Übersetzer*innen oder Sprachdienstleister*innen variieren kann, da auch die Norm *ISO 18587* für Post-Editing vor allem in freiberuflichen Szenarien weniger stark eingehalten wird als die bekanntere *ISO 17100* für Übersetzungen. Durch Ausarbeitung und Implementierung praxisfundierter Post-Editing-Ausbildungsmodule für Hochschulen soll sichergestellt werden, dass die nächsten Generationen an Übersetzer*innen auf den Umgang mit den neuen technologischen Entwicklungen im translatorischen Bereich umgehen können. Nun stellt sich die Frage, ob angehende Übersetzer*innen, die sich in der Ausbildung befinden und keine PE-Vorerfahrung haben, erfolgreich post-editieren können und der Tätigkeit auch aufgeschlossen sind.

2.4.7.1 Post-Editing durch Studierende

Studien von Jia et al. (2019), Mesa-Lao (2014) oder de Faria Pires (2020: 28ff) zeigen, dass Studierende Post-Editing in überwiegender Zahl aufgeschlossen gegenüberstehen, auch wenn sie vereinzelte Bedenken bezüglich der MÜ-Qualität und der Gefährdung von Arbeitsplätzen durch MÜ äußerten. Die größten Vorteile der Arbeitsweise sehen sie in dem möglichen Produktivitätszuwachs, dem geringeren kognitiven Aufwand im Vergleich zur Humanübersetzung sowie der Möglichkeit, sich durch die MÜ-Vorschläge für eine eigene Übersetzung inspirieren zu lassen. 53% der 64 von de Faria Pires (2020: 35) befragten Studierenden schätzten PE als positive Entwicklung für ihren zukünftigen Beruf ein, 30% standen der Entwicklung neutral gegenüber. 90% der befragten Studierenden bedauerten jedoch, dass sie sich im Rahmen ihrer Ausbildung nicht mit PE beschäftigen konnten und fühlten sich nicht auf die Anforderungen des Arbeitsmarktes vorbereitet. Wie eine Studie von Aranberri (2017) gezeigt hat, tendieren professionelle Übersetzer*innen ohne PE-Vorerfahrungen zu einem *over-editing* der maschinellen Rohübersetzung, da sie es durch ihre Berufserfahrung und vor allem in Hinblick auf den Arbeitsschritt der Revision gewohnt sind, Texte in ihrem vollem Umfang zu bearbeiten. Da beim PE von der Rohübersetzung jedoch so viel wie möglich beibehalten und nur so viel wie nötig editiert werden soll, ist hier eine

gewisse Selbstregulierung nötig. Studierende, die potenziell weniger davor zurückschrecken nur die nötigen Änderungen vorzunehmen, können auch stärker von PE profitieren, da sich ein hohes Maß an Selbstregulierung positiv auf die Produktivität beim Post-Editing auswirkt, wie eine Studie von Yang und Wang (2020: 3) zeigt. Von 43 befragten Studierenden befanden 74% in einer Studie von Yamada (2015), dass PE einfacher und schneller als Humanübersetzen ist, was auch durch den gemessenen, tatsächlichen Produktivitätszuwachs von durchschnittlich 25% bestätigt wird. Bei einer anschließenden Qualitätsevaluation der editierten Texte zeigte sich jedoch, dass 40% jener Studierenden, die PE als einfachere Tätigkeit eingestuft haben, keine zufriedenstellende Qualität erreichen konnten. Trotz tatsächlichen Produktivitätszuwächsen und einer Bevorzugung der Tätigkeit eignen sich somit nicht alle Studierenden als Post-Editor*innen, was in diesem Falle aber auch auf die fehlende Erfahrung und Ausbildung zurückzuführen ist (Yamada 2015: 57).

Guerberof Arenas et al. (2013: 214), die den Bereich des Post-Editings bereits langjährig erforscht, betont auch, dass Studierende vor allem durch den privaten Gebrauch von kommerziell verfügbaren MÜ-Systemen deren Vorteile erkennen und somit auch bereit sind, sich auf PE einzulassen, sofern die Qualität der Rohübersetzung akzeptabel ist. Mit Einführung von NMÜ-Systemen und deren Integration in PE-Arbeitsprozesse lässt sich zumindest bei Studierenden auch ein steigendes Vertrauen in die Qualität der maschinellen Rohübersetzung beobachten, da sie auch trotz fehlender Vorerfahrung größere Produktivitätsgewinne als zuvor verzeichnen können (Scansani et al. 2019: 74). Je besser die Qualität der MÜ-Systeme wird, umso mehr sinkt auch der Arbeitsaufwand beim PE, wonach diese Tätigkeit als Entlastung des eigenen Arbeitsaufwandes durch maschinelle Unterstützung gesehen und akzeptiert wird. Die 47 von Scansani et al. (2019:75) befragten Studierenden befanden PE zu 82,6% als hilfreich und zu 17,4% sogar als sehr hilfreich, keine/r der Studierenden hatte die maschinelle Rohübersetzung als störend oder unbrauchbar eingestuft. Daems et al. (2017) bestätigen in ihrer Studie, die sich mit dem Vergleich professioneller und nicht-professioneller Übersetzer*innen beschäftigt, überdies, dass die beiden Gruppen sich bereits beim Übersetzen nur in wenigen Kategorien, wie Berufserfahrung oder gewählte Übersetzungsstrategie, unterschieden, und die produzierten Übersetzungen sich in ihrer Qualität nur durch wenige Fehler unterschieden. Beim Post-Editing allgemeinsprachlicher Texte, worin keine der beiden Gruppen nennenswerte Vorerfahrung hatte, schnitten die professionellen Übersetzer*innen nicht bedeutend besser als die angehenden Übersetzer*innen ab, was darauf schließen lässt, dass die gängigen Curricula bereits eine gute Basis für PE liefern, auf der noch aufgebaut werden könnte (Daems et al. 2017: 265). In einer

vergleichbaren Studie von Aranberri et al. (2014: 27) äußerte sich die Gruppe der angehenden Übersetzer*innen gegenüber der Nützlichkeit von PE deutlich positiver als jene der professionellen Übersetzer*innen. Das Bearbeiten einer vorgegebenen Übersetzung wurde von ihnen als weniger zeitaufwändig und anstrengend eingeschätzt, vor allem bei Texten aus weniger vertrauten Fachbereichen. Bei Auswertung der editierten Texte mittels der Metriken BLEU und TER ließen sich bei beiden Gruppen beinahe identische Werte feststellen, die um weniger als 2% voneinander abwichen (Aranberri et al. 2014: 30), was sich auch mit der Studie von Daems et al. (2017) deckt.

Sofern richtig eingesetzt, kann PE auf Basis einer Übersetzungsausbildung für angehende Übersetzer*innen als Werkzeug fungieren, durch das sie die im Vergleich zu ihren professionellen Kolleg*innen fehlende vieljährige Berufserfahrung ausgleichen können. Durch Implementierung von PE-Modulen in den Curricula wäre es auch denkbar, dass diese Gruppe bereits unmittelbar nach absolvierter Ausbildung beim PE besser als manche professionelle Übersetzer*innen abschneiden könnte (Aranberri et al. 2014: 31).

2.4.7.2 Post-Editing in der Übersetzungsausbildung

Die *European Master's in Translation* (EMT) Expert*innengruppe definiert innerhalb ihres Kompetenzrahmens, der Ausbildungsstätten dabei helfen soll, angehende Übersetzer*innen auf die Anforderungen des Marktes vorzubereiten, folgende übersetzungsbezogene Kompetenz: „Apply post-editing to MT output using the appropriate post-editing levels and techniques according to the quality and productivity objectives, and recognise the importance of data ownership and data security issues." (EMT 2017: 8) Auch wenn diese Kompetenz bereits seit 2017 als wichtig gilt, wird sie bis dato nur von wenigen europäischen Hochschulen im Rahmen eigener Curricula gelehrt. Dass eine reine Übersetzungsausbildung nicht ausreichend auf PE vorbereitet, zeigt Yamada (2015: 58) in seiner Studie mit Übersetzungsstudierenden ohne PE Ausbildung. Insgesamt herrscht eine gewisse Korrelation zwischen den Noten der Studierenden und der Qualität, die ihre post-editierten Texte aufweisen, da 86% aller Studierenden mit mäßigen oder schlechten Noten im Rahmen der Studie auch qualitativ nicht zufriedenstellende Texte produzierten. Da das Kompetenzprofil für PE, wie in Kapitel 2.4.6.1 erwähnt wurde, zu einem großen Teil aus Übersetzungskompetenzen besteht, kann davon ausgegangen werden, das Studierende mit schlechten Noten im Übersetzen auch als Post-Editor*innen ungeeignet sind. Dennoch fanden sich bei den qualitativ nicht zufriedenstellenden Texten auch 52% der Studierenden mit ausgezeichneten Noten, wonach ebenso davon ausgegangen werden muss, dass nicht jede/r

angehende Übersetzer*in auch als Post-Editor*in geeignet ist, zumindest nicht ohne spezifische PE-Ausbildung (Yamada 2015: 59).

O'Brien (2002) beschäftigte sich bereits vor 20 Jahren mit der Gestaltung einer universitären PE-Ausbildung und betonte dabei vor allem die Wichtigkeit der technischen Versiertheit in Bezug auf die Funktionsweise von MÜ-Systemen, das Durchführen von PE-Aufgaben sowie auf das Beherrschen grundlegender Programmierkenntnisse (Nitzke 2019: 45). Auch Doherty und Kenny (2014: 286) sehen in ihrem Vorschlag eines PE-Curriculums eine umfassende Beschäftigung mit der Geschichte, den Arten und der Evaluierung von MÜ-Systemen vor sowie Kurse über Pre-Editing und kontrollierte Sprachen, und zuletzt das Durchführen von Post-Editing-Aufgaben. Durch ihre Befragung von über 200 Lehrenden diverser translationswissenschaftlicher Hochschulen versuchten auch Cid und Ventura (2020) zu eruieren, welche Inhalte diese für ein PE-Curriculum als sinnvoll erachten würden. Die drei wichtigsten Inhalte, für die sich rund 50% aller Befragten aussprachen, waren der Überblick über Arten und Funktionsweise diverser MÜ-Systeme, vor allem neuronale, die Beschäftigung mit den zwei Arten leichtes und vollständiges PE sowie mit praktischen und sprachpaarspezifischen PE-Aufgaben (Cid & Ventura 2020: 235). Die wichtigsten Kompetenzen, die Studierende durch dieses Curriculum erwerben sollen, sind den Befragten zufolge die Fähigkeit ein vollständiges PE mit zufriedenstellender Qualität durchzuführen, die Fähigkeit typische Fehler der vor allem neuronalen MÜ-Systeme zu erkennen und die Fähigkeit einschätzen zu können, an welchen Stellen editiert werden muss und an welchen nicht. Laut eigener Berufserfahrung der Befragten werden den Studierenden bei Berufseinstieg primär die Aufgaben des Post-Editings, der Qualitätskontrolle und Revision post-editierter Texte und die Evaluierung von maschinell übersetzten Texten anhand von Fehlerkategorien zukommen (Cid & Ventura 2020: 36). Aus der Befragung ging ebenfalls hervor, dass bei bereits existierenden PE-Curricula, die meist zwischen 2015 und 2020 eingeführt wurden, überwiegend mittels CAT-Tools wie *SDL Trados Studio* oder *MemoQ* post-editiert wird, und dass für diese Kurse auch keine eigenen MÜ-Systeme trainiert, sondern cloudbasierte Systeme wie *DeepL* oder *Google Translate* zum Einsatz kommen (Cid & Ventura 2020: 240). Die praktische Komponente dieser Curricula ist meist aufgabenbasiert und nicht projektbasiert, Studierende erhalten also nur einzelne PE-Aufträge und haben nicht die Möglichkeit, gesamte PE-Projekte im Laufe eines Semester zu simulieren. Konttinen et al. (2020: 201) betonen jedoch die Wichtigkeit einer praxisnahen Ausbildung und schlagen daher zwei aufeinander aufbauende Kurse, je ein Semester lang und im Umfang von 10 ECTS-Punkten, vor, in denen ein PE-Projekt simuliert werden soll. Studierende sollen dabei ein

fiktives Übersetzungsbüro gründen, Aufgaben verteilen, Arbeits- und Qualitätssicherungsprozesse nach ISO 17100 entwickeln und anschließend 12 fiktive PE-Aufträge bearbeiten. Neben PE-Kompetenzen sollen Studierende auch auf den Arbeitsmarkt vorbereitet werden, indem sie lernen zusammenzuarbeiten, Abgabefristen einzuhalten, Honorare zu verhandeln, auf unvorhergesehene Probleme zu reagieren und ein Gefühl für realisierbare Ziele zu entickeln (Konttinen et al. 2020: 202).

Die Universitat Autònoma de Barcelona gilt in Europa als Vorreiter, da sie bereits 2009 ein Modul zu PE und MÜ in ihr Masterstudium Translation integrierte und dieses seither stets weiterentwickelte und auch für NMÜ-Systeme adaptierte. In der ersten Hälfte des insgesamt 16-stündigen Moduls erlernen Studierende Funktionsweise und Arten von MÜ-Systemen, wie diese trainiert werden und welche häufigen Fehler auftreten können. In der zweiten Hälfte wird auf die verschiedenen Arten von PE eingegangen, welche Richtlinien zu beachten sind, welche Strategien angewandt werden können und wie diese Dienstleistung auf dem Arbeitsmarkt zu verrechnen ist. Im Laufe des gesamten Moduls müssen die Studierenden PE-Aufgaben verschiedenster Art durchführen, etwa monolingual, mit Hilfe eines Online-PE-Tools, oder mittels eines vertrauten CAT-Tools (Guerberof Arenas & Moorkens 2019: 223). Das Modul zielt darauf ab, Studierenden die mit PE verbundenen möglichen Produktivitätsgewinne zu verdeutlichen und zu zeigen, dass PE mit steigendem Wissen über MÜ-Systeme und den typischen Fehlern, die sie produzieren, kognitiv weniger aufwändig und somit produktiver werden kann. Studierenden soll ein neuer Zugang zu Übersetzungsqualität vermittelt werden, die beim PE eher anhand von terminologischer und inhaltlicher Korrektheit als anhand von gutem Stil gemessen wird, damit sie auch die Wichtigkeit erkennen, nicht zu viele Änderungen an einer Rohübersetzung vorzunehmen (Guerberof Arenas & Moorkens 2019: 223f).

Wie aus diesem Kapitel hervorgeht, existieren bereits zahlreiche Rahmenwerke, die die Implementierung theoretisch und praktisch ausgerichteter PE-Curricula beschreiben. Auch wenn seit 2015 einige Hochschulen, wie die Universitäten von Bologna oder Genf, zumindest einzelne PE-Lehrveranstaltungen anbieten, wird die Disziplin anderswo gar nicht unterrichtet. Mit dem 2020 erneuerten Bachelorstudium *Transkulturelle Kommunikation* der Universität Wien wurde zwar die Vorlesung *Maschinelle Translation* eingeführt, eine tatsächlich PE-spezifische Lehrveranstaltung findet sich jedoch auch im Masterstudium *Translation* nicht. Auch das englischsprachige Masterprogramm *Multilingual Technologies,* welches in Kooperation der Universität Wien mit der Fachhochschule FH Campus Wien ab dem Wintersemester 2022 angeboten wird und Kenntnisse in den Bereichen Sprachtechnologien,

Entwicklung maschineller Übersetzungssysteme und *Human-Computer-Interaction* vermitteln soll, umfasst laut aktuellem Curriculum keine PE-Kurse (ZTW 2021).

2.4.8 Ausblick über die weitere Entwicklung

Auch wenn PE nur auf Basis maschinell übersetzter Texte funktionieren kann, so handelt es sich dabei um eine menschenzentrierte Tätigkeit, bei der die translatorische und technische Expertise der Post-Editor*innen die wichtigsten Voraussetzungen sind (Vieira 2019: 330). Dabei ist es auch wichtig, PE als spezialisierten Prozess unabhängig vom Humanübersetzen zu betrachten, der sich in seinen Anforderungen und Abläufen teils sehr stark unterscheidet. Post-Editor*innen müssen hier einige neue Kompetenzen erlernen, etwa die Vertrautheit mit der Funktion und den häufigsten Fehlern von MÜ-Systemen, das Beherrschen der vier Änderungstypen Löschen, Einfügen, Ersetzen und Verschieben oder auch ein Maß an Selbstregulierung, um einschätzen zu können, an welchen Stellen die Rohübersetzung editiert werden muss und an welchen nicht, um den größtmöglichen Produktivitätszuwachs zu erzielen (do Carmo & Moorkens 2020: 49). Auch wenn PE, wie in Kapitel 2.4.6 erwähnt, in erster Linie von professionellen Übersetzer*innen ausgeführt wird, wurde im Rahmen dieser Arbeit der Begriff Post-Editor*in bevorzugt, um hervorzuheben, dass es sich um eine eigenständige Tätigkeit handelt. Dennoch ist es denkbar, dass mit dem technologischen Fortschritt die Grenzen zwischen MÜ und Humanübersetzung sowie zwischen Übersetzer*innen und Post-Editor*innen immer mehr verschwimmen und die beiden Bezeichnungen zu einer verschmelzen, sobald sie sich in ihrer Funktion und Tätigkeit genug ähneln. Bereits jetzt wird PE in der Sprachindustrie als Teil des Dienstleistungsportfolios von Übersetzer*innen gesehen, was für eine Entwicklung in diese Richtung spricht (Vieira 2019: 330).

Post-Editing kann unter der Voraussetzung der richtigen Kombination aus MÜ-System, Sprachpaar und Textsorte die Produktivität bei gleichbleibender oder sogar höherer Qualität des Zieltextes steigern und helfen, das stets steigende Volumen an zu übersetzenden Texten zu bewältigen (Koponen 2016: 142). Auch wenn PE-spezifische Normen und Richtlinien wie von ISO (2017) oder TAUS (Massardo et al. 2016) nur vereinzelt implementiert werden, wird PE immer mehr zum Arbeitsprozess bei Sprachdienstleister*innen. Diese Entwicklung sorgte auch für ein größeres Interesse seitens der Translationswissenschaft und führte zur Entwicklung einzelner PE-Module oder -Curricula in translationswissenschaftlichen Ausbildungen, die kommende Generationen an Übersetzer*innen auf die neuen Anforderungen der Sprachindustrie vorbereiten sollen. Die Interaktion zwischen Mensch und Computer ist im 21. Jahrhundert eine zentrale Komponente

der Sprachindustrie geworden und ist momentan als Kooperation zu gleichen Teilen zu verstehen, da der Mensch ohne Hilfe des Computers seine Produktivitätsquote nicht erreichen könnte, und da der Computer ohne menschliche Nachbearbeitung nicht in der Lage ist, publikationsreife Übersetzungen anzufertigen, was sich auch in den kommenden Jahren nicht ändern dürfte (Koponen 2016: 143). Auch wenn mit der Einführung von NMÜ-Systemen große Fortschritte in Bezug auf Zieltextqualität gemacht werden konnten, so ist die Verwendung von neuronalen Netzen dennoch eine Disziplin in ihren Anfängen. Mit der stetigen Weiterentwicklung und Erforschung von Faktoren wie Fehlertypologien, der Anwendbarkeit für ressourcenarme Sprachen und der Akzeptanz von NMÜ und PE seitens der Übersetzer*innen wird sich auch das PE daran orientiert weiterentwickeln und noch besser erforscht werden (Guerberof Arenas 2020: 349).

3 Aktueller Forschungsstand

Seit 2014 lässt sich ein intensives Forschungsinteresse im Bereich des Post-Editings in Bezug auf Produktivität, Aufwand und das Potential von Studierenden als Post-Editor*innen beobachten. Die Vielzahl der publizierten Studien kann als Indiz für die Wichtigkeit des PE innerhalb der Translationswissenschaft gesehen werden sowie für die Erkenntnis, dass angehende Übersetzer*innen darin ausgebildet werden sollten. In diesem Kapitel wird eine Auswahl der relevantesten Forschungsarbeiten präsentiert, die auch als Inspiration für das Forschungsdesign dieser Arbeit dienen sollen.

Durch das Fehlen einer Post-Editing-Ausbildung an vielen Hochschulen stellt sich einerseits die Frage, ob Übersetzungswissen und -erfahrung auch auf PE übertragbar sind. In einer Studie von Aranberri et al. (2014) wurden die Unterschiede im Produktivitätszuwachs zwischen professionellen Übersetzer*innen und Personen mit Fremdsprachenkenntnissen untersucht. Je sechs Übersetzer*innen und sechs Lai*innen post-editierten zwei aus dem Englischen in ihre Erstsprache Baskisch maschinell übersetzte Texte und füllten einen Fragebogen bezüglich Produktivität und Schwierigkeit der Texte aus. Die finalen Texte wurden mittels der Qualitätsmetriken *Bilingual Evaluation Understudy (BLEU)* und *Translation Edit Rate (TER)* bewertet. Im Durchschnitt konnten die Übersetzer*innen ihre Produktivität bei den Texten um 28% steigern, die Lai*innen sogar um 45%. Die Auswertung der Fragebögen ergab, dass die Lai*innen gegenüber PE offener eingestellt waren und den möglichen Produktivitätsgewinn erkannten. Trotz des gemessenen Produktivitätszuwachses äußerten sich die Übersetzer*innen negativer und beschrieben das Post-Editing als störend und zeitaufwändig. Auch Gaspari et al. (2014) untersuchten die Unterschiede zwischen subjektiver Wahrnehmung und objektiver Messung von Produktivität beim PE in einer Studie mit 20 Übersetzer*innen. Die Teilnehmer*innen bearbeiteten bidirektionale maschinelle Übersetzungen in den Sprachpaaren Deutsch-Englisch und Niederländisch-Englisch und füllten anschließend einen Fragebogen zu ihrer Einschätzung aus. In beiden Sprachpaaren wurden Aufwand, Geschwindigkeit und Produktivität des PE von den Teilnehmer*innen als geringer und schlechter eingestuft als eine Humanübersetzung. Generell zeigte sich eine klare Bevorzugung des klassischen Humanübersetzens ohne Beihilfe von MÜ und eine Ablehnung des PE. Die berechnete Bearbeitungszeit pro Ausgangssprachenwort zeigte jedoch eine erzielte Produktivitätssteigerung von 3,5% bis 13,6% Prozent, und auch die eruierten *BLEU* und *TER*-Werte der post-editierten Texte waren höher als vergleichbare Humanübersetzungen. Nachdem also davon ausgegangen werden kann, dass PE in den meisten Fällen einen Produktivitätszuwachs mit sich bringt, stellt sich nun die Frage nach der

optimalen Qualität der rohen MÜ. In einer Studie mit 24 Übersetzer*innen untersuchte Guerberof Arenas (2014a) die Produktivitäts- und Qualitätsunterschiede zwischen dem Post-Editing von maschinell übersetzen Segmenten im Vergleich zum Post-Editing von 85-94%en Fuzzy Matches eines Translation Memorys. Die Teilnehmer*innen erhielten einen Text, der aus dem Englischen in ihre Erstsprache Spanisch übersetzt wurde. Ein Drittel der Segmente wurde maschinell übersetzt, ein Drittel stammte aus einem Translation Memory und ein Drittel wurde freigelassen und musste von den Teilnehmer*innen selbst übersetzt werden. Mit 17 Wörtern pro Minute, einer Zeitersparnis von 37% im Vergleich zur Humanübersetzung, schnitt das PE der MÜ-Segmente am besten ab und auch beim PE der Fuzzy Match-Segmente wurden 15,9 Wörter pro Minute und 32% Zeitersparnis erzielt. Der finale Text wies in den MÜ- und Fuzzy Match-Segmenten durchschnittlich 550 Fehler auf, in den humanübersetzten Segmenten jedoch 948.

Wie in Kapitel 2.4.5.2 bereits erwähnt wurde, sollte der kognitive Aufwand beim Post-Editing optimalerweise am Individuum und nicht am Text gemessen werden. Lacruz und Shreve (2014) untersuchten daher das Pausenverhalten der Teilnehmer*innen, da die Annahme besteht, dass Pausen Indizien für gesteigerten kognitiven Aufwand sind. Die Eingaben von drei Teilnehmer*innen wurden während des Post-Editings von vier Texten mittels eines Keyloggers aufgezeichnet. Die erzeugten Zeitmarken ermöglichten es, folgende Werte zu berechnen: *average pause ratio (APR)* = Durchschnittszeit pro Pause / Durchschnittszeit pro Wort; *pause ratio (PR)*; pause to word ratio *(PWR)* = Anzahl von Pausen im Segment/ Anzahl der Wörter im Segment. Diese Werte ließen erkennen, dass Segmente, in denen der kognitive Aufwand höher ist, immer Pausen-Cluster aufweisen, und, dass vor dem Durchführen einer Änderung vermehrt Pausen auftreten. Vieira (2014) wiederum untersuchte den kognitiven Aufwand beim PE hinsichtlich der Arbeitsgedächtniskapazität und der Kompetenz der Ausgangssprache. Dafür wurde ein Eyetracker eingesetzt, der die Fixationen, also die Betrachtungen, der 13 Teilnehmer*innen während der Post-Editing-Aufgabe maß. Die Studie zeigte, dass Teilnehmer*innen mit schlechterer Ausgangssprachenkompetenz mindestens 30% aller Fixationen auf den Ausgangstext richteten und diese Segmente im abschließenden Fragebogen als besonders schwierig bewerteten. Jene Teilnehmer*innen mit besserer Sprachkompetenz und folglich höherer Arbeitsgedächtniskapazität waren stärker auf den Zieltext fokussiert und schnitten in den Punkten Produktivität und Qualität besser ab. Herbig et al. (2019) führten die bis dato umfangreichste Studie zum kognitiven Aufwand beim PE durch und zeigten auf, dass eine Vielzahl an Messergebnissen in Verbindung miteinander gute Indikatoren für kognitiven

Aufwand sein können. Die Leistungen von zehn Teilnehmer*innen wurden während einer PE Aufgabe mit folgenden Hilfsmitteln aufgezeichnet: mit einem Keylogger, um Wörter pro Minute und Pausen zu berechnen, mit einem Eyetracker, der Fixationen misst, mit einem Band zur Herzfrequenzmessung, mit dem Microsoft Band v2, ein Armband zur Messung von Schweiß und Hauttemperatur, mit zwei Webcams, die Mimik und Offenheitsgrad der Augenlider messen sowie dem Kinect v2, einem Motion-Tracker, der die Nähe der Teilnehmer*innen zum Bildschirm misst. Zusätzlich mussten die Teilnehmer*innen nach jedem editierten Segment eine Einschätzung zum Aufwand auf einer Likert-Skala von 1-9 abgeben. Das Experiment zeigte, dass subjektive Einschätzungen von Aufwand nicht mit tatsächlich gemessenem Aufwand übereinstimmen, und dass eine Vielzahl an Messergebnissen deutliche Korrelationen erkennen lässt.

Nachdem durch einige Forschungsarbeiten die negative Einstellung von Übersetzer*innen gegenüber des Post-Editings aufgezeigt wurde, hat sich der Schwerpunkt der Forschung in den letzten Jahren auf Studierende als Post-Editor*innen verschoben. Auf Grund der oftmals fehlenden Ausbildungsmöglichkeiten stellt sich die Frage, in welchem Ausmaß Studierende für das Erfüllen von PE Aufträgen geeignet sind und mit welcher Einstellung sie dies tun. In einer Studie mit 28 Studierenden ließ Yamada (2019) diese einen aus dem Englischen ins Japanische maschinell übersetzten Text post-editieren und befragte sie anschließend. 71% empfanden PE einfacher als Humanübersetzen, und 79% gaben an, eine Reduktion an Aufwand festgestellt zu haben. Eine anschließende Qualitätsauswertung zeigte, dass die Teilnehmer*innen weniger Änderungen vornahmen, diese aber mit höherem kognitiven Aufwand verbunden waren. Rund 40% der Teilnehmer*innen produzierten akzeptable Zieltexte und zeigten Potenzial, gute Post-Editor*innen zu werden. In einer ähnlichen Studie befragt de Faria Pires (2020) 64 Studierende im Rahmen einer PE Aufgabe zu ihrer Wahrnehmung von und Einstellung zum Post-Editing. Beim Editieren des vom Englischen ins Französische maschinell übersetzten Textes wurden die Augenbewegungen der Teilnehmer*innen zusätzlich durch einen Eyetracker aufgezeichnet. Die Auswertung der Fragebögen, die die Teilnehmer*innen vor und nach Erledigen der Aufgabe ausgefüllt haben, zeigten ähnliche Ergebnisse wie bei Yamada (2019). Die Studierenden erkannten die Wichtigkeit von und den Produktivitätsgewinn durch PE, jedoch zeigte die Auswertung der Texte, dass die Teilnehmer*innen keine einheitlichen Strategien verwendeten. Dies verdeutlicht die Wichtigkeit der Einführung von PE Kursen an Hochschulen, um der zukünftigen Generation an Post-Editor*innen das notwendige Wissen zu vermitteln.

4 Forschungsdesign der wissenschaftlichen Studie

Im Rahmen dieser Studie erhielten die Teilnehmer*innen eine Post-Editing-Aufgabe, bei der sie einen Text, der mit *DeepL* vom Englischen ins Deutsche maschinell übersetzt wurde, unter gewissen Vorgaben post-editieren mussten. Als Methode für diese wissenschaftliche Studie wurde also eine Mischung aus einer Korrelationsstudie und einem Quasi-Experiment gewählt. Zum einen ist nur eine kleine Anzahl an Teilnehmer*innen vorhanden, die nicht randomisiert zugeteilt, sondern anhand identischer Merkmale, primär der Zugehörigkeit zu der Gruppe „Studierende", ausgewählt wurde. Weiters fehlen in dieser Studie eine abhängige und unabhängige Variable sowie eine Kontrollgruppe. Es soll also nicht das Ziel sein, Kausalaussaugen zu treffen, sondern multivariate Variablenzusammenhänge aufzudecken und zu strukturieren. Ziel der Studie ist es, die objektiven Messergebnisse durch Eyetracking gemeinsam mit den subjektiven Einschätzungen der Teilnehmer*innen aus den Interviews zu analysieren, um so Aussagen über den mit der Aufgabe verbundenen Aufwand treffen zu können. Die Auswertung erfolgt wegen der oben erwähnten Limitationen der Studie primär mittels deskriptiver Statistik und qualitativer Inhaltsanalyse und soll auch dazu dienen, die in der Einleitung gestellte Forschungsfrage zu beantworten und einige Annahmen zu bestätigen oder zu widerlegen. Das Forschungsdesign ist durch mehrere Arbeiten zu ähnlichen Fragestellungen inspiriert, primär orientiert es sich jedoch an den beiden Arbeiten von Mesa-Lao (2014) und de Faria Pires (2020), die sich ebenfalls mit Studierenden und Post-Editing beschäftigten und Eyetracking als Erhebungstechnik verwendeten.

4.1 Profil der Teilnehmer*innen

Da sich diese Arbeit mit der Eignung von Studierenden als Post-Editor*innen beschäftigt, wurden auch die drei Teilnehmer*innen für diese Studie nach bestimmten Kriterien ausgewählt. Die Teilnehmer*innen waren alle zwischen 25 und 26 Jahre alt und hatten das Bachelorstudium Transkulturelle Kommunikation am Zentrum für Translationswissenschaft abgeschlossen. Zwei Teilnehmer*innen befanden sich zum Zeitpunkt der Studie im letzten Semester des Masterstudiums Translation mit dem Schwerpunkt Fachübersetzen und Sprachindustrie, ein/e Teilnehmer*in hatte das Studium mit demselben Schwerpunkt innerhalb der letzten sechs Monate abgeschlossen. Die Teilnehmer*innen gaben alle Deutsch als ihre Erstsprache auf dem Niveau C2 an, und Englisch als ihre Zweitsprache mit dem Niveau C1. Zwei Teilnehmer*innen gaben an, seit mehr als 12 Monaten in einem Unternehmen für intralinguale Audiodeskription und Untertitelung zu arbeiten und dort teilweise auch Erfahrungen im Bereich Übersetzen und Revision gesammelt zu haben. Ein/e

Teilnehmer*in hatte zuvor über ein Jahr in der Marketingabteilung einer Bank gearbeitet und gab jedoch keine Berufserfahrung im translatorischen Bereich an. Im Zuge des ersten Interviews bestätigten alle Teilnehmer*innen über keinerlei Post-Editing-Ausbildung zu verfügen, ein/e Teilnehmer*in gab jedoch an, im Laufe des Studiums öfters laienhaft post-editiert zu haben. Die ausgewählten Teilnehmer*innen erfüllten für diese Studie also insofern die wichtigsten Kriterien, weil sie durch ihre Ausbildung die notwendige theoretische und praktische Übersetzungskompetenz erlernt haben, durch die Vertiefung im Schwerpunkt Fachübersetzen auch mit CAT-Tools und computergestützter Translation vertraut sind und sie alle über keinerlei Post-Editing-Ausbildung verfügen. Im Rahmen dieser Studie führten also alle Teilnehmer*innen zum allerersten Mal ein Post-Editing unter realistischen, praxisnahen Auftragsbedingungen durch und waren bei der Wahl der Vorgangsweise und Anwendung von Strategien gänzlich auf sich gestellt.

4.2 Auswahl des Textes

Bei der Auswahl des englischsprachigen Ausgangstextes wurden die folgenden Aspekte berücksichtigt. Erstens sollte der ausgewählte Text einige Merkmale wie eine längere Syntax, Mehrdeutigkeiten von Wörtern sowie idiomatische Wendungen beinhalten, die selbst bei NMÜ-Systemen wie *DeepL* zu Fehlern oder Ungereimtheiten in der Übersetzung führen können. Dies stellt auch sicher, dass die Teilnehmer*innen ausreichend Änderungen vornehmen können und die Studie in dieser Hinsicht aussagekräftig ist. Gleichzeitig sollte der Text inhaltlich nicht zu fordernd sein, indem beispielsweise ein übermäßiger Rechercheaufwand an Fachvokabular nötig ist, da dies nicht das Ziel dieser Studie ist. Aranberri et al. (2014: 23f) und Lacruz et al. (2014: 79) weisen in ihren Forschungsarbeiten darauf hin, dass Texte zum allgemeinen Gebrauch mit wenig Fachvokabular am besten geeignet sind und die Textauswahl sich an dem Wissensstand der Teilnehmer*innen orientieren sollte. Für diese Studie wurden daher vorerst fünf verschiedene Zeitungsartikel aus der Rubrik *Kommentar* diverser englischsprachiger Tageszeitung ausgewählt, die sich mit den gesellschaftspolitisch aktuellen Themen *Recht auf Reparatur, Fast Fashion, Kryptowährungen* und *Arbeitsbedingungen bei Amazon* befassen. Nachdem die Texte probeweise mit *DeepL* übersetzt wurden, wurden die Zieltexte auf Zeichensetzung, Zahlen und Redewendungen oder Wortspiele überprüft, um die Übersetzung mit dem größten Bearbeitungspotenzial für die Studie zu finden. So wurde schlussendlich der Kommentar *Embracing cryptocurrency* vom 16.06.2021 aus der indischen, englischsprachigen Tageszeitung *The Hindu* ausgewählt.

Zweitens sollte die Textlänge mit dem aktuellen Kompetenzstand der Teilnehmer*innen bewältigbar sein. Laut Robert (2013: 32f) ist von professionellen Post-Editor*innen eine Produktivität von 437,5 Wörtern pro Stunde zu erwarten, von Studierenden hingegen 300 bis 370 Wörter. Dieser Richtwert wurde bereits in einer ähnlichen Studie von de Faria Pires (2020) gewählt, deren Ergebnisse aufzeigen, dass Texte mit einer maximalen Länge von 370 Wörtern von Studierenden innerhalb einer Stunde problemlos post-editiert werden können. Der für die vorliegende Studie ausgewählte Text wurde auf eine Länge von 361 Wörter (2.402 Zeichen inkl. Leerzeichen) gekürzt und hat eine durchschnittliche Wortanzahl von 22,6 pro Satz. Durch die längeren Sätze sollte für die Teilnehmer*innen auch ein höherer Schwierigkeitsgrad bei der Aufgabe erzeugt werden, da das Vergleichen mehrzeiliger Sätze zwischen Ausgangs- und Zieltext kognitiv aufwändiger ist als bei kurzen Sätzen.

Drittens sollte der Text der Lesekompetenz der Teilnehmer*innen entsprechen und nicht übermäßig schwierig zu verstehen sein. Die relative Schwierigkeit des Textes wurde anhand der Lesbarkeitsindexe *Flesch-Reading-Ease* und *Flesch-Kincaid-Grade-Level* eruiert, die auch in einer Studie von Yamada (2015) zur Anwendung kamen. Beide Indexe errechnen anhand der durchschnittlichen Satzlänge und Silbenanzahl pro Wort einen Wert, der auf die Schwierigkeit des Textes schließen lässt, wobei der *Reading-Ease* Wert zwischen 0 und 100 und der *Grade-Level* Wert zwischen 5 und 18 liegen kann. Der ausgewählte Text hat einen *Reading-Ease* Wert von 23,5 und einen *Grade-Level* Wert von 16,2 erzielt, wonach der Text für *College Graduates*, also für Hochschulabsolvent*innen, geeignet ist und insgesamt als *very difficult to read* gilt.

4.3 Ablauf der Studie und Erhebungstechniken

Der ausgewählte Text wurde, wie bereits erwähnt, mit *DeepL* maschinell übersetzt und sollte durch die Teilnehmer*innen anschließend in M*emoQ Translator Pro 9* post-editiert werden. Dafür wurde als CAT-Tool bewusst *MemoQ* ausgewählt, weil dieses am Zentrum für Translationswissenschaft im Masterschwerpunkt Fachübersetzen und Sprachindustrie primär gelehrt wird und folglich davon ausgegangen werden kann, dass alle Teilnehmer*innen ausreichend damit umgehen können. *DeepL* wurde als MÜ-System gewählt, da es laut Heinisch und Lušicky (2019: 44) von 69% von Studierenden und Absolvent*innen bevorzugt wird, und da es aus persönlicher Erfahrung von Lehrenden des Zentrums für Translationswissenschaft in mehreren Lehrveranstaltungen sowohl im Bachelor- als auch im Masterstudium zur Anwendung kommt. Außerdem gaben alle drei Teilnehmer*innen im Interview an, *DeepL* im privaten und beruflichen Alltag regelmäßig zu verwenden.

Der ausgewählte Text wurde als zweisprachige XLIFF-Datei aufbereitet, denn dadurch sind Ausgangstext und der maschinell übersetzte Zieltext auf Satz- bzw. Segmentebene aligniert und die Bearbeitung gestaltet sich für alle Teilnehmer*innen identisch. Vor Abspeichern der Datei wurde der Segmentstatus für den gesamten Text auf *Nicht begonnen* gestellt, anschließend wurde die XLIFF-Datei exportiert. Abschließend wurden alle 16 Segmente des Textes in ein Translation Memory gespeichert und als TMX-Datei exportiert. Am Tag der Studie, die im *HAITrans Eyetracking-Labor* am Zentrum für Translationswissenschaft durchgeführt wurde, wurde ein eigenes *MemoQ* Projekt für alle Teilnehmer*innen angelegt. Dabei wurde zuerst die zweisprachige XLIFF-Datei importiert, welche die Teilnehmer*innen später bearbeiten sollten, sowie die TMX-Datei, die als Master-Translation Memory fungiert und ebenfalls die maschinelle Übersetzung des Textes beinhaltet. Die Übereinstimmungen für jedes der Segmente aus dem Master-Translation Memory wurden in *MemoQ* am rechten Bildschirmrand angezeigt und ermöglichten es den Teilnehmer*innen, den ursprünglichen Übersetzungsvorschlag für ein Segment einzusehen, auch wenn sie diesen im Zieltextfeld bereits überschrieben oder gelöscht hatten. Außerdem wurde in jedem Projekt noch ein leeres Arbeits-Translation Memory eingefügt, in welchem die finale Version des post-editierten Textes jedes/r Teilnehmer*in abgespeichert wurde. Zuletzt wurden in *MemoQ* die Funktionen *Bearbeitungszeit aufzeichnen* und *Änderungen nachverfolgen* aktiviert, da diese sämtliche Änderungen am Text und die Bearbeitungszeit pro Segment während der Studie aufzeichneten. In einem letzten Schritt wurde der Teilnehmer-Bildschirm so eingerichtet, dass die rechte Bildschirmhälfte vom *MemoQ*-Fenster ausgefüllt wurde und die linke Hälfte von einem Browser-Fenster, das den Teilnehmer*innen während der Aufgabe für Recherchezwecke und die Verwendung von Online-Wörterbüchern zur Verfügung stand.

Anschließend wurde jede/r Teilnehmer*in zu seinem/ihrem Termin durch den Forschungsleiter begrüßt, über den groben Ablauf der Studie informiert und gebeten eine Einverständniserklärung zu unterzeichnen sowie Angaben zur eigenen Person auszufüllen. Dann wurde mit jedem/r Teilnehmer*in ein erstes Leitfadeninterview durchgeführt, in dem neben allgemeinen Fragen zu Ausbildung und beruflicher sowie übersetzerischer Vorerfahrung vor allem erforscht werden sollte, wie die Teilnehmer*innen zu Post-Editing stehen, welche Erfahrungen sie damit bereits haben und wie sie dessen Entwicklung einschätzen. Dabei erhielten alle Teilnehmer*innen dieselben Fragen aus einem festgelegten Katalog, wobei darauf geachtet wurde, die Teilnehmer*innen möglichst lange frei erzählen zu lassen und bei Gelegenheit auch Folgefragen zu stellen. Nach Durchführen dieses ersten

Interviews erhielt jeder/e Teilnehmer*in vom Forschungsleiter eine genaue Einweisung bezüglich des Ablaufes und der Regeln der Post-Editing Studie. Die Teilnehmer*innen wurden gebeten, die Studie wie einen echten Post-Editing-Auftrag zu behandeln, bei dem das Ziel die Produktion eines publikationsreifen, hochqualitativen und möglichst fehlerfreien Textes ist. Die Teilnehmer*innen wurden darauf hingewiesen, dass es auf Grund der akzeptablen Qualität des maschinell übersetzten Textes nicht nötig ist, Zieltextsegmente gänzlich zu löschen und diese selbst zu übersetzen. Die Teilnehmer*innen sollten sich hingegen darauf konzentrieren, in erster Linie Änderungen an den maschinell übersetzten Zieltextsegmenten vorzunehmen, wie es dem Charakter des Post-Editings entspricht. Bezüglich der technischen Aspekte wurden die Teilnehmer*innen gebeten, jedes abgeschlossene Segment in *MemoQ* zu bestätigen und auch auf die Wichtigkeit hingewiesen, das *MemoQ*-Fenster auf dem PC nicht zu verschieben oder zu schließen. Den Teilnehmer*innen wurde erklärt, dass sie das Browserfenster für sämtliche Suchmaschinen und Online-Wörterbuch zur Recherche verwenden dürfen, jedoch keinerlei maschinelle Übersetzungssysteme wie *DeepL* oder *Google Translate.* Den Teilnehmer*innen wurde kein Zeitlimit gesetzt. Nach dieser Einweisung nahmen die Teilnehmer*innen vor dem Teilnehmer-PC Platz, auf dem die Post-Editing-Aufgabe in *MemoQ* sowie das Browserfenster bereits geöffnet waren. Der *EyeLink® Portable Duo* Eyetracker, der sich vor den Teilnehmer*innen zwischen Tastatur und Bildschirm befand, wurde durch den Forschungsleiter am Host-PC in einem separaten Raum eingerichtet. Dafür wurde in der zum Eyetracker dazugehörigen Softwareandwendung *WebLink* ein neues Projekt für alle Teilnehmer*innen angelegt, welches aus den zwei Schritten *Camera Setup* (Kalibrierung) und *Screen Recording* (Bildschirmaufnahme) bestand. Für den ersten Schritt wurden die Abstände vom Kopf der Teilnehmer*innen bis zum Eyetracker und zum Bildschirm gemessen und eingegeben sowie ein zweistufiger Kalibrierungstest durchgeführt, der für alle Teilnehmer*innen abhängig von Sitzposition und Körpergröße notwendig war. Anschließend erhielten die Teilnehmer*innen mit Starten des zweiten Schrittes die Freigabe, mit dem Bearbeiten der Aufgabe zu beginnen, wobei auch eine Stoppuhr gestellt wurde, um die Arbeitsdauer aller Teilnehmer*innen zu erfassen. Nach Absolvieren der Aufgabe füllten die Teilnehmer*innen zwei Likert Skalen zur Schwierigkeit des Ausgangstextes und zur Qualität der maschinellen Übersetzung aus und wurden in einem zweiten Interview gefragt, wie ihnen das Post-Editing gefallen hat, welche Probleme bei der Aufgabe aufgetreten sind, welche Strategien sie angewandt haben, wie sie ihre eigene Produktivität einschätzen und ob sich ihre Meinung zu Post-Editing durch diese Studie geändert hat.

4.4 Auswertungsverfahren

Die Auswertung der gesammelten empirischen Daten soll auf Grund der kleinen Teilnehmer*innenanzahl primär mittels deskriptiver Statistik durchgeführt werden. Zum einen trifft dies auf jene Daten zu, die im Rahmen dieser Studie als „objektive Messergebnisse" bezeichnet werden. Diese wurden in *Microsoft Excel* aufbereitet und anschließend in tabellarischer Form oder mittels Diagrammen visualisiert. Dazu zählen etwa die durch den Eyetracker aufgezeichneten proprietären *Eyetracking Data File* (EDF)-Dateien, die in das zugehörige Programm *WebLink* importiert wurden, welches ein Exportieren ausgewählter Variablen in die Excel-Tabellen *„Trial Report"* und *„Interest Area Report"* ermöglichte. Der *Trial Report* bezieht sich auf die gesamte Aufnahme eines/r Teilnehmer*in und umfasst Variablen, wie beispielsweise die Anzahl und Dauer der Fixationen und Lidschläge und die Größe der Pupillen. Für das Erstellen des *Interest Area Reports* wurden mittels *WebLink* innerhalb der Bildschirmaufnahmen vier rechteckige Interessensbereiche definiert, da alle Teilnehmer*innen in einer statischen Umgebung, bestehend aus *MemoQ*- und Browserfenster, arbeiteten. Die vier Bereiche waren das gesamte Browserfenster (1) und im *MemoQ*-Fenster die Spalte des Ausgangstextes (2), die Spalte des Zieltextes (3) sowie der Bereich, in dem Translation Memory Treffer angezeigt wurden (4). Nach dem Export wurden *im Interest Area Report* die ausgewählten Variablen, wie etwa die durchschnittliche Pupillengröße, die Anzahl und Dauer der Fixationen und die Verweildauer für jeden der vier Interessensbereiche gesondert angegeben und ermöglichen so einen Vergleich der Vorgangsweisen der Teilnehmer*innen. Außerdem wurden zwei Screenshots der gesamten Aufzeichnung jedes/r Teilnehmer*in in *WebLink* gemacht, die zum einen die Position aller Fixationen und ihrer Dauer in Form blauer Kreise auf einem repräsentativen Hintergrundbild zeigen sowie die übersichtlichere Verteilung aller Fixationen in Form einer *Fixation Map*, die die durch die Teilnehmer*innen betrachteten Bereiche nach ihrer Häufigkeit in rot, gelb oder grün farblich hervorhebt. Für die Auswertung der Arbeitszeit wurden zweierlei Datensätze herangezogen. Zum einen wurde vor Beginn der Aufgabe in *MemoQ* die Funktion *Bearbeitungszeit aufzeichnen* aktiviert, welche für jedes Segment die Zeit aufzeichnet, die auf Tastatur- oder Mauseingaben entfällt. Es wurde somit nur jene Zeit gemessen, die auf ein aktives Bearbeiten des Textes entfällt, jedoch nicht auf das Lesen. Diese Zeit wurde pro Teilnehmer*in und pro Segment manuell aus der *MemoQ-XLIFF* Datei in eine Excel-Tabelle übertragen und gab auch Aufschluss darüber, an welchen Segmenten die Teilnehmer*innen am längsten arbeiteten. Zusätzlich wurde aus dem *Interest Area Report* für jeden der vier Interessensbereiche die Summe der Dauer aller Fixationen, auch Verweildauer genannt,

entnommen. Diese gab Aufschluss darüber, wie lange die Teilnehmer*innen jeden der vier Bereiche im Laufe der gesamten Aufnahme betrachteten. Zuletzt wurde mittels der gesamten Arbeitszeit pro Teilnehmer*in und der Wortanzahl des bearbeiteten Textes die Wörter pro Minute und Wörter pro Stunde berechnet, die jede/r Teilnehmer*in erzielen konnte.

Der post-editierte Zieltext wurde für jede/n Teilnehmer*in mitsamt der nachverfolgten Änderungen aus *MemoQ* in verschiedene Dateiformate exportiert. Anschließend wurde der Text als RTF-Tabelle herangezogen und dort sämtliche Änderungen jedes/r Teilnehmer*in in jedem Segment manuell gezählt und für die Summe an Änderungen im gesamten Text addiert. Als Änderung sind hier die im Kapitel 2.4.3 beschriebenen Vorgänge *Löschen*, *Hinzufügen*, *Verschieben* und *Ersetzen* zu verstehen, die in diesem Fall auf Wort- beziehungsweise Phrasenebene gezählt wurden. Die Teilnehmer*innen wurden im Interview nach Ausführen der Aufgabe auch darum gebeten einzuschätzen, wie viele Änderungen sie im Durchschnitt pro Segment vorgenommen hatten. Neben diesem Wert ist zuletzt auch noch die Anzahl an Segmenten, in denen keine Änderung vorgenommen wurde, aufgeführt. Darüber hinaus wurden die post-editierten Texte aller Teilnehmer*innen mit den maschinell übersetzten Rohversionen desselben Textes unter Verwendung der Qualitätsmetriken *METEOR*, *TER* und *BLEU* sowie der *Levenshtein Editierdistanz* verglichen. Die errechneten Werte sagen in diesem Fall jedoch nichts über die Qualität des post-editierten Textes aus, sondern nur über die Qualität des verwendeten MÜ-Systems. Da jedoch alle vier Metriken darauf ausgelegt sind die Äquivalenz oder Verschiedenheit der Zeichenfolgen von zwei Texten zu messen, lassen diese Werte darauf schließen, wie stark die maschinelle Rohübersetzung durch die Teilnehmer*innen verändert wurde. Die vier Metriken wurden in der auf der Programmiersprache *Python* basierenden Programmierumgebung *Google Colab* mittels eines Notebooks berechnet. Zu den subjektiven Einschätzungen der Teilnehmer*innen zählen etwa die Likert Skalen bezüglich der Qualität der maschinellen Übersetzung und der Schwierigkeit des Ausgangstextes, die die Teilnehmer*innen nach Durchführen der Aufgabe auf einem Zettel ausfüllten und die anschließend durch gestapelte Säulendiagramme visualisiert wurden. Die zwei Leitfadeninterviews, die mit jedem/r der Teilnehmer*innen geführt wurden, wurden nach gängigen Konventionen wörtlich transkribiert und mit Zeilennummern versehen. Durch den einheitlichen Ablauf des Interviews mit allen Teilnehmer*innen konnten auch die Antworten einfach nach Frage bzw. Thema gruppiert werden. Mittels einer qualitativen Inhaltsanalyse nach Gläser und Laudel (2010) war es anschließend möglich, etwaige Gemeinsamkeiten und Unterschiede im Antwortverhalten der Teilnehmer*innen festzustellen und zu beschreiben.

5 Ergebnisse

Die Präsentation der Ergebnisse in den folgenden Unterkapiteln orientiert sich an Krings (2001) Typologie von temporalem, technischem und kognitivem Aufwand. Die erhobenen Daten sollen abschließend durch eine Inhaltsanalyse der mit den Teilnehmer*innen geführten Interviews ergänzt werden und auch einen Einblick in deren Einstellungen zu Post-Editing aufzeigen. Der Einfachheit halber werden die Teilnehmer*innen 1, 2 und 3 in diesem Kapitel als T1, T2 und T3 bezeichnet. Der Großteil der Ergebnisse wird in Bezug auf die 16 Segmente mit durchschnittlich 22,6 Wörtern, aus denen der zu bearbeitende, 361 Wörter lange Text bestand, sowie in Bezug auf die vier Interessensbereiche, die bei Auswertung der Eyetracking-Daten festgelegt wurden, präsentiert.

5.1 Temporaler Aufwand

Die Arbeitszeit der Teilnehmer*innen ist besonders in Hinblick auf deren Aufteilung auf die verschiedenen Segmente und Interessensbereiche aussagekräftig. In Tabelle 1 ist die durchschnittliche Bearbeitungszeit pro Segment zu sehen, daneben die gesamte Bearbeitungszeit aller 16 Segmente in *MemoQ*. Ebenso ist die Arbeitszeit der Teilnehmer*innen zu sehen, die für die gesamte Post-Editing-Aufgabe benötigt wurde, wozu auch die Recherche im Browser zählt, sowie die Einschätzungen der Teilnehmer*innen aus den Interviews bezüglich der Dauer, die sie ihrem Gefühl nach für die Aufgabe benötigt hatten. T1 erzielte die längste Bearbeitungsdauer pro Segment und folglich auch die längste Arbeitsdauer für den gesamten Text. In Hinblick auf die Gesamtdauer hat T1 nur 8 Minuten und 47 Sekunden nicht mit dem Editieren des Textes verbracht. T2 erzielte hingegen die kürzeste Bearbeitungsdauer pro Segment und für den Text insgesamt, von rund 50 Minuten gesamter Arbeitszeit entfielen nur 24 Minuten auf das Editieren des Textes. T3 weist eine ausgewogenere Aufteilung zwischen Bearbeitungszeit und Gesamtzeit auf. Auch wenn die gemessene gesamte Arbeitszeit zeigt, dass alle Teilnehmer*innen etwa dieselbe Dauer für das Post-Editing benötigten, so unterscheiden sich die Schätzungen der Teilnehmer*innen durchaus. T1 konnte die eigene Arbeitszeit beinahe exakt einschätzen, T2 schätzte diese 10 Minuten, T3 sogar 25 Minuten länger ein als tatsächlich gemessen.

Teilnehmer	*Ø / Segment*	*Segmente / Gesamt*	*Gesamt*	*Schätzung / Interview*
T1	00:02:42	00:43:11	00:51:58	00:50:00
T2	00:01:31	00:24:12	00:50:16	01:00:00
T3	00:01:53	00:30:09	00:46:10	01:10:00

Tabelle 1: Arbeitszeiten in Stunden-Minuten-Sekunden

Wie durch Abb. 5 sichtbar wird, variiert die durchschnittliche Bearbeitungsdauer pro Segment stark. Für sechs der 16 Segmente beträgt die Bearbeitungszeit weniger als 60 Sekunden, für die Segmente 6, 10 und 16 wurde hingegen mit mindestens dreieinhalb bis maximal acht Minuten die meiste Zeit aufgewandt. Dabei ist zu beachten, dass die durchschnittliche Wortanzahl pro Segment, also pro Satz, für den gesamten Zieltext 22,6 betrug. Die Segmente 6, 10 und 16 zählen hier mit einer Länge von je 42, 31 und 33 Wörtern auch zu den drei längsten Segmenten im gesamten Text, was sich in der Bearbeitungszeit widerspiegelt.

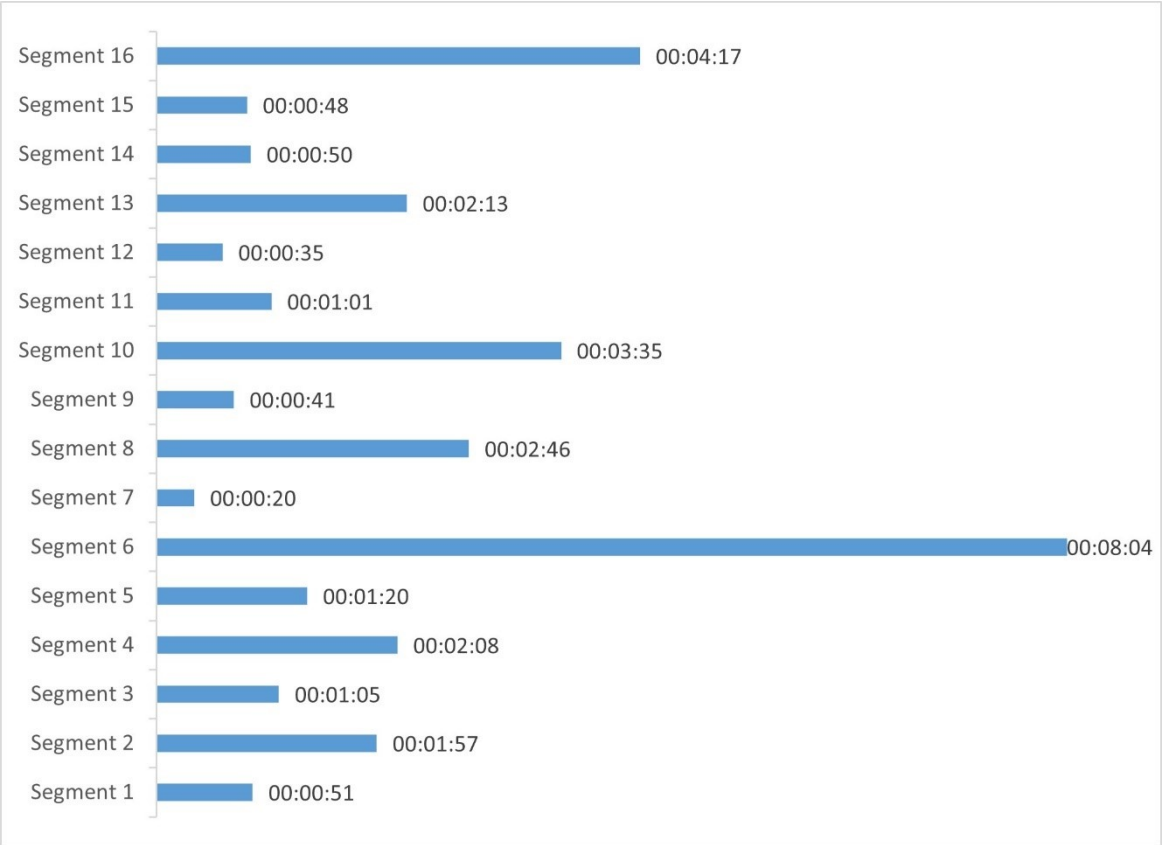


Abbildung 5: Durchschnittliche Bearbeitungszeit aller Teilnehmer*innen pro Segment

Während sich Tabelle 1 und Abb. 5 auf die Dauer beziehen, die von den Teilnehmer*innen für die Bearbeitung des Textes mit Maus und Tastatur aufgewandt wurden, so zeigt Abb. 6 die Gesamtdauer aller Fixationen aus der Eyetracking-Aufnahme und deren Aufteilung auf die vier Interessensbereiche Browser, Ausgangstext, Zieltext und Feld mit Translation Memory Treffern. In Übereinstimmung mit der Bearbeitungsdauer aus Tabelle 1 ist auch hier erkennbar, dass die Verweildauer von T1s Fixationen primär auf Ausgangs- und Zieltext entfällt und nur sechs Minuten mit Betrachten des Browsers verbracht wurden. T2 weist hingegen eine deutlich längere Verweildauer im Browser auf und betrachtete Ausgangs- und Zieltext dafür weniger lange als T1 und T3. Ebenso ist erkennbar, dass alle drei Teilnehmer*innen das Feld mit den Translation Memory Treffern kaum betrachteten.

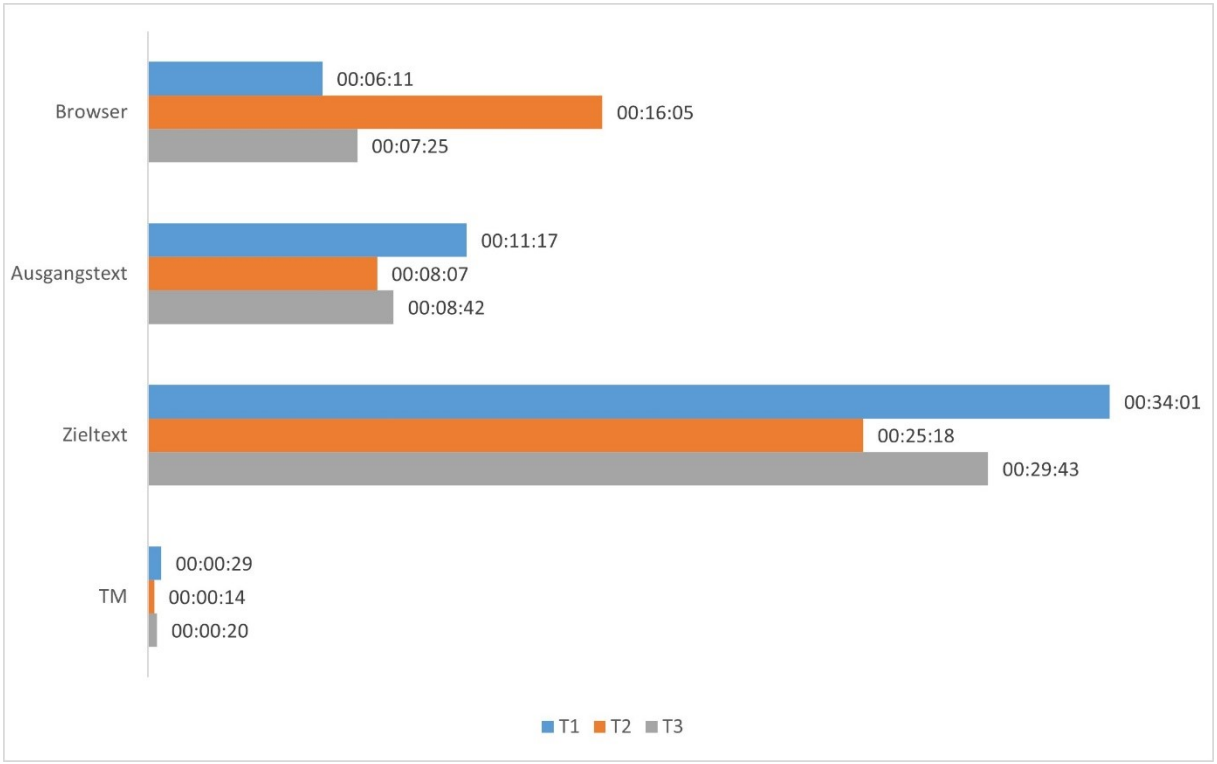


Abbildung 6: Verweildauer pro Interessensbereich

Anhand der gesamten Arbeitszeit aus Tabelle 1 konnten die Wörter pro Minute und Stunde berechnet werden, die jede/r der Teilnehmer*innen erzielen konnte. Mit durchschnittlich sieben Wörtern pro Minute und 440 Wörtern pro Stunde waren die Teilnehmer*innen, wie in Tabelle 2 dargestellt, besonders produktiv, da nach Robert (2013: 32) selbst professionelle Post-Editor*innen im Durchschnitt nur 437,5 Wörter pro Stunde schaffen. Trotz dieser hohen Produktivität ist zu beachten, dass die Qualität der editierten Texte in dieser Studie nicht evaluiert wurde und daher von einer hohen Produktivität nicht automatisch auf eine gute Qualität geschlossen werden kann.

Teilnehmer	*Wörter pro Minute*	*Wörter pro Stunde*
T1	6,99	419,3
T2	7,19	431,81
T3	7,83	469,84
Mittelwert	7,34	440,3

Tabelle 2: Produktivität

5.2 Technischer Aufwand

Besonders aussagekräftig für den technischen Aufwand, der mit der Post-Editing-Aufgabe verbunden war, sind die in Tabelle 3 dargestellten Änderungen, die die Teilnehmer*innen im Zieltext vornahmen. Für den gesamten Text betrug die durchschnittliche Anzahl an Änderungen pro Segment 2. T1 nahm mit durchschnittlich 3,1 die meisten Änderungen vor, T2 mit 1,4 die wenigsten. Interessant ist hier auch der Vergleich der gemessenen Änderungen mit der Einschätzung der Teilnehmer*innen. T1 schätzte nach Durchführen des Post-Editings, durchschnittlich 3 bis 5 Änderungen vorgenommen zu haben, was dem gemessenen Wert sehr nahe kommt. Auch T3 lag mit einer Schätzung von 2 Änderungen bei 1,6 tatsächlichen Änderungen recht nahe. T2 überschätzte den Wert der gemessenen Änderungen jedoch deutlich.

Teilnehmer	*Ø Änderungen / Segment*	*Schätzung / Interview*
T1	3,1	3 bis 5
T2	1,4	3 bis 5
T3	1,6	2
Mittelwert	2	/

Tabelle 3: Durchschnittliche Anzahl an Änderungen pro Teilnehmer*in

In Abb. 7 ist erkennbar, wie viele Änderungen in jedem der 16 Segmente durchschnittlich vorgenommen wurden. In Übereinstimmung mit den Bearbeitungszeiten pro Segment aus Abb. 5 wurde in den Segmenten 6, 10 und 16 nicht nur am längsten editiert, sondern auch die meisten Änderungen vorgenommen. Auch wenn die durchschnittliche Anzahl an Änderungen unter den Teilnehmer*innen teils stärker variierte, so wurden in Segment 6 von T1, T2 und T3 je 7, 10 und 8 Änderungen vorgenommen.

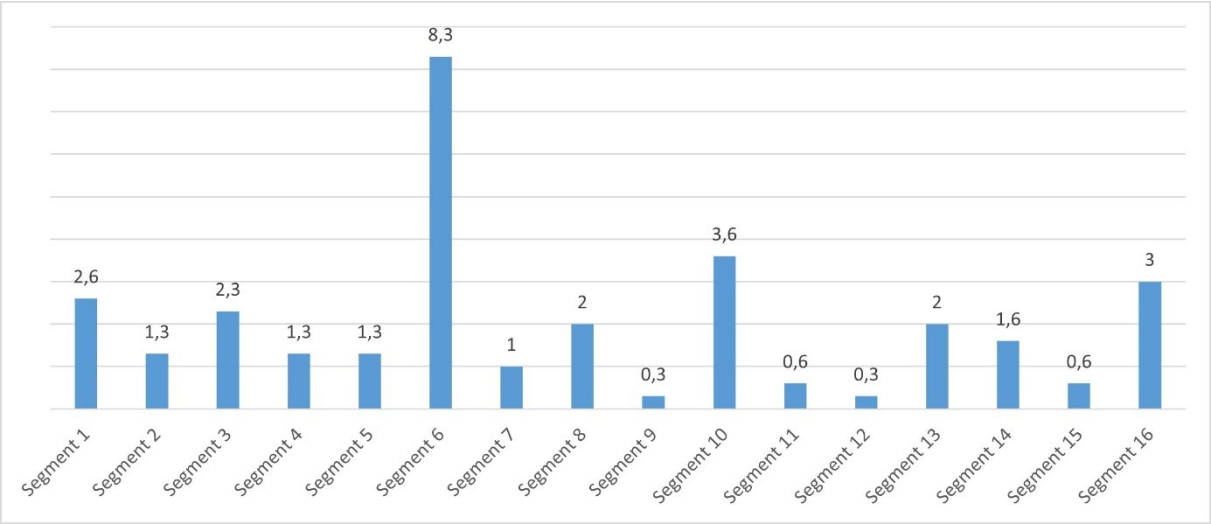


Abbildung 7: Durchschnittliche Anzahl an Änderungen aller Teilnehmer*innen pro Segment

Die gesamte Anzahl an Änderungen, die die Teilnehmer*innen im Text vornahmen, ist in Tabelle 4 dargestellt und steht in gewissem Zusammenhang mit der Bearbeitungszeit aus Tabelle 1. T1 nahm bei weitem die meisten Änderungen vor und hatte auch die längste Bearbeitungszeit in *MemoQ*. T2 nahm etwas weniger Änderungen als T3 vor, jedoch verbrachte T2 auch deutlich weniger Zeit mit dem Bearbeiten des Textes. Darüber hinaus ist ersichtlich, dass T1 in jedem der 16 Segmente mindestens eine Änderung vornahm und somit den ganzen Text post-editierte. T2 nahm hingegen nur in der Hälfte aller Segmente Änderungen vor, und auch T3 übernahm in einigen Segmenten die maschinelle Übersetzung ohne Änderung. Dabei ist interessant, dass T2 und T3 in den Segmenten 7, 9, 12 und 14 keine Änderungen vornahmen, da die maschinelle Übersetzung scheinbar zufriedenstellend war. T1 nahm in diesen vier Segmenten im Durchschnitt hingegen je 2,5 Änderungen vor.

Teilnehmer	*Änderungen*	*Segm. ohne Änderungen*
T1	49	0/16
T2	23	8/16
T3	26	5/16

Tabelle 4: Anzahl an Änderungen im gesamten Text

Die in Tabelle 5 angeführten Werte der Metriken TER, BLEU, METEOR sowie der Levenshtein Editierdistanz, sind ebenfalls aussagend dafür, in welchem Umfang der Text durch die Teilnehmer*innen editiert wurde.

Teilnehmer	*Levenshtein*	*L. Ähnlichkeit*	*TER*	*BLEU*	*METEOR*
T1	461	86%	20,63	0,69	0,7
T2	198	93%	12,18	0,86	0,77
T3	298	90%	12,93	0,83	0,76

Tabelle 5: Evaluierung der post-editierten Texte durch Metriken

BLEU und METEOR vergeben Werte zwischen 0 und 1, je näher der Wert an 1 liegt, umso ähnlicher sind die maschinelle Rohübersetzung und der post-editierte Text. Im Falle der post-editierten Texte gilt jedoch: umso mehr Änderungen beim Post-Editing an dem Text vorgenommen wurden, umso stärker geht der Wert in Richtung 0. TER vergibt hingegen einen Wert zwischen 0 und 100, der für den Anteil an vorgenommenen Änderungen steht und als Prozentwert verstanden werden kann. Hier bedeutet also ein höherer Wert auch eine höhere Anzahl an Änderungen. Der Wert der Levenshtein Distanz steht für die Anzahl an Änderungen zwischen den beiden Texten, die als Zeichenketten evaluiert werden. Zur besseren Verständlichkeit wird die Ähnlichkeit hier auch mit einem Prozentwert angegeben.

Die Werte zeigen für jede/n der Teilnehmer*innen ein einheitliches Bild, da T1, wie aus Tabelle 4 bekannt ist, insgesamt die meisten Änderungen vornahm und somit auch die höchsten Levenshtein- und TER-Werte sowie die niedrigsten BLEU- und METEOR-Werte erzielte. T2 nahm insgesamt die wenigsten Änderungen am Text vor, wonach dieser der maschinellen Rohübersetzung am ähnlichsten ist und so auch die niedrigsten Levenshtein- und TER Werte sowie die höchsten BLEU- und METEOR-Werte.

5.3 Kognitiver Aufwand

Nach den diversen Messergebnissen bezüglich Arbeitsdauer und durchgeführten Änderungen sollen nun die durch den Eyetracker aufgezeichneten Daten präsentiert werden, die mitunter am besten auf den kognitiven Aufwand während der Post-Editing-Aufgabe schließen lassen. Zuerst sollen jedoch die Einschätzungen der Teilnehmer*innen bezüglich der Schwierigkeit des englischen Ausgangstextes und der Qualität der maschinellen Übersetzung im Zieltext erwähnt werden, die stark subjektive Indikatoren für den kognitiven Aufwand sind.

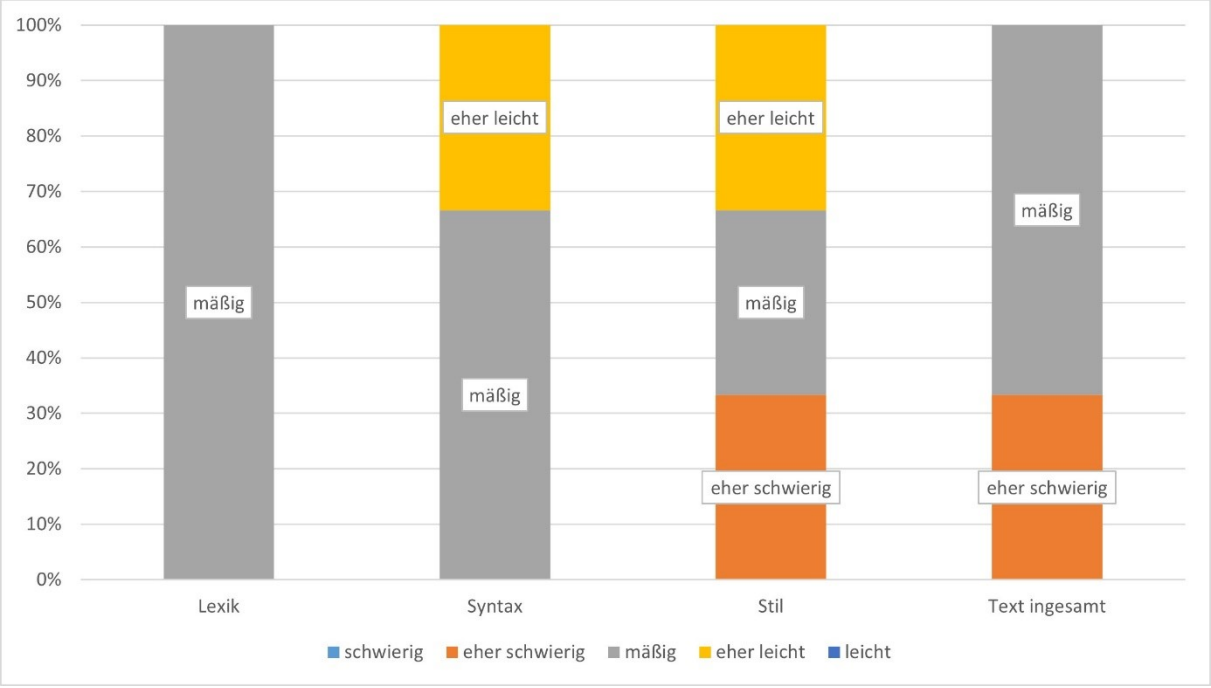


Abbildung 8: Einschätzung der Schwierigkeit des Ausgangstextes

Wie in Abb. 8 ersichtlich ist, waren sich die Teilnehmer*innen lediglich einig, dass die Lexik des Ausgangstextes als mäßig einzuschätzen ist. Auch die Syntax wurde als mäßig bis eher leicht eingeschätzt, der gesamte Text als mäßig bis eher schwierig. Bezüglich des Stils gab jede/r der drei Teilnehmer*innen eine andere Einschätzung ab, was auch die Subjektivität dieses Textbestandteils hervorhebt. Nachdem die Teilnehmer*innen in drei der vier Kategorien überwiegend *mäßig* wählten, kann davon ausgegangen werden, dass der

ausgewählte Text sowohl sprachlich als auch inhaltlich dem Kompetenz- und Wissensstand der Teilnehmer*innen entsprach.

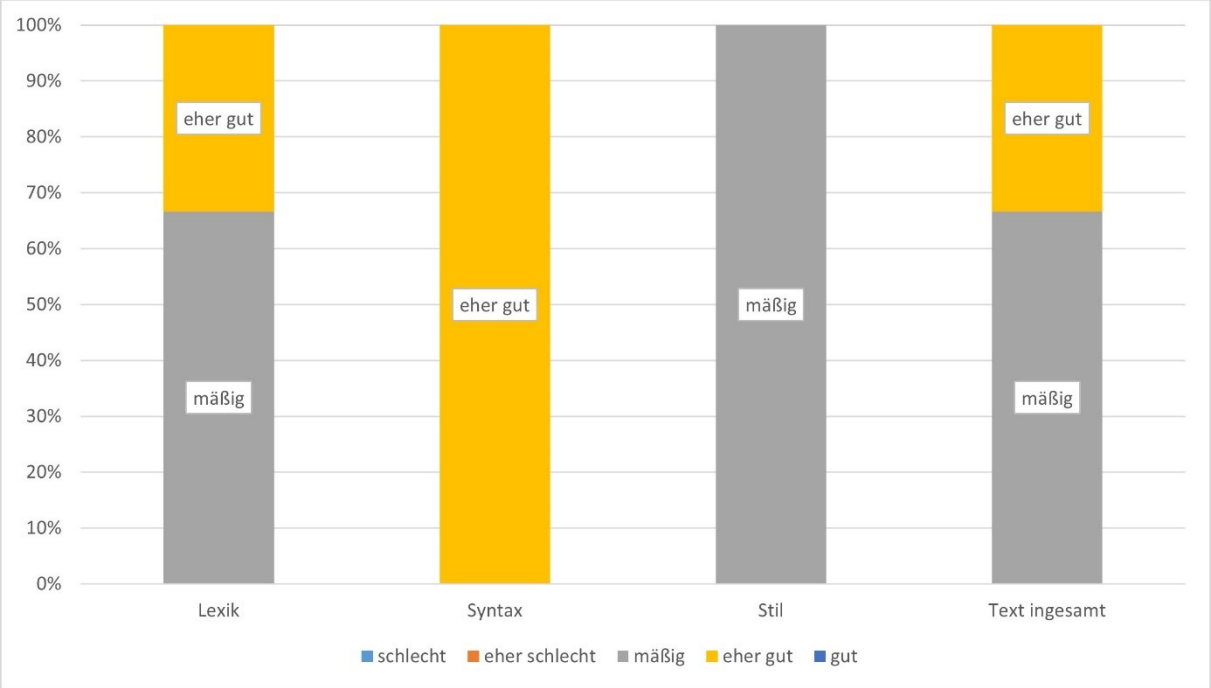


Abbildung 9: Einschätzung der Qualität des maschinell übersetzten Zieltextes

Die Einschätzungen der Teilnehmer*innen bezüglich der Qualität des maschinell übersetzten Zieltextes in Abb. 9 fallen hingegen etwas einheitlicher aus, die Syntax wurde von allen als eher gut, der Stil als mäßig eingeschätzt. Die Lexik und der Text insgesamt wurden ebenfalls überwiegend als mäßig bis eher gut eingeschätzt. Auch hier kann davon ausgegangen werden, dass das gewählte MÜ-System *DeepL* in Kombination mit dem vorliegenden Ausgangstext eine Übersetzung erzeugen konnte, die von allen Teilnehmer*innen tendenziell als mäßig bis eher gut eingestuft wurde und somit eine gute Grundlage für die Post-Editing-Aufgabe darstellte, da nicht unverhältnismäßig viele Änderungen durchgeführt werden mussten.

In Tabelle 6 sind diverse Messergebnisse des Eyetrackers aufgelistet, die sich auf die Fixationen aus der gesamten Aufnahme, also während der gesamten Arbeitszeit der Teilnehmer*innen, beziehen. Diese Werte sind in Bezug auf die gesamte Aufnahme weniger aussagekräftig, sie dienen jedoch als Referenz für die nachfolgenden Abbildungen und Tabellen, die sich gesondert auf die Fixationen in den vier Interessensbereiche beziehen. Für die gesamte Aufnahme kann festgestellt werden, dass T2 die höchste Anzahl an Fixationen hatte, diese im Durchschnitt jedoch am kürzesten waren. T3 hatte hingegen die wenigsten Fixationen, diese waren im Schnitt jedoch umso länger. Die durchschnittliche Pupillengröße pro Fixation variiert unter den Teilnehmer*innen stark, ist jedoch auf Grund von Faktoren wie

der subjektiven Lichtsensibilität nicht direkt vergleichbar und soll in Tabelle 9 pro Teilnehmer*in und pro Interessensbereich gesondert präsentiert werden.

Teilnehmer	*Anzahl / Fixationen*	*Ø Dauer / Fixation*	*Ø Pupillengröße / Fixation*
T1	6865	324 ms	2,56 mm
T2	8283	281 ms	5,32 mm
T3	6325	374 ms	4,88 mm

Tabelle 6: Fixationen insgesamt

Angelehnt an die Verweildauer pro Interessensbereich aus Abb. 6 sollen Abb. 10, 11 und 12 nun die Aufteilung der Fixationen auf die vier Interessensbereiche für jede/n Teilnehmer*in veranschaulichen. T1 konzentrierte einen Großteil der Fixationen auf Ausgangs- und Zieltext, insgesamt rund 88%. In den Bildschirmbereich des Internetbrowsers, der zur Recherche während des Post-Editings diente, entfielen jedoch nur 11%, beziehungsweise 817 der insgesamt 6.865 Fixationen. Außerdem wurde die Spalte, die den Zieltext beinhaltete und in der Änderungen vorgenommen wurden, von T1 dreimal so oft betrachtet wie die Spalte mit dem Ausgangstext. Auf den Bereich mit den Translation Memory Treffern entfielen beinahe keine Fixationen.

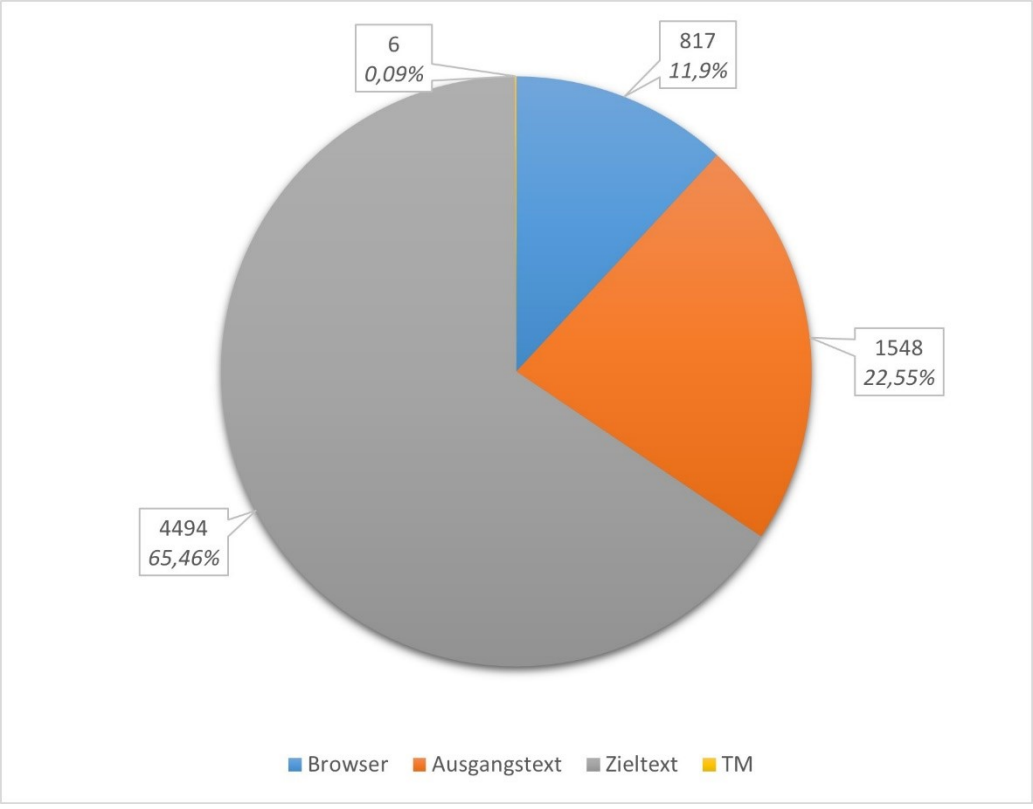


Abbildung 10: Anzahl und Aufteilung von Fixationen pro Interessensbereich: T1

Bei T2 fielen nur rund 67% aller Fixationen auf Ausgangs- und Zieltext, dafür allerdings rund 32%, also 2.658 der insgesamt 8.283 Fixationen, auf den Browser. Wie T1 betrachtete T2 den Zieltext dreimal so häufig wie den Ausgangstext, und der Bereich mit den Translation-Memory Treffern wurde auch von T2 kaum betrachtet.

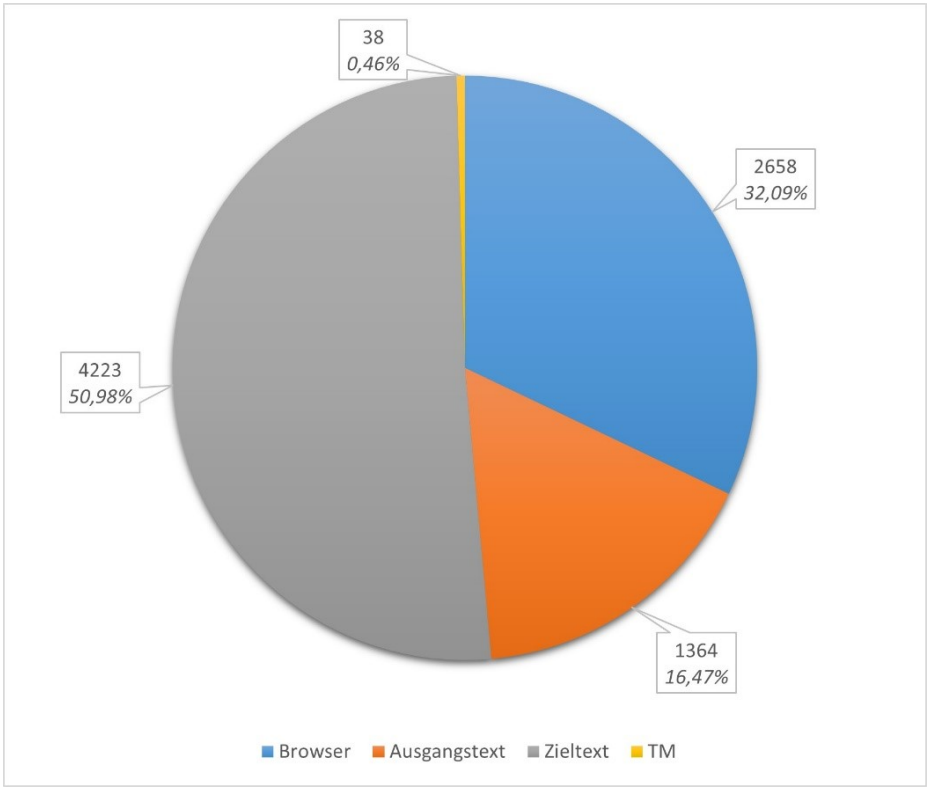


Abbildung 11: Anzahl und Aufteilung von Fixationen pro Interessensbereich: T2

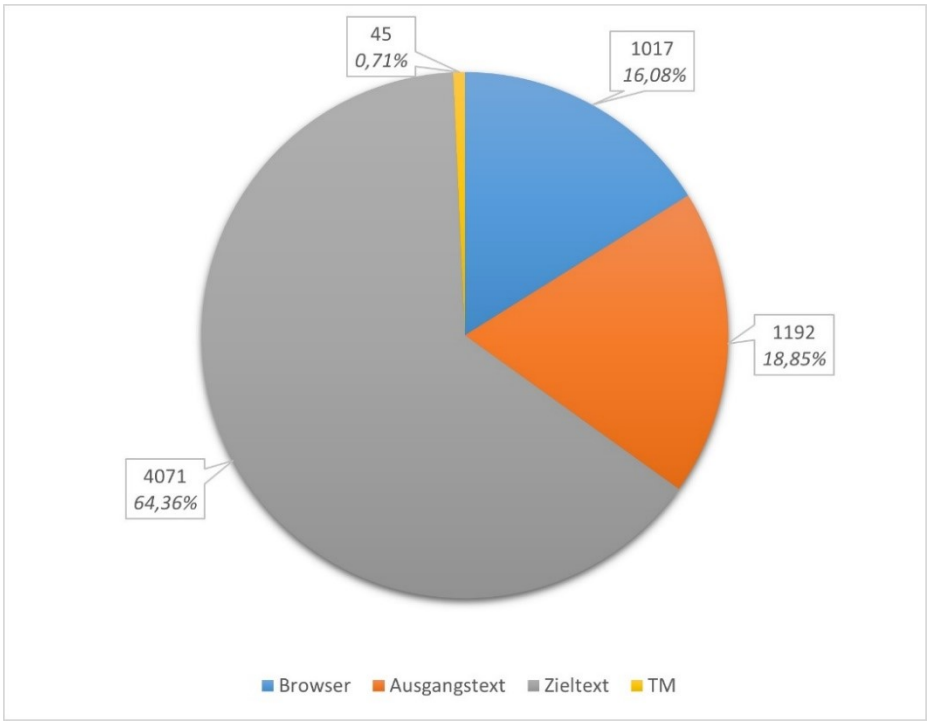


Abbildung 12: Anzahl und Aufteilung von Fixationen pro Interessensbereich: T3

Bei T3 entfielen ähnlich wie bei T1 rund 83% aller Fixationen auf Ausgangs- und Zieltext, auf den Browser hingegen 16%, also etwas mehr als bei T1. T3 betrachtete den Zieltext sogar 3,5-mal häufiger als den Ausgangstext und hatte ebenfalls kaum Fixationen im Translation-Memory Bereich.

Aus den Eyetracking-Daten lassen sich ebenfalls alle Wechsel zwischen Interessensbereichen herauslesen, die in Tabelle 7 dargestellt sind. Konkret ist darunter jede Instanz zu verstehen, bei der auf eine Fixation in einem Interessensbereich A eine Fixation in einem Interessensbereich B folgt. Der/die Teilnehmer*in hat somit die visuelle Aufmerksamkeit auf ein neues Objekt in einem neuen Bereich geleitet und versucht aus diesem Objekt visuelle Informationen zu extrahieren. Auch wenn im Fall der drei Teilnehmer*innen keine genauen Angaben über die Interessensbereiche, zwischen denen gewechselt wurde, vorliegen, so kann davon ausgegangen werden, dass beispielsweise einige der vorgenommenen Wechsel zwischen Ausgangs- und Zieltext geschehen sind, da diese beim Post-Editing regelmäßig miteinander verglichen werden müssen.

Teilnehmer	*Wechsel zwischen Interessensbereichen*	*Ø Anzahl / Fixationen vor Wechsel*
T1	1203	5,7
T2	1619	5,1
T3	612	10,3

Tabelle 7: Wechsel zwischen Interessensbereichen

Die Anzahl der Wechsel ist dabei in Relation zur Anzahl der gesamten Fixationen pro Teilnehmer*in aus Tabelle 6 zu verstehen. Daraus lässt sich auch auf die durchschnittliche Anzahl an Fixationen schließen, die ein/e Teilnehmer*in innerhalb eines Interessensbereichs gemacht hat, bevor in einen anderen gewechselt wurde. T1 und T2 haben mit 5,7 und 5,1 Fixationen pro Interessensbereich während der gesamten Aufnahme relativ viele Wechsel zwischen den Bereichen vorgenommen. T3 hat hingegen deutlich weniger Wechsel vorgenommen und pro Bereich vor einem Wechsel rund 10 Fixationen gemacht. Aus Tabelle 6 geht in Kombination mit diesem Wert ebenfalls hervor, dass T3 die durchschnittlich höchste Fixationsdauer hat und somit im Vergleich zu den anderen beiden Teilnehmer*innen pro Bereich besonders viele und besonders lange Fixationen vornahm.

Teilnehmer	*Browser*	*Ausgangstext*	*Zieltext*	*TM*
T1	225	257	393	282
T2	269	236	326	250
T3	238	241	358	273
Mittelwert	244	245	359	268

Tabelle 8: Ø Fixationsdauer pro Interessensbereich in Millisekunden

Die durchschnittliche Fixationsdauer in den verschiedenen Interessensbereichen zeigt bei allen Teilnehmer*innen ein recht einheitliches Bild, wie in Tabelle 8 zu sehen ist. Die durchschnittliche Dauer von 244 beziehungsweise 245 ms in den Bereichen Browser und Ausgangstext, in denen die Teilnehmer*innen primär lesen mussten, deckt sich mit der von Jakobsen (2019: 401) festgestellten durchschnittlichen Fixationsdauer beim Lesen von 200-250 ms. Mit 359 ms sind die Fixationen im Bereich des Zieltextes, in dem sowohl gelesen als auch geschrieben werden musste, jedoch durchschnittlich 42% länger als in allen anderen Bereichen. Die Kombination an kognitivem und technischem Aufwand, die beim Betrachten und Bearbeiten des Zieltextes entsteht, wirkt sich direkt auf die Fixationsdauer aus und sorgt dafür, dass diese umso mehr ansteigt, je höher der Aufwand für den/die Bearbeiter*in ist.

Teilnehmer	*Browser*	*Ausgangstext*	*Zieltext*	*TM*
T1	3,25	2,49	2,45	2,62
T2	6,46	5,19	5,32	5
T3	5,69	4,72	4,75	4,52

Tabelle 9: Ø Pupillengröße pro Interessensbereich in Millimeter

In Tabelle 9 soll ergänzend zu den in Tabelle 6 gelisteten durchschnittlichen Pupillengrößen pro Fixation im gesamten Text nun auch auf die Größen in den vier Interessensbereichen eingegangen werden. Alle drei Teilnehmer*innen weisen bei Fixationen in den Bereichen Ausgangstext, Zieltext und TM relativ einheitliche Pupillengrößen auf, bei Fixationen auf den Browser waren die Pupillen hingegen durchschnittlich 25% größer als in allen anderen Bereichen. Wie in Kapitel 2.4.5.2 erwähnt wurde, konnte ein Zusammenhang zwischen höherem kognitivem Aufwand und erweiterten Pupillen nicht belegt werden. Weiters wird die in Tabelle 8 aufgelistete längere Fixationsdauer im Zieltextbereich nicht durch einen größeren Pupillendurchmesser bestätigt.

Abschließend sollen nun noch zwei Screenshots präsentiert werden, die sämtliche Fixationen, die die Teilnehmer*innen während der gesamten Arbeitszeit machten, visuell darstellen und auch deren Aufteilung auf die vier festgelegten Interessensbereiche Browser, Ausgangstext, Zieltext und Translation-Memory Treffer zeigen. Da die durch T3 erzielten Messergebnisse in den meisten Fällen zwischen jenen von T1 und T2 lagen und somit als relativ repräsentativ gesehen werden können, wurden auch die zwei Screenshots von der Aufnahme von T3 ausgewählt. In Abb. 13 sind alle 6.325 Fixationen, die T3 während der gesamten Aufnahme machte, als blaue Kreise zu sehen, sowie deren Position im Verhältnis zu den Interessensbereichen.

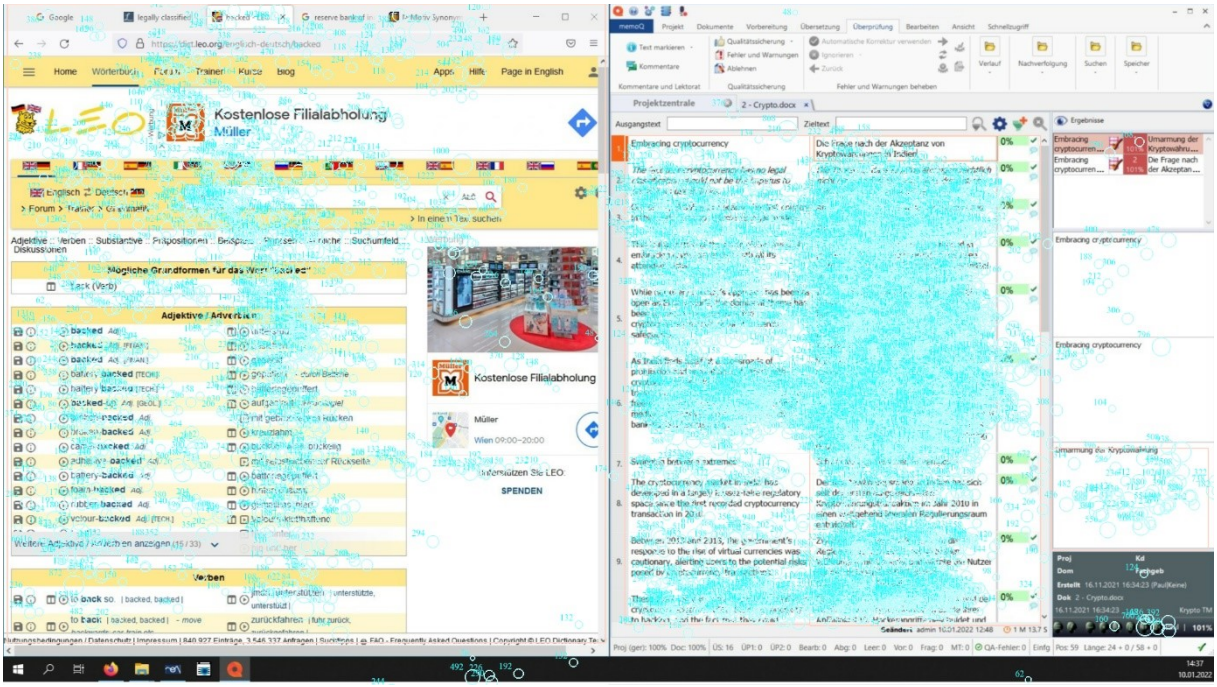


Abbildung 13: Position aller Fixationen: T3

Abb. 14 basiert auf den in Abb. 13 enthaltenen Informationen bezüglich Position und Häufigkeit von Fixationen und stellt diese als leichter verständliche *Heatmap* dar. In den grün eingefärbten Bereichen traten eher weniger Fixationen auf, der rote Punkt in der Zieltextspalte war hingegen der Bereich mit der höchsten Fixationsdichte innerhalb der gesamten Aufnahme. Hier ist zu erwähnen, dass die Ausgangs- und Zieltextspalten über eine Scrollfunktion verfügen. T3 konnte also das Segment, das er/sie gerade bearbeitete, immer an derselben Position relativ zum Rest des *MemoQ*-Fensters platzieren. In diesem Fall war diese Position in etwa auf Höhe der Bildschirmmitte.

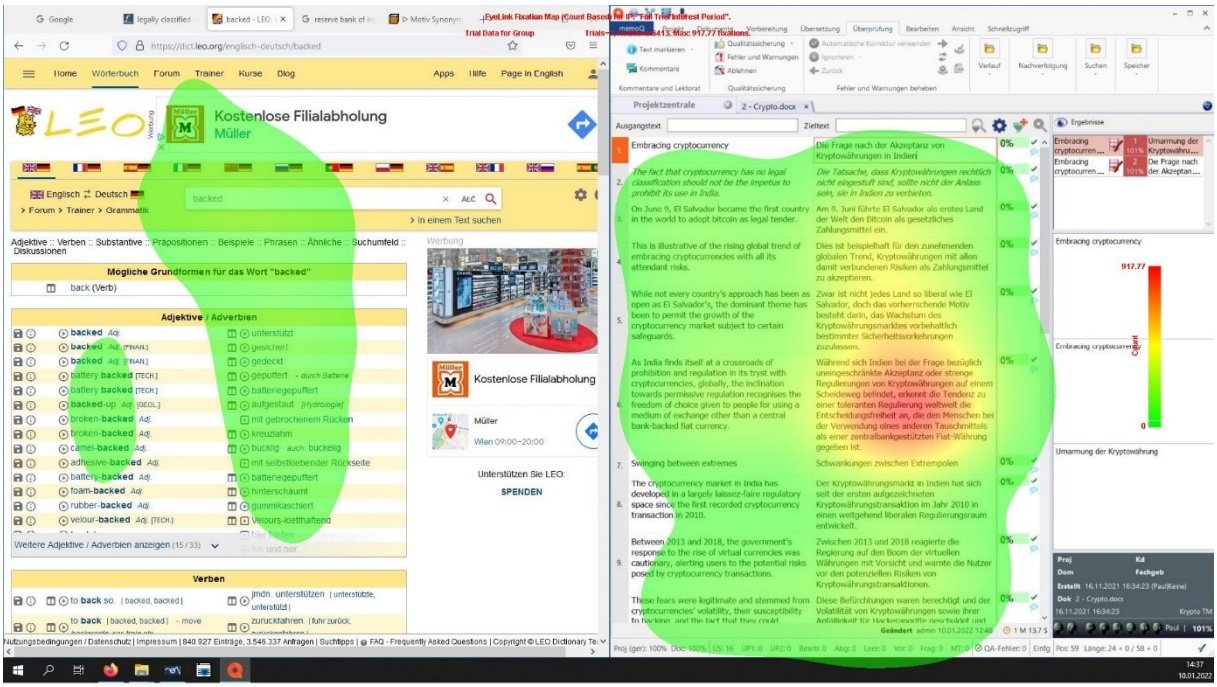


Abbildung 14: Fixationsdichte als Heatmap: T3

5.4 Interviews

Die nachfolgende Auswertung der Interviews, die mit den Teilnehmer*innen vor und nach der Post-Editing-Aufgabe geführt wurden, sollen einen Einblick in ihre subjektive Einstellung gegenüber PE bieten und in der anschließenden Diskussion der Ergebnisse dazu dienen, bestimmte Korrelationen mit den objektiven Messergebnissen festzustellen. Bei der Präsentation der Ergebnisse wurde darauf geachtet, die Teilnehmer*innen so oft wie möglich im Wortlaut zu zitieren.

5.4.1 Vor der Post-Editing-Aufgabe

Zum Einstieg des Interviews wurden die Teilnehmer*innen gefragt, was sie unter dem Begriff Post-Editing verstehen. Alle drei gaben dabei eine ähnliche Definition, die einerseits einen durch eine Maschine übersetzten Text und zum anderen eine menschliche Nachbearbeitung dieses Textes umfasste. Das Ziel von PE ist laut T3, dass die Übersetzung „sprachlich so nahe an der natürlichen Sprache ist, wie wenn es ein echter Mensch geschrieben hätte“ (T3: 3f). Als Mehrwert, den PE mit sich bringt, gab T2 an, dass es „eben Zeit und dadurch auch Geld spart“ (T2: 11). T1 und T3 gaben weiters an, im Rahmen ihrer universitären Ausbildung PE mehrmals angewandt zu haben, etwa für „Hausübungen und Aufträge und Prüfungen“ (T1: 17) oder „wenn man ganz dringend was gebraucht hat und keine Zeit gehabt hat“ (T3: 11). T3 nutzte die Vorschläge maschineller Übersetzungssysteme jedoch primär als Inspirationsquelle oder für das Vergleichen mit der selbst angefertigten Übersetzung, somit war T1 der/die

einzige Teilnehmer*in, der/die tatsächlich etwas Erfahrung mit laienhaftem PE hatte. T2 gab hingegen an, noch nie PE durchgeführt zu haben, erwähnte jedoch im Rahmen des Berufs, dass „bei den Untertiteln, die ich übersetzt habe, da waren mal welche dabei, die haben für mich sehr maschinell übersetzt ausgeschaut“ (T2: 14f). T2 gab dafür an, mit der Revision von Übersetzungen und Texten gut vertraut zu sein und diese im Rahmen des Berufs regelmäßig durchzuführen. T1 und T3 erwähnten, Revisionen nur im Kontext von Lehrveranstaltungen für andere Studienkolleg*innen durchgeführt zu haben, jedoch nie im beruflichen Kontext.

Auf die Frage, ob sie eine PE-Ausbildung erhalten hätten, antworteten alle drei Teilnehmer*innen mit einem klaren Nein. T1 gab an, sich das PE „selber beigebracht“ (T1: 22) und es „nach Gefühl gemacht“ (T1: 24) zu haben. T1 gab weiters an, PE gegenüber „sehr aufgeschlossen“ (T1:26) zu sein. T2 erwähnte, eine „recht neutrale Einstellung“ (T2: 30) zu haben und äußerte Bedenken bezüglich des Verlustes von Arbeitsplätzen für Übersetzer*innen durch stets besser werdende MÜ-Systeme. Trotzdem erkannte T2 die Vorteile von PE und meinte, „es ist natürlich sowohl zeit- also auch kostensparend, es wird mehr oder weniger nichts anderes übrigbleiben als das“ (T2: 29f). Auch T3 teilte diese pragmatische Sichtweise und gab an, PE gegenüber „eher aufgeschlossen“ (T3: 31) zu sein, da „die Arbeitslast nicht zu bewältigen ist ohne solche Möglichkeiten“ (T3: 32f). Weiters bestätigte T3, während der Studienzeit regelmäßig MÜ-Systeme verwendet zu haben, etwa um „zu schauen, ob ich auf einem ähnlichen Niveau, beziehungsweise ob mein Niveau sogar besser ist wie das von der Machine Translation“ (T3: 43f). Auch T1 gab an MÜ für private Zwecke zu verwenden, wie etwa für *gisting* auf fremdsprachigen Internetseiten. T2 gab an *DeepL* im beruflichen Alltag vereinzelt als Hilfestellung für schwierige Textpassagen bei der Übersetzung zu verwenden, betonte jedoch, dass die MÜ eher „als Inspiration, nicht als Vollübersetzung“ dient (T2: 47f).

Die Teilnehmer*innen wurden darüber informiert, dass der Text zum Thema Kryptowährungen, den sie anschließend post-editieren werden, mit *DeepL* übersetzt wurde. T1 und T2 erwarteten sich von der MÜ eine „relativ gute Qualität“ (T2: 52), vor allem in Bezug auf die korrekte Verwendung der Terminologie, gaben jedoch an, dass es am ehesten in Bezug auf den Stil zu Fehlern und Problemen kommen könnte. T3 erwartete sich von der MÜ eine mäßige Qualität, hatte jedoch Bedenken wegen der terminologischen Konsistenz des Textes und merkte an, dass „das Ganze vielleicht nicht so harmonisch ist, wie wenn es eben ein Mensch machen würde“ (T3: 51f). Die Teilnehmer*innen wurden auch gefragt, ob für sie persönlich Post-Editing oder Humanübersetzen mit einem höheren kognitiven Aufwand, also einer höheren mentalen Anstrengung während des Arbeitens, verbunden ist. T1 und T3

entschieden sich beide für das Humanübersetzen und nannten den höheren Aufwand durch Vorbereitung als Grund, „weil man dann die ganze Terminologie erst kennenlernen muss, und insofern wäre es dann anstrengend, sich in das Ganze einzuarbeiten“ (T3: 60ff). T2 bezeichnete PE als mental aufwändigere Arbeitsweise und betonte, beim Humanübersetzen mehr Vertrauen in die eigene Leistung und ein „von Grund auf [...] positiveres Gefühl“ (T2: 67) zu haben. PE wäre laut T2 mit einem höherem Aufwand verbunden, denn er/sie wäre „einfach genauer und würde nochmal mehr nachschauen, ob das jetzt wirklich die richtigen Entsprechungen sind“ (T2: 70f).

Bei der Frage, ob die Teilnehmer*innen voraussichtlich mit Post-Editing oder Humanübersetzen schneller beziehungsweise produktiver sein können, entschieden sich T1 und T3 für Post-Editing. Nicht nur würde man sich so einen Großteil des Tippens ersparen, sondern auch die Zeit, die man mit Finden von Übersetzungsvorschlägen verbringt, „weil man generell als Mensch, als Humanübersetzer, kritischer gegenüber dem ist, was man selber hinschreiben würde, als wie wenn es eh schon dasteht“ (T3: 69ff). T2 meinte jedoch, selbst voraussichtlich beim Humanübersetzen schneller als beim PE zu sein und führte dies auf ein größeres Vertrauen in die eigene Übersetzungskompetenz zurück. Beim Editieren der maschinellen Übersetzung würde T2 laut eigener Angabe zu viel Zeit mit dem Suchen von Fehlern verbringen und mehr Durchgänge als nötig machen, „um sicher zu sein, dass ich nichts übersehen habe, was da jetzt falsch sein könnte“ (T2: 78). Abschließend wurden die Teilnehmer*innen mit der fiktiven Situation konfrontiert, einen Auftrag für die Übersetzung eines Textes zu erhalten, der publiziert werden soll, und bei dem ihnen offensteht, ob sie diesen selbst übersetzen möchten, oder ob sie den Text zuerst maschinell übersetzen lassen und anschließend post-editieren möchten. Alle Teilnehmer*innen entschieden sich in diesem Fall dafür, den Text lieber selbst zu übersetzen, insofern sie dafür ausreichend Zeit erhalten. T2 und T3 begründeten diese Entscheidung mit dem finanziellen Aspekt, da sie für eine Humanübersetzung auch mehr verlangen könnten als für ein Post-Editing. Obwohl T3 im Laufe des Interviews angab, dass das Humanübersetzen voraussichtlich langsamer und mental anstrengender als Post-Editing sei, entschied sich T3 in diesem Szenario dennoch dafür, „weil es natürlich eine persönliche Herausforderung ist, so einen Text zu übersetzen und auch spannend, und deshalb haben wir das Ganze ja studiert“ (T3: 76ff).

5.4.2 Nach der Post-Editing-Aufgabe

Nach Durchführen der Aufgabe wurden die Teilnehmer*innen zuerst gefragt, ob ihnen das Post-Editing an sich gefallen hat. T1 und T3 bejahten dies und T3 führte weiter an, dass er/sie Post-Editing als eine Sonderform des Übersetzen erachtet, denn „selbst wenn man [die Übersetzung] nicht hinschreibt, im Kopf passieren ähnliche Vorgänge“ (T3: 9f). T2 stand dieser Frage mit gespaltener Meinung gegenüber, denn „generell macht es schon Spaß“ (T2: 3), jedoch gab T2 auch an, dass einige wenige problematische Stellen in der Übersetzung den gesamten Arbeitsablauf gestört haben. „Es ist einfach frustrierend, wenn du nicht und nicht auf irgendeine Lösung draufkommst, die dich zufriedenstellt“. (T2: 6f) Bezüglich Problemen, die spezifisch bei Post-Editing auftreten, gaben T1 und T2 die durch die MÜ vorgegebene Syntax an. „Man ist natürlich sehr an den maschinell übersetzten Text gebunden von der Satzstruktur.“ (T1: 5f) T2 erwähnte auch, dass das Vorhandensein einer Übersetzung die eigene Handlungsfähigkeit einschränkt: „Wenn du einmal den Text eben so liest und die Sätze schon so formuliert sind, dann kommst du vielleicht [...] gar nicht mehr auf andere Lösungsansätze.“ (T2: 24ff) Alle drei Teilnehmer*innen erwähnten zusätzlich, dass Segment 6 besonders schwierig zu editieren war, und dass sie mit ihren Lösungen für das Segment unzufrieden seien, da es laut T3 „lexikalisch einfach irgendwie komisch war, auch von den Wörtern her, weil es einfach überhaupt nicht zusammengepasst hat“ (T3: 14f). T2 meinte hingegen, dass in Segment 6 „meistens nicht mal die lexikalischen Sachen das Mega-Problem [waren]“ (T2: 33f), dafür die größten Schwierigkeiten vom vorgegebenen Satzbau ausgingen. T2 meinte diesbezüglich, es „wäre vielleicht einfach gewesen, wenn ich es selber übersetzt hätte“ (T2: 29). Alle Teilnehmer*innen erwähnten außerdem, dass es eine Herausforderung war, „dass man halt auch nicht zu viel vom Text verändert, dass dann der persönliche Stil wieder mit einfließt“ (T3: 3f). T2 fühlte sich durch die Vorgabe, nur so viel wie nötig zu editieren, besonders eingeschränkt: „Ich hätte da noch viel mehr editieren können. [...] Und das ist halt schwierig, da zu sagen wann ist es ein Fehler für dich, und wann ist es einfach etwas, das ich anders formulieren würde.“ (T2: 126ff). Auch T3 klagte über fehlendes Einschätzungsvermögen bezüglich der idealen Anzahl an Änderungen, „weil du denkst dir: Ist das jetzt zu viel, was ich geändert habe? Ist das zu wenig? Kann ich das lassen oder muss ich das doch ändern?“ (T3: 55f).

Bei der Frage nach der Vorgehensweise oder Strategie, die die Teilnehmer*innen beim Post-Editing angewandt haben, lassen sich auch verschiedene Ansätze beobachten. T1 gab beispielsweise an, zuerst Ausgangs-, dann Zieltext in chronologischer Abfolge Segment für Segment gelesen zu haben, dann überprüft zu haben, „ob das auch stimmig ist,

grammatikalisch und auch stilistisch“ (T1: 14f) und anschließend Änderungen vorgenommen zu haben. T2 und T3 wählten hingegen eine andere Vorgehensweise, da sie sich zuerst den maschinell übersetzten Zieltext in seiner Gesamtheit durchlasen, um zu sehen, „ob das eben sprachlich auf Deutsch jetzt so stimmig ist, dass es relativ natürlich klingt“ (T3: 19f). Anschließend lasen sie den Zieltext erneut, jedoch diesmal Segment für Segment und verglichen ihn mit dem Ausgangstext, vor allem an Stellen, an denen ihnen „etwas ganz komisch vorgekommen ist“ (T3: 21). Die Einschätzungen bezüglich der eigenen Arbeitsdauer und der durchschnittlichen Anzahl an Änderungen pro Segment, um die die Teilnehmer*innen gebeten wurden, wurden bereits in Tabelle 1 und Tabelle 3 aufgelistet und zeigten, dass nur T1 in der Lage war, die eigene Leistung präzise einzuschätzen. Die Teilnehmer*innen wurden an dieser Stelle erneut gefragt, ob sie nach Durchführen der Aufgabe Humanübersetzen oder Post-Editing als kognitiv aufwändiger einschätzen würden. T1 und T2 schätzten Humanübersetzen als mental anstrengender ein, da beim Post-Editing durch die vorgegebene Übersetzung eine gewisse Entlastung entsteht, „weil trotzdem einige Sätze dabei waren, die ich wirklich so unterschrieben hätte“ (T2: 66f). T3 bezeichnete beide Arbeitsweisen als gleichermaßen anstrengend, da „du zwar [beim Post-Editing] den vorgegebenen Text hast, aber dir dann eben Gedanken machen musst, wie sagt man das auf eine natürliche Art“ (T3: 45f). Alle Teilnehmer*innen waren sich jedoch einig, dass sie den vorliegenden Text mit Post-Editing deutlich schneller als mit Humanübersetzen fertigstellen konnten. Auch hier wurden die Teilnehmer*innen erneut mit der fiktiven Situation konfrontiert, dass sie den Auftrag erhalten, den soeben bearbeiteten Text entweder selbst zu übersetzen oder zu post-editieren, da dieser publiziert werden soll. Nur T1 entschied sich in dieser Situation für PE, T2 und T3 würden es bevorzugen den Text selbst zu übersetzen, „weil stilistisch schon ein paar Sachen drinnen waren, die mich echt gestört haben“ (T3: 65f) und „weil ich dann vielleicht auf andere Lösungsansätze kommen würde eventuell und einfach ein besseres Gewissen hätte dabei, einfach wenn ich weiß, dass ich das selber gemacht habe“ (T2: 77ff). T2 beklagte hier das Fehlen der eigenen PE-Kompetenz, die es auch ermöglichen würde, den Text mit einem geschulteren Blick für Fehlertypen zu editieren. T2 entschied sich auch gegen PE, „weil ich Angst hätte, dass ich irgendetwas übersehen habe“ (T2: 81).

Die Frage, ob sich die eigene Meinung gegenüber PE nach Durchführen der Aufgabe geändert hätte, verneinte T1, da er/sie nach wie vor positiv eingestellt war. Auch T2 blieb bei einer neutralen bis positiven Einstellung gegenüber PE, ebenso wie T3, der/die jedoch anmerkte, dass Post-Editing schwieriger und eine größere Herausforderung sei als ursprünglich gedacht. Die Teilnehmer*innen wurden auch gebeten, ihre Einschätzungen

bezüglich der zukünftigen Entwicklung von PE sowie von sämtlichen damit verbundenen Gefahren für den Übersetzungsberuf abzugeben. T1 blickt dieser Zukunft positiv entgegen, sieht Post-Editing als Chance und betont, „dass ich eigentlich ein Befürworter bin, dass Mensch und Maschine [...] zusammen harmonieren“ (T1: 70ff). T3 befürchtete, dass durch die Verbreitung von PE die Preise am Übersetzungsmarkt noch stärker gedrückt werden, als es ohnehin bereits der Fall ist, und auch, dass die Rolle von Übersetzer*innen dadurch an Wert verliert, da „Post-Editing uns [...] schlechter macht [...] als wir eigentlich sind“ (T3: 81). T2 befürchtete hingegen, dass Kund*innen mit der Verbreitung von PE die Tätigkeit von Übersetzer*innen weniger wertschätzen würden, da sie sich auch über die genauen Abläufe der Tätigkeit und die damit verbundenen Herausforderungen nicht im Klaren sind: „Leute denken [sich] ‚Du kriegst eh einen fertigen Text, der ist wahrscheinlich eh ur gut, bist eh gleich fertig damit‘, aber so ist es halt einfach nicht.“ (T2: 110ff). T2 erwähnte auch, dass trotz der Zusammenarbeit zwischen Mensch und Maschine immer noch der Mensch die volle Verantwortung für jede Übersetzung trägt, die er liefert, und somit auch der Mensch die dominante Rolle in dieser Beziehung übernimmt. Alle Teilnehmer*innen bestätigten abschließend, dass sie sich während ihrer Ausbildung spezifische Lehrveranstaltungen zu Post-Editing gewünscht hätten und es für sinnvoll erachten würden, solche in Zukunft am Zentrum für Translationswissenschaft einzuführen. T1 erwähnte, dass während des Studiums Post-Editing als Teil mancher Übersetzungskurse behandelt wurde, jedoch fehlte die Vermittlung theoretischer Grundlagen. T3 erwähnte: „Ich habe die Entwicklung auch in unserem Studienzeitraum bemerkt, eben auch als *DeepL* aufgekommen ist [...], dass das immer prominenter wird in unserer Berufssparte.“ (T3: 90ff). Eine Post-Editing-Ausbildung zu erhalten, würde laut T3 nicht nur dabei helfen, diese Arbeitsweise effizienter anwenden zu können, sondern auch die eigene Tätigkeit gegenüber Kund*innen zu rechtfertigen: „Wenn ich die Maschine nachbearbeite, kostet es trotzdem mehr, weil es eben qualitativ hochwertiger ist.“ (T3: 100f). Laut T2 führt an einer Einführung von PE-Kursen in der Übersetzungsausbildung kein Weg vorbei, da es „ein immer größer werdender Teil wird von dem, was wir mal machen müssen“ (T2: 92ff). T2 sieht auch ausreichend Möglichkeiten, einzelne PE-Module in bereits existierende Übersetzungskurse zu integrieren. Diese sollten Studierenden die wichtigsten Strategien vermitteln, damit „man sich dann auch sicher sein kann, dass man sich im Text die richtigen Sachen anschaut, die die halt wichtig sind und sich nicht an Kleinigkeiten aufhängt“ (T2: 123ff). Abschließend zieht T2 folgendes Fazit: „Die Uni Wien braucht unbedingt einen Post-Editing Kurs im Master Fachübersetzen.“ (T2: 134f)

6 Diskussion

Obwohl alle Teilnehmer*innen eine recht ähnliche Zeitdauer zwischen 46 und 51 Minuten für die Post-Editing-Aufgabe benötigten, unterscheidet sich die Bearbeitungszeit der Segmente in *MemoQ* vor allem zwischen T1 und T2 stark, was darauf schließen lässt, dass hier gänzlich unterschiedliche Strategien gewählt wurden. Die Segmente 6, 10 und 16 wurden hingegen von allen Teilnehmer*innen auffällig lange bearbeitet und geben ein erstes Indiz dafür, dass diese Segmente besonders schwierig zu editieren waren. Auch wenn die Teilnehmer*innen im Interview nach der PE-Aufgabe alle angaben, dass PE schneller als Humanübersetzen sei, so überschätzten T2 und T3 ihre eigene Arbeitszeit deutlich. T1 konnte die tatsächlich benötigte Zeit wohl auf Grund der eigenen Vorerfahrung mit PE besser einschätzen. Die Teilnehmer*innen erzielten mit durchschnittlich 440 Wörtern pro Stunde eine äußerst gute Produktivität, die mit jener professioneller Post-Editor*innen vergleichbar ist (Koponen 2016: 134; Robert 2013: 32f). Nachdem die editierten Texte jedoch nicht hinsichtlich ihrer Qualität evaluiert wurden, ist es möglich, dass die Teilnehmer*innen zwar produktiv waren, aber Texte von schlechter Qualität erzeugt haben. Auch wenn die Übersetzungskompetenz der Teilnehmer*innen auf Grund ihrer Ausbildung mindestens als gut eingeschätzt werden kann, wurde im Theorieteil dieser Arbeit bereits erwähnt, dass Studierende, die gute Übersetzungen anfertigen, nicht automatisch gute Post-Editor*innen sind und in vielen Fällen sogar Zieltexte von belegbar schlechter Qualität erzeugen.

Bezüglich der durchschnittlichen Anzahl an vorgenommenen Änderungen pro Segment gab T1, der/die bereits die Arbeitszeit korrekt verorten konnte, auch hier die genaueste Einschätzung der eigenen Leistung ab. T2 und T3 überschätzten die tatsächliche Anzahl, was jedoch auf die starken Schwankungen pro Segment zurückzuführen ist. Wie in den Interviews erwähnt wurde, lieferte die Übersetzung von *DeepL* einige Segmente, die ohne jegliche Änderung übernommen werden konnten, was auch für die Qualität dieses MÜ-Systems spricht. Auch wenn die drei Segmente mit den meisten Änderungen mit durchschnittlich 35 Wörtern auch die längsten waren, so kann nicht pauschal auf eine höhere Anzahl an Änderungen in längeren Segmenten geschlossen werden, da die vier Segmente, in denen von T2 und T3 keine Änderungen vorgenommen wurden, mit durchschnittlich 22 Wörtern auch relativ lang waren. Die Tatsache, dass T2 nur in der Hälfte aller Segmente Änderungen vornahm, deckt sich auch mit der kurzen Bearbeitungszeit, die T2 für den gesamten Text aufwandte. T1 verbrachte hingegen die längste Dauer mit dem Bearbeiten und nahm die meisten Änderungen vor, was auch durch die Metriken bestätigt wird. Für die Anwendung im Bereich Post-Editing ist voraussichtlich die Translation Edit Rate (TER) am

aussagekräftigsten, da sie nach Moorkens (2015) nicht nur auf den technischen Aufwand, also auf die vorgenommenen Änderungen, schließen lässt, sondern nach Lacruz und Shreve (2014) und O'Brien (2011) auch ein Indikator für den kognitiven Aufwand ist, da ein höherer TER-Wert in einer langsameren kognitiven Verarbeitungsgeschwindigkeit resultiert.

Die Einschätzungen, die die Teilnehmer*innen vor Durchführen der PE-Aufgabe bezüglich der Qualität der maschinellen Übersetzung durch *DeepL* abgaben, decken sich auch mit den Angaben, die sie diesbezüglich nach der Aufgabe machten. Die Teilnehmer*innen erwarteten sich eine gute Qualität und wurden in dieser Annahme auch bestätigt, sie scheinen somit gut mit dem maschinellen Übersetzungssystem vertraut zu sein und realistische Erwartungen zu haben, da sie auch angaben, *DeepL* regelmäßig im privaten Alltag zu nutzen. Die von den Teilnehmer*innen geäußerten Bedenken bezüglich Lexik und Stil der MÜ waren ebenso gerechtfertigt, da sie diese beiden Aspekte im Nachhinein nur als überwiegend mäßig beurteilten. Obwohl die Teilnehmer*innen in den nachträglichen Interviews alle erwähnten, sich durch die vorgegebene Satzstruktur der MÜ in der eigenen Arbeitsweise behindert zu fühlen und Schwierigkeiten dabei hatten, alternative Formulierungen zu finden, so bewerteten alle Teilnehmer*innen die Syntax als eher gut. Dies zeigt, dass selbst eine gute maschinelle Übersetzung zu Konflikten mit dem persönlichen Stil und den Vorlieben der Post-Editor*innen führen kann. Generell bestätigten die Teilnehmer*innen, dass die maschinelle Übersetzung dieses Textes im Großen und Ganzen brauchbar war, und es nur wenige problematische Stellen beim Editieren gab, diese jedoch umso frustrierender und zeitaufwändiger waren. Alle Teilnehmer*innen gaben an, mit dem finalen Text nicht gänzlich zufrieden zu sein. Sie äußerten auch den Wunsch, an manchen Stellen die MÜ zu löschen und lieber selbst übersetzt zu haben. Es ist denkbar, dass die Teilnehmer*innen mit einer hybriden Form von Post-Editing, bei der für jedes Segment MÜ-Vorschläge herangezogen werden können, bei Belieben jedoch auch selbst übersetzt werden kann, besser hätten arbeiten können. Vor Durchführen der Aufgabe waren T1 und T3 bereits davon überzeugt, dass PE schneller und weniger kognitiv aufwändig als Humanübersetzen ist, T2 behauptete das Gegenteil. Nach Durchführen der Aufgabe änderte jedoch T2 die eigene Meinung und erkannte ebenfalls den möglichen Produktivitätszuwachs durch PE und schätzte es als weniger kognitiv aufwändig ein und das, obwohl T2 im nachträglichen Interview mehrmals angab, beim Bearbeiten gewisser Segmente frustriert gewesen zu sein.

Verteilung und Anzahl der Fixationen auf die vier Interessensbereiche zeigen, dass vor allem T1 und T2 unterschiedliche Arbeitsweisen und Strategien wählten. Da T1 die meiste Zeit mit Betrachten von Ausgangs- und Zieltext verbrachte, die meisten Änderungen vornahm

und den Browser verhältnismäßig wenig nutzte, kann davon ausgegangen werden, dass T1 nicht nur mit dem Thema des Textes vertraut war, sondern auch, dass T1 größeres Vertrauen in die Qualität der maschinellen Übersetzung sowie in die eigene PE-Vorerfahrung hatte als T2 und T3. T2 verbrachte hingegen deutlich mehr Zeit im Browser und nahm insgesamt die wenigsten Änderungen vor. Laut eigener Aussage hatte T2 zwar Vertrauen in die Qualität und Korrektheit der maschinellen Übersetzung, jedoch gab T2 an, auf Grund der eigenen fehlenden PE-Vorerfahrung Angst zu haben, Fehler im Text zu übersehen und verbrachte daher umso mehr Zeit mit der Recherche einzelner Wörter und Phrasen. Die Aufteilung der Fixationen von T3 wirkt im Vergleich zu den anderen Teilnehmer*innen ausgewogen und lässt keine nennenswerten Auffälligkeiten erkennen. Interessant ist auch, dass der Bereich mit den Translation-Memory Treffern, in dem die Teilnehmer*innen den maschinellen Übersetzungsvorschlag in seiner Originalform einsehen konnten, selbst wenn sie diesen in der Zieltextspalte bereits überschrieben hatten, von allen Teilnehmer*innen so gut wie nicht betrachtet wurde. Auch wenn die Teilnehmer*innen über die Funktion dieses Bereichs unmittelbar vor Durchführen der Aufgabe aufgeklärt wurden, ist es denkbar, dass sie diesen trotz ihrer Vertrautheit mit *MemoQ* zuvor nie benutzt und während der Aufgabe auch darauf vergessen hatten. Es ist aber ebenso denkbar, dass die Teilnehmer*innen sehr wohl über den Bereich Bescheid wussten, ihn jedoch bewusst nicht betrachteten, da er von keiner Relevanz für sie war. Die maschinelle Übersetzung in der Zieltextspalte wurde von den Teilnehmer*innen wahrscheinlich nur als Vorschlag gesehen, der nach Belieben verändert werden kann, bis er die Form eines akzeptablen Satzes angenommen hat. Dabei ist es nicht wichtig, den selbst editierten Satz mit der ursprünglichen, maschinell übersetzten Version zu vergleichen, sondern viel eher diesen mit dem Ausgangstext zu vergleichen und schrittweise anzupassen.

Auch wenn aus den Interviews hervorging, dass die Teilnehmer*innen unterschiedliche Vorgehensweisen beim Editieren anwendeten, etwa die Reihenfolge, in der Ausgangs- und Zieltext gelesen wurden, und auch wenn gezeigt wurde, dass die Teilnehmer*innen unterschiedlich viele Änderungen vornahmen, so fielen bei allen eindeutig die meisten Fixationen in den Bereich des Zieltextes. Nicht nur wurden in diesem Bereich die Änderungen am Text vorgenommen, sondern auch die meiste Denkarbeit geleistet. Da die durchschnittliche Fixationsdauer in diesem Bereich bei allen Teilnehmer*innen um rund 42% länger war als in anderen Bereichen, kann dies nach Moorkens (2015) als eindeutiges Indiz für einen höheren kognitiven Aufwand in diesem Bereich gesehen werden und gilt innerhalb der durch den Eyetracker erfassten Daten gemeinsam mit der Anzahl an Fixationen pro

Interessensbereich als bester Indikator für kognitiven Aufwand bei Post-Editing. Die durchschnittliche Pupillengröße pro Fixation lieferte hingegen kein Indiz für einen höheren kognitiven Aufwand im Bereich des Zieltextes, jedoch wurde ein Zusammenhang zwischen den beiden Variablen auch nie belegt. Interessant ist jedoch, dass die Pupillen aller Teilnehmer*innen bei Betrachten des Browsers um rund 25% größer als in den anderen Bereichen waren. Nachdem angenommen wird, dass sich die Pupillen als Reaktion auf einen unerwarteten visuellen Reiz erweitern, könnte der große Durchmesser im Bereich des Browsers folgendermaßen erklärt werden. Die drei Bereiche Ausgangstext, Zieltext und TM befanden sich im *MemoQ*-Fenster, welches mehr oder wenig statisch war und über die gesamte Arbeitsdauer dasselbe Layout und Erscheinungsbild hatte. Der Internetbrowser war hingegen ein dynamischer Bereich, da die Teilnehmer*innen stets zwischen Internetseiten hin und her wechselten oder neue öffneten, wobei jede Seite bezüglich ihrer Gestaltung und ihres Layouts anders war. Folglich mussten die Teilnehmer*innen die stets neu auftretenden, unerwarteten visuellen Reize verarbeiten, was als kognitiv anstrengender gesehen werden kann als das Betrachten eines statischen, unveränderten Reizes. In Bezug auf den kognitiven Aufwand beim Post-Editing scheint die Pupillengröße jedoch keine aussagekräftige Variable zu sein.

Somit lässt sich zusammenfassen, dass Studierende ohne einschlägige Post-Editing-Ausbildung gänzlich unterschiedliche Strategien und Vorgehensweisen wählen, jedoch ähnlich viele Änderungen in denselben Segmenten vornehmen und ihre visuelle Aufmerksamkeit primär auf den Zieltext konzentrieren. Die stärksten Unterschiede in den Messergebnissen konnten zwischen T1, der/die über etwas PE-Vorerfahrung verfügte und von den Vorteilen der Arbeitsweise überzeugt war, und T2, der/die keinerlei Vorerfahrung hatte und der Arbeitsweise anfangs skeptisch gegenüberstand, festgestellt werden. Da T1 bereits während des Studiums vereinzelt laienhaft post-editierte, hatte er/sie eine eigene Strategie entwickelt und auch weniger Hemmungen, Änderungen vorzunehmen, wonach T1 doppelt so viel editierte wie T2. T2 gab hingegen an, deutlich mehr Hemmungen gehabt zu haben, Änderungen vorzunehmen und zu viel zu editieren sowie zu befürchten, Fehler in der maschinellen Übersetzung zu übersehen. Dies könnte auch der Grund sein, warum T2 so viel mehr Zeit im Browser als die anderen Teilnehmer*innen verbrachte, da er/sie einen größeren Teil der maschinellen Übersetzung auf Fehler überprüfte und dabei auch mehrere Durchgänge machte. Da eine längere Dauer für die Recherche aufgebracht wurde, nahm T2 insgesamt auch die wenigsten Änderungen am Text vor. Zusätzlich erwähnte T2 unsicher gewesen zu sein, was im Rahmen eines Post-Editing als Fehler zu klassifizieren sei, weshalb T2 lieber zu

wenig als zu viel editierte. Bei T1 konnte insgesamt eine starke Bevorzugung von PE gegenüber dem Humanübersetzen festgestellt werden, T2 und T3 erkannten hingegen sämtliche Vorteile von PE an und betonten dessen Wichtigkeit für die Zukunft des Übersetzungsberufes, schlussendlich bevorzugten T2 und T3 dennoch das Humanübersetzen, da für sie der damit verbundene Spaß und die Selbsterfüllung deutlich wichtiger als jegliche Produktivitätsgewinne wären.

Abschließend sollen die wichtigsten Ergebnisse diese Studie mit jenen von de Faria Pires (2020) und Mesa-Lao (2014) verglichen werden, die ein ähnliches Forschungsdesign aufweisen und in Tabelle 10 dargestellt sind. Bei de Faria Pires hatten vor Durchführen des Post-Editing 53% der Teilnehmer*innen eine positive, 30% eine neutrale und 17% eine negative Einstellung gegenüber PE. In dieser Studie äußerten 67% eine neutrale und 33% eine positive Einstellung. Nach Durchführen der Aufgabe verdoppelte sich die Anzahl an Teilnehmer*innen mit negativer Einstellung bei de Faria Pires, im Falle dieser Studie änderte hingegen niemand seine/ihre Einstellung. Bei de Faria Pires waren vor Durchführen der Aufgabe 54%, danach 77% der Teilnehmer*innen überzeugt, dass Post-Editing ihre Produktivität steigern kann, und auch in dieser Studie waren zuerst 67% und nachher 100% der Teilnehmer*innen dieser Ansicht.

	Diese Studie	*de Faria Pires (2020)*	*Mesa-Lao (2014)*
AT	19%	15%	30%
ZT	60%	38%	47%

Tabelle 10: Durchschnittliche Verweildauer in den Bereichen Ausgangs- und Zieltext

Die durchschnittliche Verweildauer der Teilnehmer*innen in Ausgangs- und Zieltext in Tabelle 10 zeigt keine klaren Übereinstimmungen mit den anderen Studien, lässt jedoch erkennen, dass auch hier der Zieltext deutlich länger betrachtet wurde, bei Mesa Lao 1,5-mal, bei de Faria Pires 2,5-mal und in dieser Studie sogar 3-mal länger als der Ausgangstext.

	Diese Studie	*de Faria Pires (2020)*	*Mesa-Lao (2014)*
AT	24% [1368]	29% [1422]	33% [673]
ZT	76% [4262]	71% [3500]	67% [1220]

Tabelle 11: Durchschnittliche Aufteilung und Anzahl von Fixationen auf Ausgangs- und Zieltext

Die durchschnittliche Aufteilung der Fixationen auf Ausgangs- und Zieltext in Tabelle 11 deckt sich hingegen mit den anderen beiden Studien, auch wenn diese auf Grund der kürzeren Textlänge eine geringere Anzahl an durchschnittlichen Fixationen aufweisen. Bei Mesa-Lao

entfielen 2-mal, bei de Faria Pires 2,5-mal und in dieser Studie sogar 3-mal so viele Fixationen auf den Zieltext wie auf den Ausgangstext.

Auch wenn durch diese Studie ähnliche Ergebnisse erzielt werden konnten und einige äußerst aufschlussreiche Beobachtungen bezüglich Präferenzen bei der Arbeitsweise und Wahl der Strategien der Teilnehmer*innen gemacht werden konnten, so ist die Aussagekraft dieser Studie dennoch limitiert. Zum einen ist die Teilnehmer*innenanzahl von drei Personen zu niedrig, um die Ergebnisse als repräsentativ für die breite Masse an Studierenden ohne Post-Editing-Ausbildung zu sehen. Weiters wurde von allen Teilnehmer*innen derselbe Text bearbeitet, auch hier würde eine größere Variation und Anzahl an Texten eine bessere Aussagekraft bewirken. Auch der Produktivitätszuwachs, den Studierende durch PE im Vergleich zu Humanübersetzen erzielen können, wurde aus Zeitgründen nicht erfasst, da die Teilnehmer*innen im Rahmen der Studie keinen Text erhielten, den sie selbst übersetzen mussten. Weiters wurde die Qualität der editierten Zieltexte nicht evaluiert. Es soll auch erwähnt werden, dass die Leistung der Teilnehmer*innen im Rahmen einer simulierten Situation unter Laborbedingungen erfasst wurde und somit gewissermaßen verfälscht ist. Zum einen waren sich die Teilnehmer*innen der Präsenz des Eyetracker bewusst und mussten auf Grund der feinen Kalibrierung des Geräts durch den Forschungsleiter während der Aufgabe mehrere Male gebeten werden, die Sitzposition anzupassen, um von der Kamera wieder korrekt erfasst zu werden. Außerdem erwähnten die Teilnehmer*innen im Interview, dass sie das Post-Editing in einer echten Situation nicht wie bei der Studie in einer Sitzung durchgeführt hätten, sondern den Text in mehreren Durchgängen, eventuell auch über mehrere Tage hinweg, bearbeitet hätten, wodurch sie auch mehr Zeit für die Lösung schwieriger Segmente gehabt hätten. Auch wenn das durch die Teilnehmer*innen durchgeführte Post-Editing in seiner Aussagekraft durch einige Faktoren limitiert ist, so konnten dennoch äußerst interessante Unterschiede in der Wahl der Strategie sowie Gemeinsamkeiten in Bezug auf die schwierigsten Segmente und die Verteilung der visuellen Aufmerksamkeit als Indikator für kognitiven Aufwand beim Post-Editing aufgezeigt werden, die sich auch mit den Ergebnissen vergleichbarer Studien decken.

7 Fazit

Mit der aktuellen Leistungsfähigkeit von NMÜ-Systemen und ihrer breiten Verfügbarkeit für viele Nutzer*innengruppen ist davon auszugehen, dass diese auch in den kommenden Jahren immer mehr an Bedeutung gewinnen und auch Übersetzer*innen dabei helfen werden, den stets wachsenden Arbeitsaufwand zu bewältigen. In der europäischen Sprachindustrie macht NMÜ bereits 30% des Marktanteils für Übersetzungen aus, und auch immer mehr große Sprachdienstleister*innen setzen auf die Weiterentwicklung spezialisierter, leistungsstarker Systeme. Doch auch frei verfügbare Systeme wie *DeepL* sind vor allem unter Studierenden beliebt und immer häufiger ein Bestandteil von translatorischen Ausbildungen. Studierende sind durch die regelmäßige Verwendung auch schon mit den typischen Fehlern, die neuronale Systeme beim Übersetzen produzieren können, vertraut und verfügen dadurch bereits über eine wichtige Post-Editing-Kompetenz. Mit der steigenden Leistungsfähigkeit der Systeme erwarten sich Kund*innen jedoch auch schnellere Auftragsabwicklungen und niedrigere Preise von den Übersetzer*innen, was sich insgesamt negativ auf die Übersetzungsbranche auswirkt. Übersetzer*innen sehen sich damit konfrontiert, sich diesen neuen Entwicklungen anzupassen und etwa Post-Editing in ihr Kompetenzportfolio aufzunehmen, um auch Kund*innen über dessen Besonderheiten aufklären und die eigene Dienstleistung preislich rechtfertigen zu können. Außerdem dürfen die Limitationen von NMÜ-Systeme nicht übersehen werden, wie etwa fehlende Parallelkorpora in seltenen Sprachkombinationen, oder die für das Training der Systeme aufwändigen Hardwareanforderungen.

Als Begleiterscheinung der maschinellen Übersetzung hat auch Post-Editing in den vergangenen 30 Jahren an Bedeutung gewonnen und konnte mit Einführung von Normen wie der *ISO 1858* oder den *TAUS Post-Editing Richtlinien* einheitlich durchgeführt werden, da sich daraus die zwei Arten des leichten und vollständigen Post-Editings etablierten. Das für Post-Editing benötigte Kompetenzprofil baut zu einem großen Teil auf jenem für Übersetzen auf, erweitert es jedoch um einige Faktoren, weshalb auch eine spezifische Ausbildung an translationswissenschaftlichen Hochschulen sinnvoll wäre. Professionelle Übersetzer*innen, die eine solche Ausbildung nicht erhalten haben, erkennen zwar die Wichtigkeit und die Vorteile der Arbeitsweise, sind PE gegenüber jedoch skeptisch und bevorzugen in vielen Fällen das Humanübersetzen. Es konnte jedoch aufgezeigt werden, dass diese Einstellung meist auf die Vertrautheit mit der Tätigkeit zurückzuführen ist, da Übersetzer*innen mit spezifischer Ausbildung oder Vorerfahrung Post-Editing deutlich effektiver nutzen können. Übersetzer*innen, die die Vorteile aus erster Hand erleben konnten, sind PE demnach viel aufgeschlossener als jene ohne Vorerfahrung, die durch die neuartige und unvertraute

Arbeitsweise eventuell frustriert wurden. Es konnte durch mehrere Studien bewiesen werden, dass sowohl angehende als auch professionelle Übersetzer*innen durch Post-Editing einen deutlichen Produktivitätszuwachs im Vergleich zum Humanübersetzen erzielen können. Weiters zeigten einige Studien, dass das Messen des Aufwands beim Post-Editing anhand des Individuums mittels Eyetracking einen spannenden und noch recht jungen Forschungsbereich darstellt, dessen Ergebnisse für die weitere Entwicklung und Optimierung von PE-Arbeitsabläufen und Ausbildungen maßgeblich sein könnten. Auch für den empirischen Teil dieser Arbeit stellte sich Eyetracking als besonders aufschlussreiche Erhebungsmethode dar, die es erlaubt, die Forschungsfrage bezüglich des Aspekts, der beim Post-Editing mit dem höchsten temporalen, technischen und kognitiven Aufwand verbunden ist, zu beantworten. So konnte durch die verschiedenen Variablen bewiesen werden, dass das Editieren des Zieltextes insgesamt mit dem höchsten Aufwand verbunden ist. Nicht nur war die Verweildauer der Studierenden in diesem Bereich deutlich länger als bei Ausgangstext oder Browser, sondern es war durch die vielen vorgenommenen Änderungen, die auch durch die diversen Metriken belegt wurden, ein hoher technischer Aufwand zu verzeichnen. Auch die Daten der Eyetracking-Aufnahme bestätigten eine überwiegende Verteilung der Fixationen auf den Zieltext, der im Schnitt dreimal so häufig wie der Ausgangstext betrachtet wurde. Zudem ist die 42% längere durchschnittliche Fixationsdauer im Zieltextbereich als eindeutiges Indiz für einen hohen kognitiven Aufwand zu verstehen.

In Bezug auf die in der Einleitung gestellten Unterfragen kann durch die im Rahmen der Studie geführten Interviews festgestellt werden, dass die teilnehmenden Studierenden dem Post-Editing relativ offen, teils sogar äußerst positiv gegenüberstehen. Spätestens nach Durchführen der PE-Aufgabe waren alle drei von dem möglichen Produktivitätszuwachs überzeugt und gaben trotz ihrer fehlenden Ausbildung an, dass Post-Editing weniger kognitiv anstrengend als Humanübersetzen sei. Weiters konnte durch die Studie festgestellt werden, dass die Studierenden durch ihre fehlende PE-Ausbildung unterschiedliche Strategien beim Editieren des Textes anwendeten. Während manche wie bei einer Übersetzung zuerst die Ausgangs- und dann die Zieltextsegmente in chronologischer Abfolge lasen und bearbeiteten, konzentrierten sich andere zuerst auf den gesamten maschinell übersetzten Zieltext, bevor sie diesen mit dem Ausgangstext verglichen. Zuletzt konnte noch festgestellt werden, dass die Studierenden sich während ihrer Ausbildung zumindest einige Kurse gewünscht hätten, die ihnen die theoretischen und praktischen Grundlagen des Post-Editing vermittelt hätten, und sie die Einführung derartiger Kurse in die Curricula von translationswissenschaftlichen Hochschulen in Zukunft begrüßen würden. Die zu Beginn dieser Arbeit aufgestellten

Annahmen, dass die Studierenden den Großteil ihrer Fixationen auf den Zieltext richten würden, sowie dass sie durch Post-Editing ein hohes Maß an Produktivität erzielen könnten, ließen sich auf Grund der geringen Teilnehmer*innenanzahl nicht eindeutig bestätigen, jedoch konnten klare Tendenzen in diese Richtung beobachtet werden. Auf die Fixationen im Zieltext wurde bereits bei Beantwortung der Forschungsfrage eingegangen. Die von den Studierenden erzielte Produktivität, gemessen anhand der Wörter pro Minute und Stunde, fiel überraschend hoch aus und war sogar mit jenem Wert vergleichbar, der von professionellen Post-Editor*innen zu erwarten ist. Für die dritte und letzte Annahme, dass Studierende übermäßig viele Änderungen am Text vornehmen würden, konnten in dieser Studie keine Tendenzen festgestellt werden. Nur ein/e Teilnehmer*in nahm Änderungen in allen Segmenten des Textes vor, wohingegen die anderen beiden im Durchschnitt nur die Hälfte aller Segmente bearbeiteten. Darüber hinaus gaben diese beiden Teilnehmer*innen in den Interviews an, diesbezüglich eher zögerlich vorgegangen zu sein, da sie befürchteten sonst Änderungen in korrekten Segmenten vorzunehmen und dadurch Zeit zu vergeuden.

Wie durch viele Studien bereits belegt wurde und auch durch diese Studie bestätigt werden konnte, haben Studierende eine überwiegend positive Einstellung gegenüber Post-Editing. Sie sehen diese neue Disziplin als einen wichtigen Bestandteil ihres künftigen Kompetenzportfolios, auch wenn sie Bedenken bezüglich des Verlustes von Arbeitsplätzen durch Automatisierung und einer Abwertung des Übersetzerberufes haben. Weiters bringen Studierende durch ihre Übersetzungsausbildung bereits die Basis für eine darauf aufbauende PE-Ausbildung mit, auch wenn derartige Kurse und Module mit aktuellem Stand in Europa nur an manchen Hochschulen angeboten werden. Eine praxisorientierte Ausbildung, die sich an den aktuellen Erkenntnissen der PE-Forschung orientiert, könnte Studierende auf diese neue Disziplin vorbereiten, wie es auch bereits seit 2017 von der EMT Expert*innengruppe gefordert wird. Weiters könnten diese Erkenntnisse in die Entwicklung eigener PE-Tools oder die Verbesserung von Schnittstellen in CAT-Tools mit MÜ einfließen, wodurch auch mehr professionelle Übersetzer*innen von den Vorteilen und der Sinnhaftigkeit eines leicht zu bedienenden PE Tools überzeugt werden können. In den kommenden Jahren und Jahrzehnten kann anhand der in dieser Arbeit gewonnenen Erkenntnisse davon ausgegangen werden, dass der Übersetzerberuf durch eine Kooperation von Mensch und Maschine geprägt sein wird, da beide durch die Anforderungen des Marktes und die technischen Limitationen nach wie vor aufeinander angewiesen sein werden, und diese *Translator-Computer-Interaction* auch weiterhin ein unabdingbarer Faktor in der Sprachindustrie des 21. Jahrhunderts bleiben wird.

Bibliographie

Agarwal, Abhaya & Lavie, Alon (2008). METEOR, M-BLEU and M-TER: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output. In: Callison-Burch, Chris; Koehn, Phillip & Monz, Christof (Hg.) *Proceedings of the Third Workshop on Statistical Machine Translation*. Columbus: ACL, 115-118.

ALPAC (1966). *Languages and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council.* Washington, D.C.: National Academy of Sciences. In: https://www.nap.edu/read/9547/chapter/1 (Stand: 28.09.2021)

Aranberri, Nora (2017). What Do Professional Translators Do When Post-Editing for the First Time? First Insight into the Spanish-Basque Language Pair. *HERMES* 56, 89-110.

Aranberri, Nora; Labaka, Gorka; de Ilarraza, Arantza Diaz & Sarasola, Kepa (2014). Comparison of post-editing productivity between professional translators and lay users. In: O'Brien, Sharon; Simard, Michel & Specia, Lucia (Hg.) *Proceedings of the Third Workshop on Post-editing Technology and Practice (WPTP-3).* Vancouver: AMTA, 20-34.

Alves, Fabio; Szpak, Karina Sarto; Goncalves, José Luiz; Sekino, Kyoko; Aquino, Marceli; Araujo e Castro, Aquino; Koglin, Arlene; de Lima Fonseca, Norma B. & Mesa-Lao, Bartolomé (2016). Investigating cognitive effort in post-editing: A relevance-theoretical approach. In: Grucza, Sambor & Hansen-Schirra, Silvia (Hg.) *Eyetracking and Applied Linguistics.* Berlin: Language Science Press, 109-141.

Aziz, Wilker & Specia, Lucia (2012). PET: a Tool for Post-editing and Assessing Machine Translation. In: Cettolo, Mauro; Federico, Marcello; Specia, Lucia & Way, Andy (Hg.) *Proceedings of the 16th EAMT Conference.* Trentino: EAMT, 99-100.

Banerjee, Satanjeev & Lavie, Alon (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Goldstein, Jade; Lavie, Alon & Lin, Chin-Yew (Hg.) *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization*. Ann Arbor: ACL, 65–72.

Bar-Hillel, Yehoschua (1959). *Report on the State of Machine Translation in the United States and Great Britain.* Jerusalem: Hebrew University.

Bruhn, Manfred (2019). *Qualitätsmanagement für Dienstleistungen: Handbuch für ein erfolgreiches Qualitätsmanagement. Grundlagen – Konzepte – Methoden*. 11. Auflage Berlin/Heidelberg: Springer.

Bundgaard, Kristine (2017). Translators Attitudes Towards Translator-Computer Interaction – Findings from a Workplace Study. In: *HERMES* 56, 125-144.

Cadwell, Patrick; Castilho, Sheila; O'Brien, Sharon & Mitchell, Linda (2016). Human factors in machine translation and post-editing among institutional translators. *Translation Spaces* 5 (2). Amsterdam: John Benjamins, 222–243.

Cadwell, Patrick; O'Brien, Sharon & Teixeira, Carlos S. C. (2018). Resistance and accommodation: factors for the (non-) adop-tion of machine translation among professional translators. *Perspectives: Studies in Translatology* 26 (3), 301–21.

Carl, Michael; Dragsted, Barbara; Elming, Jakob; Hardt, Daniel & Jakobsen, Arnt Lykke (2011). The process of post-editing: A pilot study. In: Bernadette Sharp (Hg.) *Proceedings of the 8th International NLPSC workshop. Special theme: Human-machine interaction in translation.* Frederiksberg: Samfundslitteratur, 131–142.

Castilho, Sheila; Doherty, Stephen; Gaspari, Federico & Moorkens, Joss (2018). Approaches to human and machine translation quality assessment. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) *Translation Quality Assessment. From Principles to Practice.* Cham: Springer, 9-38.

Chatzikoumi, Eirini (2019). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, 26 (2), 137-161.

Cid, Clara Ginovart (2020a). Language industry views on the profile of the post-editor. *Translation Spaces* 9 (2), 283-313.

Cid, Clara Ginovart (2020b). The professional profile of a post-editor according to LSCs and linguists: A survey-based research. *HERMES* 60, 171-190.

Cid, Clara Ginovart & Ventura, Carme Colominas (2020). The MT post-editing skill set. In: Koponen, Maarit; Mossop, Brian; S. Robert, Isabelle & Scocchera, Giovanna (Hg.) *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes*. London: Routledge, 226-246.

Clark, John D. (1972). *Ignition! An Informal History of Liquid Rocket Propellants*. New Brunswick: Rutgers University Press.

Canfora, Carmen & Ottmann, Angelika (2020). Risks in neural machine translation. *Translation Spaces* 9 (1), 58-77.

Daems, Joke; Vandepitte, Sonia; Hartsuiker, Robert J. & Macken, Lieve (2017). Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators. *Meta: Journal des traducteurs* 62 (2), 245-270.

de Faria Pires, Loïc (2020). Master's students' post-editing perception and strategies. *FORUM* 18, 26-44.

Depraetere, Ilse ; De Sutter, Nathalie & Tezcan, Arda (2014). Post-Edited Quality, Post-Editing Behaviour and Human Evaluation: A Case Study. In: O'Brien, Sharon; Balling, Laura Winther; Carl, Michael; Simard, Michel & Specia, Lucia (Hg.) *Post-Editing of Machine Translation: Processes and Applications.* Newcastle-upon-Tyne: Cambridge Scholars Publisher, 78-109.

do Carmo, Félix & Moorkens, Joss (2020). Differentiating editing, post-editing and revision. In: Koponen, Maarit; Mossop, Brian, Robert; S. Robert, Isabelle & Scocchera Giovanna (Hg.) *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes.* London: Routledge, 35-49.

Doherty, Stephen & Kenny, Dorothy (2014). The design and evaluation of a statistical machine translation syllabus for translation students. *The Interpreter and Translator Trainer* 8 (2), 295-315.

Dorr, Bonnie & Snover, Matthew (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In: Dorr, Bonnie; Snover, Matthew; Schwartz, Rich; Micciulla, Linnea & Makhoul, John (Hg.) *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers.* Cambridge: AMTA, 223-231.

Duchowski, Andrew T. (2017). *Eye Tracking Methodology: Theory and Practice*. 3. Auflage Cham: Springer.

Edmundson, Harold P. & Hays, David G. (1958). Research methodology for machine translation. *Mechanical Translation* 5 (1), 8-15.

EMT (2017). *Competence framework 2017*. Brüssel: Europäische Komission.

Fernández, Gerardo; Biondi, Juan; Castro, Silvia & Agamennoni, Osvaldo E. (2016). Pupil size behavior during online processing of sentences. *Journal of Integrative Neuroscience* 15 (4): 1-12.

Forcada, Mikel L. (2010). Machine translation today. In: Gambier, Yves & van Doorslaer, Luc (Hg.) *Handbook of Translation Studies. Volume 1*. Amsterdam/Philadelphia: John Benjamins, 215-223.

Forcada, Mikel L. (2017). Making sense of neural machine translation. *Translation Spaces*, 6 (2), 291-309.

Forcada, Mikel L. & Neco, Ramón (1997). Asynchronous translations with recurrent neural nets. In: *Proceedings of International Conference on Neural Networks Vol. 4*, 2535-2540.

Gaspari, Federico; Toral, Antonio; Naskar, Sudip Kumar; Groves, Declan & Way, Andy (2014). Perception vs Reality: Measuring Machine Translation Post-Editing Productivity. In: O'Brien, Sharon; Simard, Michel & Specia, Lucia (Hg.) *Proceedings*

of the Third Workshop on Post-editing Technology and Practice (WPTP-3). Vancouver: AMTA, 60-73.

Gaspari, Federico; Almaghout, Hala & Doherty, Stephen (2015). A survey of machine translation competences: insights for translation technology educators and practitioners. *Perspectives: Studies in Translatology*, 1–26.

Gläser, Jochen & Laudel, Grit (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen.* 4. Auflage, Wiesbaden: VS.

Guerberof Arenas, Ana (2013). What do professional translators think about post-editing? *Journal of Specialised Translation* 31, 75-95.

Guerberof Arenas, Ana (2014a). Correlations between productivity and quality when post-editing in a professional context. *Machine Translation* 28 (3-4), 165–186.

Guerberof Arenas, Ana (2014b). The Role of Professional Experience in Post-editing from a Quality and Productivity Perspective. In: O'Brien, Sharon; Balling, Laura Winther; Carl, Michael; Simard, Michel & Specia, Lucia (Hg.) *Post-Editing of Machine Translation: Processes and Applications.* Newcastle-upon-Tyne: Cambridge Scholars Publisher, 51-77.

Guerberof Arenas, Ana (2020). Pre-editing and post-editing. In: Angelone, Erik; Ehrensberger-Dow, Maureen & Massey, Gary (Hg.) *The Bloomsbury Companion to Language Industry Studies*. London: Bloomsbury Academic, 333-360.

Guerberof Arenas, Ana; Depraetere, Heidi & O'Brien, Sharon (2013). What we know and what we would like to know about post-editing. *Revista Tradumatica* 10, 211-218.

Guerberof Arenas, Ana & Moorkens, Joss (2019). Machine translation and post-editing training as part of a master's programme. *Journal of Specialised Translation* 31, 217-238.

Guerrero Romeo, Lucía (2018). From a Discreet Role to a Co-Star: The Post-Editor Profile Becomes Key in the PEMT Workflow for an Optimal Outcome. In: AsLing, The International Association for Advancement in Language Technology (Hg.) *Proceedings of the 40th Conference Translating and the Computer*. Genf: Université de Genève.

Heinisch, Barbara & Lušicky, Vesna (2019). User expectations towards machine translation: A case study. In: Forcada, Mikel; Way, Andy; Tinsley, John; Shterionov, Dimitar; Rico, Celia & Gaspari, Federico (Hg.) *Proceedings of the 17th Machine Translation Summit – Volume 2: Translator, Project and User Tracks.* Association of Computational Linguistics, 42-48.

Herbig, Nico; Pal, Santanu; Vela, Mihaela; Krüger, Antonio & van Genabith, Josef (2019). Multi-modal indicators for estimating perceived cognitive load in post-editing of machine translation. *Machine Translation* 33 (1-2), 91-115.

Hickey, Sarah & Agulló García, Belén (2021). *The 2021 Nimdzi 100: The Ranking of Top 100 Largest Language Service Providers.* https://www.nimdzi.com/nimdzi-100-top-lsp/#where-do-we-go-from-here (Stand: 11.10.2021).

ISO (2015a). *Qualitätsmanagementsysteme - Grundlagen und Begriffe (ISO 9000:2015).* Brüssel: Europäisches Komitee für Normung.

ISO (2015b). *Übersetzungsdienstleistungen – Anforderungen an Übersetzungsdienstleistungen (ISO 17100:2015).* Brüssel: Europäisches Komitee für Normung.

ISO (2017). *Übersetzungsdienstleistungen - Posteditieren maschinell erstellter Übersetzungen - Anforderungen (ISO 18587:2017).* Brüssel: Europäisches Komitee für Normung.

Jakobsen, Arnt Lykke (2019). Translation technology research with eye tracking. In: O'Hagan, Minako (Hg.) *The Routledge Handbook of Translation and Technology.* Milton: Routledge, 398-416.

Jia, Yanfang; Carl, Michael & Wang, Xiangling (2019). "How Does the Post-Editing of Neural Machine Translation Compare with From-Scratch Translation? A Product and Process Study." *Journal of Specialised Translation* 31, 60–86.

Johnson, Melvin; Schuster, Mike; Le, V. Quoc; Krikun, Maxim; Wu, Yonghui; Chen, Zhifeng; Thorat, Nikhil; Viéga, Fernanda; Wattenberg, Martin; Corrado, Greg; Hughes, Macduff & Dean, Jeffrey (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5, 339-351.

Kalchbrenner, Nal & Blunsom, Phil (2013). *Recurrent Convolutional Neural Networks for Discourse Compositionality*. Cornell: Cornell University Press.

Kenny, Dorothy (2019). Machine Translation. In: Rawling, Piers & Wilson, Philip (Hg.) *The Routledge Handbook of Translation and Philosophy*. New York: Routledge, 428–445.

Koby, Geoffrey S.; Fields, Paul; Hague, Daryl R.; Lommel, Arle & Melby, Alan (2014). Defining Translation Quality. *Revista Tradumatica* 12, 413-420.

Koehn, Philipp (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Koehn, Philipp (2020). *Neural Machine Translation*. Cambridge: Cambridge University Press.

Konttinen, Kalle; Salmi, Leena & Koponen, Maarit (2020). Revision and post-editing competences in translator education. In: Koponen, Maarit; Mossop, Brian, Robert; S. Robert, Isabelle & Scocchera Giovanna (Hg.) *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes*. London: Routledge, 187-202.

Koponen, Maarit (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation* 25, 131-148.

Krings, Hans P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Postediting Processes*. Kent: The Kent State University Press.

Kurz, Christopher (2014). Maschinelle Übersetzung – Vom streng geheimen Forschungsprojekt zum alltäglichen Einsatz. In: Baumann, Klaus-Dieter & Kalverkämper, Hartwig (Hg.) *Theorie und Praxis des Dolmetschens und Übersetzens in fachlichen Kontexten*. Berlin: Frank & Timme, 397-426.

Lacruz, Isabel & Shreve, Gregory M. (2014). Pauses and Cognitive Effort in Post-editing. In: O'Brien, Sharon; Balling, Laura Winther; Carl, Michael; Simard, Michel & Specia, Lucia (Hg.) *Post-Editing of Machine Translation: Processes and Applications.* Newcastle-upon-Tyne: Cambridge Scholars Publisher, 246-274.

Läubli, Samuel & Green, Spence (2019). Translation technology research and human–computer interaction (HCI). In: O'Hagan, Minako (Hg.) *The Routledge Handbook of Translation and Technology.* Milton: Routledge, 370-383.

Levenshtein Vladimir. (1966). Binary codes capable of correcting deletions, insertions, and reversals*. Soviet Physics – Doklady* 10 (8), 707–710.

Massardo, Isabella; van der Meer, Jaap; O'Brien, Sharon; Hollowood, Fred; Aranberri, Nora & Drescher, Katrin (2016). *TAUS MT Post-Editing Guidelines*. Amsterdam: TAUS Signature Editions.

Melamed, Dan; Green, Ryan & Turian, Joseph P. (2003). Precision and recall of machine translation. In: Association for Computational Linguistics (Hg.) *Companion Volume of the Proceedings of HLT-NAACL 2003 - Short Papers.* Edmonton: ACL, 61-63.

Mesa-Lao, Bartolomé (2014). Gaze behaviour on source texts: an exploratory study comparing translation and post-editing. In: O'Brien, Sharon; Balling, Laura Winther; Carl, Michael; Simard, Michel & Specia, Lucia (Hg.) *Post-Editing of Machine Translation: Processes and Applications.* Newcastle-upon-Tyne: Cambridge Scholars Publisher, 219-240.

Moorkens, Joss (2017). Under Pressure: Translation in Times of Austerity. *Perspectives: Studies in Translatology* 25 (3): 464–477.

Moorkens, Joss (2018). Eye tracking as a measure of cognitive effort for post-editing of machine translation. In: Walker, Callum & Federici, Frederico M. (Hg.) *Eye Tracking*

and Multidisciplinary Studies on Translation. Benjamins Translation Library 143, 55–69.

Nauth, Danny (2012). *Durch die Augen meines Kunden: Praxishandbuch für Usability Tests mit einem Eyetracking System*. Hamburg: Diplomica.

Nießen, Sonja; Och, Franz Josef; Leusch, Gregor & Ney, Hermann (2000). An evaluation tool for machine translation: Fast evaluation for MT research. In: Gavrilidou, Maria; Carayannis, George & Markantonatou, Stella (Hg.) *Proceedings of the Second International Conference on Language Resources and Evaluation*. Athens: ELRA.

Nitzke, Jean (2019). *Problem solving activities in post-editing and translation from scratch. A multi-method study*. Berlin: Language Science Press.

Nitzke, Jean; Hansen-Schirra, Silvia & Canfora, Carmen (2019). Risk management and post-editing competence. *Journal of Specialised Translation* 31, 239-259.

Nord, Christiane (2006). Translationsqualität aus funktionaler Sicht. In: Schippel, Larisa (Hg.). *Übersetzungsqualität: Kritik –Kriterien –Bewertungshandeln*. Berlin: Frank & Timme, 11-30.

O'Brien, Sharon (2002). Teaching post-editing: A proposal for course content. In: European Association for Machine Translation (Hg.) *6th EAMT Workshop. Teaching Machine Translation*, 99-106.

O'Brien, Sharon (2006). Eye Tracking and Translation Memory Matches. *Perspectives: Studies in Translatology* 14 (3): 185–205.

O'Brien, Sharon (2011). Towards predicting post-editing productivity. *Machine Translation* 25, 197-215.

O'Brien, Sharon (2012). Translation as human–computer interaction. *Translation Spaces* 1 (1), 101-122.

Orr, David B. & Small, Victor H. (1967). Comprehensibility of Machine-aided Translations of Russian Scientific Documents. *Mechanical Translation and Computational Linguistics* 10 (1-2), 1-10.

Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing (2002). BLEU: A method for automatic evaluation of machine translation. In: Isabelle, Pierre; Charniak, Eugene & Lin, Dekang (Hg.) *Proceedings of ACL-2002: 40th Annual meeting of the Association for Computational Linguistics*. Philadelphia: ACL, 311–318.

Pérez Macías, Lorena (2020). What do translators think about post-editing? A mixed-methods study of translators' fears, worries and preferences on machine translation post-editing. *Revista Tradumatica* 18, 11-32.

Poibeau, Thierry (2017). *Machine translation*. Cambridge, Massachusetts: MIT Press.

Porsiel, Jörg (2020). Welcome to the Machine! Zwischen Goldgräberstimmung und der Suche nach dem Heiligen Gral. In: Porsiel, Jörg (Hg.) *Maschinelle Übersetzung für Übersetzungsprofis*. Berlin: BDÜ Weiterbildungs- und Fachverlags GmbH, 2-10.

Rajs, Arkadiusz; Aleksiewicz, Mariusz; Banaszak-Piechowska, Agnieszka & Gospodarczyk, Jacek (2016). Methodology and applications of eyetracking. *Journal of education, health and sport*. 6 (4), 115-121.

Reiß, Katharina & Vermeer, Hans J. (1984). *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: De Gruyter.

Robert, Anne-Marie (2013). Vous avez dit post-éditrice ? Quelques éléments d'un parcours personnel. *Journal of Specialised Translation* 19, 29–40.

Routier, Johann; Mitchell, Linda & Silva, David (2013). The ACCEPT Post-Editing Environment: A Flexible and Customisable Online Tool to Perform and Analyse Machine Translation Post-Editing. In: O'Brien, Sharon; Simard, Michael & Specia, Lucia (Hg.) *Proceedings of the 2nd Workshop on Post-editing Technology and Practice.* Nizza: ACL, 119-128.

Sakamoto, Akiko (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu's concepts. *Journal of Specialised Translation* 31, 201-216.

Sakamoto, Akiko & Yamada, Masaru (2020). Social groups in machine translation post-editing: A SCOT analysis. *Translation Spaces* 9 (1), 78-97.

Sin-wai, Chan (2019). *The Human Factor in Machine Translation.* London: Routledge.

Scansani, Randy; Bernardini, Silvia; Ferraresi, Adriano & Bentivogli, Luisa (2019). Do translator trainees trust machine translation? An experiment on post-editing and revision. In: Forcada, Mikel; Way, Andy; Tinsley, John; Shterionov, Dimitar; Rico, Celia & Gaspari, Federico (Hg.) *Proceedings of the 17th Machine Translation Summit – Volume 2: Translator, Project and User Tracks.* Association of Computational Linguistics, 73-79.

Schmalz, Antonia (2019). Maschinelle Übersetzung. In: Wittpahl, Volker (Hg.) *Künstliche Intelligenz. Technologie, Anwendung, Gesellschaft*. Berlin: Springer, 194-210.

Schubert, Klaus (1999). Zur Automatisierbarkeit des Übersetzens. In: Gil, Alberto; Haller, Johann; Steiner, Erich & Gerzymisch-Arbogast, Heidrun (Hg.) *Modelle der Translation – Grundlagen für Methodik, Bewertung, Computermodellierung.* Frankfurt: Lang, 423-441.

Serra, Richard & Weyergraf, Clara (1980). *Richard Serra: Interviews, Etc. 1970-1980.* New York: The Hudson River Museum.

Shannon, Claude & Weaver, Warren (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Silva, Roberto (2014). Integrating Post-Editing MT in a Professional Translation Workflow. In: O'Brien, Sharon; Balling, Laura Winther; Carl, Michael; Simard, Michel & Specia, Lucia (Hg). *Post-Editing of Machine Translation: Processes and Applications.* Newcastle-upon-Tyne: Cambridge Scholars Publisher, 24-51.

Somers, Harold (2003). *Computers and translation: a translator's guide*. John Benjamins, Amsterdam.

SR Research (2021). EyeLink Portable Duo. https://www.sr-research.com/eyelink-portable-duo/ (Stand: 16.12.2021).

Stein, Daniel (2018). Machine Translation: Past, Present and Future. In: Rehm, Georg; Sasaki, Felix; Stein, Daniel & Witt, Andreas (Hg.) *Language Technologies for a Multilingual Europe*. Berlin: Language Science Press, 5-17.

Torrejón, Enrique & Rico, Celia (2013). Skills and Profile of the New Role of the Translator as MT Post-editor. *Revista Tradumatica* 10, 166-178.

van der Meer, Jaap & Ruopp, Achim (2014). *TAUS Machine Translation Market Report*. Amsterdam: TAUS Signature Editions.

Veale, Tony & Way, Andy (1997). Gaijin: A bootstrapping approach to example-based machine translation. *International conference on recent advances in natural language processing*, 239-244.

Vidal, Sergi Alvarez; Oliver, Antoni & Badia, Toni (2020). Post-editing for professional translators: Cheer or fear? *Revista Tradumatica* 18, 50-69.

Vieira, Lucas Nunes (2014). Indices of cognitive effort in machine translation post-editing. *Machine Translation* 28 (3-4), 187–216.

Vieira, Lucas Nunes (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation* 30 (1-2), 41–62.

Vieira, Lucas Nunes (2019). Post-editing of machine translation. In: O'Hagan, Minako (Hg.) *The Routledge Handbook of Translation and Technology.* Milton: Routledge, 319-335.

Vieira, Lucas Nunes; Alonso, Elisa & Bywood, Lindsay (2019). Introduction: Post-editing in practice – Process, product and networks. *Journal of Specialised Translation* 31, 75-95.

Wadhwani, Preeti & Gankar, Saloni (2021). *Machine Translation Market Size By Technology. Forecast, 2021 – 2027.* Selbyville: Global Market Insights. In: https://www.gminsights.com/industry-analysis/machine-translation-market-size (Stand: 14.10.2021).

Way, Andy (2018). Quality Expectations of Machine Translation. *Translation Quality Assessment*, 1-22.

Weaver, Warren (1949). *Translation*. Carlsbad: The Rockefeller Foundation.

Yamada, Masaru (2015). Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing settings. *Machine Translation* 29 (1), 49–67.

Yamada, Masaru (2019). The impact of Google neural machine translation on post-editing by student translators. *Journal of Specialised Translation* 31, 87-106.

Yang, Yanxia & Wang, Xiangling (2020). Predicting student translators' performance in machine translation post-editing: interplay of self-regulation, critical thinking, and motivation. *Interactive Learning Environments*, 1-15.

Zaretskaya, Anna (2017). Machine Translation Post-Editing at TransPerfect – the 'Human' Side of the Process. *Revista Tradumatica* 15, 116-123.

ZTW (2021). Mastercurriculum "Multilingual Technologies". https://transvienna.univie.ac.at/studium/masterstudium-translation/masterprogramm-multilingual-technologies/ (Stand: 21.01.2022).

Abstract (Deutsch)

Aus diversen translationswissenschaftlichen Forschungsarbeiten über Post-Editing geht hervor, dass professionelle Übersetzer*innen Post-Editing sowie maschineller Übersetzung skeptisch gegenüberstehen und selbst trotz möglicher Produktivitätsgewinne das Humanübersetzen bevorzugen. Davon ausgehend kann die Frage gestellt werden, ob die neue Generation an Übersetzer*innen, etwa Masterstudierende, für das Post-Editing durch ihre translatorische Ausbildung geeignet ist und vor allem, ob sie die Tätigkeit in einem positiveren Licht als ihre erfahrenen Kolleg*innen sehen. Allerdings verfügen nur wenige Studierenden über eine Post-Editing-Ausbildung, wodurch ihnen nötige Kompetenzen fehlen. Die Forschung könnte jedoch beim Erstellen von Post-Editing Curricula und Kursen helfen, und auch diese Arbeit soll dazu einen kleinen Beitrag leisten, indem der temporale, technische und kognitive Aufwand, der bei Studierenden während eines Post-Editings auftritt, gemessen und analysiert wird.

Dafür wurde eine Studie mit drei Masterstudierenden durchgeführt, die kurz vor Ende ihrer Ausbildung am Zentrum für Translationswissenschaft stehen, über keine Post-Editing-Ausbildung verfügen und einen Text, der mit *DeepL* vom Englischen ins Deutsche maschinell übersetzt wurde, unter gewissen Vorgaben im CAT-Tool *MemoQ* post-editieren mussten. Dabei wurde die Bearbeitungszeit in den diversen Segmenten aufgezeichnet, die Anzahl und Verteilung von Änderungen in den Segmenten erfasst und die Augenbewegungen der Teilnehmer*innen mit einem Eyetracker gemessen. Dieser gab Aufschluss über Anzahl, Dauer und Verteilung von Fixationen innerhalb von vier festgelegten Interessensbereichen, nämlich Browserfenster, Ausgangstextspalte, Zieltextspalte und Bereich mit Translation-Memory Treffern. Aus den gewonnenen Daten ging hervor, dass der Bereich des Zieltextes für Studierende mit dem höchsten temporalen, technischen und kognitiven Aufwand verbunden war. Zusätzlich ging aus den mit den Teilnehmer*innen geführten Interviews hervor, dass sie dem Post-Editing bereits vor Durchführung der Aufgabe generell aufgeschlossen waren und spätestens nach dem Post-Editing die trotz fehlender Ausbildung möglichen Produktivitätsgewinne erkannten.

Abstract (Englisch)

In keeping with the topic of this master's thesis, the following abstract was machine translated from German using DeepL and then post-edited by the author.

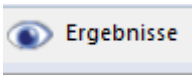
Various research papers on post-editing have shown that professional translators tend to be sceptical about post-editing and machine translation and usually prefer human translation in spite of possible productivity gains. This begs the question whether a new generation of translators, such as master's students, would also be suited for post-editing due to their translation training and, above all, whether they are more open to this new activity than their experienced peers. However, few students have received post-editing training and therefore lack the necessary skills. Still, research projects could help in creating post-editing curricula and courses, an effort towards which this thesis will also make a small contribution by measuring and analysing the temporal, technical and cognitive effort that students experience during post-editing.

To this end, a study was carried out with three master's students who are nearing the end of their training at the Centre for Translation Studies in Vienna, who have received no post-editing training and who were asked to post-edit a text in *MemoQ* that had been machine-translated from English into German using *DeepL*. During the experiment, the time spent editing the various segments was recorded, as well as the number and distribution of changes made in the segments. The participants' eye movements were measured using an eye tracker, which provides information about the amount, duration, and distribution of fixations within four defined areas of interest, namely the web browser, the source text column, the target text column, and the area containing translation memory matches. The data obtained revealed that for the students the target text column was associated with the highest temporal, technical and cognitive effort. In addition, the interviews conducted with the students revealed that they generally approved of post-editing even before having performed the task and unanimously recognised the possible productivity gains associated with post-editing despite their lack of prior training.

Anhang

Anhang A: Anleitung der Studie für die Teilnehmer*innen

Anleitung für die Post-Editing Studie

- Behandle die Aufgabe wie einen echten Post-Editing-Auftrag.
- Du wurdest damit beauftragt, ein vollständiges Post-Editing des Textes durchzuführen. Die Norm ISO 18587 definiert dieses als „process of post-editing […] to obtain a product comparable to a product obtained by human translation […]".
- Von der maschinellen Übersetzung soll so viel wie möglich beibehalten werden und nur so viel wie nötig an den relevanten Stellen gelöscht und editiert werden.
- Ein maschinell übersetztes Segment darf nur gänzlich gelöscht werden und von dir eigenständig neu übersetzt werden, wenn es mehrere schwerwiegende Fehler beinhaltet. Stilistische Präferenzen, z.B. bezüglich Formulierungen, sind kein Grund für ein vollständiges Löschen.
- Achte darauf, dass am Ende deiner Arbeit alle Segmente mit der Tastenkombination *Strg+Enter* in das Translation Memory gespeichert wurden, und dass neben allen Segmenten ein grünes Häkchen zu sehen ist.
- Im *MemoQ*-Kästchen rechts außen kannst du unter 

die maschinelle Übersetzung für das ausgewählte Segment einsehen, wenn du sie im Textfeld überschrieben oder gelöscht hast.
- Das *MemoQ*-Fenster, in dem du arbeitest, sollte unter keinen Umständen verschoben oder geschlossen werden.
- Bitte gendere im Text einheitlich mit * z.B.: „*jede/r Mitarbeiter*in*"
- Neben dem *MemoQ*-Fenster ist ein Browser-Fenster geöffnet, das du zur Recherche und für Online-Wörterbücher verwenden darfst. Dazu zählt auch Linguee.
- Das Verwenden von Online-MÜ-Systemen wie DeepL ist untersagt.
- Es gibt bei dieser Aufgabe kein Zeitlimit. Als Richtwert werden aber rund 60 Minuten angegeben.
- Bei technischen Problemen oder Fragen bitte den Forschungsleiter rufen und nicht vom PC aufstehen.
- Nach Abschließen der Aufgabe bitte sitzen bleiben und dem Forschungsleiter Bescheid geben.

Anhang B: Englischer Ausgangstext und deutsche Übersetzung von DeepL

ID	Englisch	Deutsch
1	Embracing cryptocurrency	Umarmung der Kryptowährung
2	*The fact that cryptocurrency has no legal classification should not be the impetus to prohibit its use in India.*	*Die Tatsache, dass Kryptowährungen rechtlich nicht klassifiziert sind, sollte nicht der Anlass sein, ihre Verwendung in Indien zu verbieten.*
3	On June 9, El Salvador became the first country in the world to adopt bitcoin as legal tender.	Am 9. Juni wurde El Salvador zum ersten Land der Welt, das Bitcoin als gesetzliches Zahlungsmittel einführte.
4	This is illustrative of the rising global trend of embracing cryptocurrencies with all its attendant risks.	Dies ist bezeichnend für den zunehmenden globalen Trend, Kryptowährungen mit allen damit verbundenen Risiken zu akzeptieren.
5	While not every country's approach has been as open as El Salvador's, the dominant theme has been to permit the growth of the cryptocurrency market subject to certain safeguards.	Zwar ist nicht jedes Land so offen wie El Salvador, doch das vorherrschende Thema ist, das Wachstum des Kryptowährungsmarktes vorbehaltlich bestimmter Sicherheitsvorkehrungen zuzulassen.
6	As India finds itself at a crossroads of prohibition and regulation in its tryst with cryptocurrencies, globally, the inclination towards permissive regulation recognises the freedom of choice given to people for using a medium of exchange other than a central bank-backed fiat currency.	Während sich Indien bei seinem Rendezvous mit Kryptowährungen an einem Scheideweg zwischen Verbot und Regulierung befindet, erkennt die Tendenz zu einer permissiven Regulierung weltweit die Wahlfreiheit an, die den Menschen bei der Verwendung eines anderen Tauschmittels als einer zentralbankgestützten Fiat-Währung gegeben ist.
7	Swinging between extremes	Schwanken zwischen den Extremen
8	The cryptocurrency market in India has developed in a largely laissez-faire regulatory space since the first recorded cryptocurrency transaction in 2010.	Der Kryptowährungsmarkt in Indien hat sich seit der ersten aufgezeichneten Kryptowährungstransaktion im Jahr 2010 in einem weitgehend laissez-faire Regulierungsraum entwickelt.
9	Between 2013 and 2018, the government's response to the rise of virtual currencies was cautionary, alerting users to the potential risks posed by cryptocurrency transactions.	Zwischen 2013 und 2018 reagierte die Regierung auf den Aufstieg der virtuellen Währungen mit Vorsicht und warnte die Nutzer vor den potenziellen Risiken von Kryptowährungstransaktionen.
10	These fears were legitimate and stemmed from cryptocurrencies' volatility, their susceptibility to hacking, and the fact that they could potentially facilitate criminal activities such as money laundering, terrorist financing and tax evasion.	Diese Befürchtungen waren berechtigt und rührten von der Volatilität der Kryptowährungen, ihrer Anfälligkeit für Hackerangriffe und der Tatsache her, dass sie möglicherweise kriminelle Aktivitäten wie Geldwäsche, Terrorismusfinanzierung und Steuerhinterziehung erleichtern könnten.
11	Instead of developing a regulatory framework to address these issues, the Reserve Bank of India (RBI), in April 2018, effectively imposed a ban on cryptocurrency trading.	Anstatt einen Regulierungsrahmen zu entwickeln, um diese Probleme anzugehen, verhängte die Reserve Bank of India (RBI) im April 2018 ein Verbot für den Handel mit Kryptowährungen.
12	This ban was overturned by the Supreme Court in 2020.	Dieses Verbot wurde vom Obersten Gerichtshof im Jahr 2020 gekippt.

13	The court reasoned that there were alternative regulatory measures short of an outright ban through which the RBI could have achieved its objective of curbing the risks associated with cryptocurrency trading.	Das Gericht begründete dies damit, dass es alternative regulatorische Maßnahmen gab, mit denen die RBI ihr Ziel, die mit dem Kryptowährungshandel verbundenen Risiken einzudämmen, hätte erreichen können, ohne ein vollständiges Verbot auszusprechen.
14	After swinging between the extremes of non-interference and prohibition, a clue as to India's next move lies in the draft Cryptocurrency and Regulation of Official Digital Currency Bill, 2021.	Nachdem Indien zwischen den Extremen der Nichteinmischung und des Verbots hin- und hergeschwankt ist, gibt der Gesetzesentwurf "Cryptocurrency and Regulation of Official Digital Currency Bill, 2021" einen Hinweis auf den nächsten Schritt.
15	The draft Bill proposes to criminalise all private cryptocurrencies while also laying down the regulatory framework for an RBI-backed digital currency.	Der Gesetzesentwurf sieht vor, alle privaten Kryptowährungen zu kriminalisieren und gleichzeitig den Regulierungsrahmen für eine von der RBI unterstützte digitale Währung festzulegen.
16	The Minister of State for Finance, in response to a question in Parliament, stated that regulatory bodies do not have a legal framework to directly regulate private cryptocurrencies owing to their imprecise legal nature in India.	Der Staatsminister für Finanzen erklärte in seiner Antwort auf eine Anfrage im Parlament, dass die Regulierungsbehörden keinen rechtlichen Rahmen haben, um private Kryptowährungen aufgrund ihrer unpräzisen rechtlichen Natur in Indien direkt zu regulieren.

Anhang C: Interviews vor Durchführen der Post-Editing-Aufgabe

Teilnehmer*in 1 – Interview 1

10.01.2022, 11:19, Zentrum für Translationswissenschaft

Was fällt dir zum Begriff „Post-Editing“ spontan ein?

Eine qualitativ bessere Übersetzung durch Überarbeitung.

Und diese Übersetzung ist wie entstanden?

Durch die Maschine. Also die Übersetzung des Ausgangstextes sozusagen wurde durch die Maschine in den Zieltext transferiert und dann überarbeitet. Und je nach Aufwand, da gibt es ja zwei Arten des Post-Editing, genau, ja.

Hast du schon einmal ein Post-Editing eines maschinell übersetzten Textes durchgeführt?

[lacht] Ja, genug.

Erzähl mal darüber.

Ja also ich finde es super einfach. Es braucht natürlich Übung sozusagen, das ist eh klar, aber ich finde es erspart extrem viel Zeit und wenn man wirklich schnell eine Übersetzung anfertigen möchte und gerne mit MÜ arbeitet, dann ist das sicher eine sehr gute Aufgabe, um sozusagen zeitsparend und kosteneffizient eine Übersetzung anzufertigen.

Und wie hast du es genutzt bis jetzt?

Im Rahmen der Uni sage ich immer dazu. Also im Berufsleben nicht. [lacht] Leider nicht. Nur für Hausübungen und Aufträge und Prüfungen.

Hast du schon einmal eine Revision/Korrektur eines humanübersetzten Textes durchgeführt?

Ja, ja. Ist vorgekommen in Übersetzungskursen, sowohl im Bachelor als auch Master.

Hast du eine Post-Editing-Ausbildung (Fachvertiefung an der Uni, Workshop, Online-Kurs) absolviert?

Nein nichts. Im Prinzip eigentlich selber beigebracht, ja.

Hast du beim PE bewusst einen Plan oder eine Strategie angewandt?

Im Normalfall nein, da habe ich es nach Gefühl gemacht.

Bist du dem PE aufgeschlossen oder abgeneigt?

Sehr aufgeschlossen.

Nutzt du maschinelle Übersetzungssysteme wie DeepL in deinem privaten und professionellen Alltag?

Ja, ja.

Mehr privat als professionell?

Nur privat im Prinzip.

Für welchen Zweck nutzt du maschinelle Übersetzungssysteme am meisten?

Ja zum Beispiel Internetseiten wo man sich Rezensionen durchlesen will in einer anderen Sprache, Booking, Tripadvisor zum Beispiel, finde ich sehr sinnvoll. Also wenn ich schnell eine Info brauche.

Du wirst anschließend einen Zeitungsartikel zum Thema Kryptowährung post-editieren, der mit DeepL EN-DE übersetzt wurde. Welche Qualität erwartest du dir von dem maschinell-übersetzten Zieltext bezüglich Korrektheit und Stil?

Also ich glaube, dass DeepL in diesem Bereich ganz gut abschneidet. Ich glaube für das Finanzwesen, für Texte aus diesem Bereich ist DeepL ganz gut geeignet.

Du erwartest dir also, dass du an groben Fehlern gar nicht so viel bearbeiten musst?

Von groben Fehlern nicht, eher vom Sprachstil. Vielleicht auch von der Terminologie, wobei ich habe in meiner Masterarbeit ja selber gemerkt, dass die Terminologie teilweise schon sehr gut ist von DeepL, also ja.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich mit einem höheren kognitiven Aufwand verbunden, sprich, welche ist mental anstrengender? PE oder Humanübersetzung?

Das Humanübersetzen. Also für mich schon, für andere vielleicht nicht.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich schneller? PE oder Humanübersetzung?

Auf jeden Fall Post-Editing.

Angenommen du erhältst einen Übersetzungsauftrag und darfst entscheiden, ob du diesen mittels Humanübersetzung oder Post-Editing von MÜ erledigen möchtest: Welche Arbeitsweise wählst du?

Ja da kommt es darauf wieviel der Übersetzer, die Übersetzerin verlangt, ich muss mir natürlich immer den Kosten-Nutzen Faktor anschauen und auch für was die Übersetzung verwendet wird. Also sagen wir mal wenn das publiziert wird, würde ich mir schon überlegen, ob ich das nicht vielleicht ganz humanübersetzen lasse oder nicht. Wenn es jetzt für den Eigengebrauch ist, würde ich sagen Post-Editing reicht, und ja es kommt darauf an wo das publiziert wird, wenn das jetzt ein kleiner Text ist der irgendwo im Internet publiziert wird, dann kann ich das vielleicht noch post-editieren. Wenn das jetzt irgendwie was Gedrucktes ist, was Längeres, dann würde ich eher sagen humanübersetzen. Das kann man nicht einfach so sagen, das müsste man sich im Einzelfall anschauen. Da habe ich keine konkrete Antwort dazu.

Teilnehmer*in 2 – Interview 1

12.01.2022, 12:10, Zentrum für Translationswissenschaft

Was fällt dir zum Begriff „Post-Editing“ spontan ein?

Also ich würde sagen es ist ein Text, der vorübersetzt ist von Mensch oder Maschine, das ist ja eigentlich egal, der dann im Nachhinein eben bearbeitet wird. Also in Bezug auf Rechtschreibung, Grammatik, vielleicht auch ob die richtigen Wortentsprechungen gefunden worden sind, also nicht nur vielleicht sondern ziemlich sicher. Also ja, solche Sachen, wie so eine Korrektur halt.

Und was glaubst du ist das Ziel von Post-Editing, was soll es für einen Mehrwert bringen?

Es geht vielleicht schneller, wenn man es von einer Maschine übersetzen lässt und dann im Nachhinein eben einen Menschen nochmal drüber schauen lässt, weil eben viele Maschinen mittlerweile sehr gut übersetzen. Ja, also es ist vielleicht das der Mehrwert, dass es eben Zeit und dadurch auch Geld spart.

Hast du schon einmal ein Post-Editing eines maschinell übersetzten Textes durchgeführt?

Ich glaube jetzt nicht. Also nicht so, dass ich es gewusst hätte. Vielleicht hin und wieder, bei den Untertiteln, die ich übersetzt habe, da waren mal welche dabei, die haben für mich sehr maschinell übersetzt ausgeschaut, aber jetzt nicht so dass ich gewusst hätte „Das ist jetzt eine maschinelle Übersetzung gewesen und ich mache jetzt Post-Editing“, das war jetzt nicht bewusst.

Hast du schon einmal eine Revision/Korrektur eines humanübersetzten Textes durchgeführt?

Ja, das ist mein Job, mehr oder weniger.

Das heißt du bist ziemlich vertraut mit dem Revisions- oder Korrekturarbeitsprozess?

Ja schon, also jetzt eben nicht nur Übersetzungen, sondern auch so normale Texte mache ich öfter.

Hast du eine Post-Editing-Ausbildung (Fachvertiefung an der Uni, Workshop, Online-Kurs) absolviert?

Nein, nein.

Bist du dem PE aufgeschlossen oder abgeneigt?

Ja sicher aufgeschlossen. Ich denke mir, wenn es, also natürlich ist es für uns als Übersetzer nicht gut, wenn uns die ganze Arbeit weggenommen wird [lacht], aber es ist natürlich sowohl zeit- also auch kostensparend, es wird mehr oder weniger nichts anderes übrigbleiben als das. Ich habe da eigentlich eine recht neutrale Einstellung.

Glaubst du fällt diese Arbeit für den Menschen weg, oder ist es einfach nur eine Umschichtung, also was ein Mensch vorher übersetzt hätte, muss er jetzt post-editieren?

Ja ich glaube es ist eine andere Rolle, es ist nicht so, dass ein komplettes Arbeitsfeld verloren geht, du kannst dann einfach das andere machen. Die Frage ist nur, wie groß dann noch die Nachfrage ist. Also ob dann wirklich für alle Übersetzer, die jetzt aktiv übersetzen, auch wirklich die Arbeit auch überbleibt die gemacht werden muss, oder ob viele dann irgendwie umlernen müssen oder etwas anderes machen müssen. Das kann ich dir nicht beantworten, das weiß ich nicht [lacht].

Nutzt du maschinelle Übersetzungssysteme wie DeepL in deinem privaten und professionellen Alltag?

Ja, hin und wieder.

Für welche Zwecke nutzt du maschinelle Übersetzungssysteme am meisten?

Also ich verwende vor allem DeepL hin und wieder, vor allem für Französisch, wenn ich da einfach nicht weiß, wie ich das am besten ausdrücken kann. Aber meistens halt auch um zu schauen, was mir das so ausspuckt, nicht dass ich das dann Hundertprozent verwenden würde, weil wir alle wissen, dass das meistens nicht die genauen Entsprechungen sind, oder zumindest oft. Ja, also ich würde eher sagen als Inspiration, nicht als Vollübersetzung.

Du wirst anschließend einen Zeitungsartikel zum Thema Kryptowährung post-editieren, der mit DeepL EN-DE übersetzt wurde. Welche Qualität erwartest du dir von dem maschinell-übersetzten Zieltext bezüglich Korrektheit und Stil?

Ich würde sagen eigentlich eine relativ gute Qualität. Weil normalerweise sind die Texte, die DeepL aus dem Englischen ins Deutsche übersetzt, oder auch ins Englische, das ist eigentlich egal, eigentlich sind die meistens ganz gut. Also ich erwarte mir da jetzt keine riesengroßen Fehler.

Auch mit der Terminologie?

Also es kann schon sein, aber ich denke mir, dass diese ganze Kryptowährungsterminologie eh noch relativ neu ist, von da her gibt es relativ viele

Texte, die sich DeepL da nehmen kann, also denke ich sollte das nicht so ein großes Problem sein, aber das werde ich gleich sehen.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich mit einem höheren kognitiven Aufwand verbunden, sprich, welche ist mental anstrengender? PE oder Humanübersetzung?

Das ist eine gute Frage. Wahrscheinlich eher der maschinell übersetzte Text, weil ich habe da das Gefühl, da muss ich von Anfang an mehr aufpassen was jetzt passiert sein kann. Bei einem Humanübersetzer habe ich eher das Vertrauen, dass das was geschrieben ist auch wirklich stimmt. Da hätte ich von Grund auf ein positiveres Gefühl bei einer Humanübersetzung.

Also auch wenn du die Humanübersetzung selbst machst.

Ja, ja. Also bei einer maschinellen Übersetzung wäre ich einfach genauer und würde nochmal mehr nachschauen, ob das jetzt wirklich die richtigen Entsprechungen sind.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich schneller? PE oder Humanübersetzung?

Ich glaube die Humanübersetzung. Eben weil ich glaube ich bei der maschinellen Übersetzung mehr nochmal drüber schauen würde. Also bei einer Humanübersetzung würde ich eher denken „Das passt jetzt so“, vielleicht noch einen zweiten Durchgang machen. Bei der maschinellen Übersetzung würde ich nochmal drüber gehen um sicher zu sein, dass ich nichts übersehen habe, was da jetzt falsch sein könnte. Also ich glaube es ist eine Einstellungssache bei mir.

Angenommen du erhältst einen Übersetzungsauftrag und darfst entscheiden, ob du diesen mittels Humanübersetzung oder Post Editing von MÜ erledigen möchtest: Welche Arbeitsweise wählst du?

Ich würde ihn lieber selber übersetzen, weil ich dafür auch bezahlt werde. Also natürlich kommt es immer darauf an, wieviel man bezahlt bekommt und wie ernst das Thema ist, und wie terminologisch anspruchsvoll, das ist eh klar. Aber ich würde ihn trotzdem selber übersetzen, eben mit Hilfe von irgendwelchen Wörterbüchern online halt, aber nicht maschinell übersetzen lassen.

Teilnehmer*in 3 – Interview 1

13.01.2022, 14:25, Zentrum für Translationswissenschaft

Was fällt dir zum Begriff „Post-Editing“ spontan ein?

Also beim Post-Editing ist es so, es gibt eine Machine Translation und die wird dann im Nachhinein so bearbeitet, damit sie sprachlich so nahe an der natürlichen Sprache ist, wie wenn es ein echter Mensch geschrieben hätte.

Das heißt das Ziel ist...

Ja es soll natürlich klingen, und die Qualität soll natürlich auch angemessen sein, je nachdem was es halt betrifft.

Hast du schon einmal ein Post-Editing eines maschinell übersetzten Textes durchgeführt?

Ja also so, ja, privat und manchmal für die Uni.

Für Hausübungen oder?

Wenn man ganz dringend was gebraucht hat und keine Zeit gehabt hat, ja. [lacht]

Also zum Zeitsparen?

Genau.

Und wie ist es dir damit gegangen?

Eigentlich finde ich, dass die, zum Beispiel wenn wir jetzt genauer sprechen über DeepL zum Beispiel, dass die schon ziemlich gut sind, vor allem was zum Beispiel Englisch betrifft. Und dass man halt auf kleine Feinheiten vielleicht aufpassen muss, aber im Groben sind die schon echt gut.

Hast du solche post-editierten Texte als Hausübung auch abgegeben?

Nein, also ich habe meistens schon persönlich, also meinen eigenen Draft angefertigt, und dann nochmal kurz drüber laufen lassen, und so habe ich das gemacht.

Hast du schon einmal eine Revision/Korrektur eines humanübersetzten Textes durchgeführt?

Ja, von Studienkollegen halt auch, in Übungen, wenn man es so gegenseitig überprüft hat.

Aber nichts im beruflichen Alltag?

Nein, nein.

Hast du eine Post-Editing-Ausbildung (Fachvertiefung an der Uni, Workshop, Online-Kurs) absolviert?

Nö, leider nein. [lacht]

Bist du dem PE aufgeschlossen oder abgeneigt?

Eher aufgeschlossen, weil vor allem kann ich mir vorstellen, dadurch dass ich zwar noch nicht im Beruf bin, aber man das von allen Seiten hört, dass einfach die Arbeitslast nicht zu bewältigen ist ohne solche Möglichkeiten.

Also das heißt du bist aufgeschlossen, aber eher weil es so sein muss?

Naja, aus Zeitmangel vermutlich, oder natürlich auch weil es vielleicht Texte gibt, die insofern jetzt nicht so anspruchsvoll sind, sondern einfach nur nochmal kurz drüber gelesen wird und man dann sagt „Ja passt, kann veröffentlicht werden“.

Nutzt du maschinelle Übersetzungssysteme wie DeepL in deinem privaten und professionellen Alltag?

Nein, eigentlich gar nicht.

Zur Studienzeit schon?

[lacht] Ja, da schon. Eben auch zum Üben für die Modulprüfung, um gegenzuchecken, zu schauen, ob ich auf einem ähnlichen Niveau, beziehungsweise ob mein Niveau sogar besser ist wie das von der Machine Translation.

Das heißt du hättest eh schon recht hohe Erwartungen an das, was die Maschine ausspuckt?

[lacht] Ja eigentlich schon denke ich.

Du wirst anschließend einen Zeitungsartikel zum Thema Kryptowährung post-editieren, der mit DeepL EN-DE übersetzt wurde. Welche Qualität erwartest du dir von dem maschinell-übersetzten Zieltext bezüglich Korrektheit und Stil?

Insofern denke ich nicht so schlecht, aber weil vielleicht bei so Kryptowährungen die ganze Terminologie vielleicht ziemlich neu ist und das Ganze vielleicht nicht so harmonisch ist, wie wenn es eben ein Mensch machen würde.

Glaubst du das, weil die Maschine von vielen verschiedenen Texten viele verschiedene Termini holt?

Ja, genau, wegen der Konsistenz.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich mit einem höheren kognitiven Aufwand verbunden, sprich, welche ist mental anstrengender? PE oder Humanübersetzung?

[Pause] Hm. Es ist insofern anstrengend, wenn man zum Beispiel überhaupt keine Ahnung hat von dem Thema Kryptowährung vermutlich, weil man dann die ganze Terminologie erst kennenlernen muss, und insofern wäre es dann anstrengend, sich in das Ganze einzuarbeiten und dann den Text zu übersetzen. Was natürlich aber auch Voraussetzung ist beim Post-Editing, dass man über das Thema Bescheid weiß, weil sonst erkennt man ja nicht, ob der Computer Fehler gemacht hat. Aber ich würde glaube ich schon sagen mental anstrengender ist das Humanübersetzen.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich schneller? PE oder Humanübersetzung?

Schneller [Pause] vermutlich mit Post-Editing, einfach weil ich mir schon mal die Tippzeit erspare, und auch weil man generell als Mensch, als Humanübersetzer, kritischer gegenüber dem ist, was man selber hinschreiben würde, als wie wenn es eh schon dasteht.

Angenommen du erhältst einen Übersetzungsauftrag und darfst entscheiden, ob du diesen mittels Humanübersetzung oder Post Editing von MÜ erledigen möchtest: Welche Arbeitsweise wählst du? Wieso hast du dich so entschieden?

Kommt darauf an, und zwar mal auf die Zeit, wieviel Zeit ich habe. Wenn ich genügend Zeit habe, würde ich es natürlich gerne selber machen, weil es natürlich eine persönliche Herausforderung ist, so einen Text zu übersetzen und auch spannend, und deshalb haben wir das Ganze ja studiert. Und natürlich auch der Kostenfaktor, je nachdem was bezahlt wird dafür. Ja.

Also du würdest lieber Humanübersetzen auch, weil du mehr verlangen kannst?

Ja, wenn ich dann auch genügend Zeit hätte. Wenn das jetzt nur so eine kurze Aufgabe zwischendrin ist und ich habe keine Zeit, dann sage ich „Ja passt, ich mache Post-Editing und geht schon."

Anhang D: Interviews nach Durchführen der Post-Editing-Aufgabe

*Teilnehmer*in 1 – Interview 2*

10.01.2022, 12:58, Zentrum für Translationswissenschaft

Hat dir die Arbeitsweise des Post-Editing gefallen und Spaß gemacht?

Ja hat es.

Sind während des Post-Editings Schwierigkeiten oder Probleme aufgetreten, die du bei einer Humanübersetzung desselben Textes nicht gehabt hättest?

Ja, man ist natürlich sehr an den maschinell übersetzten Text gebunden von der Satzstruktur. Also ich glaube, wenn ich das normal humanübersetzt hätte, hätte ich mir in manchen Fällen andere Satzstrukturen überlegt und auch so übersetzt dann.

Das heißt also du hast dich brav an die Vorgabe gehalten, so viel wie möglich von der Rohübersetzung zu übernehmen?

Ja genau, manche Sachen habe ich zum Beispiel stilistisch schon ausgebessert, aber ich habe geschaut, dass ich nicht zu viel ausbessere.

Welche Vorgehensweise oder Strategie hast du beim Post-Editing des Textes primär gewählt? Hast du zuerst den Ausgangs-, dann den Zieltext gelesen?

Ja genau, zuerst Ausgangstext dann Zieltext, dann habe ich mir angeschaut, ob das auch stimmig ist, grammatikalisch und auch stilistisch, und so habe ich dann ausgebessert.

Du bist also den Text von oben nach unten durchgegangen?

Ja, und manchmal bin ich auch zurückgegangen, wenn ich gemerkt habe „Ah gut, da wurde etwas so übersetzt und dann so“.

Damit es einheitlich bleibt?

Genau, damit es einheitlich bleibt. Da habe ich zum Beispiel bei einem Segment „Extremwerte“ einmal gehabt, und dann habe ich das auf „Extrempole“ umgetauscht. Ich musste da noch einmal zurückgehen, weil ich das ja schon davor so gehändert gehabt habe.

Der von dir soeben post-editierte Text war 361 Wörter lang. Von Studenten ist eine durchschnittliche Post-Editing Produktivität von 350 Wörtern pro Stunde zu erwarten. In Bezug auf diesen Wert (350 Wörter/Stunde): Wie schätzt du deine Produktivität bei der soeben absolvierten Aufgabe ein?

Wahrscheinlich durchschnittlich produktiv. Es gibt bestimmt Leute, die das schneller machen, ich arbeite generell immer langsamer.

Würdest du sagen du hast circa so diese 350 Wörter pro Stunde, also du hast eine Stunde gebraucht?

Ja vielleicht ein bisschen weniger, so 50 Minuten vielleicht.

Die durchschnittliche Wortanzahl pro Segment in diesem Text war 22,6. Wie viele Änderungen (Löschen, Hinzufügen, Verschieben oder Ersetzen von Wörtern) glaubst du hast du im Durchschnitt pro Segment vorgenommen?

Drei bis fünf Änderungen im Durchschnitt würde ich sagen.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich mit einem höheren kognitiven Aufwand verbunden, sprich, welche ist mental anstrengender? PE oder Humanübersetzung?

Das Post-Editing war auch anstrengend, aber wenn ich das humanübersetzt hätte, hätte ich auf jeden Fall länger gebraucht. Also ich bleibe dabei.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich schneller? PE oder Humanübersetzung?

Ja schon Post-Editing. Aber ich habe nicht damit gerechnet, dass die Sätze so lang sind. [lacht]

Angenommen du erhältst einen Übersetzungsauftrag und darfst entscheiden, ob du diesen mittels Humanübersetzung oder Post Editing von MÜ erledigen möchtest: Welche Arbeitsweise wählst du? Wieso hast du dich so entschieden?

Würde ich Post-Editing wählen.

Warst du zufrieden mit dem Text, den du produziert hast?

Nicht ganz, vielleicht bei einem Segment habe ich mir schwer getan. Segment 6 um genau zu sein. Da klingt es noch immer, da habe ich den Satz noch immer nicht so, wie ich ihn mir eigentlich gewünscht hätte. Ich habe dann einfach keine gescheite Lösung gefunden [lacht]. Genau.

Hat sich deine persönliche Meinung von Post-Editing nach Durchführen dieser Aufgabe geändert?

Ist gleichgeblieben.

Welche Gefahren entstehen deiner Meinung nach durch den Fortschritt von Post-Editing maschineller Übersetzungen für den Übersetzerberuf?

Nein, ich sehe es eher als Chance, als Chance da viel mehr in diese Richtung zu gehen in Zukunft.

Findest du ist das eine Chance, die zum Beispiel bei uns im Studium gut gelehrt und genutzt wird? Also über Post-Editing zu lernen?

Ja auf jeden Fall, und auch das zu üben sollte viel mehr integriert werden in die Kurse.

Hat dir das gefehlt im Masterstudium?

Schon eigentlich, es gibt keinen konkreten Kurs dazu, wenn macht man es nur in anderen Übungen.

Siehst du in zehn Jahren ein Zusammenleben von Mensch und Maschine bezüglich Post-Editing?

Schon ja. Nach meiner Masterarbeit habe ich ja auch gesagt und geschrieben, dass ich eigentlich ein Befürworter bin, dass Mensch und Maschine eigentlich zusammenharmonieren.

Teilnehmer*in 2 – Interview 2

12.01.2022, 13:27, Zentrum für Translationswissenschaft

Hat dir die Arbeitsweise des Post-Editing gefallen und Spaß gemacht?

Nein [lacht] nein, das hat überhaupt keinen Spaß gemacht. Am Anfang vielleicht schon noch, also so generell macht es schon Spaß, aber es macht dann keinen Spaß, wenn du weißt, dass du diesen Text nicht liegenlassen kannst und dir später nochmal drüber Sorgen machen kannst, was da für Baustellen waren teilweise. Es waren eh nicht viele, aber die die halt waren... Es ist einfach frustrierend, wenn du nicht und nicht auf irgendeine Lösung draufkommst, die dich zufriedenstellt. Ich bin mir sicher, wenn das irgendwer liest, dann denkt sich der „Bist du wahnsinnig, was hat die Person da geschrieben“. Aber ich bin einfach auf keine besser Lösung draufgekommen.

Wenn das jetzt nicht im Rahmen von diesem Experiment gewesen wäre, hättest du den Text liegengelassen und wärst später nochmal gekommen?

Ja, hundertprozentig, weil bei mir ist das meistens so, mir fällt dann, wenn ich an irgendwas anderes denke, eine bessere Formulierung ein und ich denke mir so „Ah ja, das kann ich hinschreiben“. Aber jetzt natürlich war halt nicht die Zeit da einfach.

Sind während des Post-Editings Schwierigkeiten oder Probleme aufgetreten, die du bei einer Humanübersetzung desselben Textes nicht gehabt hättest?

Also die Probleme im Text bestehen ja trotzdem, es ist nur halt so, wenn du schon einen Text vor dir hast, dass du dann nicht mehr wirklich an andere Lösungsweisen denkst, weil du dann schon so in diesem Satzbau drinnen bist, dass du den gar nicht mehr ändern würdest. Also vielleicht wären es, also die Probleme wären sicher da gewesen, vielleicht nur auf eine andere Weise. Ich hätte trotzdem nicht gewusst, wie ich es besser formulieren könnte, wahrscheinlich auch wenn ich es selber übersetzt hätte.

Hast du dich gefangen gefühlt durch das, was schon dagestanden ist?

Ja, ich glaube gar nicht, dass es bewusst ist, aber wenn du einmal den Text eben so liest und die Sätze schon so formuliert sind, dann kommst du vielleicht, kann sein, gar nicht mehr auf andere Lösungsansätze, denke ich mir. Das ist vielleicht in dem Sinn schwerer etwas zu post-editieren. Natürlichen waren die meisten Sätze eh okay, da waren jetzt keine großen Fehler, aber diese zwei drei Sätze, die mich halt so fertig gemacht haben, das wäre vielleicht einfach gewesen, wenn ich es selber übersetzt hätte.

Segment 6 war schwierig oder?

War das der mit „Rendezvous“?

Ja genau der mit „Rendezvous“ und „Verbot“.

Ja, schrecklich. Also wie gesagt, es waren meistens nicht mal die lexikalischen Sachen das Mega-Problem, weil dann schaust du halt eine andere Entsprechung, das geht dann

meistens irgendwie. Aber der Satzbau, den du halt im Deutschen nicht hundertprozentig nachmachen kannst, aber auch nicht wirklich hundertprozentig verändern kannst, so dass es dann Sinn macht, das ist so frustrierend manchmal. Ja [lacht] und das war halt bei dem Segment ganz schlimm.

Welche Vorgehensweise oder Strategie hast du beim Post-Editing des Textes primär gewählt?

Also ich habe mir als Erstes nur den Zieltext angeschaut, einfach weil ich als Erstes mal schauen wollte wie der deutsche Text so im Allgemeinen einfach ist. Dann bin ich glaube ich jedes Segment einzeln durchgegangen, wo ich mir den Zieltext zuerst angeschaut habe und dann den Ausgangstext, ob das auch Sinn macht. Also ich habe eigentlich mit dem Zieltext angefangen.

Und sonst war die Arbeitsweise recht linear, dass du von einem Segment zum nächsten gegangen bist?

Ja ich würde schon sagen, glaube ich schon.

Der von dir soeben post-editierte Text war 361 Wörter lang. Von Studenten ist eine durchschnittliche Post-Editing Produktivität von 350 Wörtern pro Stunde zu erwarten. In Bezug auf diesen Wert (350 Wörter/Stunde): Wie schätzt du deine Produktivität bei der soeben absolvierten Aufgabe ein?

Also ich würde eh sagen so genau eine Stunde. Ich habe einmal auf die Uhr geschaut, aber das weiß ich nicht. Also ich habe zweimal auf die Uhr geschaut, da war circa eine halbe Stunde vorbei, aber das war halt Mittendrinnen irgendwann, ich habe davor und danach noch ein bisschen Zeit gebraucht, also ich würde sagen eine Stunde müsste es gewesen sein. Glaube ich.

Die durchschnittliche Wortanzahl pro Segment in diesem Text war 22,6. Wie viele Änderungen (Löschen, Hinzufügen, Verschieben oder Ersetzen von Wörtern) glaubst du hast du im Durchschnitt pro Segment vorgenommen?

Ich glaube, wenn ich jetzt den gesamten Text hernehme, weil es waren ja recht viele Segmente, also einige Segmente, die ich genau so gelassen habe wie sie waren, würde ich vielleicht sagen drei bis fünf, dann aufgerechnet. Ja, so würde ich das verorten.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich mit einem höheren kognitiven Aufwand verbunden, sprich, welche ist mental anstrengender? PE oder Humanübersetzung?

Anstrengender ist sicher die Humanübersetzung, immer noch. Ja, weil trotzdem einige Sätze dabei waren, die ich wirklich so unterschrieben hätte, also ich meine sicher hätte ich es anders formuliert, wenn ich es jetzt selber gemacht hätte, aber das kommt einfach dadurch, dass ich ein Mensch bin [lacht] und dass das eine Maschine ist, aber so im Allgemeinen finde ich war es jetzt nicht so schlimm. Es war okay. [lacht]

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich schneller? PE oder Humanübersetzung?

Bei dem Text hätte ich länger gebraucht, wenn ich ihn selber übersetzt hätte. Ja.

Angenommen du erhältst einen Übersetzungsauftrag und darfst entscheiden, ob du diesen mittels Humanübersetzung oder Post Editing von MÜ erledigen möchtest: Welche Arbeitsweise wählst du?

Ich würde ihn trotzdem selber übersetzen, immer noch. Einfach wie gesagt weil ich dann vielleicht auf andere Lösungsansätze kommen würde eventuell und einfach ein besseres Gewissen hätte dabei, einfach wenn ich weiß, dass ich das selber gemacht habe.

Bei Post-Editing hättest du ein schlechteres Gewissen, weil...?

Weil ich Angst hätte, dass ich irgendetwas übersehen habe. Nicht, weil ich jetzt kein Vertrauen habe, dass es nicht funktioniert, sondern weil ich da eher das Gefühl hätte „Vielleicht habe ich irgendwas übersehen" was halt sonst nicht passiert wäre, wenn ich das von Anfang an selbst gemacht hätte und jedes Wort, wo ich mir unsicher war, nachgeschaut hätte. Wenn ich jetzt ein Wort sehe, was da steht, schaue ich vielleicht nicht nach, nicht unbedingt wenn ich mir denke das könnte passen und lese einfach drüber, das kann halt bei einer Übersetzung von mir selbst eigentlich nicht passieren.

Also ist es fehlendes Vertrauen in deine Post-Editing Kompetenz, die du ja nicht hast.

Ja, genau.

Würdest du es sinnvoll finden, dass Kurse zu Post-Editing bei uns in das Masterstudium aufgenommen werden?

Ja sicher, weil wir wissen alle, dass es ein immer größer werdender Teil wird von dem, was wir mal machen müssen, oder es ist halt überhaupt einfach sinnvoll, wenn man es kann. Also jetzt nicht nur weil man das als Beruf machen will, sondern auch einfach so ist es gut zu wissen, wie das funktioniert. Und wir hätten ja genug Übungen, in denen man das irgendwie einbauen könnte. Wäre schon super.

Du warst vorher neutral bis aufgeschlossen. Hat sich deine persönliche Meinung von Post-Editing nach Durchführen dieser Aufgabe geändert?

Nein, es ist immer noch gleich. Es war zwar schwerer, als ich es mir vorgestellt habe, aber es ist eh klar. Natürlich kannst du jetzt nicht irgendeinen Text nehmen, der super einfach ist, weil sonst hat das ja auch keinen Mehrwert. Es muss ja auch Stellen geben, die auch schwieriger sind. Ich habe mir natürlich gewünscht, dass es leichter wird [lacht], aber nein, von der Schwierigkeit war es eh angemessen würde ich sagen.

Welche Gefahren entstehen deiner Meinung nach durch den Fortschritt von Post-Editing maschineller Übersetzungen für den Übersetzerberuf?

Ja, vielleicht dass die Leute das noch weniger wertschätzen, wieviel Zeit man braucht dafür. Weil es ist jetzt schon so, dass sich viele denken „Ja du setzt dich halt hin, brauchst eine Stunde und dann kriegst du ur viel Geld dafür und hast im Endeffekt keine Denkarbeit geleistet, keine wirkliche Anstrengung undso dahinter", und ich glaube das wird dann halt noch weniger werden wahrscheinlich. Also dass sich Leute denken „Du kriegst eh einen fertigen Text, der ist wahrscheinlich eh ur gut, bist eh gleich fertig damit", aber so ist es halt einfach nicht. Vor allem wenn du irgendwelche terminologisch anspruchsvollen Sachen hast einfach. Oder es kann auch die Syntax

sein, wie man gerade gesehen hat. Ja, und sowas ist halt etwas, das wissen die meisten nicht. Aber woher auch.

Du meinst die Kunden würden denken du bist das Helferlein zur Maschine, aber in echt ist es eigentlich immer noch umgekehrt, die Maschine ist dein Helferlein. Weil du musst immer noch darauf schauen, dass das was rauskommt...

Du trägst immer noch die Verantwortung im Endeffekt. Und wie gesagt, gerade wenn es jetzt um unsere Sprachen geht, würde ich mal sagen die großen Sprachen Deutsch, Englisch, Französisch, vielleicht noch Spanisch, glaube ich ist es einfach unweigerlich, dass es immer mehr Leute machen werden das Post-Editing. Und da wäre es dann trotzdem nochmal wichtiger, dass man das auch lernt auf der Uni, und dass man sich dann auch sicher sein kann, dass man sich im Text die richtigen Sachen anschaut, die die halt wichtig sind und sich nicht an Kleinigkeiten aufhängt. Weil das wäre meine Sache gewesen, ich hätte da noch viel mehr editieren können aber habe es nicht gemacht, weil du mir gesagt hast am Anfang „Wenn es passt, dann soll man es lassen", aber ja ich hätte noch viel mehr machen können. Und das ist halt schwierig, da zu sagen wann ist es ein Fehler für dich, und wann ist es einfach etwas, das ich anders formulieren würde. Das ist halt schwierig, dass man das irgendwann lernt, oder hört mal irgendwann. Es ist halt auch ein sehr junges Tätigkeitsfeld, um da eine gescheite Ausbildung machen zu können, weil natürlich kannst du sagen wir machen heute Post-Editing, aber es gibt einfach wenig Erfahrungswerte irgendwie. Aber trotzdem wäre es halt gut, wenn man es irgendwo mal mitbekommen würde. [Pause] Die Uni Wien braucht unbedingt einen Post-Editing Kurs im Master Fachübersetzen. [beide lachen]

Teilnehmer*in 3 – Interview 2

13.01.2022, 15:48, Zentrum für Translationswissenschaft

Wie hat dir diese Aufgabe, die Arbeitsweise des Post-Editing gefallen?

Wie gesagt, es war mal wieder nett [lacht] zu übersetzen und an und für sich ist es ja doch gar nicht so leicht, weil man muss ja dann schauen, dass man halt auch nicht zu viel vom Text verändert, dass dann der persönliche Stil wieder mit einfließt undso.

Du hast gesagt es war nett mal wieder zu übersetzen: Ist Post-Editing für dich eine Art von Übersetzen?

Ja schon, in dem Sinne, dass man schon den geistigen Prozess, also man schreibt es jetzt nicht direkt alles hin was man sich denkt, aber der geistige Prozess ist ungefähr derselbe. Finde ich, für mich halt und ja, wie gesagt, selbst wenn man es nicht hinschreibt, im Kopf passieren ähnliche Vorgänge würde ich sagen.

Sind während des Post-Editings Schwierigkeiten oder Probleme aufgetreten, die du bei einer Humanübersetzung desselben Textes nicht gehabt hättest?

Ja, und zwar war ein Segment das sehr, wie sagt man, das nicht sehr grammatikalisch, also grammatikalisch schon, aber lexisch einfach irgendwie komisch war, auch von den Wörtern her, weil es einfach überhaupt nicht zusammengepasst hat.

Das Segment mit der Tendenz?

Ja, das sechste Segment, das hat vorne und hinten nicht gepasst.

Welche Vorgehensweise oder Strategie hast du beim Post-Editing des Textes primär gewählt?

Also ich habe mir zuerst einmal den Zieltext angeschaut und geschaut, ob das eben sprachlich auf Deutsch jetzt so stimmig ist, dass es relativ natürlich klingt. Und dann, wenn mir etwas ganz komisch vorgekommen ist, habe ich einmal mit dem Ausgangstext verglichen und habe mir das eben angeschaut und überlegt, wie ich das eben ausbessern könnte, damit es stimmiger wird.

Das heißt du hast mehr auf den Zieltext geschaut?

Im Prinzip schon.

Der von dir soeben post-editierte Text war 361 Wörter lang. Von Studenten ist eine durchschnittliche Post-Editing Produktivität von 350 Wörtern pro Stunde zu erwarten. In Bezug auf diesen Wert (350 Wörter/Stunde): Wie schätzt du deine Produktivität bei der soeben absolvierten Aufgabe ein?

Das hat jetzt circa, ja ich glaube ein bisschen mehr als eine Stunde gedauert, glaube ich.

Also ein bisschen über sechzig Minuten glaubst du?

Ja, ein bisschen, so siebzig Minuten.

Die durchschnittliche Wortanzahl pro Segment in diesem Text war 22,6. Wie viele Änderungen (Löschen, Hinzufügen, Verschieben oder Ersetzen von Wörtern) glaubst du hast du im Durchschnitt pro Segment vorgenommen?

Also im Durchschnitt, ich habe nicht in jedem Segment etwas geändert, weil die ganz kurzen Segmente haben eigentlich relativ gut gepasst, aber so pro Segment würde ich schon sagen so zwei Änderungen.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich mit einem höheren kognitiven Aufwand verbunden, sprich, welche ist mental anstrengender? PE oder Humanübersetzung?

Ich bräuchte zwar den direkten Vergleich jetzt mit einer Übersetzung, die ich selbst vor kurzem hergestellt hätte, aber ich würde sagen es ist eigentlich relativ gleich trotzdem.

Aber das, was die beiden Arbeitsweisen anspruchsvoll macht, sind verschiedene Dinge.

Ja, auf jeden Fall, weil du zwar eben den vorgegebenen Text hast, aber dir dann eben Gedanken machen musst, wie sagt man das auf eine natürliche Art. Weil normalerweise, wenn du das selber schon übersetzt, dann kommt es ja theoretisch schon natürlicher heraus, weil das ja deine Muttersprache ist. Und so musst du erst kurz überlegen, so „Hä? Da passt etwas nicht“ und dann „Was passt nicht?“.

Du hast vorher erwähnt du musstest dich zusammenreißen nicht zu viel zu ändern. Hast du dich ein bisschen gefangen gefühlt durch das, was schon dagestanden ist?

Schon. Es kommt glaube ich auf den Auftrag an, wie sehr, oder auch auf die Zeit, wie sehr man das ganze verändert, beziehungsweise wie sehr der Kunde will, dass man das eben natürlicher macht und zielsprachengemäßer, aber es stimmt schon, dass du

irgendwie ein bisschen gefangen bist, weil du denkst dir: „Ist das jetzt zu viel, was ich geändert habe? Ist das zu wenig? Kann ich das lassen oder muss ich das doch ändern?“.

Nach deiner persönlichen Einschätzung, in der Sprachrichtung Englisch → Deutsch: Welche der beiden Arbeitsweisen ist für dich voraussichtlich schneller? PE oder Humanübersetzung?

Ja ich glaube Post-Editing. Obwohl, wenn es da um einen echten Auftrag gegangen wäre, wäre ich das vielleicht schon zwei-, dreimal durchgegangen, vielleicht einmal drüber schlafen und liegen lassen, nicht in einer Sitzung alles machen.

Angenommen du erhältst einen Übersetzungsauftrag und darfst entscheiden, ob du diesen mittels Humanübersetzung oder Post Editing von MÜ erledigen möchtest: Welche Arbeitsweise wählst du?

Ich hätte es fast selber gemacht, weil stilistisch schon ein paar Sachen drinnen waren die mich echt gestört haben [lacht], die ich schon gerne selber ausgebessert hätte aber eben nicht gemacht habe.

Hat sich deine persönliche Meinung von Post-Editing nach Durchführen dieser Aufgabe geändert?

Ja schon ein bisschen, weil wie wir vorher gesagt haben, habe ich vorher gesagt, dass es vielleicht einfacher ist, aber es ist trotzdem eine Herausforderung auf anderer Ebene einfach.

Welche Gefahren entstehen deiner Meinung nach durch den Fortschritt von Post-Editing maschineller Übersetzungen für den Übersetzerberuf?

Ja, insofern als das in unserem Beruf die Preise so gedrückt werden, was eh schon der Fall ist, dass man kaum noch davon leben kann, beziehungsweise auch unseren Beruf eher so vom Niveau her abstuft, wenn man sagt „Ja das kann eh der Computer auch machen“, aber in Wirklichkeit, selbst beim Post-Editing findet man immer noch Sachen, die man verbessern könnte, auch wenn man es nicht unbedingt machen sollte, aber man kann immer einen noch besseren Text herstellen. Und deshalb ist es vielleicht schon so, dass Post-Editing uns eigentlich schlechter macht [lacht] als wir eigentlich sind.

Glaubst du wird es dadurch auch noch schwieriger, den eigenen Beruf gegenüber dem Kunden zu rechtfertigen?

Ja, ich denke schon. Weil es ist ja jetzt schon so, dass viele sagen „Ja, hätte ich eh selber auch machen können“, aber es ist halt einfach nicht so, dass nur weil man ein paar Brocken Englisch kann, dass man das einfach so machen kann.

Du hast gesagt du hast auch keinerlei Post-Editing-Ausbildung erhalten: Hättest du dir das gewünscht im Masterstudium?

Definitiv. Weil ich glaube das ist über unseren Studienzeitraum immer, also ich kann nicht sagen, wie es davor war, aber ich habe die Entwicklung auch in unserem Studienzeitraum bemerkt, eben auch als *DeepL* aufgekommen ist und das Ganze, dass das immer prominenter wird in unserer Berufssparte, und somit wäre es auch hilfreich, eben auch damit man gegenüber Menschen argumentieren kann, wenn eben Kunden kommen und sagen „Ja, ich tu das selber in DeepL, dann passt das eh“. Damit ich das

selber besser erklären kann, auch wo die Fehler liegen zum Beispiel bei DeepL, vielleicht auch den technischen Hintergrund erklären können. Die Maschine macht das nach einem Algorithmus und das ist nicht hundertprozentig fehlerfrei, was natürlich der Mensch auch nicht ist, aber ja. So halt, damit man sich auch besser rechtfertigen kann. Und auch damit man selber einmal kennenlernt, wie das Ganze funktioniert, und dann eben sagen kann: „Ja, ich koste aus dem Grund mehr, oder selbst wenn ich die Maschine nachbearbeite kostet es trotzdem mehr, weil es eben qualitativ hochwertiger ist, aus dem Grund, weil die Maschine den Fehler macht, den Fehler macht, den Fehler macht.“