universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Die Qualität der maschinellen Übersetzung im Rechtsbereich anhand einer Fehlertypologie am Beispiel Russisch-Deutsch“

verfasst von / submitted by

Kristina Dunina

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 070 360 331
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Translation Russisch Deutsch
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Dagmar Gromann

Inhaltsverzeichnis

Tabellenverzeichnis

Abbildungsverzeichnis

0 Einleitung

Die Zuverlässigkeit der maschinellen Übersetzungssysteme ist einer der wichtigsten Einflussfaktoren für die weitere Entwicklung der maschinellen Übersetzung, die unmittelbar den translatorischen Bereich betrifft. Auch wenn künstliche neuronale Netze die Qualität der maschinellen Übersetzung erheblich verbessert haben, bleibt die neuronale maschinelle Übersetzung (NMÜ) nicht fehlerfrei. Um festzustellen, welche Systeme für welche Sprachen und/oder Textsorten geeignet sind, ist es notwendig, diese Systeme regelmäßig zu testen, um die Qualität maschinell erstellter Übersetzungen zu analysieren. Aus diesem Grund gewinnt die Evaluierung der maschinellen Übersetzung von Jahr zu Jahr mehr an Bedeutung. Diese Tendenz lässt sich ebenso durch die zunehmende Nutzung der maschinellen Übersetzung in verschiedenen Bereichen des Lebens erklären.

Die vorliegende Arbeit setzt sich mit der Qualität der maschinellen Übersetzung mit dem Schwerpunkt Rechtsübersetzung auseinander. Im Mittelpunkt der Arbeit steht folgende Frage: Welche Fehlerarten treten in der maschinellen Übersetzung von Rechtstexten vom Russischen ins Deutsche am häufigsten auf? Es wird angenommen, dass in den Übersetzungen von neuronalen maschinellen Übersetzungssystemen Terminologiefehler sehr oft vorkommen, da die korrekte Terminologie eine große Herausforderung nicht nur für die NMÜ, sondern auch für die Rechtsübersetzung an sich darstellt. Da die richtige Terminologiewahl eine entscheidende Rolle in der Bewertung der Übersetzungsqualität spielt (Ahrend 2006: 37), wird vermutet, dass die NMÜ ungenügende Qualitätsergebnisse liefern wird. Diese Grundannahmen werden untersucht.

Die Arbeit setzt sich zum Ziel, die Methoden und Ansätze zur Bewertung der maschinellen Übersetzungsqualität zu beschreiben, die Komplexität der Rechtsübersetzung zu erläutern, die Qualität von neuronalen maschinellen Übersetzungssystemen in dem Sprachpaar Russisch-Deutsch zu bewerten, typische Fehler der neuronalen maschinellen Übersetzung festzustellen und auf der Grundlage der analysierten Fehler sowohl Translator*innen, die NMÜ verwenden, als auch den Entwickler*innen der NMÜ-Systeme Empfehlungen für den Einsatz und Verbesserungsmöglichkeiten der NMÜ zu geben.

Obwohl die Übersetzungsqualität ein zentrales Konzept in der Translationswissenschaft ist, gibt es in der Übersetzungstheorie noch keine allgemein akzeptierte Definition dieses Begriffs, da viele verschiedene Ansätze vorgestellt wurden (z. B. Reiß & Vermeer 1984; Nord 1993; Schmitt 1999; Nida 2001; Ahrend 2006; Nord 2006; House 2014; Koby et al. 2014 etc.), die sich mit diesem Thema beschäftigen. Dies hat zur Folge, dass über die Jahre verschiedene

Qualitätsansätze erschienen sind, welche die Übersetzungsqualität aus unterschiedlichen Perspektiven betrachten und bewerten. Zwar gibt es zahlreiche Studien (Killman 2014; Azer & Aghayi 2015; Wiesmann 2019; Vieira et al. 2021; Welnitzova et al. 2021), die maschinelle Übersetzungsqualität evaluieren, allerdings werden häufig andere Sprachkombinationen untersucht. Es werden hauptsächlich maschinelle Übersetzungen ins Englische oder aus dem Englischen in eine romanische Sprache oder auch Übersetzungen im Sprachpaar Englisch-Deutsch analysiert. Das Sprachpaar Russisch-Deutsch wird dadurch nicht selten vernachlässigt.

Killman (2014) befasst sich mit der Genauigkeit des statistischen maschinellen Übersetzungssystem Google Translate. Der Wissenschaftler untersucht aus dem Spanischen ins Englische maschinell übersetzte Rechtsbegriffe und wählt eine Evaluierungsmethode nach zwei Kriterien: Flüssigkeit und Adäquatheit (engl. „fluency", „adequacy") (Killman 2014: 94). Azer und Aghayi (2015) evaluieren maschinell erstellte Übersetzungen von sechs unterschiedlichen Texten (darunter ein Rechtstext) aus dem Englischen ins Persische. In dieser Studie wird die Übersetzungsqualität von den MÜ-Systemen Google Translate und Padideh mittels der sechs folgenden Evaluierungskategorien gemessen: Verständlichkeit, Wiedergabetreue, Kohärenz, Benutzerfreundlichkeit, Annehmbarkeit und Lesezeit (Azer & Aghayi 2015: 228f.). Wiesmann (2019) beschäftigt sich mit der neuronalen MÜ von Rechtstexten und untersucht, inwieweit das NMÜ-System DeepL und das mit der MÜ ausgestattete CAT-System MateCat für Rechtsübersetzungen im Sprachpaar Italienisch-Deutsch nützlich sein können (Wiesmann 2019: 120). Wiesmann (2019) führt eine manuelle Evaluierung durch, die sich auf zwei Hauptkriterien bezieht: die Verständlichkeit und Aussagekraft des Zieltextes und die Übereinstimmung zwischen dem Ausgangs- und Zieltext (Wiesmann 2019: 118). Vieira et al. (2021) setzen sich mit den Auswirkungen unkontrollierter Einsätze der MÜ in den Bereichen Medizin und Recht auseinander. Auf der Grundlage einer kritischen Literaturrecherche analysieren Vieira et al. (2021) themenrelevante offizielle Dokumente und Forschungsarbeiten. Welnitzova et al. (2021) befassen sich mit der Qualität der maschinellen Übersetzung von Rechtstexten im Sprachpaar Slowakisch-Englisch und analysieren aufgetretene Fehler. Dabei wird eine vom MÜ-System Google Translate maschinelle Übersetzung eines 16-seitigen Strafrechtskodexes mit einer menschlichen Übersetzung verglichen und anhand der Multidimensionalen Qualitätsmetrik (MQM) evaluiert (Welnitzova et al. 2021: 224).

Im Rahmen der vorliegenden Masterarbeit werden vier Rechtstexte bzw. Textabschnitte aus verschiedenen Rechtsthematiken für die Untersuchung ausgewählt. Die ausgesuchten Ausgangstexte werden maschinell übersetzt. Für die Erstellung der maschinellen

Übersetzungen wird der Online-Übersetzter DeepL verwendet, dessen MÜ-System auf der Technologie der neuronalen Netze basiert. Die Qualität der maschinell übersetzten Texte wird nach der Norm SAE-J2450 (SAE 2001) manuell evaluiert. Aufgetretene Fehler werden von drei ausgewählten Annotator*innen kategorisiert. Schließlich wird die Übereinstimmung der Annotator*innen mittels des Kappa-Wertes berechnet.

Die vorliegende Untersuchung ist für die Optimierung der Qualität maschineller Übersetzungen, die mithilfe von NMÜ-Systemen angefertigt werden, und insbesondere für die Sprachrichtung Russisch-Deutsch, relevant. Es ist notwendig, die Qualitätsdynamik der MÜ-Systeme regelmäßig zu prüfen und die Methoden zur Qualitätsbewertung zu verbessern, um die MÜ-Systeme weiterzuentwickeln. Daraus ergibt sich die Aktualität der Arbeit. Darüber hinaus ist die geplante Untersuchung ebenso für die Testung des neuronalen maschinellen DeepL-Übersetzungssystems von Bedeutung, weil dadurch untersucht wird, ob dieses NMÜ-System für die Übersetzung von Texten aus der Rechtsthematik in der oben erwähnten Sprachkombination geeignet ist bzw. welche Herausforderungen für die Technologie in diesem Bereich bestehen.

Die vorliegende Masterarbeit ist in acht Kapitel unterteilt. Die Kapitel 1, 2 und 3 befassen sich mit den theoretischen Grundlagen des Themas dieser Arbeit. Kapitel 1 gibt einen kurzen Überblick über die Geschichte der maschinellen Übersetzung. Des Weiteren werden in diesem Kapitel verschiedene Arten der MÜ erläutert, es wird auf ihre Vor- und Nachteile eingegangen und die meistverbreiteten MÜ-Systeme werden vorgestellt. Kapitel 2 thematisiert die Evaluierung der MÜ. Dabei werden das Konzept der Übersetzungsqualität und die Möglichkeiten einer automatischen und manuellen Evaluierung von maschinell erstellten Übersetzungen diskutiert. Kapitel 3 ist der Problematik der Rechtsübersetzung gewidmet. Im Fokus dieses Kapitels stehen Rechtstexte und die besonderen Eigenschaften der Rechtssprache. Die für das Thema dieser Masterarbeit relevanten und aktuellen Studien werden in Kapitel 4 vorgestellt. Kapitel 5 beschäftigt sich mit dem praktischen Teil der Arbeit. In diesem Zusammenhang wird das Forschungsdesign, nämlich der Gegenstand und die genaue Vorgehensweise der Untersuchung sowie die Auswertungsmethode, detailliert präsentiert. In dem anschließenden Kapitel wird das Untersuchungskonzept umgesetzt und somit die beschriebene Untersuchungsmethode angewendet (Kapitel 6). Zugleich werden die Evaluierungsergebnisse ausführlich dargestellt. Kapitel 7 dient der Analyse und der Interpretation der Untersuchungsergebnisse sowie der Beantwortung der Forschungsfrage. An dieser Stelle werden die Problembereiche der Untersuchung diskutiert und Vorschläge für weitere Forschungen vorgelegt. Darüber hinaus werden auf der Grundlage der Analyse

Empfehlungen sowohl für Translator*innen, die NMÜ-Systeme verwenden, als auch für die Entwickler*innen solcher Systeme gegeben. Die Arbeit wird mit einer Zusammenfassung der gewonnenen Erkenntnisse abgeschlossen (Kapitel 8).

1 Maschinelle Übersetzung

Das vorliegende Kapitel liefert einen Überblick über die Geschichte der maschinellen Übersetzung (MÜ) und befasst sich mit der Definition der MÜ. Zudem werden in diesem Kapitel verschiedene Arten der MÜ sowie deren Vor- und Nachteile erläutert und die bekanntesten kommerziellen, frei verfügbaren MÜ-Systeme vorgestellt.

1.1 Geschichte und Definition der maschinellen Übersetzung

Die Beschäftigung mit der Geschichte der maschinellen Übersetzung (MÜ) ist praktisch und theoretisch relevant. Sie dient der Orientierung in der Gegenwart, damit der Fortschritt der Wissenschaft verfolgt werden kann. Bevor die Geschichte der MÜ in diesem Kapitel erläutert wird, ist es sinnvoll, diesen Begriff zu definieren.

Das Automatic Language Processing Advisory Committee (ALPAC) definiert die maschinelle Übersetzung wie folgt: „‚Machine Translation' presumably means going by algorithm from machine-readable source text to useful target text, without recourse to human translation or editing" (ALPAC 1966: 19). Haverkort (1991) äußert sich folgendermaßen zum Thema: „Nur bei der maschinellen Übersetzung wird die Übersetzung eigenständig vom Computer angefertigt. Diese Eigenständigkeit schließt übrigens nicht aus, dass die vom Computer erzeugten Rohübersetzungen noch nachredigiert werden müssen" (Haverkort 1991: 9).

In der vorliegenden Masterarbeit wird die Definition von Zimmermann (2004: 475) verwendet. Unter maschineller Übersetzung versteht Zimmermann (2004: 475) „die vollautomatische Übersetzung eines Textes in natürlicher Sprache in eine andere natürliche Sprache". Darüber hinaus findet man bei Zimmermann (1990) folgende Definition des maschinellen Übersetzungssystems:

> Ein maschinelles Übersetzungssystem (MT-System: Machine Translation) liegt nur dann vor, wenn der Übersetzungsprozess – ausgehend von einem maschinenlesbaren Quelltext – ohne menschliche Hilfe abläuft und zu einem Zieltext führt, dessen Qualität für Informationszwecke mindestens „gut genug" ist. (Zimmermann 1990: 265)

Nachdem somit der Begriff „maschinelle Übersetzung" (MÜ) erklärt wurde, kann nun auf die Geschichte der MÜ eingegangen werden.

1.1.1 Die ersten Schritte und Erfolge der maschinellen Übersetzung

Die ersten Überlegungen zur maschinellen Übersetzung gab es bereits im 17. Jahrhundert. Allerdings wurde die Umsetzung der Ideen erst mit dem Erscheinen eines elektromechanischen Computers Mitte der 1940er Jahre denkbar (Hutchins 1997a: 14). Man kann nicht eindeutig sagen, wer zuerst auf die Idee einer von Computern durchgeführten Übersetzung menschlicher Sprachen kam. Wenn man sich jedoch mit der Geschichte der maschinellen Übersetzung genauer auseinandersetzt, stößt man auf Gespräche und Korrespondenz zwischen Andrew D. Booth, Norbert Wiener und Warren Weaver (Weaver 1947). Im März 1947 schrieb Warren Weaver an Norbert Wiener und traf sich mit Andrew Booth, woraufhin er beide auf die Möglichkeit des Einsatzes von Computern für Übersetzungen hinwies (Hutchins 1997b: 196). Am 4. März desselben Jahres schrieb Warren Weaver, der zur damaligen Zeit der Direktor der naturwissenschaftlichen Abteilung der Rockefeller Foundation war, folgende Nachricht an seinen Freund, den Mathematiker Norbert Wiener, der sich mit Kybernetik beschäftigte:

> Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while. (Weawer 1947: 1)

Weaver hatte Wiener während des Krieges kennengelernt, als beide in der militärischen Forschung tätig waren: Weaver beschäftigte sich mit Ballistik und Wiener mit Radar und Vorhersagetheorie (Hutchins 1997c: 22).

Zwei Jahre danach, im Juli 1949, schrieb Weaver das Memorandum (Weaver 1955). Das war die erste Publikation, die Forscher*innen in Amerika und Westeuropa bekannt war, in welcher eine Möglichkeit einer durch einen modernen elektronischen Computer durchgeführten Übersetzung vorgeschlagen wurde (Schwartz 2018: 165). Weavers Ziel war es, die bereits geleistete Arbeit zusammenzufassen und die Richtungen vorzuschlagen, in welche die Forschung in diesem Bereich gehen könnte (Hutchins 1997b: 205). Der Forscher gab zu, dass die maschinelle Übersetzung literarischer Texte eine zu große Herausforderung sein könnte, behauptete jedoch, dass maschinelle Übersetzungen generell für die Übersetzung technischer Dokumente immer noch nützlich sein könnten (Schwartz 2018: 165). Dieses Memorandum löste ein großes Forschungsinteresse aus, und Anfang der 1950er Jahre gab es eine große Anzahl an Forschungsgruppen, die in Europa und den USA sich mit dem Thema MÜ beschäftigen, was eine beträchtliche finanzielle Investition bedeutete (Arnold et al. 1994: 13).

Yehoshua Bar-Hillel wurde 1951 vom Massachusetts Institute of Technology (MIT) als erster Vollzeitforscher in MÜ angestellt (Liu & Zhang 2015: 106). Im Jahr 1952 hielt er die erste Konferenz über maschinelle Übersetzung am MIT ab. Diese Konferenz brachte Wissenschaftler, die bereits mit der MÜ in Kontakt standen, mit denjenigen zusammen, die an weiterer Forschung auf diesem Gebiet interessiert sein konnten (Liu & Zhang 2015: 106).

1954 gab es schon die erste öffentliche Demonstration der Durchführbarkeit maschineller Übersetzung: Leon Dostert von der Georgetown University und Peter Sheridan von der International Business Machines Corporation (IBM) nutzten die Maschine IBM701, um eine maschinelle Übersetzung öffentlich zu demonstrieren (Chan 2015: 3). Die Maschine verwendete nur 250 Wörter und 6 Grammatikregeln für die Übersetzung von 49 Sätzen aus dem Russischen ins Englische (Hutchins 2004: 102; Liu & Zhang 2015: 106). Obwohl die getestete Maschine nur ein sehr begrenztes Vokabular und sehr einfache Grammatik verwendete, waren die Ergebnisse sehr überzeugend (Hutchins 2000a: 1060). Dies führte zu einer umfassenden Finanzierung der MÜ-Projekte in den USA und gab den Anreiz für weitere Forschungen in diesem Bereich weltweit (Hutchins 2000a: 1060).

1.1.2 Frustration und der ALPAC-Bericht

Allerdings löste diese öffentliche Demonstration eines MÜ-Systems nicht nur ein großes öffentliches Interesse, sondern auch eine große Kontroverse aus (Hutchins 2004: 102). Denn obwohl das Übersetzungsresultat bezüglich des Informationsbedarfs für viele Forschungsteilnehmer*innen zufriedenstellend war, wiesen die untersuchten MÜ-Systeme eine schlechte Übersetzungsqualität auf (Hutchins 2000a: 1060). Die Enttäuschung wuchs, als die Forscher*innen auf semantische Probleme stießen, für die es keine einfachen Lösungen gab (Hutchins 2000a: 1060). Die Sponsor*innen waren frustriert und zweifelten an der Möglichkeit, den Übersetzungsprozess zu automatisieren (Arnold et al. 1994: 13).

Bar-Hillel kritisierte die Theorie der MÜ und erklärte in seinem Bericht (1959), dass eine vollautomatische qualitativ hochwertige MÜ unmöglich sei (Bar-Hillel 1959: 38). Als Grundlage des Berichts dienten seine Besuche mehrerer wichtiger Forschungszentren der MÜ in den USA und persönliche Gespräche mit den Mitgliedern der Forschungsgruppen. Einige Forschungsgruppen stellten Informationen über die Namen und Anzahl der Forscher*innen, ihren Hintergrund und ihre Qualifikationen, das Budget und zukünftige Pläne der Forschung in diesem Bereich zu Verfügung (Bar-Hillel 1959: 2). Bar-Hillels Hauptanliegen war es, zu begründen, dass der Großteil der MÜ-Forschung auf dem falschen Weg sei:

> It was not sufficiently realized that the gap between such an output, for which only with difficulty the term „translation“ could be used at all, and high-quality translation proper, i.e., a translation of the quality produced by an experienced human translator, was still enormous, and that the problems solved until then were indeed many but just the simplest ones, whereas the „few“ remaining problems were the harder ones — very hard indeed. (Bar-Hillel 1959: 4)

Hutchins (2000b: 307) bezeichnete Bar-Hillels Bericht (Bar-Hillel 1959) als eine der einflussreichsten Publikationen im Bereich der MÜ.

Bereits 1964 waren die Sponsor*innen der US-Regierung angesichts des mangelnden Fortschritts besorgt und gründeten das Automatic Language Processing Advisory Committee (ALPAC) (ALPAC 1966; Arnold et al. 1994: 13; Hutchins 2000a: 1060; Chan 2015: 3). Die Kommission bestand aus sieben Experten, die den Stand der maschinellen Übersetzung untersuchen sollten (ALPAC 1966; Chan 2015: 4). In dem berühmten ALPAC-Bericht (1966) kam die Kommission zu dem Schluss, dass die maschinelle Übersetzung langsam, nicht ganz genau und teuer ist und „there is no immediate or predictable prospect of useful machine translation“ (ALPAC 1966: 32). Die ALPAC-Kommission sah keinen Bedarf für weitere Investitionen in die MÜ-Forschung in den USA und dies führte dazu, dass die staatliche Finanzierung in den USA zum Stillstand kam (Arnold et al. 1994: 13; Hutchins 2000a: 1060; Chan 2015: 4). Das bedeutete das Ende der MÜ-Forschung in den USA für mehr als ein Jahrzehnt und hatte große Auswirkungen auch in der Sowjetunion und in Europa (Hutchins 2000a: 1060).

Arnold et al. (1994) sind überzeugt, dass die Schlussfolgerung des Berichts missverstanden wurde (Arnold et al. 1994: 14). Laut den Forscher*innen scheint die Annahme, dass es keine Perspektiven für eine erfolgreiche MÜ gebe, auf einer engen Vorstellung der hochqualitativen vollautomatischen maschinellen Übersetzung beruht zu haben (Arnold et al. 1994: 14). Arnold et al. (1994) sind davon überzeugt, dass diese Annahme nicht sehr aussagekräftig ist, wenn man die durch MÜ gewonnene Zeit des Übersetzungsprozesses in Betracht zieht (Arnold et al. 1994: 14). Dennoch ist die Zahl der an der MÜ-Forschung beteiligten Gruppen und Personen dramatisch gesunken (Arnold et al. 1994: 14). Die MÜ-Forschung wurde jedoch in Kanada, Frankreich und Deutschland fortgesetzt (Hutchins 2000a: 1060).

1.1.3 Die Rückkehr der maschinellen Übersetzung: die 1970er und 1980er Jahre

In den 1970er Jahren wurde MÜ aus unterschiedlichen Gründen und für unterschiedliche Sprachen wieder gefragt. Die wichtigsten Gründe dafür waren die administrativen und kommerziellen Erfordernisse mehrsprachiger Gemeinschaften und der multinationale Handel

(Hutchins 2000a: 1060). Der kanadischen Regierung war ihre bikulturelle Politik von großer Bedeutung, daher waren englisch-französische Übersetzungen erforderlich (Hutchins 2000a: 1060). In den europäischen Ländern gab es eine wachsende Nachfrage nach Übersetzungen aus und in alle Sprachen der Europäischen Gemeinschaft (Liu & Zhang 2015: 107). Gesucht wurde in erster Linie nach kostengünstigen maschinengestützten Übersetzungssystemen, die technische Dokumentationen in wichtigen Sprachen des internationalen Handels übersetzen könnten (Hutchins 2000a: 1060). Peter Toma, als Mitglied des Georgetown-Projekts, entwickelte System Translation (SYSTRAN) für den operationellen Einsatz bei der United States Air Force (USAF) und der National Aeronautics and Space Administration (NASA) (Toma 1977). Danach wurde SYSTRAN von der Kommission der Europäischen Gemeinschaften (KEG) erworben und eingesetzt (Toma 1997: 34). Dank der KEG wurde der Grundstein für das EUROTRA-Projekt gelegt, das auf den Entwicklungen der Gruppe SUSY (Saarbrücker Übersetzungssystem) basierte (Chan 2015: 25). In Kanada erschien das METEO-System (Chevalier et al. 1978) zur Übersetzung von Wetterberichten, das von der TAUM-Gruppe an der Universität Montreal (Traduction Automatique à l'Université de Montréal) entwickelt wurde (Arnold et al. 1994: 14; Hutchins 2000a: 1060).

Ab Mitte der 1970er Jahre wurde auch die MÜ-Forschung in den Vereinigten Staaten aufgenommen. Die Panamerikanische Gesundheitsorganisation (PAHO) begann mit der Entwicklung des spanisch-englischen MÜ-Systems SPANAM (Vasconcellos & León 1985). Die USAF finanzierte die Arbeit am METAL-System (Bennett & Slocum 1985) am Linguistics Research Center an der Universität von Texas in Austin (Arnold et al. 1994: 14). Die Mormonenkirche in Provo, Utah (USA) war an Bibelübersetzungen interessiert und finanzierte MÜ-Forscher*innen an der Brigham Young University (Arnold et al. 1994: 14). Die Forschung führte schließlich zu den Systemen WEIDNER (Hundt 1982) und ALPS. In Europa waren die GETA-Gruppe (Vauquois 1979) in Grenoble und die SUSY-Gruppe (Maas 1977) in Saarbrücken bekannt. Neue Computertechnologie der späten 70er Jahre verbesserte auch die maschinelle Übersetzung.

Die Forschung in den 1980er Jahren hatte drei Schwerpunkte: Transfersysteme, neue Arten der Interlinguasysteme und die korpusbasierte MÜ-Forschung (Liu & Zhang 2015: 107). Die breite Verfügbarkeit von Mikrocomputern und Textverarbeitungssoftware öffnete einen kommerziellen Markt für günstigere MÜ-Systeme nicht nur in den USA und Europa (ALPS, Weidner, Linguistic Products, Tovna und Globalink etc.), sondern auch in Japan (Sharp, NEC, Oki, Mitsubishi und Sanyo) (Hutchins 2000a: 1060). Es erschienen weitere

mikrocomputergestützte Systeme in Asien, Osteuropa und der Sowjetunion (Hutchins 2000a: 1060).

Ende der 1980er Jahre arbeiteten Peter Brown und seine Kolleg*innen der IBM-Forschungsgruppe ein Konzept der statistischen maschinellen Übersetzung (SMÜ) aus und präsentierten die Testergebnisse ihres SMÜ-Systems (Brown et al. 1988). In der Geschichte der MÜ war dies ein wichtiger Wendepunkt, da die statistische MÜ die regelbasierte MÜ dadurch ablöste.

Japanische Forschungsgruppen begannen korpusbasierte Methoden zu verwenden (Hutchins 2000a: 1061). Nagao (2003) schlug die Idee der beispielbasierten maschinellen Übersetzung (BBMÜ) vor. Damals bezeichnete der Forscher diese Art der MÜ als „machine translation by example-guided inference“ sowie „machine translation by the analogy principle“ (Nagao 2003). Zur damaligen Zeit begannen ebenso erste Forschungsansätze im Bereich der Übersetzung gesprochener Sprache, die sich mit der Integration von Spracherkennung und -synthese sowie Übersetzungsmodulen beschäftigten (Hutchins 2000a: 1061).

Zusammenfassend lässt sich sagen, dass die Systeme, die in den späten 70er Jahren auf dem Markt waren, ihre Vor- und Nachteile hatten. Einer Reihe von Forschungsprojekten, die zu diesem Zeitpunkt begonnen wurden, ist es gelungen, gut funktionierende, kommerzielle Systeme zu schaffen, was bedeutete, dass die MÜ sich als Forschungsgebiet und Technologieanwendung gut etabliert hat (Arnold et al. 1994: 15). Die Projekte, die in den 70er und 80er Jahren begonnen wurden, entwickelten sich zu vollständigen kommerziellen Systemen.

1.1.4 Weiterentwicklung der vorhandenen Übersetzungssysteme: 1990er Jahre

In den 1990er Jahren, mit dem Aufkommen und der Entwicklung des Internets, wurde es möglich, maschinelle Übersetzungssysteme weiterzuentwickeln. Die Forschung wurde im Bereich der korpusbasierten und regelbasierten maschinellen Übersetzung fortgesetzt. Einige Forschungsgruppen vor allem in Asien beschäftigen sich mit hybriden Ansätzen, welche die Kombination verschiedener MÜ-Ansätze in einem MÜ-System ermöglichen (Hutchins 2000a: 1061). Während die MÜ-Forschung in Russland und Osteuropa aufgrund der politischen Situation fast zum Stillstand kam, erlebte die Forschung in den USA einen Aufschwung (Hutchins 2000a: 1061). Allerdings wurden in Russland MÜ-Produkte wie PROMT und STYLUS angeboten (Sokolova 1994; 1999). Zu Beginn entwickelte das Unternehmen PROMT hauptsächlich die MÜ-Technologie, die das Herzstück der PROMT-Produkte bildete. Später

wurde von PROMT eine vollständige Palette von Übersetzungslösungen angeboten: maschinelle Übersetzungssysteme und -dienste, Wörterbücher, Translation-Memory-Systeme, Data-Mining-Systeme (Chan 2015: 7). In Westeuropa wurden Arbeitsstellen für professionelle Translator*innen geschaffen. Zu den wichtigen Schwerpunkten gehörten die Entwicklung von sprachkontrollierten und domänenbeschränkten Systemen (Hutchins 2000a: 1061). In Asien nahm die Forschung in Japan, China, Taiwan und Korea weiter zu (Hutchins 2000a: 1061).

Ende der 1990er beschäftigte sich die MÜ-Forschung mit der Erweiterung des wortbasierten SMÜ-Modells. Infolgedessen wurde ein neues SMÜ-Modell entwickelt, das nicht nur einzelne, sondern auch zusammengehörige Wörter erkennen und übersetzen konnte (Schwartz 2018: 177). Dieses neue SMÜ-Modell wurde als phrasenbasierte SMÜ bezeichnet (Koehn et al. 2003). Das erste phrasenbasierte SMÜ-System hieß Pharaoh (Koehn 2004) und bestand lediglich aus einem phrasenbasierten Decoder, was das Trainieren der Übersetzungsmodelle unmöglich machte (Schwartz 2018: 177). Des Weiteren war die Entwicklung einer größeren Anzahl von Funktionen sowie die Unterstützung unterschiedlicher Dokumentformate und zusätzlicher Sprachen für diesen Zeitraum charakteristisch (Chan 2015: 8).

1.1.5 Neue Entwicklungen: die Forschung der 2000er Jahre bis heute

Anfang der 2000er Jahre wurde das phrasenbasierte SMÜ-System Pharaoh (Koehn 2004) durch ein verbessertes System namens Moses (Koehn et al. 2007) ersetzt, das in der Lage war, Übersetzungsmodelle mittels der phrasenbasierten SMÜ-Techniken zu trainieren (Koehn et al. 2007: 177). Mitte der 2000er Jahre löste die phrasenbasierte SMÜ die wortbasierte SMÜ fast vollständig ab (Schwartz 2018: 177). Die MÜ-Forschung konzentrierte sich darauf, ob es möglich wäre, statistische und regelbasierte Methoden zu kombinieren, sodass syntaktische und grammatische Regeln in ein statistisches maschinelles Übersetzungssystem integriert werden können (Schwartz 2018: 178).

Gleichzeitig wurde versucht, die MÜ anhand neuronaler Netze durchzuführen. Dabei ging es darum, dass die Wortbedeutung indirekt über einen sprachunabhängigen hochdimensionalen Vektor dargestellt wird (Schwartz 2018: 178). Obwohl die Integration neuronaler Netze in maschinelle Übersetzungssysteme der MÜ-Forschung neue Impulse gab, wurden neuronale Netze anfangs nur oberflächlich eingesetzt, da sie nur wenig erforscht waren (Stahlberg 2020: 343). Ab Mitte der 2000er Jahre wurden rekurrente neuronale Netze vorgestellt (Kalchbrenner & Blunsom 2013). Allerdings wurden neuronale Netze zu diesem

Zeitpunkt nur als Komponenten eines klassischen SMÜ-Systems betrachtet (Stahlberg 2020: 343).

Einige Jahre später schaffte es die neuronale maschinelle Übersetzung (NMÜ), sich von der SMÜ zu trennen, was für die Entwicklung der MÜ einen enormen Fortschritt bedeutete. Nun verwendeten NMÜ-Systeme ausschließlich neuronale Netze, die den Ausgangssatz direkt in den Zielsatz umwandeln konnten (Stahlberg 2020: 343). Die Evaluierung der MÜ wird zu einem wichtigen Bereich in der MÜ-Forschung (Liu & Zhang 2015: 109). Aufgrund des Globalisierungsprozesses weltweit steigt der Bedarf an verschiedenen MÜ-Anwendungen wie Speech-to-Speech-Translation, Fotoübersetzung, Webübersetzung etc. (Liu & Zhang 2015: 109f.).

1.2 Arten der maschinellen Übersetzung

Es gibt unterschiedliche maschinelle Übersetzungsstrategien. Einer der ersten Ansätze zur Entwicklung der maschinellen Übersetzung ist die regelbasierte maschinelle Übersetzung (RBMÜ). Die statistische maschinelle Übersetzung (SMÜ) ist eine weitere MÜ-Strategie, bei der die Übersetzung aus statistischen Modellen generiert wird, deren Parameter durch die Analyse eines zweisprachigen Textkörpers bestimmt werden. Es erschienen neue Technologien, die auf der Verwendung neuronaler Netze basieren. Die neuronale maschinelle Übersetzung (NMÜ) ist ein einziges großes neuronales Netzwerk (mit Millionen von künstlichen Neuronen), das den gesamten MÜ-Prozess modelliert (Kalchbrenner & Blunsom 2013). In diesem Kapitel werden die Arten der MÜ genauer beschrieben.

1.2.1 Regelbasierte maschinelle Übersetzung

Regelbasierte maschinelle Übersetzung (RBMÜ) ist der klassische Ansatz in der MÜ. Vauquois (1976) beschäftigte sich mit den Methoden und Prozessen der RBMÜ und fasste sie in einem Dreieck zusammen, was als das Vauquois-Dreieck bezeichnet wird (Vauquois 1976: 131). Das Vauquois-Dreieck (1976) beschreibt die regelbasierte maschinelle Übersetzung als einen Prozess, der drei wichtige Phasen umfasst: Analyse, Transfer und Generierung (Vauquois 1976: 130f.). Ein regelbasiertes System besteht grundsätzlich aus bestimmten Regeln und einem Wörterbuch, das morphologische und semantische Informationen beinhaltet (Lagarda et al. 2009: 2017). Diese Regeln werden von Fachexpert*innen bzw. Linguist*innen geschrieben. Die Ergebnisse unterscheiden sich je nach dem Sprachpaar und dem Schwierigkeitsgrad des Textes (Stein 2009: 8). Für die regelbasierten Systeme der MÜ ist die Zahl der Zielsprachen

von Bedeutung, da dadurch der Ablauf des Übersetzungsprozesses bestimmt wird (Haverkort 1991: 9).

1.2.1.1 Direkter Ansatz

Wenn es nur darum geht, Übersetzungen aus einer Sprache in eine andere wortwörtlich durchzuführen, werden bestimmte charakteristische Merkmale der Zielsprache bei der Analyse der Ausgangssprache beachtet (Haverkort 1991: 9). Mit anderen Worten wird die Analyse der Ausgangssprache durch die Zielsprache bestimmt (Haverkort 1991: 9). Ein syntaktisches Element wird „oberflächlich an die Satzstellung der Zielsprache angepasst" (Stein 2009: 8). Der Vorteil des direkten Systems liegt in der Reduzierung des Aufwandes für die Analyse. Jedoch ist das Systems nur auf ein bestimmtes Sprachpaar beschränkt und muss im Fall einer zusätzlichen Sprache erweitert werden, d. h., der Übersetzungsalgorithmus muss umgeschrieben werden (Haverkort 1991: 9). Ein weiteres Problem besteht darin, dass viele Wörter polysem sind und Mehrwortlexeme sowie Kollokationen nicht wörtlich übersetzt werden können (Stein 2009: 8).

1.2.1.2 Transfer-Ansatz

Wenn es sich um Übersetzungen in mehrere Sprachen oder einen weiteren Ausbau des Übersetzungssystems handelt, ist es sinnvoller, die Analyse der Ausgangssprache durchzuführen (Haverkort 1991: 9). Vauquois (1976) beschreibt die Transfer-Methode als einen Prozess, der drei Hauptvorgänge umfasst: die Analyse des Ausgangstextes, den Transfer einer Ausgangs- zu einer Zielrepräsentation und schließlich die Erzeugung des Zieltextes (Vauquois 1976: 130f.).

In der ersten Phase wird der Ausgangstext analysiert, was zu einer Zwischenrepräsentation (engl. „intermediate representation"), z. B. dem Syntaxbaum der Ausgangssprache oder der semantischen Repräsentation, führt. In der zweiten Phase wird die ausgangssprachliche Zwischendarstellung in eine entsprechende zielsprachliche Zwischendarstellung transformiert. In der dritten Phase wird der zielsprachliche Text aus der zielsprachlichen Zwischendarstellung generiert (Schwartz 2018: 173). Im deutschsprachigen Raum werden diese drei Phasen als Analyse, Transfer und Synthese bezeichnet (Stein 2009: 8; Haverkort 1991: 9). Schwartz (2018) bezeichnet diesen Transfer-Ansatz als „the analysis-transfer-generation paradigm" (Schwartz 2018: 173).

1.2.1.3 Interlingua-Ansatz

Haverkort (1991: 10) versteht unter einer Interlingua eine „sprachunabhängige Zwischensprache“ und erklärt, dass es sich bei diesem Ansatz um eine erweiterte Analyse der Ausgangssprache handelt. Das Ziel des Interlingua-Ansatzes ist es, die Transfer-Phase zu vermeiden und stattdessen interlinguale Zwischenpräsentationen zu verwenden (Schwartz 2018: 174). Genauer gesagt wird der Ausgangstext einer gründlichen Analyse unterzogen, wodurch eine sprachunabhängige Zwischenpräsentation entsteht (Schwartz 2018: 174). Allerdings unterschieden sich die aus dieser Analyse sich ergebenden Zwischenrepräsentationen von der Form eines ausgangssprachlichen Satzes und geben nur den Inhalt wieder (Haverkort 1991: 9). Dies bietet eine Möglichkeit, einen Satz in jeder ausgesuchten Zielsprache direkt ohne Transfer zu generieren (Haverkort 1991: 9).

Haverkort (1991: 10) sieht darin zwei Nachteile: das Fehlen von kontrastiven Kenntnissen der Ausgangs- und Zielsprache sowie die daraus resultierende Erschwernis der Synthese. Stein (2009) bezeichnet die Interlingua-Übersetzung als ein „utopisches Ideal“ und erklärt, dass es keine Universalsprache gibt, die eine absolut sprachneutrale Kodierung linguistischer Informationen ermöglicht (Stein 2009: 8).

1.2.1.4 Vorteile und Nachteile der regelbasierten maschinellen Übersetzung

Ein Vorteil der RBMÜ ist, dass sie sowohl auf syntaktischer als auch auf semantischer Ebene tief analysieren kann (Charoenpornsawat et al. 2002: o. S.). Da natürliche Sprachen sehr komplex sind, ergeben sich Schwierigkeiten und Probleme für die MÜ. Zu den häufigsten Nachteilen der RBMÜ gehören die Frage nach der Hinlänglichkeit von Regeln und Wörterbüchern, die Methode und die Kosten für deren Erstellung, den Umgang mit Mehrdeutigkeiten und idiomatischen Ausdrücken in der Sprache sowie die Anpassbarkeit des Systems in neuen Domänen (Shiwen & Xiaojing 2015: 192f.). Charoenpornsawat et al. (2002) sehen zwei wesentliche Nachteile des regelbasierten Ansatzes, nämlich, dass das regelbasierte System viel linguistisches Wissen erfordert und es unmöglich ist, linguistische Regeln für alle Sprachen zu schreiben (Charoenpornsawat et al. 2002: o. S.).

1.2.2 Statistische maschinelle Übersetzung

Brown et al. (1988) definieren den statistischen Ansatz als einen Ansatz, bei dem die Informationen statistisch aus großen Datenbanken extrahiert werden (Brown et al. 1988: 810). Die Methode von Brown et al. (1988) basiert auf einer großen Anzahl von Texten, die paarweise

zu Verfügung stehen. Dies bedeutet, dass solche Paralleltexte den Parallelkorpus eines statistischen Systems ausmachen und Übersetzungen anhand von statistischen Wahrscheinlichkeitsmodellen aus diesem Korpus generiert werden (Brown et al. 1988: 810; Azzano 2009: 20). Im Vergleich zur RBMÜ sind für die SMÜ nicht grammatische und syntaktische Regeln, sondern große parallele Korpora von Bedeutung (Stein 2009: 9; Somers 2003: 4).

1.2.2.1 Wortbasierte statistische maschinelle Übersetzung

Die wortbasierte SMÜ ist die ursprüngliche Variante der SMÜ, die von Browns Team vorgeschlagen wurde (Brown et al. 1988, 1993). Koehn (2010) bezeichnet diesen Typ der SMÜ als „a simple model for machine translation that is based solely on lexical translation, the translation of words in isolation" (Koehn 2010: 81). Er untersuchte die wortbasierte SMÜ und stellte fest, dass es sich dabei um die Häufigkeitsverteilung der Übersetzungen von Lexemen (engl. „lexical translation probability distribution") handelt (Koehn 2010: 118). Der Wissenschaftler führt ein gutes Beispiel an, wie die wortbasierte SMÜ funktioniert: Wenn wir z. B. eine englische Übersetzung für das deutsche Wort „Haus" benötigen, brauchen wir eine große Sammlung deutscher Texte mit deren Übersetzungen ins Englische, damit wir zählen können, wie oft das Wort „Haus" in jede der vorhandenen Varianten übersetzt wird (Koehn 2010: 81).

Nach Meinung von Koehn (2010) sind fehlende Informationen über die Übereinstimmung zwischen Eingabe- und Ausgabe-Wörtern bei dieser Art des maschinellen Übersetzungssystems nachteilig, was zum Problem der unvollständigen Daten führt (Koehn 2010: 118). Allerdings kann dieses Problem durch den Erwartungs-Maximierungs-Algorithmus (engl. „expectation maximization algorithm") gelöst werden (Koehn 2010: 118). Stein (2009) sieht ebenfalls Nachteile darin, dass jedem Wort in der Ausgangssprache ein Wort in der Zielsprache zugeordnet werden muss, was dazu führt, dass Wortkombinationen nicht als eine Einheit zusammen übersetzt werden können (Stein 2009: 12).

1.2.2.2 Phrasenbasierte statistische maschinelle Übersetzung

Zu den phrasenbasierten Modellen der SMÜ gehören Modelle, die kleine Wortsequenzen am Stück übersetzen (Koehn 2010: 127). Stein (2009) und Schwartz (2018) betonen, dass trotz der Bezeichnung „phrasenbasiert" diese Übersetzungsmodelle auch viele Phrasenpaare enthalten, die nicht den linguistischen Konstituenten entsprechen (Stein 2009: 12; Schwartz 2018: 177).

Anders als den wortbasierten Modellen gelingt es phrasenbasierten Modellen, mehrere Wörter (auch nur mit einem Wort) zu übersetzen und umgekehrt (Stein 2009: 12). Stein (2009: 12) und Koehn (2010: 128) teilen die Meinung, dass in der phrasenbasierten SMÜ das Problem der Zweideutigkeit – dank dem erweiterten Kontext – besser gelöst wird. Koehn (2010) stellt fest, dass die Verfügbarkeit großer Korpora für die SMÜ-Modelle von Bedeutung ist, da dadurch auch lange Phrasen bzw. ganze Sätze gelernt werden können (Koehn 2010: 128).

1.2.2.3 Vorteile und Nachteile der statistischen maschinellen Übersetzung

Vorteile der SMÜ-Systeme sind, dass sie recht schnell zu erstellen und relativ günstig sind, da solche Systeme keine linguistischen Regeln benötigen (Stein 2009: 12). Was die Modellierung angeht, ist diese nicht kompliziert, wenn man sie mit den regelbasierten Modellen vergleicht (Stein 2009: 12). Wenn man über genügend und ausreichend große Parallelkorpora für das Training des Systems verfügt, können ziemlich gute Übersetzungsergebnisse erzielt werden. Wie im vorigen Unterkapitel schon erwähnt, tritt das Problem der Zweideutigkeit in der SMÜ kaum auf (Stein 2009: 12f.).

Jedoch ist zu berücksichtigen, dass Fehlerquellen in der SMÜ nicht zu identifizieren sind, weil das Trainingsmaterial unübersichtlich ist. Zudem lassen sich die Berechnungen nicht mehr nachvollziehen (Stein 2009: 13). Stein (2009: 13) stellt heraus, dass große strukturelle Unterschiede der beiden Sprachen (etwa ihrer Syntax und Flexion) die Übersetzungsqualität beeinträchtigen. Auch seien zweisprachige Korpora in der Regel sehr fachspezifisch und passten daher nicht zu jedem Kontext (Stein 2009: 13).

1.2.3 Beispielbasierte maschinelle Übersetzung

Ähnlich wie die SMÜ verwendet die beispielbasierte maschinelle Übersetzung (BBMÜ) einen Korpus von Paralleltexten. Allerdings geht es dabei nicht um das Wahrscheinlichkeitsprinzip, sondern um die Analogie (Stein 2009: 13). Anhand einer großen Anzahl von Beispielsätzen mit den entsprechenden Übersetzungen erfasst das BBMÜ-System die Ähnlichkeiten und Unterschiede des jeweiligen Beispielsatzpaares (Nagao 2003: 351). Anschließend wird ein weiteres Beispielsatzpaar erstellt, das sich nur durch ein Wort vom ursprünglichen Paar unterscheidet, und das MÜ-System wird nach jeder Wortsubstitution informiert (Nagao 2003: 351).

Schwartz (2018) zufolge ist das beispielbasierte MÜ-System ein Modell, das zwischen regelbasierten und statistischen korpusbasierten Ansätzen der MÜ platziert werden kann (Schwartz 2018: 176). Er stellt fest, dass die beispielbasierte MÜ auf dem Prinzip des Fuzzy Matching beruht und daher mit Translation-Memory-Systemen (TMS) eine gewisse Ähnlichkeit aufweist (Schwartz 2018: 176). Somers (2003) spricht ebenfalls davon, dass die BBMÜ oft mit TMS in Verbindung gebracht wird; allerdings ist dies seiner Meinung nach nur teilweise korrekt, da die BBMÜ im Vergleich zu TMS ausschließlich maschinell funktioniert (Somers 2003: 5).

Einerseits zeigen die BBMÜ-Systeme den realen Sprachgebrauch, beinhalten die tatsächlich vorkommenden Konstruktionen und sind leicht zu erweitern (Wong & Webster 2015: 146). Anderseits leiden einige Sprachpaare unter Ressourcenmangel. Für die Entwicklung solcher MÜ-Systeme ist das ein Problem, das nur teilweise gelöst wurde. Wong und Webster (2015) weisen darauf hin, dass es bis jetzt kein MÜ-System gebe, das rein beispielbasiert und öffentlich zugänglich wäre (Wong & Webster 2015: 146).

1.2.4 Kontextbasierte MÜ

Die kontextbasierte maschinelle Übersetzung (KBMÜ) ist ein korpusbasierter Ansatz der MÜ, der weder bestimmte linguistische Regeln noch Paralleltexte benötigt (Carbonell et al. 2006: 19). Das KBMÜ-System verwendet ein zweisprachiges Vollformenwörterbuch und einen hochentwickelten Decoder, der den Kontext durch lange *n*-Gramme (aufeinanderfolgende einzelne Segmente eines zerlegten Textes) berücksichtigt (Carbonell et al. 2006: 19). Das Vollformenwörterbuch ermöglicht die Festlegung aller möglichen Übersetzungsvarianten eines Wortes (Stein 2009: 14). Anhand von *n*-Grammen vergleicht das KBMÜ-System Übersetzungsvarianten mit dem Zielkorpus und übernimmt diejenige Variante, die über mehr Matches verfügt (Stein 2009: 14).

Carbonell et al. (2006) stellen zwei wichtige Vorteile der KBMÜ heraus: höhere Genauigkeit aufgrund der Fähigkeit des Modells, lange *n*-Gramme zu dekodieren, und die Fähigkeit des Systems, schnell auf neue Sprachpaare erweitert zu werden (Carbonell et al. 2006: 19). Darüber hinaus eignen sich die KBMÜ-Modelle gut für solche Sprachen, die über wenig linguistische Ressourcen bzw. wenig Paralleltexte verfügen. Die Fähigkeit des KBMÜ-Systems, den Kontext beizubehalten, macht sie vorteilhaft (Tripathi & Sarkhel 2010: 391). Zudem können übersetzte Segmente nach dem Prinzip der Aussagewahrscheinlichkeit als hoch

oder niedrig bewertet werden, sodass nur Sätze mit niedriger Wahrscheinlichkeit überprüft werden müssen (Tripathi & Sarkhel 2010: 391).

1.2.5 Neuronale maschinelle Übersetzung

Forcada (2017) bezeichnet die NMÜ als einen neuen Ansatz der korpusbasierten MÜ, der auf riesigen Parallelkorpora von ausgangssprachlichen Segmenten und deren Übersetzungen trainiert wird (Forcada 2017: 292). Solche Korpora sind so groß, dass sie Millionen von Übersetzungseinheiten enthalten können (Forcada 2017: 292). Obwohl die NMÜ der SMÜ ähnelt, benötigt das NMÜ-System weniger Vorverarbeitungsschritte, bevor ein Übersetzungsmodell erstellt wird (Luong 2016: 7). Außerdem wird in der NMÜ ein völlig anderer Berechnungsansatz verwendet, nämlich neuronale Netze (Forcada 2017: 292).

Neuronale Netze müssen lineare Modelle modifizieren, weil solche Modelle sonst komplexere Beziehungen zwischen den Merkmalen nicht definieren können (Koehn 2020: 67). Linearen Modelle sind wiederum ein wichtiger Bestandteil der statistischen maschinellen Übersetzung. Koehn (2020) erklärt, dass die erste Möglichkeit für die Modifikation von linearen Modellen die Verwendung von mehreren Schichten ist (Koehn 2020: 68). Anstatt die Ausgabe direkt aus der Eingabe zu berechnen, werden versteckte Schichten (engl. „hidden layers“) eingeführt (Koehn 2020: 68). Man nennt diese Schichten „versteckt“, weil nur die Werte der Eingabe- und Ausgabeschicht beobachtet werden können, nicht jedoch das Verhalten der dazwischen liegenden Schichten (Koehn 2020: 68). Luong (2016) findet es vorteilhaft, dass das NMÜ-System direkt vom Lernziel aus durchgehend trainiert werden kann (Luong 2016: 7).

1.2.5.1 Encoder-Decoder-Prinzip

Laut Forcada (2017) ähneln die meisten NMÜ-Systeme einer Textvervollständigungsvorrichtung, wobei Informationen mittels Repräsentationen der einzelnen Wörter im Satz unter Einbeziehung des Kontextes erhoben werden (Forcada 2017: 296). Der Decoder übernimmt die Funktionen eines Textvervollständigers und legt für jedes mögliche Wort im Zielvokabular die Wahrscheinlichkeit fest, dass dieses Wort eine Fortsetzung des bereits produzierten Wortes ist (Forcada 2017: 296).

Bei Luong (2016: 8) findet man die folgende Erläuterung des Encoder-Decoder-Prinzips: Ein Encoder konvertiert einen konkreten Ausgangssatz in einen Bedeutungsvektor.

Genau genommen verwendet der Decoder den Satzvektor in der Ausgangssprache, um die Übersetzung zu generieren (Luong 2016: 8).

Koehn (2020: 125) beschreibt die Funktionsweise des Encoder-Decoder-Modells als die Vorhersage des nächsten Wortes auf Basis des vorherigen unter Berücksichtigung der komprimierten Vektordarstellung des Ausgangssatzes. Um dieses Modell zu trainieren, müssen die Ausgangs- und Zielsätze verkettet werden (Koehn 2020: 125). Beim Decodieren wird der Ausgangssatz eingegeben, um mit dem Vorhersagen so lange fortzufahren, bis ein Satzende-Token vorhersagt wird (Koehn 2020: 125).

1.2.5.2 Vorteile und Nachteile der neuronalen maschinellen Übersetzung

Ein Vergleich der NMÜ mit anderen Ansätzen der MÜ zeigt, dass dieser Ansatz einfacher ist, weil man dank der großen Netzwerke und einer End-to-End-Übersetzung mehrere Komponenten und Verarbeitungsschritte einsparen kann (Bentivogli et al. 2016). Laut Bentivogli et al. (2016) ermöglicht die NMÜ eine effizientere Nutzung von Personal- und Datenressourcen (Bentivogli et al. 2016: 257). Allerdings kritisieren sie die mangelnde Transparenz des NMÜ-Prozesses:

> Indeed, it represents a further step in the evolution from rule-based approaches that explicitly manipulate knowledge, to the statistical/data-driven framework, still comprehensible in its inner workings, to a sub-symbolic framework in which the translation process is totally opaque to the analysis. (Bentivogli et al. 2016: 257)

Wu et al. (2016) behaupten, dass die NMÜ-Systeme früher Probleme mit ihrer Genauigkeit aufwiesen, wenn sie auf sehr großen Datensätzen trainiert wurden (Wu et al. 2016: 1f.). Die Forscher*innen nennen drei Gründe dafür: langsames Training des Systems und langsame Inferenzgeschwindigkeit, ineffektiver Umgang mit seltenen Wörtern und das Problem unvollständiger Übersetzung (Wu et al. 2016: 2).

Einen weiteren Nachteil aktueller NMÜ-Architekturen sehen Britz et al. (2017) im enormen Berechnungsaufwand für das Training solcher Systeme (Britz et al. 2017: 1442). Die Forscher*innen verweisen auf die Arbeit von Wu et al. (2016) und erklären, dass das Training auf Datensätzen mit mehreren Millionen Beispielen Dutzende von Grafikprozessoren erfordert (Britz et al. 2017: 1442).

1.2.6 Hybride Ansätze

Da die menschlichen Sprachen sehr komplex sind, kann kein einziges MÜ-System alle sprachlichen Besonderheiten berücksichtigen. Werthmann und Witt (2014) sind der Meinung,

dass die Möglichkeiten jedes MÜ-Modells limitiert sind (Werthmann & Witt 2014: 99). Zu den häufigsten Schwierigkeiten der MÜ gehören laut Tripathi und Sarkhel (2010) linguistische Unregelmäßigkeiten, Mehrdeutigkeit sowie mangelnde Universalität der Grammatik und des Lexikons (Tripathi & Sarkhel 2010: 392). Um diese Probleme lösen zu können, werden hybride Ansätze entwickelt. Unter hybriden Systemen verstehen Werthmann und Witt (2014) solche Systeme, die mit mehreren diversen MÜ-Modellen gleichlaufend betrieben werden, mit dem Ziel, die Vorzüge verschiedener MÜ-Methoden in einem MÜ-System zu kombinieren und die Probleme eines bestimmten Modells zu lösen (Werthmann & Witt 2014: 99). Stein (2009) ist der Ansicht, dass hybride Ansätze vor allem für die SMÜ hilfreich sind, da dadurch gewisse Probleme wie Sprachen mit kleinen Korpora oder Probleme mit unterschiedlicher Syntax und Flexion gelöst werden können (Stein 2009: 14).

Costa-Jussa und Fonollosa (2015) verschafften einen guten Überblick über die gebräuchlichsten hybriden Ansätze. Die Wissenschaftler*innen unterscheiden zwischen der regelbasierten und korpusbasierten Hybridisierung (Costa-Jussa & Fonollosa 2015: 5f.). Zu den hybriden regelbasierten Ansätzen gehören die Verwendung eines Korpus zum Aufbau des RBMÜ-Systems, die Verwendung von korpusbasierten Tools zur Bewertung des RBMÜ-Outputs sowie die statistische Nachbearbeitung eines RBMÜ-Outputs (Costa-Jussa & Fonollosa 2015: 5). Bei den hybriden korpusbasierten Ansätzen können linguistische Regeln integriert werden oder es bestehen Möglichkeiten einer Kombination von mehreren korpusbasierten MÜ-Ansätzen (Costa-Jussa & Fonollosa 2015: 6). Costa-Jussa und Fonollosa (2015) erwähnen zwei Optionen einer Integration von Regeln: Die Verwendung von Regeln im Pre-/Post-Processing oder die Integration von Wörterbüchern / Regeln in das Kernmodell eines Systems (Costa-Jussa & Fonollosa 2015: 6).

In den 2000ern war SYSTRAN ein gutes Beispiel für ein hybrides MÜ-System, das auf dem regelbasierten Ansatz beruhte und statistische Komponenten wie statistisches Post-Editing verwendete (Dugast et al. 2007). Smith und Clark (2009) entwickelten ein hybrides Modell für das phrasenbasierte SMÜ-System Moses, indem sie es mit beispielbasierten Komponenten ergänzten.

1.3 Verfügbare Systeme der maschinellen Übersetzung

Heutzutage gibt es zahlreiche Systeme der maschinellen Übersetzung. Dieses Kapitel konzentriert sich auf vier bekannte kommerzielle MÜ-Systeme: SYSTRAN, Google Translate, DeepL und PROMT.

1.3.1 SYSTRAN

System Translation (SYSTRAN) wurde als System von Peter Toma entwickelt und 1968 als Firma gegründet (Toma 1977). Laut SYSTRANs Website ist dieses MÜ-Softwaresystem in der Lage, sowohl die mehrsprachige Kommunikation zu optimieren als auch die Produktivität von Unternehmen zu erhöhen (Systran o. J.). Ursprünglich war das MÜ-System SYSTRAN regelbasiert (Dugast et al. 2007: 221). Im Laufe der Jahre wurde sein klassisches MÜ-System durch die Integration neuer Technologien verbessert (Senellart et al. 2001; Dugast et al. 2007). So wurde SYSTRAN mit korpusbasierten Tools ergänzt (Senellart 2006) und mit einem statistischen Post-Editing-System kombiniert (Dugast et al. 2007). 2016 brachte SYSTRAN sein neues MÜ-System „Pure Neural™ Machine Translation" heraus, das auf künstlichen neuronalen Netzen und Deep Learning basierte. Laut SYSTRANs Entwickler*innen (Systran 2018: 1) sollte dieses neue System in der Lage sein, eine gute Übersetzungsqualität zu liefern, die einer menschlichen Übersetzung nahekommt (Systran 2018: 1).

1.3.2 Google Translate

Google startete sein eigenes Projekt der Online-Übersetzung im Jahr 2001 (Levy 2011: 63). Zu Beginn seiner Entwicklung verfügte der Online-Übersetzer von Google über ein regelbasiertes MÜ-System (Ghasemi & Hashemian 2016: 14). Allerdings wiesen die Übersetzungsergebnisse eine sehr schlechte Qualität auf, weshalb ein weiterer Verbesserungsbedarf dieses Übersetzungssystems bestand (Levy 2011: 63). 2006 bekam das Übersetzungsprojekt von Google seinen Namen „Google Translate", das ein statistisches Modell auf Basis von Paralleltexten verwendete. 2011 entwickelte die Forschungsgruppe von Google eine App, die in der Lage war, die gesprochene Ausgangssprache automatisch zu erkennen und den Inhalt in der Zielsprache mündlich wiederzugeben (Turovsky 2016). 2014 rief das Google-Team die Initiative „Translate Community" ins Leben, die es mehrsprachigen Nutzer*innen ermöglichte, eigene Übersetzungen vorzuschlagen sowie eine vorhandene Übersetzung zu bewerten (Kelman 2014). Seit 2016 verwendet Google Translate neuronale Methoden, die erstmals vom Google-Forschungsteam präsentiert wurden (Wu et al. 2016).

Derzeit ermöglicht Google Translate maschinelle Übersetzung und maschinelles Dolmetschen. Das System kann einzelne Wörter, verschiedene Textsorten und Webseiten und sogar Texte, die mit einer Kamera aufgenommen wurden, übersetzen (Turovsky 2016). Word Lens ist eine App von Google Translate, die Speisekarten, Straßenschilder etc. in 28 Sprachen anhand einer Kamera offline übersetzen kann (Turovsky 2016).

1.3.3 DeepL-Übersetzer

DeepL ist ein maschineller Online-Übersetzer der gleichnamigen Firma aus Köln, der 2017 präsentiert wurde. Das Team von DeepL-Entwickler*innen, unter der Leitung von Jaroslaw Kutylowski, behauptet, es sei ihnen gelungen, künstliche neuronale Netze zu entwickeln, die im Stande sind, den Sinn eines Ausgangssatzes samt seinen Feinheiten besser als bisherige Systeme zu verstehen und diesen sinngemäß wiederzugeben (DeepL 2018a). Der DeepL-Übersetzer verfügt über zwei Online-Versionen: eine kostenlose sowie eine kostenpflichtige Version „DeepL Pro".

Am Anfang unterstützte das DeepL-System nur neun Sprachen. Im Laufe der Zeit wurde der Online-Übersetzer mit weiteren Sprachen schrittweise ergänzt, was momentan in der Summe 27 Sprachen (inkl. Sprachvarietäten) macht (Stand: 04.10.2022). DeepL unterstützt auch docx- und pptx-Dateiformate unter Beibehaltung der Originalformatierung (inkl. Fußnoten, Bildunterschriften etc.) (DeepL 2018b). Seit Mai 2020 bietet der DeepL-Übersetzer die Möglichkeit, das Glossar individuell zu gestalten (DeepL 2020a).

Laut Evaluierungsergebnissen der Blindtests (DeepL 2020b) übersetzt das NMÜ-System DeepL effizienter und präziser als andere MÜ-Übersetzungssysteme wie Google Translate, Amazon Translate sowie Microsoft Translator (DeepL 2020b). Die Entwickler*innen von DeepL behaupten, dass ihre mathematischen und methodischen Verbesserungen bzw. Innovationen im Bereich der neuronalen Netze zu solch guten Ergebnissen geführt haben (DeepL 2020b).

Im Fall von DeepL werden Paralleltexte bzw. Übersetzungsbeispiele von Linguee für das Training verwendet. Linguee ist ein Online-Wörterbuch, das eine große Menge von hochwertigen Übersetzungsbeispielen beinhaltet (Linguee o. J.). Am 26. September 2019 bekam das DeepL-Unternehmen eine Auszeichnung für den ersten deutschen KI-Ehrenpreis (DeepL 2019). Eine detaillierte Beschreibung des DeepL-Modells wird vom Unternehmen aus Konkurrenzgründen nicht zu Verfügung gestellt.

1.3.4 PROMT

PROject Machine Translation (PROMT) wurde 1991 als eine weitere Version der ursprünglichen MÜ-Software STYLUS von Svetlana Sokolova und ihrem Team in Russland gegründet (Sokolova 1999: 9). PROMTs erstes MÜ-System war regelbasiert und bot maschinelle Übersetzungen für das Sprachpaar Englisch-Russisch (Sokolova 1999: 9). Ab den 2000er Jahren entwickelte PROMT ein hybrides System, das regelbasierte und statistische

Algorithmen kombinierte. 2005 teilte PROMT auf seiner Website mit, dass die NASA sich nach der Testung möglicher Alternativen für die Übersetzungstechnologien von @promt Professional 7.0 entschied (PROMT 2005).

PROMT bekam die höchste Punktzahl in der ersten Runde des europäischen Projekts Covid-19 MLIA Eval (Casacuberta 2021: 8ff., 15). Dieses Projekt setzte sich zum Ziel, die von verschiedenen Entwickler*innen präsentierten MÜ-Systeme zu evaluieren (Covid-19 MLIA Eval o. J.).

2019 präsentierte das PROMT-Unternehmen seine PROMT Neural Translation Server 21 und PROMT Master NMT 21, die als nächste Generation der MÜ-Lösungen für das Übersetzen gelten (PROMT 2020). PROMT Neural Translation Server verwendet eine hybride Technologie, die den regelbasierten Ansatz mit Machine Learning kombiniert, das auf rekurrenten neuronalen Netzen (RNNs) basiert (PROMT 2019). Die PROMT-Entwickler*innen sind überzeugt, dass diese Kombination der verschiedenen MÜ-Ansätze eine sehr genaue und flüssige Übersetzung ermöglicht, da dieses MÜ-System sehr gut Fachbegriffe bzw. Unternehmensterminologien und den Stil der Trainingsdaten lernt und daher konsistente Übersetzungen anfertigen kann (PROMT 2019). Unter anderem garantiert PROMT Neural Translation Server die Diskretion und den Datenschutz der übersetzten Daten, da es in das Unternehmensnetzwerk der Kund*innen integriert ist und problemlos ohne Internetverbindung funktioniert (PROMT 2019). PROMT Master NMT ähnelt funktionsmäßig der oben beschriebenen Software (PROMT 2020).

2 Evaluierung der maschinellen Übersetzung

Die Evaluierung der MÜ-Qualität ist ein sehr kontroverses Thema in der Sprachdienstleistungsbranche, da sie die finanzielle Seite beeinflusst (Läubli et al. 2020: 264, White 2003: 212). Ein MÜ-System zu erforschen, zu entwerfen und zu implementieren kostet viel Geld und Zeit, und noch teurer ist es, das System zu pflegen (White 2003: 212). In den letzten Jahren wurden viele Evaluierungstechniken entwickelt, die für alle Arten der MÜ-Systeme verwendet werden können. Darüber hinaus wirft der Begriff der Übersetzungsqualität in der Translationswissenschaft viele Fragen auf, da es unterschiedliche Ansichten der Forscher*innen darüber gibt, was eine gute Übersetzungsqualität ausmacht. In diesem Kapitel werden verschiedene Ansätze zur MÜ-Evaluierung vorgestellt. Zudem wird die Frage der Übersetzungsqualität aufgegriffen.

2.1 Übersetzungsqualität

Zilahy (1963) befasste sich mit dem literarischen und technischen Übersetzen und bezeichnete eine Übersetzung als gut, wenn sie die gleiche Wirkung wie das Original erwecken kann (Zilahy 1963: 285). Kandler (1963) betont die Wichtigkeit bestimmter Kriterien für die Evaluierung der Qualität:

> Consequently, translations cannot be simply judged as right or wrong. We should rather have a scale of valuation according to the degree of coincidence of the interpretability of the translation with the interpretability of the original, and we should not forget that quality cannot possibly be assessed apart from the purpose of the translation. (Kandler 1963: 295)

Den Begriff „Qualität“ kann man unterschiedlich definieren. Schmitt (1999) versteht darunter „weder etwas Absolutes noch das maximal Machbare, sondern die Erfüllung definierter Erwartungen“ (Schmitt 1999: 397).

Reiß und Vermeer (1984) publizierten die sogenannte Skopostheorie, die sich an der Übereinstimmung von Informationen und Kommunikationsziel orientiert. Eine Übersetzung wird als eine kommunikative Handlung betrachtet, die durch einen konkreten Zweck bestimmt wird (Reiß & Vermeer 1984). Nord (1993) vertritt das Konzept „des funktionalen Übersetzens“, das auf der Skopostheorie von Reiss und Vermeer (1984) basiert. Die Wissenschaftlerin erklärt, dass der Übersetzungszweck die kommunikative Situation berücksichtigen muss, für welche die Übersetzung benötigt wird (Nord 2006: 15). Nord (2006) zufolge ist die „Funktionsgerechtigkeit“ ein wichtiges Kriterium der Übersetzungsqualität (Nord 2006: 16).

House (2014) ist der Ansicht, dass die Frage, was eine gute Übersetzungsqualität ist, nicht einfach zu beantworten ist. Dazu schreibt sie:

> Any statement about the quality of a translation implies a conception of the nature and goals of translation, in other words it presupposes a theory of translation. And different theoretical stances must lead to different concepts of translational quality, to different ways of going about assessing (retrospectively) the quality of a translation and different ways of ensuring (prospectively) the production of a translation of specified qualities. (House 2014: 241)

House (2014) stellt verschiedene Ansätze vor, bei denen die Übersetzungsqualität aus unterschiedlichen Perspektiven betrachtet und dadurch ganz unterschiedlich wahrgenommen und bewertet wird. Sie teilt diese Ansätze in zwei Gruppen ein: psychosoziale sowie text- und diskursorientierte Ansätze (House 2014: 242, 245).

Psychosoziale Ansätze befassen sich mit mentalistischen Ansichten (engl. „mentalist views“) und reaktionsbasierten Ansätzen. Bei den mentalistischen Ansichten geht es um die Annahme, dass die Qualität einer Übersetzung von den subjektiven Entscheidungen eines/einer Übersetzers/Übersetzerin abhängt, was die Interpretation betrifft (House 2014: 243). Diese Entscheidungen basieren auf der Intuition und Erfahrung der Übersetzer*innen. Zu den reaktionsbasierten Ansätzen gehören verhaltensorientierte und funktionsorientierte Konzepte.

Bei dem verhaltensorientierten Konzept verweist House (2014) auf das Nidas-Prinzip der dynamischen Äquivalenz (Nida 2001: 225), das davon ausgeht, dass eine gute Übersetzung zu einer äquivalenten Reaktion führen muss. Das bedeutet, dass die Reaktion der Rezipient*innen auf den Ausgangstext und auf seine Übersetzung gleich sein soll (House 2014: 244). Zu dieser Kategorie gehört ebenso die funktionsorientierte Ansicht von Reiß und Vermeer (1984), die oben schon beschrieben wurde.

Text- und diskursorientierte Ansätze umfassen laut House (2014) deskriptiv-historische Übersetzungsstudien, postmodernistische und dekonstruktivistische Ansätze sowie linguistisch orientierte Ansätze (House 2014: 245, 246, 247).

Nach Meinung von Schmitt (1999) sollten Auftraggeber*innen im Übersetzungsauftrag das erforderliche Niveau der Übersetzungsqualität angeben:

> […] z. B. kann spezifiziert werden, daß eine „korrekturgelesene, fehlerfreie Übersetzung“ zu liefern ist. Dies ist nur dann aussagekräftig, wenn auch spezifiziert wird, was unter „korrekturgelesen“ und „fehlerfrei“ zu verstehen ist. (Schmitt 1999: 397)

Ahrend (2006) beschäftigte sich mit der Bewertung von Fachübersetzungen und spricht über allgemeine und praxisorientierte Kriterien für die Feststellung der Übersetzungsqualität. Ihm zufolge gehören Vollständigkeit, Genauigkeit, Übersetzungstreue, Verständlichkeit und Stilsicherheit zu den allgemeinen Kriterien der Übersetzungsqualität (Ahrend 2006: 31). Was die

praxisorientierten Kriterien angeht, teilt Ahrend (2006) die Ansicht von Reiß und Vermeer (1984) und Nord (1993, 2006) und stellt fest, dass der Übersetzungszweck im Vordergrund steht (Ahrend 2006: 32, 34). Für die fachtextspezifischen Kriterien der Übersetzungsqualität spiele die korrekte Fachterminologie eine entscheidende Rolle (Ahrend 2006: 37).

Koby et al. (2014) unterscheiden zwischen einer breit gefassten und einer eng gefassten Definition des Begriffs „Übersetzungsqualität" (Koby et al. 2014: 416). Im breiteren Sinne wird unter einer guten Übersetzungsqualität die Erfüllung aller für die Zielgruppe und den Übersetzungszweck erforderlichen Anforderungen an Genauigkeit und Sprachflüssigkeit sowie die Berücksichtigung aller von den Auftraggeber*innen und Endbenutzer*innen festgelegten Spezifikationen verstanden (Koby et al. 2014: 416). Im engeren Sinne handelt es sich bei einer qualitativen Übersetzung um eine volle Übertragung der Kernaussage des Ausgangstextes in den Zieltext unter Verwendung der korrekten Grammatik und der Berücksichtigung des Stils, sodass der Zieltext an kulturelle Besonderheiten der Zielgruppe angepasst wird und sich als Original lesen lässt (Koby et al. 2014: 416f.).

In Diskussionen zur Übersetzungskritik kommen die Begriffe „Adäquatheit" und „Äquivalenz" sehr oft vor. Reiß (1989) definiert Adäquatheit als eine zielorientierte Auswahl der Sprachzeichen unter Berücksichtigung des Übersetzungszwecks (Reiß 1989: 163). Unter Äquivalenz versteht die Wissenschaftlerin „die Relation der Gleichwertigkeit von Sprachzeichen" in jedem der beiden Sprachsysteme (Reiß 1989: 163).

Läubli (2020) analysiert die Wichtigkeit der Übersetzungsqualität aus zwei verschiedenen Sichtweisen. Aus der Perspektive der Kund*innen ist die Übersetzungsqualität von Bedeutung, denn es wird dadurch bestimmt, ob eine Übersetzung den konkreten Zweck erfüllt. Aus der Perspektive der professionellen Translator*innen ist die Übersetzungsqualität deshalb wichtig, weil damit der Preis begründet und bestimmt wird (Läubli 2020: 11).

Zusammenfassend lässt sich sagen, dass es nicht möglich ist, den Begriff „Qualität" genau zu definieren, weil die Wahrnehmung der Qualität in gewissem Maße subjektiv ist. Allerdings gibt es verschiedene Qualitätsstufen und -kategorien, die man genau untersuchen kann. Daraus lässt sich schließen, dass der Verwendungszweck für die Auswahl von Qualitätsstufen bzw. Qualitätskategorien entscheidend ist.

2.2 Automatische Evaluierung

Generell gibt es zwei Möglichkeiten, die Qualität der maschinellen Übersetzung zu evaluieren, und zwar die automatische und die manuelle Evaluierung der MÜ. Zur automatischen

Evaluierung der MÜ gehören Ansätze, die maschinelle Übersetzungen ohne menschliche Beteiligung bewerten (Chatzikoumi 2020: 140). Bei der automatischen Evaluierung werden verschiedene Metriken verwendet. Chatzikoumi (2020: 141) teilt alle automatischen Metriken der MÜ-Evaluierung in drei Kategorien ein:

- Metriken, die auf Referenzübersetzungen basieren und einen numerischen Wert für die MÜ-Ausgaben ermitteln (Papineni et al. 2002; Banerjee & Lavie 2005; Lita et al. 2005; Callison-Burch et al. 2007). Sie basieren auf dem Grad der Ähnlichkeit mit Referenzübersetzungen;
- Metriken der Qualitätsbewertung (engl. „quality estimation metrics“) (Specia et al. 2010, 2013; Chatzikoumi 2020), welche die MÜ-Ausgaben nach Qualitätsstufen klassifizieren und
- diagnostische Evaluierung auf Basis von Checkpoints (Zhou et al. 2008).

Eine der bekanntesten Metriken ist die Bilingual Evaluation Understudy (BLEU) (Papineni et al. 2002; Callison-Burch et al. 2007). Diese Metrik zeichnet sich durch einen hohen Grad an Übereinstimmung mit der menschlichen Bewertung aus. Die Metrik basiert auf der Annahme, dass eine maschinelle Übersetzung umso besser ist, je näher sie an einer professionellen Humanübersetzung liegt (Callison-Burch et al. 2006: 249). Die BLEU-Metrik berechnet die Übereinstimmung für einzelne Segmente oder Sätze, indem sie diese Segmente mit einer Reihe hochqualitativer Referenzübersetzungen vergleicht.

Die Metric for Evaluation of Translation with Explicit Ordering (METEOR) wurde entwickelt, um eine gute Übereinstimmung mit den Expert*innenbewertungen auf Wort- oder Satzebene zu erreichen (Banerjee & Lavie 2005: 66f.). Die METEOR-Metrik basiert auf der Verwendung von *n*-Grammen und einer statistischen Auswertung des Ausgangstextes (Banerjee & Lavie 2005: 67).

Lita et al. (2005) haben eine Gruppe trainierbarer Metriken für die MÜ-Evaluierung namens Broad Learning and Adaptation for Numeric Criteria (BLANC) vorgestellt. Die Wissenschaftler*innen betrachten die BLANC-Metriken als eine Kombination aus der BLEU, ROUGE und anderen Metriken. Lita et al. (2005) entwarfen einen Algorithmus, der eine vollständige Überlappungssuche über unterschiedlich große und nicht zusammenhängende Wortsequenzen (engl. „skip-n-grams“) durchführen kann (Lita et al. 2005: 741).

Es sollte noch einmal betont werden, dass die oben erwähnten Metriken nur in solchen Fällen gut funktionieren können, wenn die entsprechenden Referenzübersetzungen zu Verfügung stehen (Specia et al. 2010: 40). Mangels der entsprechenden Referenzübersetzungen

werden weiterhin Versuche unternommen, die Qualität der maschinell übersetzten Texte vorauszusagen und zu bewerten. Dafür verwendet man Metriken der Qualitätsbewertung (engl. „quality estimation metrics“) (Specia et al. 2010). Diese Metriken sind keine Metriken der MÜ-Evaluierung per se, aber dienen als deren Ersatz und als eine Alternative für die Qualitätsbewertung (Chatzikoumi 2020: 144). Die Aufgabe solcher Metriken besteht darin, die Qualität der Ausgabe durch eine vorgegebene Eingabe abzuschätzen, ohne über Informationen über die erwartete Ausgabe zu verfügen (Specia et al. 2010: 40). Specia et al. (2010) untersuchten das Problem der Voraussage der Übersetzungsqualität auf Satzebene ohne Referenzübersetzungen und verwendeten Klassifikations- und Regressionsalgorithmen sowie unabhängige Sprach-, Ressourcen- und MÜ-System-Merkmale (Specia et al. 2010: 40). Den Forscher*innen ist es gelungen, eine Metrik ohne Referenzübersetzungen zu verwenden, die eine gute Korrelation mit der menschlichen Bewertung erhalten hat (Specia et al. 2010: 46). Aus den Experimenten geht hervor, dass die Schwierigkeit des Ausgangssatzes einen starken Einfluss auf die Gesamtqualität der Übersetzungen hat (Specia et al. 2010: 46). Specia et al. (2010) betonen, dass das Ziel dieser Metrik nicht darin besteht, die anderen Metriken der MÜ-Evaluierung zu ersetzen, sondern darin, eine Möglichkeit einer Qualitätsbewertung zu bieten, wenn keine Referenzübersetzungen verfügbar sind (Specia et al. 2010: 46).

Es gibt noch eine Möglichkeit, die MÜ automatisch zu evaluieren, und zwar eine diagnostische Evaluierung anhand von Checkpoints (Zhou et al. 2008). Zhou et al. (2008) stellten eine diagnostische Evaluierungsplattform vor, die eine multifaktorielle Evaluierung auf der Grundlage automatisch konstruierter Checkpoints ermöglicht (Zhou et al. 2008: 1121). Unter einem Checkpoint verstehen die Forscher*innen „a linguistically motivated unit (e.g. an ambiguous word, a noun phrase, a verb-obj. collocation, a prepositional phrase etc.), which are pre-defined in a linguistic taxonomy for diagnostic evaluation“ (Zhou et al. 2008: 1121). Lita et al. (2005) sehen generell zwei große Herausforderungen, mit denen die automatische Evaluierung der MÜ konfrontiert ist: Der Mangel an Merkmalen, die sowohl die Satzstruktur als auch die lokale Bedeutung erfassen, und die Schwierigkeit, diese unterschiedlichen Merkmale in einem Algorithmus zu kombinieren (Lita et al. 2005: 740).

2.3 Manuelle Evaluierung

Grundsätzlich handelt es sich bei der manuellen Evaluierung um die Überprüfung einer Übersetzung auf Fehler gemäß einem bestimmten Qualitätsstandard durch professionelle Evaluierer*innen (Bewerter*innen) (Läubli 2020: 12). Die ersten Methoden der manuellen

Evaluierung basierten auf den vom ALPAC (1966) vorgeschlagenen Bewertungskategorien, nämlich auf der Verständlichkeit (engl. „intelligibility“) und Wiedergabetreue (engl. „fidelity“) (ALPAC 1966: 67). Bei der Verständlichkeit wird darauf geachtet, dass die zu bewertende Übersetzung für die Zielgruppe gleichermaßen verständlich ist wie für die Leser*innen des Originals. Die Wiedergabetreue bezieht sich darauf, dass die Bedeutung des Originaltextes in der Übersetzung so wenig wie möglich verändert wird (Han et al. 2016: 2).

Zusätzlich können weitere Evaluierungskategorien in Betracht gezogen werden: Lesbarkeit (engl. „readability“) und Verständlichkeit (engl. „comprehensibility“) sowie Benutzerfreundlichkeit (engl. „usability“) und Annehmbarkeit (engl. „acceptability“) von Übersetzungen (Castilho et al. 2018: 17).

Mithilfe der Lesbarkeit kann man die Merkmale der sprachlichen Komplexität eines Textes in einer bestimmten Sprache definieren (Castilho et al. 2018: 18). Dabei spielen solche sprachlichen Elemente wie die Worthäufigkeit, Satzlänge sowie das Format und der Abstand eine wichtige Rolle. Castilho et al. (2018) erklären, dass die Lesbarkeit sich auf die Leichtigkeit bezieht, mit der ein bestimmter Text gelesen werden kann (Castilho et al. 2018: 18).

Durch die Verständlichkeit wird angegeben, inwieweit ein übersetzter Text verstanden werden kann (Popović 2020: 256). Im Vergleich zur Lesbarkeit hängt die Verständlichkeit vom Zielpublikum ab. Die Lesbarkeit hingegen hängt von einem bestimmten Text ab (Castilho et al. 2018: 19).

Durch das Konzept der Annehmbarkeit kann festgestellt werden, ob der Zieltext den Bedürfnissen und Erwartungen der Leser*innen bzw. Benutzer*innen entspricht (Castilho et al. 2018: 20). Roturier (2006) schreibt über diese Kategorie Folgendes: „Acceptability does not only refer to the relevance a text has for its receiver, but also to the manner in which its textual characteristics are going to be accepted, tolerated, or rejected by its receivers“ (Roturier 2006: 4).

Unter dem Begriff Benutzerfreundlichkeit versteht Byrne (2014) die Leichtigkeit, mit der Benutzer*innen bzw. Leser*innen auf Informationen zugreifen und diese verarbeiten können, um ihre gewünschten Aufgaben zu erledigen (Byrne 2014: 94). Die Internationale Organisation für Normung (ISO) definiert Benutzerfreundlichkeit als „extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use“ (ISO 2010: 3).

Bentivogli et al. (2018) sind der Ansicht, dass die MÜ-Evaluierung, die auf Post-Editing basiert, sowie die Evaluierung, die auf direkter Bewertung (engl. „direct assessment“) basiert, sich unter vielen manuellen MÜ-Evaluierungsmethoden am besten etablierte (Bentivogli et al.

2018: 62). In der Post-Editing-basierten Evaluierung werden die MÜ-Ergebnisse nachbearbeitet bzw. korrigiert. Dafür wird entweder ein Ausgangssatz oder eine Referenzübersetzung verwendet.

Die Evaluierung, die auf direkter Bewertung basiert, besteht darin, die von Menschen durchgeführten Bewertungen von Übersetzungen zu sammeln (Bentivogli et al. 2018: 62). Die Bewerter*innen vergeben eine Punktzahl zwischen 0 und 100, je nachdem, für wie gut sie die Übersetzung halten. Der entsprechende Ausgangstext oder die Referenzübersetzung sind vorhanden. Graham et al. (2013) stellen den folgenden Vorteil der direkten Bewertung fest:

> A direct estimation method of assessment, as opposed to ranking outputs from best-to-worst, has the advantage that it includes in annotations not only that one output is better than another, but also the degree to which that output was better than the other. (Graham et al. 2013: 33)

Chatzikoumi (2020) stellt die manuellen Metriken vor, die für die Evaluierung der Qualität des Outputs verwendet werden können, und liefert den notwendigen Hintergrund für die Projekte zur Bewertung der maschinellen Übersetzung. Die Wissenschaftlerin unterscheidet zwischen den Evaluierungsmethoden, die auf einer direkt geäußerten Beurteilung basieren – die sogenannten Directly Expressed Judgment Methods (DEJ-basierte Methoden) – und den Evaluierungsmethoden, die nicht auf einer direkt geäußerten Beurteilung basieren (Non-DEJ-basierte Methoden) (Chatzikoumi 2020: 146). Unter DEJ-basierten Methoden versteht Chatzikoumi (2020) die Evaluierungsmethoden, bei denen Bewerter*innen die Übersetzungsqualität hinsichtlich der Genauigkeit und Flüssigkeit direkt äußern (Chatzikoumi 2020: 147). In diesem Fall wird der Ausgangstext mit dem Zieltext oder der Zieltext mit der Referenzübersetzung verglichen (Chatzikoumi 2020: 147). Bei den Non-DEJ-basierten Methoden werden halbautomatische Metriken sowie Klassifizierung, Analyse und Korrektur von MÜ-Outputs verwendet (Chatzikoumi 2020: 147). Die Bewerter*innen sind in den Prozess der Evaluierung miteinbezogen, allerdings äußern sie ihre Beurteilung nur indirekt.

Zu den bekanntesten Möglichkeiten der manuellen Evaluierung zählen die Metriken der Fehlerklassifizierung und Fehleranalyse (Chatzikoumi 2020: 150), auf die im nächsten Unterkapitel genauer eingegangen wird.

2.3.1 Metriken der Fehlerklassifizierung und Fehleranalyse

Da die manuelle Evaluierung auf die subjektive Entscheidung der Bewerter*innen angewiesen ist, kommt es nicht selten vor, dass die Meinungen verschiedener Bewerter*innen voneinander sehr stark abweichen (Lommel et al. 2014: 457). Aus diesem Grund entstand in der

Translationswissenschaft in den 90er Jahren der Wunsch nach einer universellen objektiven Metrik für die Evaluierung der Übersetzungsqualität. Allerdings ist es praktisch unmöglich, eine universelle Evaluierungsmetrik zu entwickeln, da der Skopos bei jeder Übersetzung unterschiedlich ist und der Begriff Qualität bis jetzt nicht klar und universell definiert werden kann (Lommel et al. 2014: 457). Obwohl das Thema der Übersetzungsqualität sehr heikel ist, ist es Forscher*innen gelungen, unterschiedliche Metriken zu schaffen, die zwar nicht universell sind, aber für bestimmte Bereiche hilfreich sein könnten.

Die Notwendigkeit, Metriken zur Bewertung der Übersetzungsqualität zu entwickeln, hat ebenfalls eine wirtschaftliche Grundlage: In einigen Branchen ist die Dokumentation von Produkten und Dienstleistungen so umfangreich, dass eine klassische Qualitätsprüfung zu teuer und zeitaufwändig wäre (Stejskal 2006: 15). Mittels der Qualitätsmetriken wurde es möglich, die gesamte Übersetzungsqualität zu beurteilen und wiederkehrende Probleme zu identifizieren (Stejskal 2006: 15f.). Zu den bekanntesten Qualitätsmetriken der manuellen Evaluierung gehören The LISA Quality Assurance, SAE J2450 und The Multidimensional Quality Metrics (MQM) (Läubli 2020: 12f.).

2.3.1.1 Das LISA-QA-Modell

Das Qualitätssicherungsmodell The LISA Quality Assurance (QA) wurde 1995 von der Localization Industry Standards Association (LISA 2004) entwickelt. Das ist ein eigenständiges Tool für Lokalisierungsprojekte, das für die Produktdokumentation und Benutzeroberfläche angewendet wird (Mateo 2014: 77). Dieses Modell besteht aus zahlreichen anpassbaren Vorlagen, Formularen und Protokollen, die in eine Zugriffsdatenbank eingebettet sind (Stejskal 2006: 16). Das LISA-QA-Modell besteht aus acht Fehlerkategorien. Allerdings gibt es nur eine Kategorie, die linguistische Aspekte berücksichtigt (Jiménez-Crespo 2009: 74). Die Kategorie behandelt sieben Fehlertypen: falsche Übersetzung, Genauigkeit, Terminologie, Sprache, Stil, Lokalisierung und Konsistenz (Mateo 2014: 78). Insgesamt definiert das LISA-Modell 20 bis 123 Fehlertypen und drei Schweregrade und ist somit der erste Qualitätsstandard, der sich in der Übersetzungsbranche durchsetzte (Läubli 2020: 12f.). Um akzeptabel zu sein, darf der Zieltext keine kritischen Fehler enthalten und das Verhältnis zwischen Fehlerpunkten und Gesamtwörtern darf einen bestimmten Wert nicht überschreiten (Mateo 2014: 78). Lommel et al. (2014) machen darauf aufmerksam, dass die Localization Industry Standards Association vor der Insolvenz im Jahr 2011 mit der Entwicklung einer verbesserten Version des LISA-QA-

Modells begann. Allerdings gibt es diesbezüglich keine weiteren Informationen, da diese Vereinigung nicht mehr existiert (Lommel et al. 2014: 457).

2.3.1.2 Die SAE-J2450-Metrik

Eine weitere Metrik für die Bewertung der Übersetzungsqualität namens SAE-J2450 kommt aus der Automobilindustrie. Ursprünglich diente die Metrik dazu, lediglich die Qualität der Übersetzungsergebnisse für Automobilunternehmen zu bewerten. Allerdings wurde sie später für die Bewertung der Übersetzungsqualität in verschiedenen Bereichen angewendet. Im Gegensatz zum LISA-QA-Modell legt die SAE-J2450-Metrik den Fokus ausschließlich auf die sprachliche Qualität exklusive des Stils und der Formatierung etc. (Läubli 2020: 12f.). Insgesamt beinhaltet die Qualitätsmetrik SAE-J2450 sieben Fehlerkategorien: falsche Benennung, falsche Bedeutung, Auslassung, Strukturfehler, Rechtschreibfehler, Interpunktionsfehler und sonstige Fehler (SAE 2001). Darüber hinaus wird auch der Schweregrad jedes konkreten Fehlers angegeben. Die Metrik kann unabhängig von der Ausgangssprache oder der Übersetzungsform angewendet werden, sei es eine menschliche, computergestützte oder maschinelle Übersetzung (Stejskal 2006: 16).

Unter falscher Benennung wird fehlende und/oder falsche Fachterminologie verstanden. Es geht ebenfalls um eine falsche Übersetzung von Abkürzungen, Akronymen, Zahlen, Eigennamen (wie z. B. Markennamen, Ortsnamen und Personennamen etc.) (SAE 2001: 5). Eine falsche Benennung kann sowohl ein einziges Wort als auch eine Wortkombination bzw. eine Phrase, die als eine lexikalische Einheit betrachtet wird, betreffen. Ein wichtiger Aspekt ist die Einhaltung der terminologischen Konsistenz. Wenn ein bestimmter Begriff, Gegenstand oder Sachverhalt im Zieltext nicht einheitlich und konsistent verwendet wird, wird es als eine falsche Benennung betrachtet (Atesman o. J.: 5).

Die Fehlerkategorie „falsche Bedeutung“ weist darauf hin, dass die Bedeutung des übersetzten Textabschnittes mit der Bedeutung des Ausgangstextabschnittes nicht übereinstimmt und/oder das genaue Gegenteil davon ist, was im Zieltext impliziert wird (Dalla-Zuanna 2010: 23). Wenn durch eine falsche Zeichensetzung, etwa durch einen Kommafehler, die Bedeutung des Ausgangssatzes verändert wird, zählt dieser Fehler nicht als Interpunktionsfehler, sondern wird der Kategorie „falsche Bedeutung“ zugeordnet (Atesman o. J.: 10).

Bei der Kategorie „Auslassung“ geht es um „eine zusammenhängende Textstelle“, die im Zieltext fehlt (Dalla-Zuanna 2010: 24). Das können ein Wort, ein Satz, ein Segment oder

ein Absatz etc. sein (Dalla-Zuanna 2010: 24). Es ist zu beachten, dass fehlende Füll- und Blähwörter nicht unter die Auslassung fallen, da sie den Sinn eines Textelementes nicht beeinträchtigen und somit nicht als Fehler bewertet werden.

Zu Strukturfehlern gehören Grammatik-, Syntax-, Semantik-, Kollokations- und Wortbildungsfehler (Dalla-Zuanna 2010: 24; Atesman o. J.: 8). In folgenden Fällen spricht man von Strukturfehlern:

- Der Zieltext enthält eine falsche Satzstruktur (z. B. einen Relativsatz, obwohl eine Verbalphrase erforderlich wäre) (SAE 2001: 6);
- die Wörter der Zielsprache sind korrekt, aber in der falschen linearen Reihenfolge gemäß den syntaktischen Regeln der Zielsprache (SAE 2001: 6);
- Fehler bei Kasus, Geschlecht, Zahl, Zeitform, Flexion, Präfix, Suffix und Infix (SAE 2001: 7);
- falsche Artikel / Konjunktionen / Adverbien / Präpositionen (Atesman o. J.: 8);
- falsche Kollokation: „eine Kombination von Wörtern, die semantisch zwar zutreffen, aber konventionell nicht miteinander verknüpft werden“ (Atesman o. J.: 8) und
- der Satz ergibt keinen Sinn, obwohl er syntaktisch und grammatikalisch richtig ist.

Es handelt sich um Rechtschreibfehler, wenn die Schreibweise gegen die akzeptierten Rechtschreibnormen der Zielsprache verstößt und / oder beim Textfragment ein falsches Schriftsystem benutzt wird (SAE 2001: 8). Fehlende Buchstaben, Tippfehler etc. gelten als Rechtschreibfehler (Atesman o. J.: 9). Des Weiteren wird die aktuelle Rechtschreibung beachtet, die mit alten Normen der Rechtschreibung nicht kombiniert werden darf.

Eine Zeichensetzung, die gegen die Interpunktionsregeln der Zielsprache verstößt und / oder im Text nicht einheitlich eingesetzt wird, wird als Interpunktionsfehler betrachtet (Atesman o. J.: 10). Wenn durch die Zeichensetzung die Bedeutung des Ausgangssatzes verfälscht wird, bedeutet dies, dass dieser Fehler nicht als ein Interpunktionsfehler zählt, sondern in die Kategorie „falsche Bedeutung“ gehört (Atesman o. J.: 10). Dieses Problem kann auftreten, wenn ein Beistrich an einer falschen Stelle gesetzt wird.

Fehler, die nicht eindeutig den oben beschriebenen Fehlerkategorien zuzuordnen sind, werden als sonstige Fehler kategorisiert. Es ist wichtig, noch einmal zu erwähnen, dass die Qualitätsmetrik SAE-J2450 stilistische Fehler und falsche Formatierungen nicht in Betracht zieht. Aus diesem Grund werden solche Fehler außer Acht gelassen.

Der Schweregrad der Fehler wird in zwei Unterkategorien eingeteilt: leichter und schwerer Fehler. Es handelt sich um einen schweren Fehler, wenn dieser Übersetzungsfehler mit großem Risiko oder möglichen Fehlhandlungen (Sachschaden, Personenschaden etc.) für Endbenutzer*innen verbunden ist und / oder eine erhebliche Auswirkung auf die Bedeutung des übersetzen Textes hat (Atesman o. J.: 12, SAE 2001:3, Dalla-Zuanna 2010: 24). Darüber hinaus kann ein schwerer Fehler ebenso für eine Irritation bei Endbenutzer*innen sorgen (Atesman o. J.: 12). Ein leichter Fehler hingegen wird als „ein Fehler, der zu keiner bzw. zu einer geringfügigen Irritation des Anwenders ohne die Gefahr einer Fehlhandlung führt", definiert (Dalla-Zuanna 2010: 24).

2.3.1.3 Das Multidimensional Quality Metrics Framework

Das Multidimensional Quality Metrics (MQM) Framework wurde, dank dem von der Europäischen Union geförderten Projekt QTLaunchPad, entwickelt, um die Mängel, die bisher bei der Qualitätsbewertung aufgetreten sind, zu beheben (Lommel et al. 2014: 458). Das ursprüngliche Ziel der Metrik war es, das LISA-QS-Modell zu aktualisieren, deshalb übernimmt MQM viel aus den Grundlagen des LISA-Modells (Lommel et al. 2014: 458). MQM ermöglicht die Festlegung von benutzerdefinierten Qualitätsmetriken durch die Auswahl von Fehlerkategorien und bietet Zuordnungen (Mapping) zur SAE-J2450-Metrik (Läubli 2020: 12f.). Bewertungen werden entweder mit einer eigenständigen Software oder innerhalb von Übersetzungswerkbänken durchgeführt, die die Anzahl der von den Bewerter*innen markierten Fehler aufzeichnen und, je nach Metrik, eine endgültige Punktzahl durch Kombination von Fehleranzahl und Gewichtung berechnen (Läubli 2020: 12f.).

Lommel et al. (2014) erklären, dass MQM im Unterschied zu anderen Metriken der Qualitätsbewertung so konzipiert ist, dass es flexibel und mit anderen Standards kompatibel ist, um die Qualitätsbewertung in den gesamten Dokumentenproduktionszyklus integrieren zu können (Lommel et al. 2014: 458). Im Mittelpunkt der MQM-Metrik steht eine hierarchische Auflistung von verschiedenen Aspekten bzw. Problembereichen: „The issues were restricted to just those dealing with language and format, leaving out issues from the LISA QA Model that deal with project quality" (Lommel et al. 2014: 458). MQM analysiert mehr als 100 Arten von Qualitätsproblemen. Auf der obersten Ebene ist MQM in folgende wichtige Bereiche unterteilt: Genauigkeit, Design, Flüssigkeit, Internationalisierung, Stil, Terminologie, Wahrhaftigkeit, Kompatibilität und Sonstiges (Lommel 2015).

2.4 Vor- und Nachteile der automatischen und manuellen Evaluierung der MÜ

Läubli et al. (2020) behaupten, dass eine gründliche manuelle Evaluierung trotz ihrer hohen Kosten bevorzugt werden soll, obwohl die automatischen Evaluierungsmethoden in der Systementwicklung sehr wichtig sind (Läubli et al. 2020: 654). Eine ähnliche Meinung vertreten Bentivogli et al. (2018), wenn sie sagen, dass die menschliche Evaluierung als primär zu betrachten ist, auch wenn sie kostspielig und zeitaufwendig ist (Bentivogli et al. 2018: 62).

Callison-Burch et al. (2007) äußern sich zu diesem Thema wie folgt: „While automatic measures are an invaluable tool for the day-to-day development of machine translation systems, they are an imperfect substitute for human assessment of translation quality“ (Callison-Burch et al 2007: 139). Es lässt sich aus vielen wissenschaftlichen Untersuchungen schließen, dass die Hauptnachteile der manuellen MÜ-Evaluierung darin bestehen, dass sie teurer und zeitaufwändiger ist als eine automatische Evaluierung. Allerdings gilt die manuelle Evaluierung weiterhin nach so vielen Jahren der MÜ-Forschung als sehr zuverlässig.

Gornostay (2008) verschafft einen guten Überblick über die Nachteile und Vorteile sowohl der automatischen als auch der manuellen Metriken zur MÜ-Evaluierung. Zu den Vorteilen der automatischen Metriken gehören laut der Wissenschaftlerin Aspekte wie Preisgünstigkeit, Schnelligkeit und Wiederverwendbarkeit (Gornostay 2008: 4). Außerdem ermöglichen die automatischen Qualitätsmetriken den Vergleich mehrerer Systeme sowie die Systemüberwachung und -entwicklung (Gornostay 2008: 4). Die Evaluierungsergebnisse der automatischen Qualitätsmetriken können ebenso von Nutzen sein, da dadurch die zu bearbeitenden Stellen angezeigt werden (Gornostay 2008: 4). Allerdings sind solche Metriken nicht makellos: Die aufwändige Textvorbereitung sowie Unzuverlässigkeit der automatischen Evaluierungsergebnisse und Schwierigkeiten bei der Fehlerklassifizierung zählen zu den bekannten Nachteilen der automatischen Metriken (Gornostay 2008: 4).

Die Durchführung einer hochwertigen Analyse des Bewertungsprozesses gilt als ein großer Vorteil der MÜ-Evaluierung anhand von manuellen Metriken (Gornostay 2008: 4). Nachteile dieser Metriken sind, dass sie kostspielig, zeit- und arbeitsaufwendig sowie nicht wiederverwendbar sind (Gornostay 2008: 4). Zudem müssen die bilingualen Bewerter*innen vor der Evaluierung in das Prozedere eingewiesen werden, und es ist nicht garantiert, dass sie alle Evaluierungskategorien richtig interpretieren und nicht verwechseln (Gornostay 2008: 4). White (2003) vertritt die Meinung, dass es keine perfekte Evaluierungsmethode geben wird:

> No one evaluation method will fit all needs. The obvious, and probably most important people in the world of MT, as it is practiced today, are translators, and the people who need the translations — the information consumers, if you will. There are several other groups of people,

> however, who have a stake in the success of one or more aspects of MT (e.g. end-users, managers, developers, vendors, and investors.) (White 2003: 220)

Obwohl alle aktuell vorhandenen Evaluierungsmethoden Mankos aufweisen, sind Lommel et al. (2014) der Meinung, dass die Evaluierung der Übersetzungsqualität nicht nur für kommerzielle Übersetzungen wichtig ist, sondern auch eine wichtige Rolle für die Entwicklung der maschinellen Übersetzung spielt, da dadurch festgestellt wird, wie sich die vorgenommenen Systemänderungen auf die Übersetzungsqualität des jeweiligen MÜ-Systems auswirken (Lommel et al. 2014: 456).

3 Die Komplexität der Rechtsübersetzung

Der vorliegende Abschnitt widmet sich den Besonderheiten und Schwierigkeiten der Rechtsübersetzung. Dabei werden die Themenbereiche Rechtssprache, Rechtstexte und Rechtsübersetzung unterschieden. Kapitel 2.1 beschäftigt sich mit den Eigenschaften und Merkmalen der Rechtssprache und befasst sich mit der Wechselbeziehung zwischen Sprache, Recht und Rechtssystem. Im darauffolgenden Unterkapitel wird auf unterschiedliche Funktionen der Rechtstexte eingegangen. Des Weiteren werden hierzu diverse funktionsbasierte Textklassifikationen vorgestellt. Das letzte Unterkapitel beschreibt die Schwierigkeiten und Problemstellen der Rechtsübersetzung und greift zum Teil das Thema der Übersetzungskompetenzen im Rechtsbereich auf.

3.1 Besonderheiten der Rechtssprache

Der Begriff Rechtssprache ist so komplex, dass er auf unterschiedliche Weise definiert werden kann. Mushchinina (2009) versteht darunter alle sprachlichen Phänomene, die auf dem Rechtsgebiet vorkommen (Mushchinina 2009: 42). Cao (2007) betrachtet die Rechtssprache als eine Sonderform der Gemeinsprache bzw. der Standardsprache, die für juristische Zwecke verwendet wird (Cao 2007:15). Berūkštienė (2016: 96) bezeichnet die Rechtssprache als eine Fachsprache. Da die Sprache und das Recht kommunikative Funktionen haben, muss die Rechtssprache in der Lage sein, das rechtliche Handeln zu garantieren (Mushchinina 2009: 33, 46). Cao (2007) ist ebenso der Meinung, dass die Rechtssprache die gesamte Kommunikation in einem rechtlichen Kontext umschließt (Cao 2007:10).

Da sich Rechtssysteme einzelner Länder voneinander deutlich unterscheiden, besitzt jede Rechtssprache ihren eigenen Stil, spezielle Fachterminologie und Ausdrucksweise (Cao 2007; Dumitrescu 2014). Dennoch gibt es ebenso typische Merkmale, welche die Rechtssprache im Allgemeinen von der Gemeinsprache differenzieren (Dumitrescu 2014; Cao 2007). Zu den allgemeinen Merkmalen der Rechtssprache gehören nach Dumitrescu (2014) veraltete Syntax bzw. syntaktische Konstruktionen, lange komplexe Sätze, zusammengesetzte Phrasen und bestimmte Wortauswahl (Dumitrescu 2014: 505). Den juristischen Stil schriftlicher Rechtssprache beschreibt Cao (2007) als einen sachlichen formalen Stil, in dem Rechte und Pflichten mittels Deklarativsätzen verkündet werden (Cao 2007: 22). Des Weiteren spricht die Wissenschaftlerin ebenso über das Problem der Länge und Komplexität der Sätze, was auf die Komplexität der Sachverhalte zurückzuführen ist (Cao 2007:21). Eine der wichtigsten Besonderheiten der Rechtssprache ist die Verwendung der spezifischen komplexen

Rechtsterminologie (Cao 2007: 20). Ein interessanter Aspekt der Rechtssprache wird von Mushchinina (2009) aufgegriffen, indem sie über ihre Dualität spricht (Mushchinina 2009: 35). Damit meint die Wissenschaftlerin zum einen die starke Gebundenheit der Rechtssprache an das Regelungssystem. Zum anderen spricht Mushchinina (2009: 35) davon, dass die Rechtssprache einen Freiraum für Interpretationen bietet, wodurch eine Weiterentwicklung des Rechts im Einklang mit der aktuellen Rechtslage ermöglicht wird.

Den Stil der deutschen Rechtssprache charakterisiert Cao (2007) als deutlich und klar und begründet dies mit der Tatsache, dass die Entwicklung des deutschen Rechts immer systematisch und logisch sowie abstrakt und konzeptuell verlief (Cao 2007: 22). Dies spiegelt sich in der deutschen Rechtsterminologie wider und macht sie dadurch ebenso abstrakt mit einem hohen Anteil an Substantiven. Das Besondere an der deutschen Rechtssprache ist die Verwendung von mehreren attributiven Adjektiven (Cao 2007: 21). An dieser Stelle ist es wichtig zu erwähnen, dass es erhebliche Unterschiede zwischen der deutschen Rechtssprache Deutschlands, Österreichs, Belgiens, Liechtensteins und der Schweiz gibt, da alle Standardvarietäten der deutschen Rechtssprache an ein anderes Rechtssystem gebunden sind (Daum 2003: 38). Muhr (2009) analysiert rechtsterminologische Unterschiede zwischen der österreichischen und deutschen Rechtssprache (RS) und kategorisiert sie wie folgt:

1) Begriffslücke, wenn der Begriff in Form sowie Inhalt in der anderen RS fehlt;
2) Begriffslücke, wenn ein bestimmtes Konzept in der anderen RS fehlt, jedoch ein Begriff mit einer Teil-Entsprechung bezüglich der Funktionalität und des Inhalts vorliegt;
3) Benennungslücke, wenn das Konzept in der anderen RS vorliegt, aber eine vergleichbare Benennung fehlt;
4) Unterschiedliche Benennungen für ein identisches Konzept;
5) Inhaltliche Äquivalenz mit einem geringen formalen Unterschied;
6) Teilsynonymie;
7) Äquivalenter Hauptbegriff mit nur teiläquivalenten Nebenbegriffen;
8) Falsche Freunde und
9) Mehrfachrelationen (Muhr 2009: 211–214).

Derartige Unterschiede kann man gut an folgenden Beispielen sehen: Landeshauptmann (AUT) vs. Ministerpräsident (DEU) vs. Regierungspräsident (CHE); Bezirksgericht (AUT) vs. Amtsgericht (DEU); Landesgericht (AUT) vs. Landgericht (DEU) (Pohl 2011: 67). Der Begriff Totschlag wird beispielweise in Österreich als „eine mit einer zwingenden Strafmilderung verbundene besondere Form des vorsätzlichen Tötungsdeliktes“ definiert (Wikipedia 2021). In

Deutschland hingegen versteht man darunter „die vorsätzliche Tötung eines Menschen, die weder die Strafandrohung erhöhenden Kriterien für Mord noch die die Strafandrohung mindernde Kriterien für eine Tötung auf Verlangen erfüllt“ (Wikipedia 2022). Daher sollen sich Translator*innen dieser rechtsterminologischen Unterschiede bewusst sein und die Terminologie der Zielkultur im Hinblick auf die entsprechende Rechtsordnung verwenden.

3.2 Funktion und Typologie der Rechtstexte

Ein Rechtstext wird von Berūkštienė (2016) als ein Text verstanden, der in einer Rechtssprache verfasst wurde und sowohl von Rechtsexpert*innen als auch von Lai*innen für juristische Zwecke in einem rechtlichen Umfeld verwendet wird (Berūkštienė 2016: 95). Aus der Sicht von Šarčevic (1997) ist ein Rechtstext ein kommunikatives Ereignis (engl. „communicative occurrence“), das zu einem bestimmten Zeitpunkt und an einem bestimmten Ort produziert wird und eine bestimmte Funktion erfüllt (Šarčevic 1997: 9).

Berūkštienė (2016) spricht von den drei wichtigsten Funktionen der Rechtstexte, nämlich über eine informative, regulative und präskriptive Funktion (Berūkštienė 2016: 112). Dabei betont die Wissenschaftlerin die Wichtigkeit der präskriptiven Funktion für Rechtstexte, da es sich meistens um die Festlegung und das Vorschreiben von Verpflichtungen, Verboten oder Erlaubnissen handelt (Berūkštienė 2016: 112). Zudem befasst sie sich mit den Eigenschaften der Rechtstexte und stellt fest, dass sie gewisse funktionale, strukturelle und sprachliche Merkmale aufweisen, wodurch sich diese Texte von anderen Textsorten unterscheiden lassen (Berūkštienė 2016: 95, 111f.). Dies ist auf die Besonderheiten der Rechtssprache zurückzuführen, die im vorigen Unterkapitel besprochen wurden.

Da der Rechtsbereich sehr umfangreich ist, gibt es keine allgemeine Texttypologie der Rechtstexte, die alle rechtlichen Aspekte berücksichtigt. Aus diesem Grund können Rechtstexte auf unterschiedliche Weise klassifiziert werden, je nachdem, auf welchem Prinzip die Klassifizierung beruht. Berūkštienė (2016:112) nennt folgende mögliche Kriterien für die Klassifikationen dieser Art von Texten: 1) die Funktion der Rechtstexte, 2) ihre Einsatzsituation, 3) ihre Rechtsgebiete, 4) Subgruppen der Rechtsexpert*innen und 5) die Formalität des Stils und die Form der Medien. Daum (2003) ist der Meinung, dass im Vordergrund einer Klassifizierung der Rechtstexte auch weitere diverse Aspekte stehen können, wie Textfunktion, Inhalt, Abstraktion oder Konkretheit sowie Generalisierung oder Individualisierung (Daum 2003: 37). Allgemein gesehen, sind es textexterne oder textinterne Faktoren, die als Basis für schon vorhandene Typologien der Rechtstexte dienen (Berūkštienė

2016: 112). Eine detaillierte Ausführung aller von den Forscher*innen vorgeschlagenen Klassifikationen würde den Rahmen dieser Arbeit sprengen. Aus diesem Grund wird an dieser Stelle nur auf die Texttypologie der Rechtstexte nach ihrer Funktion eingegangen.

Cao (2007) unterscheidet zwischen präskriptiven und deskriptiven Rechtstexten. Laut der Wissenschaftlerin dienen präskriptive Rechtstexte normativen Zwecken, um rechtliche Sachverhalte festzustellen und Rechte und Pflichten zu begründen (Cao 2007: 10). Deskriptive Rechtstexte hingegen erfüllen informative Zwecke. Dabei geht es um rechtswissenschaftliche Werke, Rechtsberatungen, juristische Korrespondenz etc. (Cao 2007: 10).

Šarčevic (1997) schlägt eine ausführlichere Klassifikation nach dem gleichen Prinzip vor. Die Forscherin teilt Rechtstexte in 1) primär präskriptive, 2) primär deskriptive und ebenso präskriptive und 3) rein deskriptive Texte ein (Šarčevic 1997: 11). Zu der ersten Kategorie gehören nach Šarčevic (1997: 11) Verträge, Gesetze, Vereinbarungen, Verordnungen etc. Die zweite Kategorie beinhaltet gerichtliche Entscheidungen, Anträge, Berufungen, Klagen, Schriftsätze, Petitionen etc. (Šarčevic 1997: 11). Die letzte Kategorie aus dieser Klassifizierung bezieht sich auf die von Rechtswissenschaftler*innen verfassten Werke und umfasst beispielsweise Lehrbücher, Artikel, Rechtsgutachten (Šarčevic 1997: 11).

Daum (2003) stellt eine Einteilung der Rechtstexte vor, welche die Funktion und den Inhalt berücksichtigt. Aus seiner Sicht lassen sich Texte des Rechtsbereichs in Normtexte, rechtsanwendende Entscheidungen, Formulartexte und informative Texte einteilen (Daum 2003: 35). Normtexte befassen sich mit allgemeinen und individuellen Normen oder mit einer Kombination aus beiden Normen. Unter den allgemeinen Normtexten versteht Daum (2003: 36) Gesetze, Verordnungen und weitere Rechtsnormen, die den öffentlichen Bereich regulieren. Die individuellen Normtexte befassen sich wiederum mit dem Privatrecht. Zu dieser Kategorie gehören Verträge und Vereinssatzungen (Daum 2003: 36). Allgemeine Geschäftsbedingungen betrachtet er als einen kombinierten Normtext, in dem die allgemeinen und individuellen Normen miteinander verbunden sind (Daum 2003: 36). Zu rechtsanwendenden Entscheidungen zählt Daum (2003: 36) Beschlüsse, Urteile sowie verschiedene Verwaltungsakte. Formulartexte sind Mustertexte und gerichtliche sowie behördliche (Antrags-)Formulare (Daum 2003: 36). Informative Texte umfassen wiederum Texte, die rechtliche Informationen beinhalten, wie z. B. Rechtsratgeber, Lehrbücher, Merkblätter (Daum 2003: 37).

Berūkštienė (2016: 112) betont die Wichtigkeit der Textsortenbestimmung eines jeweiligen Rechtstextes für die Erstellung einer qualitativen Übersetzung mittels der für die jeweilige Textsorte angemessenen Übersetzungsstrategie.

3.3 Problembereiche der Rechtsübersetzung

Die Rechtsübersetzung sollte man nicht als eine automatische Wiedergabe eines Konzepts aus einer Ausgangssprache in eine Zielsprache betrachten, sondern als die Übertragung aus einem Rechtssystem in das andere (Cao 2007: 24, 55; Daum 2003: 38). Nach Cao (2007) gehört die Übersetzung von Rechtskonzepten zu einem der Hauptprobleme der Rechtsübersetzung (Cao 2007: 32). Unter Rechtskonzepten versteht die Wissenschaftlerin abstrakte Bezeichnungen für die grundlegenden Rechtsvorstellungen und Regeln eines Rechtssystems, welche die Systematisierung und Strukturierung des Rechts erleichtern (Cao 2007: 54). Rechtskonzepte verdeutlichen die Relation zwischen Recht und Sprache und helfen rechtliche Fragen zu klären (Mushchinina 2009: 33).

Mushchinina (2009), Chirilă (2014) und Dumitrescu (2014) betrachten richtige Terminologie und Phraseologie als die größte Herausforderung von Rechtsübersetzungen. Dumitrescu (2014) stellt fest, dass die Rechtsterminologie oft nicht nur Rechtsbegriffe, sondern auch Fachbegriffe aus anderen Bereichen umfasst (Dumitrescu 2014: 504). Als Beispiel führt die Forscherin (2014: 504) die Zulassung eines Medikaments an. Cao (2007: 53f.) hebt folgende vier terminologische Problembereiche der Rechtsübersetzung hervor: 1) Rechtskonzepte und ihre Äquivalenz, 2) Rechtsbegriffe, die mit Gesetzen und Rechtsinstitutionen in Verbindung stehen, 3) rechtliche Bedeutungen und Synonyme sowie 4) terminologische Undeutlichkeit und Ambiguität.

Dumitrescu (2014) weist darauf hin, dass terminologische Probleme durch die Abhängigkeit der Rechtsterminologie von einem bestimmten Rechtssystem entstehen. Die Wissenschaftlerin erklärt, dass es zwar Länder mit der gleichen Amtssprache gibt, jedoch verfügt jedes Land über ein eigenes Rechtssystem und somit über eigene Rechtsterminologie (Dumitrescu 2014: 503). Der Grund hierfür liegt in den kulturellen und historischen Unterschieden. Gotti (2016) spricht davon, dass es aufgrund kultureller Unterschiede keine einheitliche Version desselben Rechtsdokuments für alle Sprachen gibt, was er als größtes Problem für die Rechtsübersetzung ansieht. Er weist darauf hin, dass Rechtsdokumente in jeder Gesellschaft ähnliche Themen behandeln und es daher notwendig wäre, die formale Übereinstimmung zwischen den unterschiedlichen landesspezifischen Versionen desselben Dokuments zu wahren (Gotti 2016: 7). Allerdings ist Dumitrescu (2014) der Überzeugung, dass nicht der Stil des Ausgangstextes, sondern stilistische Regeln des Zieltextes für die Anfertigung einer Rechtsübersetzung entscheidend sind (Dumitrescu 2014: 506).

Da professionelle Übersetzer*innen von Rechtstexten viele Aspekte berücksichtigen, sehr viel Rechercheaufwand betreiben und mit dem Thema und Inhalt des Dokumentes vertraut sein müssen, ist die Rechtsübersetzung besonders schwierig. Außerdem tragen Rechtsübersetzer*innen eine große Verantwortung, da eine fehlerhafte Übersetzung zu einem Prozessverlust oder zu Haftungsproblemen führen kann. Die Qualität einer guten Rechtsübersetzung hängt unbestritten von Kompetenzen der Übersetzer*innen ab. Chirilă (2014: 492) und Dumitrescu (2014: 504) erwähnen drei Herausforderungen in der Rechtübersetzung bzw. drei wichtige Kompetenzen eines / einer Rechtsübersetzers / Rechtsübersetzerin: 1) die Beherrschung des spezifischen Schreibstils der Rechtstexte in der Zielsprache, 2) die Vertrautheit mit der relevanten Terminologie und 3) Grundkenntnisse der Rechtssysteme des Ausgangs- und Ziellandes. Grundsätzlich ist es beim Übersetzen von Rechtstexten erforderlich, den rechtlichen Status und den kommunikativen Zweck der Ausgangs- und Zieltexte festzustellen (Cao 2007: 10). Darüber hinaus muss die Struktur eines Rechtstextes berücksichtigt werden. Unter strukturellen Eigenschaften versteht Berūkštienė (2016: 98) das Format, den Aufbau, die Gliederung und den Zusammenhang zwischen verschiedenen Teilen eines Rechtstextes. Gotti (2016) spricht von der durchzuführenden Konzeptanalyse des Ausgangstextes für die Feststellung der Unterschiede zwischen zwei oder mehreren Rechtssystemen und für die Suche nach einem passenden Äquivalent (Gotti 2016: 7). Zusammenfassend lässt sich sagen, dass die Schwierigkeiten der Rechtsübersetzung hauptsächlich mit den Besonderheiten der Rechtssprache zusammenhängen, die im Unterkapitel 2.1 zu finden sind.

4 Aktueller Stand der Forschung

Es gibt eine geringe Anzahl an Forschungsarbeiten, die sich mit der maschinellen Übersetzung von Rechtstexten in der Sprachkombination Russisch-Deutsch beschäftigen. In diesem Kapitel werden wissenschaftliche Arbeiten vorgestellt, die sich mit der Qualitätsbewertung maschinell übersetzter Rechtstexte in verschiedenen Sprachpaaren befassen.

Killman (2014) untersuchte die Genauigkeit der maschinell erstellten Rechtsübersetzungen von Google Translate für das Sprachpaar Spanisch-Englisch. Das Ziel der Studie bestand darin, die Genauigkeit des Vokabulars bzw. der Terminologie und Phraseologie zu überprüfen und festzustellen, ob die SMÜ im Rechtsbereich hilfreich und sinnvoll wäre. Killman (2014) nahm an, dass die Verwendung der MÜ für das Übersetzen von Rechtstexten lohnenswert sein könnte, wenn die Ergebnisse eine angemessene Menge an korrekten Übersetzungen lieferten, was für Übersetzer*innen eine große Hilfe bei zeitaufwändigen Recherchen bedeuten würde (Killman 2014: 85). Für die Studie wurde eine breite Stichprobe von spanischen Rechtsbegriffen und Kollokationen (621 Vokabeln) erstellt, die aus verschiedenen Texten von Urteilen des spanischen Verfassungsgerichtshofs zusammengestellt wurden (Killman 2014: 85). Anhand einer vom Autor angefertigten Referenzübersetzung wurde die maschinell erstellte Übersetzung satzweise nach den zwei Kriterien Flüssigkeit und Adäquatheit evaluiert.

Die Ergebnisse der Untersuchung zeigten, dass es dem damaligen SMÜ-System Google Translate gelungen ist, in knapp über 64 % der Fälle die Terminologie und Kollokationen korrekt zu übersetzen (Killman 2014: 94f.). Angesichts dieser Ergebnisse kam Killman (2014) zum Schluss, dass, obwohl die SMÜ immer noch Probleme mit dem Erkennen des richtigen Kontexts aufwies, die gelungenen Übersetzungen auf den Fortschritt im Bereich der MÜ bzw. den Erfolg der phrasenbasierten statistischen Methoden deuteten (Killman 2014: 96). Laut dem Forscher hängen die Übersetzungsergebnisse direkt davon ab, ob die Datenbank genügend entsprechende Paralleltexte enthält (Killman 2014: 96).

Azer und Aghayi (2015) befassten sich mit der Bewertung der Übersetzungsqualität der MÜ-Systeme Google Translate und Padideh. Die Forscher untersuchten maschinell erstellte Übersetzungen aus dem Englischen ins Persische und gingen der Frage nach, ob maschinelle Übersetzungen unterschiedlicher Textsorten aus der Perspektive der Bewerter*innen verständlich und akzeptabel waren (Azer & Aghayi 2015: 226f.). Dabei wurde genau analysiert, wie akzeptabel und nützlich maschinell übersetzte Texte welcher Textsorte für die Bewerter*innen waren (Azer & Aghayi 2015: 226). Für die Untersuchung wurden sechs

unterschiedliche Textsorten ausgewählt: eine Kindergeschichte, ein Text aus dem politischen Bereich, ein Text aus der Computerwissenschaft, ein Rechtstext, ein Gedicht und eine Webseite (Azer & Aghayi 2015: 229). Die Evaluierung wurde auf der Grundlage der von Van Slype (1979) vorgeschlagenen Evaluierungskriterien von 16 Bewerter*innen manuell durchgeführt.

Die Wissenschaftler Azer & Aghayi (2015) entschieden sich für folgende sechs Kriterien, welche die Makroevaluierung (Van Slype 1979: 12f.), sprich die Bewertung der Gesamtleistung des Systems, umfassten: Verständlichkeit (engl. „intelligibility"), Wiedergabetreue (engl. „fidelity"), Kohärenz (engl. „coherence"), Benutzerfreundlichkeit (engl. „usability"), Annehmbarkeit (engl. „acceptability") und Lesezeit (engl. „reading time") (Azer & Aghayi 2015: 228f.). Den Bewerter*innen wurden Fragebogen mit verschiedenen Skalen für die jeweilige Evaluierungskategorie vorgelegt. Demzufolge wurden die Sätze in der Kategorie „Verständlichkeit" anhand einer Vier-Punkte-Skala bewertet: unverständlich (0 Punkte), wenig verständlich (1 Punkt), ziemlich gut verständlich (2 Punkte) und sehr gut verständlich (3 Punkte) (Azer & Aghayi 2015: 228). Hierbei gab es mehrere Versionen der zu bewertenden Sätze (eine Version pro Bewerter*in): Originalsätze, maschinell übersetzte Sätze inklusive oder exklusive des Post-Editings sowie humanübersetzte Sätze mit oder ohne Revision (Azer & Aghayi 2015: 228).

Was die Kategorie Wiedergabetreue betraf, so erhielten die Bewerter*innen Textproben, die aus den Ausgangs- und Zieltexten bestanden und satzweise überprüft werden mussten. Für die Bewertung wurde ebenso eine Vier-Punkte-Skala nach einem ähnlichen Punktevergabeprinzip verwendet: vollständig oder fast vollständig falsch (0 Punkte), kaum originalgetreu (1 Punkt), ziemlich originalgetreu (2 Punkte) und vollständig oder fast vollständig originalgetreu (3 Punkte) (Azer & Aghayi 2015: 228).

In den Kategorien Kohärenz, Benutzerfreundlichkeit und Annehmbarkeit wurde die Qualität der maschinell erstellten Übersetzungen durch eine direkte Befragung der Bewerter*innen gemessen (Azer & Aghayi 2015: 228f.). Die letzte Kategorie Lesezeit maß die für das Lesen des jeweiligen maschinell übersetzten Textes aufgebrachte Zeit (Azer & Aghayi 2015: 229).

Die Ergebnisse zeigten, dass die maschinellen Übersetzungen von Google Translate und Padideh eine mittlere oder zum Teil unterdurchschnittliche Qualität aufwiesen (Azer & Aghayi 2015: 236). Genauer betrachtet lag der Gesamtdurchschnittswert in den Kategorien Verständlichkeit, Wiedergabetreue und Kohärenz für Google Translate bei 47,77 %, 49 % sowie 55 % und für Padideh bei 23,33 %, 29,22 % und 36,66 % (Azer & Aghayi 2015: 236). Die besten Evaluierungsergebnisse in Verständlichkeit, Wiedergabetreue und Kohärenz

erzielte der computerwissenschaftliche Text (Azer & Aghayi 2015: 237). Der Rechtstext hingegen zählte zu der Textsorte mit den schlechtesten Ergebnissen in den Kategorien Wiedergabetreue und Kohärenz und belegte den vorletzten Platz in der Kategorie Verständlichkeit (Azer & Aghayi 2015: 237). Azer und Aghayi (2015) verglichen die Evaluierungsergebnisse der beiden MÜ-Systeme in den Kategorien Benutzerfreundlichkeit, Annehmbarkeit und Lesezeit und kamen zum Schluss, dass Google Translate darin deutlich besser als Padideh abschnitt (Azer & Aghayi 2015: 237). Allerdings stellten sie fest, dass die beiden MÜ-Systeme zum Zeitpunkt der Untersuchung bei der Erstellung von maschinellen Übersetzungen aus dem Englischen ins Persische in allen sechs Textsorten noch gravierende Schwächen aufwiesen (Azer & Aghayi 2015: 236).

Wiesmann (2019) setzte sich mit der neuronalen MÜ von Rechtstexten auseinander. Die Wissenschaftlerin beschäftigte sich mit zwei Forschungsfragen, und zwar, inwieweit die NMÜ in der Lage war, Texte aus dem Bereich der Gesetzgebung, der Rechtspraxis und der Rechtstheorie so qualitativ hochwertig zu übersetzen, dass sich der Aufwand für das Post-Editing auf ein Minimum beschränken würde, und welche Rolle die maschinelle Übersetzung im Fachübersetzungskurs Italienisch-Deutsch an der Universität Bologna spielen sollte (Wiesmann 2019: 128). Schließlich wurden folgende Texte zum Übersetzen ausgewählt: ein Auszug aus einem Gesetz, ein juristischer Aufsatz, eine Vollmacht, ein notarieller Grundstückskaufvertrag, eine Klageschrift und ein zivilrechtliches Urteil. Übersetzt wurde vom Italienischen ins Deutsche. Für die Beantwortung der Forschungsfragen verwendete Wiesmann (2019) den MÜ-Dienst DeepL sowie das CAT-System MateCat, das die Integrierung der MÜ ermöglicht. Die Evaluierung der ausgesuchten Texte wurde von den Teilnehmenden des Übersetzungskurses an der Universität Bologna manuell durchgeführt. Zu den Evaluierungskriterien der Studie gehörten die Verständlichkeit und Aussagefähigkeit des Zieltextes sowie die Übereinstimmung zwischen dem Ausgangs- und Zieltext unter Berücksichtigung der jeweiligen Übersetzungssituation (Wiesmann 2019: 118, 130). In Hinblick auf den Zieltext wurde die morphologische, syntaktische, lexikalische und semantische Richtigkeit evaluiert (Wiesmann 2019: 130f.). Beim Vergleich des Ausgangs- und Zieltextes bestand die Aufgabe der Evaluator*innen darin, Punkte von 0 bis 3 für jeden Satz zu vergeben, wobei 0 „ganz schlecht“ und 3 „sehr gut“ bedeutete.

Die Ergebnisse zeigten, dass der juristische Aufsatz und der Auszug aus einem Gesetz sowohl in der Kategorie Verständlichkeit und Aussagefähigkeit des Zieltextes als auch in der Evaluierungskategorie Übereinstimmung zwischen dem Ausgangs- und Zieltext unter Berücksichtigung der jeweiligen Übersetzungssituation die höchste Punktezahl erreichten.

Allerdings zeigte die zweite Kategorie insgesamt deutlich schwächere Ergebnisse als die erste. In den meisten Fällen (abgesehen von der Vollmacht und dem Gerichtsurteil) erzielte DeepL bessere Übersetzungsergebnisse als MateCat. Angesichts der Ergebnisse kommt Wiesmann (2019: 148) zum Schluss, dass die Übersetzungen aus dem Italienischen ins Deutsche, die mithilfe von DeepL und MateCat erzeugt wurden, insgesamt noch unbefriedigend waren. Die Wissenschaftlerin stellte fest, dass zum Zeitpunkt der Untersuchung die Entwicklung der neuronalen maschinellen Übersetzung noch nicht weit genug fortgeschritten war, um maschinell übersetzte juristische Texte mit nur minimalem Einsatz von Post-Editing zu verwenden (Wiesmann 2019: 149). Was die zweite Forschungsfrage bezüglich der Verwendung der MÜ im Fachübersetzungskurs an der Uni Bologna angeht, ist Wiesmann (2019: 149) der Ansicht, dass es nicht sinnvoll ist, dem Post-Editing von maschinell übersetzten Rechtstexten noch mehr Aufmerksamkeit zu widmen. Laut der Wissenschaftlerin ist es wichtig, dass die Studierenden sich der derzeitigen Grenzen der MÜ bewusst sind und sich den juristischen Fachkenntnissen widmen (Wiesmann 2019: 149). Wiesmann (2019: 149) war überzeugt, dass die Beschäftigung mit der Fachliteratur ein besseres Verständnis davon schafft, ob eine Übersetzung flüssig, verständlich und sinnvoll ist und inwieweit ein Ausgangs- mit seinem Zieltext übereinstimmt.

Vieira et al. (2021) untersuchten die Auswirkungen des Missbrauchs von MÜ in medizinischen und juristischen Fachbereichen. Die Wissenschaftler*innen führten eine kritische Literaturrecherche durch, um eine qualitative Meta-Analyse der Risiken und des Potenzials von MÜ als Kommunikationsmittel in den beiden Bereichen zu präsentieren (Vieira et al. 2021: 1515). Vieira et al. (2021) untersuchten zwei Aspekte, nämlich wie MÜ heutzutage in medizinischen und juristischen Bereichen wahrgenommen und eingesetzt wird und wie MÜ die Kommunikation in diesen Bereichen beeinflusst (Vieira et al. 2021: 1516). Als Grundlage für die Untersuchung dienten offizielle Dokumente und veröffentlichte Forschungsarbeiten. Die Wissenschaftler*innen beschäftigten sich mit unbearbeiteten bzw. nicht posteditierten maschinellen Übersetzungen.

Was lediglich die Analyse des MÜ-Einsatzes im Rechtsbereich betrifft, so weisen die Ergebnisse darauf hin, dass das Problembewusstsein hinsichtlich der von MÜ ausgehenden potenziellen Risiken in diesem Bereich sehr unterschiedlich ausgeprägt ist. Die Forscher*innen führen Beispiele von unkontrollierter und kontrollierter Verwendung von MÜ in einigen Behörden an. So befassten sich einige Gerichte mit der Problematik des MÜ-Einsatzes und stellten anschließend Hinweise zur Verwendung von MÜ bereit. Darüber hinaus wurden nicht-posteditierte maschinelle Übersetzungen bei manchen Justizbehörden nicht akzeptiert und als

eine nicht konforme Übersetzungsmethode bezeichnet (Vieira et al. 2021: 1522). Es gibt jedoch auch Beispiele von Umständen, in denen die Verwendung von MÜ-Systemen ohne gründliche Recherche und Überprüfung der Fachtermini Menschen in heikle Situationen brachte, wie zum Beispiel beim Widerruf eines Schuldbekenntnisses (Vieira et al. 2021: 1522).

Die allgemeine Untersuchung zeigte, dass in der Verwendung von MÜ für zu übersetzende Texte und / oder für zu dolmetschende Gespräche mit dem Schwerpunkt Recht und Medizin die Komplexität einer bestimmten Sprache oder eines konkreten Sprachpaars in den meisten Fällen nicht berücksichtigt wurde. Ähnlich wie Wiesmann (2019) sind Vieira et al. (2021) der Meinung, dass Menschen eine bessere Vorstellung bzw. Kenntnis von Stärken und Schwächen von MÜ benötigen. Daher empfehlen die Forscher*innen (Vieira et al. 2021: 1515), interdisziplinäre Studien durchzuführen, damit diejenigen, die im Gesundheits- und Rechtsbereich tätig sind, von Übersetzungswissenschaftler*innen über die Schwachstellen der MÜ anhand von empirischen Daten informiert werden können. Die Forschungsgruppe kommt ebenfalls zu der Feststellung, dass die aktuellen MÜ-Systeme mit Vorsicht zu verwenden sind, da nicht überprüfte maschinell erstellte Übersetzungen zur Verschärfung sozialer Ungleichheiten führen und bestimmte Gruppen einem höheren Risiko aussetzen können (Vieira et al. 2021: 1515).

Welnitzova et al. (2021) untersuchten die Qualität der maschinellen Übersetzung von Rechtstexten im Sprachpaar Slowakisch-Englisch. Die Studie zielte darauf ab, gravierende Fehler eines maschinell übersetzten Rechtstextes in der erwähnten Sprachrichtung zu identifizieren. Die Forscher*innen entschieden sich für die Analyse einer Rückübersetzung, wobei für den Ausgangstext eine Humanübersetzung aus dem Slowakischen ins Englische angefertigt wurde, die folgend aus dem Englischen ins Slowakische maschinell übersetzt wurde (Welnitzova et al. 2021: 223f.). Bei der Suche nach einem MÜ-System für die Erstellung einer maschinellen Übersetzung fiel ihre Wahl auf Google Translate. Was die Textauswahl betrifft, so wählten Welnitzova et al. (2021) einen in slowakischer Sprache verfassten Strafrechtskodex im Umfang von 16 Normseiten aus (Welnitzova et al. 2021: 224).

Mithilfe der Multidimensionalen Qualitätsmetrik (MQM) wurde die Qualitätsevaluierung durchgeführt. Dabei wurde der Schwerpunkt der Untersuchung auf die Ebene der Flüssigkeit (engl. „fluency“) gelegt, die aus folgenden Kategorien bestand: Mehrdeutigkeit, Zeichenkodierung, Kohärenz, Kohäsion, Korpuskonformität, Duplikation, Grammatik, grammatisches Register, Inkonsistenz, Verzeichnis, unerlaubte Zeichen, Beleidigung, Musterproblem, Sortierung, Rechtschreibung, Typografie und Unverständlichkeit (Welnitzova et al. 2021: 224).

Welnitzova et al. (2021) berechneten die Fehlerhäufigkeit und posteditierten die fehlerhaften Segmente. Anschließend wurde die maschinell erstellte Übersetzung mit der Humanübersetzung verglichen. Um die Unterschiede und Übereinstimmung der beiden Versionen zu kontrollieren und den Eintritt sowie die Beziehungen der MQM-Flüssigkeit-Fehlerkategorien zu analysieren, wurden unterschiedliche Berechnungsmethoden verwendet, nämlich der Kendall-Koeffizient, Cochrane-Q-Test, Tukey-HSD-Test und Cohens Kappa (Welnitzova et al. 2021: 225).

Laut den Untersuchungsergebnissen zählten die festgestellten Terminologie-, Flexions- und Konjunktionsfehler zu den häufigsten Fehlern in der analysierten Übersetzung (Welnitzova et al. 2021: 229). Darüber hinaus konnten 237 Kongruenz- und 98 Mehrdeutigkeitsfehler identifiziert werden (Welnitzova et al. 2021: 229). Welnitzova et al. (2021) stellten fest, dass obwohl die analysierte maschinelle Übersetzung gut verständlich war, viele Segmente korrigiert werden mussten (Welnitzova et al. 2021: 229). Angesichts der Ergebnisse ziehen die Wissenschaftler*innen das Fazit, dass die maschinelle Übersetzung des Rechtstextes aufgrund seiner Komplexität in Bezug auf die Terminologie, Genauigkeit und Eindeutigkeit gründlich revidiert und nachbearbeitet werden muss (Welnitzova et al. 2021: 229).

Wenn man die oben erwähnten Arbeiten miteinander vergleicht, so kommt man zum Schluss, dass die Komplexität der Rechtssprache und vor allem die richtige Terminologie für MÜ-Systeme eine Herausforderung darstellt. Ein wichtiger Punkt bei der Evaluierung der Ergebnisse ist die Frage, wofür, wie und für wen MÜ eingesetzt wird.

5 Methodik

Dieser Abschnitt gehört zum praktischen Teil der Arbeit und beschreibt die Vorgehensweise der geplanten Untersuchung. Im Unterkapitel 5.1 werden die Methode und das Konzept der Untersuchung vorgestellt. Darauf aufbauend werden in den nächsten Unterkapiteln die Text-, System- und Sprachauswahl erklärt. Nachfolgend wird auf die ausgesuchte Fehlertypologie eingegangen. Im Fokus des Unterkapitels 5.5 steht die Auswahl der Annotator*innen, welche die maschinell übersetzten Texte auf Fehler überprüfen. Abschließend wird der gesamte Evaluierungsprozess beschrieben.

5.1 Methode und Konzept der Untersuchung

Um die Forschungsfrage beantworten zu können, wurde ein Konzept für die empirische Untersuchung entwickelt. Im Mittelpunkt der Untersuchung steht die Übersetzungsqualität. Des Weiteren soll analysiert werden, welche Arten von Fehlern maschinell übersetze Texte am häufigsten aufweisen. Es sollte schließlich überprüft werden, ob falsche Terminologie die Fehlerkategorie ist, in der die meisten Fehler zu finden sind.

Es werden vier Texte aus dem Rechtsbereich aus dem Russischen ins Deutsche übersetzt. Für die Erstellung maschineller Übersetzungen wird der neuronale maschinelle Online-Übersetzter DeepL verwendet. Die maschinellen Übersetzungen werden anhand der Fehlertypologie der SAE-J2450 manuell evaluiert. Drei Annotator*innen überprüfen die Übersetzungsqualität der DeepL-Übersetzungen, indem sie von ihnen identifizierte sprachliche Fehler analysieren und kategorisieren. Jede*r Annotator*in bekommt eine Word-Datei mit den hierfür erforderlichen Informationen. Diese gliedern sich in einen theoretischen und einen praktischen Teil. Der theoretische Teil erläutert die Regeln der Annotation und der Beschreibung der Fehlerkategorien. Der praktische Teil widmet sich unmittelbar der Annotation der übersetzten Texte bzw. der Fehlerklassifizierung. Insgesamt wird den Annotator*innen in der Datei Folgendes zur Verfügung gestellt:

- vier Textausschnitte in Originalsprache (Russisch) aus einem Kaufvertrag, Bestimmungen zur Eheschließung in Russland, zur Verfassung der Russischen Föderation und zu Änderungen der Datenschutzgrundverordnung in Russland (siehe Anhang A1);
- vier zu evaluierende maschinelle Übersetzungen der oben erwähnten Texte (siehe Anhang A1);

- die Vorgehensweise der Evaluierung bzw. der Textannotation und Analyse samt einer genauen Beschreibung der Fehlerkategorien (siehe Anhang A2, A3);
- Textaufbereitung in Form einer Tabelle für den besseren Vergleich zwischen dem Ausgangs- und Zieltext und
- vier Annotationsblätter zur Identifizierung der Fehler (siehe Anhang A4).

Weitere Details zur Untersuchung werden in den nächsten Unterkapiteln präsentiert.

5.2 Text- und Systemauswahl

Die grundlegenden Fragen bei der Textauswahl bezogen sich auf die Textsorte und die Textmenge bzw. Textlänge. Da MÜ-Systeme eine Tendenz aufweisen, bei unterschiedlichen Textsorten ungleiche Qualitätsergebnisse zu liefern (Marx et al. 1998: 1222), wurde beschlossen, möglichst unterschiedliche Texte auszuwählen. Die Texte sollten zu differierenden Textsorten oder Stilen gehören. Für die Untersuchung wurden vier Texte bzw. Textabschnitte mit jeweils ca. 100 Wörtern pro Textabschnitt ausgewählt (siehe Anhang A1). Danach wurden sie maschinell in Deutsche übersetzt und analysiert.

Text 1 ist ein Ausschnitt aus einem KFZ-Kaufvertrag. Der Ausgangstext besteht aus 104 Wörtern und wurde von einer russischen Automobil-Website (Kurumadom.ru o.J.) zur Kennzeichenprüfung heruntergeladen, die über eine große Anzahl an Mustern für Formulare und Dokumente verfügt. Laut der in Unterkapitel 3.2 vorgestellten Texttypologien der Rechtstexte nach Šarčevic (1997) und Daum (2003) wird Text 1 als primär präskriptiver individueller Normtext klassifiziert (Šarčevic 1997: 11; Daum 2003: 36).

Text 2 ist ein Abschnitt aus den Bestimmungen zur Eheschließung in Russland (Semenenko 2020). Der Ausgangstext beinhaltet 96 Wörter. Der Text wurde in Form einer schriftlichen rechtlichen Beratung zum Thema Eheschließung auf einer Webseite eines russischen Juristen gefunden (Semenenko 2020). Nach der Texttypologie von Šarčevic (1997) und Daum (2003) handelt es sich bei Text 2 um einen reinen deskriptiven informativen Text (Šarčevic 1997: 11; Daum 2003: 37).

Text 3 ist ein Abschnitt aus der Verfassung der Russischen Föderation, der sich mit den Grundzügen der verfassungsmäßigen Ordnung befasst (Constitution.ru 2001). Der Ausgangstext enthält 92 Wörter und ist online zu finden. Der Texttypologie von Šarčević (1997) und Daum (2003) zufolge gehört Text 3 zu primär präskriptiven allgemeinen Normtexten (Šarčevic 1997: 11; Daum 2003: 36).

Text 4 ist ein Presseartikel (Interfax.ru 2015), in dem eine neue Datenschutzgrundverordnung erläutert wird. Der Ausgangstext besteht aus 106 Wörtern und wird nach Šarčević (1997) und Daum (2003) als ein rein deskriptiver informativer Text klassifiziert (Šarčevic 1997: 11; Daum 2003: 37).

Was die Systemauswahl betrifft, so gehörten die Verwendung von neuronalen Netzen, freie Verfügbarkeit des MÜ-Systems im Internet und die Unterstützung der russischen und deutschen Sprache zu den wichtigsten Kriterien der Systemauswahl. Da die Evaluierungsergebnisse der Blindtests (DeepL 2020b) zeigten, dass das NMÜ-System DeepL in der Lage war, effektiver und genauer als andere MÜ-Übersetzungssysteme zu übersetzen, gewann der DeepL-Übersetzer das Interesse der vorliegenden Untersuchung.

5.3 Sprachauswahl

In dieser Arbeit werden ausschließlich maschinelle Übersetzungen aus dem Russischen ins Deutsche untersucht. Diese Sprachrichtung wurde aus verschiedenen Gründen gewählt. Erstens machte die Zugehörigkeit der ausgewählten Sprachen zu unterschiedlichen Sprachgruppen (slawische und germanische) diese Sprachkombination interessant. Die erheblichen Unterschiede zwischen den beiden Sprachsystemen werden zwangsläufig die Übersetzungsqualität des NMÜ-Systems beeinflussen, was für die Untersuchungsanalyse relevant ist. Zweitens scheint es sinnvoll zu sein, Deutsch als Zielsprache für einen deutschsprachigen Raum festzulegen, da die vorliegende Arbeit dadurch über die Risiken der unkontrollierten Verwendung von MÜ und Übersetzungsfehler in dieser Sprachrichtung informiert und für österreichische Behörden, die mit der deutschen Sprache nicht mächtigen russischsprachigen Ausländer*innen kommunizieren, relevant sein kann. Angesichts der aktuellen weltpolitischen Lage steigt der Bedarf an Übersetzungen aus dem Ukrainischen und Russischen ins Deutsche, was Menschen ohne Deutschkenntnisse dazu zwingt, mehr zu MÜ zu greifen. Daher ist das ausgewählte Sprachpaar nach Meinung der Verfasserin der vorliegenden Masterarbeit berechtigt.

5.4 Fehlertypologie

Für die Evaluierung der Übersetzungsqualität maschinell übersetzter Texte wurde die Fehlertypologie der SAE-J2450-Metrik ausgewählt. Die Metrik sollte die Übersetzungsqualität evaluieren und messen können, unabhängig von der Ausgangssprache, der Zielsprache und der Art der Übersetzung (menschliche oder maschinelle Übersetzung) (Dalla-Zuanna 2010). Diese

Qualitätsmetrik eignet sich ebenso gut für die manuelle Evaluierung der maschinellen Übersetzung von Rechtstexten, da sie sich ausschließlich mit linguistischen Aspekten befasst und für die Suche nach Terminologiefehlern nützlich ist. Zudem besitzt die Rechtssprache ähnliche Eigenschaften wie die technische Fachsprache, da sie beide präzise und unflexibel sind. Anhand dieser Metrik werden Fehler in sieben Fehlerkategorien eingeordnet: falsche Benennung, falsche Bedeutung, Auslassung, Strukturfehler, Rechtschreibfehler, Interpunktionsfehler und sonstiger Fehler (SAE 2001; Dalla-Zuanna 2010). Die genaue Beschreibung der Fehlerkategorien wurde in Unterkapitel 2.3.1.2 präsentiert.

Gegenüber dem Standard der Metrik wurden minimale Abänderungen vorgenommen. Laut SAE-J2450-Metaregeln soll ein Fehler als schwerer Fehler eingestuft werden, auch wenn sich der / die Annotator*in bezüglich der starken negativen Auswirkungen des Fehlers nicht sicher ist. Bei der vorliegenden Evaluierung wird der Schweregrad eines Fehlers im gleichen Fall jedoch als leichter Fehler bewertet. Des Weiteren wird Semantik nicht in der Kategorie Strukturfehler inkludiert, um die Überschneidung mit falscher Bedeutung und falscher Benennung zu vermeiden.

5.5 Profil der Annotator*innen

Eine bedeutende Rolle bei der Auswahl der Annotator*innen spielen die Sprachkenntnisse. Wichtig ist, dass sowohl russische als auch deutsche Erstsprachler*innen fürs Annotieren ausgesucht werden. Um eine Tendenz zur Übereinstimmung der Annotation zu berechnen, haben insgesamt drei Annotator*innen vier Texte annotiert. Die Annotator*innen wurden nach folgenden Kriterien ausgewählt:

- Abschluss des Masterstudiums Translation mit Deutsch als Erstsprache und fließendem Russisch;
- Abschluss des Masterstudiums Translation mit Russisch als Erstsprache und fließendem Deutsch.

Auf Basis dieser Kriterien wurden drei Annotator*innen ausgewählt, wodurch insgesamt drei Annotationen für jeden Text erstellt wurden. Dadurch sollte es ermöglicht werden, dass die maschinellen Übersetzungen sowohl aus der Sicht der Erstsprachler*innen als auch der Nicht-Erstsprachler*innen betrachtet werden.

5.6 Auswertung der Ergebnisse

Nach der Textüberprüfung und Übermittlung der Ergebnisse durch die Annotator*innen wird analysiert, wie die untersuchten Zieltexte laut SAE-J2450 eingeschätzt wurden. Dafür wird die Gesamtanzahl der annotierten Fehler in jeder Kategorie mit dem numerischen Schweregrad des Fehlers multipliziert. Die numerischen Ergebnisse aller Kategorien werden summiert und die Summe der Gesamtpunkte wird durch die Anzahl der Wörter im Ausgangstext dividiert (siehe Tab. 1).

Tabelle 1: Berechnung des Evaluierungswertes nach SAE-J2450, adaptiert nach SAE (2001: 10)

Fehlerkategorie	**Anzahl der leichten Fehler**		**Anzahl der schweren Fehler**		**Summe der Fehlerpunkte pro Fehlerkategorie**
Falsche Benennung (FBN)	… *2	+	… *5	=	
Falsche Bedeutung (FBD)	… *2	+	… *5	=	
Auslassung (A)	… *2	+	… *4	=	
Strukturfehler (SRF)	… *2	+	… *4	=	
Rechtschreibfehler (RF)	… *1	+	… *3	=	
Interpunktionsfehler (IF)	… *1	+	… *2	=	
Sonstige Fehler (SF)	… *1	+	… *3	=	
		Gesamtsumme aller Fehlerpunkte:			

Anzahl der Wörter im Ausgangstext: _____	**Evaluierungswert** = Die Gesamtsumme aller Fehlerpunkte geteilt durch die Anzahl der Wörter im Ausgangstext

Auf diese Weise wird der Qualitätswert nach SAE-J2450 berechnet. Die SAE-J2450 legt keinen Grenzwert für die Akzeptanz der jeweiligen Übersetzung fest. Somit bleibt der Grenzwert jeder Firma oder jedem/jeder Evaluierer*in selbst überlassen. Die vorliegende Untersuchung richtet sich nach dem Grenzwert des Volkswagen-Konzerns, der 0,03 beträgt (Dalla-Zuanna 2010: 24). Zum Schluss wird ein Kappa-Wert berechnet. Neben der numerischen Auswertung wird es ebenfalls noch eine Detailauswertung geben, bei der die Fehler näher analysiert werden.

5.7 Zuverlässigkeit der Annotator*innen

Durch die Berechnung des Kappa-Wertes kann man feststellen, wie hoch die Übereinstimmung bzw. Reliabilität zwischen den Bewerter*innen ist. Da die Annotation der Texte in diesem Fall durch mehr als zwei Annotator*innen durchgeführt wird, muss Fleiss' Kappa berechnet werden (Fleiss 1971). Genauer gesagt eignet sich Fleiss' Kappa für eine beliebige Anzahl von Bewerter*innen, die aus mehr als zwei Personen besteht, und ermöglicht ebenfalls die Bewertung verschiedener Einheiten von unterschiedlichen Beurteiler*innen trotz einer fixierten

Anzahl von Bewerter*innen (Fleiss 1971: 378). Im Grunde genommen ist Fleiss' Kappa eine Erweiterung der Cohens-Kappa-Formel (Cohen 1960: 40).

Cohens Kappa wird hingegen nur für zwei Beurteiler*innen verwendet und sieht folgendermaßen aus: $k = \frac{p_0 - p_c}{1 - p_c}$. Dabei ist p_0 der Anteil der Einheiten, bei denen sich die Rater*innen bzw. Annotator*innen einig waren, und p_c der Anteil der Einheiten, bei denen eine Übereinstimmung durch Zufall erwartet wird (Cohen 1960: 39). Der Koeffizient *k* ist entweder der Anteil der zufällig erwarteten Unstimmigkeiten, die nicht vorkommen, oder der Anteil der Übereinstimmung, nachdem die zufällige Übereinstimmung aus der Betrachtung ausgeschlossen wurde (Cohen 1960: 40).

Für die Berechnung von Fleiss' Kappa wird die Formel für drei Annotator*innen angepasst:

$$p0 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{n(n-1)}\sum_{j=1}^{k}(n_{ij}^2 - n_{ij})\right)$$

$$pc = \sum_{j=1}^{k} p_j^2$$

Nach Fleiss (1971: 378) steht *N* für die Gesamtzahl der Gegenstände; *n* für die Anzahl der Bewertungen pro Gegenstand; *k* steht für die Anzahl der Kategorien, in denen die Bewertungen durchgeführt werden. Die Anzahl *i = 1, ..., N* steht für die Gegenstände und *j = 1, ..., k* für die Kategorien der Skala; n_{ij} steht für die Anzahl der Bewerter*innen, die das Subjekt *i* der Kategorie *j* zuordneten (Fleiss 1971: 378). Landis und Koch (1977:165) schlagen folgende Interpretation der relativen Stärke der Übereinstimmung vor:

- < 0,00 = niedrige Übereinstimmung (poor agreement);
- 0,00–0,20 = geringe Übereinstimmung (slight agreement);
- 0,21–0,40 = ausreichende Übereinstimmung (fair agreement);
- 0,41–0,60 = mittelmäßige Übereinstimmung (moderate agreement);
- 0,61–0,80 = erhebliche Übereinstimmung (substantial agreement) und
- 0,81–1,00 = beinahe volle Übereinstimmung (almost perfect agreement).

6 Ergebnisse

In diesem Kapitel werden die Ergebnisse der Untersuchung präsentiert. Zunächst wird ein Überblick über die Ergebnisse der maschinell übersetzten Texte nach der SAE-J2450-Metrik gegeben. Anschließend werden die in den maschinellen Übersetzungen aufgetretenen Fehler detaillierter analysiert. Dabei stehen die schweren Fehler im Fokus. Abschließend wird auf den Kappa-Wert der annotierten Übersetzungen eingegangen.

6.1 Evaluierungswerte der maschinell übersetzten Texte nach SAE-J2450

Die Übersetzungsqualität der ausgewählten Texte (siehe Anhang A1) wurde mittels der SAE-J2450-Metrik manuell gemessen. Eine Absolventin des translatorischen Masterstudiums an der Universität Wien mit Deutsch als Erstsprache und fließendem Russisch sowie eine weitere Absolventin des gleichen Studienganges mit Russisch als Erstsprache und fließendem Deutsch annotierten die maschinell übersetzten Texte anhand von festgelegten Fehlerkategorien der oben erwähnten Metrik (siehe Anhang A3) (inklusive der Annotation der Verfasserin dieser Masterarbeit). Dies ermöglichte eine Berechnung des SAE-J2450-Evaluierungswertes. Tabelle 2 zeigt die Anzahl der festgestellten Fehler pro Fehlerkategorie für den jeweiligen Text sowie die Summe der Fehlerpunkte. Darüber hinaus sind die Informationen zum Mittelwert der evaluierten Texte durch SAE-J2450 ebenso in Tabelle 2 zu sehen.

Tabelle 2: Evaluierungsblatt der SAE-J2450-Metrik

Annotator*in 1																
Kategorie		FBN		FBD		A		SRF		RF		IF		SF		Summe der Fehlerpunkte	SAE-J2450-Wert
Schweregrad		*2	*5	*2	*5	*2	*4	*2	*4	*1	*3	*1	*2	*1	*3		
Fehleranzahl	Text 1	3	2	1		2		1								24	**0,23**
	Text 2	2	3	3	2	2								2		41	**0,42**
	Text 3	12						2				2				30	**0,32**
	Text 4	3	4	6				2								42	**0,39**
																Total: 137	

Annotator*in 2

Kategorie		FBN		FBD		A		SRF		RF		IF		SF		Summe der Fehlerpunkte	SAE-J2450-Wert
Schweregrad		*2	*5	*2	*5	*2	*4	*2	*4	*1	*3	*1	*2	*1	*3		
Fehleranzahl	Text 1	1	1			1		1						1		12	**0,11**
	Text 2	3			1	1	1	2		1		2				24	**0,25**
	Text 3	1	6		1											37	**0,4**
	Text 4	1			2	3		1						1		21	**0,19**
																Total: 94	

Annotator*in 3

Kategorie		FBN		FBD		A		SRF		RF		IF		SF		Summe der Fehlerpunkte	SAE-J2450-Wert
Schweregrad		*2	*5	*2	*5	*2	*4	*2	*4	*1	*3	*1	*2	*1	*3		
Fehleranzahl	Text 1	2	2			3		1								22	**0,21**
	Text 2	8	2	1	1			1		1				1		37	**0,38**
	Text 3	9			1			4	1	1		2				38	**0,41**
	Text 4	3	1	2	3	2		1				1		3		40	**0,37**
																Total: 137	

Mittelwert der Evaluierung

Text 1 = **0,18**
Text 2 = **0,35**
Text 3 = **0,37**
Text 4 = **0,31**

Abkürzungen:

FBN = Falsche Benennung
FBD = Falsche Bedeutung
A = Auslassung
SRF = Strukturfehler
RF = Rechtschreibfehler
IF = Interpunktionsfehler
SF = Sonstige Fehler

Aus Tabelle 2 wird ersichtlich, dass Text 1 (KFZ-Kaufvertrag) die besten Ergebnisse der Gesamtevaluierung mit einem Mittelwert von 0,18 erzielte. Text 3 (Verfassung der Russischen Föderation) hingegen zeigte die schlechtesten Resultate mit einem Mittelwert von 0,37. Text 4 mit den Bestimmungen zur Eheschließung in Russland belegte den zweiten Platz mit 0,31. Danach folgte Text 2 (Datenschutzgrundverordnung) mit einem Mittelwert von 0,35. Es soll noch einmal erwähnt werden, dass sich die vorliegende Untersuchung an dem Grenzwert des Volkswagen-Konzerns orientiert, der 0,03 beträgt (Dalla-Zuanna 2010: 24). Somit sind nicht mehr als 3 Fehlerpunkte pro 100 Wörter des übersetzten Textes erlaubt. Die untersuchten Texte

waren nicht deutlich länger als 100 Wörter: 104 Wörter (Text 1), 96 Wörter (Text 2), 92 Wörter (Text 3) und 106 Wörter (Text 4). Allerdings lassen die Ergebnisse in Tabelle 2 darauf schließen, dass die Fehlerpunkte deutlich über diesem Grenzwert liegen. Durch die Annotation der Texte konnte die Fehleranzahl und die Fehlerkategorie mit der größten Fehleranzahl bestimmt werden (siehe Abb.1).

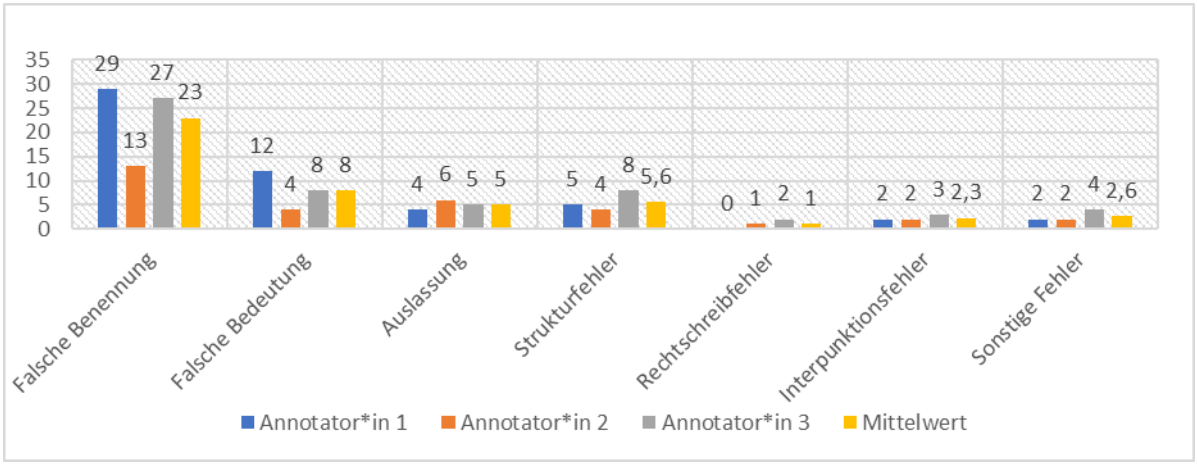


Abbildung 1: Summe der Fehler pro Fehlerkategorie

Insgesamt wurden im Durchschnitt 47 Fehler festgestellt. Es zeigte sich, dass die Mehrheit der Fehler zur Kategorie „falsche Benennung“ gehörte. Die geringste Fehleranzahl war in der Kategorie „Rechtschreibfehler“ zu finden. Allerdings gab es keine Fehlerkategorie, die keinen einzigen Fehler beinhaltete.

Wenn man dazu den Prozentsatz berechnet (siehe Abb. 2), so stellt man fest, dass fast die Hälfte aller Fehler zur falschen Benennung gehörte. Die Kategorie „falsche Bedeutung“

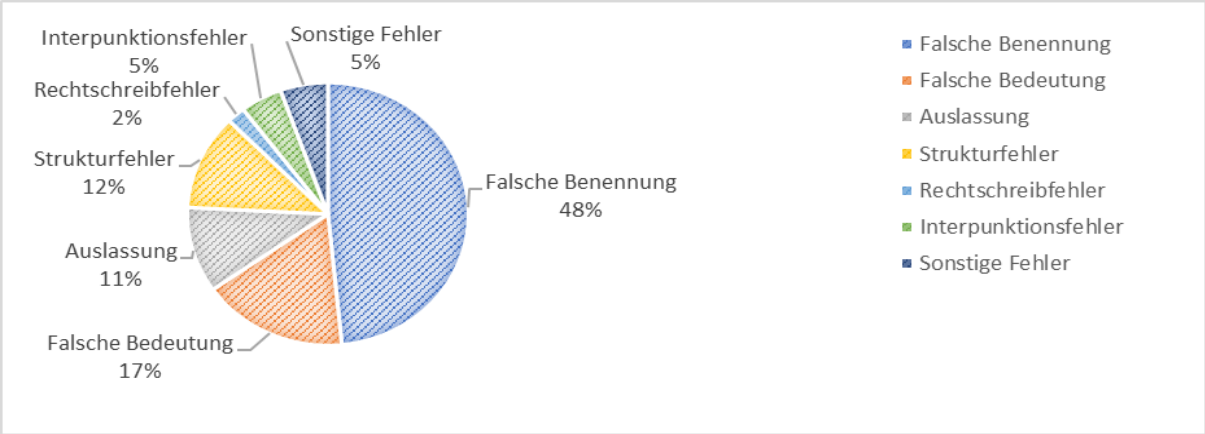


Abbildung 2: Prozentsatz der aufgetretenen Fehler nach Fehlerkategorie

machte annähernd ein Fünftel aller Fehler aus. Die Kategorien „Auslassung“ und „Strukturfehler“ hatten fast den gleichen prozentualen Anteil, nämlich ein Zehntel aller Fehler.

Interpunktionsfehler und sonstige Fehler repräsentierten nur 5 % der Fehler. Die seltensten Fehler waren Rechtschreibfehler; sie machten nur 2 % aller Fehler aus.

Im Durchschnitt wurden 80 der 123 untersuchten Satzsegmente mittels DeepL fehlerfrei übersetzt. Dies entspricht 65 % aller für die Untersuchung ausgewählten Segmente. Was ganze Sätze betrifft, so wurden beispielsweise von 24 überprüften Sätzen nur 4,6 Sätze fehlerfrei übersetzt.

6.2 Fehleranalyse der maschinell übersetzten Texte

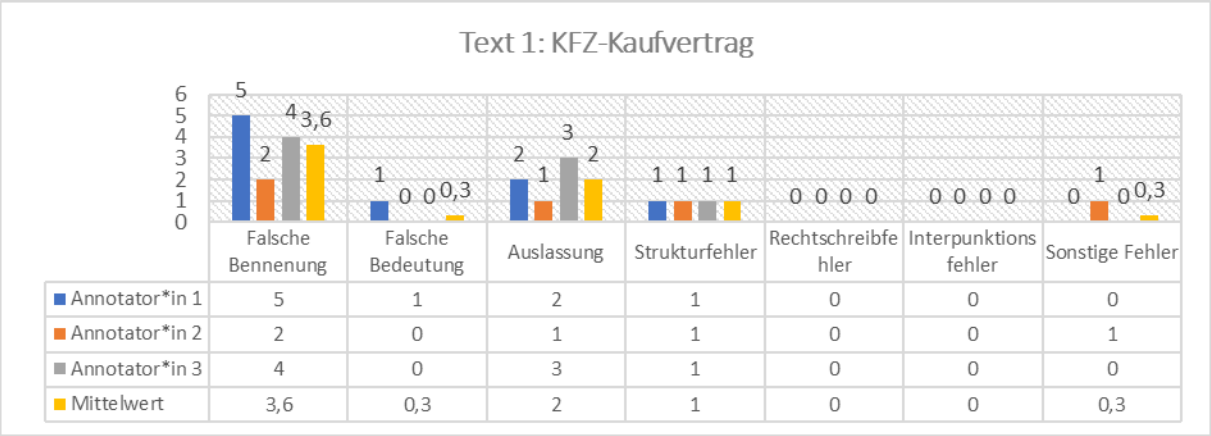
Im Allgemeinen wurden folgende Fehler festgestellt (von denen mehrere in dieselbe Fehlerkategorie fallen): falsche Terminologie, Auslassungen, falsche Kollokationen, falsche Bedeutung, Rechtschreibfehler, Numerusfehler, nicht gegenderte Substantive, Interpunktionsfehler, falsche Artikel und im Ausgangstext nicht vorhandene Ergänzungen.

Als sonstige Fehler wurden folgende Fehler von den Annotator*innen eingestuft: doppelte Benennung, doppelte Übersetzung, fehlendes Gendern, fehlende Anpassung der Benennung an die Zielkultur und im Ausgangstext nicht vorhandene Ergänzungen.

Die Fehleranalyse ergab, dass bei allen maschinell übersetzten Texten nicht nur leichte, sondern auch schwere Fehler festgestellt wurden. Fehler mit solchem Schweregrad änderten den Sinn des Zielsatzes, sodass sie zu Missverständnissen führten und daher gravierende Folgen für das Zielpublikum haben konnten. Manche Formulierungen konnten vom Zielpublikum für unangemessen oder sogar beleidigend gehalten werden. Im folgenden Unterkapitel wird jeder einzelne Text (siehe Anhang 1) detaillierter betrachtet und aufgetretene schwere Fehler werden kommentiert.

6.2.1 Fehleranalyse von Text 1: der KFZ-Kaufvertrag

Text 1 (KFZ-Kaufvertrag) zeichnete sich durch die besten Übersetzungsergebnisse verglichen mit den anderen maschinell übersetzten Texten aus. Aus Abb. 3 ist zu ersehen, dass Fehler der Kategorien „falsche Benennung“ sowie „Auslassung“ die größte Herausforderung für das NMÜ-System von DeepL in Text 1 waren. Auffällig ist, dass alle Annotator*innen zu einer einstimmigen Beurteilung kamen, was die Fehler in den Kategorien Strukturfehler, Rechtschreibfehler und Interpunktionsfehler angeht. Wie Abb. 3 zeigt, wurde kein einziger Rechtschreib- oder Interpunktionsfehler im Zieltext gefunden. Im Durchschnitt wurden insgesamt 7,3 Fehler in der vorliegenden maschinellen Übersetzung gefunden.



	Falsche Bennenung	Falsche Bedeutung	Auslassung	Strukturfehler	Rechtschreibfehler	Interpunktionsfehler	Sonstige Fehler
Annotator*in 1	5	1	2	1	0	0	0
Annotator*in 2	2	0	1	1	0	0	1
Annotator*in 3	4	0	3	1	0	0	0
Mittelwert	3,6	0,3	2	1	0	0	0,3

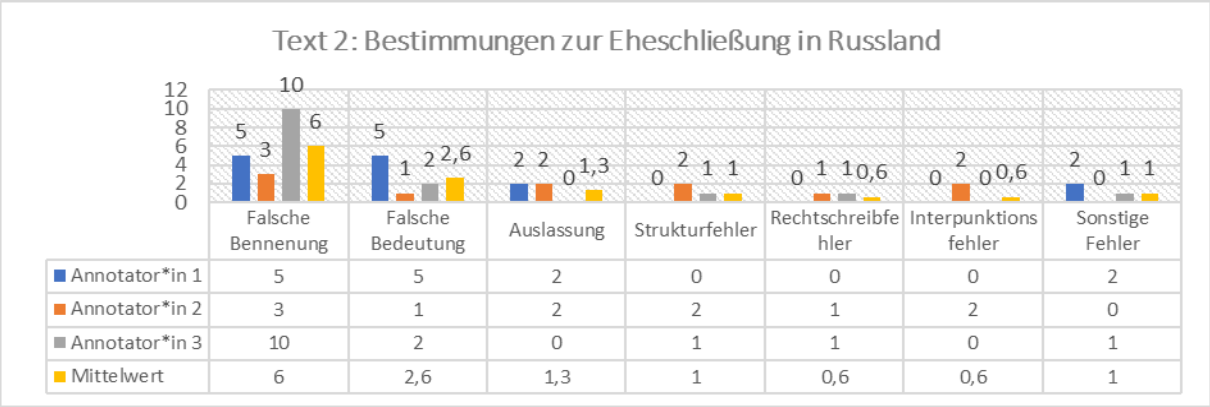
Abbildung 3: Anzahl der Fehler in Text 1

Zu den schweren Fehlern des Textes gehörten etwa die folgenden:

1) Die Wortkombination „Предмет договора“ wurde ins Deutsche als „Gegenstand des Abkommens“ übersetzt, was als eine falsche Terminologie betrachtet werden kann. Es handelt sich hier nicht um ein Abkommen, sondern um einen Vertrag.
2) Der Satz „Транспортное средство не находится под арестом“ wurde als „das Fahrzeug ist nicht verhaftet“ im Deutschen wiedergegeben. Jedoch kann man ein Fahrzeug nicht verhaften. Die richtige Übersetzung wäre: „Das Fahrzeug wird nicht beschlagnahmt“.
3) Im Satz „Gelingt es den Parteien nicht, eine außergerichtliche Einigung zu erzielen, so wird die Angelegenheit von einem Gericht nach dem geltenden Recht der Russischen Föderation überprüft“ hat sich das DeepL-System für das Wort „Angelegenheit“ entschieden, wobei es im Zieltext um einen Rechtsfall bzw. Rechtsstreit geht.

6.2.2 Fehleranalyse von Text 2: Bestimmungen zur Eheschließung in Russland

Text 2 beinhaltete fast doppelt so viele Fehler in der Kategorie „falsche Benennung“ wie der vorherige Text 1. Die zweithöchste Anzahl von Fehlern gehörte zur Kategorie „falsche Bedeutung“; auch in dieser Kategorie wurde ein Anstieg der Fehlerzahl beobachtet (siehe Abb. 4). Zu den anderen Fehlerkategorien ist der Abb. 4 zu entnehmen, dass die Verteilung der Fehler in den restlichen Kategorien ziemlich einheitlich war, nämlich zwischen 0,6 und 1,3 Fehlern pro Text. Die durchschnittliche Anzahl aller Fehler im Zieltext lag bei 13,3 Fehlern.



	Falsche Bennenung	Falsche Bedeutung	Auslassung	Strukturfehler	Rechtschreibfehler	Interpunktionsfehler	Sonstige Fehler
Annotator*in 1	5	5	2	0	0	0	2
Annotator*in 2	3	1	2	2	1	2	0
Annotator*in 3	10	2	0	1	1	0	1
Mittelwert	6	2,6	1,3	1	0,6	0,6	1

Abbildung 4: Anzahl der Fehler in Text 2

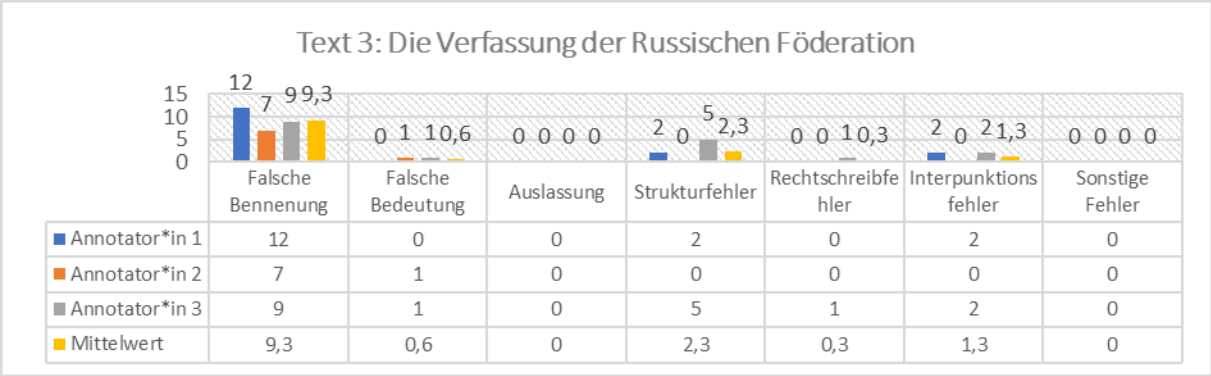
Nun sollen die groben Fehler der maschinellen Übersetzung des vorliegenden Textes genauer besprochen und durch Beispiele illustriert werden:

1) Die Wortverbindung „Родственники по прямой и нисходящей линии“ wurde ins Deutsche als „Nachkommen oder Nachkommen von Geschwistern“ übersetzt. Das ist eine völlig falsche Terminologie, die zu einer falschen Bedeutung des Satzes führt, mithin ein sehr schwerer Fehler. Die korrekte Übersetzung wäre „Verwandte gerader Linie sowie der Seitenlinie“.
2) Der Begriff „Слабоумие“ wurde als „Demenz“ übersetzt. Es muss erklärt werden, dass der Begriff des Ausgangssatzes alle angeborenen und erworbenen Behinderungen umfasst, welche die geistigen Fähigkeiten beeinträchtigen. Der Begriff des Zielsatzes hingegen bezeichnet nur Demenz, d. h. eine spezifische erworbene geistige Behinderung. Ein guter Übersetzungsvorschlag hier wäre „mentale Beeinträchtigung“.
3) Die Wortverbindung „дееспособность граждан“ wurde vom DeepL-System als „die Geschäftsfähigkeit von Bürgern“ übersetzt. Dabei ist die Geschäftsfähigkeit nur ein Unterbegriff. Gemeint wird in diesem Fall die Handlungsfähigkeit. Des Weiteren soll das Wort „Bürger“ gegendert werden, um eine Diskriminierung zu vermeiden.

6.2.3 Fehleranalyse von Text 3: die Verfassung der Russischen Föderation

Wie bei den vorangegangenen Texten enthielt die Kategorie „falsche Benennung“ ebenso in Text 3 die größte Anzahl an Fehlern (siehe Abb. 5). Mit durchschnittlich 9,3 Fehlern lag diese Fehlerkategorie bei allen von DeepL übersetzten Texten an der Spitze. Wie Abb. 5 zeigt, war die Verteilung der Fehler innerhalb der restlichen Kategorien nicht so einheitlich, wie es bei Text 2 der Fall war. In den Kategorien „sonstige Fehler“ sowie „Auslassung“ wurden keine

Fehler festgestellt. In der Kategorie „Strukturfehler“ war ein deutlicher Anstieg an Fehlern zu verzeichnen, verglichen mit den Texten 1 und 2. Im Durchschnitt wurden insgesamt 14 Fehler erfasst.



	Falsche Bennenung	Falsche Bedeutung	Auslassung	Strukturfehler	Rechtschreibfehler	Interpunktionsfehler	Sonstige Fehler
Annotator*in 1	12	0	0	2	0	2	0
Annotator*in 2	7	1	0	0	0	0	0
Annotator*in 3	9	1	0	5	1	2	0
Mittelwert	9,3	0,6	0	2,3	0,3	1,3	0

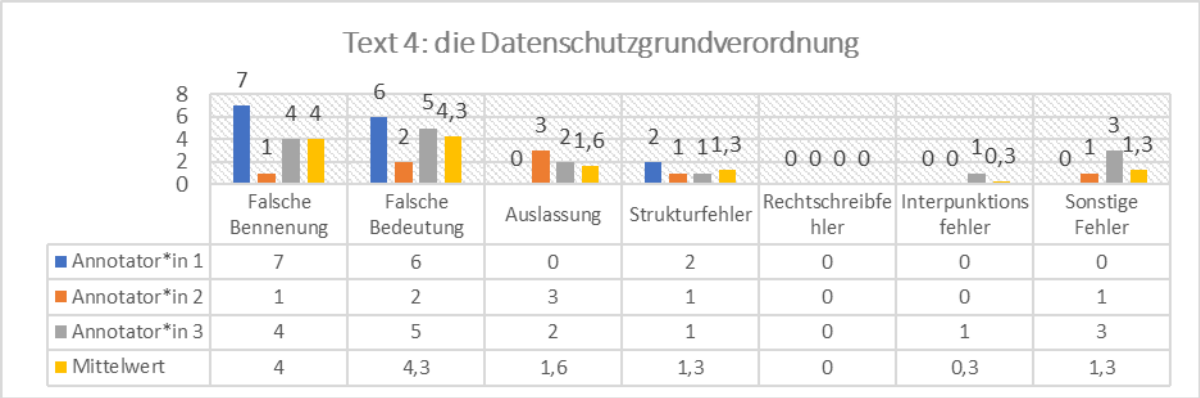
Abbildung 5: Anzahl der Fehler in Text 3

Nun werden die schweren Fehler des vorliegenden übersetzten Textes analysiert:

1) Der Staatsname „Российская Федерация“ wurde ins Deutsche als „die Rußländische Föderation“ übersetzt. Dies ist eine falsche Bezeichnung des Staates, die von den russischen Staatsbürger*innen als beabsichtigt oder beleidigend empfunden werden kann.
2) Bei der Übersetzung des Satzes „Российская Федерация состоит из […] автономной области“ als „Die Rußländische Föderation besteht aus […] autonomen Oblasten“ tritt ein schwerer Strukturfehler auf. Es geht hier ausschließlich um ein einziges autonomes Gebiet (oder eine autonome Oblast). Die Verwendung der Pluralform für das Wort „Oblast“ ist sehr verwirrend, da die Russische Föderation nur über ein autonomes Gebiet verfügt. Des Weiteren wiederholt sich eine falsche Benennung des Staates in diesem Satz erneut.
3) Das Segment „Равноправие и самоопределение народов в Российской Федерации“ wird mit falscher Bedeutung ins Deutsche übertragen: „Die Gleichheit und Selbstbestimmung der Völker der Rußländischen Föderation“. Im Ausgangstext geht es um die Völker, die auf dem Territorium der Russischen Föderation leben. Allerdings gehören diese Volksgruppen Russland nicht. Daher ist die Verwendung eines Genitivs in diesem Fall nicht gerechtfertigt. Eine richtige Übersetzung wäre „Die Gleichheit und Selbstbestimmung der Völker in der Russischen Föderation“.

6.2.4 Fehleranalyse von Text 4: die Datenschutzgrundverordnung

Durchschnittlich wurden 13 Fehler in der maschinellen Übersetzung des Textes 4 festgestellt. Überraschenderweise war es nicht die Kategorie „falsche Benennung“, sondern die Kategorie „falsche Bedeutung“, die bei der Anzahl der Fehler an erster Stelle stand (siehe Abb. 6). Mit



	Falsche Bennenung	Falsche Bedeutung	Auslassung	Strukturfehler	Rechtschreibfehler	Interpunktionsfehler	Sonstige Fehler
Annotator*in 1	7	6	0	2	0	0	0
Annotator*in 2	1	2	3	1	0	0	1
Annotator*in 3	4	5	2	1	0	1	3
Mittelwert	4	4,3	1,6	1,3	0	0,3	1,3

Abbildung 6: Anzahl der Fehler in Text 4

einem kleinen Vorsprung von 0,3 Fehlern lag sie in Führung. Vergleicht man alle vier übersetzten Texte miteinander, so kann man erkennen, dass die Kategorie „falsche Bedeutung“ im Text 4 am stärksten repräsentiert war. Dahingegen wurden in diesem Text kaum Interpunktionsfehler und keine Rechtschreibfehler gefunden. In den übrigen Fehlerkategorien sind die Fehler mit einem Maximalwert von 1,6 Fehlern fast gleich verteilt, wie in Abb. 6 ersichtlich ist.

Folgende schwere Fehler wurden in dem übersetzten Text identifiziert:

1) Die Wortkombination „Персональные данные россиян“ wurde ins Deutsche als „Personenbezogene Daten von Russen“ übersetzt. Allerdings handelt es sich im Text nicht um Russen als Volk, sondern um russische Staatsangehörige. Deshalb wäre als korrekte Übersetzung „personenbezogene Daten russischer Staatsangehöriger“ angemessen. Generell sollte man im rechtssprachlichen Kontext zwischen einer Volkszugehörigkeit und einer Staatsangehörigkeit unterscheiden können, da ein falscher Begriff eine abwertende bzw. herabsetzende Wirkung erzeugen kann.
2) Das Satzsegment „Уточнение персональных данных“ wurde ins Deutsche als „personenbezogene Daten werden korrigiert“ mit einer falschen Bedeutung übertragen. Im Ausgangssatz geht es um eine Präzisierung und evtl. eine Ergänzung personenbezogener Daten. Damit sind Aktualisierung und Änderung der Daten gemeint. Das Korrigieren der Daten hat nicht dieselbe Bedeutung wie „Präzisierung der Daten“.

Dies kann zu einer falschen oder fehlenden Handlung führen, deshalb wurde dieser Fehler als schwerer eingestuft.

3) Das übersetzte Segment „jede Information, die sich auf eine Person bezieht, die durch sie definiert wird“, ergibt keinen Sinn. Im Ausgangssatz sind „alle Informationen, die einer bestimmten Person zugeordnet werden“ gemeint. Im russischen Text steht „любая информация, относящаяся к определяемому ею физлицу“.
4) Bei der Übersetzung des russischen Satzes „В обновленном законе пересмотрено само понятие персональных данных“ ins Deutsche als „in dem aktualisierten Gesetz wurde der Begriff der personenbezogenen Daten selbst überarbeitet“, kommen einige Fehler vor. Der übersetzte Satz gibt den Sinn des Ausgangssatzes falsch wieder. In diesem konkreten Fall wurde der Begriff „personenbezogene Daten“ einfach neu definiert.

6.3 Übereinstimmung der Annotationen

Um festzustellen, wie hoch die Übereinstimmung zwischen den Annotator*innen beim Annotieren der übersetzten Texte war, und um somit den Objektivitätsgrad der Ergebnisse einzuschätzen, wurde der Kappa-Wert berechnet. Dafür wurde eine Excel-Datei erstellt, die acht Tabellenblätter enthielt. Das erste Tabellenblatt beinhaltete alle durchgeführten Annotationen von drei Annotator*innen. Weitere Tabellenblätter wurden für jede einzelne Fehlerkategorie erstellt und zeigten genau, in welchen Segmenten Fehler aufgetreten waren. Dies ermöglichte es, einen guten Überblick über die Übereinstimmung unter den Annotator*innen in jeder Fehlerkategorie zu erhalten. Alle Segmente der vier Texte wurden zusammengefügt und entsprechend nummeriert, wodurch eine Liste von insgesamt 123 Segmenten generiert wurde. Mit 0 wurden diejenigen Segmente markiert, in denen laut Annotator*innen kein Fehler in einer angegebenen Fehlerkategorie vorlag. Mit 1 wurden Segmente markiert, in denen ein*e Annotator*in einen Fehler festgestellt und einer Fehlerkategorie zugeordnet hat. Wichtig ist, hier zu betonen, dass die Markierung mit 1 nicht auf die Anzahl der Fehler hindeutete, sondern lediglich die Tatsache darstellte, dass Fehler in diesem Segment aufgetreten waren. Dabei spielte der Schweregrad der Fehler keine Rolle und wurde aus diesem Grund nicht berücksichtigt. Nach der Einteilung der Segmente in 0 und 1 konnte für jede Fehlerkategorie ein Kappa-Wert berechnet werden. Schließlich wurde ein Gesamtwert berechnet, da dieser für die Aussagekraft der Übereinstimmung entscheidend war. Tabelle 3 zeigt die Ergebnisse der Kappa-Berechnung.

Tabelle 3: Kappa-Wert der annotierten Übersetzungen pro Fehlerkategorie

Fehlerkategorie	**Kappa-Wert**	**Interpretation**
Falsche Benennung	0,465	mittelmäßige Übereinstimmung (engl. „moderate agreement“)
Falsche Bedeutung	0,242	ausreichende Übereinstimmung (engl. „fair agreement“)
Auslassung	0,406	ausreichende Übereinstimmung (engl. „fair agreement“)
Strukturfehler	0,02	geringe Übereinstimmung (engl. „slight agreement“)
Rechtschreibfehler	0,328	ausreichende Übereinstimmung (engl. „fair agreement“)
Interpunktionsfehler	0,272	ausreichende Übereinstimmung (engl. „fair agreement“)
Sonstige Fehler	0,0889	geringe Übereinstimmung (engl. „slight agreement“)
Gesamtwert	0,361	ausreichende Übereinstimmung (engl. „fair agreement“)

Wie Tabelle 3 zeigt, ergab der Wert der Gesamtberechnung eine ausreichende Übereinstimmung. Allerdings wiesen Strukturfehler und sonstige Fehler eine geringe Übereinstimmung auf. Hingegen erzielte die Fehlerkategorie „falsche Bedeutung“ den besten Kappa-Wert der gesamten Berechnung.

7 Diskussion

Um die Qualität der mittels des neuronalen maschinellen Übersetzungssystems erstellten Übersetzungen zu bewerten, wurden die übersetzten Texte anhand der SAE-J2450-Qualitätsmetrik evaluiert. Im Fokus der Untersuchung stand nicht nur das numerische Ergebnis, sondern auch die Fehleranalyse.

Die Untersuchungsergebnisse zeigten, dass keiner der maschinell übersetzten Texte die SAE-J2450-Qualitätsprüfung bestanden hat, da die Gesamtergebnisse ein schlechtes Ergebnis aufwiesen. Die maschinelle Übersetzung des KFZ-Kaufvertrages (Text 1) war dabei die beste der vier Übersetzungen.

Wenn die maschinell erstellten Übersetzungen aus der Perspektive der in Kapitel 3.2 vorgestellten Texttypologien der Rechtstexte nach Šarčevic (1997) und Daum (2003) betrachtet werden, so lässt sich bemerken, dass die Übersetzungsqualität des NMÜ-Systems nicht von der Funktion des jeweiligen Textes, sondern von dem ausgewählten Thema bzw. dem Inhalt abhängt, was sich auf den Schwierigkeitsgrad der Terminologie auswirkt (Šarčevic 1997: 11; Daum 2003: 35). Diesbezügliche Ergebnisse waren überraschend, da die beste und die schlechteste Übersetzung zu primär präskriptiven Normtexten gehörten (Šarčevic 1997: 11; Daum 2003: 35). Der kleine Unterschied lag lediglich in der Generalisierung und Individualisierung dieser Normtexte, da der Kaufvertrag (Text 1) die Eigenschaften eines individuellen Normtextes und die Verfassung der Russischen Föderation (Text 3) hingegen die Eigenschaften eines allgemeinen Normtextes aufwies (Daum 2003: 36). Es lässt sich daher vermuten, dass der Unterschied in den Unterkategorien der Normtexte die Ergebnisse des NMÜ-Systems DeepL stark beeinflusst. Die zweitbeste und die zweitschlechteste Übersetzung gehörten ebenso zu der gleichen Kategorie der Rechtstexte, genauer gesagt, zu den rein deskriptiven informativen Texten (Šarčevic 1997: 11; Daum 2003: 35). Daraus ergibt sich, dass diese Art der Rechtstexte einen mittleren Schwierigkeitsgrad für das oben erwähnte NMÜ-System darstellen.

Zu Beginn der Untersuchung wurde davon ausgegangen, dass falsche Terminologie bzw. die Kategorie „falsche Benennung“ die höchste Fehlerzahl aufweisen würde. Diese Annahme ließ sich nach der Auswertung der Annotationen bestätigen. Laut den Annotationsergebnissen gehörten 48 % aller festgestellten Fehler zur Kategorie „falsche Benennung“. Diese Fehlerkategorie belegt nach Anteil der Fehler bei allen Texten mit einer Ausnahme von Text 4 den ersten Platz. Der Kappa-Wert der Fehlerkategorie „falsche Benennung“ zeigte eine gute Übereinstimmung, was bestätigt, dass die Annotator*innen in

Bezug auf die markierten falschen Benennungen gleicher Meinung waren. Die hohe Fehleranzahl in dieser Kategorie könnte sich durch Schwierigkeiten bei der Erstellung eines geeigneten Korpus von Paralleltexten für das Training des NMÜ-Systems im Sprachpaar Russisch-Deutsch erklären lassen. Die Tatsache, dass das DeepL-System Paralleltexte und Übersetzungsbeispiele von der Linguee-Webseite benutzt (siehe Kapitel 1.3.3), die allerdings keine Paralleltexte für die Sprachkombination Russisch-Deutsch anbietet, lässt vermuten, dass das DeepL-NMÜ-System einen Mangel an sprachlichen Ressourcen bzw. an geeigneten Paralleltexten für den Rechtsbereich für diese Sprachkombination aufweist.

Bei der Analyse der Ergebnisse wurde festgestellt, dass es bei der Klassifizierung der Fehler signifikante Unterschiede zwischen den Annotator*innen gab. Fehlerkategorien wie Strukturfehler und sonstige Fehler wiesen niedrige Kappa-Werte auf, was auf eine geringe Übereinstimmung zwischen den Annotator*innen hindeutet. Da die Kategorie „sonstige Fehler“ Fehler enthielt, die keiner anderen Fehlerkategorie entsprachen, war die Übereinstimmung in dieser Kategorie erwartungsgemäß gering. Überraschend war die geringe Übereinstimmung bei Strukturfehlern, da es sich hierbei um eine sehr gut geregelte Kategorie handelt, die durch grammatische, syntaktische und semantische Regeln bestimmt wird und Fehler aus diesen Regelsystemen umfasst. Betrachtet man die genaue Beschreibung der Strukturfehler-Kategorie (siehe Kapitel 2.3.1.2) und die festgestellten Strukturfehler, so stellt man fest, dass diese Kategorie ebenso Kollokationsfehler umfasst, die von den Annotator*innen häufig mit falscher Benennung verwechselt wurden. Diese Verwechslung hat zu einer falschen Kategorisierung der Fehler und folglich zu einer geringen Übereinstimmung in der Kategorie Strukturfehler geführt. Darüber hinaus war es ebenso schwierig beim Annotieren, zwischen den Kategorien „falsche Benennung“ und „falsche Bedeutung“ zu unterscheiden, da diese Kategorien gegenseitig verbunden waren, d. h., eine falsche Benennung führte häufig zu einer falschen Bedeutung des Satzes. Das Problem der Verwechselung der Evaluierungskategorien ist für die manuelle Evaluierung nicht neu und zählt zu ihren bekannten Nachteilen, da es sich um eine subjektive Entscheidung und Wahrnehmung der Bewerter*innen handelt (Gornostay 2008: 4). Weniger Verwirrung stiften die Kategorien „Rechtschreib-“ sowie „Interpunktionsfehler“. Verglichen mit den anderen Kategorien dürfen sie keinen signifikanten negativen Einfluss auf die maschinellen Übersetzungen haben, da sie durchschnittlich insgesamt nicht mehr als 1 bis 2,3 Fehler beinhalteten.

Im Laufe des maschinellen Übersetzens der Texte wurde beobachtet, dass das DeepL-System bei Eingabe einzelner Sätze bessere Übersetzungsergebnisse als bei Eingabe des gesamten Textes zur Übersetzung liefert. Dies bedeutet, dass dieses NMÜ-System weniger

Übersetzungsfehler produziert, wenn der gesamte Text segmentiert bzw. wenn jeder Satz separat vom System übersetzt wird. Dies gilt vor allem für die Terminologie. Dieser Aspekt bedarf weiterer Untersuchungen.

Es wird dringend davon abgeraten, maschinell erstellte Übersetzungen von Rechtstexten ohne Überprüfung und anschließende Nachbearbeitung einzureichen, da jede in dieser Arbeit untersuchte Übersetzung schwerwiegende Fehler beinhaltete. Zu ähnlichen Schlussfolgerungen kamen Wiesmann (2019) und Welnitzova et al. (2021), die feststellten, dass neuronale maschinelle Übersetzung noch nicht verlässlich und ausreichend vorangekommen war, um maschinell übersetzte Rechtstexte ohne Nachbearbeitungsaufwand zu verwenden (Wiesmann 2019: 149), obwohl die übersetzten Texte im Allgemeinen gut verständlich waren (Welnitzova et al. 2021:228). Alle von Wiesmann (2019) untersuchten maschinellen Übersetzungen sowie die maschinell übersetzten Texte in dieser Arbeit wiesen terminologische Inkonsistenzen und terminologische Fehler auf, was derartige Einstellungen zur Qualität der maschinellen Übersetzung der Rechtstexte zum Teil erklärte (Wiesmann 2019: 140). Azer und Aghayi (2015) kamen ebenso zum Schluss, dass die von ihnen untersuchten MÜ-Systeme noch erhebliche Mängel aufwiesen, da sie nur mittelmäßige oder sogar unterdurchschnittliche Qualität lieferten. (Azer & Aghayi 2015: 236). Vieira et al. (2021) bestätigten die Gefahr unkontrollierter MÜ-Einsätze im Rechtsbereich, die zu Fehlentscheidungen führen können (Vieira et al. 2021: 1522).

Obwohl die Untersuchungsergebnisse der verwandten Arbeiten (siehe Kapitel 4) viele Parallelen zu den Ergebnissen dieser Arbeit zeigen, gibt es einige Unterschiede. Verglichen mit den Ergebnissen von Welnitzova et al. (2021: 228), zeigten die Ergebnisse der vorliegenden Arbeit, dass das DeepL-System sehr selten Probleme mit der Flexion oder Konjunktion hatte. Wiesmann (2019) berichtete über unübersetzte Wörter, falsche Übersetzungen von Abkürzungen, Wörter, die keinen Sinn ergaben, oder englische Wörter, die in einer Übersetzung aus dem Italienischen ins Deutsche keinen Sinn ergaben, sowie Ergänzungen, die im Zieltext nicht vorkamen, und falsche Wortreihenfolge sowie falsche Zeitform (Wiesmann 2019: 137f.). Diese Fehlerarten wurden in dieser Arbeit nicht festgestellt. Dies könnte einerseits auf den Fortschritt des NMÜ-Systems DeepL hinweisen, andererseits könnte die Diskrepanz zwischen den Fehlerergebnissen darauf zurückzuführen sein, dass die in dieser Arbeit untersuchten Texte eine geringere Komplexität aufwiesen als die von Wiesmann (2019). Überraschend optimistisch waren die Schlussfolgerungen von Killman (2014), der die Ansicht vertrat, dass das von ihm untersuchte MÜ-System im Bereich der Rechtsübersetzung konstant gute Ergebnisse erzielen konnte und daher für Rechtsübersetzer*innen bezüglich der Rechtsterminologie hilfreich wäre (Killman 2014: 96).

Da sich die vorliegende Arbeit lediglich auf die Qualität der sprachlichen Komponenten der maschinell übersetzen Texte beschränkt, können keine Angaben zum Stil sowie zur Formatierung, Genauigkeit oder Flüssigkeit der Übersetzungen gemacht werden. Daher ist es zu empfehlen, eine weitere Studie für dasselbe Sprachpaar, jedoch unter Verwendung anderer Qualitätsmetriken (siehe Kapitel 2.3.1), durchzuführen, die eine gründlichere Untersuchung dieser Aspekte ermöglicht.

8 Fazit

In der vorliegenden Masterarbeit wurde die Qualität neuronaler maschineller Übersetzungen von Rechtstexten in der Sprachrichtung Russisch-Deutsch bewertet. Darüber hinaus wurden theoretische Grundlagen maschineller Übersetzung aufgearbeitet, Methoden und Ansätze zur Bewertung der Qualität maschineller Übersetzungen vorgestellt und die Besonderheiten der Übersetzung von Rechtstexten erläutert.

Die Qualitätsleistung des NMÜ-Systems DeepL wurde anhand der Fehlertypologie der SAE-J2450-Metrik bewertet und die aufgetretenen Fehler wurden im Detail untersucht. Die Untersuchung der maschinell erstellten Übersetzungen zeigte, dass das untersuchte NMÜ-System im Bereich der Rechtsübersetzung eine ungenügende Übersetzungsqualität lieferte, was möglicherweise zum Teil auf die Zugehörigkeit der ausgewählten Sprachen zu unterschiedlichen Sprachgruppen (slawisch und germanisch) zurückzuführen ist.

Fast die Hälfte aller festgestellten Fehler (48 %) gehörte zur Kategorie „falsche Benennung“, bestand also in der Verwendung falscher Terminologie. Dies deutet eindeutig darauf hin, dass das Problem der richtigen Terminologie dem NMÜ-System DeepL große Schwierigkeiten bereitet. Die Rechtschreib- und Interpunktionsfehler scheinen jedoch kein ernsthaftes Problem mehr für dieses System zu sein, woraus sich schließen lässt, dass es die Rechtschreib- und Interpunktionsregeln gut trainiert und gelernt hat. Die in den übersetzten Texten festgestellten schweren Fehler zeigten, welche Risiken und Konsequenzen mit der Verwendung der MÜ ohne Nachbearbeitung verbunden sind. Vor allem sollte darauf geachtet werden, dass das MÜ-System DeepL die geschlechtergerechte Sprache nicht beherrscht, was von betroffenen Personen als diskriminierend wahrgenommen werden könnte. Translator*innen, welche die maschinelle Übersetzung einsetzen, wird nach der Analyse der Forschungsergebnisse empfohlen, im Vorfeld eine sorgfältige Terminologiearbeit durchzuführen und sich über die Vor- und Nachteile des MÜ-Einsatzes zu informieren.

Das NMÜ-System DeepL scheint seine „Konzentration“ beim Übersetzen des gesamten Textes zu verlieren und nicht mehr in der Lage zu sein, den Kontext gut zu verstehen. Dieser Aspekt bedarf weiterer Forschung. Es wäre sinnvoll, eine Studie durchzuführen, in der die MÜ-Qualität in Abhängigkeit von der Textmenge untersucht wird (satzweise und gesamter Text), da größere Textmengen offensichtlich zu keinem besseren Verständnis des Kontexts für das NMÜ-System führen. Solche Abhängigkeit der Qualität eines System-Outputs von der Textmenge sollte man als Translator*in bei der Nutzung von DeepL ebenso berücksichtigen.

Die Untersuchungsergebnisse lassen annehmen, dass die meisten Schwierigkeiten des DeepL-Systems mit der mangelnden Zuverlässigkeit des verwendeten Korpus zusammenhängen. Daher wäre es ratsam, dass DeepL-Entwickler*innen qualifizierte Expertin*innen und Translator*innen für die Erstellung eines guten Korpus mit überprüften Übersetzungen oder Paralleltexten involvieren. Mit Hilfe der durchgeführten Fehleranalyse verfügen DeepL-Entwickler*innen nun über Möglichkeiten, die Ursachen der aufgetretenen Fehler zu untersuchen und Korrekturen oder Verbesserungen vorzunehmen. Daher wäre interdisziplinäre Forschung angebracht, die eine Untersuchung der Fehler nicht nur aus der linguistischen Sicht, sondern auch aus der Perspektive der Computerlinguistik ermöglicht. Dadurch kann ebenso die technische Seite bzw. können die Arbeitsschritte des Systems untersucht werden.

Bibliographie

Ahrend, Klaus (2006). Kriterien für die Bewertung von Fachübersetzungen. In: Schippel, Larisa (Hg.) *Übersetzungsqualität: Kritik–Kriterien–Bewertungshandeln*. Berlin: Frank & Timme, 31–42.

ALPAC (1966). Language and Machines: Computers in Translation and Linguistics: a Report (No. 1416). Modern Language Assoc. National Academy of Sciences, National Research Council.

Arnold, Douglas; Balkan, Lorna; Humphreys, R. Lee; Meijer, Siety & Sadler, Louisa (eds.) (1994). *Machine Translation: an Introductory Guide.* Manchester/Oxford: NCC Blackwell.

Atesman (o.J.). Evaluierung von Übersetzungen nach SAE-J2450. http://www.ates-man.info/wp-content/uploads/2015/10/SAE-Qualit%C3%A4tsmetrik.pdf (Stand: 04.10.2022).

Azer, Haniyeh Sadeghi & Aghayi, Mohammad Bagher (2015). An evaluation of output quality of machine translation (Padideh Software vs. Google Translate). *Advances in Language and Literary studies* 6 (4), 226–237.

Azzano, Dino (2009). CAT und MÜ – Getrennte Welten? *The Journal for Language Technology and Computational Linguistics* 24 (3), 19 –36.

Banerjee, Satanjeev & Lavie, Alon (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Association for Computational Linguistics: Proceedings of the Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65–72.

Bar-Hillel, Yehoshua (1959). Report on the state of machine translation in the United States and Great Britain. Jerusalem: Hebrew University.

Bennett, Winfield S. & Slocum, Jonathan (1985). The LRC machine translation system. *Computational linguistics* 11 (2-3), 111–121.

Bentivogli, Luisa; Bisazza, Arianna; Cettolo, Mauro & Federico, Marcello (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, November 1–5, 2016. Austin, Texas, 257–267.

Bentivogli, Luisa; Cettolo, Mauro; Federico, Marcello & Federmann, Christian (2018). Machine translation human evaluation: An investigation of evaluation based on Post-editing and its relation with Direct Assessment. *Proceedings of the 15th International Workshop on Spoken Language Translation*. Bruges, Belgium, 62–69.

Berūkštienė, Donata (2016). Legal discourse reconsidered: Genres of legal texts. *Comparative legilinguistics* vol. 28/2016, 89–117.

Britz, Denny; Goldie, Anna; Luong, Minh-Thang & Le, Quoc (2017). Massive Exploration of Neural Machine Translation Architectures. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, September 7–11, 2017. Copenhagen, Denmark, 1442–1451.

Brown, Peter F.; Cocke, John; Della Pietra, Stephen; Della Pietra, Vincent J.; Jelinek, Frederick; Mercer, Robert L. & Roossin, Paul S. (1988). A Statistical Approach to French/English Translation. *Proceedings of the Second International Conference on Computer-Assisted Information Retrieval*, March 21–25, 1988. Massachusetts Institute of Technology, Cambridge, USA, 810–829.

Brown, Peter F.; Della Pietra, Stephen; Della Pietra, Vincent J. & Mercer, Robert L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics* 19 (2), 263–311.

Byrne, Jody (ed.) (2006). *Technical translation. Usability strategies for translating technical documentation*. Dordrecht: Springer.

Callison-Burch, Chris; Osborne, Miles & Koehn, Philipp (2006). Re-evaluation the role of bleu in machine translation research. *11th Conference of the European Chapter of the Association for Computational Linguistics*, April 2006, Trento, Italy, 249–256.

Callison-Burch, Chris; Fordyce, Cameron; Koehn, Philipp; Monz, Christof & Schroeder, Josh (2007). (Meta-) evaluation of machine translation. *Association for Computational Linguistics: Proceedings of the Second Workshop on Statistical Machine Translation*, June, Prague, 136–158.

Cao, Deborah (2007). *Translating Law*. Clevedon/Buffalo/Toronto: Multilingual Matters Ltd.

Carbonell, Jaime; Klein, Steven; Miller, David; Steinbaum, Mike; Grassiany, Tomer & Frey, Jochen (2006). Context-based machine translation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, August 8–12, 2006. Cambridge, Massachusetts, USA, 19–28.

Casacuberta, Francisco; Ceausu, Alexandru; Choukri, Khalid; Deligiannis, Miltos; Domingo, Miguel; Garcıa-Martınez, Mercedes; Heranz, Manuel; Papavassiliou, Vassilis; Piperidis, Stelios; Prokopidis, Prokopis & Roussis, Dimitris (2021). *The Covid-19 MLIA@ Eval initiative: Overview of the machine translation task*.

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 (eds.) *Translation Quality Assessment. From Principles to Practice*. Cham: Springer International Publishing, 9–38.

Chan, Sin-Wai (2015). Development of Translation Technology: 1967-2013. In: Chan, Sin-Wai (ed.) *The Routledge Encyclopedia of Translation Technology*. Abingdon/New York: Routledge, 3–31.

Charoenpornsawat, Paisarn; Sornlertlamvanich, Virach & Charoenporn, Thatsanee (2002). Improving translation quality of rule-based machine translation. *Proceedings of the COLING workshop on Machine translation in Asia*, September 2002, Volume 16, 1–6.

Chatzikoumi, Eirini (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering* 26 (2), 137–161.

Chevalier, Monique; Dansereau, Jules & Poulin, Guy (1978). *TAUM-METEO: description du système.* TAUM/Université de Montréal.

Chirilă, Camelia (2014). Errors and difficulties in translating legal texts. *Management Strategies Journal* 26 (4), 487–492.

Cohen, Jacob (1960). A coefficient agreement for nominal scales. *Educational and psychological measurements* 20 (1), 37–46.

Constitution.ru (2001). Die Verfassung der Russischen Föderation. http://www.constitution.ru/10003000/10003000-3.htm (Stand: 04.10.2022).

Costa-Jussa, Marta R. & Fonollosa, José A. R. (2015). Latest trends in hybrid machine translation and its applications. *Computer Speech & Language* 32 (1), 3–10.

Covid-19 MLIA Eval. (o.J.) Aims and Scope. http://eval.covid19-mlia.eu/ (Stand: 04.10.2022).

Dalla-Zuanna, Jean-Marc (2010). Direkte Qualitätsmessung: Evaluierung fremdsprachiger technischer Dokumentation im Bereich Vertrieb; After Sales der Marke Volkswagen Pkw. *Mitteilungen für Dolmetscher und Übersetzer* 2 (2010), 20–25.

Daum, Ulrich (2003). Übersetzen von Rechtstexten. In: Schubert, Klaus (Hg.) *Übersetzen und Dolmetschen: Modelle, Methoden, Technologie*. Tübingen: Gunter Narr Verlag, 33–46.

DeepL (2018a). Willkommen zum DeepL Blog! *DeepL Blog*, 15.02.2018. https://www.deepl.com/de/blog/20180215 (Stand: 04.10.2022).

DeepL (2018b). Dokumente übersetzen mit DeepL. *DeepL Blog*, 16.07.2018. https://www.deepl.com/de/blog/20180716 (Stand: 04.10.2022).

DeepL (2019). DeepL ist der erste Gewinner des Deutschen KI-Ehrenpreises. *DeepL Blog*, 4.10.2019. https://www.deepl.com/de/blog/20191004 (Stand: 04.10.2022).

DeepL (2020a). Passen Sie den DeepL Übersetzer mit Ihrem individuellen Glossar an. *DeepL Blog*, 6.05.2020. https://www.deepl.com/de/blog/20200506 (Stand: 04.10.2022).

DeepL (2020b). Erneuter Durchbruch bei der KI-Übersetzungsqualität. *DeepL Blog*, 6.02.2020. https://www.deepl.com/de/blog/20200206 (Stand: 04.10.2022).

Dugast, Loïc; Senellart, Jean & Koehn, Philipp (2007). Statistical Post-Editing on SYSTRAN's Rule-Based Translation System. *Association for Computational Linguistics: Proceedings of the Second Workshop on Statistical Machine Translation*, June 2007, Prague, 220–223.

Dumitrescu, Adela-Elena (2014). Difficulties and strategies in the process of legal texts translation. *Management Strategies Journal* 26 (4), 502–506.

Fleiss, Joseph L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin* 76 (5), 378–382.

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

Ghasemi, Hadis & Hashemian, Mahmood (2016). A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations. *English Language Teaching* 9 (3), 13–17.

Gornostay, Tatiana (2008). Machine translation evaluation. *Machine Translation Course of the Nordic Graduate School of Language Technology*, October 23, Sweden, 1–6.

Gotti, Maurizio (2016). The translation of legal texts: Interlinguistic and intralinguistic perspectives. *ESP Today (English for Specific Purposes)* 4 (1), 5–21.

Graham, Yvette; Baldwin, Timothy; Moffat, Alistair & Zobel, Justin (2013). Continuous measurement scales in human evaluation of machine translation. *Association for Computational Linguistics: Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, August 8-9, Sofia, Bulgaria, 33–41.

Han, Aaron Li-Feng; Wong, Derek F. & Chao, Lidia S. (2016). Machine translation evaluation: A survey. *arXiv preprint arXiv:1605.04515*.

Haverkort, Kurt (1991). Was Übersetzer schon immer über maschinelle Übersetzung wissen wollten, sich aber nicht zu fragen trauten. *Lebende Sprachen*, 1 (91), 8–12.

House, Juliane (2014). Translation quality assessment: Past and present. In: House, Juliane (ed.) *Translation: A multidisciplinary approach*. London: Palgrave Macmillan, 241–264.

Hundt, Michael G. (1982). Working with the WEIDNER machine-aided translation system. In: Lawson, Veronica (ed.) *Practical experience with machine translation*. North-Holland Publishing Company, 45–51.

Hutchins, John W. (1997a). First steps in mechanical translation. In: Teller Virginia & Sundheim, Beth (eds.) *MT Summit VI: Past, Present, Future, Proceedings*, 29 October – 1 November. San Diego, California, USA, 14–23.

Hutchins, John W. (1997b). From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology. *Machine Translation* 12 (3), 195–252.

Hutchins, John W. (1997c). Fifty years of the computer and translation. *Machine Translation Review* 6, 22–24.

Hutchins, John W. (2000a). Machine translation. In: Ralston, Anthony; Reilly, Edwin D. & Hemmendinger, David (eds.) *Encyclopedia of computer science* (4th edition). Basingstoke/Oxford: Nature Publishing Group, 1059–1066.

Hutchins, John W. (2000b). Yehoshua Bar-Hillel: a philosopher's contribution to machine translation. In: Hutchins, John W. (ed.) *Early years in machine translation*. Amsterdam: John Benjamins, 299–312.

Hutchins, John W. (2004). The Georgetown-IBM experiment demonstrated in January 1954. In: Frederking, Robert E. & Taylor, Kathryn B. (eds.) *Machine translation: From real users to research. Proceedings of the 6th conference of the Association for Machine Translation in the Americas*, September 28–October 2, 2004. Washington, USA, 102–114.

Interfax.ru (2015). In Russland trat das Gesetz über personenbezogene Daten in Kraft (russ.: В России вступил в силу закон о персональных данных). https://www.interfax.ru/business/463850 (Stand: 04.10.2022).

ISO (2010). Ergonomics of Human-System Interaction – Part 210: Human-centred design for interactive systems. ISO 9241-210:2010.

Jiménez-Crespo, Miguel A. (2009). The evaluation of pragmatic and functionalist aspects in localization: towards a holistic approach to Quality Assurance. *The Journal of Internationalization and Localization* 1 (1), 60–93.

Kalchbrenner, Nal & Blunsom, Phil (2013). Recurrent continuous translation models. *Association for Computational Linguistics: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 18-21 October 2013. Seattle, Washington, USA, 1700–1709.

Kandler, Günther (1963). On the problem of quality in translation: Basic considerations. In: Cary, Edmond & Jumpelt, Rudolf W. (eds.) *La qualité en matière de traduction – Quality in translation*. Actes du IIIème Congrès de la Fédération Internationale des Traducteurs. Oxford: Pergamon Press, 291–298.

Kelman, Sveta (2014). Translate Community: Help us improve Google Translate! *Google Translate Blog*. https://translate.googleblog.com/2014/07/translate-community-help-us-improve.html (Stand: 04.10.2022).

Killman, Jeffrey (2014). Vocabulary accuracy of statistical machine translation in the legal context. In: O'Brien, Sharon; Simard, Michel & Specia, Lucia (eds.) *Association for Machine Translation in the Americas: Proceedings of the 3d Workshop on Post-Editing Technology and Practice*, October 22–26. Vancouver, 85–98.

Koby, Geoffrey S.; Fields, Paul; Hague, Daryl; Lommel, Arle & Melby, Alan (2014). Defining Translation Quality. *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació* 12, 413–420.

Koehn, Philipp (2004). Pharaoh: a beam search decoder for phrase-based statistical machine translation models. In: Frederking, Robert E. & Taylor, Kathryn B. (eds.) *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas*. Berlin/Heidelberg/New York: Springer, 115–124.

Koehn, Philipp (2010). *Statistical machine translation*. Cambridge: Cambridge University Press.

Koehn, Philipp (2020). *Neural machine translation*. Cambridge University Press.

Koehn, Philipp; Hoang, Hieu; Birch, Alexandra; Callison-Burch, Chris; Federico, Marcello; Bertoldi, Nicola; Cowan, Brooke; Shen, Wade; Moran, Christine; Zens, Richard; Dyer, Chris; Bojar, Ondřej; Constantin. Alexandra & Herbst, Evan (2007). Moses: Open source toolkit for statistical machine translation. *Association for Computational Linguistics: Proceedings of Demo and Poster Sessions*, June 2007. Prague, 177–180.

Koehn, Philipp; Och, Franz J. & Marcu, Daniel (2003). Statistical phrase-based translation. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology* vol. 1, May 2003, 48–54.

Kurumadom.ru (o.J.). Das Formular eines KFZ-Kaufvertrages. https://kurumadom.ru/wp-content/uploads/2019/07/%D0%B4%D0%BE%D0%B3%D0%BE%D0%B2%D0%BE%D1%80-%D0%BA%D1%83%D0%BF%D0%BB%D0%B8-%D0%BF%D1%80%D0%BE%D0%B4%D0%B0%D0%B6%D0%B8.doc (Stand: 04.10.2022).

Lagarda, Antonio L.; Alabau, Vicent; Casacuberta, Francisco; Silva, Roberto & Díaz-de-Liaño, Enrique (2009). Statistical post-editing of a rule-based machine translation system. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the*

North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, June 2009. Boulder, Colorado, 217–220.

Landis, Richard J. & Koch, Gary G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159–174.

Läubli, Samuel (2020). Machine Translation for Professional Translators. *Zurich Open Repository and Archive*, Doctoral dissertation, University of Zurich.

Läubli, Samuel; Castilho, Sheila; Neubig, Graham; Sennrich, Rico; Shen, Qinlan & Toral, Antonio (2020). A set of recommendations for assessing human–machine parity in language translation. *Journal of Artificial Intelligence Research* 67, 653–672.

Levy, Steven (2011). *In the Plex: How Google Thinks, Works, and Shapes Our Lives*. New York: Simon & Schuster.

Linguee (o.J.). Homepage der Webseite. https://www.linguee.de/ (Stand 10.08.2021).

LISA (2004). LISA Best Practice Guide: Quality Assurance – The Client Perspective. https://ot2009.files.wordpress.com/2009/05/5-lisa-best-practice-guide.pdf (Stand: 14.09.2022).

Lita, Lucian Vlad; Rogati, Monica & Lavie, Alon (2005). Blanc: Learning evaluation metrics for MT. *Association for Computational Linguistics: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, October 2005. Vancouver, 740–747.

Liu, Qun & Zhang, Xiaojun (2015). Machine translation: general. In: Chan, Sin-Wai (ed.) *The Routledge Encyclopedia of Translation Technology* 1. Publ. Abingdon: Routledge, 105–119.

Lommel, Arle (2015). Multidimensional Quality Metrics (MQM) Definition: Issue types (normative). *Deutsches Forschungszentrum für Künstliche Intelligenz GmbH.* http://www.qt21.eu/mqm-definition/definition-2015-06-16.html#issue_types (Stand: 11.07.2021)

Lommel, Arle; Uszkoreit, Hans & Burchardt, Aljoscha (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica* 12, 455–463.

Luong, Minh-Thang (2016). Neural machine translation. Doctoral dissertation, Stanford University.

Maas, Heinz Dieter (1977). The Saarbrücken automatic translation system (SUSY). *Proceedings of the Third European Congress on Information Systems and Networks: Overcoming the Language Barrier* vol. 1, 585–592.

Marx, Jutta; Smith, Nancy & Staudinger, Bernhard (1998). Some Problems in the Evaluation of the Russian-German Machine Translation System MIROSLAV. *Proceedings of the first International Conference on Language Resources & Evaluation* vol. 2, 28-30 May. Granada, Spain, 1219–1226.

Mateo, Roberto Martinez (2014). A deeper look into metrics for translation quality assessment (TQA): a case study. *Miscelánea: A Journal of English and American Studies* 49, 73–96.

Muhr, Rudolf (2009). Die Unterschiede in der Rechtsterminologie Österreichs und Deutschlands und die Folgen für die Rechtssprache Deutsch im Rahmen der Europäischen Union. *Muttersprache*, 119 (3), 199–216.

Mushchinina, Maria (2009). Rechtsterminologie – ein Beschreibungsmodell: Das russische Recht des geistigen Eigentums. *Forum für Fachsprachen-Forschung* Band 78. Berlin: Frank & Timme GmbH.

Nagao, Makoto (2003). A framework of a mechanical translation between Japanese and English by analogy principle. In: Nirenburg, Sergei; Somers, Harold L. & Wilks, Yorick A. (eds.) *Readings in Machine Translation*. Cambridge: The MIT Press, 351–354.

Nida, Eugene A. (2001). Dynamic equivalence in translating. In: Chan, Sin-Wai & Pollard, David E. (eds.) *An Encyclopaedia of Translation: Chinese–English/ English–Chinese*. Hong Kong: Chinese University Press, 223–230.

Nord, Christiane (1993). *Einführung in das funktionale Übersetzen: am Beispiel von Titeln und Überschriften*. Tübingen: UTB.

Nord, Christiane (2006). Translationsqualität aus funktionaler Sicht. In: Schippel, Larisa (Hg.) *Übersetzungsqualität: Kritik–Kriterien–Bewertungshandeln* (Vol. 8). Berlin: Frank & Timme, 11–30.

Papineni, Kishore; Roukos, Salim; Ward Todd & Zhu, Wei-Jing (2002). BLEU: A method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, July 2002. Philadelphia, USA, 311–318.

Pohl, Heinz-Dieter (2011). Österreichisches Deutsch. Überlegungen zur Diskussion um die deutsche Sprache in Österreich. *Klagenfurter Beiträge zur Sprachwissenschaft* 358 (37/38), 63–124.

Popović, Maja (2020). Relations between comprehensibility and adequacy errors in machine translation output. *Proceedings of the 24th Conference on Computational Natural Language Learning*, 256–264.

PROMT (2005). PROMT Selected by NASA Astronauts. https://www.promt.com/press/news/promt-selected-by-nasa-astronauts/ (Stand: 04.10.2022).

PROMT (2019). PROMT Neural Translation Server takes machine translation to a new level. https://www.promt.com/press/news/promt-neural-translation-server-takes-machine-translation-to-a-new-level/ (Stand: 04.10.2022).

PROMT (2020). PROMT presented new solutions for corporate users. https://www.promt.com/press/news/promt-presented-new-solutions-for-corporate-users-/ (Stand: 04.10.2022).

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

Reiß, Katharina (1989). Adäquatheit und Äquivalenz. *HERMES-Journal of Language and Communication in Business* 3, 161–177.

Roturier, Johann (2006). An investigation into the impact of controlled English rules on the comprehensibility, usefulness, and acceptability of machine-translated technical documentation for French and German users. Doctoral dissertation, Dublin City University.

SAE (2001). Surface vehicle recommended practice: SAE J2450 Translation Quality Metric. http://www.apex-translations.com/documents/sae_j2450.pdf (Stand: 04.10.2022).

Šarčevic, Susan (1997). *New approach to Legal Translation.* The Hague: Kluwer Law International.

Schmitt, Peter A. (1999). Qualitätsmanagement. In: Snell-Hornby, Mary; Hönig, Hans G.; Kußmaul, Paul & Schmitt, Peter A. (Hg.) *Handbuch Translation*. 2. verbesserte Auflage. Stauffenburg, 394–399.

Schwartz, Lane (2018). The history and promise of machine translation; In: Lacruz, Isabel & Jääskeläinen, Riitta (eds.) *Innovation and Expansion in Translation Process Research*. Vol. 43, American Translators Association scholarly monograph series. Amsterdam/Philadelphia: John Benjamins Publishing Company, 161–190.

Semenenko, Mikhail (2020) Eheschließung: Die Bestimmumgen, das Eherecht und eheliche Pflichten (russ.: Брак: Условия заключения брака, права и обязанности супругов). https://vk.com/@lawyer_consultation_crimea-brak-usloviya-zaklucheniya-braka-prava-i-obyazannosti-suprug (Stand: 04.10.2022).

Senellart, Jean (2006). Boosting linguistic rule-based MT systems with corpus-based approaches. *Global Autonomous Language Exploitation PI Meeting*.

Senellart, Jean; Dienes, Péter & Varadi, Tamás (2001). New generation Systran translation system. *Proceedings of Machine Translation Summit VIII*, September 18–22. Santiago de Compostela, Spain.

Shiwen, Yu & Xiaojing, Bai (2015). Rule-based machine translation. In: Chan, Sin-Wai (ed.) *The Routledge Encyclopedia of Translation Technology*. Abingdon/New York: Routledge, 186–200.

Smith, James & Clark, Stephen (2009). EBMT for SMT: a new EBMT-SMT hybrid. In: Forcada, Mikel L. & Way, Andy (eds.) *Proceedings of the 3rd International Workshop on Example-Based Machine Translation*, 12–13 November. Dublin, Ireland, 3–10.

Sokolova, Svetlana (1994). STYLUS – the MT product line for Russian: the current state. *Proceedings of the International conference "Machine translation: ten years",* 12–14 November. Cranfield University, England.

Sokolova, Svetlana (1999). The PROMT Translation Technology for Russian and other languages. *Workshop of the European Association for Machine Translation: EU and the new languages*, April 22–23. Prague, 9–10.

Somers, Harold (2003). An Overview of EBMT. In: Carl, Michael & Way, Andy (eds.) *Recent advances in Example-Based Machine Translation*. Dordrecht/Boston/London: Kluwer Academic Publishers, 3–57.

Specia, Lucia; Raj, Dhwaj & Turchi, Marco (2010). Machine translation evaluation versus quality estimation. *Machine translation* 24 (1), 39–50.

Specia, Lucia; Shah, Kashif; De Souza, Jose G. C. & Cohn, Trevor (2013). QuEst – A translation quality estimation framework. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, August 4–9. Sofia, Bulgaria, 79–84.

Stahlberg, Felix (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research* 69, 343–418.

Stein, Daniel (2009). Maschinelle Übersetzung – ein Überblick. *The Journal for Language Technology and Computational Linguistics* 24 (3), 5–18.

Stejskal, Jiri (2006). Quality Assessment in Translation. *The ATA Chronicle* Oktober 2006, 12–16.

SYSTRAN (2018). SYSTRAN celebrates its golden anniversary as a machine translation company by looking back at their most memorable milestones. https://www.systransoft.com/download/press-releases/systran-50-years-of-innovation-in-machine-translation-2018-04-10.pdf (Stand: 04.10.2022).

SYSTRAN (o.J.) About SYSTRAN: A pioneer and global leader in translation solutions. https://www.systransoft.com/systran/ (Stand: 04.10.2022).

Toma, Peter (1977). Systran as a multilingual machine translation system. *Proceedings of the Third European Congress on Information Systems and Networks, Overcoming the language barrier*, 3–6 May, vol. 1. München: Verlag Dokumentation, 569–581.

Toma, Peter (1997). My first 30 years with MT. *MT Summit VI, Machine Translation: Past, Present, Future, Proceedings* 29, 33–34.

Tripathi, Sneha & Sarkhel, Juran Krishna (2010). Approaches to machine translation. *Annals of Library and Information Studies* 57 (December 2010), 388–393.

Turovsky, Barak (2016). Ten years of Google Translate. Google Blog. https://www.blog.google/products/translate/ten-years-of-google-translate/ (Stand: 04.10.2022).

Van Slype, Georges (1979). Critical study of methods for evaluating the quality of machine translation. *Prepared for the Commission of European Communities Directorate General Scientific and Technical Information and Information Management. Report BR, 19142*.

Vasconcellos, Muriel & León, Marjorie (1985). SPANAM and ENGSPAN: machine translation at the Pan American Health Organization. *Computational Linguistics* 11 (2–3), 122–136.

Vauquois, Bernard (1976). Automatic translation – a survey of different approaches. *Statistical Methods in Linguistics*, 127–135.

Vauquois, Bernard (1979). The Evolution of Oriented Software and Formalized Linguistic Models for Automatic Translation or Machine-Aided Translation of Natural Languages. *Doc GETA*, Grenoble, France.

Vieira, Lucas Nunes; O'Hagan, Minako & O'Sullivan, Carol (2021). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society* 24 (11), 1515–1532.

Weaver, Warren (1947). Letters to Norbert Wiener. *Reproduced by permission of the Rockefeller Foundation Archives.* https://aclanthology.org/www.mt-archive.info/Weaver-1947-original.pdf (Stand: 04.10.2022).

Weaver, Warren (1955). Translation. In: Locke, William N. & Booth, Andrew Donald (eds.) *Machine Translation of Languages*. *Fourteen Essays*. Cambridge: The MIT Press, 15–23.

Welnitzova, Katarina; Jakubickova, Barbara & Králik, Roman (2021). Human-Computer Interaction in Translation Activity: Fluency of Machine Translation. *RUDN Journal of Psychology and Pedagogics* 18 (1), 217–234.

Werthmann, Antonina & Witt, Andreas (2014). Maschinelle Übersetzung – Gegenwart und Perspektiven. In: Stickel, Gerhard (Hg.) *Translation and Interpretation in Europe. Contributions to the Annual Conference 2013 of EFNIL in Vilnius.* Frankfurt am Main/Berlin/Bern/Bruxelles/New York/Oxford/Wien: Lang, 79–103.

White, John S. (2003). How to evaluate machine translation. In: Somers, Harold (ed.) *Computers and Translation. A translator's guide*. John Benjamins B.V., 211–244.

Wiesmann, Eva (2019). Machine Translation in the Field of Law: A Study of the Translation of Italian Legal Texts into German. *Comparative Legilinguistics* 37, 117–153.

Wikipedia (2021). Definition des Begriffs „Totschlag“ in Österreich. https://de.wikipedia.org/wiki/Totschlag_(%C3%96sterreich) (Stand: 15.09.2022).

Wikipedia (2022). Definition des Begriffs „Totschlag“ in Deutschland. https://de.wikipedia.org/wiki/Totschlag_(Deutschland) (Stand: 15.09.2022).

Wong, Billy Tak-ming & Webster, Jonathan J. (2015). Example-based machine translation. In: Chan, Sin-Wai (ed.) *The Routledge encyclopedia of translation technology*. Abingdon/New York: Routledge, 137–151.

Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le, Quoc V.; Norouzi, Mohammad; Macherey, Wolfgang; Krikun, Maxim; Cao, Yuan; Gao, Qin; Macherey, Klaus; Klingner, Jeff; Shah, Apurva; Johnson, Melvin; Liu, Xiaobing; Kaiser, Łukasz; Gouws, Stephan; Kato, Yoshikiyo; Kudo, Taku; Kazawa, Hideto; Stevens. Keith; Kurian, George; Patil, Nishant; Wang, Wei; Young, Cliff; Smith, Jason; Riesa, Jason; Rudnick, Alex; Vinyals, Oriol; Corrado, Greg; Hughes, Macduff & Dean, Jeffrey (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv: Computation and Language*, 1–23.

Zhou, Ming; Wang, Bo; Liu, Shujie; Li, Mu; Zhang, Dongdong & Zhao, Tiejun (2008). Diagnostic evaluation of machine translation systems using automatically constructed linguistic check-points. *Proceedings of the 22nd International Conference on Computational Linguistics*, August 2008. Manchester, 1121–1128.

Zilahy, Simon Pietro (1963). Quality in Translation. In: Cary, Edmund & Jumpelt, Rudolf Walter (eds.) *La qualité en matière de traduction Actes du 3e Congrès de la Fédération internationale des Traducteurs*. / Proceeding of the 3rd Congress of the International Federation of Translators. Oxford: Pergamon Press, 285–290.

Zimmermann, Harald H. 1990. Linguistisch-technische Aspekte der maschinellen Übersetzung. In: Buder, Marianne; Rehfeld, Werner & Seeger, Thomas (Hg.) *Grundlagen der praktischen Information und Dokumentation*. 3., völlig neu bearbeitete Auflage. München: K.G. Saur, 264–274.

Zimmermann, Harald H. 2004. Maschinelle und Computergestützte Übersetzung. In: Kuhlen, Rainer; Seeger, Thomas & Strauch, Dietmar (Hg.) *Grundlagen der praktischen Information und Dokumentation.* 5., völlig neu gefasste Auflage. München: K.G. Saur, 475–480.

Anhang

A1 Die untersuchten Ausgangs- und Zieltexte

Text 1: KFZ-Kaufvertrag

Originale russische Version des Textes 1 (Kurumadom.ru o.J.):

Договор купли-продажи транспортного средства

[…]

2. Предмет договора.

2.2 Продавец гарантирует, что указанное в п 2.1. настоящего договора транспортное средство (ТС) не заложено, не находится в споре, под арестом, не обременено правами третьих лиц, не является предметом каких-либо иных сделок, ограничивающих право покупателя на распоряжение им, а также то, что данное ТС полностью и надлежащим образом оформлено для реализации на территории РФ.

[…]

4. Прочие условия

4.3. Споры и разногласия , которые могут возникнуть из настоящего договора решаются путем переговоров между сторонами. В случае, если стороны не придут к согласию во внесудебном порядке, то дело подлежит рассмотрению судом в соответствии с действующим законодательством РФ.

Maschinelle Übersetzung des Textes 1:

Vertrag über den Kauf und Verkauf eines Fahrzeugs

[...]

2. Gegenstand des Abkommens.

2.2 Der Verkäufer garantiert, dass das in Punkt 2.1. genannte Fahrzeug nicht verpfändet, nicht angefochten, nicht verhaftet, nicht mit Rechten Dritter belastet, nicht Gegenstand anderer Transaktionen ist, die das Verfügungsrecht des Käufers einschränken, und dass das Fahrzeug in der Russischen Föderation vollständig und ordnungsgemäß zum Verkauf zugelassen ist.

[...]

4. Andere Bedingungen und Konditionen

4.3 Streitigkeiten und Meinungsverschiedenheiten, die sich aus diesem Vertrag ergeben können, werden durch Verhandlungen zwischen den Parteien beigelegt. Gelingt es den Parteien

nicht, eine außergerichtliche Einigung zu erzielen, so wird die Angelegenheit von einem Gericht nach dem geltenden Recht der Russischen Föderation überprüft.

Text 2: Bestimmungen zur Eheschließung in Russland

Originale russische Version des Textes 2 (Semenenko 2020):

Вступление в брак предполагает ряд обязательных условий, куда входят:

- достижение 18 летнего возраста (в силу возникших исключительных случаев российское законодательство допускает регистрацию брака в возрасте 16 лет);
- отсутствие ранее заключенного брака у будущих супругов (российское законодательство запрещает полигамный (бигамный) брак, или многоженство
- отсутствие родственных и семейных связей у обеих сторон (то есть брак не может быть зарегистрирован между братьями и сестрами, родственниками по прямой и нисходящей линии, а также, между усыновителями и усыновленными);
- дееспособность граждан, которые хотят зарегистрировать брак (то есть в брак не могут вступить лица, признанные судом недееспособными в результате слабоумия либо психологических расстройств).

Maschinelle Übersetzung des Textes 2:

Die Eheschließung bringt eine Reihe von Voraussetzungen mit sich, unter anderem:

- Vollendung des 18. Lebensjahres (in einigen Ausnahmefällen erlaubt die russische Gesetzgebung die Registrierung von Ehen im Alter von 16 Jahren);
- Fehlen einer früheren Ehe der künftigen Ehegatten (die russische Gesetzgebung verbietet polygame (bi-game) Ehen oder Polygamie);
- Fehlen von Verwandtschaftsverhältnissen (d. h. Eheschließungen zwischen Geschwistern, Nachkommen oder Nachkommen von Geschwistern oder zwischen Adoptiveltern und Adoptivkindern können nicht eingetragen werden);
- Die Geschäftsfähigkeit von Bürgern, die ihre Ehe registrieren lassen wollen (d.h. Personen, die von einem Gericht aufgrund von Demenz oder psychischen Störungen für geschäftsunfähig erklärt wurden, dürfen keine Ehe eingehen).

Text 3: Die Verfassung der Russischen Föderation

Originale russische Version des Textes 3 (Constitution.ru 2001):

Раздел первый. Основные положения
Глава 1. Основы конституционного строя
Статья 51.
1. Российская Федерация состоит из республик, краев, областей, городов федерального значения, автономной области, автономных округов - равноправных субъектов Российской Федерации.
2. Республика (государство) имеет свою конституцию и законодательство. Край, область, город федерального значения, автономная область, автономный округ имеет свой устав и законодательство.
3. Федеративное устройство Российской Федерации основано на ее государственной целостности, единстве системы государственной власти, разграничении предметов ведения и полномочий между органами государственной власти Российской Федерации и органами государственной власти субъектов Российской Федерации, равноправии и самоопределении народов в Российской Федерации.

Maschinelle Übersetzung des Textes 3:

Abschnitt eins. Allgemeine Bestimmungen.
Kapitel 1: Grundzüge der verfassungsmäßigen Ordnung
Artikel 51.
1. Die Rußländische Föderation besteht aus Republiken, Krajen, Oblasten, Städten von föderaler Bedeutung, autonomen Oblasten und autonomen Kreisen, die gleichberechtigte Subjekte der Rußländischen Föderation sind.
2. Eine Republik (Staat) hat ihre eigene Verfassung und Gesetzgebung. Ein Kraj, eine Oblast, eine Stadt von föderaler Bedeutung, eine autonome Oblast oder ein autonomer Okrug hat eine eigene Satzung und Gesetzgebung.
3. Die föderale Struktur der Rußländischen Föderation beruht auf ihrer staatlichen Integrität, der Einheit des Systems der Staatsgewalt, der Abgrenzung der Zuständigkeiten und Befugnisse zwischen den Organen der Staatsgewalt der Rußländischen Föderation und den Organen der Staatsgewalt der Subjekte der Rußländischen Föderation sowie der Gleichheit und Selbstbestimmung der Völker der Rußländischen Föderation.

Text 4: Änderungen bei der Datenschutzgrundverordnung in Russland

Originale russische Version des Textes 4 (Interfax.ru 2015):

Во вторник вступили в силу поправки к закону "О персональных данных", которые требуют локализации персональных данных на территории России. Любая российская или зарубежная компания, ориентированная на работу с российскими пользователями, должна обеспечивать запись, систематизацию, накопление, хранение, уточнение персональных данных россиян с использованием баз данных, находящихся на территории РФ.

В обновленном законе пересмотрено само понятие персональных данных. Теперь это не только ФИО и, например, адрес и год рождения, а любая информация, относящаяся к определяемому ею физлицу.

За соблюдением этих норм будет следить Роскомнадзор, которому поручено создание реестра нарушителей законодательства о персональных данных, а также дано право блокировать сайты тех компаний или организаций, которые включены в этот реестр.

Maschinelle Übersetzung des Textes 4:

Am Dienstag traten Änderungen des Gesetzes über personenbezogene Daten in Kraft, wonach personenbezogene Daten in Russland lokalisiert werden müssen. Jedes russische oder ausländische Unternehmen, das mit russischen Nutzern zusammenarbeiten will, muss sicherstellen, dass personenbezogene Daten von Russen mit Hilfe von Datenbanken in Russland erfasst, systematisiert, gespeichert, aufbewahrt und korrigiert werden.

In dem aktualisierten Gesetz wurde der Begriff der personenbezogenen Daten selbst überarbeitet. Jetzt sind es nicht mehr nur der vollständige Name, die Adresse und das Geburtsjahr einer Person, sondern jede Information, die sich auf eine Person bezieht, die durch sie definiert wird.

Die Einhaltung dieser Vorschriften wird von Roskomnadzor überwacht, die mit der Erstellung eines Registers von Verstößen gegen die Rechtsvorschriften über personenbezogene Daten beauftragt wurde und das Recht erhalten hat, die Websites von Unternehmen oder Organisationen zu sperren, die in dem Register aufgeführt sind.

A2 Anleitung zum Annotieren

1. Markieren Sie bitte in der jeweiligen Tabelle jene Stelle im deutschen Zieltext in Rot, die einen Fehler beinhaltet.
2. Tragen Sie dann bitte in der Zeile der jeweiligen Textstelle in der Tabelle eine „1“ in die entsprechende Fehlerkategorie.
3. Wenn ein Satzsegment mehr als nur einen Fehler beinhaltet, sollen die Fehler aufsummiert werden (z. B. bei zwei Fehlern derselben Art “2” eintragen).
4. Wenn ein Fehler sich mehrmals wiederholt, so wird er jedes Mal als Fehler gewertet. Das heißt der Fehler wird noch einmal markiert, kategorisiert und in die Tabelle eingetragen.
5. Lesen Sie bitte zuerst die Beschreibung der Fehlerkategorien, bevor Sie diese in die Tabelle eingeben.
6. Prüfen Sie bitte jede Übersetzung zuerst auf Interpunktionsfehler. Schauen Sie sich zuerst den übersetzten Text an, danach vergleichen sie ihn mit dem Original. Für einen gründlichen Vergleich wurden in der Textaufbereitung die Übersetzung und der Originaltext einander gegenübergestellt. Falls es Interpunktionsfehler gibt, finden Sie das entsprechende Satzsegment in der Annotation zum Text, markieren Sie die Fehler und tragen sie diese in die Tabelle unter der Fehlerklassifizierung „Interpunktion“ (IF) ein.
7. Danach kann der Rest der Fehler überprüft werden. Dafür wurde in der Annotation jedes Satzsegment separat eingeführt.
8. Fehler sollen nur markiert und eingetragen und nicht ausgebessert werden. Nur im Fall von „Sonstigen Fehlern“ wird ein kurzer Kommentar mit einer Erklärung benötigt.
9. Bitte überprüfen Sie nur die deutschen Satzsegmente. Der russische Text dient nur als Referenz und entspricht nicht immer dem Satzbau des Originalsatzes, sondern wird zur Überprüfung von Rechtschreibfehlern, falschen Benennungen oder Bedeutungen, Strukturfehlern oder Auslassungen verwendet.
10. Wenn ein Fehler gleichzeitig zu mehreren Fehlerkategorien gehört, annotieren Sie bitte alle zutreffenden Fehlerkategorien in der Zeile des Textsegments mit „1“.
11. Bitte vergessen Sie nicht, den Schweregrad des jeweiligen Fehlers anzugeben. Dafür wurde farblich visuell „leichter Fehler“ in Blau und „schwerer Fehler“ in Rot in der Tabelle markiert.
12. Wenn Sie sich nicht sicher sind, ob es ein schwerer oder ein leichter Fehler ist, bewerten Sie diesen bitte als „leichter Fehler“.

A3 Beschreibung der Fehlerkategorien nach SAE-J2450

Falsche Benennung (FBN)

- fehlende und/oder falsche Fachterminologie;
- falsche Übersetzung von Abkürzungen, Akronymen, Zahlen, Eigennamen (wie z. B. Markennamen, Ortsnamen und Personennamen etc.) (SAE 2001: 5).
- Nichteinhaltung der terminologischen Konsistenz: Wenn ein bestimmter Begriff, Gegenstand oder Sachverhalt im Zieltext nicht einheitlich und konsistent verwendet wird, wird der Fehler ebenfalls als „falsche Benennung“ klassifiziert (Atesman o. J.: 5).

Falsche Bedeutung (FBD)

Die Bedeutung des übersetzten Textabschnittes stimmt mit der Bedeutung des Ausgangstextabschnittes nicht überein und/oder ist das genaue Gegenteil davon, was im Zieltext impliziert wird (Dalla-Zuanna 2010: 23).

Auslassung (A)

„Eine zusammenhängende Textstelle“, die im Zieltext fehlt (Dalla-Zuanna 2010: 24). Das können ein Wort, ein Satz, ein Segment oder ein Absatz etc. sein (Dalla-Zuanna 2010: 24).

Strukturfehler (SF)

In diese Kategorie gehören Grammatik-, Syntax-, Kollokations- und Wortbildungsfehler (Dalla-Zuanna 2010: 24; Atesman o. J.: 8).

Alle Fehler dieser Kategorie müssen als „Strukturfehler“ markiert werden und brauchen keine genaue Präzisierung. Die Kategorie „Strukturfehler“ umfasst folgende Fehler wie:

- Der Zieltext enthält eine falsche Satzstruktur (z. B. einen Relativsatz, obwohl eine Verbalphrase erforderlich wäre) (SAE 2001: 6);
- die Wörter der Zielsprache sind korrekt, aber in der falschen linearen Reihenfolge gemäß den syntaktischen Regeln der Zielsprache (SAE 2001: 6);
- Fehler bei Kasus, Geschlecht, Zahl, Zeitform, Flexion, Präfix, Suffix und Infix (SAE 2001: 7);
- falsche Artikel/ Konjunktionen/ Adverbien/ Präpositionen/ (Atesman o. J.: 8);
- falsche Kollokation: „eine Kombination von Wörtern, die semantisch zwar zutreffen, aber konventionell nicht miteinander verknüpft werden“ (Atesman o. J.: 8) und;
- der Satz ergibt keinen Sinn, obwohl er syntaktisch und grammatikalisch richtig ist.

Rechtschreibfehler (RF)

- die Schreibweise verstößt gegen die akzeptierten Rechtschreibnormen der Zielsprache und/oder
- beim Textfragment wird ein falsches Schriftsystem benutzt (SAE 2001: 8) sowie
- fehlende Buchstaben, Tippfehler etc. (Atesman o. J: 9).

Interpunktionsfehler (IF)

Die Zeichensetzung des Zieltextes verstößt gegen die Interpunktionsregel der Zielsprache oder/und wird im Text nicht einheitlich eingesetzt (Atesman o. J.: 10).

Sonstiger Fehler (SF)

Alle aufgetretenen Fehler, die nicht eindeutig den oben beschriebenen Fehlerkategorien zuzuordnen sind, sollten als „sonstige Fehler" markiert werden. Wenn Sie einen Fehler als sonstigen Fehler annotieren, fügen Sie bitte eine kurze Beschreibung der Art des Fehlers in die Spalte „Kommentar" in der Zeile des jeweiligen Textsegments ein.

Schweregrad der Fehler

Zusätzlich zur Identifizierung und Kategorisierung der Fehler soll ein Schweregrad für jeden markierten Fehler festgelegt werden. SAE-J2450 unterscheidet zwischen einem **schweren** und einem **leichten** Fehler.

Es handelt sich eindeutig um einen **schweren Fehler**, wenn dieser Übersetzungsfehler mit großem Risiko oder möglichen Fehlhandlungen (Sachschaden, Personenschaden etc.) für Endbenutzer*innen verbunden ist oder/und eine erhebliche Auswirkung auf die Bedeutung des übersetzen Textes hat (Atesman o. J.: 12, SAE 2001:3, Dalla-Zuanna 2010: 24). Darüber hinaus kann ein schwerer Fehler ebenso für eine Irritation bei Endbenutzer*innen sorgen (Atesman o. J.: 12).

Ein **leichter Fehler** wird als „ein Fehler, der zu keiner bzw. zu einer geringfügigen Irritation des Anwenders ohne die Gefahr einer Fehlhandlung führt" definiert (Dalla-Zuanna 2010: 24).

A4 Annotationsblätter

Annotationsblatt Text 1: Kaufvertrag (Kurumadom.ru o.J.)

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
1	Договор	Vertrag															
	-	über															
	купли-продажи	den Kauf und Verkauf															
	транспортного средства	eines Fahrzeugs															
2	Предмет договора.	Gegenstand des Abkommens.															
3	Продавец	Der Verkäufer															
	гарантирует, что	garantiert, dass															
	указанное в п 2.1. настоящего договора транспортное средство (ТС)	das in Punkt 2.1. genannte Fahrzeug															
	не заложено,	nicht verpfändet,															
	не находится в споре,	nicht angefochten,															
	под арестом,	nicht verhaftet,															
	не обременено	nicht mit Rechten															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	правами третьих лиц,	Dritter belastet,															
	не является предметом каких-либо иных сделок,	nicht Gegenstand anderer Transaktionen ist,															
	ограничивающих право покупателя на распоряжение им,	die das Verfügungsrecht des Käufers einschränken,															
	а также то, что данное ТС	und dass das Fahrzeug															
	на территории РФ	in der Russischen Föderation															
	полностью и надлежащим образом	vollständig und ordnungsgemäß															
	оформлено для реализации	zum Verkauf zugelassen ist.															
4	Прочие условия	Andere Bedingungen und Konditionen															
5	Споры и разногласия,	Streitigkeiten und Meinungsverschiedenheiten,															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	которые могут возникнуть из настоящего договора	die sich aus diesem Vertrag ergeben können,															
	решаются путем переговоров между сторонами.	werden durch Verhandlungen zwischen den Parteien beigelegt.															
6	В случае, если стороны не	Gelingt es den Parteien nicht,															
	придут к согласию во внесудебном порядке,	eine außergerichtliche Einigung zu erzielen,															
	то дело	so wird die Angelegenheit															
	судом	von einem Gericht															
	в соответствии с действующим законодательством РФ	nach dem geltenden Recht der Russischen Föderation															
	подлежит рассмотрению	überprüft.															

Annotationsbaltt Text 2: Bestimmungen zur Eheschließung in Russland (Semenenko 2020)

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
1	Вступление в брак	Die Eheschließung															
	предполагает ряд обязательных условий,	bringt eine Reihe von Voraussetzungen mit sich,															
	куда входят:	unter anderem:															
2	- достижение 18 летнего возраста	-Vollendung des 18. Lebensjahres															
	(в силу возникших исключительных случаев	(in einigen Ausnahmefällen															
	российское законодательство допускает	erlaubt die russische Gesetzgebung															
	регистрацию брака	die Registrierung von Ehen															
	в возрасте 16 лет);	im Alter von 16 Jahren)															
3	отсутствие	- Fehlen															
	ранее заключенного брака	einer früheren Ehe															
	у будущих супругов	der künftigen Ehegatten															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		Kommentar bei „Sonstigen Fehlern“
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	(российское законодательство	(die russische Gesetzgebung															
	запрещает	verbietet															
	полигамный (бигамный) брак, или многоженство	polygame (bi-game) Ehen oder Polygamie)															
4	отсутствие родственных и семейных связей у обеих сторон	- Fehlen von Verwandtschaftsverhältnissen															
	(то есть брак между братьями и сестрами,	(d. h. Eheschließungen zwischen Geschwistern,															
	родственниками по прямой и нисходящей линии,	Nachkommen oder Nachkommen von Geschwistern															
	а также, между усыновителями и усыновленными	oder zwischen Adoptiveltern und Adoptivkindern															
	не может быть зарегистрирован)	können nicht eingetragen werden)															
5	- дееспособность граждан,	- Die Geschäftsfähigkeit von Bürgern,															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		Kommentar bei „Sonstigen Fehlern“
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	которые хотят зарегистрировать брак	die ihre Ehe registrieren lassen wollen															
	(то есть лица, признанные судом недееспособными в результате слабоумия либо психологических расстройств,	(d.h. Personen, die von einem Gericht aufgrund von Demenz oder psychischen Störungen für geschäftsunfähig erklärt wurden,															
	не могут вступить в брак).	dürfen keine Ehe eingehen).															

Annotationsblatt Text 3: Die Verfassung der Russischen Föderation (Constitution.ru 2001)

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
1	Раздел первый.	Abschnitt eins.															
2	Основные положения.	Allgemeine Bestimmungen.															
3	Глава 1.	Kapitel 1.															
	Основы конституционного строя	Grundzüge der verfassungsmäßigen Ordnung															
4	Статья 51.	Artikel 51.															
5	1. Российская Федерация	1. Die Rußländische Föderation															
	состоит из	besteht aus															
	республик,	Republiken,															
	краев,	Krajen,															
	областей,	Oblasten,															
	городов федерального значения,	Städten von föderaler Bedeutung,															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	автономной области,	autonomen Oblasten															
	автономных округов - равноправных субъектов Российской Федерации.	und autonomen Kreisen, die gleichberechtigte Subjekte der Rußländischen Föderation sind.															
6	2. Республика (государство)	2. Eine Republik (Staat)															
	имеет	hat															
	свою	ihre eigene															
	конституцию и законодательство.	Verfassung und Gesetzgebung.															
7	Край,	Ein Kraj,															
	область,	eine Oblast,															
	город федерального значения,	eine Stadt von föderaler Bedeutung,															
	автономная область,	eine autonome Oblast															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	автономный округ	oder ein autonomer Okrug															
	имеет	hat															
	свой	eine eigene															
	устав и законодательство.	Satzung und Gesetzgebung.															
8	3. Федеративное устройство Российской Федерации	3. Die föderale Struktur der Rußländischen Föderation															
	основана	beruht auf															
	ее государственной целостности,	ihrer staatlichen Integrität,															
	единстве системы государственной власти,	der Einheit des Systems der Staatsgewalt,															
	разграничении предметов ведения	der Abgrenzung der Zuständigkeiten															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	и полномочий между органами государственной власти Российской Федерации	und Befugnisse zwischen den Organen der Staatsgewalt der Rußländischen Föderation															
	и органами государственной власти субъектов Российской Федерации,	und den Organen der Staatsgewalt der Subjekte der Rußländischen Föderation															
	равноправии	sowie der Gleichheit															
	и самоопределении	und Selbstbestimmung															
	народов в Российской Федерации.	der Völker der Rußländischen Föderation.															

Annotationsblatt Text 4: Änderungen bei der Datenschutzgrundverordnung in Russland (Interfax.ru 2015)

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
1	Во вторник	Am Dienstag															
	вступили	traten															
	поправки к закону	Änderungen des Gesetzes															
	"О персональных данных",	über personenbezogene Daten															
	в силу	in Kraft,															
	которые	wonach															
	требуют локализации персональных данных на территории России.	personenbezogene Daten in Russland lokalisiert werden müssen.															
2	Любая российская или зарубежная компания,	Jedes russische oder ausländische Unternehmen,															
	ориентированная на работу с российскими пользователями,	das mit russischen Nutzern zusammenarbeiten will,															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	должна обеспечивать	muss sicherstellen,															
	персональных данных россиян	dass personenbezogene Daten von Russen															
	с использованием баз данных,	mit Hilfe von Datenbanken															
	находящихся на территории РФ.	in Russland															
	запись,	erfasst,															
	систематизацию,	systematisiert,															
	накопление,	gespeichert,															
	хранение,	aufbewahrt															
	уточнение	und korrigiert werden.															
3	В обновленном законе	In dem aktualisierten Gesetz															
	-	wurde															
	само понятие персональных данных.	der Begriff der personenbezogenen Daten selbst															
	пересмотрено	überarbeitet.															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
4	Теперь это не только	Jetzt sind es nicht mehr nur															
	ФИО	der vollständige Name,															
	и, например, адрес	die Adresse															
	и год рождения,	und das Geburtsjahr einer Person,															
	а любая информация,	sondern jede Information,															
	относящаяся к определяемому ею физлицу.	die sich auf eine Person bezieht, die durch sie definiert wird.															
5	За соблюдением этих норм	Die Einhaltung dieser Vorschriften															
	будет следить Роскомнадзор,	wird von Roskomnadzor überwacht,															
	создание реестра нарушителей	die mit der Erstellung eines Registers von Verstößen															

Satz N.	Satzsegment Original	Satzsegment Übersetzung	Fehlerklassifizierung														Kommentar bei „Sonstigen Fehlern“
			Falsche Benennung		Falsche Bedeutung		Auslassung		Strukturfehler		Rechtschreibfehler		Interpunktionsfehler		Sonstiger Fehler		
			L	S	L	S	L	S	L	S	L	S	L	S	L	S	
	законодательства о персональных данных,	gegen die Rechtsvorschriften über personenbezogene Daten															
	которому поручено	beauftragt wurde															
	а также дано право	und das Recht erhalten hat,															
	сайты тех компаний или организаций	die Websites von Unternehmen oder Organisationen															
	блокировать	zu sperren,															
	которые включены в этот реестр.	die in dem Register aufgeführt sind.															

Abstract (Deutsch)

Die vorliegende Arbeit setzt sich mit der Qualität der maschinellen Übersetzung mit dem Schwerpunkt Rechtsübersetzung auseinander. In der Arbeit wird analysiert, welche Arten von Fehlern maschinell übersetze Texte am häufigsten aufweisen. Es wird überprüft, ob falsche Terminologie die Fehlerkategorie ist, in der die meisten Fehler zu finden sind.

Für die Untersuchung werden vier Texte aus dem Rechtsbereich aus dem Russischen ins Deutsche maschinell übersetzt. Für die Erstellung maschineller Übersetzungen wird der neuronale maschinelle Online-Übersetzter DeepL verwendet. Die maschinellen Übersetzungen werden anhand der Fehlertypologie der SAE-J2450 manuell evaluiert.

Die Untersuchung der maschinell erstellten Übersetzungen zeigt, dass das untersuchte NMÜ-System im Bereich der Rechtsübersetzung eine ungenügende Übersetzungsqualität liefert, da keiner der maschinell übersetzten Texte die SAE-J2450-Qualitätsprüfung bestanden hat. Fast die Hälfte aller festgestellten Fehler (48 %) gehört zur Kategorie „falsche Benennung“, die sich mit der Verwendung der falschen Terminologie befasst. Dies deutet darauf hin, dass das Problem der richtigen Terminologie dem NMÜ-System DeepL große Schwierigkeiten bereitet. Die Untersuchungsergebnisse lassen annehmen, dass die meisten Schwierigkeiten des DeepL-Systems mit der mangelnden Zuverlässigkeit des verwendeten Korpus zusammenhängen.

Abstract (English)

The present paper examines the quality of machine translation with a focus on legal translation. The paper attempts to analyse which types of errors are most frequently found in machine-translated texts. It is examined whether incorrect terminology is the error category in which most errors occur.

For the investigation, four texts from the field of law are machine translated from Russian into German. The online neural machine translator DeepL is used to create machine translations. The machine translations are evaluated manually using the error typology of SAE-J2450.

The examination of the machine translations shows that the investigated neural machine translation system provides insufficient translation quality in the field of legal translation, as none of the machine translated texts passed the SAE-J2450 quality test. Almost half of all errors found (48%) belong to the category “Wrong Term”, which deals with the use of the wrong terminology. This indicates that the problem of correct terminology causes great difficulties for the neural machine translation system DeepL. The research results suggest that most of the difficulties of the DeepL system are related to the lack of reliability of the corpus used.