



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

**Pre-Editing als Vorschlag zur Verbesserung Neuronaler
Maschinelles Übersetzung**

verfasst von / submitted by

Dott.ssa Fiorenza Finetti

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 070 348 331

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Translation Italienisch Deutsch

Betreut von / Supervisor:

Ass.-Prof. Mag. Dr. Dagmar Gromann, BSc

Danksagung

Zuallererst möchte ich meiner Betreuerin Ass.-Prof. Mag. Dr. Dagmar Gromann, BSc meine Dankbarkeit für ihre wertvolle Unterstützung vor und während der Verfassung der vorliegenden Masterarbeit aussprechen. Ich möchte mich bei ihr für ihre hilfreichen Ratschläge und für ihre Hilfsbereitschaft und Professionalität bedanken.

Ich möchte mich herzlich bei meinen Studienkolleginnen und vor allem Freundinnen bedanken, weil sie unter allen Umständen zu mir stehen. Ich freue mich sehr, dass ich diesen Weg gemeinsam mit ihnen beschreiten konnte. Danke an meine Mitbewohnerinnen, die die Schreibphase dieser Arbeit mit mir erlebt haben und die mich ständig motiviert haben. Insbesondere möchte ich Barbara danken, die nicht nur immer da für mich war, sondern auch das Lektorat der vorliegenden Arbeit übernommen hat.

Il Grazie più grande va alla mia famiglia. Per il sostegno e la spinta morale, per la fiducia costante e per non aver mai smesso di credere in me. Sopra ogni cosa sono profondamente grata ai miei genitori per avermi permesso di arrivare dove sono oggi. Questa laurea la dedico a loro. Infine ringrazio tutti coloro che seppur a distanza hanno saputo appoggiare le mie scelte, sempre.

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	iii
Abbildungsverzeichnis	iv
Tabellenverzeichnis.....	v
1. Einleitung.....	1
2. Grundlagen.....	6
2.1. Maschinelle Übersetzung: Entwicklung.....	6
2.1.1. Verschiedene Formen von MÜ	10
2.1.2. Neuronale Maschinelle Übersetzung.....	12
2.1.2.1. Architekturen und Funktionsweisen.....	13
2.1.2.2. Aktuelle Herausforderungen der NMÜ	15
2.1.2.3. Verfügbare NMÜ-Systeme: ein Vergleich.....	17
2.1.3. Hybride Ansätze	19
2.2. Qualität der Übersetzung.....	20
2.2.1. SAE J2450.....	21
2.2.2. Multidimensional Quality Metrics (MQM)	22
2.2.3. Göpferichs Modell.....	24
2.3. Qualität maschineller Übersetzung.....	24
2.3.1. Menschliche Evaluierung der MÜ.....	26
2.3.1.1. TQA-Kriterien.....	27
2.3.1.2. Ranking.....	28
2.3.1.3. Fehlerklassifizierung	28
2.3.1.4. Linguistische Evaluierung: Test Suites	31
2.3.1.5. Evaluierung auf Dokumentenbasis.....	32
2.3.2. Automatische Evaluierung der MÜ.....	32
2.3.2.1. Metriken.....	34
2.3.2.2. Quality Estimation Ansätze	36
2.4. MTranslatability.....	39
2.5. Pre-Editing.....	41
2.5.1. Tools für Pre-Editing	42
2.6. Best-Worst-Scaling	43
3. Aktueller Stand der Forschung.....	47

4. Forschungsmethodik.....	50
4.1. Auswahl der Texte	51
4.2. Fehlertypologie	52
4.3. Pre-Editing-Methoden	53
4.3.1. Auswahl der Methoden	54
4.3.2. Praktischer Einsatz der Methoden	55
4.4. Maschineller Übersetzungsprozess	55
4.4.1. Neuronale Übersetzungssoftware: SYSTRAN Translate	55
4.5. Evaluierung durch Best-Worst-Scaling	58
5. Ergebnisse der Datenerhebung.....	62
5.1. Profile der Proband*innen	70
6. Diskussion.....	72
6.1. Die Zahlen als Antwort	72
6.2. Bedeutung der Ergebnisse	73
6.3. Best-Worst-Scaling: Vorteile und Nachteile	75
6.4. Bezug auf andere Studien	76
6.5. Bemerkungen zur Methode.....	77
7. Fazit	79
Bibliographie.....	81
Anhang 1.....	90
Ausgangstext 1 und 8.....	90
Ausgangstext 2 und 3.....	93
Ausgangstext 4 und 12.....	95
Ausgangstext 5 und 7.....	96
Ausgangstext 6	100
Ausgangstext 9	101
Ausgangstext 10.....	103
Ausgangstext 11.....	105
Anhang 2.....	107
Abstract.....	125

Abkürzungsverzeichnis

KI = Künstliche Intelligenz

MÜ = Maschinelle Übersetzung

SMÜ = Statistische Maschinelle Übersetzung

NMÜ = Neuronale Maschinelle Übersetzung

KNN = Künstliche Neuronale Netze

CAT = engl. *Computer-Assisted-Translation*

ALPAC = engl. *Automatic Language Processing Advisory Committee*

TMS = Translation-Memory-System

GPU = engl. *Graphic Processing Unit*, Grafikprozessor(en)

RNN = Rekursive Neuronale Netze, engl. *Recurrent Neural Network*

CNN = Konvolutionsnetze, engl. *Convolutional Neural Network*

TQA = engl. *Translation Quality Assessment*

SAE = engl. *Society of Automotive Engineers*

MQM = engl. *Multidimensional Quality Metrics*

WER = engl. *Word Error Rate*

BLEU = engl. *Bilingual Evaluation Understudy*

METEOR = engl. *Metric for Evaluation of Translation with Explicit ORdering*

TER = engl. *Translation Error Rate*

QE = engl. *Quality Estimation*

CE = engl. *Confidence Estimation*

EEA = engl. *EasyEnglishAnalyzer*

BWS = engl. *Best-Worst Scaling*

AT = Ausgangstext

ZT = Zieltext

K = Kombination (aus Pre-Editing-Methoden)

PE = Pre-Editing(-Methode)

kP = kein Pre-Editing

P = Proband*in

Abbildungsverzeichnis

Abbildung 1: Phasen der empirischen Studie	5
Abbildung 2: MQM Kern. (Lommel et al. 2015:o. S.).....	23
Abbildung 3: Fehlerklassifizierung (Vilar et al. 2006:699).....	29
Abbildung 4: Angepasste Fehlertypologie für die vorliegende wissenschaftliche Arbeit	53
Abbildung 5: Screenshot der Weboberfläche von SYSTRAN Translate, praktisches Beispiel (SYSTRAN Translate Weboberfläche 2020)	56
Abbildung 6: Screenshot der möglichen Übersetzungsvarianten von SYSTRAN Translate (SYSTRAN Translate Weboberfläche 2020)	57
Abbildung 7: Verteilung der Gesamtbewertungen pro Pre-Editing-Methode.....	70

Tabellenverzeichnis

Tabelle 1: Liste der angewandten Pre-Editing-Methoden	54
Tabelle 2: Einteilung der AT und ZT für das Best-Worst-Scaling	60
Tabelle 3: Bewertungen aller Übersetzungen und jeweiligen Pre-Editing-Methoden nach allen Proband*innen.....	62
Tabelle 4: Gesamtpunkte pro Pre-Editing-Methode	63
Tabelle 5: Inkonsistente Bewertungen Proband*in 1 und 5	64
Tabelle 6: Gesamtbewertungen Pre-Editing-Methode 3	64
Tabelle 7: Detail der fünf Bewertungen für 10.4 und 11.2.....	65
Tabelle 8: Vergleich Übersetzungsvarianten für AT 10 ohne Pre-Editing und mit Methode 3	66
Tabelle 9: Vergleich Übersetzungsvarianten für AT 11 ohne Pre-Editing und mit Methode 3	67
Tabelle 10: Gesamtbewertungen Pre-Editing-Methode 5K.....	68
Tabelle 11: Gesamtbewertungen kein Pre-Editing.....	68
Tabelle 12: Gesamtpunkte pro Pre-Editing-Methode pro Proband*in.....	73

1. Einleitung

„Human beings are generally not prone to confuse the execution of a command with the execution of a political leader, and we expect our translations to reflect that.”¹

Mathias Winther Madsen

Künstliche Intelligenz (KI) und ihre praktischen Anwendungen – Maschinen bzw. Algorithmen – kommen seit einigen Jahren in verschiedenen Bereichen zum Einsatz. Gabriel (2019) hat sich aber die Frage gestellt, in welchen von diesen Bereichen es möglich sein wird, dass die künstliche Intelligenz einige menschliche Beschäftigungen automatisieren wird (2019:95). Ziele ihres Einsatzes von Seite der Unternehmen sind eine Reduktion der Kosten aber auch eine eindeutige Steigerung der Kundenzufriedenheit und eine bessere Qualität, was die Prozesse und Ergebnisse betrifft (2019:96). Zu den von der Europäischen Kommission vorgeschlagenen Zielen zählt auch eine höhere Zugänglichkeit der Übersetzungen (vgl. Lazzari & Steinbiss 2006:25). Besonders betroffen von der aktuellen Forschung ist der Bereich der Dienstleistungen wie zum Beispiel die Finanzwirtschaft, Medienindustrie und Gesundheitswirtschaft (vgl. Tractica 2016:9; vgl. Gabriel 2019:97). Insbesondere wird jedoch der Bereich der Sprachübersetzung erforscht, welcher als praktische Anwendung von künstlicher Intelligenz gilt (vgl. Gabriel 2019:97).

Eine Verbindung zwischen KI und Sprachübersetzung bietet die maschinelle Übersetzung (MÜ), mit dem Ziel, eine Sprache und ihre Formulierungen mit den im Kontext passenden Wörtern in eine andere Sprache zu übertragen. Nach Kenny (2018) ist unter „maschineller Übersetzung“ ein Prozess gemeint, bei dem die interlinguale Umwandlung eines Textes durch eine Maschine bzw. einen Computer durchgeführt wird. Selbst wenn die Funktionsfähigkeit der Maschine auf menschlichen Anstrengungen vor oder nach der Laufzeit der Maschine basiert (vgl. Kenny 2018:429). Grundlage der Entwicklung maschineller Übersetzung war und ist noch heute die stetig wachsende Globalisierung und der damit verbundene hohe Übersetzungsbedarf. Gleich in ihren Anfangszeiten war es aber schon klar, dass die maschinelle Übersetzung gewisse Grenzen aufwies: „Ohne jegliche Alltagserfahrung und ohne ein zumindest rudimentäres ‚Verständnis‘ für die Zusammenhänge im Satz oder gar

¹ (Madsen, 2009:6)

im gesamten Text ist es dem Computer nicht möglich, Mehrdeutigkeiten aufzulösen” (Schmalz 2019:195).

Die Frage der Qualität und des Nutzens der maschinellen Übersetzung ist immer mit ihrer Entwicklung seit den 1950er Jahren einhergegangen und seitdem haben sich viele Forscher*innen damit auseinandergesetzt. Flanagan (1994) behauptet, dass tatsächlich verschiedene Übersetzungen des gleichen Textes als korrekt bzw. akzeptabel betrachtet werden können. Außerdem sind Fehler im Zieltext oft schwierig einzugrenzen, da eine maschinelle Übersetzungssoftware mehrere Textstellen falsch übersetzen könnte und dies zu einer Fehlerkette führen würde (1994:65f.). In einem Bericht aus dem Jahre 2006 hat sogar die Europäische Kommission die Humanübersetzung als qualitativ hochwertiger als die maschinelle Übersetzung eingeschätzt; die stärkere Position der MÜ am Markt könne ausschließlich durch eine qualitative Verbesserung ermöglicht werden (vgl. Lazzari & Steinbiss 2006:25). Madsen (2009) hat hervorgehoben, dass der Mangel an Metaphern und Metonymien einer ganz automatischen maschinellen Übersetzung verhindern wird, das Qualitätsniveau einer Humanübersetzung zu erreichen. Außerdem wird angenommen, dass Maschinen ihre Fehler nicht erkennen können, da sie nicht in der Lage sind, einzuschätzen, welche ihrer Resultate korrekt sind (2009:102). Nach Burchardt und Porsiel (2017:16) kann die MÜ auf sprachlicher Ebene „nicht oder nur sehr eingeschränkt [...] fehlertolerant sein, ahnen, interpretieren, vermuten, assoziieren, antizipieren oder gar paraphrasieren, zwischen den Zeilen lesen, Unübersetzbares (richtig) verarbeiten oder auch nur Sinnvolles von Sinnlosem trennen“. Folgende weitere Forschungsbeispiele haben die zu niedrige Qualität der MÜ deutlich gemacht. Stankevičiūtė et al. (2017:75) behaupten, dass die maschinelle Übersetzung viele Schwierigkeiten auf syntaktischer und semantischer Ebene bereitet. Weiterhin ist Ausgangspunkt der Studie von Miyata und Fujita (2017:47), dass maschinelle Übersetzungssysteme derzeit noch keine Rohübersetzungen liefern können, die qualitativ vergleichbar mit menschlichen sind.

Mit anderen Worten bleibt die Frage der Qualität noch unbeantwortet, obwohl stetig neue Entwicklungen maschineller Übersetzung stattfinden. Zu den verschiedenen Ansätzen der Computerlinguistik zählen bis heute regelbasierte, statistische und hybride Übersetzungssysteme. Seit dem Jahr 2016 stellt die Neuronale Maschinelle Übersetzung (NMÜ) den aktuellen Stand der Forschung dar. Dieser jüngste Fortschritt solle eine Annäherung der maschinellen Übersetzung an eine menschliche Übersetzung ermöglichen. Dementsprechend sollen natürlichere Satzstrukturen und weniger Fehler im Vergleich zur vorigen statistischen maschinellen Übersetzung erzielt werden in Bezug auf Syntax und

Wortreihenfolge (vgl. Schmalz 2019:198f.). Der Grund dafür ist, dass die neuronale maschinelle Übersetzung auf Deep Learning, und zwar auf künstlichen neuronalen Netzen (KNN) basiert, die Wissen aus großen Datenmengen eigenständig extrahieren (2019:194).

Die Intention dieser Masterarbeit ist eine Verbesserung der Qualität der neuronalen maschinellen Übersetzung zu erreichen. Die bestehenden Fehler in den maschinellen Resultaten sollten durch das Präeditieren der Ausgangstexte durch bestimmte Methoden möglichst vollständig aufgehoben werden. Mit Präeditieren wird gemeint, dass ein Ausgangstext spezifischen terminologischen und stilistischen Veränderungen bzw. Regeln unterworfen wird, um eine bessere Qualität des entsprechenden maschinell erstellten Zieltextes zu erzielen (vgl. Guerberof Arenas 2019:3). Bei dieser Masterarbeit wird der Fokus auf die Vorbereitungs- und nicht auf die Nachbereitungsphase eines Textes gelegt. Dies stellt einen Perspektivenwechsel in diesem Bereich dar. Als Annahme dieser Arbeit gilt, dass die Resultate der maschinellen Übersetzungssysteme nicht „lieferbereit“ sind. Aus diesem Grund wird im Laufe dieser Forschungsarbeit untersucht, ob die Resultate von maschinellen Übersetzungssystemen im Bereich Biologie qualitativ hochwertiger sein können, wenn bestimmte Pre-Editing-Methoden und Kombinationen daraus für das Sprachpaar Englisch-Italienisch angewandt werden. Das menschliche bzw. manuelle Präeditieren einiger Ausgangstexte sollte eventuell zur Erstellung einer Liste bestimmter sprachspezifischer Methoden dienen. Dies könnte eine Verringerung der maschinellen Übersetzungsfehler für das oben erwähnte Sprachpaar gewährleisten ohne Einsatz von Post-Editing-Strategien.

Ein zentrales Konzept dieser Arbeit, das die Begriffe Pre-Editing und Qualität zusammenlegt, ist „MTranslatability“. Nach Bernth und Gdaniec (2001) wird damit der Grad der Übersetzbarkeit eines Ausgangstextes durch maschinelle Übersetzungssysteme bezeichnet. Im Rahmen ihrer Studie wird eine Liste von möglichen Korrekturen des Ausgangstextes unter Anwendung von Pre-Editing-Strategien vorgestellt, zwecks der Optimierung der Qualität maschineller Resultate. Die Effektivität der angewandten Pre-Editing-Methoden im Bereich Biologie für die gewählte Sprachkombination wird untersucht, die Studie wird anschließend die Richtung der Tendenzen aufzeigen.

Die Evaluierung der Qualität der Zieltexte nach Einsatz von Pre-Editing wird menschlich durchgeführt, um die Vorstellungen der Proband*innen über den Grad der Qualität einer Übersetzung direkt mit den maschinellen Ergebnissen zu vergleichen. Menschliche Evaluierungen stellen tatsächlich noch den Standard für die Qualitätsbewertung dar, da die Richtigkeit einer Übersetzung direkt beurteilt werden kann (vgl. Lo Presti 2016:25). Im Gegensatz zu bisherigen Evaluierungen von Qualität auf Basis von Fehlertypologien oder

Qualitätsdimensionen, kommt in dieser Arbeit die Methode des Best-Worst-Scalings (Louviere & Woodworth 1990) zum Einsatz. Hierbei wählen Proband*innen aus mehreren angegebenen Übersetzungen, hier auf Basis verschiedener Pre-Editing-Methoden, die beste und die schlechteste Übersetzung aus.

Die Motivation für diese Masterarbeit ist ihre Relevanz sowohl für Übersetzer*innen als auch für Lai*innen. Einerseits, was die Übersetzer*innen angeht, wäre eine verbesserte maschinelle Qualität mit einer Verringerung des Arbeitsaufwandes und mit beschleunigten Übersetzungsprozessen ohne Qualitätsverluste verbunden. Andererseits wäre das Pre-Editing eine praktische Anwendungsmöglichkeit für Lai*innen, die maschinelle Online-Übersetzungssysteme oft für eine schnelle und kostengünstige Übersetzung verwenden. In beiden Fällen könnte das Pre-Editing in maschinellen Übersetzungstools direkt integriert werden, insbesondere da MÜ als Plug-In für CAT-Tools schon verfügbar ist.

Die vorliegende wissenschaftliche Arbeit gliedert sich prinzipiell in theoretischen Rahmen und empirische Studie. Im theoretischen Rahmen werden im zweiten Kapitel die Grundlagen der Arbeit präsentiert. Genauer gesagt, werden in Kapitel 2.1. die Anfänge und die langjährige Evolution der maschinellen Übersetzung dargelegt. Die verschiedenen Formen der MÜ werden in Kapitel 2.1.1. vorgestellt. Das ganze Kapitel 2.1.2. ist der NMÜ gewidmet, es werden die Architekturen und Funktionsweisen der NMÜ erläutert sowie ihre Herausforderungen und über die am Markt verfügbaren unterschiedlichen NMÜ-Systeme. In Kapitel 2.2. werden die gängigen Standards und Metriken zur Evaluierung der Übersetzungsqualität vorgestellt, während in Kapitel 2.3. der Fokus auf die Evaluierung der Qualität maschineller Übersetzung liegen wird. Insbesondere werden an diesen Stellen die Kriterien bzw. Methoden zur menschlichen Evaluierung und die Ansätze zur Quality Estimation, also zur automatischen Evaluierung der MÜ und NMÜ erläutert. Als Kern sowie auch Start- und Endpunkt dieser Arbeit wird in Kapitel 2.4. der Begriff der MTranslatability erklärt, um im Folgenden (in Kapitel 2.5.) den damit eng verbundenen Terminus des Pre-Editing zu erklären. Nach der Beschreibung der meistverwendeten Evaluierungsmethoden der MÜ und NMÜ in den vorherigen Kapiteln (2.2. und 2.3.) werden in Kapitel 2.6. das Best-Worst-Scaling und dessen Anwendungsbereiche vorgestellt, um es als potenzielle neue Evaluierungsmethode im Bereich der Translationswissenschaft vorzuschlagen. Im dritten Kapitel werden die Studien und deren Ergebnisse präsentiert, die den aktuellen Stand der Forschung über Pre-Editing darstellen. Im 4. Kapitel wird die Forschungsmethodik der Arbeit beschrieben und deswegen gilt dieses Kapitel als Übergang vom theoretischen Rahmen zur

empirischen Studie deren Phasen untenstehend mithilfe eines Diagramms (siehe Abbildung 1, S. 5) gezeigt werden.

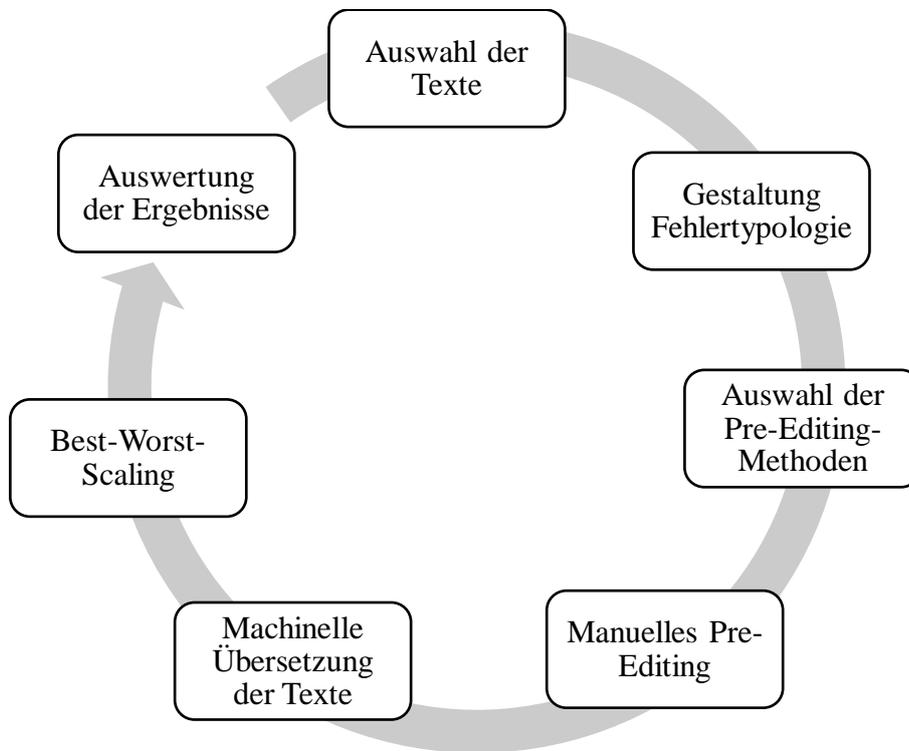


Abbildung 1: Phasen der empirischen Studie

Wie in Abbildung 1 zu sehen ist, wird die empirische Studie in folgende Phasen gegliedert: Auswahl der Ausgangstexte (Englisch), Gestaltung einer individuell angepassten Fehlertypologie und Auswahl von Pre-Editing-Methoden basierend auf den Fehlern in den Ausgangstexten, entsprechendes manuelles Pre-Editing, maschinelle Übersetzung ins Italienische durch ein NMÜ-System und menschliche Evaluierung durch Best-Worst-Scaling. Nach der Evaluierungsphase werden die Ergebnisse bzw. Daten im fünften Kapitel dargestellt und ausgewertet und in Kapitel 6. diskutiert. Zu guter Letzt werden die wichtigsten Schlussfolgerungen der Arbeit in Kapitel 7 gezogen.

2. Grundlagen

Auf den nächsten Seiten werden die Grundlagen dieser Masterarbeit beschrieben. Die Geschichte der maschinellen Übersetzung ist ein wichtiger Bestandteil der Forschung im Rahmen der MÜ. Die Fortschritte und das Scheitern in dem Bereich sind eng mit deren Entwicklungsgeschichte verbunden. Diese ist auch essenziell für künftige Entscheidungen, die die MÜ und deren Anwendungen anbelangen. Ausgangspunkt der Forschung sind die noch bestehenden Schwächen maschineller Übersetzung zu überwinden und die potenziellen Lösungen zu finden. Daher wird im Folgenden von den verschiedenen Formen maschineller Übersetzung und vor allem von den wichtigsten Aspekten neuronaler maschineller Übersetzung (NMÜ) die Rede sein, da sie aktuell im Fokus der Forscher*innen steht. Ein Vergleich zwischen den aktuellen NMÜ-Systemen wird folgen. Was heißt Qualität der Übersetzung? Wie unterscheidet sie sich von der Qualität maschineller Übersetzung? In den nächsten Kapiteln wird der Begriff genauer erörtert mit einer Erklärung der aktuell geltenden Qualitätsstandards. Im Rahmen dieser Arbeit spielt die Qualitätsevaluierung auch eine wichtige Rolle. Deswegen werden die gängigsten Evaluierungsmethoden beschrieben, bei denen prinzipiell zwischen menschlichen und automatischen Methoden differenziert wird. Nach der Beschreibung der Ansätze und Metriken der automatischen Qualitätsevaluierung werden die Begriffe von Pre-Editing, MTranslatability (Bernth & Gdaniec 2001) und Best-Worst-Scaling (Louviere & Woodworth 1990) eingeführt, da sie als wichtige Grundlagen dieser Masterarbeit dienen. Pre-Editing nimmt eine zentrale Rolle ein, da es zum Perspektivenwechsel im Bereich der Verbesserung der Qualität maschineller Übersetzung gehört. Der Begriff der MTranslatability wird als Basis für die Messung des Grades der Übersetzerbarkeit von Texten durch ein MÜ-System betrachtet. Die Methode des Best-Worst-Scalings als Bewertungsmethode und deren Anwendungsbereiche werden im Folgenden vorgestellt.

2.1. Maschinelle Übersetzung: Entwicklung

Die Entwicklungsgeschichte der maschinellen Übersetzung geht auf die Jahre nach dem Zweiten Weltkrieg zurück, wie Stein (2009) erklärt:

Im Zweiten Weltkrieg spielte die Dechiffrierung feindlicher Funksprüche eine wichtige Rolle. Für das Knacken des Codes der deutschen ENIGMA war in erster Linie das britische Team um Alan Turing in Bletchley Park verantwortlich. Mittels statistischer Methoden, ausgewertet von auf Relais basierenden Rechenmaschinen, legten die Wissenschaftler hier, ohne es zu wissen, den Grundstein für die praktische MÜ. Auf den in Bletchley Park gewonnenen Erfahrungen aufbauend führten

Warren Weaver und Andrew Booth einen Briefwechsel, der als Geburtsstunde der MÜ gilt. (Stein 2009:6)

Einen wichtiger Entwicklungsstartpunkt maschineller Übersetzung bot Warren Weaver (1949). In seinem Memorandum schlug er vor, die modernen elektronischen Computer und ihre Geschwindigkeit, Kapazität und logische Flexibilität im Bereich der Übersetzung anzuwenden (1949:4). Obwohl sich auch Andrew Booth mit dem mechanisierten Nachschlagen eines Wortes befasste, wurde immer noch über eine Wort-für-Wort Übersetzung berichtet: Die Schwierigkeiten, die mehrdeutige Wörter, Kontext und Wortstellung bereiteten, wurden nicht behandelt. Aus diesem Grund wurden von Weaver einige Lösungen vorgeschlagen, und zwar vor allem eine N Anzahl von Wörtern, die sich gleich links oder rechts des betrachteten Wortes befinden, mitzurechnen, um Ambiguität zu reduzieren (1949:6-9). Nach Weaver (1949) war also eine statistische semantische Studie zunächst durchzuführen, wobei die Kryptografie auch eine maßgebliche Rolle spielte. Von wesentlicher Bedeutung war die Struktur von allen Sprachen tiefergehend zu analysieren, um deren Gemeinsamkeiten festzustellen (vgl. Weaver 1949:11). Die Forschungsarbeit Weavers gelang im Jahr 1954 ein Durchbruch in Bezug auf die maschinelle Übersetzung. IBM und die Georgetown University in New York kooperierten und verwirklichten ein MÜ-System in der Sprachkombination Russisch-Englisch (vgl. Hutchins 2000:6).

Die Forschung in diesem Bereich verbreitete sich nicht nur in den USA, sondern global. Nach einer gewissen Pionierzeit wurde der anfängliche Fokus auf die wörterbuchbasierten Systeme auf die Analyse und Umwandlung der syntaktischen Strukturen gerichtet (vgl. Hutchins 2000:2). Der erste Forscher, der sich mit syntaktischen Strukturen beschäftigte, war Noam Chomsky (1957) mit seiner Theorie der Unabhängigkeit zwischen Syntax und Semantik: „He provided a different theory of linguistic structure, stated in a quasi-mathematical notation that fit well with the emerging field of language engineering” (Madsen 2009:23). Chomsky und seine generative Transformationsgrammatik stellen einen Meilenstein im Bereich der Linguistik dar und haben die Forschung im Bereich der maschinellen Übersetzung beeinflusst (2009:28). Zu diesen innovativen Zeiten gehörte zudem die Entwicklung des SERNA-Systems für die maschinelle Übersetzung aus dem Russischen ins Englische. Im Jahr 1959 wurde das System im Pentagon getestet (vgl. Toma 1977:570). Trotz der stetigen Forschung im Bereich maschineller Übersetzung fand keine Verbesserung statt, da die Schwierigkeiten noch bei der Semantik lagen (vgl. Hutchins 2000:9). Ferner verbreitete Bar-Hillel (1960) Skepsis und Pessimismus gegenüber der MÜ, da er schlussfolgerte, dass eine komplett automatische und

zugleich qualitativ hochwertige Übersetzung wegen der semantischen Ambiguitäten noch unerreichbar war.

Ein wichtiger Wendepunkt wurde im Jahr 1966 mit dem Automatic Language Processing Advisory Committee Bericht (ALPAC 1966) erreicht. Der Bericht wurde im Jahr 1964 von dem US-Verteidigungsministerium, dem National Science Foundation und dem Central Intelligence Agency (CIA) verlangt, die unter dem Joint Automatic Language Processing Group (JALPG) zusammengestellt wurden. Zweck des Berichts war es in dem Bereich der automatischen Sprachverarbeitung bzw. Übersetzung zu forschen, um die Kosten konsequent zu reduzieren und die Resultate zu verbessern. Der Bericht zeigte, dass Qualität, Geschwindigkeit und Kosten die Kernbereiche der Übersetzung darstellten (1966:16ff.). Vor allem was die Qualität angeht, wurde dem Bericht ein Appendix hinzugefügt, in dem die Resultate einer von John B. Carroll durchgeführten empirischen Studie präsentiert wurden. Bei der empirischen Studie wurden die Faktoren „Verständlichkeit“ und „Treue“ mithilfe von einer 9-Punkte-Skala menschlich evaluiert. Die Ergebnisse zeigten, dass es die Möglichkeit gab, die Qualität sowohl von humanen als auch von maschinellen Übersetzungen zuverlässig zu bewerten. Allerdings fehlte noch ein System, das eine beliebige Übersetzung einfach und verlässlich evaluieren vermochte (1966:67-75). Die Schlussfolgerung des ALPAC-Berichts über die Zukunft der MÜ-Forschung war negativ. Trotz des erwähnten Verbesserungspotenzials der Übersetzungsqualität wurde die damalige maschinelle Übersetzung ohne unmittelbare künftige Nützlichkeit eingeschätzt (1966:32). Laut Hutchins (1996) verurteilte der ALPAC-Bericht die maschinelle Übersetzung für die darauffolgenden zehn Jahre. In den USA wurde die MÜ vom breiten Publikum und von der wissenschaftlichen Gemeinschaft als ein Misserfolg betrachtet. Kenny (2018) zufolge ist der ALPAC-Bericht noch heute von Interesse, da damit der erste Evaluierungsversuch der maschinellen Übersetzung eingeführt und einige Faktoren angegeben wurden, die diese Bewertung insgesamt beeinflusst haben – darunter die damalige große Verfügbarkeit von (humanen) Übersetzer*innen und die niedrigen Kosten ihrer Dienstleistungen (2018:431).

Trotz der Ernüchterung der Post-ALPAC-Zeit, ging die Entwicklung maschineller Übersetzung weiter bis zu einem neuen Aufschwung ab den 1970er Jahren (vgl. Hutchins 2000:12). Diesbezüglich gilt es zu erwähnen, dass in den Folgezeiten des ALPAC-Berichts ein noch heute wichtiges MÜ-System entwickelt wurde. Genauer gesagt, wurde das SYSTRAN-System im Jahr 1968 von Peter Toma gegründet. Nach Toma (1977) stellte SYSTRAN der dritten Generation der Nachkomme des SERNA-Systems aus dem Jahre 1959 dar (1977:570). Was den Verlauf des universellen SYSTRAN-Systems betraf, ließen sich folgende positive

Aspekte feststellen: die Fähigkeit, ganze Konkordanz-Korpora zu erstellen, die wichtig auf syntaktischer und semantischer Ebene sind. Zudem die Fähigkeit, Wörter in einem universellen Wörterbuch nachzuschlagen (vgl. Toma 1977:573). Außerdem ermöglichte das System die simultane Übersetzung in mehrere Sprachen und konnte Mehrdeutigkeitsprobleme durch Disambiguierung der Ausgangssprache auf verschiedenen Ebenen lösen. Dieses System konnte sogar mit Weltwissen, semantischen und pragmatischen Informationen erweitert werden (1977:576-580). Seit den 1970er Jahren wurde SYSTRAN von der amerikanischen Regierung verwendet und seit dem Jahr 1976 bis 2010 von der Europäischen Kommission² (vgl. Kenny 2018:432). Seit dem Jahr 1996 ist SYSTRAN auch online verfügbar (vgl. Yang & Lange 1998:276). Dieses System zählt zu den Grundlagen dieser Masterarbeit, da es im empirischen Teil verwendet wird, um die Tendenzen bezüglich der Forschungsfrage der Arbeit aufzuzeigen.

Zwischen den 1980er und den 2000er Jahren folgte die Entwicklung neuer Technologien und Systeme – beeinflusst von der Einführung des Heimcomputers. Die breitere technologische Umgebung ab den 1990er Jahren ermöglichte die Forschung im Bereich der MÜ wieder. Im Übrigen stand die MÜ mit der Entwicklung des Internets plötzlich für Millionen Benutzer*innen zur Verfügung (vgl. Kenny 2018:433). Seit Mitte der 1990er Jahre wurde die maschinelle Übersetzung zu einem kostenlosen, online verfügbaren Massenprodukt, jedoch aus professioneller Sicht nicht gleich interessant. Die zwei Hauptgründe dafür waren, dass die Qualität maschineller Resultate unzureichend war für einen Post-Editing-Einsatz und dass verschiedene Translation Memorys als Unterstützung für Übersetzer*innen schon auf dem Markt zu finden waren. Nichtsdestotrotz wurden Translation-Memory- und maschinelle Übersetzungssysteme in einer Umgebung eingesetzt, wo beide Technologien für die Verbesserung der Übersetzungsqualität von Nutzen waren (2018:434).

Ab den 2000er Jahren stellte die neu entwickelte statistische maschinelle Übersetzung (SMÜ) den Stand der Forschung dar, bis zum Jahr 2016, als die statistischen Systeme durch die neuronalen Systeme ersetzt wurden (vgl. Kenny 2018:434; vgl. Bentivogli et al. 2016:o. S.). Die neuronale maschinelle Übersetzung stellt den gegenwärtigen Stand der Forschung in dem Bereich der MÜ dar und es wird kontinuierlich weitergeforscht.

² Wagner (2014) gibt einen Überblick über das Übersetzen für die Europäische Union und den Einsatz von SYSTRAN.

2.1.1. Verschiedene Formen von MÜ

Stein (2009) und Werthmann und Witt (2014) haben einen Überblick über die verschiedenen Ansätze zur MÜ gegeben, wobei der markanteste Unterschied in der Differenzierung zwischen regelbasierten und statischen MÜ-Systemen liegt. Der Grund der Entwicklung unterschiedlicher Ansätze besteht darin, dass sich „die MÜ an der Schnittstelle von Linguistik, Computerlinguistik, Übersetzungswissenschaften und Informatik“ befindet (Werthmann & Witt 2014:86).

Der erste Ansatz zur MÜ ist **regelbasiert**, bei dem die unterschiedlichen linguistischen Ebenen einer Ausgangssprache analysiert und repräsentiert werden; auf dieser Repräsentation der Sprache werden nachher Übersetzungsregeln angewandt. Zu diesem Ansatz gehören sowohl die direkte, im Sinne einer Wort-für-Wort-Übersetzung, als auch die indirekte Übersetzung, welche in Transfer und Interlingua unterteilt wird. Diese letzten zwei Strategien sind komplexer als die direkte Übersetzung. Im dreiphasigen Prozess des Transfers (Analyse, Transfer und Generierung) wird die Satzebene mit deren syntaktischen Strukturen betrachtet, während bei der Interlingua-Übersetzung nur Analyse und Generierung stattfinden. Die Repräsentationen beim Transfer sind sprachspezifisch, die Zwischenrepräsentation der Interlingua-Übersetzung hingegen sollte unabhängig von der Ausgangssprache erfolgen und zur Generierung unterschiedlicher Zielsprachen dienen (vgl. Werthmann & Witt 2014:87-91). Allerdings ist bis dato noch keine universale und sprachunabhängige Repräsentation für diesen letzten Ansatz zu finden (vgl. Stein 2009:8). Die Qualität der Übersetzung bei den jeweiligen Ansätzen ist sehr unterschiedlich, allerdings ist die Komplexität des Ansatzes kein Beweis der höheren Qualität der Resultate (2009:8). Einen speziellen Fall des Interlingua-Ansatzes (vgl. Stein 2009:14) für beschränkte Fachgebiete stellt die **wissensbasierte** MÜ dar, wobei das System nicht nur linguistische Regeln dekodieren, sondern auch „[...] das allgemeine Weltwissen über einzelne Dinge, Sachverhalte, Ereignisse, Vorgehensweisen [...] aus Enzyklopädien, Ontologien, Semantischen Netzen, Thesauri etc.“ erschließen sollte (Werthmann & Witt 2014:99).

Im Vergleich zur regelbasierten MÜ, die auf der aufwändigen Erstellung von bilingualen Wörterbüchern und Regelsets basiert, extrahiert die **korpusbasierte** maschinelle Übersetzung Informationen aus zwei- oder mehrsprachigen manuell oder automatisch alignierten Korpora. Eine Alignierung ermöglicht die Markierung von den inhaltlichen Entsprechungen zwischen Korpora auf lexikalischer und syntaktischer Ebene. Die korpusbasierten Ansätze werden in reinstatistische, beispiel- und kontextbasierte MÜ

klassifiziert (vgl. Werthmann & Witt 2014:91f.). Bei der reinen **statistischen** Übersetzung (SMÜ) werden die parallelen Korpora verwendet, um die statistische Wahrscheinlichkeit der Übersetzung eines Ausgangsterms (eines Wortes, einer Phrase oder eines Satzes je nach Ebene) in einen bestimmten Zielterm zu berechnen. Das Ergebnis der Wahrscheinlichkeitsberechnungen ist dann ein selbstlernendes Übersetzungsmodell mit allen potenziellen zielsprachlichen Übersetzungen und deren entsprechenden Wahrscheinlichkeitsgraden (2014:92f.). Zuerst erfolgt eine Decodierung der zielsprachlichen Wahrscheinlichkeiten mittels des Übersetzungsmodells, dann wird ein Sprachmodell der Zielsprache benötigt, das auf einsprachigen Korpora basiert, um kontextuell passende Übersetzungsalternativen weiter zu decodieren und zu ordnen und grammatikalisch korrekte Resultate zu erhalten. Bei der SMÜ wird zwischen wortbasierter (betrachtet die Wortebene) und phrasenbasierter (betrachtet die Phrasenebene) maschineller Übersetzung unterschieden. Im Vergleich zu dem regelbasierten ist der statistische Ansatz kostengünstiger und nicht so zeitaufwändig; außerdem werden weder linguistisches Wissen noch sprachspezifische Grammatikregeln dafür benötigt. Diese Systeme können einfach erweitert werden, obwohl dafür eine große Menge geeigneter und zuverlässiger Korpora bereitzustellen ist (2014:95f.). Als Nachteil erzielt die SMÜ noch keine befriedigenden Ergebnisse mit Sprachkombinationen, die eine sehr unterschiedliche Struktur haben (vgl. Stein 2009:13). Die **beispielbasierte** MÜ basiert auf dem Konzept des Translation-Memory-Systems (TMS). Dabei handelt es sich um die Wiederverwendung von ähnlichen oder gleichen Segmenten aus Inhalten, die schon übersetzt worden waren, mit dem Ziel, den Übersetzungsprozess zu automatisieren. Dieser Ansatz weist eine höhere Qualität der Übersetzung auf, da die Texte für die parallelen Korpora von Fachübersetzer*innen übersetzt worden sind. Hochwertige Resultate können gewährleistet werden, sofern die Ausgangs- und Zieltextdaten auf die gleiche Domäne zurückgehen. Bei diesem System werden nützliche zusammengehörende ausgangs- und zielsprachliche Beispiele aus dem Korpus extrahiert und in einer Datenbank gesammelt, um folglich abgerufen und wiederverwendet zu werden. Unter den gespeicherten Sätzen und Redewendungen mit entsprechenden Übersetzungen wird nach den besten Matches gesucht, um sie zusammenzufügen und möglichst korrekte Zieltexte zu erhalten (vgl. Werthmann & Witt 2014:96f.). Im Verhältnis zum statistischen und beispielbasierten Ansatz stützt sich ein **kontextbasierter** Ansatz auf zielsprachlichen einsprachigen Korpora und auf einem bilingualen Vollformenwörterbuch. Das heißt, dass die Qualität der Übersetzung steigt, je umfangreicher diese zwei Komponenten sind. Bei diesem Ansatz werden verschiedene Übersetzungsvarianten aussortiert, mit dem Ziel, potenzielle Ambiguitäten im Text zu

vermeiden und eine Übersetzung im Kontext zu ermöglichen (vgl. Werthmann & Witt 2014:98).

Der Forschungsprozess im Bereich Maschinelle Übersetzung erweitert sich kontinuierlich und aktuell liegt ein neuer korpusbasierter Ansatz im Fokus der Forscher*innen: die kürzlich entwickelte neuronale maschinelle Übersetzung (NMÜ). Dieser Ansatz dient als Grundlage für diese Masterarbeit und wird dementsprechend im nächsten Kapitel 2.1.2. genauer erläutert.

2.1.2. Neuronale Maschinelle Übersetzung

Wie bereits erwähnt, stützt sich diese Arbeit auf die neuronale maschinelle Übersetzung, die in den letzten Jahren versucht hat, den bisherigen statistischen Ansatz zu ersetzen. Dabei handelt es sich um einen neuen korpusbasierten Ansatz, der mit umfangreichen Korpora aus Ausgangssprachlichen und Zielsprachlichen Segmentenpaaren trainiert wird. Wie der Name suggeriert basiert der Ansatz auf künstlichen neuronalen Netzen. Sie bestehen aus tausenden vernetzten, künstlichen Einheiten, die gleich wie menschliche Neuronen funktionieren, indem ihre Aktivierung von den Impulsen anderer Neuronen und der Festigkeit ihrer Verbindung abhängt (vgl. Forcada 2017:291ff.). Wichtige Elemente der neuronalen Netze sind daher ihre nicht lineare Aktivierungsfunktion und die Verwendung von mehreren Schichten im Trainingsprozess, darunter eine versteckte Schicht (engl. *hidden layer*). Die Bezeichnung „Deep Learning“ basiert sogar auf der Möglichkeit, mehrere versteckte Schichten bei neuronalen Netzen aufzustapeln, um eine bessere Leistung zu erzielen (vgl. Koehn 2017:8ff.). Dementsprechend wird bei der NMÜ die Verbindung zwischen einem Quell- bzw. Eingabetext in einer Ausgangssprache und dessen Zielsprachlichen Übersetzung direkt aufgebaut. Die NMÜ verfügt über die Fähigkeit, das syntaktische und semantische Wissen zu lernen, indem der globale Satzkontext für die Übersetzung betrachtet wird (vgl. Crego et al. 2016:o.S.).

Dieser neue Ansatz wird häufig mit dem statistischen verglichen, sowohl was die Stärken und Schwächen beider Ansätze als auch was die Ähnlichkeiten und Unterschiede in ihren Funktionsweisen betrifft. Im Rahmen neuronaler maschineller Übersetzung werden für alle Sprachpaare sowohl vollständige Sätze als auch einzelne Wörter mit ihren grammatikalischen oder inhaltlichen Abhängigkeiten im Satz betrachtet. Die Ergebnisse zeigen natürlichere Satzstrukturen und weniger Fehler im Vergleich zur statistischen maschinellen Übersetzung in Bezug auf die Syntax. Trotz der Unterschiede bezüglich Wortreihenfolge in verschiedenen Sprachen, bedarf die NMÜ keiner zusätzlichen Sprachmodelle (vgl. Schmalz

2019:198f.). Bentivogli et al. (2016) heben auch hervor, dass linguistisch betrachtet die NMÜ weniger morphologische und lexikalische Fehler als die phrasenbasierte SMÜ aufzeigt. Des Weiteren haben Burchardt et al. (2017) in ihrer linguistischen Evaluierung der verschiedenen regelbasierten, phrasenbasierten und neuronalen MÜ-Systeme beobachtet, dass die online verfügbaren Systeme eine deutlich sichtbare Verbesserung beim Übergang von phrasenbasierten zu neuronalen Systemen gezeigt haben (2017:169). Auf diesen Vergleich wird hier aber nicht näher eingegangen, hierzu werden weitere Lektüren³ empfohlen. Der Fokus dieses Kapitels ist eine Vorstellung der NMÜ und deren unterschiedlichen Architekturen und Funktionsweisen. In einem weiteren Schritt folgt eine Kurzbeschreibung hybrider Ansätze sowie der Herausforderungen in der NMÜ. In Kapitel 2.3. wird ein Überblick der Qualitätsevaluierung maschineller Übersetzung, insbesondere der NMÜ gegeben.

2.1.2.1. Architekturen und Funktionsweisen

Im Rahmen neuronaler maschineller Übersetzung werden verschiedene Arten von neuronalen Netzen bzw. Architekturen aufgebaut, wie sich aus den nachfolgenden Beispielen erschließen lässt: das Encoder-Decoder-Modell (oder auch engl. *sequence-to-sequence model*), die Konvolutionsnetze (engl. *Convolutional Neural Networks (CNN)*, LeCun 1989) und das Transformer-Modell (mit einem Attention-Mechanismus, Vaswani et al. 2017) (vgl. Koehn 2017:6). Im Kontext der NMÜ wird der Ausgangstext in Wörter oder in einer Serie von Zeichensequenzen (auch Byte-Pair-Encodings (BPE)⁴, Gage 1994) umgewandelt, die anschließend parallel verteilt verarbeitet werden. Die Aktivierung jedes Sets von Neuronen wird so trainiert, dass „verteilte Repräsentationen“ (engl. *distributed representations*) von Wörtern und deren Kontexten (in dem Quell- und Zielsatz) gelernt werden. Eine Repräsentation stellt den Aktivierungsstatus jedes Neurons innerhalb einer bestimmten Gruppe von Neuronen dar, das heißt innerhalb einer Schicht (engl. *layer*). Die Übersetzung resultiert dann aus diesen Vektorrepräsentationen (vgl. Forcada 2017:293f.).

Das Training hat das Ziel, dass sich die maschinellen Übersetzungen möglichst den Referenz-Übersetzungen oder dem goldenen Standard menschlicher professioneller Übersetzungen im Trainingsdatensatz annähern (vgl. Forcada 2017:295). In der Trainingsphase werden Sequenzen – Ein- und Ausgabesequenzen – so lange wiederholt bis das Netz alle Sequenzen ausreichend beherrscht und gute Gesamtergebnisse erzielt. Im Folgenden kann das

³ In den Forschungsarbeiten von Luong (2017), Forcada (2017) und Schmalz (2019) sind ausführliche Vergleiche zwischen statistischer und neuronaler MÜ zu finden.

⁴ In 2.1.2.2. kurz erklärt.

Modell in der Testphase aufgrund der gelernten Eigenschaften auf zuvor unbekannte Sequenzen verallgemeinern. Aus diesem Grund benötigt die NMÜ sehr umfangreiche Trainingskorpora. Daher wird eine spezifische aus Grafikprozessoren (engl. *graphic processing unit*, GPU) entwickelte Hardware mit langen Trainingszeiten (von Tagen bis zu Monaten) verwendet (vgl. Forcada 2017:295f.). Bei der NMÜ wird also automatisch von den verteilten Repräsentationen der einzelnen Wörter und der Word-Embeddings „gelernt“: Spezielle neuronale Netze lernen daher, ein bestimmtes Wort innerhalb eines spezifischen Kontexts unter Berücksichtigung der links- und rechtsstehenden Wörtersequenzen wiederzugeben oder die links- und rechtsstehenden Wörtersequenzen eines bestimmten Wortes zu prognostizieren (vgl. Forcada 2017:296f.). In einem NMÜ-System erfolgt die Übersetzung durch einen Encoder und einen Decoder, die das **Encoder-Decoder-Modell** bilden. Sowohl Encoder als auch Decoder kann aus jeglicher NN-Architektur bestehen, zum Beispiel CNNs oder Transformer. Der Encoder kodiert (komprimiert) die Eingabesequenz. Ein gut trainierter Decoder weist dann ein Wort mit Maximum-Likelihood durch einen Wahrscheinlichkeitsvektor einer Position zu, in mehreren Schichten und wiederholt dies bis zur kompletten Übersetzung des Satzes (2017:297f.). Der Decoder betrachtet also nicht nur die letzte Repräsentation, sondern auch die ganze Repräsentationssequenz vom Encoder, mithilfe eines zusätzlichen Sets an neuronalen Verbindungen und Schichten (2017:299). Diese Architektur ist daher die direkte Erweiterung neuronaler Sprachmodelle (vgl. Koehn 2017:47).

Die Verwendung von rekursiven neuronalen Netzen weist einige Nachteile auf, vor allem bei dem Encoding, weil sie einen langen sequenziellen Prozess benötigen. Außerdem wird eine parallele und gleichzeitige Verarbeitung aller Wörter verhindert und deshalb die Verwendung der GPU-Kapazitäten eingeschränkt. Diesbezüglich haben sich kürzlich die **Konvolutionsnetze** (CNN) verbreitet, die schon für die maschinelle Bildverarbeitung erfolgreich angewandt wurden (vgl. Koehn 2017:83). Der Hauptunterschied zum bisherigen Modell liegt in der Verwendung von Konvolutionsschichten (engl. *convolutional layers*) anstatt rekursiver neuronaler Netze. Eine Konvolution bzw. eine Faltung verbindet die Repräsentation einer Eingabesequenz limitierter Kontextwörter in einer einzelnen Repräsentation mithilfe einer Matrix. Die Wiederholung dieses Prozesses führt zu einer Satzrepräsentation in einem einzelnen Vektor, das heißt, zur parallelen Wörterverarbeitung. Auch bei dieser Architektur wird ein Attention-Mechanismus eingesetzt (2017:83-86).

Das **Transformer-Modell** stellt die aktuelle NMÜ-Architektur dar, indem es das erste Modell ist, das sich völlig auf den Self-Attention-Mechanismus stützt, um die Repräsentationen der Ein- und Ausgabe zu berechnen, ohne Verwendung der sequenzalignierten rekursiven

neuronalen Netze oder der Konvolution. Dieses Modell besteht aus aufgestapelten Self-Attention-Schichten sowohl bei dem Encoder als auch bei dem Decoder (vgl. Vaswani et al. 2017:2). Das Training dieses Modells erfolgt schneller, da die Berechnungskomplexität jeder Self-Attention-Schicht sowie der Berechnungsumfang sehr niedrig sind. Außerdem geht der Self-Attention-Mechanismus besser mit Langzeitabhängigkeiten (engl. *long-range dependencies*) im Netz um (2017:6,10). Dieser Mechanismus bietet die Möglichkeit, jede Eingabewortrepräsentation mit Kontextwörtern zu verfeinern, die zur Disambiguierung dienen (vgl. Koehn 2017:87).

Belinkov et al. (2017) haben sich mit einigen wichtigen Fragen beschäftigt, um die NMÜ-Architektur besser zu verstehen, da den Forscher*innen noch unklar bleibt, was die NMÜ-Modelle über die Ausgangs- und Zielsprache im Laufe des Trainingsprozesses eigentlich lernen. Was lernen die Modelle über Morphologie? Welche Wirkung haben die unterschiedlichen Repräsentationen (und zwar auf Zeichen- oder Wortebene) auf das tatsächliche Lernen? Was wird über die syntaktische und semantische Struktur einer Sprache durch die verschiedenen Schichten gelernt? (2017:861) Als Schlussfolgerungen der Studie und Evaluierung der Repräsentationsqualität im Bereich der NMÜ von Belinkov et al. (2017) ist Folgendes zu bemerken: Die Repräsentationen der linguistischen Einheiten durch die neuronalen Netze sind von entscheidender Bedeutung, um qualitativ hochwertige Übersetzungen zu bekommen (auf die Definition von Qualität wird in Kapitel 2.2. näher eingegangen). Zudem, was das Erlernen der Morphologie angeht, sind die Repräsentationen auf Zeichenebene geeigneter vor allem in Bezug auf seltene oder unbekannte Wörter. Die Morphologie wird in den niedrigeren Schichten des neuronalen Netzes gelernt, während die Übersetzungsleistung eher in den tieferen Netzen verbessert wird (2017:868f.). Ein besseres Verständnis der Herausforderungen, mit denen sich die NMÜ-Forscher*innen konfrontiert sehen, dient zur Entwicklung und letztendlich zur Verbesserung der NMÜ-Resultate.

2.1.2.2. Aktuelle Herausforderungen der NMÜ

Die oben erwähnten Herausforderungen beziehen sich meistens auf die Trainingsphase, die ein ausschlaggebender Teil der NMÜ darstellt. Diesbezüglich ist das Verhalten der NMÜ-Modelle unbeständig, wenn sie unter Bedingungen arbeiten, die stark von den Trainingsbedingungen abweichen (vgl. Koehn 2017:90).

Eine erste Herausforderung stellen die **Domänenadaption** und ihre Implementierung dar, da sie als Kern der MÜ-Systementwicklung für spezifische Anwendungsfälle gelten.

Tatsächlich können Wörter in verschiedenen Domänen⁵ unterschiedlich übersetzt und ihre Bedeutung kann in unterschiedlichen Stilen ausgedrückt werden. Dies führt daher zu unpassenden Übersetzungen (engl. *domain mismatch*). Die Trainingsdaten sind oft nur domänenfremd (engl. *out-of-domain*) verfügbar und die Qualität der Ausgabe ist niedriger im Vergleich zur Verwendung domänenspezifischer (engl. *in-domain*) Trainingsdaten. Hierzu sind die Ausgaben zwar fließend, aber sie hängen nicht mit der Eingabe zusammen (vgl. Koehn 2017:91ff.).

Die zweite Herausforderung ist die **Trainingsdatenmenge**. Die NMÜ hat eine steilere Lernkurve und kann mit einer größeren bzw. zunehmenden Trainingsdatenmenge gut arbeiten. Allerdings sind „beschränkte“ Trainingskorpora problematisch, die nur über wenige Millionen Wörter verfügen (vgl. Koehn 2017:95). Koehn und Knowles (2017) haben festgestellt, dass die Systeme noch keine befriedigenden Resultate in Bezug auf ressourcenarme Szenarien bzw. Sprachen liefern im Vergleich zur besseren Qualität in ressourcenreichen Szenarien.

Die neuronalen maschinellen Übersetzungssysteme verfügen zwar über ein kleineres Vokabular (zum Beispiel hinsichtlich SMÜ-Systeme), jedoch erzielen sie bessere Resultate bei der Übersetzung von sehr seltenen oder unbekanntem Wörtern (engl. *out-of-vocabulary words*, OOV), weil sie mit dem Byte-Pair-Encoding arbeiten (vgl. Koehn & Knowles 2017: o. S.). Wie von Sennrich et al. (2016) eingeführt, werden seltene und unbekannte Wörter mit dem Byte-Pair-Encoding als Sequenzen von kleineren Einheiten dargestellt, also wird das Byte-Pair-Encoding für die Wortsegmentierung verwendet. Nichtsdestotrotz stellt die **Übersetzung von seltenen stark flektierten Wörtern**, wie zum Beispiel Adjektive und Verben, eine Herausforderung für die NMÜ dar (vgl. Koehn & Knowles 2017: o. S.).

Die **Übersetzung von längeren Sätzen**, schon mit mehr als 60 Wörtern, ergibt keine guten Resultate in Bezug auf die Qualität. Längere Sätze, zum Beispiel 80 oder mehr Wörter, ergeben noch kürzere Resultate, die eine noch niedrigere Qualität aufweisen (vgl. Koehn & Knowles 2017 o. S.).

Zu den Herausforderungen der NMÜ gehören auch das **Wort-Alignment-Modell** und das Beam-Search-Decoding. Der Attention-Mechanismus ist tatsächlich eine Herausforderung, da er nicht in allen Fällen die Rolle eines Wort-Alignment-Modells übernimmt, sondern ihre Funktionen weichen in der NMÜ voneinander ab (vgl. Koehn & Knowles 2017:o. S.). Das **Beam-Search-Decoding** für das Prognostizieren der Ausgabewörter bzw. des ganzen

⁵ Die übliche Definition von Domäne in der MÜ lautet: „a *domain* is defined by a corpus from a specific source, and may differ from other *domains* in topic, genre, style, level of formality, etc.“ (Koehn 2017:91, Kursivschrift im Original)

Ausgabesatzes je nach Wahrscheinlichkeit funktioniert am besten mit einer begrenzten Anzahl von Ausgabewörtern (engl. *beam*); eine größere Anzahl verschlechtert die Übersetzungsergebnisse (vgl. Koehn & Knowles 2017:o. S.; vgl. Koehn 2017:54-57).

Die hier aufgelisteten Herausforderungen beziehen sich vor allem auf die Entwicklung der NMÜ und derer Systeme bzw. Softwares, haben aber einen direkten Einfluss auf die Qualität maschineller Resultate. Zudem wirken sich diese Herausforderungen auf die Arbeit professioneller Übersetzer*innen aus. Nicht nur benötigt die NMÜ ein langwieriges Training und eine spezifische Hardware (wie GPU), sondern sind auch die verfügbaren Toolkits oft schwer zu installieren, konfigurieren und benutzen, da Übersetzer*innen Ad-Hoc-Kompetenzen und eine adäquate Hardware brauchen (vgl. Forcada 2017:302). Miteinbezogen werden aber auch die Post-Editor*innen, weil die semantische Natur der bei der NMÜ gelernten Repräsentationen zu einigen Fehlern führt, die in der Post-Editing-Phase bearbeitet werden müssen. Dazu gibt es terminologische Inkonsistenzen, Fehlübersetzungen oder falsch angebrachte Formatierungstags. Diesbezüglich zählt es zu den Aufgaben der Post-Editor*innen, Lösungen für die neuen „kreativen“ maschinellen Resultate zu finden. Die Systeme, die auf Zeichenebene arbeiten, schaffen tatsächlich Resultate, die entweder akzeptablen Übersetzungen ähnlich sind, oder sie erfinden oder erzeugen teilweise übersetzte Wörter (2017:303f.).

2.1.2.3. Verfügbare NMÜ-Systeme: ein Vergleich

Heutzutage sind verschiedene NMÜ-Systeme am Markt sowie online verfügbar, die eine Dienstleistung anbieten, indem sie den Internetbenutzer*innen eine Zugangsmöglichkeit zum Wissen ermöglichen oder erleichtern, zum Beispiel weil die Informationen oft auf Englisch zu finden sind (vgl. Schmalz 2019:197). Im Folgenden werden drei aktuell verfügbare Übersetzungssysteme verglichen, die auf verschiedenen neuronalen Architekturen basieren und deren Übersetzungsqualität kontinuierlich von den Forscher*innen hinterfragt wird⁶: Google Translate (Google Translate 2020), DeepL (DeepL Translator 2020) und SYSTRAN Translate (SYSTRAN Translate Tool 2020).

Google ist hinsichtlich neuronaler Systeme Pionier und hat *Google's Neural Machine Translation* (GNMT)⁷ angeboten, mit dem Ziel, die Unterschiede zwischen Human- und statistischer maschineller Übersetzung auszugleichen (vgl. Wu et al. 2016:1). Dementsprechend kommen bei dieser neuronalen Version von **Google Translate** die ersten

⁶ Die Herausforderungen wurden in 2.1.2.2. detaillierter beschrieben.

⁷ Ab hier nur mehr als Google Translate bezeichnet.

Verbesserungen in Bezug auf Trainingszeiten, Umgang mit seltenen und unübersetzten Wörtern unter Verwendung eines Encoder-Decoder-Modells mit einem Attention-Mechanismus zur Umsetzung. Das neuronale Google Translate erzielt im Vergleich zu einem phrasenbasierten System einen höheren Genauigkeitsgrad und die Übersetzungsfehler werden für viele übliche Sprachpaare um 60% reduziert (2016:1f., 20). Nach Turovsky (2016) ist es bei diesem System möglich ganze Sätze, statt kleineren Einheiten gleichzeitig zu übersetzen. Mithilfe von dem Satzkontext kann das System die relevanteste Übersetzung schneller finden und die Grammatik anpassen, damit der Satz natürlich wie eine menschliche Übersetzung klingt. Heute bietet Google Translate Übersetzungen in 103 Sprachen und unzählige Sprachkombinationen an (vgl. Tavosanis 2019: o.S.; vgl. Google Translate Languages 2020). Bei diesem System besteht sogar die Möglichkeit, eine automatische Übersetzung praktisch in allen digitalen Umgebungen wie Apps und Internetseiten zu integrieren (vgl. Schmalz 2019:206). Die große Menge an mehrsprachigen Daten, mit denen Google Translate trainiert wird, stellt einen Vorteil für das System dar. Allerdings ist sie keine Garantie der Qualität der Daten selbst, da sie nicht verifiziert werden können (2019:203).

Wie von Schmalz (2019:199f.) hervorgehoben, ist nach kurzer Zeit ein weiteres System, **DeepL**, entwickelt worden. Es wurde von einem deutschen Unternehmen eingeführt und als Konkurrent von Google Translate betrachtet. Ein Grund dafür ist, dass die Firma bereits eine Plattform für die Suche nach Übersetzungspaaren in verschiedenen Sprachkombinationen unter dem Namen Linguee angeboten hat. Dementsprechend steht DeepL ausreichend qualitativ hochwertiges Material (Textfundus) für das Training des Algorithmus zur Verfügung. Außerdem basiert das System nicht auf rekursiven neuronalen Netzen, sondern auf Konvolutionsnetzen, die die simultane parallele Übersetzung aller Wörter ermöglichen (vgl. Merkert 2017). Der Textfundus besteht im Regelfall aus Dokumenten der UNO oder EU, die parallel in vielen Sprachen professionell übersetzt worden sind und deshalb mit einer höheren Übersetzungsqualität einhergehen, insbesondere wenn das System für bestimmte Domänen trainiert wird (vgl. Schmalz 2019:203). Im Vergleich zu Google Translate deckt DeepL nur neun Sprachen ab (vgl. Tavosanis 2019: o. S.).

Die NMÜ-Software **SYSTRAN Translate** (SYSTRAN Translate Tool 2020) arbeitet mit dem Open-Source-System OpenNMT (Klein et al. 2017) und wird von einer Gemeinschaft von Expert*innen aus der ganzen Welt trainiert. Aus datenschutzrechtlichen Gründen bleiben alle übersetzten Dateien nur für eine begrenzte Zeit gespeichert. Ein Vorteil dieses Systems besteht darin, dass im Vergleich zu anderen Generalübersetzungssystemen SYSTRAN Translate maßgeschneiderte industrie- sowie domänenspezifische Übersetzungen anbietet, die

für Fachleute von großer Bedeutung sein können (vgl. SYSTRAN Blog 2020). Im Vergleich zu dem ersten regelbasierten und nachfolgenden hybriden Ansatz (in Kapitel 2.1.3. präsentiert), bietet SYSTRAN unlängst Echtzeit-Übersetzungen in mehr als 140 Sprachkombinationen an. Dies geschieht dank seinem „*Pure Neural Machine Translation* (PNMT)-Übersetzer“, was als eine reine neuronale NMÜ-Software bezeichnet werden kann, die auf künstlichen neuronalen Netzen basiert. Das System generiert qualitativ hochwertige Übersetzungen in vielen Bereichen, insbesondere in der Automobilindustrie, in der rechtlichen Sphäre sowie dem IT- und touristischen Bereich (vgl. SYSTRAN beyond language 2020). SYSTRAN stellt den neuesten Stand der Technik dar, da es schon als erste Basis für die Entwicklung von Google Translate gedient hat (vgl. SYSTRAN Blog 2020). Crego et al. (2016) haben zum ersten Mal über die Entwicklung des Systems berichtet. Das System basiert auf einem Encoder-Decoder-Modell mit Attention-Mechanismus und bietet die Möglichkeit, die Ausgangs- bzw. Quellwörter in einem externen Wörterbuch nachschlagen zu können.

Unter Berücksichtigung des vorangehenden Vergleichs der verschiedenen NMÜ-Systeme wird im empirischen Teil SYSTRAN Translate verwendet. Die Ergebnisse werden in Kapitel 5. und 6. gesammelt und ausgewertet, um zu beobachten, ob durch Pre-Editing-Regeln für eine spezifische Domäne eine Optimierung der Qualität der NMÜ in einer spezifischen Sprachkombination erzielt werden kann.

2.1.3. Hybride Ansätze

Ziel der hybriden Ansätze ist es, alle Vorteile unterschiedlicher MÜ-Systeme zusammenzustellen und die systemspezifischen Nachteile zu umgehen. Dementsprechend vereinen die ersten hybriden Ansätze bestimmte Systeme und Komponenten anderer MÜ-Systeme (vgl. Werthmann & Witt 2014:99). Zum Beispiel ist das regelbasierte SYSTRAN-System mit Komponenten des statistischen Ansatzes integriert worden (SYSTRAN Third Quarter 2009 Revenue Release 2009). Ein anderes Beispiel ist die Kombination eines statistischen mit einem beispieldbasierten Ansatz (Smith & Clark 2009). Stein (2009:14) weist bereits im Speziellen auf hybride Ansätze hin, die regelbasierte und statistische Systeme kombinieren, da sie besonders der Verbesserung statistischer MÜ dienen. Sie betreffen vor allem diejenigen Kombinationen von Sprachen mit einer unterschiedlich starken Flexion oder mit begrenzt verfügbaren Korpora.

Letztendlich betreffen die hybriden Ansätze auch **statistische und neuronale Verfahren**, da der statistische Ansatz nicht komplett vom neuronalen ersetzt wurde (vgl.

Schmalz 2019:199). Als Beispiel wird die Studie von Grundkiewicz und Junczys-Dowmunt (2018) erwähnt. Dabei wird ein phrasenbasiertes SMÜ-System mit einem neuronalen Encoder-Decoder-Modell mit Attention-Mechanismus im Bereich der automatischen Korrektur grammatikalischer Fehler in Texten kombiniert. Ihre Resultate nähern sich der menschlichen Leistung.

Allerdings werden einige **hybride Ansätze auch nur innerhalb der NMÜ** vorgeschlagen. Luong (2017) schlägt eine neue Architektur vor, bei der die Übersetzung auf Wortebene stattfindet, aber seltene Wörter sowohl bei der Ein- als auch bei der Ausgabe auf Zeichenebene repräsentiert werden. Diese Lösung für seltene Wörter, die auf Wortebene vorkommen, weist ein kürzeres und einfacheres Training auf. Ziel dieser Kombination ist zu einem neuronalen „offenen Vokabular“ (engl. *open-vocabulary*) zu kommen (2017:75f). Chen et al. (2018) haben hingegen hybride Modelle entwickelt, um die einzelnen Architekturen (zum Beispiel Encoder-Decoder-Modell, Konvolutionsnetze und Transformer-Modell) zu überholen. Bei diesen Modellen werden die Encoder und Decoder des Transformer-Modells und eines schon optimierten Encoder-Decoder-Modells mit rekursiven neuronalen Netzen (RNMT+, Chen et al. 2018) kombiniert. Die Ergebnisse ihrer empirischen Studie haben gezeigt, dass die Kombination aus dem Transformer-Modell-Encoder und dem RNMT+-Decoder bessere maschinelle Resultate erzielt.

2.2. Qualität der Übersetzung

Übersetzen ist ein komplexer kognitiver, linguistischer, sozialer, kultureller und technologischer Prozess und die Evaluierung der Qualität seiner Resultate spiegelt die Komplexität dieses Prozesses wider, daher ist die Qualität der Übersetzung schwer messbar. Die Evaluierung der Übersetzungsqualität (engl. *translation quality assessment*, TQA) liegt aus diesem Grund im Kern der Debatte sowohl in der Übersetzungswissenschaft als auch im Bereich der Übersetzungstechnologie (vgl. Castilho et al. 2018:10). Die Ansätze zur Evaluierung der Übersetzungsqualität sind stark unterschiedlich, da die Qualität personen-, gruppen- oder kontextbezogene unterschiedliche Bedeutungen haben kann. Außerdem spielen die Art und das Ziel einer Evaluierung eine wichtige Rolle, je nachdem, ob die Evaluierung im Rahmen eines Produktionsprozesses (in der Übersetzungsindustrie) oder einer Forschungsstudie stattfindet. Im ersten Fall geht es darum, ein bestimmtes Qualitätsniveau in der Übersetzung zu erreichen und an die Endbenutzer*innen zu liefern. Im zweiten Fall ist das

Ziel einer Evaluierung eine eventuelle Verbesserung in der Übersetzungsqualität zu messen und nachvollziehen zu können, im Vergleich zu bisherigen Studien. Der Begriff der Übersetzung kommt im Bereich der Industrie und im Forschungsbereich eine unterschiedliche Bedeutung zu. Deswegen sind sich Industrie und Forschung bezüglich einer objektiven Evaluierungsmethode der Übersetzungsqualität nicht einig (vgl. Castilho et al. 2018:11). Im Folgenden werden die wichtigsten Qualitätsmetriken vorgestellt, die zur Evaluierung der Qualität der menschlichen und maschinellen Übersetzung dienen.

2.2.1. SAE J2450

Der J2450-Qualitätsstandard (SAE International J2450) wurde von der Society of Automotive Engineers (SAE) gesetzt, zur objektiven Messung der Übersetzungsqualität von Serviceinformationen bzw. -dokumentation in der Automobilindustrie (engl. *automotive service information*), unabhängig von der Ausgangs- und Zielsprache und von der Übersetzungsmethode (menschlich oder maschinell) (vgl. SAE International J2450 2001:1). Durch diesen Standard werden die Übersetzungen lediglich auf linguistischer Ebene evaluiert, insbesondere auf syntaktischer Ebene, da Fehler in Bezug auf Stil oder Register nicht berücksichtigt werden. Die allgemeine Struktur dieser Metrik besteht aus vier Elementen: sieben primären Fehlerkategorien, zwei sekundären Unterkategorien, wo Fehler als gravierend (engl. *serious*) oder gering (engl. *minor*) eingeschätzt werden, zwei Regeln in Zweifelsfällen und numerischen Schweregraden (vgl. SAE International J2450 2001:3f.). Die sieben Fehlerkategorien sind: falsche Bezeichnung (engl. *Wrong Term*, WT), wo unter Bezeichnung einzelne Wörter, Mehrwortbegriffe, Abkürzungen, Akronyme, Nummern oder Eigennamen fallen, syntaktische Fehler oder Wortstellung (engl. *Syntactic Error*, SE), Auslassung (engl. *Omission*, OM), Wortstruktur oder „Übereinstimmungsfehler“ (engl. *Word Structure or Agreement Error*, SA), Rechtschreibfehler (engl. *Misspelling*, SP), Zeichensetzungsfehler (engl. *Punctuation Error*, PE), Sonstiger Fehler (engl. *Miscellaneous Error*, ME) (2001:5-9). Die „Fehlübersetzungs“-Kategorie ist nicht vorhanden, da jeder Fehler eine Fehlübersetzung ist, die die Bedeutung des Zieltextes beeinflussen und beeinträchtigen kann. Jede Evaluator*in kennzeichnet jeden Fehler zweimal, und zwar nach Fehlerkategorie und nach gravierendem oder geringem Fehler, je nachdem, ob der Fehler das Verständnis der Übersetzungsbedeutung erschwert oder nicht. Im Zweifelsfall müssen die Evaluator*innen einen Fehler als gravierend einstufen oder im Fall eines unklaren Fehlers fällt er willkürlich in die primäre Fehlerkategorie. Dazu kommt der numerische Schweregrad. Jedem Fehler wird ein Schweregrad von 1 bis 5 zugewiesen, wobei 5 dem schwerwiegendsten Fehler entspricht und 1 einen Fehler ohne

gravierende Konsequenzen kennzeichnet (2001:2ff.). Die normalisierten Ergebnisse werden berechnet beim Addieren aller einzelnen numerischen Schweregrade und im Anschluss durch Dividieren der Gesamtsumme durch die Zahl der Wörter im Ausgangstext (vgl. SAE International J2450 2001:5).

2.2.2. Multidimensional Quality Metrics (MQM)

Diese Metrik (Multidimensional Quality Metrics (MQM), Lommel et al. 2015) wurde direkt von dem LISA (Localization Industry Standards Association) QA-Modell⁸ zur Qualitätsevaluierung der Software- und Dokumentationslokalisierung abgeleitet (vgl. Lommel 2018:113). Wie von Lommel (2018) betont wird, werden einige Prinzipien des LISA QA-Modells beibehalten. Die MQM-Metrik legt keine einzelne Metrik fest, die für alle Übersetzungen und Inhaltstypologien anwendbar sei. Stattdessen bestimmt sie ein gemeinsames flexibles Vokabular bzw. Kataloge zur Beschreibung von Übersetzungsfehlern und daher zur Angabe der richtigen Metrik, je nach Fall. Diese Metrik ist kompatibel mit vorab bestehenden Anforderungen und mit der Funktionsweise einiger Tools, damit sie am einfachsten übernommen werden kann. Des Weiteren hat die Metrik eine Baumstruktur, wo jede Fehlerkategorie hierarchisch bestimmt wird und so ein hoher Grad an Genauigkeit erzielt wird. Dementsprechend hängt die MQM-Metrik stark von den eigentlichen Übersetzungsprojektanforderungen ab, was die Evaluierung anbelangt (2018:113f.). Die MQM-Metrik (Lommel et al. 2015) bildet also einen systematischen Rahmen zur Festlegung von Qualitätsmetriken auf der Basis textlicher Merkmale. Sie eignet sich einen funktionalistischen Ansatz an. Die Qualität wird dadurch definiert, wie gut ein Text seinen kommunikativen Zweck erfüllt (vgl. Lommel et al. 2015:o. S.). Die Metrik legt viele Probleme offen und im Folgenden werden die Hauptdimensionen erläutert: Genauigkeitsgrad (engl. *Accuracy*), Design (engl. *Design*), Geläufigkeit (engl. *Fluency*), Internationalisierung(sgrad) (engl. *Internationalization*), lokale Konventionen (engl. *Locale Convention*), Stil (engl. *Style*), Terminologie (engl. *Terminology*), Wahrheit (engl. *Verity*), Kompatibilität mit älteren und ungeeigneten Metriken (engl. *Compatibility*) und Sonstiges (engl. *Other*). Außerdem sind im Allgemeinen ungefähr zwanzig Problem- bzw. Fehlertypologien in einem MQM-Kern (engl. *core*) (siehe Abbildung 2, S. 23) enthalten, die am häufigsten bei

⁸ Im Rahmen dieser Arbeit nicht beschrieben.

Qualitätsevaluierungsaufgaben vorkommen. Der Bewertungsmechanismus basiert auf der Fehleranzahl und auf vier unterschiedlichen Schweregraden der jeweiligen Fehler (vgl. Lommel et al. 2015:o. S.).

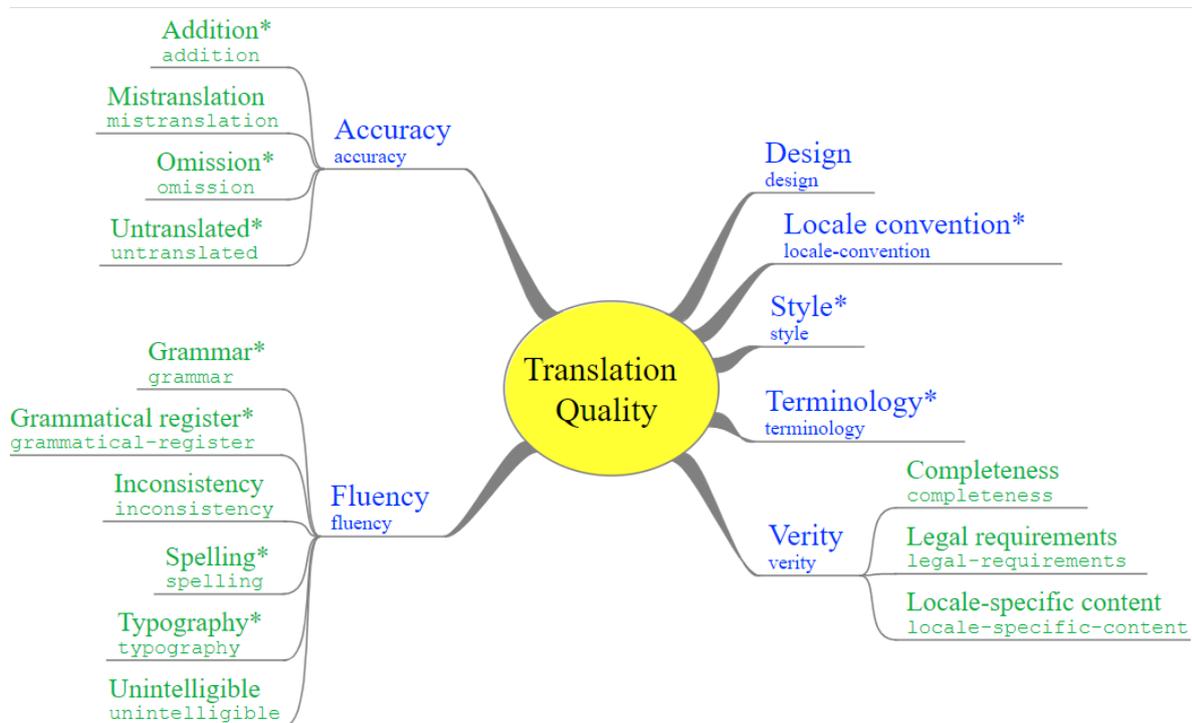


Abbildung 2: MQM Kern. (Lommel et al. 2015:o. S.)

Bei der MQM-Metrik werden beide Ansätze zur Übersetzungsevaluierung unterstützt, und zwar der holistische und der analytische Ansatz. Im Rahmen **holistischer Evaluierung** wird sich die Qualität auf allgemeine Kriterien stützen, da dieser Evaluierungsansatz die Übersetzung als Ganzes betrachte. Hingegen werden bei der **analytischen Evaluierung** einzelne Fehler berücksichtigt und analysiert (vgl. Lommel 2018:122). Hinzu kommt, dass diese zwei Ansätze verschiedenen Zwecken dienen: Beim holistischen Ansatz wird schnell entschieden, ob eine Übersetzung den Anforderungen insgesamt entspricht und es wird kein Feedback über bestimmte Fehler gegeben. Beim analytischen Ansatz jedoch werden spezifische Fehler nachgewiesen, aber es wird kein gesamtes Bild der Übersetzung vermittelt. Im Übrigen ist dieser Ansatz zeitaufwändiger und erfordert, dass die Evaluator*innen dafür trainiert werden (2018:122f.). Die **DQF** (Dynamic Quality Framework) **Fehlertypologie** ist von TAUS (Translation Automation User Society) entwickelt worden und gehört als „Unterfehlerkategorie“ zur MQM-Metrik. Nach einigen Revisionen sind MQM und DQF kompatibel geworden, das heißt, dass die sechs festgestellten Probleme bei der DQF den MQM-

Dimensionen direkt entsprechen, außer das *kudos*-Merkmal, das nur in der DQF zu finden ist und das sich auf die positiven Aspekte der Übersetzung bezieht (vgl. Lommel et al. 2015:o. S.). Zum Zeitpunkt der Abfassung dieser Arbeit sind MQM und DQF schon seit einigen Jahren adaptiert und integriert worden und werden heutzutage als neuer Standard zur Evaluierung der Übersetzungsqualität verwendet (vgl. Lommel 2018:124ff.).

2.2.3. Göpferichs Modell

Zur Qualitätsevaluierung kommen auch andere Modelle hinzu, vor allem in Bezug auf die verständliche Formulierung von Fachtexten. Solche Modelle helfen dabei, verschiedene Forschungsergebnisse zu vereinen und dienen zu didaktischen und gestalterischen Zwecken (vgl. Lutz 2017:284f.). Nach dem Hamburger Verständlichkeitsmodell (Langer et al. 1974) ist von Göpferich das **Karlsruher Verständlichkeitskonzept** (2001) eingeführt worden. Wie von Lutz (2017) erklärt, zieht das Modell den ganzen Textproduktionsprozess in Betracht und ist insbesondere für Texte im Bereich der technischen Kommunikation relevant. Es wird in drei Teile gegliedert, und zwar Auftragsdaten, Textproduktionsdaten und Text, der selbst mithilfe von sechs Verständlichkeitsdimensionen – Korrektheit, Struktur, Perzipierbarkeit, Simplizität, Motivation und Prägnanz – definiert wird. Eine wesentliche Rolle spielt bei diesem Modell die kommunikative Funktion, die sich aus dem Ziel eines Textes, seinen Adressat*innen und Sender*innen ergibt, indem sie als Anforderung zum Textproduktionsauftrag gilt und alle andere Daten bezüglich der Textgestaltung bestimmt. Als Bezugsgröße dient sie zur Textkodierung, -evaluierung und -optimierung (vgl. Lutz 2017:287).

2.3. Qualität maschineller Übersetzung

Schmalz (2019:202) weist darauf hin, dass die maschinelle Übersetzungsqualität von einigen Faktoren abhängt. Darunter fallen das ausgewählte MÜ-System, das Trainingsmaterial mit Umfang und Domänen, das ausgewählte konkrete Sprachpaar, die Textsorte und die Zieldomäne. Diesbezüglich behauptet Way (2018), dass die Übersetzungsqualität messbar sein muss, um ein höheres maschinelles Qualitätsniveau zu erreichen:

If MT system output is good enough for some areas of application, yet at the same time system developers are trying hard to improve the level of translations produced by their engines, then

translation quality – whether produced by a machine or by a human – needs to be *measurable*⁹. (Way 2018:160)

Evaluierung maschineller Übersetzung bedeutet ein Prozess der Analyse der Resultate bzw. Übersetzungen eines MÜ-Systems und der Beurteilung der Korrektheit durch eine messbare Methode. Eine MÜ-Evaluierung dient zur Verbesserung der Übersetzungen selbst, aber auch zur Weiterforschung bzw. -entwicklung der MÜ-Systeme und ermöglicht Vergleiche zwischen unterschiedlichen Systemen und Methoden (vgl. Avramidis 2019:2). Die meistverwendeten Evaluierungsmethoden maschineller Übersetzung sind die menschliche und die automatische Evaluierung. Bei der menschlichen Evaluierung werden maschinelle Resultate von humanen Evaluator*innen qualitativ beurteilt, wobei eine zuverlässige, aber auch zeit- und kostenaufwändige Evaluierung stattfindet. Bei der automatischen Evaluierung hingegen können bestimmte qualitative Merkmale maschineller Resultate mithilfe von spezifischen Programmen durch numerische Werte schnell und unbegrenzt dargestellt werden, vor allem im Rahmen des Entwicklungsprozesses. Solche Programme bzw. Metriken vergleichen die maschinellen Resultate mit menschlichen Referenzübersetzungen (2019:2). Darunter befinden sich *Word Error Rate* (WER, Nießen et al. 2000), *Bilingual Evaluation Understudy* (BLEU, Papineni et al. 2002), METEOR (Banerjee & Lavie, 2005) und *Translation Error Rate* (TER, Snover et al., 2006).

Eine erwähnenswerte Basis der gegenwärtigen Evaluierungsmethoden im Bereich der MÜ geht auf die Studie von Dorr et al. (2011) zurück, obwohl die beschriebenen Methoden Bezug auf regelbasierte und statistische MÜ nehmen, statt auf die NMÜ. Dorr et al. (2011) haben über zwei MÜ-Evaluierungsparadigmen geschrieben, und zwar die *Glass Box*- und dann die *Black Box*-Evaluierungen (2011:745). Bei der *Glass Box*-Evaluierung wird die Qualität des Systems hinsichtlich der internen Eigenschaften des Systems bewertet, wie beispielsweise der Umgang mit linguistischen Aspekten. Bei der *Black Box*-Evaluierung wird die Qualität des Systems auf der Basis der maschinellen Resultate sowie des objektiven Verhaltens des Systems gegenüber einem vordefinierten Evaluierungsset bewertet. Dabei wird kein Bezug auf die internen Übersetzungsmechanismen des Systems genommen. Dies hat maßgeblich zur Forschung im Rahmen des Systemvergleichs auf der Grundlage gleicher Evaluierungssets beigetragen (2011:745f.). Innerhalb des *Black Box*-Evaluierungsansatzes ist als Grundlage der heutigen Ansätze eine weitere Untergliederung von Dorr et al. (2011) wichtig, und zwar, die extrinsischen und intrinsischen Bewertungsmethoden. Die extrinsischen Methoden sind

⁹ Kursivschrift im Original.

aufgabenbasiert (engl. *task-based*) und zielen auf das Testen maschineller Resultate ab, in Bezug auf eine spezifische Aufgabe und auf die Nützlichkeit der MÜ selbst. Der Fokus der intrinsischen (menschlichen oder automatischen) Bewertungen hingegen liegt bei der Qualität der MÜ-Resultate, wobei es sich meistens um Qualitätsvergleiche zwischen maschinellen Resultaten und einem Set von vordefinierten qualitativ hochwertigen Referenzübersetzungen handelt (2011:746).

Unter Berücksichtigung der oberen Ausführungen zeigt sich, dass eine quantitative Qualitätsevaluierung maschineller Übersetzung trotzdem noch schwer durchführbar ist, da die Bewertungen meistens subjektiv sind und selbst automatische Evaluierungsmethode mit subjektiv ausgewählten Referenzübersetzungen verglichen werden (vgl. Schmalz 2019:201). Der Bedarf an Referenzübersetzungen, die bisweilen nicht verfügbar sind, stellt einen Nachteil der automatischen Evaluierung durch Metriken dar (vgl. Avramidis 2019:3). Aus diesem Grund wird der Begriff der *Quality Estimation* (QE, Specia et al. 2010) eingeführt, wobei eine Qualitätsevaluierung maschineller Resultate automatisch stattfindet, aber referenzunabhängig.

Im Folgenden werden beide Evaluierungsmethoden (menschliche und automatische) der maschinellen Übersetzungsqualität dargelegt. Insbesondere wird das Augenmerk auf die Methoden gelegt, die für die NMÜ anwendbar sind, denn die NMÜ ist die jüngste Entwicklung im MÜ-Bereich. Bei der automatischen Evaluierung wird zwischen Metriken und aktuellen Ansätzen der *Quality Estimation* unterschieden, die eine wichtige Grundlage der vorliegenden Masterarbeit bieten.

2.3.1. Menschliche Evaluierung der MÜ

Die menschliche Evaluierung von MÜ hat ihre Grenzen, vor allem was die Evaluator*innen angeht, wie schon von Dorr et al. (2011) eingeführt und zu einem späteren Zeitpunkt von Doherty (2017) bekräftigt wurde. Die Schwierigkeiten, mit denen man im Falle einer menschlichen Evaluierung konfrontiert wird, zeigen sich auch im Rahmen der neuronalen maschinellen Übersetzung. Dorr et al. (2011:752) behaupten, dass die Evaluator*innen unterschiedliche Meinungen über eine Übersetzung haben und daher unterschiedliche Antworten für die gleiche Bewertungsaufgabe geben können, weil sie individuell verschiedene Erfahrungen gesammelt haben. Aus diesem Grund wird üblicherweise eine Gruppe von unabhängigen Evaluator*innen benötigt, um den Mittelwert dieser Differenzen festzustellen. Des Weiteren stellt Doherty (2017) fest, dass die Evaluator*innen über keinen Grad an Expertise in dem Bereich des zu evaluierenden Textes, über keine geeigneten

Sprachkompetenzen oder spezifisches Training für die Evaluierungsaufgaben verfügen (2017:139-142). Eine wichtige Rolle spielt nach Doherty (2017:138) eine Analyse der Interrater-Reliabilität, um die Reliabilität der Evaluator*innen sowie der Studienergebnisse zu gewährleisten.

Allerdings stellen zweisprachige Evaluator*innen noch den goldenen Standard dar: Übersetzungen werden zunächst für humane Leser*innen angefertigt und deswegen sind die Leser*innen, die adäquate Personengruppe, imstande die Qualität der Übersetzungen richtig zu bewerten. Außerdem verfügen die Leser*innen über ein Weltverständnis, das ihnen ermöglicht, die Wichtigkeit eines Übersetzungsfehlers in der Praxis zu erkennen (vgl. Dorr et al. 2011:751). Aus diesem Grund werden im Folgenden die gängigsten menschlichen MÜ- bzw. NMÜ-Evaluierungsmethoden erwähnt, zusammen mit einigen Studien, die als Inspiration für diese Arbeit gelten.

2.3.1.1. TQA-Kriterien

Eine menschliche Qualitätsevaluierung (oder TQA) der MÜ, die auch auf die NMÜ zutrifft, bezieht sich vor allem auf Kriterien wie Angemessenheit (engl. *Adequacy*) und Geläufigkeit (engl. *Fluency*)¹⁰, aber kann sich auch auf sekundäre Kriterien wie Lesbarkeit (engl. *Readability*), Verständlichkeit (engl. *Comprehensibility*), Benutzerfreundlichkeit (engl. *Usability*) und Akzeptanz (engl. *Acceptability*) beziehen. Außerdem können rankingbasierte komparative Ansätze auf mehrere Übersetzungen des gleichen Ausgangstextes bzw. -segments angewandt werden (vgl. Castilho et al. 2018:17). *Adequacy* (auch engl. *Accuracy* oder *Fidelity* genannt) entspricht der Übertragung der Bedeutung aus dem Ausgangssegment in das Zielsegment und wird dementsprechend bewertet, während der Fokus von *Fluency* auf dem Zieltext liegt, wobei die Einhaltung von den zielsprachlichen Regeln und Normen berücksichtigt wird, unabhängig vom Ausgangstext. Im Regelfall geht die Bewertung von *Adequacy* mit der von *Fluency* einher und beide werden durch Ordinalskalen (Likert-Skalen, im Normalfall von Stufe 5 bis 1) auf Satz- bzw. Segmentebene evaluiert. *Adequacy* benötigt möglichst eine Evaluierung von zweisprachigen Evaluator*innen, *Fluency* hingegen kann auch von einsprachigen bzw. Erstsprachler*innen bewertet werden, die nur die Zielsprache beherrschen (2018:18). Wie Castilho et al. (2018) zeigen, wurden und werden noch heute *Readability* und *Comprehensibility* bei der Evaluierung in vielen Studien zusammengeführt.

¹⁰ Die Kriterien von *Adequacy* und *Fluency* wurden erstmals im Rahmen des *Linguistics Data Consortium* zur MÜ-Evaluierung im Jahr 2005 verwendet (LDC (2005). Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations. Revision 1.5.)

Einerseits nimmt *Readability* direkten Bezug auf die Lesbarkeit des Ausgangs- oder Zieltextes und basiert auf linguistischen Merkmalen, wie zum Beispiel Satzlänge und auf nicht linguistischen Merkmalen wie Formatierung. Andererseits basiert *Comprehensibility* auf den Leser*innen und auf ihrer Bildung und Vertrautheit mit bestimmten Textsorten (2018:18f.). Van Slype (1979) stellt *Comprehensibility* als ein Merkmal eines Textes dar, das bezeichnet, wie verständlich ein Text für eine bestimmte Leser*in ist. Die Begriffe von Lesbarkeit und Verständlichkeit sind mit dem Begriff von MTranslatability verbunden und werden im Kapitel 2.4. und anschließend entsprechend miteinbezogen.

2.3.1.2. Ranking

Ein weiterer von Castilho et al. (2018) behandelte Punkt ist das Ranking. Es wird besonders in dem Forschungsbereich als Methode verwendet, zur Evaluierung von MÜ-Resultaten aus unterschiedlichen MÜ-Systemen basierend auf demselben Ausgangstext (vgl. Castilho et al. 2018:21). Dabei werden menschliche Evaluator*innen gefragt, ausgewählte Sätze in die Zielsprache nach gegebenen Kriterien (wie zum Beispiel Geläufigkeit) einzustufen. Genauer erklärt, werden die Evaluator*innen das Ausgangssegment und zwei oder mehrere maschinell erzeugte Zielsegmente ungeordnet vorgelegt bekommen, damit sie entweder die „beste“ Übersetzung auswählen oder die „beste“ und „schlechteste“ Zielversionen reihen, nach bestimmten Kriterien. Die Ergebnisse dieser Methode sind relativ einfach zu interpretieren und die Methode selbst ist schnell durchzuführen (2018:21). Das Ranking dient als Basis der Best-Worst-Scaling Methode, die spezifisch für den empirischen Teil dieser Arbeit ausgewählt wurde und die ausführlicher in Kapitel 2.6. erklärt wird.

2.3.1.3. Fehlerklassifizierung

Die Evaluierungsmethoden von Fehlerklassifizierung und -analyse werden bei der MÜ entwickelt, damit Fehler in einem übersetzten Text identifiziert und klassifiziert werden, zum Zweck der Verbesserung im Rahmen einer Systementwicklung oder -anwendung (vgl. Popović 2018:130). Die Fehleranalyse zielt darauf ab, ein „Fehlerprofil“ der Übersetzungen zu skizzieren, wo alle Fehler nach Fehlertypologie verteilt werden. Diese Methode könnte auch bei einem Vergleich verschiedener Übersetzungen angewandt werden oder um den Einfluss unterschiedlicher Fehlertypen auf den Post-Editing-Aufwand zu analysieren (2018:131). Wie von Popović (2018) gedeutet, kann die Fehlerklassifizierung menschlich, automatisch oder semi-automatisch durchgeführt werden, aber in dieser Arbeit wird sie nur in Bezug auf die menschliche Anwendung zur Evaluierung maschinell übersetzter Texte betrachtet. Die

Fehlerklassifizierung ist ein herausfordernder Prozess. Nach einer Fehlerannotation ist eine passende Fehlertypologie, ein Set von Fehlerkategorien, zu bestimmen. Dafür müssen die wichtigen Fehlertypen für eine bestimmte Studie bzw. Aufgabe überlegt und spezifiziert werden. Des Weiteren sollten die Fehlertypen alle Vor- und Nachteile des MÜ-Systems (hier auch NMÜ-Systems) zeigen, die für eine spezifische Studie und Sprachkombination wichtig sind. Detailliertere Fehlertypen sind informativer, aber auch schwieriger zu erkennen (2018:131f.). Popović (2018:132) betont diesbezüglich, dass die Fehlertypen sowohl linguistische als auch translatorische Aspekte abdecken sollten, obwohl es keine allgemeinen Regeln zur Bestimmung der Fehlerkategorien gibt. Einige Nachteile der manuellen bzw. menschlichen Fehlerklassifizierung sind, dass sie ressourcen- und zeitaufwändig ist und dass sich die Evaluator*innen häufig nicht einig sind, insbesondere wenn viele Fehlerkategorien zu erkennen sind (2018:129).

Popović (2018) gibt einen Überblick der verschiedenen Fehlertypologien zur manuellen Klassifizierung, die in den letzten fünfzehn Jahren und in einigen Studien verwendet worden sind. Von der SMÜ-Fehlerklassifizierung von Vilar et al. (2006) (siehe Abbildung 3, S. 29) über die von Castilho et al. (2017b) verwendeten spezifisch gestalteten Fehlertypologien zum Vergleich von SMÜ- und NMÜ-Resultaten. Außerdem wird die Verwendung von der MQM-Metrik als Fehlertypologie erläutert, die zum Beispiel von Klubicka et al. (2017) für einen Vergleich zwischen phrasenbasierter und neuronaler MÜ verwendet wurde.

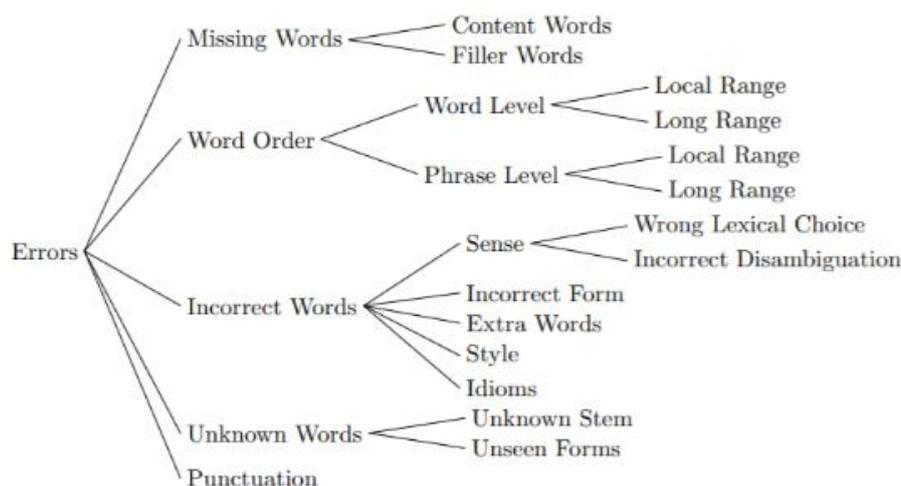


Abbildung 3: Fehlerklassifizierung (Vilar et al. 2006:699)

Die hier erwähnten Studien zeigen, dass Fehlerkategorien auch für die Analyse von NMÜ verwendet und eben adaptiert werden können. Obwohl sich die NMÜ-Fehler (im Verhältnis zu SMÜ-Fehlern) trotz inkorrektur Terminologie den menschlichen Fehlern annähern (vgl.

Yamada 2019:96), hat Flanagan (1994) eine der ersten Fehlerklassifizierungen vorgeschlagen, mit flexiblen Kategorien (vgl. Flanagan 1994:71), die noch heute in der Evaluierung maschineller Resultate von Nutzen sind (vgl. Stankevičiūtė et al. 2017:77).

<u>Category</u>	<u>Description</u>
Spelling	Misspelled word
Not Found Word	Word not in dictionary
Accent	Incorrect Accent
Capitalization	Incorrect upper or lower case
Elision	Illegal elision or elision not made
Verb inflection	Incorrectly formed verb, or wrong tense
Noun inflection	Incorrectly formed noun
Other inflection	Incorrectly formed adjective or adverb
Rearrangement	Sentence elements ordered incorrectly
Category	Category error (e.g. noun vs. verb)
Pronoun	Wrong, absent or unneeded pronoun
Article	Absent or unneeded article
Preposition	Incorrect, absent or unneeded preposition
Negative	Negative particles not properly placed or absent
Conjunction	Failure to reconstruct parallel constituents after conjunction, or failure to identify boundaries of conjoined units
Agreement	Incorrect agreement between subject-verb, noun-adjective, past participle agreement with preceding direct object, etc.
Clause boundary	Failure to identify clause boundary, or clause boundary unnecessarily added
Word Selection	Word selection error (single word)
Expression	Incorrect translation of multi-word expression

(Flanagan 1994:67)

Dazu folgert Popović (2018:139f.), dass die Klassifizierungsaufgaben vereinfacht werden können, indem die Anzahl und die genaue Definition der Fehlerkategorien gegeben werden. Dies trägt außerdem zur Zuverlässigkeit der Ergebnisse bei. Hinsichtlich der Definition der Kategorien, hängt sie von der Aufgabe selbst ab, jedoch auch von der Sprachkombination. Außerdem wird sie von der Domäne eines Textes stark beeinflusst. Comparin und Mendes (2017) bieten eine interessante Grundlagenstudie, die in der Vorbereitungsphase der empirischen Studie dieser Arbeit gute Anwendung finden konnte. Sie behaupten, dass durch die menschliche Fehlerannotation zur Evaluierung maschineller Übersetzung die Möglichkeit besteht, Strategien zu entwickeln, Fehler zu vermeiden und spezifische Regeln zur automatischen Korrektur dieser Fehler zu erstellen (obwohl in ihrer Studie Bezug auf die Post-

Editing-Phase genommen wird). Außerdem wird für ihre Studie eine auf der MQM-Metrik basierte Taxonomie verwendet, die ermöglicht hat, die wichtigsten Fehlertypen bzw. Fehlerkategorien in maschinell übersetzten Texten aus dem Englischen ins Italienische zu analysieren.

ACCURACY: errors in this category concern the relationship between the source text and the target

text and the extent to which the latter maintains the meaning and the information of the former

FLUENCY: errors in this category regard the quality of a text, assessing whether it is wellwritten and easy to read, and if it accomplishes its communication purpose in the target language

STYLE: issues concerning register and fluency

TERMINOLOGY: mistranslation of terminology

WRONG LANGUAGE VARIETY: use of a word or expression from a different language variety.

NAMED ENTITIES: wrong translation of proper nouns

FORMATTING AND ENCODING: issues concerning the segmentation of sentences and paragraphs

(Comparin & Mendes 2017:o. S.)

2.3.1.4. Linguistische Evaluierung: Test Suites

Als weitere Grundlage wird die Studie von Burchardt et al. (2017) vorgestellt, die eine linguistische Evaluierung von unterschiedlichen phrasenbasierten (statistischen), regelbasierten und neuronalen MÜ-Systemen durchgeführt haben. Die Basis für ihre Studie bildet eine manuell zusammengestellte Test Suite, die mehr als hundert linguistische Phänomene pro Sprachrichtung enthält. Der Fokus der Studie liegt auf der Evaluierung und dem Vergleich der Leistungsfähigkeit verschiedener MÜ-Systeme. Neuronale Systeme werden miteinbezogen, um den Umgang mit den vielfältigen linguistischen Phänomenen aufzuklären (2017:161f.). Die Test Suite besteht aus Testsegmenten bzw. -sätzen aus Parallelkorpora und Grammatikressourcen und für jedes Segment werden eine Phänomenkategorie und das sprachspezifische Phänomen selbst annotiert. Die Kategorien berücksichtigen eine Serie von grammatikalischen Aspekten, die einem MÜ-System Schwierigkeiten bei dem Übersetzungsprozess bereiten können (2017:160f.). Burchardt et al. (2017) haben in ihrer quantitativen und qualitativen Studie das Fazit gezogen, dass die Resultate von den NMÜ-Systemen signifikant besser sind als jene der bisherigen statistischen Systeme (2017:169). Die NMÜ-Systeme zeigen generell eine bessere Leistung auf. Bei der Sprachkombination Deutsch-Englisch und vice versa erwiesen sich bessere Resultate, was die Long-Distance-Dependencies, Mehrwort-Formulierungen, Funktionswörter, Verbvalenzen und Subordination betreffen. (2017:164). Dazu werden bei der Sprachkombination Englisch-Deutsch bessere Resultate bei Koordination, Ellipse und Terminologie erzielt (2017:167).

2.3.1.5. Evaluierung auf Dokumentenbasis

Läubli et al. (2018) haben einen weiteren Ansatz zur menschlichen (von professionellen Übersetzer*innen durchgeführten) Qualitätsevaluierung vorgeschlagen. Humanübersetzungen und NMÜ werden durch die Kriterien von *Adequacy* und *Fluency* auf Satz- und Dokumentenebene mithilfe von Evaluierungsprotokollen bewertet und paarweise eingestuft. Dies hat zum Ziel, eine Evaluierung auf Dokumentenbasis durchzusetzen. Sie heben hervor, dass die NMÜ-Resultate schon qualitativ hochwertig sind und deswegen Fehler auf Satzebene kaum oder nicht erkannt werden. Die NMÜ-Fehler sind auf Dokumentenebene trotzdem entscheidend, wenn es sich um die Qualität verschiedener Übersetzungsergebnisse handelt (2018). Läubli et al. (2018) behaupten, dass im Regelfall eine Evaluierung auf Satzebene aus Kostengründen und aus Gründen der Studienvalidität stattfindet. Allerdings bei Bewertungskampagnen wird Kontext immer von den Übersetzer*innen und den Evaluator*innen in Betracht gezogen. Die Ergebnisse von Läubli et al. (2018) weisen darauf hin, dass bei der Evaluierung auf Dokumentenebene die Evaluator*innen eine Präferenz für die Humanübersetzung gegenüber der NMÜ manifestieren. Dies gilt sowohl bei der Bewertung von *Adequacy* als auch *Fluency*, während bei der Evaluierung auf Satzebene die Evaluator*innen Schwierigkeiten haben, die Humanübersetzung von der maschinellen zu unterscheiden (2018). Laut Läubli et al. (2018) ermöglicht eine Evaluierung auf Dokumentenbasis ein besseres Verständnis des Ausgangstextes und seine entsprechenden Übersetzungen seitens der Evaluator*innen und trägt dazu bei, auf Satzebene unsichtbare Fehler erkennbar zu machen.

Laut Rysová et al. (2019) muss die Qualität einer Übersetzung in Bezug auf ihre Funktion in der menschlichen Kommunikation evaluiert werden. Trotz der rasanten Entwicklungen fällt es MÜ-Leser*innen heute noch schwer, den übersetzten Text zu lesen, da morphologische, syntaktische und terminologische Fehler auftreten. Rysová et al. (2019) schlagen eine qualitative Evaluierung der NMÜ auf Dokumentenebene vor, die zur Identifizierung, menschlichen Annotation und linguistischen Beschreibung der Fehler in der Sprachkombination Englisch-Tschechisch führt. Ihr Ziel ist, NMÜ-Systeme zu vergleichen, um das Potenzial zu eruieren, dokumentenbasierte Diskurs-Phänomene zu übersetzen.

2.3.2. Automatische Evaluierung der MÜ

Die automatische Evaluierung der MÜ ist ein florierender Forschungsbereich, der zum Ziel hat, die MÜ-Resultate mit einer oder mehreren Referenzübersetzungen zu vergleichen, sprich

qualitativ hochwertigen Humanübersetzungen (vgl. Castilho et al. 2018:25). Mit automatischen Metriken – wie *Word Error Rate* (WER, Nießen et al. 2000), *Bilingual Evaluation Understudy* (BLEU, Papineni et al. 2002), METEOR (Banerjee & Lavie, 2005) oder *Translation Error Rate* (TER, Snover et al., 2006) – wird ein Wert berechnet, der die Ähnlichkeit bzw. Übereinstimmung von Referenz und maschineller Übersetzung quantifiziert (vgl. Koehn 2010:540). Nach Specia et al. (2010) wird die Reliabilität dieser Metriken gemessen, im Zusammenhang mit menschlichen Evaluierungen, weil die Metriken zur genauen Annäherung der menschlichen Bewertungen dienen. Obwohl automatische Metriken und humane Bewertungen in eine adäquate Korrelation gebracht werden, beschränkt die Notwendigkeit von Referenzübersetzungen bei Metriken die Menge Daten, die evaluiert werden können (2010:40). Um die Grenzen der Metriken zu überwinden, kann eine „Qualitätsvorhersage“ maschineller Resultate – das heißt eine *Quality Estimation* (QE) – stattfinden: „This task [QE] consists in estimating the quality of a system’s output for a given input, without any information about the expected output, that is, without reference translations“ (Specia et al. 2010:40). In ihrer Studie schlussfolgern Specia et al. (2010:48), dass QE unterschiedliche Anwendungen im Vergleich zu den traditionellen Metriken haben kann. Zum Beispiel können „schlechte“ Übersetzungen für das Post-Editing ausgefiltert oder kann die „beste“ Übersetzung verschiedener MÜ-Systeme ausgewählt werden. Hinsichtlich der MÜ-Evaluierung zeigen die Ergebnisse von Specia et al. (2010), dass Qualitätsbewertungen auf Satzebene für unbegrenzte neue Sätze erzielt werden können, basierend auf Modellen, die mit unterschiedlichen Trainingsdaten (vom MÜ-System, von der Sprachkombination und Textdomäne) trainiert werden (2010:39): „QE is generally addressed as a supervised machine learning task using a variety of algorithms to induce models from examples of translations described through a number of features and annotated for quality.“ (Specia & Shah 2018:203). Des Weiteren korrelieren die QE-Bewertungen besser mit den menschlichen Bewertungen als die Evaluierungsmetriken (vgl. Specia et al. 2010:48). QE führt zu einem oder mehreren numerischen Werten, die die Qualität der betrachteten Übersetzungen wiedergeben. Je nach Aufgabe bzw. Studie kann sich QE auf Dokumenten-, Satz-, Phrasen- oder Wortebene beziehen (vgl. Avramidis 2019:3).

Sowohl die oben erwähnten Metriken als auch QE sind für diese Arbeit relevant, weil sie auch als Evaluierungsmethoden in Bezug auf die NMÜ angewandt werden können. In den folgenden Kapiteln werden beide Ansätze (Metriken und QE) zur automatischen Evaluierung vorgestellt.

2.3.2.1. Metriken

Die automatischen Metriken messen die Qualität maschineller Resultate beim Vergleich der Resultate eines MÜ-Systems (auch „Kandidaten“ oder „Hypothesen“ genannt) mit einem Set an Referenzübersetzungen. Diese bilden den sogenannten „goldenen Standard“ der Quelldateien, die lediglich eine Stichprobe aller möglichen richtigen Übersetzungen darstellen (vgl. Dorr et al. 2011:759). Die Qualität dieser Evaluierungsmetriken wird gemessen, indem die Korrelation aufgewiesen wird zwischen den von den Metriken zugewiesenen Qualitätswerten und den menschlich zugewiesenen Werten (gewöhnlich *Fluency* und *Adequacy*) (2011:759). Automatische Metriken weisen einige Vorteile auf. Keine zweisprachigen Evaluator*innen werden benötigt, dementsprechend ist die Verwendung von automatischen Metriken kostengünstiger im Vergleich zum menschlichen Evaluierungsprozess. Des Weiteren ermöglichen sie ein schnelles Feedback (vgl. Castilho et al. 2018:26). Die folgenden Metriken werden häufig in der Evaluierung von NMÜ-Systemen eingesetzt.

Word Error Rate (WER) wurde erstmals von Nießen et al. (2000) eingeführt, als eine der ersten automatischen Metriken zur MÜ-Evaluierung und gilt als Standard-Evaluierungsmetrik zur automatischen Spracherkennung (vgl. Dorr et al. 2011:760). Die Metrik stützt sich auf die Levenshtein-Distanz (Levenshtein 1966), welche sich aus den Wörtern der maschinellen Resultate und den Wörtern der Referenzübersetzung berechnet. Diese wird dividiert durch die Länge (N Anzahl von Wörtern) der Referenzübersetzung, mit dem Ziel, das optimale Alignment zwischen Resultaten und Referenzübersetzung zu finden. Diese Distanz wird durch die Gesamtanzahl von Lösch-Operationen (engl. *deletions*), Einfüge-Operationen (engl. *insertions*), und Ersatz-Operationen (engl. *substitutions*) berechnet. Bei Lösch-Operationen wird das Referenzwort nicht aligniert, bei Einfüge-Operationen wird ein resultierendes bzw. übersetztes Wort nicht aligniert und bei Ersatz-Operationen wird das Referenzwort mit dem übersetzten Wort aligniert (vgl. Dorr et al. 2011:760).

Bilingual Evaluation Understudy (BLEU, Papineni et al. 2002) stellt den aktuellen Standard zur automatischen MÜ-Evaluierung dar, aber auch zur Optimierung der Systemparameter. Die Verwendung mehrfacher Referenzen gilt als Hauptmerkmal der Metrik (vgl. Dorr et al. 2011:760, 762). Der BLEU-Wert einer Übersetzung wird berechnet durch die Anzahl der Wortsequenzen (engl. *n-grams*) in den maschinellen Resultaten, die in den Referenzübersetzungen zu finden sind. Das heißt, dass die BLEU-Metrik nach maximaler Genauigkeit (engl. *precision*) sucht, indem sie die Korrektheit der Resultate misst, anstatt derer

kompletten Übereinstimmung mit den Referenzen (vgl. Dorr et al. 2011:761). Wie von Papineni et al. (2002) erläutert, ist das Ziel der Metrik eine schnelle, kostengünstige und sprachunabhängige (automatische) Evaluierung (2002:311). Verschiedene „perfekte“ Übersetzungen des gleichen Ausgangssatzes zeigen Unterschiede in Wortwahl und -stellung, dementsprechend besteht die Aufgabe von BLEU in der Berechnung von den Matches nach einem Vergleich der N-Gramme in der Übersetzung und in ihrer Referenz (2002:312). Allerdings führt die Auswahl unterschiedlicher Referenzwörter für das gleiche Quellwort zu Schwierigkeiten, weil bei der Übersetzung alle oder mehrere Referenzwörter abgerufen werden könnten. Dies hätte eine schlechte Übersetzung zur Folge (vgl. Papineni et al. 2002:314f.).

Diese Metrik wird noch heute von einigen Forscher*innen kritisiert. Koehn und Monz (2006) berichten, dass eine Korrelation zwischen den BLEU-Bewertungen und den menschlichen Bewertungen schwierig zu bestätigen ist. Sie gehen davon aus, dass BLEU sehr empfindlich in Bezug auf fachsprachliche Wörter ist und deswegen keine Synonyme betrachtet (2006:109). Diesbezüglich stellt BLEU nach Dorr et al. (2011:762) keine semantische Qualität fest, weil die semantischen Fehler nicht beachtet werden (vgl. Way 2018:168). Dies widerspricht der anfänglichen Behauptung von Papineni et al. (2002:317), die hingegen über eine hohe Korrelation informieren. Wie Koehn und Monz (2006), auch Callison-Burch et al. (2006) verdeutlichen in ihrer Studie, dass BLEU keine oder kaum Korrelationen mit den menschlichen Bewertungen gewährleisten kann. Sie zeigen, dass BLEU zahlreiche Variationen von identisch bewerteten MÜ-Resultaten erlaubt und dass hohe BLEU-Bewertungen keine Garantie für eine qualitativ hochwertigere Übersetzung sind (vgl. Callison-Burch et al. 2006:249). In anderen Worten bedeutet eine bessere bzw. höhere BLEU-Bewertung nicht zwangsläufig eine effektive Verbesserung der Übersetzungsqualität (2006:255). In der gleichen Studie heben Callison-Burch et al. (2006:255f.) hervor, dass sowohl BLEU als auch andere Evaluierungsmetriken, wie METEOR, für bestimmte Aufgaben geeignet oder ungeeignet sein können.

METEOR (*Metric for Evaluation of Translation with Explicit ORdering*, Banerjee & Lavie, 2005) wurde eingeführt, um die Schwächen von BLEU zu überwinden (vgl. Banerjee & Lavie, 2005:66f.; vgl. Dorr et al. 2011:762). Im Gegensatz zu BLEU, ist METEOR stimmbar (engl. *tunable*) und Alignment-orientiert, da nur die einzelnen Wörter (engl. *unigrams*) zwischen MÜ und menschlichen Referenzübersetzungen verglichen und aligniert werden. Der Alignment-Prozess findet auf drei Ebenen statt: gleiche Wörter, Wörter mit denselben morphologischen Wurzeln und Synonyme werden aligniert. METEOR berechnet Genauigkeit und Trefferquote (engl. *recall*) und kombiniert sie (2011:762f.). Dazu berechnet die Metrik die

Wortstellung der alignierten Wörter (vgl. Banerjee & Lavie, 2005:65). Im Vergleich zu BLEU kann METEOR mehrere Referenzübersetzungen nicht direkt verwenden, aber kann die Referenzübersetzung für jedes Segment auswählen, das die beste METEOR-Bewertung erzielt (vgl. Dorr et al. 2011:763).

Translation Error Rate oder *Translation Edit Rate* (TER, Snover et al., 2006) misst die minimale Anzahl der menschlichen Editing-Operationen (engl. *shifts*) zur Änderung der maschinellen Resultate, um einen hundertprozentigen Match mit der Referenzübersetzung zu bekommen (2006:223). TER korreliert einigermaßen gut mit den Humanbewertungen, aber erzielt eine Überbewertung der eigentlichen Übersetzungsfehlerquote (2006:230). In ihrem Vergleich der BLEU-, METEOR und TER-Metrik gegen menschliche Bewertung von *Adequacy* und *Fluency* heben Snover et al. (2006) hervor, dass BLEU und TER in Bezug auf einzelne Sätze mit den menschlichen Bewertungen analog korrelieren, obwohl TER ein Viertel der Referenzübersetzungen im Verhältnis zu BLEU benötigt, um gute Ergebnisse zu erzielen (2006:228ff.).

Shterionov et al. (2018) haben das phrasenbasierte und das neuronale System von KantantMT analysiert, indem sie die entsprechenden Qualitätsbewertungen von BLEU und TER mit der menschlichen Evaluierung verglichen haben. Ihre Ergebnisse zeigen, dass beide Metriken nicht immer mit der NMÜ-Qualität übereinstimmen, die von den Expert*innen definiert wird. Sie zeigen hingegen, dass diese Metriken die NMÜ-Qualität unterbewerten, obwohl TER sich der menschlichen Evaluierung annähert (vgl. Shterionov et al. 2018:233, 232). Dazu unterstützen die Studienergebnisse die Hypothese, dass N-Gramm-basierte Metriken mit menschlichen Bewertungen der SMÜ-Qualität besser korrelieren als mit menschlichen Bewertungen der NMÜ-Qualität (vgl. Shterionov et al. 2018:233; vgl. Way 2018:169). Way (2018) erklärt dazu, dass Evaluierungsmetriken besser mit mehrfachen menschlichen Referenzübersetzungen funktionieren, trotzdem wird meistens lediglich eine einzelne Referenzübersetzung bereitgestellt (2018:166).

2.3.2.2. Quality Estimation Ansätze

Die menschliche Wahrnehmung und daher auch die Definition von Qualität sind subjektiv (vgl. Garvin 1984:32). Aus diesem Grund wird *Quality Estimation* auch als subjektive Aufgabe betrachtet. Zwecks einer Steigerung der Objektivität in der Qualitätsevaluierung, kann eine QE-Aufgabe für einen spezifischen Anwendungsfall definiert werden (vgl. Avramidis 2019:3). Die verschiedenen möglichen Aufgaben werden daher festgelegt für die Bewertung von „wahrgenommener *Adequacy*“ oder „Sinn-Behaltung“, „Post-Editing-Zeit“, „Post-Editing-

Maß“ oder „Post-Editing-Aufwand“, um einen numerischen Wert zu bekommen: einen Binärwert (annehmen/ablehnen), eine echte Nummer oder ein relatives Ranking (2019:3). Der Fokus von QE liegt somit auf den einzelnen Übersetzungen (vgl. Specia & Shah, 2018:202), wobei qualitative Merkmale identifiziert werden, die als Indikatoren einer guten Übersetzung gelten können (vgl. Avramidis 2016:147).

Wie von Avramidis (2019) erläutert, versteht man unter *Confidence Estimation* (CE, Blatz et al. 2004) ein Bewertungsprozess des ganzen Ablaufs des Übersetzungsprozesses, da das MÜ-System selbst gewöhnlich einen Wert generiert, der als Vertrauensindex (engl. *confidence*) für seine Übersetzung gilt (vgl. Specia et al. 2010:44; vgl. Avramidis 2019:3). Trotzdem ist CE eine Teilmenge von QE, weil QE zahlreiche Qualitätsindikatoren enthalten kann, darunter manchmal auch systeminterne Informationen (2019:3).

Wie bereits von Macketanz et al. (2018) erklärt, würde eine detailgetreue qualitative Evaluierung mit Bezug auf die linguistischen Aspekte zu einem besseren Verständnis der maschinellen Fehler und zur Verbesserung der MÜ-Systeme führen. Kern dieser Art von Evaluierung sind Test Suites: ausgedachte Korpora aus Testsätzen zur Prüfung linguistischer Phänomene und Aspekte. Erwähnenswert bei diesem letzten Punkt ist, dass Test Suites im Rahmen der NMÜ-Evaluierung erneut Verwendung finden, denn die NMÜ-Resultate sind qualitativ sehr hochwertig und benötigen detailgetreue Analysen (vgl. Macketanz et al. 2018:578). Als Beispiel, in der Studie von Burlot et al. (2018) werden Test Suites verwendet, um die morphologischen Fertigkeiten von MÜ- und NMÜ-Systemen auf Satzebene und in verschiedenen Sprachkombinationen kontrastiv zu evaluieren. Laut ihren Ergebnissen verfügen die Systeme, die eine gute Leistung in Bezug auf globale QE erbringen, auch über gute morphologische Fertigkeiten (2018:563). In einer anderen Studie, führen Guillou et al. (2018) eine detaillierte Evaluierung unterschiedlicher Arten von Pronomen in den NMÜ-Resultaten mithilfe von Test Suites durch. Nach Guillou et al. (2018) vereinfachen Test Suites den Evaluierungsprozess, da sie sowohl eine quantitative Evaluierung als auch eine qualitative Analyse der Herausforderungen zugleich ermöglichen (vor allem in Bezug auf die NMÜ) (2018:570f.). Ihre Studie bestätigt, dass die automatische Evaluierung durch Metriken im Vergleich zu QE ein irreführendes Bild des Verhaltens einiger Systeme offenlegt. Insbesondere NMÜ-Systeme gehen gut mit dem „Modellieren“ des Kontexts für die Übersetzung von pleonastischen Pronomen um, während anaphorische Pronomen aktuell noch einen schwierigen Fall darstellen (vgl. Guillou et al. 2018:575). Avramidis et al. (2018:246f.) heben in ihrer Schlussfolgerung die Nützlichkeit der Verwendung von Test Suites hervor, sogar im Bereich

der Evaluierung von QE-Systemen selbst. Sie argumentieren, dass eine linguistisch motivierte Evaluierung zur Weiterforschung und Verbesserung von QE führen kann.

Der Vertrauensgrad (engl. *confidence*) beim Quantifizieren der Qualität wird in der Forschung hinterfragt, da sich die Bewertungsprognose in den meisten QE-Aufgaben auf eine einzelne Übersetzung bezieht (vgl. Avramidis 2019:4). Dementsprechend wird von Avramidis (2019) ein weiterer Ansatz zur QE vorgestellt, und zwar *Comparative Quality Estimation*, der auf der automatischen Analyse von zwei oder mehreren MÜ-Übersetzungen und auf der Bewertung ihres Vergleiches basiert. *Comparative Quality Estimation* hat zum Ziel, das maschinelle Ranking verschiedener maschineller Resultate eines einzigen Ausgangssatzes zu erzielen, in Bezug auf die vergleichende Qualität der Resultate (2019:4). Dieses maschinelle Ranking wird mithilfe eines statistischen Klassifikationsmodells durchgeführt, das mit bestehenden menschlichen Rankings und zahlreichen qualitativen Kriterien eines Textes trainiert wird (2019:4). Obwohl sich die Studie von Avramidis (2019) nicht direkt auf NMÜ bezieht, könnte der Ansatz auch für die NMÜ relevant sein.

Zwecks *Quality Estimation* sind viele Systeme und Open-Source-Frameworks entwickelt worden, um die Qualität eines Übersetzungssystems zu evaluieren (vgl. Kepler et al. 2019:117). Unter den letzten Entwicklungen befindet sich OpenKiwi, ein Open-Source-Framework für QE für das gemeinsame Training und Testen von QE-Systemen auf Wort- und Satzebene. QE auf Wortebene weist jedem maschinell übersetzten Wort oder Ellipsen Qualitäts-Etiketten zu, während QE auf Satzebene darauf abzielt, die Qualität des ganzen übersetzten Satzes zu prognostizieren (2019:117f.). Unter den von OpenKiwi implementierten QE-Systemen sind folgende neuronale Systeme zu finden: QUETSCH (*Quality Estimation from scraTCH*, Kreuzer et al. 2015) für QE auf Wortebene, NuQE (*NeUral Quality Estimation*, Martins et al. 2016) auch für QE auf Wortebene und Predictor-Estimator (Kim et al. 2017) auf Satz-, Wort- und Phrasenebene. Laut Kepler et al. (2019:121) hat OpenKiwi andere Open-Source-Toolkits sowohl auf Wort- als auch auf Satzebene überholt.

2.4. MTranslatability

Wie bereits im Kapitel 2.3. erwähnt, variiert die Qualität maschineller Resultate abhängig vom System, Sprachpaar und von der Domäne und ist in ihrem Rohzustand kaum „perfekt“ (vgl. Bernth & Gdaniec 2001:175). Um eine höhere Übersetzungsqualität zu erzielen, würden manchmal wenige korrektive Maßnahmen genügen (2001:175):

[I]f only the writer had been aware of the limitations of current MT systems and had created the document with “MTranslatability” in mind. [...] Increasing awareness of how to write for MT should contribute to better user acceptance of MT once the users learn a few tricks that will enhance the output quality. (Bernth & Gdaniec 2001:175)

Der Begriff von „MTranslatability“ wurde von Bernth und Gdaniec (2001) eingeführt, um die Übersetzbarkeit eines Textes in Bezug auf die MÜ (also die „MÜbersetzbarkeit“) zu bezeichnen (2001:175). Bei der Studie wird eine systematische Analyse der Probleme verbunden mit MÜ und ihrer Systeme durchgeführt (2001:175). Bernth und Gdaniec (2001) erörtern in ihrer Studie, dass online verfügbare Dokumentation hochwahrscheinlich maschinell übersetzt werden kann, insbesondere von Benutzer*innen, die über kein oder kaum ein Verständnis der Quellsprache verfügen. Aus diesem Grund stellt laut Bernth und Gdaniec (2001) eine sinnvolle Entscheidung seitens der Autor*in dar, schon vom Anfang an, einen Text für seine künftige maschinelle Übersetzung (in vielen Sprachen) zu schreiben, damit keine zusätzlichen Kosten im folgenden Übersetzungsprozess entstehen (2001:176). Dies ist die Grundidee bei der Entwicklung von kontrollierten Sprachen (engl. *Controlled Languages*, CLs), deren Ansatz auch zu MTranslatability beiträgt, ist aber systemspezifisch im Vergleich zu den Vorschlägen von Bernth und Gdaniec (2001:176). Genauer gesagt, haben Bernth und Gdaniec (2001) einige Regeln zur Verbesserung von MTranslatability identifiziert. Besondere Aufmerksamkeit werden Grammatik, Ambiguität, Stil, Zeichensetzung, Rechtschreibung und Mark-Up in Bezug auf die englische Sprache gewidmet (2001:177). Im Rahmen dieser Arbeit haben vor allem die syntaktischen, semantischen, Ambiguitäts- und Stilfehler Relevanz, da hier erwähnenswert ist, dass sich die Vorschläge von Bernth und Gdaniec (2001) nicht auf die NMÜ beziehen, was hingegen bei dieser Arbeit der Fall ist. Mit MTranslatability ist der Begriff von *Confidence Estimation* (CE) verbunden und beide sorgen nach Stymne und Smith (2012) für eine bessere Lesbarkeit der Texte. Der Unterschied zwischen CE und MTranslatability besteht darin, dass MTranslatability als das Einschätzen im Vorhinein gilt, ob ein Text oder ein Satz einfach oder schwer maschinell übersetzbar sind. Hingegen findet die Evaluierung bei CE nach der Übersetzungsphase statt (vgl. Stymne & Smith 2012:71). MTranslatability wird vor allem bezüglich regelbasierter MÜ-Systeme angewandt, wobei die MÜ-Resultate durch das

Umschreiben des Quelltextes aufgrund bestimmter Regelsätze einer kontrollierten Sprache verbessert werden können (vgl. Stymne & Smith 2012:71).

Besonders relevant für diese Arbeit ist die Studie von Roturier (2004), wo die Effektivität eines ausgewählten Regelsatzes von kontrollierten Sprachen getestet wird. Dazu wird versucht, die häufig erscheinenden linguistischen Probleme in den MÜ-Resultaten zu isolieren und es wird ein „Pre-Post-Editing-Prozess“ vorgeschlagen, um den Post-Editing-Aufwand zu reduzieren (2004:12). Die Vorschläge von Roturier (2004) haben sich als sehr hilfreich für die Auswahl der anzuwendenden Regeln für diese Arbeit erwiesen, obwohl hier der Fokus von den kontrollierten Sprachen abweicht. Hier einige der angewandten Regeln der kontrollierten Sprache, die auch für die vorliegende Arbeit von Nutzen sind:

Repeat the head noun with conjoined articles and prepositions. (37)

Make sure that every segment can stand alone syntactically. (27)

Be very careful with the -ing words: If it is a gerund, use an article in front of it. (14). If it is introducing a new clause, use 'by' in front of it. (23). If it is modifying a noun in a non-finite clause, replace it with a relative clause. (16)

(Roturier 2004:11)

Die Regeln hängen tatsächlich davon ab, welches MÜ-System verwendet wird und müssen an den Ausgangstext angepasst werden (vgl. Roturier 2004:4f.). Das Ziel dieser Arbeit ist die Verbesserung maschineller Resultate und daher die Verringerung der Ambiguitäten im Ausgangstext. Bei MTranslatability und bei einer kontrollierten Sprache handelt es sich um bestimmte lexikalische, syntaktische oder semantische Beschränkungen (vgl. Roturier 2004:2) pro Pre-Editing-Methode.

Unter den erkannten Problemen der MÜ-Systeme kommen laut Bernth und Gdaniec (2001) folgende vor: fehlender Kontext, fehlendes Weltwissen und die fehlende Fähigkeit der Systeme die Mehrdeutigkeiten durch Kontext und Weltwissen zu vermeiden. Dazu tauchen auch die Fehler seitens der Autor*innen auf, die laut Bernth und Gdaniec (2001) mit den oben erwähnten Mängeln weiter erforscht werden müssen (2001:206). Im Folgenden ihre Vorschläge zur MTranslatability:

A. Summary of Rules

1. Avoid ungrammatical constructions.
2. Repeat final words of the left conjunct or initial words of the right conjunct, as necessary, to disambiguate the coordination.
3. Use articles with *ing*-words when they are used as nouns; or use infinitives instead of *ing*-words, depending on what you mean.

4. Rewrite *ing*-words that follow an object as a relative clause or add a suitable preposition, depending on what you mean.
5. complements of other verbs.
6. Do not omit relative pronouns; write *that* (*which*, *who*, etc.) explicitly.
7. Avoid post-modifying adjective phrases.
8. Minimize use of personal pronouns.
9. Always write the complementizer *that* explicitly.
10. Avoid long noun phrases, if possible.
11. Always write *in order to* before an infinitive in a purpose clause instead of just *to*.
12. Use one-word verbs instead of verb + particle whenever possible.
13. Avoid overly long sentences and very short sentences.
14. Avoid metaphors, idioms, slang, and dialect.
15. Avoid ellipsis.
16. Avoid passive constructions, if possible.
17. Make sure that each segment can stand alone syntactically.
18. Avoid footnotes in the middle of a segment, and make footnotes independent segments.
19. Do not include parenthesized expressions in a segment unless the segment is still valid syntactically when you remove the parentheses while leaving the parenthesised expressions.
20. Use punctuation prudently.
21. Avoid using (*s*) to indicate plural.
22. Avoid using / as in *and/or* and *user/system*.
23. Check your spelling.
24. Proofread and correct scanned documents.
25. Avoid textual content in bitmaps.
26. Use mark-up wisely.

(Bernth & Gdaniec 2001:209f.)

2.5. Pre-Editing

In den letzten fünfzehn Jahren haben sich Forscher*innen mit Pre-Editing und vor allem mit Post-Editing beschäftigt, um den Einfluss von MÜ auf die Übersetzer*innen selbst und auf ihre Arbeitsumgebung zu verstehen (vgl. Guerberof Arenas 2019:1). Der Fokus der vorliegenden wissenschaftlichen Arbeit liegt insbesondere auf Pre-Editing-Prozessen. Pre-Editing bedeutet die Verwendung von bestimmten Regeln, damit typographische und inhaltliche Fehler korrigiert werden können. Diese Regeln tragen dazu bei, kürzere Sätze zu schreiben, spezifische grammatikalische Strukturen zu gebrauchen und bestimmte Entscheidungen zu treffen, was die Semantik bzw. eine konsistente Terminologie angeht (vgl. Guerberof Arenas 2019:3). Verschiedene Studien heben hervor, dass Pre-Editing in Bezug auf spezifische Sprachen und Systeme den Post-Editing-Aufwand reduzieren kann. Jedoch gilt es zu erwähnen, dass eine Implementierung dieser Änderungen nicht immer einfach ist und in manchen Fällen trotzdem noch Post-Editing erfordert (2019:20). Allerdings wird der Post-Editing-Aufwand in der im Folgenden durchgeführten empirischen Studie nicht einfließen, da die Arbeit ausschließlich auf den Einsatz von Pre-Editing-Methoden fokussiert.

In Bezug auf die NMÜ argumentiert Guerberof Arenas (2019), dass sich NMÜ noch in ihren Anfangszeiten befindet. Die NMÜ-Ergebnisse sind zwar ermutigend, aber im Vergleich zu SMÜ lassen sich viele Inkonsistenzen feststellen, vor allem bezüglich Post-Editing und menschlicher Evaluierung. Die Inkonsistenzen betreffen Sprachkombinationen, Domänen und die unvorhersehbare Segmentverarbeitung, die gravierende Fehler wie Auslassungen, Hinzufügungen und Fehlübersetzungen enthalten können (vgl. Castilho et al. 2017a:118; vgl. Guerberof Arenas 2019:25). Die NMÜ-Fehlertypologie, der Umgang mit langen Sätzen und Terminologie, die Integration mit CAT-Tools, der Einsatz bei ressourcenarmen Sprachen müssen nach Guerberof Arenas (2019) weiter erforscht werden. Heute bedarf es noch menschlicher Eingriffe, um Übersetzungen menschlicher Qualität zu generieren (2019:25).

2.5.1. Tools für Pre-Editing

Auf dem Markt sind verschiedene Rechtschreibprüfung-, Grammatikprüfung-, Annotations-, Style Checker- sowie Controlled Language Checker-Tools verfügbar (vgl. Bernth & Gdaniec 2001:195ff.). Darunter befindet sich beispielsweise **EasyEnglishAnalyzer** (EEA) (Bernth 1997). Dabei handelt es sich um ein Autorensystem (engl. *authoring tool*), das die Ambiguität und Komplexität von Texten hervorhebt, wobei es den Autor*innen hilft, maschinell übersetzbare (engl. *MTranslatable*) Texte zu erstellen (vgl. Bernth & Gdaniec 2001:199). EEA überprüft infinite Verbalphrasen, mehrdeutige, lange und unvollständige Sätze und grammatikalische Basis-Regeln. Eine Clarity Index-Funktion fasst die Fehler im ganzen Ausgangstext zusammen und weist ihnen einen Stellenwert zu, welcher die MTranslatability des Textes definiert (2001:199f.). Das Tool schlägt mögliche automatische Korrekturen vor, die ausgewählt werden können und misst die Gesamtqualität des Textes (vgl. Bernth 1999:o. S.).

Unter anderen Tools befindet sich das online verfügbare **ReEscreve** (Barreiro & Cabral 2009), das für das Paraphrasieren im Bereich Pre-Editing für menschliche und maschinelle Übersetzung der portugiesischen Sprache geeignet ist (vgl. Barreiro & Cabral 2009:o. S.). In der hier erwähnten Studie ist ReEscreve nicht nur ein Authoring-Tool, sondern dient mit seiner Pre-Editor-Funktion zur Vereinfachung der MÜ. Es bietet eine stilistische Hilfe an, aber es reduziert die Ambiguität und die unnötige Komplexität von Ausgangstexten, deswegen ist es ein nützliches Tool für Übersetzer*innen (2009:o. S.). ReEscreve basiert auf großen Vokabularen und auf einem Set an Grammatiken und paraphrasiert ausschließlich Verbkonstruktionen. Das linguistische Wissen rund um Verbkonstruktionen kann für das

Paraphrasieren angewandt werden und qualitativ hochwertigere maschinelle Resultate erzielen (vgl. Barreiro & Cabral 2009:o. S.).

SPIDER (Barreiro 2011) findet auch Anwendung im Bereich Pre-Editing für die MÜ. Das linguistische Wissen des Tools (Vokabular und Grammatik) wird für das Paraphrasieren von Texten auf der Wort- und Phrasenebene (Mehrwort-Einheiten) angewandt. Unterschiedliche linguistische Aspekte werden vom Tool abgedeckt, Terminologie ist auch inkludiert. SPIDER wurde von ReEscribe direkt entwickelt, ist aber für viele Sprachen geeignet (vgl. Barreiro 2011:365, 367). SPIDER hat mit der Übersetzbarkeit direkt zu tun, da es den Text vorbereitet, damit die Übersetzung eine bessere Qualität aufzeigt (2011:367). Dieses Tool ermöglicht linguistisch komplexe Veränderungen eines Textes, wobei die Grammatik und der semantische Zusammenhang des Textes aufrecht erhalten bleiben. Das Tool schlägt Wörter und Ausdrücke im Kontext vor, die von den Benutzer*innen ausgewählt werden können (vgl. Barreiro 2011:367f.).

Ein automatisches Pre-Editing-Modul wurde von Yoshimi (2001) an einem MÜ-System hinzugefügt, als allgemeines Framework für die Verbesserung der Qualität der Texte (vgl. Yoshimi 2001:249). Dieses Modul weist viele Vorteile auf, indem es neue Quellsätze neu schreibt, unabhängig von den internen Informationen des Parsers. Das heißt, dass die Hinzufügung des automatischen Pre-Editing-Moduls kaum oder keine Änderungen an den bestehenden Systemteilen vornimmt. Insbesondere formuliert Yoshimi Regeln zur Umschreibung von Schlagzeilen in der Sprachkombination Englisch-Japanisch mit befriedigenden Ergebnissen (2001:249).

2.6. Best-Worst-Scaling

Die gängigsten aktuellen Metriken zur automatischen Evaluierung der MÜ wurden bereits in Kapitel 2.3.2. vorgestellt. Wie bereits erwähnt, hat sich in den letzten Jahren herauskristallisiert, dass die automatischen Metriken eigentlich nicht in der Lage sind, die Unterschiede und die Fortschritte in der Leistung der jüngsten MÜ-Systeme adäquat zu reflektieren, vor allem, was die NMÜ-Systeme angeht (vgl. Tavosanis 2009:o. S.). Die Ergebnisse der Metriken korrelieren kaum mit den menschlichen Bewertungen, die in Bezug auf die Methodenauswahl den Bezugspunkt bilden. Dementsprechend ist das Ziel der MÜ-Systeme, maschinelle Übersetzungen zu erzielen, die laut menschlichen Evaluator*innen bzw. Leser*innen über ein befriedigendes Qualitätsniveau verfügen (2009:o. S.).

Das Gesetz der vergleichenden Beurteilung wurde von Thurstone (1927a, 1927b) in Form einer Gleichung formuliert, als neues psychophysisches Gesetz zur Messung psychologischer Werte (vgl. Thurstone 1994:266, 269). Das Gesetz ist nicht nur für den Vergleich von physischen Intensitätsstimuli anwendbar, sondern auch für qualitative vergleichende Beurteilungen von Einzelphänomenen. Eine Angabe zum Objekt der Beurteilung bzw. des Vergleiches ist die erste Voraussetzung. Jedem Exemplar bzw. jeder Einzelprobe wird ein quantitatives oder qualitatives Merkmal zugewiesen, welches das Exemplar positiv oder negativ beurteilt: Grauwert, Gewicht oder Trefflichkeit. Das zugewiesene Merkmal definiert das „psychologische Kontinuum“ (engl. *psychological continuum*) für ein spezifisches Projekt (1994:266). Laut Thurstone (1994):

As we inspect two or more specimens for the task of comparison there must be some kind of process in us by which we react differently to several specimens, by which we identify the several degrees of excellence or weight or grey value in the specimens. [...] [I]t will be called here [...] *the discriminial process*¹¹ [...]. (Thurstone 1994:266)

Das Gesetz trifft auf Beurteilungen einer einzelnen Beobachter*in zu, die eine Serie von Stimuli paarweise vergleicht (engl. *paired comparison*), wobei die Elemente im Paar nicht ident beurteilt werden dürfen (vgl. Thurstone 1994:267). Durch das Gesetz werden die verglichenen Stimuli in das Kontinuum verteilt, indem der erste Stimulus „besser“ als der zweite ist (1994:269).

Die Best-Worst-Scaling-Methode (BWS, Louviere und Woodworth 1990) wurde entwickelt, damit die Proband*innen die „beste“ und „schlechteste“ Option aus einem (Auswahl-)Set auswählen konnten (vgl. Louviere et al. 2015:3). Diese Methode ermöglicht die Sammlung zusätzlicher Auswahlinformationen im Rahmen von Auswahlstudien. Vor allem ermöglicht diese Methode die Entnahme von Informationen über die weniger anziehenden Auswahloptionen und über die Funktion des angegebenen Werts der Proband*in (2015:3f.). Anders gesagt, ermöglicht die Methode ein besseres Verständnis des Prozesses, bei dem ein Individuum eine beste und schlechteste Auswahl trifft (2015:11f.). Die Methode fand in den Jahren nach der Entwicklung immer mehr Beachtung, da die Methode das Potential hat, einige der Probleme der Bewertungsskalen zu überwinden (2015:4). Das Interesse an der BWS-Methode nahm in den Bereichen Marketing und Gesundheitsökonomie zu, aber wurde parallel auch zur Persönlichkeits- und Wertemessung und in der Lebensmittel- und Weinforschung angewandt (2015:4f.). Nach derzeitigem Kenntnisstand der Verfasserin wurde die Best-Worst-

¹¹ Kursivschrift im Original.

Scaling-Methode noch nie im Bereich der Translationswissenschaft für die Qualitätsevaluierung maschineller Übersetzung verwendet.

Die BWS-Methode basiert auf der Grundidee, dass Individuen bestimmte Entscheidungen treffen, wenn sie mit einem Set von drei oder mehreren Gegenständen oder Optionen konfrontiert werden. Daher können sie die beste (engl. *best*) und die schlechteste (engl. *worst*) Option in dem Set auswählen. Hier stehen die zwei Begriffe metaphorisch für die geeigneten Begriffe, die die Endpunkten eines subjektiven Kontinuums definieren (vgl. Louviere et al. 2015:6). Hier ist aber erwähnenswert, dass „beste“ und „schlechteste“ Option den Begriffen von „akzeptabel“ (engl. *acceptable*) und „inakzeptabel“ (engl. *unacceptable*) nicht gleichgesetzt werden (2015:6).

Gemäß Louviere et al. (2015:12): „If one wants to summarize some sample or population of people, one must make various strong assumptions that may or may not be true in order to aggregate the choices of these individuals.“. Dieser Satz fasst zusammen und gibt im Voraus bekannt, was das Ziel und die Forschungsmethodik dieser Arbeit betrifft. Alle Best-Worst-Scaling-Modelle sind probabilistische diskrete Auswahlmodelle, deren Basis die von Thurstone (1927b) eingeführte Zufallsnutzentheorie bildet (engl. *random utility theory*) (vgl. Louviere et al. 2015:12). Der probabilistische Charakter der Auswahl kann zweifach interpretiert werden. Entweder sind die Forscher*innen nicht in der Lage, alle Auswahlfaktoren zu verstehen bzw. zu kennen, oder die Auswahl eines Individuums ist fehlerhaft oder inkonsistent, also variabel (2015:12).

Wie bereits erwähnt, weisen Bewertungsskalen einige Probleme auf, darunter niedrige Inter- und Intrarater-Reliabilität und eine feste Granularität (vgl. Kiritchenko & Mohammad 2017:465). Als Lösung für die Probleme und als direkter Vergleich mit der Bewertung, schlagen Kiritchenko und Mohammad (2017) die BWS-Methode als eine Variante der vergleichenden Annotationsmethode vor. Die Annotator*innen bzw. Proband*innen werden jeweils n -Optionen – das heißt ein n -Tupel, wo $n > 1$ und im Regelfall $n = 4$ – vorgelegt und werden gefragt, die beste und die schlechteste Option auszuwählen. Alle zu bewertenden Optionen werden in einem Set an m 4-Tupel (wo $m \geq N$ und N ist die Optionenanzahl) unterteilt, sodass jede Option innerhalb der verschiedenen Tupel mehrmals evaluiert wird. Nach der Annotation der Tupel wird die Punktezahl jeder Option einfach gezählt und jede Option wird eingeordnet (vgl. Kiritchenko & Mohammad 2017:466). In ihrer Studie über die Gefühlsintensität schlussfolgern Kiritchenko und Mohammad (2017:469), dass BWS viel zuverlässigere Ergebnisse im Verhältnis zur herkömmlichen Skalenbewertung erzielt, auch im

Fälle linguistischer Komplexität in den Optionen, zum Beispiel bei Negationen und Modalverben.

Die Frage der Reliabilität der BWS-Methode stellen sich auch Burton et al. (2019) in ihrer psychologischen Studie zur Messung subjektiver erster Gesichtseindrücke. Im Regelfall werden sie mithilfe von Likert-Skalen gemessen, aber die alternative BWS-Methode überholt die Likert-Skala auf individueller Ebene, indem BWS die Präferenzen der Proband*innen besser vorhersagt und eine bessere Test-Testwiederholungs-Reliabilität erzielt (vgl. Burton et al. 2019:1, 9).

3. Aktueller Stand der Forschung

In diesem Kapitel wird der aktuelle Stand der Forschung in Bezug auf Pre-Editing präsentiert. Pre-Editing stellt den Kern dieser Masterarbeit dar und nimmt eine wichtige Rolle bei dem Einsatz maschineller Übersetzung ein. Laut Porsiel (2020) erfordert MÜ tatsächlich „den Aufbau neuer bzw. die Anpassung alter Prozessketten sowie den Aufbau entsprechend qualifizierten Fachpersonals: in der Regel erfahrene Pre- und Post-Editor[*innen] sowie Computerlinguist[*innen]“ (Porsiel 2020:8). Dazu behauptet Porsiel (2020), dass Pre- (und Post-) Editing nahezu mehr Wichtigkeit als die Auswahl eines MÜ-Systems haben (vgl. Porsiel 2020:4).

Zu den Studien über den Einsatz von Pre-Editing in Bezug auf MÜ zählt die Studie von Babych et al. (2009). Mit dem Ziel, die MÜ-Qualität von Funktionsverben (engl. *light-verb construction*, LVCs) zu optimieren, schlagen sie eine menschliche Evaluierung zur systematischen Ermittlung fehlübersetzter Kontexte vor. Des Weiteren soll die Studie zur Erstellung automatischer Pre-Editing-Regeln und zu derer Funktionalitätsförderung führen (2009:36). Die Studie von Babych et al. (2009) legt ihren Fokus auf regelbasierte MÜ (Systran 5.0 und ProMT 8.0) und auf die Sprachpaare Französisch-Englisch und Russisch-Englisch. In der Studie wird auf Pre-Editing viel Wert gelegt, als eine Methode zur Förderung der Verständlichkeit der MÜ-Resultate, indem die vorhandenen Ressourcen des Systems benutzt werden können, ohne das System verändern zu müssen (2009:36f.). Im MÜ-System wird die Integration der Fähigkeit vorgeschlagen, Quellsätze mit existierenden Tools und Listen von Worten zu umschreiben, die als unübersetzbar gelten. Umschreiben bietet die Möglichkeit, zum Beispiel Wortstellung und Mehrwortkomposita, im Voraus auszubessern (vgl. Babych et al. 2009:36). Was den empirischen Teil ihrer Studie betrifft, trägt eine anfängliche Analyse zur Verständlichkeit der Funktionsverben bei, die meistens problematischen Konstruktionen zu identifizieren und auszuwählen (2009:36f.). Die Studie stützt sich auf MTranslatability (Bernth & Gdaniec 2001), die im Kapitel 2.4. beschrieben wurde und die nach Babych et al. (2009) bessere Anwendung als kontrollierte Sprachen finden kann (2009:37). Die Evaluator*innen sind englische Muttersprachler*innen (Studierende des Masterstudiums Translation), denen der Quelltext nicht gezeigt wird. Die Ergebnisse zeigen einen Rückgang der unverständlichen maschinellen Übersetzungen in die englische Sprache aus dem Russischen und Französischen um 40% (2009:36, 39).

Seretan et al. (2014) haben Pre-Editing-Technologien entwickelt, um den Quelltext (benutzergenerierte Inhalte, engl. *user-generated content*, UGC) zu korrigieren und die Übersetzbarkeit zu steigern. Im Rahmen der Studie sind linguistisch fundierte Pre-Editing-Regeln für Englisch und Französisch im Gesundheits- und technischen Bereich entwickelt worden (vgl. Seretan et al. 2014:1793). Ihr Projekt zielt darauf ab, die Qualität statistischer maschineller Übersetzung von nutzergenerierten Inhalten wie Foren, Blogs und soziale Medien zu verbessern, da solche Plattformen eine leichte Zugänglichkeit zu Informationen und einen schnellen Austausch ermöglichen. Die Übersetzung dieser Art von Inhalten bereitet einige Schwierigkeiten aufgrund des informalen und umgangssprachlichen Stils, Abkürzungen, spezifischer Terminologie und der Verwendung unregelmäßiger Grammatik und Rechtschreibung (2014:1793). Die von Seretan et al. (2014) verwendete Technologie basiert auf den Tools und Ressourcen von Acrolinx-Lingware, wobei die Textanpassung durch Rechtschreib-, Grammatik-, Terminologie- und Style Checker-Funktionen unterstützt wird. Die Regeln zur Korrektur werden komplett neu manuell definiert oder auf der Basis von bereits existierenden Acrolinx-Regeln adaptiert und können automatisch oder manuell angewandt werden (2014:1794f.). Das manuelle Pre-Editing wird von Muttersprachler*innen (Studierende des Masterstudiums Translation) durchgeführt, die für die Aufgabe präzise Anweisungen erhalten (2014:1795f.). Die Annotator*innen haben das nicht-präeditierte und das präeditierte maschinelle Resultat mit einer 5-Punkte Likert-Skala evaluiert (vgl. Seretan et al. 2014:1796). Ihre Ergebnisse zeigen, dass der Einsatz von Pre-Editing-Strategien zu qualitativ hochwertigeren SMÜ-Resultaten führt (2014:1798).

In einer ähnlichen Arbeit besteht das Ziel von Gerlach (2015) darin, die Möglichkeit zur Verbesserung der SMÜ-Übersetzung von französischen IT-Foren durch die Anwendung von Pre-Editing-Regeln zu untersuchen. Genauer gesagt, versucht Gerlach (2015:o. S.) die nützlichen Veränderungen zu identifizieren, welche auch automatisch angewandt werden können, um die MÜ zu verbessern. Die Evaluierungsmethode ist ein menschlicher Vergleich durch Crowdsourcing. Die Resultate der Evaluierung des zeitlich und technisch bedingten Post-Editing-Aufwands zeigen, dass Pre-Editing die maschinellen Resultate maßgeblich verbessert. Außerdem ist ein interaktives Pre-Editing auch den Benutzer*innen sehr zugänglich, wobei weniger Verbesserungen produziert werden als bei dem von den Expert*innen durchgeführten Pre-Editing (vgl. Gerlach 2015:o. S.). Um die Portabilität bzw. Übertragbarkeit des entwickelten Pre-Editing-Prozesses abzuschätzen, hat Gerlach (2015) andere regelbasierte MÜ-

Systeme und Foren unterschiedlicher Domänen in der Studie miteinbezogen. Zu ihren Ergebnissen gilt es zu bemerken, dass die Pre-Editing-Regeln gut übertragbar sind.

Von Miyata und Fujita (2017) wird das Potential von Pre-Editing-Strategien für die Sprachkombination Japanisch-Englisch erforscht. Dazu geben sie einen Überblick über die möglichen Pre-Editing-Methoden für die Verbesserung der MÜ-Qualität (2017:47). Mithilfe von menschlichen Editor*innen werden vier verschiedene Datensets in den Sprachpaaren Japanisch-Englisch und Japanisch-Chinesisch und Koreanisch mit einer fünfstelligen Punkteskala evaluiert und präeditiert. Das Ziel ist eine Liste der besten Pre-Editing-Methoden zu erstellen (2017:47f.). Für ihre Studie wird die Typologie von Pre-Edit-Methoden auf aktuellen Pre-Editing-Beispielen (engl. *pre-edit instances*) basieren, das heißt Paare von nicht präeditierten und präeditierten Ausgangstexten (2017:48). Die drei Fachbereiche der Ausgangstext-Sets sind: Krankenhausgespräche, online verfügbare Informationen über japanische Stadtgemeinden und japanische und englische Nachrichtenartikel (2017:49). Die Ergebnisse der Studie zeigen, dass 85% der präeditierten Quelltexte befriedigende maschinelle Resultate erzielen. Außerdem zeigt dies, dass MÜ-Systeme über ein großes Potential verfügen, wo menschliche Eingriffe stattfinden. Allgemein gesagt, werden je länger der Quelltext, desto mehr Pre-Edit-Eingriffe benötigt, um eine gute Qualität der Zieltexte zu erreichen (vgl. Miyata & Fujita 2017:49). Unter den am schwierigsten zu korrigierenden MÜ-Fehlern befinden sich Fehlübersetzungen von Eigennamen und falsche Wortwahl. Allerdings sind diese Fehler nach Miyata und Fujita (2017) nicht in der Pre-Editing-Phase anzusprechen, sondern in der Training- oder Post-Editing-Phase (vgl. Miyata & Fujita 2017:50). Um die Effektivität der Pre-Editing-Strategie innerhalb eines mehrsprachigen Settings zu überprüfen, haben Miyata und Fujita (2017) die japanischen Quelltexte ins Chinesische und ins Koreanische maschinell übersetzen lassen. Die Qualität der Sets wurde von einer Evaluator*in bewertet und als besser eingestuft als die von den Quelltexten, also wurde die Übertragbarkeit der Pre-Edit-Methoden für verschiedene Sprachen geprüft (2017:50). Dennoch „[f]urther scrutiny into the language dependency of machine-translatibility is important to justify the pre-editing approach to other target languages and domains.“ (Miyata & Fujita 2017:50). Demnach ist MTranslatibility sprachabhängig und dies gilt als Basis für bestimmte Pre-Editing-Methoden je nach Zielsprache und Domäne. Die in der Studie betrachteten Pre-Edit-Operationen reichen von der Einfügung bzw. Löschung der Zeichensetzung und Wort-Umordnung (also von der Oberfläche) bis zu syntaktischen Abwandlungen (2017:51).

4. Forschungsmethodik

Im Zuge dieser wissenschaftlichen Arbeit wird qualitativ gearbeitet, mit dem Ziel, qualitative Daten zu erheben. Allerdings werden die Ergebnisse auch quantitativ untersucht. Die erhobenen Daten beziehen sich auf die Evaluierung der Qualität der NMÜ, wobei mögliche Verbesserungsvorschläge der MTranslatability (Bernth & Gdaniec 2001) für eine bestimmte Sprachkombination analysiert und angewandt werden.

Die MTranslatability-Vorschläge, die im Kapitel 2.4. erwähnt wurden, gelten als potenzielle Pre-Editing-Regeln, die für diese Arbeit angewandt werden können. Die Auswahl der anzuwendenden Regeln für die analysierte Textsorte und Sprachkombination basiert auf verschiedenen Fehlertaxonomien, die in einer überarbeiteten MQM-Metrik (Kern) zusammengestellt und angepasst werden. Die gestaltete Fehlertypologie wird auch bei der menschlichen Evaluierung von Nutzen sein. Zu den betrachteten Taxonomien gehören die von Flanagan (1994), Vilar et al. (2006) und Comparin und Mendes (2017). Hier gilt es zu erwähnen, dass die Taxonomien insbesondere von Flanagan (1994) aber auch von Vilar et al. (2006) für die Identifizierung von Fehlern der MÜ des 21. Jahrhunderts anwendbar sind (vgl. Stankevičiūtė et al. 2017:77). Die Gestaltung der passenden Fehlertypologie für diese Arbeit wird in Kapitel 4.2. erklärt.

Angesichts der Tatsache, dass der Grad der Qualität oft schwer definierbar ist, insbesondere durch automatische Evaluierungsmethoden bzw. -softwares, fällt die Entscheidung für die Forschungsmethodik über eine manuelle Auswahl und Vorbereitung bzw. auf ein manuelles Pre-Editing der Ausgangstexte. Bei der empirischen Studie wird auch eine menschliche Evaluierung gewählt. Die empirische Studie wird sich in verschiedene Phasen gliedern: Auswahl der Ausgangstexte und entsprechende Gestaltung einer individuell angepassten Fehlertypologie der Ausgangstexte, Auswahl von Pre-Editing-Methoden auf der Basis der Ausgangstextfehler und entsprechendes manuelles Pre-Editing, maschinelle Übersetzung durch ein NMÜ-System und menschliche Evaluierung durch Best-Worst-Scaling. Darauf folgen die Datenerhebung, Datenauswertung und die finale Diskussion der Ergebnisse. Die Methodik der ganzen empirischen Studie wird in den folgenden Kapiteln ausführlicher vorgestellt.

4.1. Auswahl der Texte

Die Auswahl der Ausgangstexte und -textsorte stellt die erste Phase der empirischen Studie dar. Anstatt wissenschaftlicher Textsorten bzw. Artikel wurden verlässliche englischsprachige Internetartikel bzw. Blogs zum Thema Biolumineszenz (Biologie) gewählt. Die Quellen der ausgewählten Artikeln sind National Geographic Society, TED Conferences, Smithsonian Institution, The Washington Post Company, Hakai Magazine, Encyclopaedia Britannica, Scientific American und American Museum of Natural History (AMNH) (entsprechende Details, Beispiele und Quellen sind in Anhang 1 angeführt). Diese online verfügbaren Quellen sind ausgewählt worden, weil sich die Textsorte für bestimmte Anpassungen bzw. Pre-Editing-Methoden gut eignet. Die ausgewählten Textpassagen sind repräsentativ für die Konventionen der Textsorte (vgl. Roturier 2004:3). Das zusammengestellte Korpus der zu evaluierenden Textpassagen fasst wie bereits erwähnt viele der Komplexitäten der Textsorte zusammen. Darunter befinden sich Mehrwortkomposita, zum Beispiel *glowin-the-dark plankton* (EN)>*glowin-the-dark plancton* (IT) im AT 1 und oft auch Redewendungen, beispielsweise *ball and chain* (EN)>*palla e catena* (IT) im AT 10, die die NMÜ-Software schwer erkennen und daher falsch oder nicht übersetzen kann. Die Schwierigkeiten der Software bei der Übersetzung von Internet- und Blogartikeln hängen auch von der Fachsprachlichkeit des Bereichs Biologie ab. Zum Beispiel wird die Terminologie des Bereichs durch die Software nicht immer optimal abgedeckt, insbesondere in Bezug auf die Gattungsnamen der verschiedenen Fischspezies, die vom Phänomen der Biolumineszenz betroffen werden, zum Beispiel *anglers* (EN)>*angloers* (IT) im AT 10. Außerdem müssen die Kenntnisse über diesen Bereich noch vertieft werden. Selbst beim betrachteten Phänomen gilt es noch viele offene Fragen zu erforschen. Ein weiterer Grund für die Auswahl des Themas Biolumineszenz hat mit der Wichtigkeit der Funktion des Phänomens zu tun. Biolumineszenz, sprich die Fähigkeit von verschiedenen Meeres- aber auch Landtieren, Licht zu erzeugen, trägt hauptsächlich zur Kommunikation zwischen den Lebewesen bei, indem sie viele Lebensbereiche der erwähnten Tiere beeinflusst. Der Meinung der Verfasserin zufolge ist dies ein geeigneter Fachbereich für den Einsatz von Pre-Editing-Methoden und für deren Evaluierung, weil die meisten Texte dieser Textsorte und dieses Fachbereichs nicht für eine eventuelle maschinelle Übersetzung geschrieben wurden. Daher sind deren maschinell erstellten Übersetzungen von einer schlechten Qualität, da sie oft schwer verständlich sind.

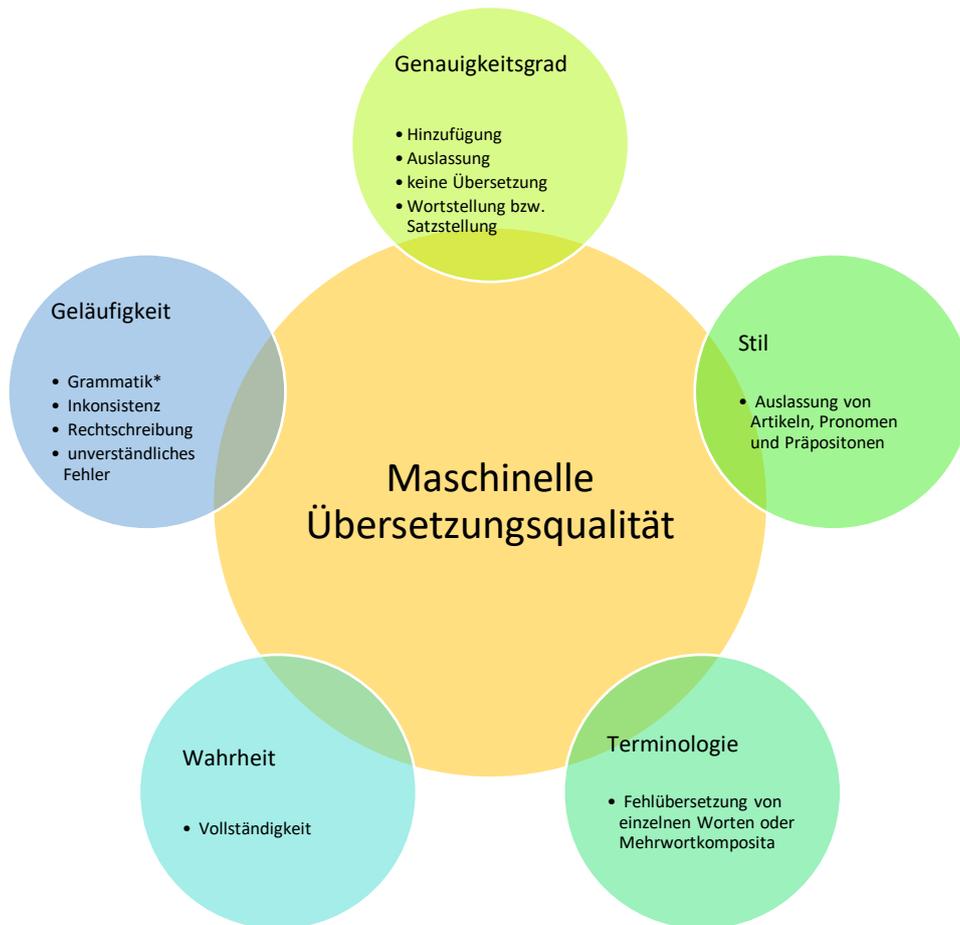
Aus jeder Quelle werden maximal zwei Textpassagen ausgewählt (aus zwei oder drei Sätzen, auf jeden Fall nicht mehr als fünf Sätze), die für die Evaluierung wichtig sind. Insgesamt

werden zehn Ausgangstexte selektiert, da sich zwei davon im Laufe der Evaluierungsphase wiederholen. Im Anhang 1 befinden sich die Quelltexten und die entsprechenden Textpassagen (unterstrichen).

4.2. Fehlertypologie

Als Basis für die Gestaltung der für die vorliegende Arbeit passenden Fehlertypologie (siehe Abbildung 4, S. 53) gilt der Kern der MQM-Metrik. Die Felder von MQM Design, grammatikalisches Register und Typographie (artverwandt mit Geläufigkeit), Locale, Rechtsvorschriften und Gebietsschema-Konventionen (artverwandt mit Wahrheit) werden von der neu erstellten Typologie ausgelassen. Besonders erwähnenswert ist die ausgelassene Kategorie von Design bzw. Formatierung und Kodierung (engl. *Formatting and Encoding*, Comparin & Mendes 2017:o.S.), die auch von Flanagan (1994:67) als Satzgrenze (engl. *Clause Boundary*) bezeichnet wird. Damit werden Mark-Up und Segmentierungsfehler bezeichnet, die im Rahmen dieser Arbeit nicht als zu evaluierende Fehler zu betrachten sind. Die Verfasserin selbst hat eventuelle Fehler bei der Extraktion der Sätze während der Verwendung der Übersetzungssoftware korrigiert, damit die Software die Sätze richtig verarbeiten konnte. Diese Vorgehensweise soll den Fokus auf die gewählten Fehlerkategorien legen. Die anderen oben erwähnten Kategorien grammatikalisches Register und Typographie werden ausgelassen, weil in den für diese Arbeit spezifisch ausgewählten Textpassagen keine Fehler im Verhältnis zu diesen Kategorien gefunden wurden. Die Kategorien Locale, die auch Comparin und Mendes (2017) in ihrer Liste von Kategorien als „falsche Sprachvarietät“ (engl. *Wrong Language Variety*) verwendet haben, Rechtsvorschriften und Gebietsschema-Konventionen wurden von der neuen Fehlertypologie ausgeschlossen, da die ausgewählte Textsorte und Textpassagen zu keinem bestimmten Übersetzungsprojekt mit Anforderungen gehören.

Die verschiedenen Taxonomien von Flanagan (1994), Vilar et al. (2006) und Comparin und Mendes (2017) werden in der MQM-Metrik integriert. Insbesondere die von Flanagan (1994:67) vorgeschlagenen Kategorien werden unter den Oberkategorien Geläufigkeit (Grammatik) und Stil hinzugefügt. Von dieser Taxonomie werden die Kategorien Akzent (engl. *Accent*), Großschreibung (engl. *Capitalization*) und Kategoriefehler (engl. *Category*) ausgeschlossen, weil sie sich im Zuge der Sprachkombination Englisch-Italienisch als nicht wichtig erwiesen haben. Zeichensetzung ist die einzige Kategorie von Vilar et al. (2006:699), die ausgeschlossen wird.



*falsche Flexion von Verben, Substantiven, Adjektiven oder Adverbien, fehlende Negation, falsche Pronomen, Artikel oder Präpositionen, Fehler bei Konjunktionen, fehlende Subjekt-Verb- und Substantiv-Adjektiv-Kongruenz

Abbildung 4: Angepasste Fehlertypologie für die vorliegende wissenschaftliche Arbeit

4.3. Pre-Editing-Methoden

Nach der Text- und Fehlerauswahl besteht die dritte Phase in der Auswahl der Pre-Editing-Methoden für die englischen Ausgangssätze. Die Auswahl basiert auf den Fehlern, die in den ausgewählten Textpassagen gefunden werden. Dies hat zum Ziel, einige Verbesserungen an den Ausgangstexten vorzuschlagen, um bessere bzw. weniger mehrdeutige maschinelle Zieltex-te in der italienischen Sprache zu erhalten.

Praktischer Einsatz der erwähnten Pre-Editing-Methoden bedeutet das tatsächliche manuelle Präeditieren der Textpassagen durch eine einzelne Pre-Editing-Methode oder eine Kombination aus zwei Methoden.

4.3.1. Auswahl der Methoden

Eine Liste von vier einzelnen Pre-Editing-Methoden und zwei Kombinationen aus zwei Pre-Editing-Methoden wird erstellt, auf der Basis der Vorschläge von Bernth und Gdaniec (2001) und von Roturier (2004). Die Regeln werden nach den Anforderungen der ausgewählten Texte angepasst (vgl. Roturier 2004:5). Die Pre-Editing-Methoden beziehen sich auf die Sprachkombination Englisch-Italienisch. Im Folgenden die Liste der angepassten Pre-Editing-Methoden (siehe Tabelle 1, S. 54):

Tabelle 1: Liste der angewandten Pre-Editing-Methoden

Pre-Editing-Methode 1	Koordinationen (bzw. Sätze) eindeutig machen, Hauptsubstantiv mit verbundenen Artikeln und Präpositionen wiederholen
Pre-Editing-Methode 2	Metaphern und Redewendungen vermeiden bzw. umschreiben
Pre-Editing-Methode 3	Keine Verben mit Partikel, sondern Verben aus einem Wort verwenden
Pre-Editing-Methode 4	Ing-Wörter umschreiben: mit einem Relativsatz ersetzen oder eine passende Präposition hinzufügen (wie „by“ zum Beispiel)
Pre-Editing-Methode 5K (Kombination)	Verwendung von (Possessiv)Pronomen vermindern + jedes Segment muss selbstständig syntaktisch stimmen
Pre-Editing-Methode 6K (Kombination)	Lange Nominalsätze vermeiden + mit Relativpronomen umschreiben

Die erste Pre-Editing-Methode (1) stammt aus der Zusammenfügung von Regel 2 („Repeat final words of the left conjunct or initial words of the right conjunct, as necessary, to disambiguate the coordination“, Bernth & Gdaniec 2001:180) und Regel 37 („Repeat the head noun with conjoined articles and prepositions“, Roturier 2004:11). Die Pre-Editing-Methode 2 ist von Regel 14 („Avoid metaphors, idioms, [...]“, Bernth & Gdaniec 2001:188) übernommen worden und Methode 3 spiegelt Regel 12 wider („Use one-word verbs instead of verb+particle whenever possible“, 2001:187). Die Pre-Editing-Methode 4 basiert auf Regel 4 („Rewrite *ing*-words that follow an object as a relative clause or add a suitable preposition, depending on what you mean“, Bernth & Gdaniec 2001:183) aber auch auf Regel 23 und 16 („Be very careful with the *-ing* words: [...] If it is introducing a new clause, use 'by' in front of it (23). If it is modifying a noun in a non-finite clause, replace it with a relative clause (16)“, Roturier 2004:11). Die letzten zwei Methoden sind Kombinationen aus Methoden, die nicht in den oben erwähnten

Methoden gehören. Die Pre-Editing-Methode 5K fügt Regel 8 („Minimize use of personal pronouns“, Bernth & Gdaniec 2001:186) und 17 („Make sure that each segment can stand alone syntactically“, 2001:191) zusammen. Regel 17 ist auch in Roturier (2004) zu finden, wird aber als Regel 27 bezeichnet („Make sure that every segment can stand alone syntactically“, 2004:11). Die Pre-Editing-Methode 5K betrifft jedoch keine Personalpronomen, sondern Possessivpronomen, die die Verständlichkeit der betrachteten Texte beeinträchtigen. Die Methode 6K ist auch eine Kombination. Sie kombiniert Regel 10 („Avoid long noun phrases, if possible“, Bernth & Gdaniec 2001:187) mit Regel 6 („Do not omit relative pronouns; write *that* (*which*, *who*, etc.) explicitly“, 2001:184), obwohl diese letzte angepasst wird, da die Relativpronomen benötigt werden, um die Nominalsätze umzuschreiben.

4.3.2. Praktischer Einsatz der Methoden

Die vorliegende empirische Studie setzt voraus, dass jede englische Textpassage durch drei unterschiedliche Pre-Editing-Methoden oder auch Kombinationen daraus präeditiert wird. Jedes Edit oder jede Kombination aus Edits wird einzeln pro Ausgangstext angewandt. Daraus wird höchstwahrscheinlich eine Übersetzung resultieren, die sich von den anderen italienischen Zieltexten des gleichen Ausgangstextes unterscheidet. Im Anhang 2 befinden sich alle Textpassagen mit den jeweiligen angewandten Pre-Edits und entsprechenden Zieltexten bzw. Übersetzungen ins Italienische, je nach Methode.

4.4. Maschineller Übersetzungsprozess

Der maschinelle Übersetzungsprozess stellt die fünfte Phase dieser empirischen Studie dar. Nach der menschlichen Text-, Fehler- und Pre-Editing-Methodenauswahl und dem manuellen bzw. menschlichen Präeditieren der ausgewählten Sätze, liegt ein maschineller Übersetzungsprozess dieser Arbeit zugrunde. Der Prozess besteht aus dem Übersetzen von bestimmten Textpassagen aus dem Englischen ins Italienische. Dieser Prozess dient dem Zweck, in einem nächsten Schritt die Resultate menschlich zu evaluieren.

4.4.1. Neuronale Übersetzungssoftware: SYSTRAN Translate

Der Hauptgrund zur Auswahl von der online verfügbaren neuronalen Übersetzungssoftware SYSTRAN Translate (SYSTRAN Translate Tool 2020) für diese wissenschaftliche Arbeit ist

die hohe Leistungsfähigkeit bei der Übersetzung in vielen Bereichen. Allerdings gibt die Software Raum für einige Verbesserungen. Die technischen Merkmale der Software sind schon im Kapitel 2.1.2.3. erläutert worden. Im Folgenden werden die Weboberfläche und Arbeitsweise der Software gezeigt, mit praktischen Beispielen (siehe Abbildung 5, S. 56):

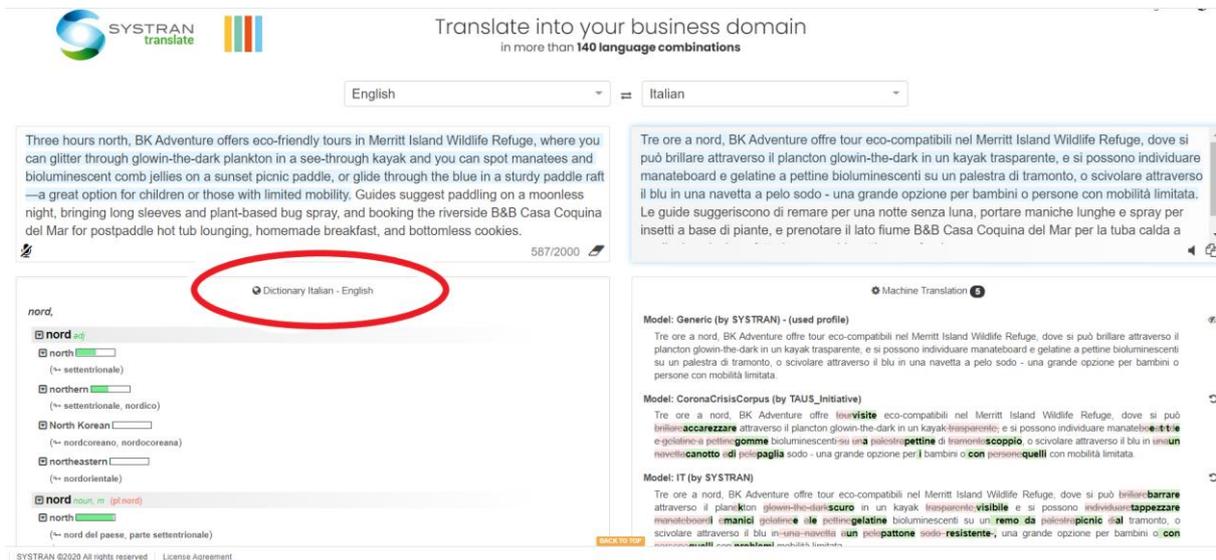


Abbildung 5: Screenshot der Weboberfläche von SYSTRAN Translate, praktisches Beispiel (SYSTRAN Translate Weboberfläche 2020)

Wie auf dem oberen Screenshot zu sehen ist, bietet die Software SYSTRAN Translate auf der linken Seite der Ergebnisse eine Vorschau des Benutzerwörterbuchs EN>IT oder IT>EN (mit einem roten Kreis markiert) an, jeweils auf Wunsch der Benutzer*innen. Besonders relevant für die vorliegende Studie ist, dass die Software auf der rechten Seite der Ergebnisse unterschiedliche Übersetzungen für jeden eingegebenen Satz anbietet (auf dem unteren Screenshot mit einem roten Kreis markiert). Genauer gesagt, jedem zu übersetzenden Satz werden bis zu vier Übersetzungsvarianten mit markierten Änderungen von der Software vorgeschlagen, die die Benutzer*innen entsprechend ihrer Präferenzen auswählen können (siehe Abbildung 6, S. 57):

Nella speciale mostra "Oceani invisibili", i visitatori si presentano in una brillante spirale "da pavimento a soffitto" di modelli di biofluorescenza sottomarina proprio come farebbero nella vita. Tra i pesci di squalo, di razze e di altri pesci che riflettono i verdi e le baccello di neon vi è un rettile luminoso: una tartaruga di mare di Hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, che attualmente è elencata come gravemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non avrebbe fatto il taglio.

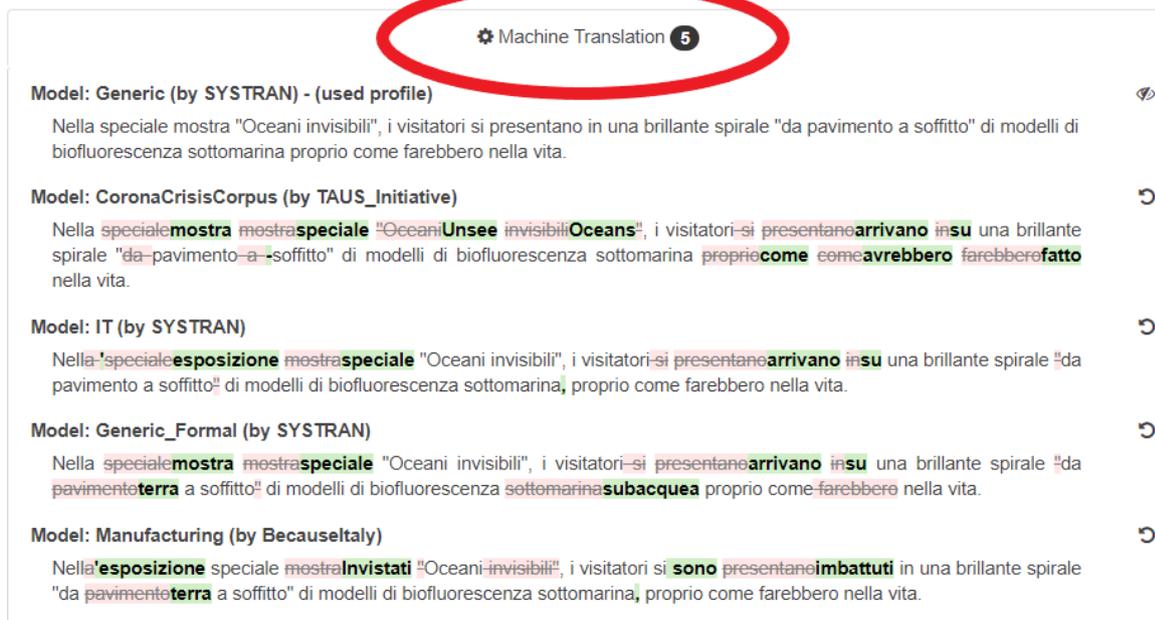


Abbildung 6: Screenshot der möglichen Übersetzungsvarianten von SYSTRAN Translate (SYSTRAN Translate Weboberfläche 2020)

Shoshan (2019) kann an dieser Stelle direkt zitiert werden, da er in einem online Beitrag über das Thema folgendes berichtet:

[...] the average NMT quality across several sentences, even in the same paragraph, will almost always look bad because there is high quality variance. That is, when running NMT on a set of sentences, the first sentence may have a perfect machine translation, the second will require some human post-editing, and the third will need to be retranslated by a human translator from scratch. (Shoshan 2019:o.S.)

Bei der praktischen Anwendung der Software SYSTRAN Translate können Shoshans (2019) Worte bestätigt werden. Dazu bereitet die Software einige Schwierigkeiten bei der Auswahl der Änderungen bzw. der einzelnen Übersetzungen selbst, wenn mehrere Sätze zu übersetzen sind. Die Auswahl anderer Übersetzungsvarianten gestaltet sich schwerfällig und jede neue Änderung an der maschinellen Übersetzung wird meistens automatisch zurückgesetzt. Das

heißt, dass nur eine Satzauswahl pro Zieltext möglich ist, um die Zurücksetzung anderer Sätze zu vermeiden. Im Rahmen dieser Arbeit hat die Verfasserin jene Übersetzungsvariante pro Satz für jeden Zieltext ausgewählt, die sich ihrer Meinung nach einer menschlichen Übersetzung annähern, das heißt die Variante, die sie als „beste“ erachtet hat.

4.5. Evaluierung durch Best-Worst-Scaling

In der letzten Phase wird die Best-Worst-Scaling-Methode zur menschlichen Evaluierung angewandt. Nach derzeitigem Kenntnisstand der Verfasserin wurde das Best-Worst-Scaling noch nie im Bereich der Translationswissenschaft für die Qualitätsevaluierung maschineller Übersetzung verwendet. Diese Methode ermöglicht eine menschliche Bewertung der Qualität der betrachteten maschinellen Übersetzungen in der Sprachkombination Englisch-Italienisch. Für diese Studie wird eine menschliche Evaluierung als richtig empfunden. Wie von Porsiel (2020) bezeichnet „Der Mensch wird im Zentrum dieser Prozesskette stehen (müssen), um weiterhin Qualität gewährleisten zu können. Denn nur Menschen können Übersetzungsqualität schnell und zweifelsfrei beurteilen“ (Porsiel 2020:9).

Im Zuge dieser Studie werden die Zieltexte von fünf Proband*innen bewertet. Die Proband*innen haben Italienisch als Erstsprache und Englisch als Arbeitssprache oder C1-Stufe des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GERS) und haben das Masterstudium Translationswissenschaft schon absolviert oder werden dies in Kürze absolvieren. Sie werden ihre Präferenzen für bestimmte Übersetzungen bekunden. Die Proband*innen sollen mindestens über Grundkenntnisse in dem Fachbereich der Texte verfügen oder sollen sich in den Fachbereich Biologie, insbesondere in das Thema Biolumineszenz einlesen, um das Verstehen der Ausgangstexte zu erleichtern.

Das Best-Worst-Scaling stellt die Methode der Datenerhebung dieser Arbeit dar. Die Proband*innen werden für jeden Ausgangstext vier Zieltexte bzw. Übersetzungen vorgelegt bekommen, aber nur zwei davon werden jeweils bewertet, weil jede Proband*in die beste und die schlechteste Übersetzung der vier auswählen wird. Dazu werden die Proband*innen angeben, wie gut die zwei Zieltexte auf einer Skala von 0 bis +4 (sehr gut) und wie schlecht auf einer Skala von -4 (vollkommen inakzeptabel) bis 0 sind. Eine 0-Bewertung wird als neutral betrachtet. Qualität ist in diesem Kontext ein Vergleichswert und keine tatsächliche qualitative Bewertung der Übersetzungen, sondern eine relative Bewertung im Vergleich zu den anderen Übersetzungen.

Für jeden Ausgangstext wird Folgendes zur Bewertung vorgelegt: der Ausgangstext selbst, eine maschinelle Übersetzung, deren AT nicht präeditiert worden ist und drei maschinelle Übersetzungen des gleichen Textes, der nach drei unterschiedlichen Pre-Editing-Methoden präeditiert worden ist. Die sechs Pre-Editing-Methoden, darunter zwei Kombinationen, werden gleichmäßig aufgeteilt, damit jede sechs Mal vorkommt und daher potenziell sechs Bewertungen bekommen wird (siehe Tabelle 2, S. 60). Einige Pre-Editing-Methoden könnten sogar mehrere Bewertungen erhalten, da nur der beste und der schlechteste Zieltext pro Ausgangstext berücksichtigt wird.

Um die Konsistenz der Antworten zu testen und zu wahren, wird eine Kombination Ausgangstext-Zieltexte und daher auch die angewandten Pre-Editing-Methoden zwei Mal vorkommen. Das heißt, dass sich AT 1, entsprechende Zieltexte 1.1, 1.2, 1.3, 1.4 und Pre-Editing-Methoden 1, 6K (Kombination, ab hier als K bezeichnet) und 2 wiederholen. AT 8 und AT 1 werden identisch sein. Außerdem wird sich ein weiterer Ausgangstext wiederholen: AT 4 und AT 12 werden gleich sein. Dementsprechend werden die Proband*innen in der Praxis zwölf Texte erhalten. Insofern kann getestet werden, ob die Proband*innen dieselben Antworten zwei Mal geben und ihre Einschätzungen konsistent sind.

Als Tool für das Best-Worst-Scaling wird eine online verfügbare Software verwendet. LimeSurvey (LimeSurvey Homepage 2020) ermöglicht die Erstellung von online Umfragen und eine einfache Erhebung der Daten. Die Daten können in der Software gespeichert und in verschiedene Formate exportiert werden, zum Beispiel Word, Excel und XML. Die Proband*innen werden ihre Bewertungen intuitiv und ohne Zeitbegrenzung angeben, die erhobenen Daten werden im Folgenden anonymisiert.

Tabelle 2: Einteilung der AT und ZT für das Best-Worst-Scaling

Ausgangstext (AT)	Pre-Editing nach Methode	MÜ	Zieltext (ZT)
AT 1	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 1 3. AT mit Pre-Editing-Methode 6K 4. AT mit Pre-Editing-Methode 2	Systran Traslate	→ZT 1.1 →ZT 1.2 →ZT 1.3 →ZT 1.4
AT 2	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 1 3. AT mit Pre-Editing-Methode 4 4. AT mit Pre-Editing-Methode 5K		→ZT 2.1 →ZT 2.2 →ZT 2.3 →ZT 2.4
AT 3	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 1 3. AT mit Pre-Editing-Methode 4 4. AT mit Pre-Editing-Methode 5K		→ZT 3.1 →ZT 3.2 →ZT 3.3 →ZT 3.4
AT 4	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 4 3. AT mit Pre-Editing-Methode 1 4. AT mit Pre-Editing-Methode 6K		→ZT 4.1 →ZT 4.2 →ZT 4.3 →ZT 4.4
AT 5	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 3 3. AT mit Pre-Editing-Methode 4 4. AT mit Pre-Editing-Methode 6K		→ZT 5.1 →ZT 5.2 →ZT 5.3 →ZT 5.4
AT 6	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 3 3. AT mit Pre-Editing-Methode 4 4. AT mit Pre-Editing-Methode 6K		→ZT 6.1 →ZT 6.2 →ZT 6.3 →ZT 6.4
AT 7	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 2 3. AT mit Pre-Editing-Methode 1 4. AT mit Pre-Editing-Methode 5K		→ZT 7.1 →ZT 7.2 →ZT 7.3 →ZT 7.4
AT 8	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 1 3. AT mit Pre-Editing-Methode 6K 4. AT mit Pre-Editing-Methode 2		→ZT 8.1 →ZT 8.2 →ZT 8.3 →ZT 8.4
AT 9	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 5K 3. AT mit Pre-Editing-Methode 2 4. AT mit Pre-Editing-Methode 3		→ZT 9.1 →ZT 9.2 →ZT 9.3 →ZT 9.4
AT 10	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 5K 3. AT mit Pre-Editing-Methode 2 4. AT mit Pre-Editing-Methode 3		→ZT 10.1 →ZT 10.2 →ZT 10.3 →ZT 10.4
AT 11	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 3 3. AT mit Pre-Editing-Methode 4 4. AT mit Pre-Editing-Methode 5K		→ZT 11.1 →ZT 11.2 →ZT 11.3 →ZT 11.4
AT 12	1. AT ohne Pre-Editing 2. AT mit Pre-Editing-Methode 6K 3. AT mit Pre-Editing-Methode 3 4. AT mit Pre-Editing-Methode 2		→ZT 12.1 →ZT 12.2 →ZT 12.3 →ZT 12.4

Zwecks der Kalkulation des Endergebnisses für jede Pre-Editing-Methode werden die durch die Bewertungen erhaltenen Gesamtpunkte addiert. Somit werden die Ergebnisse zeigen, welche Pre-Editing-Methoden die meisten positiven Bewertungen und daher den höchsten positiven Wert erzielen und welche die meisten negativen Bewertungen also den niedrigsten negativen Wert aufweisen. Beide positive und negative Ergebnisse sind wissenschaftlich relevant, denn sie heben hervor, welche Pre-Editing-Methoden die MTranslatability und daher auch die Qualität der neuronalen maschinellen Übersetzung für das Sprachpaar Englisch-Italienisch verbessern. Die erhaltenen Ergebnisse werden zudem zeigen, welche Methoden für die bereits erwähnte Sprachkombination weniger oder nicht relevant sind bzw. keine Verbesserung der NMÜ ermöglichen. Auf diese Weise kann die Richtung der Tendenzen in Bezug auf die Forschungsfrage der vorliegenden Arbeit gezeigt werden. In Kapitel 5. werden die Ergebnisse der empirischen Studie veranschaulicht.

5. Ergebnisse der Datenerhebung

Für die Datenerhebung durch das Best-Worst-Scaling wurde LimeSurvey (LimeSurvey Homepage 2020) verwendet. Aus Gründen der Klarheit werden im Folgenden (siehe Tabelle 3, S. 62) die Ergebnisse präsentiert, die die Bewertungen aller Proband*innen aufzeigen, um ihre Unterschiede und Ähnlichkeiten klarer festzulegen. Die positiven Bewertungen sind in der untenstehenden Tabelle 3 grün markiert, während die negativen rot markiert sind. Die neutralen Bewertungen (Null-Bewertungen) sind grau angestrichen. Die hellblau-markierten Zellen stellen die gleichen Ausgangs- und Zieltexte dar. Alle Antworten werden hier anonym gezeigt:

Tabelle 3: Bewertungen aller Übersetzungen und jeweiligen Pre-Editing-Methoden nach allen Proband*innen

ZT	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4	6.1	6.2	6.3	6.4	7.1
PE	kP	1.	6K	2.	kP	1.	4.	5K	kP	1.	4.	5K	kP	4.	1.	6K	kP	3.	4.	6K	kP	3.	4.	6K	kP
P1			-2	+1		-2	+3			0		-2	+1			0		+3		0		+2		-2	-1
P2		+1	-4			+2		-2	+3			-3	-3			+2	-3	+2			-3	+3			-1
P3		0		-4		-4	+1			0		-4		0		-4		+1	-2			+1	-4		
P4			+2	-3		-2	+3		+3			-1			-1	+1	-2			+3	-3	+3			
P5			+2	-4			+2	-1	+3		-3			-1		+3	-1			+2	-3	+3			

ZT	7.2	7.3	7.4	8.1	8.2	8.3	8.4	9.1	9.2	9.3	9.4	10.1	10.2	10.3	10.4	11.1	11.2	11.3	11.4	12.1	12.2	12.3	12.4	
PE	2.	1.	5K	kP	1.	6K	2.	kP	5K	2.	3.	kP	5K	2.	3.	kP	3.	4.	5K	kP	6K	3.	2.	
P1		+1		0		-3			0	+3	0			-2		+2			-3		+1		-3	
P2	+3				+1	-4			-1	+3	+1			-3	-2	+3					+1	-2		
P3	0		-4		0		-4			+3	-1			0	-4		+1		-4		-4		0	
P4	+3		-1			+2	-3	-1		+3			+2		-4	-2	+3			-2			+3	
P5	+3		-1			+2	-3		-1		+3			+3	-4	-3	+2						-3	+3

Die oberste Zeile in der Tabelle zeigt die Zieltextnummer, auf die die gleich untenstehenden Pre-Editing-Methoden (farbig markiert) angewandt worden sind. Die Abkürzung PE steht für Pre-Editing(-Methode) und kP steht für kein Pre-Editing. Das heißt, dass die die Zieltexte bzw. maschinelle Übersetzungen sind, deren Ausgangstext nicht präeditiert worden ist. Jede Proband*in ist mit einem P und einer Erkennungsnummer bezeichnet. Da fünf Proband*innen an der Umfrage teilgenommen haben, sind die Proband*innen als P1, P2, P3, P4 und P5

bezeichnet und werden ab hier dementsprechend benannt. Die erzielten Gesamtbewertungen jeder Proband*in für jede Pre-Editing-Methode werden in der folgenden Tabelle (siehe Tabelle 4, S. 63) zusammenfassend dargestellt:

Tabelle 4: Gesamtpunkte pro Pre-Editing-Methode

PE	kP	1.	2.	3.	4.	5K	6K
P1	0	-1	-2	+8	+3	-5	-6
P2	-8	+4	+3	+6	0	-6	-5
P3	0	-4	-5	-2	-5	-12	-8
P4	-7	-3	0	+2	+3	0	+8
P5	-4	0	+2	+1	-2	-3	+9
Gesamtpunkte	-19	-4	-2	+15	-1	-26	-2

Alle Bewertungen bzw. gegebenen Punkte jeder Proband*in und Pre-Editing-Methode werden addiert. Das eventuelle Dividieren der Gesamtpunkte durch die Anzahl der erhaltenen Bewertungen wird in diesem Fall nicht benötigt, da die Endergebnisse bereits bei der Beobachtung der Daten deutlich gezeigt wurden.

Die Konsistenz der Antworten hat in drei Fällen gefehlt (siehe Tabelle 5, S. 64). Die ersten zwei Fälle betreffen Proband*in 1 (violett markiert), die die sich wiederholenden Zieltexte (1.3 und 1.4) unterschiedlich bewertet hat, da 8.3 und 8.4 negativere Bewertungen bekommen haben. Der dritte Fall betrifft Proband*in 5 (orange markiert), die Zieltext 1.4 und 8.4 jeweils mit -4 und -3 bewertet hat. Die Endergebnisse der mit diesen Übersetzungen assoziierten Pre-Editing-Methoden wären trotzdem negativ ausgefallen. Sie hätten keine große Differenz in den Gesamtbewertungen dargestellt.

Tabelle 5: Inkonsistente Bewertungen Proband*in 1 und 5

ZT	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4
PE	kP	1.	6K	2.	kP	1.	6K	2.
P1			-2	+1				
P2								
P3								
P4								
P5			+2	-4				

ZT	8.1	8.2	8.3	8.4
PE	kP	1.	6K	2.
P1				
P2	+3		+1	-4
P3				
P4				
P5				

Von Tabelle 6 (siehe unten, S. 64) ist es ersichtlich, dass die Pre-Editing-Methode 3 („Keine Verben mit Partikel, sondern Verben aus einem Wort verwenden“) die am meisten positiven Bewertungen und daher den höchsten positiven Wert (+15) erzielt hat. Die Proband*innen haben am meisten die italienischen Zieltexte als „beste“ Übersetzungsvariante ausgewählt, deren englische Ausgangstexte durch die Methode 3 präeditiert worden waren.

Tabelle 6: Gesamtbewertungen Pre-Editing-Methode 3

PE	kP	1.	2.	3.	4.	5K	6K
P1	0	-1	-2	+8	+3	-5	-6
P2	-8	+4	+3	+6	0	-6	-5
P3	0	-4	-5	-2	-5	-12	-8
P4	-7	-3	0	+2	+3	0	+8
P5	-4	0	+2	+1	-2	-3	+9
Gesamtpunkte	-19	-4	-2	+15	-1	-26	-2

In Bezug auf die Pre-Editing-Methode 3 haben die einzelnen Bewertungen einen interessanten Fall gezeigt, den es im Zuge dieser Studie zu erwähnen gilt. Im Vergleich zu den Bewertungen für die anderen Zieltexte, deren Ausgangstextpassagen mit Methode 3 präeditiert wurden, gilt es die Bewertungen der Übersetzungen 10.4 und 11.2 (blau markiert, siehe Tabelle 7, S. 65) zu bemerken. Beide Übersetzungen wurden von allen fünf Proband*innen ausgewählt und die Übersetzungsvariante 11.2 wurde von allen als „beste“ wahrgenommen und daher positiv eingeschätzt. Die Übersetzungsvariante 10.4 hingegen wurde als „schlechteste“ Variante ausgewählt, negativ und von drei Proband*innen sogar am negativsten bewertet (-4), obwohl die gleiche Methode 3 auf den Ausgangstext angewandt worden ist.

Tabelle 7: Detail der fünf Bewertungen für 10.4 und 11.2

	10.4	11.2	
1	3	3	4
2	-2	+2	
3	-3	+3	
4	-4	+1	
5	-4	+3	
6	-4	+2	

Um die Gründe für die negative Bewertung des italienischen Zieltextes 10.4 aufzuzeigen, wird untenstehend (siehe Tabelle 8, S. 66) eine Vergleichstabelle angebracht, die der nicht-präeditierte englische AT 10 und seine Direktübersetzung durch die NMÜ-Software zeigt. Die rot markierten Textelemente heben das Verb mit Partikel hervor, das mit Methode 3 präeditiert worden und anschließend mit dem ganzen Text maschinell übersetzt worden ist und ZT 10.4 ergeben hat. Das Verb im Präteritum *led astray* wurde durch ein Verb aus einem Wort (Präteritum *misled*) ersetzt. Beide italienischen Übersetzungsvarianten, und zwar ZT 10.1 und 10.4 haben sich als korrekt erwiesen. In diesem Fall sind die negativen Bewertungen wahrscheinlich nicht auf das Pre-Editing des betrachteten Verbes zurückzuführen, sondern darauf, dass der italienische ZT 10.4 insgesamt von den Proband*innen als die „schlechteste“ der vier auswählbaren Varianten wahrgenommen wurde. Näher wird auf diesen Punkt in Kapitel 6.2. eingegangen.

Tabelle 8: Vergleich Übersetzungsvarianten für AT 10 ohne Pre-Editing und mit Methode 3

AT 10 kein Pre-Editing	ZT 10.1
<p>Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who led travelers astray at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic "ball and chain" relationship idiom, they attach themselves to her body, which eventually absorbs all but their testicles.</p>	<p>Come le mostruose versioni acquatiche della mitologia, la saggezza - figure luminescenti che hanno portato i viaggiatori fuori strada di notte - i pesci di angler di acque profonde tentano la loro preda con luccicanti palline che penzolano dalle bacchette di fronte alla bocca. I maschi di alcuni angloers sono miniscule rispetto alle femmine. Una volta trovata un'ondina con cui accoppiarsi, in un'inversione della misogina relazione "palla e catena", si attaccano al suo corpo, che alla fine assorbe tutti tranne i testicoli.</p>
AT 10 Pre-Editing-Methode 3	ZT 10.4
<p>Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who miled travelers at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic "ball and chain" relationship idiom, they attach themselves to her body, which eventually absorbs all but their testicles.</p>	<p>Come le mostruose versioni acquatiche della mitologia, la saggezza - figure luminescenti che hanno fuorviato i viaggiatori di notte - i pesci angler di acque profonde tentano la loro preda di luccicanti panettoni che penzolano dalle bacchette di fronte alla bocca. I maschi di alcuni angloers sono miniscule rispetto alle femmine. Una volta trovata un'ondina con cui accoppiarsi, in un'inversione della misogina relazione "palla e catena", si attaccano al suo corpo, che alla fine assorbe tutti tranne i testicoli.</p>

Tabelle 9 (siehe unten, S. 67) zeigt, dass ZT 11.2, und zwar die italienische Übersetzungsvariante, deren englischen AT mit Methode 3 präeditiert worden ist, bessere Bewertungen erzielt, als die Übersetzung (11.1) des nicht-präeditierten Ausgangstextes. Dank dieser Pre-Editing-Methode wurde die Qualität des ZTs 11.2 von den Proband*innen als besser im Vergleich zum ZT 11.1 eingestuft, besonders in Bezug auf das zweite Verb *make the cut* (auch rot markiert), das bei der ersten Übersetzung (11.1) falsch übersetzt wurde, da das Verb auch eine idiomatische Konnotation hat.

Tabelle 9: Vergleich Übersetzungsvarianten für AT 11 ohne Pre-Editing und mit Methode 3

AT 11 kein Pre-Editing	ZT 11.1
<p>In the special exhibition Unseen Oceans, visitors come upon a brilliant floor-to-ceiling spiral of models of underwater species biofluorescing just as they would in life. Among the catsharks, breams, and other fishes reflecting neon greens and reds is a glowing reptile: a hawksbill sea turtle.</p> <p>Only three years ago, this species, Eretmochelys imbricata, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have made the cut.</p> <p>Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, their main food source, from the reef's surface.</p>	<p>Nell'esposizione speciale Invisibili Oceani, i visitatori si sono imbattuti in una brillante spirale "da terra a soffitto" di modelli di biofluorescenza sottomarina, proprio come farebbero nella vita. Tra gli squali, le razze e gli altri pesci che riflettono i verdi e i rossi di neon si trova un rettile splendente: una tartaruga marina di hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, che è attualmente elencata come fortemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non avrebbe fatto il taglio.</p> <p>Le tartarughe marine delle Hawksbill giocano un ruolo essenziale nella vita delle barriere coralline, estraendo dalla superficie della barriera le spugne, la loro principale fonte alimentare.</p>
AT 11 Pre-Editing-Methode 3	ZT 11.2
<p>In the special exhibition Unseen Oceans, visitors encounter a brilliant floor-to-ceiling spiral of models of underwater species biofluorescing just as they would in life. Among the catsharks, breams, and other fishes reflecting neon greens and reds is a glowing reptile: a hawksbill sea turtle.</p> <p>Only three years ago, this species, Eretmochelys imbricata, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have survived.</p> <p>Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, their main food source, from the reef's surface.</p>	<p>Nella mostra speciale Oceani invisibili, i visitatori incontrano una brillante spirale "dal pavimento al soffitto" di modelli di biofluorescenza subacquea proprio come nella vita. Tra gli squali, le razze e gli altri pesci che riflettono i verdi e i rossi di neon si trova un rettile splendente: una tartaruga marina di hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, attualmente elencata come fortemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non sarebbe sopravvissuta.</p> <p>Le tartarughe marine delle Hawksbill giocano un ruolo essenziale nella vita delle barriere coralline, estraendo dalla superficie della barriera le spugne, la loro principale fonte alimentare.</p>

Im Vergleich zu den Resultaten für Pre-Editing-Methode 3, weist Pre-Editing-Methode 5K, und zwar die Kombination von zwei Pre-Editing-Operationen („Verwendung von (Possessiv)Pronomen vermindern + jedes Segment muss selbstständig syntaktisch stimmen“), die negativste Bewertung (-26) auf. Jede Proband*in hat alle Übersetzungen der mit Methode 5K präeditierten Textpassagen als die „schlechtesten“ Varianten ausgewählt und konsequent

negativ bewertet, mit Ausnahme einer einzelnen Bewertung, wie in Tabelle 10 (siehe unten, S. 68) klar dargestellt wird:

Tabelle 10: Gesamtbewertungen Pre-Editing-Methode 5K

PE	kP	1.	2.	3.	4.	5K	6K
P1	0	-1	-2	+8	+3	-5	-6
P2	-8	+4	+3	+6	0	-6	-5
P3	0	-4	-5	-2	-5	-12	-8
P4	-7	-3	0	+2	+3	0	+8
P5	-4	0	+2	+1	-2	-3	+9
Gesamtpunkte	-19	-4	-2	+15	-1	-26	-2

In Bezug auf diese Ergebnisse gilt es hier zu erwähnen, dass die italienischen Zieltexte, bei denen der englische Ausgangstext mit Methode 5K präeditiert wurde, sogar eine schlechtere Gesamtbewertung erzielt hat als die Zieltexte, deren Ausgangstext keine Pre-Editing-Methoden unterzogen worden ist. Dies ist in der untenstehenden Tabelle (siehe Tabelle 11, S. 68) zu sehen:

Tabelle 11: Gesamtbewertungen kein Pre-Editing

PE	kP	1.	2.	3.	4.	5K	6K
P1	0	-1	-2	+8	+3	-5	-6
P2	-8	+4	+3	+6	0	-6	-5
P3	0	-4	-5	-2	-5	-12	-8
P4	-7	-3	0	+2	+3	0	+8
P5	-4	0	+2	+1	-2	-3	+9
Gesamtpunkte	-19	-4	-2	+15	-1	-26	-2

Die Bewertungen der Zieltexte, bei denen der entsprechende englische Ausgangstext keine Pre-Editing-Operationen unterzogen worden ist, werden jedoch im Zuge dieser Arbeit nicht zur Gesamtwertung der Pre-Editing-Methoden hinzugefügt. Ein Grund dafür ist, dass die Forschungsfrage keinen direkten Bezug auf die Qualität der maschinellen Rohübersetzungen ohne präeditierte Ausgangstexte nimmt. Zur Angabe von Tendenzen bezüglich der Forschungsfrage und Stützung der Hypothesen dieser Arbeit, werden ausschließlich die Endergebnisse der angewandten Pre-Editing-Methoden für die Sprachkombination Englisch-Italienisch in Betracht gezogen. Die durchgeführten Vergleiche zwischen den Bewertungen der

Zieltexte, deren Ausgangstexte nicht präeditiert worden sind und den Bewertungen der Zieltexte, deren Ausgangstexte nach bestimmten Pre-Editing-Methoden verändert wurden, haben zum Ziel, die erzielten Gesamtbewertungen der Methoden ausführlich zu erklären.

Pre-Editing-Methode 1 („Koordinationen (bzw. Sätze) eindeutig machen, Hauptsubstantiv mit verbundenen Artikeln und Präpositionen wiederholen“), Pre-Editing-Methode 2 („Metaphern und Redewendungen vermeiden bzw. umschreiben“), Methode 6K („Lange Nominalsätze vermeiden + mit Relativpronomen umschreiben“) und Methode 4 („Ing-Wörter umschreiben: mit einem Relativsatz ersetzen oder eine passende Präposition hinzufügen (wie „by“ zum Beispiel)“) haben auch insgesamt negative Endergebnisse aufgewiesen. Im Laufe der Umfrage wurden sie jedoch nicht immer positiv oder negativ bewertet. Dies bedeutet, dass die englischen Ausgangstexte, die mit den oben erwähnten Methoden präeditiert worden sind, in maschinellen Übersetzungsvarianten resultieren, die manchmal als „beste“ und manchmal als „schlechteste“ Varianten ausgewählt und ausgewertet wurden. Die Methode 1 (Gesamtpunkte -4) hat sich letztendlich in Bezug auf die meisten negativen Bewertungen an der zweiten Stelle platziert, gefolgt von Methode 2 und 6K punktgleich (Gesamtpunkte -2) an dritter Stelle. Methode 4 (Gesamtpunkte -1) hat sich an vierter Stelle der negativ bewerteten Methoden platziert und dies signalisiert, dass diese Methode der „erfolgreicheren“ Pre-Editing-Methode 3 näherkommt, obwohl sie nicht als zweite „beste“ Pre-Editing-Methode eingeschätzt werden darf.

Jede Pre-Editing-Methode ist im Laufe der Studie bzw. der Umfrage sechs Mal aufgetreten. Jede Methode ist sechs Mal auf einen englischen Ausgangstext bzw. Textpassage angewandt worden, um unterschiedliche maschinelle Übersetzungen ins Italienische zu erhalten, die potenzielle bessere Übersetzungen sind und eventuell als „beste“ Varianten ausgewählt werden könnten. Angenommen, dass jede Methode potenziell ein Minimum von sechs Bewertungen erhalten konnte, zeigt das folgende Diagramm (siehe Abbildung 7, S. 70) die Anzahl der erhaltenen Bewertungen pro Pre-Editing-Methode. Die Zieltexte auf deren Ausgangstexte kein Pre-Editing angewandt worden ist, konnten hingegen potenziell zwölf Bewertungen erhalten. Bei jeder Übersetzungsvariantengruppe gab es den maschinellen Zieltext eines nicht-präeditierten ATs. Die Darstellung der Anzahl der Bewertungen der Zieltexte mit nicht-präeditiertem Ausgangstext dient zum Zweck, die Verteilung der Gesamtbewertungen deutlich abzubilden.

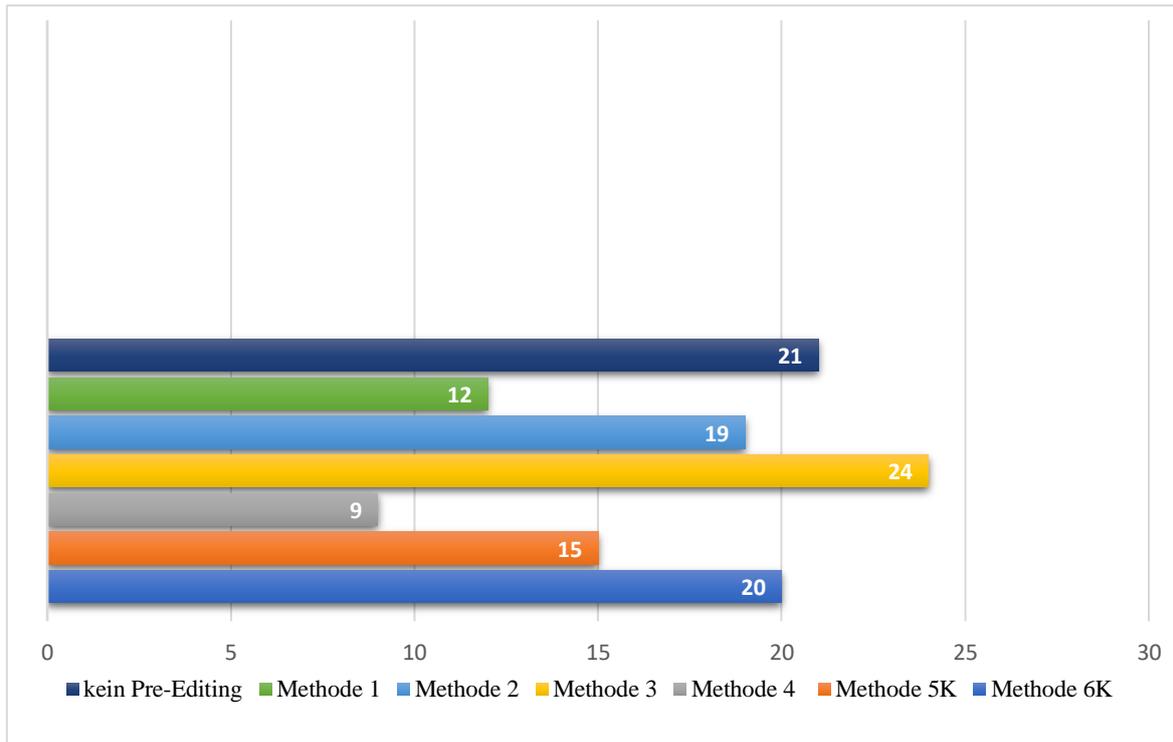


Abbildung 7: Verteilung der Gesamtbewertungen pro Pre-Editing-Methode

Die Zieltexte mit nicht-präeditiertem Ausgangstext wurden 21 Mal, die Zieltexte mit Methode 1 wurden 12 Mal bewertet. Methode 2 wurde 19 Mal bewertet; Methode 3 wurde am meisten bewertet, 24 Mal. Methode 4 wurde nur 9 Mal und Methode 5K und 6K jeweils 15 und 20 Mal bewertet. Die Anzahl der Bewertungen pro Methode trägt zur Unterstützung der oben erklärten Ergebnisse pro Pre-Editing-Methode bei, insbesondere was Methode 3 betrifft, die als „beste“ Methode mit den am meisten positiven Bewertungen festgestellt wurde.

5.1. Profile der Proband*innen

Für die empirische Studie wurden fünf Proband*innen gefragt, die im Folgenden kurz vorgestellt werden. Alle Proband*innen gehören zur Altersgruppe 18-26 Jahre. Proband*in 1 hat Italienisch als Erstsprache und verfügt über sehr gute Englisch- (C1-Stufe) und Deutschkenntnisse, die sie im Laufe ihres Bachelorstudiums erworben hat. Aktuell ist sie Student*in des Masterstudiums Translation an der Universität Wien mit Schwerpunkt Fachübersetzen und Sprachindustrie, wo sie ihre Übersetzungskenntnisse fortlaufend vertieft. Proband*in 2 ist auch Student*in mit Erstsprache Italienisch, sie wird bald ihr Masterstudium Translation mit Schwerpunkt Fachübersetzen und Sprachindustrie an der Universität Wien absolvieren. Ihr Bachelorstudium hat sie in Italien absolviert und ihre gemeldeten

Arbeitssprachen an der Universität Wien sind Italienisch, Deutsch und Englisch. Ihre Erfahrungen als Übersetzer*in konnte sie im Laufe ihrer Praktika sammeln. Proband*in 3 hat Italienisch als Erstsprache. Sie studiert Fachübersetzen und Sprachindustrie an der Universität Wien; Italienisch, Englisch und Deutsch sind ihre Arbeitssprachen. Sie konnte auch Erfahrungen als Übersetzer*in sammeln, indem sie ein Praktikum im Rahmen ihrer Ausbildung absolviert hat und sie als freiwillige Übersetzer*in für einige Nichtregierungsorganisationen gearbeitet hat. Proband*innen 4 und 5 studieren Dolmetschwissenschaft (mit Schwerpunkt Dialogdolmetschen) an der Universität Wien. Im Laufe ihres Masterstudiums konnten sie aber auch einige Übersetzungsübungen belegen. Ihre Ausbildung als Dolmetscher*innen hat schon während ihres Bachelorstudiums angefangen. Ihre Englisch- und Deutschkenntnisse konnten sie im Laufe des Studiums vertiefen. Nebst dem Studium konnten beide Berufserfahrungen im Bereich Dolmetschen sammeln.

6. Diskussion

Die Datenauswertung im Zuge dieser Arbeit erfolgte qualitativ wie die empirische Vorgehensweise. Die Schlussfolgerungen der vorliegenden Arbeit beginnen mit einer Bezugnahme auf die Forschungsfrage und auf die in der Arbeit formulierten Hypothesen (siehe Kapitel 6.1.). In Kapitel 6.2. werden die in Kapitel 5. dargestellten Ergebnisse möglichst ausführlich ausgewertet. Im dritten Kapitel dieser Diskussion (Kapitel 6.3.) werden die Vor- und Nachteile des Best-Worst-Scalings erläutert und in Kapitel 6.4. wird Bezug auf andere Studien genommen. Schlussendlich beinhaltet Kapitel 6.5. einige Anmerkungen zur Methode.

6.1. Die Zahlen als Antwort

Welche Pre-Editing-Methoden und Kombinationen aus Pre-Editing-Methoden verbessern das Resultat der maschinellen Übersetzung aus dem Englischen ins Italienische? So lautet die Forschungsfrage dieser Arbeit. Während eine klein angelegte empirische Studie keine allgemeingültige Antwort geben kann, zeigt sich doch eine klare Tendenz in den Resultaten ab. Diese bezieht sich auf die Pre-Editing-Methode 3, die sich wiederholt als „beste“ erwiesen hat. Nach diesem Resultat kann man zu der Annahme kommen, dass ausschließlich die Anwendung der Pre-Editing-Methode 3 (Verwendung von Verben aus einem Wort statt der Verwendung von Verben mit Partikel) auf einen Ausgangstext zu einem maschinell übersetzbareren Ausgangstext führen könnte. Des Weiteren könnte die Pre-Editing-Methode 3 wahrscheinlich auch qualitativ hochwertigere neuronale maschinelle Übersetzungen für die Sprachkombination Englisch-Italienisch erzielen. In diesem Zusammenhang gilt es festzuhalten, dass die Erwartungen an die Ergebnisse nicht vollends erfüllt wurden, da nur eine der sechs Pre-Editing-Methoden eine positive Endbewertung erhalten und sich als anwendbar für die Erreichung besserer Übersetzungen aus dem Englischen ins Italienische erwiesen hat. Zu den Erwartungen für diese Studie zählt, dass die Mehrheit der Pre-Editing-Methoden einen positiven Einfluss auf die MTranslatability der Texte und somit auf qualitativ hochwertigere maschinelle Resultate haben könnte. Diesbezüglich haben die Ergebnisse die formulierten Hypothesen nur teilweise unterstützt, weil lediglich eine der Pre-Editing-Methoden von Nutzen bei der NMÜ für das Sprachpaar Englisch-Italienisch sein könnte. Die anfängliche Annahme besagte, dass die maschinellen Resultate nicht als Endprodukt eines Übersetzungsprozesses betrachtet werden können und dass sie durch Pre- oder Post-Editing editiert werden müssen. Dies führt zur Schlussfolgerung, dass in fünf der sechs Fälle, in denen die ausgewählten Pre-

Editing-Methoden angewandt wurden, auf keine Post-Editing-Operation verzichtet werden kann.

Eine weitere Annahme, dass die Qualität allgemein als subjektiver Begriff betrachtet wird, kann als Grund der unterschiedlichen Bewertungen jeder Methode gelten. Deswegen ging es bei dieser Studie nicht darum, dass ein zugewiesener positiver Wert eine gute Übersetzung bedeutete, sondern eine „bessere“ Übersetzung, wobei der positive Wert die „beste“ Übersetzung aus vier Möglichkeiten definierte. Das gleiche gilt für die zugewiesenen negativen Bewertungen, wobei die negativ bewerteten Übersetzungen die „schlechteren“ sind und lediglich als die „schlechtesten“ unter die verfügbaren Varianten eingestuft wurden. Die Verbesserung der Qualität neuronaler maschineller Übersetzung konnte in dieser Arbeit nicht direkt festgestellt werden. Stattdessen wurden hier einige Verbesserungsvorschläge in die Praxis umgesetzt und ausgewertet. Diese Schlussfolgerung bezieht sich nur auf diese empirische Studie und nur auf die erwähnte Sprachkombination und ausgewählte Textsorte.

6.2. Bedeutung der Ergebnisse

Im Laufe der Vorbereitungsphase der empirischen Studie wurden bestimmte Textpassagen ausgewählt, deren maschinelle Übersetzbarkeit in Frage gestellt werden kann. Aufgrund bestimmter Fehler oder Ungenauigkeiten im Ausgangstext, kann die Übersetzung durch eine (N)MÜ-Software qualitativ mangelhaft erscheinen. Auf der Basis von den ausgesuchten Textpassagen wurde versucht, die passendsten Pre-Editing-Methoden anzuwenden, die die anfänglichen maschinellen Fehler in den Übersetzungen mindestens reduzieren konnten. Sofern die Datenerhebung vollständig war, konnte die Verfasserin beobachten, dass die Effektivität der Methoden nicht durch alle Proband*innen bewiesen worden ist (siehe Tabelle 12, S. 73):

Tabelle 12: Gesamtpunkte pro Pre-Editing-Methode pro Proband*in

PE	kP	1.	2.	3.	4.	5K	6K
P1	0	-1	-2	+8	+3	-5	-6
P2	-8	+4	+3	+6	0	-6	-5
P3	0	-4	-5	-2	-5	-12	-8
P4	-7	-3	0	+2	+3	0	+8
P5	-4	0	+2	+1	-2	-3	+9
Gesamtpunkte	-19	-4	-2	+15	-1	-26	-2

Wie schon in Kapitel 6.1. erwähnt, bedeuten diese Ergebnisse, dass jede Proband*in ihre eigene Präferenz in Bezug auf die Gesamtqualität und auf ihre eigenen Vorstellungen von Übersetzungen bekunden konnte. Ferner gehörte zur Aufgabenstellung der Proband*innen nicht die Bestimmung einer guten Übersetzung, sondern die Angabe ihrer Präferenzen in Bezug auf die vorgelegten Übersetzungen, von denen sie eine „beste“ und eine „schlechteste“ herausuchen mussten. Die positiven und die negativen Bewertungen zeigen, dass die Proband*innen wahrscheinlich verschiedene Evaluierungsstrategien umgesetzt haben bzw. verschiedene Aspekte der zu evaluierenden Übersetzungen in Betracht gezogen haben. Da sie selbst einen Wert zu jedem Punkt der Bewertungsskala zuweisen konnten (außer für die Extreme der Skala, „sehr gut“ und „vollkommen inakzeptabel“). Die obige Tabelle (siehe Tabelle 12, S. 73) zeigt auch, dass sich die Gesamtbewertungen pro Methode je nach Proband*in unterscheiden, im Sinne von Positiv- oder Negativbewertungen. Dies bedeutet, dass sich die Resultate nicht allgemein positiv oder negativ erwiesen haben.

Auf diesen Punkt sollte hier näher eingegangen werden, da die Proband*innen vier Übersetzungen vorgelegt bekommen haben: drei Übersetzungen der gleichen Textpassage, die im Voraus durch drei unterschiedliche Pre-Editing-Methoden präeditiert worden war und eine Übersetzung der gleichen Textpassage, die aber nicht präeditiert worden war. Die Verfasserin erwartete von der empirischen Studie, dass sich die negativen Bewertungen nur auf die Übersetzungen bezogen, deren Textpassage keinem Pre-Editing unterzogen worden war. Trotzdem wurden alle Pre-Editing-Methoden mindestens einmal von jeder Proband*in negativ bewertet.

Jede Proband*in hat nicht in allen Fällen für jede Übersetzungsgruppe die eine Übersetzung als „beste“ ausgewählt, deren Textpassage mit einer bestimmten Methode präeditiert worden war. Es gilt festzuhalten, dass wenn die verschiedenen Übersetzungen mehrmals ausgewählt wurden, deren Textpassagen nach identer Pre-Editing-Methode präeditiert worden waren, haben sie keine konsistenten Bewertungen erhalten. Dies bedeutet, dass die Pre-Editing-Methoden nicht konsistent bewertet wurden und dass andere Faktoren die Bewertungen beeinflusst haben. Ein konkretes Beispiel dafür stellt der Fall der Pre-Editing-Methode 3 dar (beschrieben in Kapitel 5.). Gemäß der Verfasserin liegen die Gründe darin, dass die Bewertung von vielen Aspekten in der Übersetzung abhängt, wie zum Beispiel die allgemeine Verständlichkeit des Textes. Andere Aspekte, die höchstwahrscheinlich in die Bewertung eingeflossen sind, können die grammatikalische Korrektheit, die Terminologie, der Stil und oder die generelle Übereinstimmung der Informationen in der Übersetzung mit jenen des Ausgangstexts sein. Wissenschaftlich relevant sind daher nicht nur die Positiv-, sondern

auch die Negativergebnisse. Sie geben Informationen sowohl über die Effektivität der Pre-Editing-Methoden als auch über die allgemein wahrgenommene Qualität der Übersetzungen. Dies ist hauptsächlich, was in Bezug auf Methode 5K (und auch 1, 6K, 2 und 4) passiert ist. Nicht nur haben sich die Bewertungen der verwendeten Pre-Editing-Methoden als ungenügend erwiesen, sondern die Übersetzungen waren insgesamt von niedriger Qualität, im Verhältnis zu den Erwartungen der Proband*innen.

6.3. Best-Worst-Scaling: Vorteile und Nachteile

Das Best-Worst-Scaling ist im Rahmen dieser wissenschaftlichen Arbeit von hoher Bedeutung, weil es nach derzeitigem Kenntnisstand der Verfasserin im Zuge keiner Studie im Bereich Translationswissenschaft als Evaluierungsmethode ausgewählt worden ist. Diese Methode hat am Anfang eine Herausforderung für die Verfasserin dargestellt, trotzdem hat sie sich als gute Methode zur Bestimmung von Präferenzen ergeben. Durch das Best-Worst-Scaling wurden die Proband*innen gefragt, unterschiedliche Übersetzungen zu lesen und intuitiv zu bewerten, ohne detaillierte Erklärung der für die Evaluierung anzuwendenden Skala (nur die Endpunkte wurden definiert). Dies stellt auf jeden Fall einen Vorteil dieser Methode dar, weil die Proband*innen keine vordefinierten Fehlerkategorien oder Aspekte der Übersetzungen evaluieren, sondern ihre Qualitätsvorstellungen durch einen Wert äußern mussten. Ein weiterer Vorteil, der aus der Best-Worst-Scaling-Methode abzuleiten ist, betrifft die Möglichkeit, alle Ergebnisse schnell und klar zu sammeln und auszuwerten. Die Kehrseite der Medaille über das Best-Worst-Scaling stellt, in Bezug auf diese Arbeit, der subjektive Charakter der Methode selbst dar. Alle Bewertungen spiegeln die Meinungen der Proband*innen wider, sowohl was die Qualitätsanforderungen von (maschinellen) Übersetzungen, als auch die Bedeutung der in der NMÜ zu findenden Fehler betrifft. Die Proband*innen haben der Meinung der Verfasserin nach ihre Kompetenzen als ausgebildete Translator*innen (Übersetzer*innen oder Dolmetscher*innen) in die Evaluierung eingebracht und die maschinellen Übersetzungen im gleichen Maß bewertet, wie sie im Regelfall Humanübersetzungen evaluiert hätten.

Eine freie Gestaltung der Befragung für die Proband*innen wurde durch LimeSurvey (LimeSurvey Homepage 2020) möglich, obwohl sich die Verfasserin damit zum ersten Mal beschäftigt hatte und die Software eine Schwachstelle zur Erstellung der Befragung darstellte. Angesichts der Tatsache, dass die Umfrage technische Schwierigkeiten bereitete, konnten einige Parameter für die Antworten der Proband*innen nicht spezifiziert werden. Eine

Einstellung der möglich auswählbaren Bewertungen für jeden Zieltext hätte wahrscheinlich verhindert, dass die Proband*innen die Aufgabenstellung (falsch) interpretieren könnten. Diesbezüglich haben zwei der sieben Proband*innen die Aufgabenstellung nicht verstanden, indem eine davon zwei Negativergebnisse und zwei Positivergebnisse in zwei Antwortgruppen abgegeben hat und eine zweite Proband*in beinahe durchgängig negative Bewertungen gegeben hat. Die Motivation für ihre Bewertungen war die niedrige Qualität der zu evaluierenden Übersetzungen. Aus diesem Grund mussten die zwei Proband*innen von der Studie ausgeschlossen und ersetzt werden. Bei der Evaluierung der zwei neuen Proband*innen wurde dann sichergestellt, dass die Aufgabenstellung verstanden wurde und pro Satz an Übersetzungskandidaten immer eine als vergleichsweise beste und eine als vergleichsweise schlechteste gewählt wurde. Für eine künftige Evaluierung mithilfe dieser Methode empfiehlt die Verfasserin eine technische Einschränkung der Antwortmöglichkeiten in dieser Hinsicht um Missverständnisse zu unterbinden.

6.4. Bezug auf andere Studien

In diesem Kapitel werden die aktuellen Studien im Bereich Pre-Editing mit der vorliegenden Forschungsarbeit verknüpft und die jeweiligen Ergebnisse verglichen. Keine der erwähnten Studien hat BWS als Evaluierungsmethode angewandt, sondern die meisten haben die Ergebnisse durch die gängigen Evaluierungsmetriken erzielt (wie BLEU). Die Ergebnisse des Pre-Editings von Babych et al. (2009) können mit denen dieser Studie verglichen werden, obwohl die betrachteten Sprachkombinationen (hier Englisch-Italienisch und bei Babych et al. (2009) Französisch-Englisch und Russisch-Englisch) nicht übereinstimmen und daher zu unterschiedlich geltenden Ergebnissen führen. Die Studie von Babych et al. (2009) richtet sich auf die Verbesserung der Qualität regelbasierter maschineller Übersetzung von Funktionsverben, was nicht der Fall bei der vorliegenden Studie ist. Eine Ähnlichkeit zwischen den zwei Studien besteht aber in der Analyse zur Identifizierung problematischer Verbkonstruktionen und ambiguer lexikalischer Einheiten, um Pre-Editing sinnvoll anzuwenden. Der Einsatz von Pre-Editing im Zuge der vorliegenden Arbeit hat sich ausschließlich in Bezug auf die Verben mit Partikel als positiv ergeben. Ein weiteres Ergebnis beider Studien bezieht sich auf die Verstärkung der Übersetzbarkeit des ATs, wobei auch die Verständlichkeit der Übersetzungen verbessert wird.

Bei der Studie von Seretan et al. (2014) liegt die SMÜ im Fokus. Außerdem werden auch in diesem Fall andere Sprachen miteinbezogen. Allerdings stellt die Analyse der

Übersetzung von Inhalten wie Blogs und den dazugehörigen Textkonventionen eine Gemeinsamkeit zwischen den zwei Studien dar. Die Pre-Editing-Strategien von Seretan et al. (2014) haben zu positiven Ergebnissen geführt, und zwar zur tatsächlichen Verbesserung maschineller Resultate. Diese Forschungsarbeit kann nur begrenzte positive Ergebnisse aufweisen.

In der Studie von Gerlach (2015), zeigen die Ergebnisse eine positive Tendenz auf, indem die Pre-Editing-Prozesse in unterschiedliche regelbasierte und statistische MÜ-Systeme und in unterschiedliche Bereiche übertragbar sind. Dieser Aspekt wurde im Rahmen dieser vorliegenden Arbeit nicht untersucht, aber könnte als guter Startpunkt für eine künftige Analyse gelten.

Die jüngste Forschungsarbeit über Pre-Editing stammt von Miyata und Fujita (2017). Ihre Ergebnisse zeigen einen sehr hohen Prozentsatz von befriedigenden Resultaten nach dem Einsatz von Pre-Editing auf die Ausgangstexte. Bei Miyata und Fujita (2017) wird hervorgehoben, dass Pre-Editing und menschliche Eingriffe über ein großes Potenzial verfügen. Dies bezeichnet auch eine wichtige Schlussfolgerung der vorliegenden Arbeit. Als weitere bemerkenswerte Schlussfolgerung der Studie von Miyata und Fujita (2017) und dieser Masterarbeit gilt, dass abhängig von der Länge des Ausgangstextes mehrere Pre-Editing-Operationen nötig sind, um eine bessere Zieltextqualität zu erreichen. Die Fehlübersetzung von Eigennamen und manchmal unnatürliche bzw. inkorrekte Wortwahl (vgl. Miyata und Fujita 2017) sind auch im Rahmen dieser Studie festgestellt worden. Wie im obigen Absatz angeführt, ist die Übertragbarkeit von Pre-Editing-Regeln von einer Sprachkombination auf eine andere ein guter Denkanstoß für weitere Forschungsmöglichkeiten, wobei es diese sorgfältig zu hinterfragen gilt.

6.5. Bemerkungen zur Methode

Im Zuge dieser Arbeit gilt das Best-Worst-Scaling als die Endphase der ganzen Methode, die für die empirische Studie angewandt wurde. Wie bereits im einleitenden Kapitel 1. erwähnt, besteht diese Studie aus verschiedenen Phasen, die am meisten menschlich durchgeführt wurden, mit Ausnahme des maschinellen Übersetzungsprozesses und der online zu ergänzenden Umfrage. Hinsichtlich der menschlichen bzw. manuellen Phasen hat die Verfasserin die ganze Studie schrittweise gebaut: von der Auswahl der Ausgangstexte über die Extrahierung der für die Studie nötigen Textpassagen und die Auswahl der darauf basierenden

Pre-Editing-Methoden bis hin zum Pre-Editing der Textpassagen und zur Auswahl der von der NMÜ-Software vorgeschlagenen Übersetzungsvarianten. Diese Vorgehensweise hat ermöglicht, die Studie maßgeschneidert aufzubauen und im Allgemeinen viele Aspekte der Studie selbst persönlicher zu gestalten.

Im Rahmen dieser Forschungsarbeit konnte die Verfasserin ihre Kenntnisse in vielen Bereichen vertiefen: im Bereich neuronaler maschineller Übersetzung (Entwicklung und Funktionsweisen), im Bereich Pre-Editing, und vor allem Best-Worst-Scaling als Evaluierungsmethode. Dazu konnte die Verfasserin einige technische Kenntnisse erweitern, vor allem bezüglich der Verwendung von online-Tools wie NMÜ- und Umfrage-Software.

Insbesondere in Bezug auf die angewandte Best-Worst-Scaling-Methode kann die Verfasserin ein positives Feedback geben, weil sich diese Methode als gut geeignet für die Bestimmung von Präferenzen feststellen lässt, unabhängig von den erwarteten und erzielten Ergebnissen. Es handelt sich um eine einfach interpretierbare Methode, wobei keine zu spezifischen Evaluierungskriterien oder -kategorien zu berücksichtigen sind. Diese Methode ist für qualitative Forschungsarbeiten besonders geeignet. Der Meinung der Verfasserin zufolge stellt das Best-Worst-Scaling eine sehr brauchbare Methode zur künftigen Weiterforschung der maschinellen Übersetzungsqualität.

7. Fazit

Die vorliegende Forschungsarbeit wurde zur Untersuchung der Auswirkung von Pre-Editing-Methoden auf die Verbesserung der maschinellen Übersetzung im Sprachpaar English-Italienisch angestellt.

Den Untersuchungsgegenstand dieser Arbeit bildeten ausgewählte Pre-Editing-Methoden als Verbesserungsvorschläge der neuronalen maschinellen Übersetzung für die erwähnte Sprachkombination. Die Perspektive im Bereich der Qualitätsforschung wurde zu einem früheren Schritt gewechselt: das Pre-Editing. Die Verfasserin hat sich zuerst mit dem Begriff der MTranslatability der MÜ befasst, die als wichtige Basis bzw. Voraussetzung für die vorliegende Studie zur Verbesserung der NMÜ-Qualität angesehen wurde.

Bezüglich der Methode, die zur empirischen Evaluierung dient, wurde das Best-Worst-Scaling vorgestellt. Diese Methode wurde angewandt, damit die Proband*innen ihre Präferenzen über die „besten“ und die „schlechtesten“ Übersetzungen aus verschiedenen Vierer-Sets bekunden konnten. Dazu mussten sie keine bestimmten Fehlerkategorien in die Evaluierung einfließen lassen. Die Textpassagen für die Studie kamen aus dem Bereich der Biologie. Insgesamt wurden vier einzelne Pre-Editing-Methoden und zwei Methodenkombinationen bewertet. Die Ergebnisse haben gezeigt, dass im Rahmen dieser Studie ausschließlich eine einzelne Methode die meisten positiven Ergebnisse aufweisen konnte. Dies bedeutet, dass das Präeditieren englischer Verben mit Partikel in Verben aus einem Wort vor der maschinellen Übersetzung ins Italienische recht eindeutig zur Verbesserung der Qualität neuronaler maschineller Übersetzung führen könnte. Die Negativergebnisse der Studie besitzen auch Relevanz, da sie nicht eine schlechte maschinelle Übersetzungsqualität, sondern die „schlechtesten“ Übersetzungsvarianten unter den vorgelegten bewerten. Die Bewertungen beziehen sich nicht nur auf die angewandten Pre-Editing-Methoden pro Ausgangstext, sondern auf die gesamten Übersetzungen.

Das Best-Worst-Scaling hat sich als geeignete Methode zur Einschätzung der maschinellen Qualität erwiesen, da es flexibel gestaltet werden kann. Des Weiteren haben die Proband*innen die Möglichkeit, die Resultate nach ihren Vorstellungen eigenständig zu bewerten. In Bezug auf die Qualitätsvorstellungen, hat die neuronale maschinelle Übersetzung noch keine „human parity“ bzw. Humanqualität erreicht (Porsiel 2020:2, 3). Aufgrund der unerfindlichen Fehler der NMÜ (vgl. Porsiel 2020:7), weist das Best-Worst-Scaling, nach Meinung der Verfasserin, viel Potenzial als Evaluierungsmethode im Bereich der translatorischen Qualität auf, da es subjektive Einschätzungen berücksichtigt.

In einem weiterführenden Schritt, aufgrund der gewonnen Erkenntnisse wäre es interessant eine ähnliche Studie durchzuführen, an der mehrere Proband*innen teilnehmen, sowohl Übersetzer*innen als auch Lai*innen. Auf diese Weise könnte eine breitere Sicht auf das Pre-Editing-Phänomen und somit auf die Ergebnisse ermöglicht werden. Es wäre auch interessant, die Übertragbarkeit der hier untersuchten Pre-Editing-Methoden in andere Sprachkombinationen zu untersuchen. Alternativ könnte eine künftige Studie erforschen, auf welche praktische Weise Pre-Editing-Funktionen in einer maschinellen Übersetzungssoftware integriert werden könnten, um einem breiteren Publikum bessere Übersetzungen zur Verfügung zu stellen.

Bibliographie

- ALPAC (1966). *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council (1416).
- American Museum of Natural History (AMNH, 2019). The Glowing Hawksbill Sea Turtle. American Museum of Natural History (AMNH), 09.01.2019. <https://www.amnh.org/explore/news-blogs/news-posts/the-glowing-hawksbill-sea-turtle> (Stand: 08.09.2020).
- Avramidis, Eleftherios (2016). Qualitative: python tool for MT quality estimation supporting server mode and hybrid MT. *The Prague Bulletin of Mathematical Linguistics* 106 (1), 147-158. doi: 10.1515/pralin-2016-0014.
- Avramidis, Eleftherios (2019). *Comparative Quality Estimation for Machine Translation. An Application of Artificial Intelligence on Language Technology using Machine Learning of Human Preferences*. Dissertation, Universität des Saarlandes
- Avramidis, Eleftherios; Macketanz, Vivien; Lommel, Arle & Uszkoreit, Hans. (2018). Fine-grained evaluation of Quality Estimation for Machine translation based on a linguistically motivated Test Suite. In: *Proceedings of the AMTA 2018 Workshop on Translation Quality Estimation and Automatic Post-Editing*. o.O., 243-248.
- Babych, Bogdan; Hartley, Anthony & Sharoff, Serge (2009). Evaluation-guided pre-editing of source text: improving MT-tractability of light verb constructions. In: *13th Annual Conference of the European Association for Machine Translation*. o. O., 36-43.
- Bahdanau, Dzmitry; Kyunghyun, Cho & Bengio, Yoshua (2014). Neural Machine Translation by Jointly Learning to Align and Translate, 1-15. CoRR *arXiv abs/1409.0473*.
- Banerjee, Satanj & Lavie, Alon (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements. In: *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*. Ann Arbor, Michigan, USA: Association for Computational Linguistics, 65-72.
- Bar-Hillel, Yehoshua (1960). The present status of automatic translation of languages. In: *Advances in computers* (1), 91-163.
- Barreiro, Anabela (2011). SPIDER: a system for paraphrasing in document editing and revision—applicability in machine translation pre-editing. In: *International Conference on Intelligent Text Processing and Computational Linguistics*. Berlin/Heidelberg: Springer, 365-376.
- Barreiro, Anabela & Cabral, Luis Miguel (2009). ReEscribe: a translator-friendly multi-purpose paraphrasing software tool. In: Goulet, Marie-Josée; Melançon, Christiane; Désilets, Alain & Macklovitch, Elliott (Hg.) (2009) *Proceedings of the Workshop Beyond Translation Memories: New Tools for Translators*. Château Laurier, o. S.
- Belinkov, Yonatan; Durrani, Nadir, Dalvi, Fahim; Sajjad, Hassan & Glass, James (2017). What do Neural Machine Translation Models Learn about Morphology?. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (1), 861–872. doi:10.18653/v1/P17–1080.
- Bentivogli, Luisa; Bisazza, Arianna; Cettolo, Mauro & Federico, Marcello (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. In: *The 2016 Conference on Empirical Methods on Natural Language Processing*. Austin, Texas, USA: Association for Computational Linguistics, o. S.

- Berth, Arendse (1997). EasyEnglish: A Tool for Improving Document Quality. In: *Fifth Conference on Applied Natural Language Processing*. Washington, DC, USA: Association for Computational Linguistics, 159-165.
- Berth, Arendse (1999). Controlling input and output of MT for greater user acceptance. In: *Translating and the Computer 21; Proceedings of the Twenty-first International Conference on Translating and the Computer*. London, o.S.
- Berth, Arendse & Gdaniec, Claudia (2001). MTranslatability. *Machine Translation* 16 (3), 175-218.
- Blatz, John; Fitzgerald, Erin; Foster, George; Gandrabur, Simona; Goutte, Cyril; Kulesza, Alex; Sanchis, Alberto & Ueffing, Nicola (2004). Confidence Estimation for Machine Translation. In: *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. Geneva, 315-321.
- Burchardt, Aljoscha & Porsiel, Jörg (2017). Vorwort: Was kann maschinelle Übersetzung und was nicht?. In: Porsiel, Jörg (Hg.) *Maschinelle Übersetzung: Grundlagen für den professionellen Einsatz*. Berlin: BDÜ Weiterbildungs- und Fachverlags GmbH, 11-18.
- Burchardt, Aljoscha; Macketanz, Vivien; Dehdari, Jon; Heigold, Georg; Peter, Jan-Thorsten & Williams, Philip (2017). A Linguistic Evaluation of Rule-Based, Phrase-Based, and Neural MT Engines. *The Prague Bulletin of Mathematical Linguistics* 108 (1), 159-170. doi:10.1515/pralin-2017-0017.
- Burlot, Franck; Scherrer, Yves; Ravishankar, Vinit; Bojar, Ondrej; Grönroos, Stig-Arne; Koponen, Maarit; Nieminen, Tommi & Yvon, Francois (2018). The WMT'18 Morpheval test suites for English-Czech, English-German, English-Finnish and Turkish-English. In: *Proceedings of The Third Conference on Machine Translation (WMT)* (2). Brüssel, Belgien: Association for Computational Linguistics, 550-564. doi:10.18653/v1/W18-64060.
- Burton, Nichola; Burton, Michael; Rigby, Dan; Sutherland, Clare A. M. & Rhodes, Gillian (2019). Best-worst scaling improves measurement of first impressions. *Cognitive research: principles and implications* 4 (1), 1-10. doi:10.1186/s41235-019-0183-2.
- Callison-Burch, Chris; Osborne, Miles & Koehn, Philipp (2006). Re-evaluation the role of bleu in machine translation research. In: *11th Conference of the European Chapter of the Association for Computational Linguistics*, 249-256.
- Castilho, Sheila; Moorkens, Joss; Gaspari, Federico; Calixto, Iacer; Tinsley, John & Way, Andy (2017a). Is neural machine translation the new state of the art?. *The Prague Bulletin of Mathematical Linguistics*, 108 (1), 109-120. doi: 10.1515/pralin-2017-0013.
- Castilho, Sheila; Moorkens, Joss; Gaspari, Federico; Sennrich, Rico; Sosoni, Vilelmini; Georgakopoulou, Palayota; Lohar, Pintu; Way, Andy; Barone, Antonio Valerio Miceli & Gialama, Maria (2017b). A comparative quality evaluation of PBSMT and NMT using professional translators. In: *Proceedings of MT Summit XVI* (1), Nagoya, Japan, 116-131.
- Castilho, Sheila; Doherty, Stephen; Gaspari, Federico & Moorkens, Joss (2018). Approaches to Human and Machine Translation Quality Assessment. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) *Translation Quality Assessment*. Cham: Springer International Publishing AG, 9-38. doi: 10.1007/978-3-319-91241-7.
- Chen, Mia Xu; Firat, Orhan; Bapna, Ankur; Johnson, Melvin; Macherey, Wolfgang; Foster, George; Jones, Llion; Parmar, Niki; Schuster, Mike; Chen, Zhifeng; Wu, Yonghui & Hughes, Macduff (2018). The Best of Both Worlds: Combining Recent Advances in Neural Machine Translation. CoRR *arXiv abs/1804.09849*.
- Comparin, Lucia & Mendes, Sara (2017). Using error annotation to evaluate machine translation and human post-editing in a business environment. In: *Proceedings of EAMT 2017*. Prague, o. S.

- Crego, Josep; Kim, Jungi; Klein, Guillaume; Rebollo, Anabel; Yang, Kathy; Senellart, Jean; Akhanov, Egor; Brunelle, Patrice; Coquard, Aurelien; Deng, Yongchao; Enoue, Satoshi; Geiss, Chiyo; Johanson, Joshua; Khalsa, Ardas; Khiari, Raoum; Ko, Byeongil; Kobus, Catherine; Lorieux, Jean; Martins, Leidiana; Nguyen, Dang-Chuan; Priori, Alexandra; Riccardi, Thomas; Segal, Natalia; Servan, Christophe; Tiquet, Cyril; Wang, Bo; Yang, Jin; Zhang, Dakun; Zhou, Jing & Zoldan, Peter (2016). SYSTRAN's Pure Neural Machine Translation Systems. CoRR *arXiv abs/1610.05540*.
- DeepL Translator (2020). <https://www.deepl.com/translator> (Stand: 13.06.2020).
- Doherty, Stephen (2017). Issues in human and automatic translation quality assessment. In: Kenny, Dorothy (Hg.) *Human issues in translation technology*. London: Routledge, 131-148.
- Dorr, Bonnie; Olive, Joseph; McCary, John & Christianson, Caitlin (2011). Machine Translation Evaluation and Optimization. In: Olive, Joseph; McCary, John & Christianson, Caitlin (Hg.) *Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation*. London: Springer Science & Business Media, 745–843. doi:10.1007/978-1-4419-7713-7.
- Etherton, Caitlin (o.J.). See nature's amazing glow-in-the-dark spectacles. National Geographic Society, 09.01.2019. <https://www.nationalgeographic.com/travel/lists/bioluminescence-best-glow-in-dark-spots-around-world/> (Stand: 08.09.2020).
- Flanagan, Mary (1994). Error classification for MT evaluation. In: *Technology Partnerships for Crossing the Language Barrier: Proceedings of the First Conference of the Association for Machine Translation in the Americas*. o. O., 65-72.
- Forcada, Mikel L. (2017). Making sense of neural machine translation. *Translation Spaces* 6 (2), 291–309. doi:10.1075/ts.6.2.06for.
- Gabriel, Peter (2019). Einleitung: KI ohne Grenzen?. In: Wittpahl, Volker (Hg.) *Künstliche Intelligenz*. Berlin Heidelberg: Springer, 95–98. doi:10.1007/978-3-662-58042-4.
- Gage, Philip (1994). A new algorithm for data compression. *C Users Journal* 12 (2), 23-38.
- Garvin, David A. (1984). What Does "Product Quality" Really Mean?. *Sloan management review*, 25-43.
- Gerlach, Johanna (2015). *Improving statistical machine translation of informal language: a rule-based pre-editing approach for French Forums*. Doktorarbeit, Universität Genf. doi:10.13097/archive-ouverte/unige:73226
- Google Translate (2020). <https://translate.google.com/> (Stand: 13.06.2020).
- Google Translate Languages (2020). <https://translate.google.com/intl/en/about/languages/> (Stand: 13.06.2020).
- Göpferich, Susanne (2001). Von Hamburg nach Karlsruhe: Ein kommunikationsorientierter Bezugsrahmen zur Bewertung der Verständlichkeit von Texten. *Fachsprache (1990)*, 23 (3-4), 117-138.
- Grundkiewicz, Roman & Junczys-Dowmunt, Marcin (2018). Near human-level performance in grammatical error correction with hybrid machine translation. o. O.: o. S. CoRR *arXiv abs/1804.05945*.
- Guerberof Arenas, Ana (2019). Pre-editing and post-editing. In: Angelone, Erik; Ehrensberger-Dow, Maureen & Massey, Gary (Hg.) *The Bloomsbury Companion to Language Industry Studies*. United Kingdom: Bloomsbury Companions, Bloomsbury Academic, 1-42.
- Guillou, Liane; Hardmeier, Christian; Lapshinova-Koltunski, Ekaterina & Loaiciga, Sharid (2018). A Pronoun Test Suite Evaluation of the English–German MT Systems at WMT 2018. In: *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*. Brüssel: Association for Computational Linguistics, 570-577.

- Harder, Ben (2010). Many organisms rely on bioluminescence to breed, feed and evade predators. *The Washington Post*, 11.10.2010. <https://www.washingtonpost.com/wp-dyn/content/article/2010/10/11/AR2010101104495.html> (Stand: 08.09.2020).
- Hutchins, John W. (1996). ALPAC: the (in)famous report. In: *MT News International* (14), 9-12.
- Hutchins, John W. (2000). The first decades of machine translation: overview, chronology, sources. In: Hutchins, John W. (Hg.) *Early years in machine translation: memoirs and biographies of pioneers* (Vol. 97). Amsterdam/Philadelphia: John Benjamins Publishing, 1-16.
- Jabr, Ferris (2010). Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life. *Scientific American Inc.*, 05.08.2010. <https://www.scientificamerican.com/article/edith-widder-bioluminescence/> (Stand: 08.09.2020).
- Kedmy, Dan (2016). Gallery: How to speak jellyfish. TED Conferences, LLC., 05.01.2016. <https://ideas.ted.com/gallery-how-to-speak-jellyfish/> (Stand: 08.09.2020).
- Kenny, Dorothy (2018). Machine translation. In: *The Routledge Handbook of Translation and Philosophy*. o. O: Routledge, 428-445.
- Kepler, Fábio; Trénous, Jonay; Treviso, Marcos; Vera, Miguel & Martins, André F. T. (2019). OpenKiwi: An open source framework for quality estimation. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Florenz: Association for Computational Linguistics, 117-122. doi: 10.18653/v1/P19-3020.
- Kim, Hyun; Lee, Jong-Hyeok & Na, Seung-Hoon (2017). Predictor-estimator using multilevel task learning with stack propagation for neural quality estimation. In: *Proceedings of the Second Conference on Machine Translation*. Kopenhagen: Association for Computational Linguistics, 562-568.
- Kiritchenko, Svetlana & Mohammad, Saif (2017). Best-Worst Scaling More Reliable than Rating Scales: A Case Study on Sentiment Intensity Annotation. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vancouver: Association for Computational Linguistics, 465-470. doi:10.18653/v1/P17-2074.
- Klein, Guillaume; Kim, Yoon; Deng, Yuntian; Senellart, Jean & Rush, Alexander M. (2017). Openmt: Open-source toolkit for neural machine translation. *CoRR arXiv abs/1701.02810*.
- Klubicka, Filip; Toral, Antonio & Sánchez-Cartagena, Víctor M. (2017). Fine-grained human evaluation of neural versus phrase-based machine translation. *The Prague Bulletin of Mathematical Linguistics* 108, 121–132. doi: 10.1515/pralin-2017-0014.
- Koehn, Philipp (2010). Enabling Monolingual Translators: Post-Editing vs. Options. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, Los Angeles, California: Association for Computational Linguistics, 537-545.
- Koehn, Philipp (2017). Neural Machine Translation. John Hopkins University. *CoRR arXiv abs/1709.07809*.
- Koehn, Philipp & Monz, Christof (2006). Manual and automatic evaluation of machine translation between European languages. In: *Proceedings of the Workshop on Statistical Machine Translation*, New York: Association for Computational Linguistics, 102-121. doi:10.3115/1654650.1654666.
- Koehn, Philipp & Knowles, Rebecca (2017). Six Challenges for Neural Machine Translation. *CoRR arXiv abs/1706.03872*.

- Kreutzer, Julia; Shigehiko Schamoni & Stefan Riezler (2015). Quality Estimation from ScraTCH (QUETCH): Deep Learning for Word-level Translation Quality Estimation. In: *Proceedings of the Tenth Workshop on Statistical Machine Translation*. Lissabon, Portugal: Association for Computational Linguistics, 316-322. doi:10.18653/v1/W15-3037.
- Langer, Inghard; Friedemann Schulz von Thun & Reinhard Tausch (1974): *Verständlichkeit in Schule, Verwaltung, Politik, Wissenschaft: Mit einem Selbsttrainingsprogramm zur verständlichen Gestaltung von Lehr- und Informationstexten*. München/Basel: Reinhardt.
- Läubli, Samuel; Sennrich, Rico & Volk, Martin (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. In: Riloff, Ellen; Chiang, David; Hockenmaier, Julia & Tsujii, Jun'ichi (Hg.) *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Bruxelles, Belgien: The Association for Computational Linguistics, 4791-4796. CoRR *arXiv abs/1808.07048*.
- Lazzari, Gianni & Steinbiss, Volker (2006). Human language technologies for Europe. *ITC IRST/TC-Star project report*. Trento/Aachen: Accipio Consulting.
- LDC (2005). Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Translations. Revision 1.5.
- LeCun, Yann (1989). Generalization and network design strategies. In: Pfeifer, Rolf; Schreier, Zoltan; Fogelman-Soulié, Francois & Steels, Luc (Hg.) *Connectionism in perspective*, 19. Zürich, Schweiz: Elsevier, 143-155.
- Levenshtein, Vladimir I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady* 10 (8), 707-710.
- LimeSurvey Homepage (2020). <https://www.limesurvey.org/en/> (Stand: 16.09.2020).
- Lo Presti, Rosita (2016). *Menschliche und automatische Evaluation von Übersetzungen von Fachtexten in Google Translate*. Masterarbeit, Universität Wien
- Lommel, Arle (2018). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) *Translation Quality Assessment*. Cham: Springer International Publishing AG, 109-128. doi: 10.1007/978-3-319-91241-7.
- Lommel, Arle; Burchardt, Aljoscha & Uszkoreit, Hans (Hg.) (2015) *Multidimensional Quality Metrics (MQM) Definition*. Deutsches Forschungszentrum für Künstliche Intelligenz GmbH. <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html> (Stand: 23.06.2020).
- Long, Kat (2015). Life Bustles in the Long Arctic Night. *Hakai Magazine*, 04.11.2015. <https://www.hakaimagazine.com/news/life-bustles-long-arctic-night/> (Stand: 08.09.2020).
- Louviere, Jordan J. & Woodworth, George G. (1990). Best-Worst Scaling: A Model for Largest Difference Judgments, Working Paper. Faculty of Business, University of Alberta.
- Louviere, Jordan J.; Flynn, Terry N. & Marley, Anthony Alfred John (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge University Press.
- Luong, Minh-Thang (2017). *Neural Machine Translation*. Doktorarbeit, Stanford University
- Lutz, Benedikt (2017). Modelle für die verständliche Fachkommunikation. Das Spannungsfeld zwischen wissenschaftlichem Anspruch und praktischer Anwendbarkeit. *trans-kom* 10 (3), 284-314.
- Macketanz, Vivien; Avramidis, Eleftherios; Burchardt, Aljoscha & Uszkoreit, Hans (2018). Fine-grained evaluation of German-English Machine Translation based on a Test Suite. In: *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, 578-587. doi:10.18653/v1/W18-6436.

- Madsen, Mathias Winther (2009). *The limits of machine translation*. Masterarbeit, Universität Kopenhagen.
- Martins, André F. T.; Astudillo, Ramón; Hokamp, Chris & Kepler, Fabio (2016). Unbabel's Participation in the WMT16 Word-Level Translation Quality Estimation Shared Task. In: *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*. Berlin: Association for Computational Linguistics, 806-811.
- Merkert, Pina (2017). Maschinelle Übersetzer: DeepL macht Google Translate Konkurrenz. *Heise Online*, 29.08.2017. <https://www.heise.de/newsticker/meldung/Maschinelle-Uebersetzer-DeepL-macht-Google-Translate-Konkurrenz-3813882.html> (Stand: 15.09.2020).
- Miyata, Rei & Fujita, Atsushi (2017). Dissecting human pre-editing toward better use of off-the-shelf machine translation systems. In: *Proceedings of the 20th annual conference of the European association for machine translation (EAMT)*. o. O., 47-52.
- Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) (2018). *Translation Quality Assessment*. Cham: Springer International Publishing AG. doi: 10.1007/978-3-319-91241-7
- Nießen, Sonja; Och, Franz Joseph.; Leusch, Gregor & Ney, Hermann (2000). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. In: *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC'00)*. Athen: European Language Resources Association (ELRA), 39-45.
- Pallardy, Richard (o.J.). Everything's Illuminated: 6 Bioluminescent Organisms. *Encyclopaedia Britannica Inc.* <https://www.britannica.com/list/6-bioluminescent-organisms> (Stand: 08.09.2020)
- Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In: *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 311-318.
- Popović, Maja (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) *Translation Quality Assessment*. Cham: Springer International Publishing AG, 129-158. doi: 10.1007/978-3-319-91241-7.
- Porsiel, Jörg (2020). Welcome to the Machine! Zwischen Goldgräberstimmung und der Suche nach dem Heiligen Gral. In: Porsiel, Jörg (Hg.) *Maschinelle Übersetzung für Übersetzungsprofis*. Berlin: BDÜ Weiterbildungs- und Fachverlags GmbH, 2-10.
- Roturier, Johann (2004). Assessing a set of Controlled Language rules: Can they improve the performance of commercial Machine Translation systems. In: *Proceedings of the International Conference Translating and the Computer 26*. London: Aslib, 1-14.
- Rysová, Kateřina; Rysová, Magdaléna; Musil, Tomáš; Poláková, Lucie & Bojar, Ondřej (2019). A Test Suite and Manual Evaluation of Document-Level NMT at WMT19. o. O.: o. S. CoRR *arXiv abs/1908.03043*.
- SAE International J2450 (2001). Translation Quality Metric J2450_200112. Warrendale, USA: SAE International (Vehicle E E System Diagnostic Standards Committee). doi: 10.4271/J2450_200112.
- Schmalz, Antonia (2019). Maschinelle Übersetzung. In: Wittpahl, Volker (Hg.) *Künstliche Intelligenz*. Berlin Heidelberg: Springer, 194-211. doi:10.1007/978-3-662-58042-4.
- Sennrich, Rico; Haddow, Barry & Birch, Alexandra (2016). Neural Machine Translation of Rare Words with Subword Units. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (1): Long Papers*. Berlin: Association for Computational Linguistics, 1715-1725. doi: 10.18653/v1/P16-1162.

- Seretan, Violeta; Bouillon, Pierrette, & Gerlach, Johanna (2014). A Large-Scale Evaluation of Pre-editing Strategies for Improving User-Generated Content Translation. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. o. O: European Language Resources Association (ELRA), 1793-1799.
- Shoshan, Ofer (2019). Bridging the Gap Between NMT's Theoretical Promise and Practical Limitations. *Slator*, 23.10.2019. <https://slator.com/sponsored-content/bridging-the-gap-between-nmts-theoretical-promise-and-practical-limitations/> (Stand: 08.09.2020).
- Shterionov, Dimitar; Superbo, Riccardo; Nagle, Pat; Casanellas, Laura; O'dowd, Tony & Way, Andy (2018). Human versus automatic quality evaluation of NMT and PBSMT. *Machine Translation* 32 (3), 217-235. doi:10.1007/s10590-018-9220-z.
- Smith, James & Clark, Stephen (2009). EBMT for SMT: A New EBMT-SMT Hybrid. In: *Proceedings of the 3rd International Workshop on Example-Based Machine Translation*. o. O., 3-10.
- Snover, Matthew; Dorr, Bonnie; Schwartz, Richard; Micciulla, Linnea; Makhoul, John & Weischedel, Ralph (2006). A Study of Translation Error Rate with Targeted Human Annotation. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*. Cambridge, MA, USA: The Association for Machine Translation in the Americas, 223-231.
- Specia, Lucia & Shah, Kashif (2018). Machine Translation Quality Estimation: Applications and Future Perspectives. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) *Translation Quality Assessment*. Cham: Springer International Publishing AG, 201-236. doi: 10.1007/978-3-319-91241-7.
- Specia, Lucia; Raj, Dhvaj & Turchi, Marco (2010). Machine translation evaluation versus quality estimation. *Machine translation* 24 (1), 39-50.
- Stankevičiūtė, Gilvilė; Kasperavičienė, Ramunė & Horbačauskienė, Jolita (2017). Issues in Machine Translation. *International Journal on Language, Literature and Culture in Education* 4 (1), 75–88. doi:10.1515/llce–2017–0005.
- Stein, Daniel (2009). Maschinelle Übersetzung – ein Überblick. In: Seewald–Heeg, Uta & Stein, Daniel (Hg.) *Maschinelle Übersetzung – von der Theorie zur Anwendung. Journal for Language Technology and Computational Linguistics (JLCL)* 24 (3). Gesellschaft für Sprachtechnologie und Computerlinguistik, 5–18.
- Symne, Sara & Smith, Christian (2012). On the Interplay between Readability, Summarization, and MTranslatability. In: *Swedish Language Technology Conference (SLTC) 2012*, 71-72.
- Sutskever, Ilya; Vinyals, Oriol; and Le, Quoc V. (2014). Sequence to Sequence Learning with Neural Networks. In: Ghahramani, Zoubin; Welling, Max; Cortes, Corinna; Lawrence, Neil D. & Weinberger, Kilian Q. (Hg.) *Advances in Neural Information Processing Systems* 27. o. O: Curran Associates, Inc., 3104–3112.
- SYSTRAN beyond language (2020). Vorreiter und globaler Leader im Bereich von Übersetzungslösungen. <https://www.systransoft.com/de/systran/> (Stand: 23.04.2020).
- SYSTRAN Blog (2020). 5 reasons why SYSTRAN should be your preferred translation website. <https://blog.systransoft.com/5-reasons-why-systran-translate-preferred-translation-website/> (Stand: 23.04.2020).
- SYSTRAN Third Quarter 2009 Revenue Release (2009). <https://www.systransoft.com/download/financial-releases/SYSTRAN.2009.quarter3.revenue.pdf> (Stand: 10.06.2020).
- SYSTRAN Translate Tool (2020). <https://translate.systran.net/translationTools/text?lang=en> (Stand: 23.04.2020).

- SYSTRAN Translate Weboberfläche (2020). <https://translate.systran.net/translationTools/text> (Stand: 11.09.2020).
- Tavosanis, Mirko (2019). Valutazione umana di Google Traduttore e DeepL per le traduzioni di testi giornalistici dall'inglese verso l'italiano. In: *Proceedings of the Sixth Italian Conference on Computational Linguistics. Vol. 2481*. o. S.
- The Ocean Portal Team (2018). Bioluminescence. Smithsonian Institution. <https://ocean.si.edu/ocean-life/fish/bioluminescence> (Stand: 08.09.2020).
- Thurstone, Louis L. (1927a). Psychophysical analysis. *The American journal of psychology* 38 (3), 368-389.
- Thurstone, Louis L. (1927b). A law of comparative judgment. *Psychological review* 34 (4), 273-286.
- Thurstone, Louis L. (1994). A law of comparative judgment. *Psychological review* 101 (2), 266-270.
- Toma, Peter (1977). Systran as a multilingual machine translation system. In: *Proceedings of the Third European Congress on Information Systems and Networks, Overcoming the language barrier*. München: Verlag Dokumentation, 569–581.
- Tractica (2016). Artificial Intelligence: 10 Trends to Watch in 2017 and Beyond. Boulder, CO, USA: Tractica LLC, 1-23.
- Turovsky, Barak (2016). *Found in translation: More accurate, fluent sentences in Google Translate.* *Google Blog*, 15.11.2016. <https://www.blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/> (Stand: 13.06.2020).
- Van Slype, Georges (1979). Critical study of methods for evaluating the quality of machine translation (final report). Prepared for the Commission of the European Communities. Brüssel: Bureau Marcel van Dyke.
- Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Lukasz & Polosukhin, Illia (2017). Attention is all you need. In: *31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach, CA, USA, 1-15. CoRR arXiv abs/1706.03762.
- Vilar, David; Xu, Jia; D'Haro, Luis Fernando & Ney, Hermann (2006). Error analysis of statistical machine translation output. In: *Proceedings of 5th international conference on Language Resources and Evaluation (LREC 2006)*. Genoa, 697–702.
- Wagner, Emma (Hg.) (2014). *Translating for the European Union*. o. O: Routledge.
- Way, Andy (2018). Quality expectations of machine translation. In: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (Hg.) *Translation Quality Assessment*. Cham: Springer International Publishing AG, 159-178. doi: 10.1007/978-3-319-91241-7.
- Weaver, Warren (1949). Memorandum on Translation. 1-12. www.mt-archive.info/Weaver-1949.pdf (Stand: 15.09.2020).
- 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/Brüssel/New York/Oxford/Wien: Lang, 79-103.
- Wittpahl, Volker (Hg.) (2019). *Künstliche Intelligenz*. Berlin Heidelberg: Springer. doi:10.1007/978-3-662-58042-4.
- 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. o. O., 1-23. CoRR *arXiv abs/1609.08144*.
- Yamada, Masaru (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation* 31, 87-106.
- Yang, Jin & Lange, Elke D. (1998). SYSTRAN on AltaVista a user study on real-time machine translation on the Internet. In: Farewell, David; Gerber, Laurie & Hovy, Eduard (Hg.) *Machine Translation and the Information Soup: Proceedings of the Third Conference of the Association for Machine Translation in the Americas (AMTA), Vol. 1529*. Berlin Heidelberg: Springer, 275-285.
- Yoshimi, Takehiko (2001). Improvement of Translation Quality of English Newspaper Headlines by Automatic Pre-editing. *Machine translation* 16 (4), 233-250.

Anhang 1

Quelltexte und entsprechende Details und Quellen. Die jeweiligen in der Studie verwendeten Textpassagen sind unterstrichen.

Ausgangstext 1 und 8

© 1996-2015 National Geographic Society, © 2015- 2020 National Geographic Partners, LLC. All rights reserved

T R A V E L

See nature's amazing glow-in-the-dark spectacles

Visit these eight incredible bioluminescent phenomena around the world.

B Y C A I T L I N E T H E R T O N

Perhaps there is no greater magic than catching fireflies in a Pennsylvania blackberry thicket or walking on a river beach at night only to find each footprint glowing in the sand. Bioluminescence is a glorious phenomenon, showing up on different branches of the evolutionary tree over 40 separate times. Research into bioluminescent molecules even contributes to today's research of cancer, HIV, and malaria. And while we know that bioluminescence results from a light-releasing "luciferin" molecule, scientists still don't completely understand why and how the behavior occurs.

But we do know we love it. From Puerto Rico to the Maldives, spots around the world are entertaining an increasing number of glow-in-the-dark tourists. So reserve your tour ahead of time and leave your cell phone at home. Here are eight locations where you can spend your time staring at something that glows all night whether it's plugged in or not.

Fireflies & treetop strolls, South Africa

Every summer, fireflies twinkle through the leaf litter to find a mate in South Africa's 1,304-acre Kirstenbosch National Botanical Garden. Traveling to see fireflies is like traveling to see the northern lights in Iceland: Nature's timetable is about as reliable as your favorite airline has been lately. That said, the last three weeks of October are peak season.

The evening tour promises a golden Table Mountain sunset, a guided hike past night-blooming drumstick flowers into a forest of fireflies, and a final stroll along the Boomslang—a 40-foot-high tree canopy walkway. Five minutes away, the 207-room Vineyard Hotel & Spa boasts solar-panel energy, luxurious rooms, and a garden café complete with resident tortoises.

Termite mound click beetles, Brazil

Ecologist Douglas Trent has spent over 38 years studying Brazil's unique fauna and developing responsible locally-run ecotours. In addition to cloud forest hikes and wild jaguar river cruises, his company, Focus Tours, offers small-group 4x4 excursions into Emas National Park in Brazil's UNESCO-protected Cerrado—a tropical savanna Trent describes as "surprisingly delightful and rarely visited."

Emas is home to maned wolves, pampas deer, anaconda, giant anteaters, and over 200 bird species. A tour during the winter rainy season promises a chance to see the world's highest

concentration of termite mounds sparkle with thousands of bioluminescent *Pyrearinus termitilluminans*—click beetle larvae looking to lure and dine upon termites fleeing the mound. Using a tripod and a long exposure, the mounds are an impressive show to capture on camera—even if they can't be seen from space.

The land where everything glows, Japan

In Japan, bioluminescent millipedes, mushrooms, and weeping stones illuminate the dark, and cities release fireflies all summer long like some shimmering, nonstop wedding send-off. Roughly 50 firefly species reside in the country—perhaps best observed from Onnenai's peaceful wetland boardwalk on an uncrowded late-July weeknight.

Further north in Toyama, an entire museum is devoted to *hotaruika*, or firefly squid, whose “counter-illumination” functions as an invisibility cloak from hungry sharks and whales. For seven dollars, museum visitors can brush up on squid anatomy and even stick their hands into a glowing squid-y touch pool. Late March through May, the museum also offers 3 a.m. boat tours of Toyama Bay, where working fishermen hoist tangled neon-blue hotaruika from the sea. Stay at nearby Oyado Nono Hotel for free late-night ramen and cypress onsen baths.

Bioluminescent beach worms, UK

Cozy jersey sheets, doe-eyed Jersey cows, New Jersey—they all trace back to the original cliff-cuffed English isle of Jersey, just west of Normandy. Jersey boasts the largest rocky intertidal area in Europe and is known for its bioluminescent bristle worms. Jersey Walk Adventures offers bioluminescent tours all year. For the brightest experience, owners Derek and Trudie Hairon recommend wading on a moonless summer night when the Milky Way is gleaming above and the water below is “on fire with millions of flashes.”

After a Michelin-starred dinner (try Samphire) and a night of wading shimmering shoegazing (JWA rents wellies for only one euro (U.S. \$1.15), visitors can stay the night in one of Jersey's unique Heritage Lets, choosing between a historic fort, a cider barn, the 16th-century Elizabeth Castle, a seaside folly, or (perhaps the most convenient for bioluminescent-beach-walking) the solar-powered 18th-century Seymour Tower, only accessible during low tide.

Synchronous fireflies, Malaysia

There are synchronized fireworks, synchronized baton twirlers, synchronized swimmers ... and then there are synchronized fireflies. The rare *Pteroptyx tener* subspecies is found among the berembang mangrove trees of Malaysia's Selangor River, just a one-hour drive northwest of Kuala Lumpur.

Many local firefly tours use electric boats or traditional sampan boats in order to preserve natural habitat while reducing air, water, and noise pollution. As the state government, the Forest Research Institute of Malaysia, and local community members continue conservation efforts, visitors can help by leaving behind cameras and phones: electric lights distract long-lost lightning bugs from finding a mate.

Museum exhibits, plankton, & comb jellies, Florida

Dr. Rivah Winter, marine biologist and curator at Miami's Frost Museum of Science, describes “Creatures of Light” as a familyfriendly bioluminescent immersion with “larger-than-life-sized, biologically accurate models and interactive experiences that make the unfamiliar spectacular.” On view through April 21, the exhibit showcases bioluminescent mushrooms, glowworm caves, live flashlight fish, and Sharon Berebichez's glowing marine sculptures, all accompanied by an original symphony soundtrack.

Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glow-in-the-dark plankton in a see-through kayak, spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and bottomless cookies.

Bioluminescent marine life, Cayman Islands

Visitors to the Cayman Islands can gape at bioluminescence and fluorescence both above and below the waves by scuba diving, kayaking, snorkeling, glass-bottom-paddling, paddle-boarding, submarine, electric boat, and even by pool noodle. Caymanian Tom Watling has guided kayakers aged 6 to 96 on Cayman Kayaks' bioluminescent tours out of Rum Point. The main attraction, he says, are the sparkling photoplankton, "about 2,000 of them in every teaspoon of seawater"—that and the "disco shrimp," tiny bioluminescent crustaceans that, in a manner simultaneously divine and disgusting, ooze specks of blue, bioluminescent vomit in order to attract a mate.

Seven miles west of Bio Bay, the Kimpton Seafire Resort & Spa sports solar panels, rainwater cisterns, endemic landscaping, clean-lined ocean-view rooms, locally-sourced fare, and dark, turtle-friendly beaches.

Glowworms, Australia

December through March, hundreds of glowworms dangle sticky bioluminescent threads to lure and capture unlucky flies in the caves of Springbrook National Park's Natural Bridge. The glowworms—actually the larvae of a small fly, *Arachnocampa flava*, not worms—are found only in Australia and New Zealand. Springbrook's easy half-mile loop trail to the cave also promises spectacular stargazing and a chance to spot luminous mushrooms, fireflies, ancient hoop pines and snoozing koalas throughout the Gondwana Rainforest.

All visitors should abstain from flash photography, smoking, spelunking, actively spraying insect repellent, swimming, and shining flashlights directly on the glowworms, as this inhibits their wellbeing. Admission to the park is free, and guided tours are strongly encouraged. Nearby accommodations include cozy chalets, English Garden guesthouses, rainforest retreats, farm stays, and one park-designated camping area.

Cait Etherton is a Virginia-based writer and frequent contributor to National Geographic Travel. Follow her journey on Twitter @carryoncait.

PUBLISHED JANUARY 9, 2019

Quelle: <https://www.nationalgeographic.com/travel/lists/bioluminescence-best-glow-in-dark-spots-around-world/>. (Stand: 08.09.2020).

Ausgangstext 2 und 3

IDEAS.TED.COM

SCIENCE

Gallery: How to speak jellyfish

Jan 5, 2016 / Dan Kedmey

Ocean explorer Edith Widder shares the astoundingly sophisticated language of bioluminescent creatures, which can light up the depths of the seas like a “snow storm.”

The most spectacular gatherings of bioluminescent creatures occur in the twilight region of the sea, between 200 and 1,000 meters down, where squid, fish, shrimp and jellyfish find the ideal lighting conditions to spark up a conversation. Deep sea explorer Edith Widder first encountered one of these deep sea light shows while exploring the Santa Barbara Channel in a pressurized diving suit in 1984. “The first time was kind of peculiar because it was almost like a snow storm of light,” she says. “Other times it’s been little puffs of what looks like blue smoke. It’s different every time you dive.” But what’s truly fascinating about these light shows is they’re not just for entertainment; this is how the creatures are communicating. Take a look at some of the more sophisticated displays she’s encountered in her past three decades of dives.

(BILD)

All night diner

Angler fish use glowing lures to attract unsuspecting prey, such as shrimp, squid or fish. The prey are attracted to something that looks tasty, but instead of eating dinner, they become dinner. “These fish actually swallow the prey whole,” Widder says, “and they’ll eat things bigger than themselves. They’re kind of like boa constrictors, so their stomachs will expand to take on pretty large prey if they happen on it, because food is sufficiently rare that they’re not fussy about what they eat.”

(BILD)

I’m outta here

When a copepod detects a nearby threat, this tiny crustacean releases bioluminescent chemicals from glands on its legs to distract it, much in the way a fighter jet fires off flares to send a guided missile off course.

(BILD)

Single male seeking mate

Bioluminescent ostracods are called sea fireflies because they use their light to attract a mate; but instead of flashing code the way regular fireflies do, they release lightproducing chemicals in a series of little blobs of mucus that hang in the water. The spacing of the blobs is species-specific, so a female can recognize a male of her own kind and swim to the end of the line to hook up.

(BILD)

Take that!

The Janicella shrimp spews bioluminescence out of its mouth like a fire breathing dragon — temporarily blinding a predator while it backflips away into the darkness.

(BILD)

Bet you can't see me

Many ocean dwellers produce light from their bellies that exactly matches the color and intensity of sunlight some 800 meters below the surface. Underwater, the bioluminescence from this Abralia squid's belly lights blur together to form the perfect cloaking device, hiding its silhouette from upward-looking predators.

(BILD)

Help!

Once caught by a predator, this jellyfish's only hope for survival is to use its lights to attract the attention of a larger predator to attack its attacker and give it a chance to escape.

(BILD)

Touch me and I'll slime you

When attacked, this pelagic sea cucumber flashes a warning sign: Prepare to get slimed. Just as a bank uses dye packs on banknotes to foil robbers, the sea cucumber's glowing goo stains a predator long after an attack, making them easy targets in turn.

Footage courtesy of Edith Widder / Ocean Research and Conservation Association.

ABOUT THE AUTHOR

Dan Kedmey is a staff writer at TED, dedicated to covering any idea worth spreading. He previously wrote about business and technology for TIME Magazine. Dan has a Master's degree in literary reportage from New York University and a Bachelor's in economics from Carleton College.

Quelle: <https://ideas.ted.com/gallery-how-to-speak-jellyfish/>. (Stand: 08.09.2020).

Ausgangstext 4 und 12

Hakai Magazine Coastal Science and Societies www.hakaimagazine.com

(BILD)

Life Bustles in the Long Arctic Night

The Arctic's sunless season is far busier than anyone imagined.

by Kat Long

November 4, 2015 | 700 words, about 3 minutes

This story is over 4 years old.

In the fall of 1819, William Edward Parry and his crew prepared to spend the winter in the Canadian Arctic. For this team of British explorers, the unforgiving cold and round-the-clock darkness were obstacles to be overcome in their search for the Northwest Passage. Looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. “It was a novelty to us,” he later wrote in the account of the expedition, “to see any living animal in this desolate spot.”

From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. But now, researchers are discovering that the explorers had it all wrong.

A pair of new papers describing the latest discoveries reveals that the Arctic night is far from desolate: it's alive with activity. It's a time when microbes and animals feed, grow, and reproduce in the polar ocean, with special adaptations allowing them to thrive in the dark.

Many Arctic creatures have evolved to make use of light that is invisible to the human eye. Meanwhile, “an astonishing number” have developed bioluminescence, says Jørgen Berge, lead author of both new studies. Others live by moonlight, or by the soft red and green glow of the aurora borealis.

Berge, a marine biologist with UiT The Arctic University of Norway, is no stranger to life above the Arctic Circle. Berge has spent 13 years studying zooplankton in Svalbard, a group of mountainous, ice-covered islands far off the northern coast of Norway. Yet while cruising around a fjord in January 2012, he leaned over the side of his small

boat, peered into the water, and saw something new.

“I saw a fantastic cosmos of blue-green light in the middle of the fjord,” Berge says. “That's when I decided we needed to study the polar night.”

Berge and his colleagues have since discovered algae that grow without normal sunlight, and zooplankton that maintain their circadian cycles even without the Sun's cues. And with these minuscule organisms forming the base of the food chain, larger organisms have adapted to prey on them.

Berge found healthy seabirds in winter with stomachs full of krill and *Calanus glacialis*, a fat-rich copepod that he calls the “avocado of the ocean.” How the birds located their prey in the dark remains a mystery, but he suspects that they spot the planktons' faint bioluminescence.

Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule. “For me, seeing the bivalves grow through the polar night was the most surprising,” says biologist William G. Ambrose, Jr., a coauthor of the Current Biology study.

Life in the polar night is “far more active than we ever thought,” says Carin Ashjian, an oceanographer at Woods Hole Oceanographic Institute, who was not involved with the new papers. “This is a paradigm shift.”

Now, researchers are working to determine if the burgeoning life they see is common across the Arctic, or if it’s an abundance particular to Svalbard. Then, they’ll need to figure out the whys and hows. “It appears you can’t understand the system fully by just focusing on the period of time when the Sun is up,” Ambrose says wryly.

But the rapid onset of polar warming, driven by anthropogenic climate change, means that the time for discovery may be running out. “We need to first understand the processes currently occurring. We call this the baseline,” says Kim S. Last, who worked on both of the new papers. “Without a baseline we have no way of assessing change in the biological system.”

Like the old explorers, scientists still have much to learn about what goes on when the Sun goes down.

“We have lifted the curtain on the darkness,” Berge says, “and now we’re watching the active players on stage.”

Kat Long is a freelance journalist covering science, history, and culture. Her work has appeared in The Washington Post, Scientific American, Smithsonian.com, Slate, and other publications in the US and Canada. She lives in New York City.

Quelle: <https://www.hakaimagazine.com/news/life-bustles-long-arctic-night/>. (Stand: 08.09.2020).

Ausgangstext 5 und 7

T E C H

Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life

With help from a special Webcam, a marine biologist exposes undersea bioluminescence and helps illuminate the shady world of deep-sea life

By Ferris Jabr on August 5, 2010

If you trawl a net through the ocean's depths, chances are just about every living thing you haul to the surface will be able to glow. Marine biologists estimate that between 80 and 90 percent of deep-sea creatures are bioluminescent—they produce light through chemical processes.

Like the deep sea itself, the reasons why many of these organisms flash, twinkle and gleam remain mysterious to science. But in the past decade marine biologist Edith "Eddie" Widder has

tackled more puzzles about undersea bioluminescence than any other researcher. Widder has pioneered new technologies designed specifically to study bioluminescence and, in the process, discovered new species and recorded footage of never-before observed animal behavior. Now, as president and senior scientist of the Ocean Research and Conservation Association (ORCA), Widder is currently preparing to launch a new, cheaper and more portable version of her unique deep-sea camera, which she uses to observe rarely glimpsed ocean life completely undisturbed. She has plans to create a high-resolution color camera as well.

Seeing the light

In the 1980s, while completing her PhD in neurobiology at the University of California, Santa Barbara, Widder spent a lot of time studying the cellular activity of bioluminescent dinoflagellates, a variety of marine plankton. Inside certain cells of bioluminescent organisms an enzyme called luciferase catalyzes a light-producing reaction between a pigment called luciferin and oxygen. Different species have different variants of the enzyme and pigment, but oxygen is the essential ingredient that unites them all. Once Widder became an expert on both the biology of bioluminescence and the sophisticated equipment used to measure it, she began joining trawling expeditions that dragged nets to collect glowing ocean creatures and analyzed the light they emitted. On one of the research expeditions, Widder had the chance to test an atmospheric deep-sea diving suit called WASP, which resembles a submersible suit of armor. It was her first dive, but it would not be her last.

WASP allowed Widder to immerse herself in the ocean's bioluminescence—to really see the ubiquity of undersea illumination firsthand. One dive was particularly memorable. "I was trying to take some readings with a meter," Widder says, "when suddenly the whole inside of the suit lit up blue." Widder had brushed up against a chain of siphonophores—a colony of jellyfish relatives—sparking their light show. "It was breathtaking, absolutely breathtaking," Widder says. "Especially if you know how much energy is involved in that kind of photon production—it's so energetically costly for the siphonophores."

That experience helped Widder decide what to do with her PhD: make marine bioluminescence the center of her life's work, no matter what it took. She couldn't understand why more scientists were not investigating how and why ocean organisms produce light. And she wanted to answer all her own questions. But first, she needed to design instruments with which she could study bioluminescence—instruments that did not yet exist.

Glimmering gadgets

Although submarines and submersibles like WASP taught Widder a lot about bioluminescence and marine life in general, she wanted to observe but not disturb: no bulky vehicles, no lights, no humans in waterproof suits of armor—just the fish and an invisible observer. So she designed a camera called Eye-in-the-Sea. Widder first tried to secure funding for the project in 1994, but could not convince any single agency of the project's potential. Instead, she pieced together the camera bit by bit over many years with a series of collaborators. The instrument began as a student project for the Harvey Mudd College Engineering Clinic in the fall of 2000. Later, funding from the National Oceanic and Atmospheric Administration (NOAA) and the Monterey Bay Aquarium Research Institute (MBARI) enabled Widder to build the camera's frame and purchase an underwater battery. In 2002 Widder conducted preliminary tests in the Monterey Submarine Canyon off the California coast.

The Eye-in-the-Sea relies solely on an external source of far-red light to illuminate its surroundings. Most deep-sea creatures cannot see red light, because their eyes evolved to focus on the shorter blue and green wavelengths that travel furthest through water. Thanks to this fact,

red light allows the camera to observe deep-sea organisms without disturbing them. Because red light is readily absorbed by seawater, the Eye uses an ultrasensitive black-and-white camera that amplifies the dim far-red-light illumination and is also sensitive enough to record any bioluminescent creatures swimming within range.

But Widder didn't leave the camera alone and hope glimmering creatures would swim by. Instead, she used her knowledge about bioluminescence to design a unique lure. Widder created an electronic jellyfish that mimics the light shows of living bioluminescent jellyfish using a circle of 16 blue LEDs that flash in patterns. Specifically, the robot jellyfish mimics the atolla jellyfish (*Atolla wyvillei*), which employs light displays as a distress call when confronted by predators, signaling for even bigger predators to interfere and possibly rescue them. Her idea is that the robo-jelly would attract these same big marine predators to the camera.

Fishy footage

In 2004 Widder deployed the camera in the Gulf of Mexico. What she discovered there proved beyond a doubt how effective the Eye-in-the-Sea could be when paired with its electric jelly companion.

Fish swam up to and all around the underwater camera completely undisturbed. After four hours of collecting footage, Widder switched on the electronic jellyfish, which began its pinwheel light show. A mere 86 seconds after it started glowing the robot attracted a squid around two meters long—a species that no scientist had ever documented before. Widder thinks this is exactly the kind of big marine predator a living bioluminescent jellyfish might attract when under attack by smaller predators, in the hopes of saving itself.

In 2007 Widder—who received a 2006 MacArthur Fellows Program "genius" grant—deployed the Eye-in-the-Sea on three separate occasions during a nine-day cruise in the Bahamas. The camera recorded several different species of deep-sea sharks, including giant six-gill sharks rummaging on the seafloor—a never-before observed behavior. Widder thinks the fish were searching for giant isopods—ocean crustaceans that look something like a massive terrestrial wood louse (more commonly known as pillbugs or roly pollies).

Last year, Widder installed the Eye at a depth of almost 1,000 meters on the floor of Monterey Bay in California where it is plugged into the Monterey Accelerated Research System (MARS)—a high-tech networked underwater observatory. A cable links the camera to a command center on land, negotiating the constant exchange of power and video footage. It is not just a camera on the ocean floor; it is the first ever deep-sea Webcam. "It's basically a plug at the bottom of the ocean at around 990 meters, like a power bar with six outlets," Widder says. "This is the first stage of trying to wire the ocean."

Currently, Widder's postdoctoral fellow Erika Raymond of MBARI is reviewing images captured by the Eye in Monterey Bay and much of the older footage as well, searching for any new species or interesting phenomena—such as attacks by Humboldt squid on pig and seal carcasses Widder dropped near the Webcam to attract marine predators. Footage from the Bahamas, for example, revealed a kind of bioluminescent communication in which something rhythmically responded to the electronic jellyfish. "We got something to talk back to us in the Bahamas," Widder says. "It was very consistent—I think it was a shrimp."

Light conversation

In fact, communication is one of the main functions of underwater bioluminescence, Widder explains, enabling marine creatures to locate prey, attract mates and avoid predators. "What you [have] got to realize is that the open ocean environment is the largest habitat on the planet and

the animals out there have to play all the same games that animals play on land," Widder says. "They need places to hide, too, and ways to communicate and see each other—whether they are looking for food or sex. But they spend most of their life in this twilight realm. So they evolved lights."

Over the years, Widder's curiosity has helped science better understand how many of the strangest bioluminescent ocean creatures communicate and interact with one another. Consider the cookie-cutter shark: sure, it sounds adorable, but this shark is a devious hunter—and Widder suspects bioluminescence is a key part of its arsenal. The shark's bioluminescent stomach acts as an invisibility cloak that conceals its own shadow by matching the intensity of light shining from above—a trick called counterillumination. But a dark collar on its throat mimics the silhouette of a fish, luring bigger fish. When one of those hungry predators gets close enough to what it thinks is its next meal, it winds up as a meal itself: The shark takes a vicious chomp out of its deceived prey, leaving a cookie cutter– shaped wound.

If that is not strange enough, there's also the scaleless dragonfish (*Pachystomias microdon*), which can emit and perceive red light. Because most deep-sea creatures cannot see in this color, the dragonfish uses its bioluminescent organs like an infrared sniper scope to hunt, illuminating its surroundings without its prey noticing anything unusual.

Widder recently discovered that an octopus she calls the red balloon octopus (*Stauroteuthis syrtensis*) has bioluminescent suckers that are morphing into genuine light organs. "It corroborates the view of how bioluminescence evolves," Widder says. "It's an open-ocean octopus, so it doesn't really need suckers anymore to cling onto things on the bottom. But in the dark depth of [the] ocean, it needs light." Widder thinks the octopus uses its shimmering suckers to attract prey, like small crustaceans called copepods.

One tiny critter still baffles Widder: the shining tubeshoulder fish (*Sagamichthys abei*), which squirts not just chemical light—but entire cells—from its shoulder when stimulated. "Why that one fish would eject whole cells seems incredibly costly and odd," Widder says. "It may have evolved from some other mechanism we don't know anything about. I'm still mystified by that one."

Always on the glow

Soon, Widder plans to launch an updated version of the deep-sea Webcam, which is cheaper, lighter and more portable. "We're just completing a new version of the Eye-in-the-Sea camera system, which has been a collaboration between myself and Justin Marshall [of the University of Queensland in Australia]," Widder says. "We brainstormed about how to make the lowest cost version of the camera as possible, so we could have a lot of them. My version of it is about to come out of the shop in a matter of weeks. The original version is kind of cumbersome, whereas the new version is what we call a lander system: You throw it off the back of a boat, it floats down to the bottom, lands on the bottom, and keeps running the whole time. This new version will be used in a number of different scenarios, like looking at goliath grouper spawning behavior."

Widder also has plans to create a high-resolution version of the Webcam that can record in color, which would better help her distinguish between different kinds of bioluminescence. Although most deep-sea creatures emit light in the blue or green wavelengths, underwater bioluminescence comes in many more colors. "Blue is what travels farthest through seawater, but there are lots of different colors I am curious about: red, orange, yellow, violet and even extending into the ultraviolet," Widder says.

The applications of research on bioluminescence also excite Widder's curiosity. "We are using bioluminescence in a number of different ways to protect the ocean," she says. "We are using

bioluminescent bacteria to detect toxins in ocean sediments, develop pollution gradient maps and perform water quality monitoring. Since bioluminescence in bacteria is directly linked to the respiratory chain, anything that depresses respiration—like toxins—depresses illumination. That way, we can tell you exactly how unhealthy a sediment is."

A B O U T T H E A U T H O R (S)

Ferris Jabr

Ferris Jabr is a contributing writer for *Scientific American*. He has also written for the *New York Times Magazine*, the *New Yorker* and *Outside*.

Credit: Nick Higgins

Quelle: <https://www.scientificamerican.com/article/edith-widder-bioluminescence/>. (Stand: 08.09.2020).

Ausgangstext 6

Many organisms rely on bioluminescence to breed, feed and evade predators

By Ben Harder

Monday, October 11, 2010; 5:00 PM

(Seite 1)

For a moment, I cradled the Milky Way in my hands. There, cupped in my open palms, lay dozens upon dozens of glittering stars. One by one, they faded to black. Yet all I had to do was scoop a hand into the water surrounding us, and I held another fistful of starlight.

My wife and I were bobbing in a kayak in one of Puerto Rico's famed bioluminescent bays. It was night, and while there were true stars above us, the ones in my grasp were really dinoflagellates - single-celled organisms that emit light, or bioluminesce, whenever they're disturbed.

Bioluminescence is as widespread as it is wild and mysterious. Jack-o'-lantern mushrooms, flashlight fish and fireflies are among the multitude of organisms that bioluminesce. Scientists are still finding previously unknown examples of the phenomenon, especially at sea, where bioluminescent species are particularly varied and abundant. In parts of the ocean, 80 to 90 percent of sea creatures make light or harbor microbes that do so.

Experts are studying not only why marine life-forms developed the capacity to flash or glow, but also how and when they do it.

The "why" is worth pondering. If you're a defenseless bit of biota whose major preoccupation is hiding from predators, shining a light on your location doesn't immediately stand out as a winning strategy. And yet, bioluminescent organisms can protect themselves by making light. One way, believe it or not, is camouflage.

Not far beneath the waves lies a realm that's in almost perpetual twilight. Only a little sunshine or moonlight filters down, and everything is bathed in a faint blue. In this world, an easy way to catch prey is simply to look up - and attack any appetizing morsel silhouetted against the light above.

That's where bioluminescence can provide a chameleonesque disguise. Nearly all krill, the tiny crustaceans that are food for everything from small fish to massive baleen whales, have eyelike structures called photophores on their undersides. The photophores give off light of similar color and intensity to that shining down from above, making the krill semi-invisible from below.

Bioluminescence can also help fend off an attack. Some species glow as a warning that they're poisonous; others confuse or blind attackers by releasing clouds or jets of lightemitting chemicals while trying to escape.

Some marine organisms, if attacked, produce what scientists call a burglar alarm - a visual 911 call designed to attract a larger fish that's only too happy to gobble up the flashy prey's assailant.

This is thought to be the case with the dinoflagellates I encountered. By lighting up, the dinoflagellates, which are sought after as food by small crustaceans known as copepods, may be trying to draw a copepod-eating fish to the scene. In lab experiments, copepods tend to scoot away when dinoflagellates luminesce, suggesting the light display does work as a deterrent.

In general, any flash of light in the deep ocean is a signal that action's afoot. Like men in a bar fight, some organisms will flee a fracas and others will charge toward it. In a lecture posted on TED.com in April, marine bioluminologist and MacArthur "genius grant" awardee Edith Widder showed footage of a squid previously unknown to science; she had lured it to her camera using a flashing light that resembled a jellyfish in distress.

Bioluminescence is used not only to avoid getting eaten but also, sometimes for the opposite purpose. In the open ocean, a steady rain of organic debris - mostly dead marine organisms and their fecal matter - drifts down through the water. To survive, the bacteria aboard this sinking stuff desperately need to find their way into the gut of a new host organism. Glowing is a great way to get the attention of something that's willing to stomach the waste heap you call home.

© 2010 The Washington Post Company

Quelle:<https://www.washingtonpost.com/wp-dyn/content/article/2010/10/11/AR2010101104495.html>. (Stand: 08.09.2020).

Ausgangstext 9

Smithsonian Institution

BIOLUMINESCENCE

Introduction

by The Ocean Portal Team

You may have seen the sparkle of fireflies on a summer's night. The fireflies produce light through a chemical reaction in their glowing abdomens, a process known as bioluminescence. But did you know that seascapes can also glow and glitter thanks to the light producing abilities of **many marine organisms**? Some fish dangle a lighted lure in front of their mouths to attract

prey, while some squid shoot out bioluminescent liquid, instead of ink, to confuse their predators. Worms and tiny crustaceans also use bioluminescence to attract mates.

Humans primarily see bioluminescence triggered by a physical disturbance, such as waves or a moving boat hull, that gets the animal to show their light off, but often animals light up in response to an attack or in order to attract a mate. Bioluminescent organisms live throughout the water column, from the surface to the seafloor, from near the coast to the open ocean. In the deep sea, bioluminescence is extremely common, and because the deep sea is so vast, bioluminescence may be the most common form of communication on the planet!

[...]

Protection

(BILD)

This fish is using counterillumination to disappear. At left it stands out against the light above it. At right, with bioluminescent structures lit, it blends in. (*Smithsonian Institution*)

Often animals use a strong flash of bioluminescence to scare off an impending predator. The bright signal can **startle and distract** the predator and cause confusion about the whereabouts of its target. From small copepods to the larger vampire squid, this tactic can be very useful in the deep-sea. The "green bomber" worm (*Swima bombiviridis*) and four other similar worm species from the polychaete family **release a bioluminescent "bomb"** from their body when in harms way. These deep sea worms live close to the sea bottom and were only discovered in 2009. Some animals such as the deep-sea squid *Octopoteuthis deletron* even **detach their bioluminescent arms**, which stick to and probably distract their predators. All this commotion could also serve as a burglar alarm, attracting larger predators to the scene. In certain cases a predator might only get a bite of their prey, and the evidence will keep glowing from within its stomach.

Bioluminescence can also be used to help camouflage with the use of **counterillumination**. Photophores on the bottom side of an animal can match the dim light coming from the surface, making it harder for predators searching for prey from below to see what they are looking for.

Quelle: <https://ocean.si.edu/ocean-life/fish/bioluminescence>. (Stand: 08.09.2020).

Ausgangstext 10

Encyclopaedia Britannica, Inc.

Everything's Illuminated: 6 Bioluminescent Organisms

WRITTEN BY

Richard Pallardy

Richard Pallardy received a B.A. in English from Illinois State University in 2005. He was a research editor with Encyclopaedia Britannica, Inc. from 2008 to 2016 and worked on Britannica Blog from 2010...

BILD

time-lapse photo of fireflies Time-lapse image of fireflies in the Catskill Mountains, New York. Firefly light is a cold light, with approximately 100 percent of the energy given off as light and only a minute amount as heat. Only light in the visible spectrum is emitted. *s58y*

Light-producing, or bioluminescent, organisms occur across the spectrum of life—get it, spectrum? There are blinking bacteria, flaming fungus, shimmering squid, and flashing fish. (Interestingly, fish are the only bioluminescent vertebrates and there are no plants that produce light.) Nearly all of this light is produced by the oxidization of the molecule luciferin, a reaction catalyzed by the enzyme luciferase. Some creatures produce their own light while others house bacteria that do it for them. They glow and flash for a variety of purposes—repelling predators, attracting predators *of* their predators, seducing potential mates, and luring prey. These fairy lights are in fact a coruscating dance of life, death, and sex.

Jellyfish

BILD

water jelly (*Aequorea victoria*) The medusa stage of the bioluminescent water jelly or crystal jelly (*Aequorea victoria*) in the waters off the west coast of North America. The species is harvested for its luminescent protein called aequorin, which has been used in medical research as a fluorescent marker protein. *Denise Allen*

More than half of all jellyfish species produce some kind of bioluminescence, largely as a deterrent to predators. Some even secrete lambent mucus to further confuse potential executioners. *Aequorea victoria*, the jellyfish pictured here, produces a green glow around the rim of its bell by passing its initially blue light through a compound known as green fluorescent protein (GFP). The gene encoding for this protein was synthesized in a lab and is now artificially inserted into the DNA sequences of other species so that certain areas of their bodies might be more easily identified by scientists. This work won the 2008 Nobel Prize for Chemistry.

Squid

BILD

bobtail squid (*Euprymna berryi*) *Silke Baron*

Many species of squid produce bioluminescence, using it for a variety of purposes. Some deep-sea squid spurt glowing ink or mucus to confuse their predators. Others, like the adorably tiny bobtail squid, pictured here, use their light in even more sophisticated ways. The bobtail squid possess small cavities on the underside of its body that are home to bioluminescent bacteria. When it emerges from the sand at night, the openings of the cavities expand and contract in response to the amount of moonlight penetrating the waves above so that its silhouette is less visible to predators lurking below.

Anglerfish

BILD

Anglerfish *Linophryne bicornis*, a species of bearded sea devil. Bearded sea devils are a type of anglerfish. *Painted especially for Encyclopædia Britannica by Tom Dolan, under the supervision of Loren P. Woods, Chicago Natural History Museum*

Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who led travelers astray at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. Other fish combing the murk for food are attracted to the fleshy lanterns and, when they approach, are inhaled by the beast at the other end. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic “ball and chain” relationship idiom, they attach themselves to her body, which eventually absorbs all but their testicles.

Lanternfish

BILD

Lantern fish (*Symbolophorus veranys*) *Encyclopædia Britannica, Inc.*

The lanternfishes are among the most numerous fish in the sea. Each of the more than 200 species has a unique pattern of light-generating photophores, used in signaling and mating. The photophores consist of lens-shaped, transparent scales covering the tissue that emanates light; these lenses help to focus and amplify the light. Recent research has shown that lanternfishes are diversifying more rapidly than bristlemouths, another bioluminescent deep sea species. Unlike lanternfishes, bristlemouths only use their bioluminescence as a camouflage mechanism.

Firefly

BILD

Firefly *Terry Priest*

Who didn't participate in the innocent childhood ritual of interrupting firefly orgies and stuffing the frightened insect lovers into make-shift “lanterns” that soon expired because someone forgot to punch holes in the top? Indeed, crashing the mating dance of these benignly beaming

beetles is quintessential to the idyllic summers of youth. And callous children aren't the only risk to fireflies a'courtin'. The females of some species mimic the light signals of other species, luring the males to the ground, and devouring them.

Fungi

BILD

bitter oyster (*Panellus stipticus*) Bitter oyster (*Panellus stipticus*), a saprophytic fungus capable of bioluminescence. Classified in the family Mycenaceae, the species is found in Eurasia, Australia, and North America. *Ylem*

Over 70 species of fungus are known to bioluminesce. Unlike other light-producing organisms, they glow around the clock, though of course their eerie glimmer is most visible in the dark. Most brightly luminescent fungi grow in the tropics, where their fruiting bodies—the mushroom, or sporocarp—may be highly visible. The mycelium, or body, of some other species may also glow, lighting up the rotting wood on which they subsist in a phenomenon known colloquially as “foxfire.” Fungi may glow in order to repel predators, to attract insects that may help them disperse their spores, or as an accidental byproduct of their metabolism of wood.

Quelle: <https://www.britannica.com/list/6-bioluminescent-organisms>. (Stand: 08.09.2020)

Ausgangstext 11

The Glowing Hawksbill Sea Turtle

by AMNH on Jan 9, 2019 3:19 pm

BILD

Some researchers think the biofluorescent glow of the shells of Hawksbill sea turtles (*Eretmochelys imbricata*) is caused by algae. B. navez/Wikimedia Commons

In the special exhibition *Unseen Oceans*, visitors come upon a brilliant floor-to-ceiling spiral of models of underwater species biofluorescing just as they would in life. Among the catsharks, breems, and other fishes reflecting neon greens and reds is a glowing reptile: a hawksbill sea turtle.

Only three years ago, this species, *Eretmochelys imbricata*, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have made the cut. But on a night dive during a July 2015 expedition to the Solomon Islands in search of biofluorescent corals, Museum researchers discovered that the hawksbill sea turtle glows, too.

They were using a special camera equipped with blue lights when, about 65 feet (20 meters) down, a bright-colored turtle suddenly popped into view like a “UFO ... a bright red and green spaceship,” recalls Museum Research Associate David Gruber. Scientists suspect that some, if not all, of the red glow might be produced by algae living on the turtles' shells—but more research is needed.

Meanwhile, another group of scientists at the Museum has spent more than a decade studying hawksbill sea turtles at the Palmyra Atoll National Wildlife Refuge, a remote ring of islets and associated marine habitats in the Pacific Ocean.

BILD - Hawksbill sea turtles help provide better access to corals for reef fish to feed by plucking out sponges, the turtles' main food source, from the reef's surface. Courtesy of Christian Gloor/Wikimedia Commons

When the research program began in 2005—run by the Museum’s Center for Biodiversity and Conservation (CBC) along with scientists from the City University of New York, U.S. Geological Survey, and National Institute for Standards and Technology—very little was known about the highly migratory species and its life at the atoll.

Since then, in-water habitat studies using acoustic and satellite telemetry have advanced understanding of these turtles—as well as that of Palmyra’s green sea turtles (*Chelonia mydas*), their genetics, interactions with their habitats, feeding habits, and the health of both populations, including their distribution and movement around the atoll and the broader Pacific.

“By studying the ecology and behavior of the healthy turtles at this critical foraging ground, we are able to deepen understanding of these endangered species and also inform conservation management plans for the Atoll and across the Pacific,” says Eleanor Sterling, Jaffe Chief Conservation Scientist at the CBC.

Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, their main food source, from the reef’s surface. This in turn provides better access for reef fish to feed and may give the corals a favorable edge in the competition with the sponges for space.

A version of this story appeared in the Winter issue of the Member magazine, Rotunda.

Quelle:

<https://www.amnh.org/explore/news-blogs/news-posts/the-glowing-hawksbill-sea-turtle>.
(Stand: 08.09.2020).

Anhang 2

Englische Textpassagen mit jeweiligen angewandten Pre-Edits und entsprechenden resultierenden Übersetzungen ins Italienische, je nach Methode.

Farbenlegende Pre-Editing-Methoden

- Pre-Editing-Methode 1** = Koordinationen (bzw. Sätze) eindeutig machen, Hauptsubstantiv mit verbundenen Artikeln und Präpositionen wiederholen
- Pre-Editing-Methode 2** = Metaphern und Redewendungen vermeiden bzw. umschreiben
- Pre-Editing-Methode 3** = Keine Verben mit Partikel, sondern Verben aus einem Wort verwenden
- Pre-Editing-Methode 4** = Ing-Wörter umschreiben: mit einem Relativsatz ersetzen oder eine passende Präposition hinzufügen (wie „by“ zum Beispiel)
- Pre-Editing-Methode (Kombination) 5** = Verwendung von (Possessiv) Pronomen vermindern + jedes Segment muss selbstständig syntaktisch stimmen
- Pre-Editing-Methode (Kombination) 6** = Lange Nominalsätze vermeiden + mit Relativpronomen umschreiben

<p>AT 1 kein Pre-Editing</p> <p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glowin-the-dark plankton in a see-through kayak, spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and bottomless cookies.</p> <p>Quelle: Etherton, Caitlin (o.J.). See nature’s amazing glow-in-the-dark spectacles. National Geographic Society, 09.01.2019. https://www.nationalgeographic.com/travel/lists/bioluminescence-best-glow-in-dark-spots-around-world/ (Stand: 08.09.2020).</p>	<p>ZT 1.1</p> <p>Tre ore a nord, BK Adventure offre tour eco-compatibili nel Merritt Island Wildlife Refuge, dove è possibile accarezzare il plancton glowin-the-dark in un kayak trasparente, spot manatebots e gelatine bioluminescenti su una paletta da picnic al tramonto, o scivolare attraverso il blu in un robusto gommone - una grande opzione per bambini o persone a mobilità limitata. Le guide suggeriscono di recarsi in una notte senza luna, portando maniche lunghe e spray di insetti a base vegetale, e prenotare la linea fluviale B&B Casa Coquina del Mar per rilassarsi nella vasca idromassaggio, colazione fatta in casa e biscotti imbottigliati.</p>
<p>AT 1 Pre-Editing-Methode 1</p> <p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glowin-the-dark plankton in a see-through kayak and you can spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and bottomless cookies.</p>	<p>ZT 1.2</p> <p>Tre ore a nord, BK Adventure offre dei tour eco-compatibili nel Merritt Island Wildlife Refuge, dove si può brillare attraverso il plancton glowin-the-dark in un kayak trasparente e si possono vedere manatebotti e gelatine bioluminescenti su una paletta da picnic al tramonto, o scivolare attraverso il blu in un robusto veicolo a leva, una grande opzione per i bambini o quelli con mobilità limitata. Le guide suggeriscono di recarsi in una notte senza luna, portando maniche lunghe e spray di insetti a base vegetale, e prenotare la linea fluviale B&B Casa Coquina del Mar per rilassarsi nella vasca idromassaggio, colazione fatta in casa e biscotti imbottigliati.</p>
<p>AT 1 Pre-Editing-Methode 6 (K)</p> <p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through plankton that glows in the dark in a see-through kayak, spot manatees and bioluminescent comb jellies on a picnic paddle at sunset, or glide through the blue in a sturdy paddle raft—a great option for</p>	<p>ZT 1.3</p> <p>Tre ore a nord, BK Adventure offre visite eco-compatibili nel Merritt Island Wildlife Refuge, dove si può accarezzare il plancton che risplende all'oscuro in un kayak trasparente, manateette a posto e gelatine a pettine bioluminescenti in una pala da picnic al tramonto, o scivolare attraverso il blu in una navetta soffice - una grande</p>

children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for lounging in a post-paddle hot tub , homemade breakfast, and bottomless cookies.	opzione per i bambini o con quelli con mobilità limitata. Le guide suggeriscono di recarsi in una notte senza luna, portando maniche lunghe e spray di insetti a base vegetale, e prenotare la linea fluviale B&B Casa Coquina del Mar per rilassarsi in una vasca idromassaggio post-remo , colazione fatta in casa e biscotti imbottigliati.
--	---

AT 1 Pre-Editing-Methode 2	ZT 1.4
Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glow-in-the-dark plankton in a see-through kayak, spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and a bottomless pit of cookies .	Tre ore a nord, BK Adventure offre tour eco-compatibili nel Merritt Island Wildlife Refuge, dove è possibile accarezzare il plancton glow-in-the-dark in un kayak trasparente, spot manatees e gelatine bioluminescenti su una paletta da picnic al tramonto, o scivolare attraverso il blu in un robusto gommone - una grande opzione per bambini o persone a mobilità limitata. Le guide suggeriscono di remare per una notte senza luna, portare maniche lunghe e spray per insetti a base di piante, e prenotare il lato fiume B&B Casa Coquina del Mar per una vasca da bagno calda dopo la pagaia, colazione fatta in casa e una buca senza fondo di biscotti .

AT 2 kein Pre-Editing	ZT 2.1
Bet you can't see me Many ocean dwellers produce light from their bellies that exactly matches the color and intensity of sunlight some 800 meters below the surface. Underwater, the bioluminescence from this Abralia squid's belly lights blur together to form the perfect cloaking device, hiding its silhouette from upward-looking predators. Quelle: Kedmey, Dan (2016). Gallery: How to speak jellyfish. TED Conferences, LLC., 05.01.2016. https://ideas.ted.com/gallery-how-to-speak-jellyfish/ (Stand: 08.09.2020).	Scommetto che non puoi vedermi Molti abitanti dell'oceano producono luce dalla pancia che corrisponde esattamente al colore e all'intensità della luce solare a circa 800 metri sotto la superficie. Sotto l'acqua, la bioluminescenza della pancia di questo calamaro dell'Abralia si illumina per formare il perfetto dispositivo di occultamento, nascondendo la sua silhouette da predatori dall'aspetto ascendente.

AT 2 Pre-Editing-Methode 1	ZT 2.2
Bet you can't see me Many ocean dwellers produce light from their bellies that exactly matches the color and intensity of sunlight some 800 meters below the surface. Underwater, the	Scommetto che non puoi vedermi Molti abitanti dell'oceano producono luce dalla pancia che corrisponde esattamente al colore e all'intensità della luce solare a circa 800 metri sotto la superficie. Sotto l'acqua, la pancia bioluminescente si illumina dalla

bioluminescent belly lights from this Abralia squid's blur together to form the perfect cloaking device, hiding its silhouette from upward-looking predators.	sfocatura di questo calamaro Abralia per formare il perfetto dispositivo di occultamento, nascondendo la sua silhouette da predatori dall'aspetto ascendente.
AT 2 Pre-Editing-Methode 4	ZT 2.3
Bet you can't see me Many ocean dwellers produce light from their bellies that exactly matches the color and intensity of sunlight some 800 meters below the surface. Underwater, the bioluminescence from this Abralia squid's belly lights blur together to form the perfect cloaking device, which hides its silhouette from upward-looking predators.	Scommetto che non puoi vedermi Molti abitanti dell'oceano producono luce dalla pancia che corrisponde esattamente al colore e all'intensità della luce solare a circa 800 metri sotto la superficie. Sotto l'acqua, la bioluminescenza della pancia di questo calamaro dell'Abralia si illumina per formare il dispositivo di occultamento perfetto, che nasconde la sua silhouette da predatori dall'alto.
AT 2 Pre-Editing-Methode 5 (K)	ZT 2.4
Bet you can't see me Many ocean dwellers produce light from their bellies that exactly matches the color and intensity of sunlight some 800 meters below the surface. Underwater, the bioluminescence from this Abralia squid's belly lights blur together to form the perfect cloaking device, hiding the silhouette of the squid from upward-looking predators.	Scommetto che non puoi vedermi Molti abitanti dell'oceano producono luce dalla pancia che corrisponde esattamente al colore e all'intensità della luce solare a circa 800 metri sotto la superficie. Sotto l'acqua, la bioluminescenza della pancia di questo calamaro dell'Abralia si illumina per formare il dispositivo di mascheramento perfetto, nascondendo la silhouette del calamaro dai predatori dall'aspetto più alto.

AT 3 kein Pre-Editing	ZT 3.1
Touch me and I'll slime you When attacked, this pelagic sea cucumber flashes a warning sign: Prepare to get slimed. Just as a bank uses dye packs on banknotes to foil robbers, the sea cucumber's glowing goo stains a predator long after an attack, making them easy targets in turn. Quelle: Kedmey, Dan (2016). Gallery: How to speak jellyfish. TED Conferences, LLC., 05.01.2016. https://ideas.ted.com/gallery-how-to-speak-jellyfish/ (Stand: 08.09.2020).	Toccami e ti rimpicciolisco Quando viene attaccato, questo cetriolo di mare pelagico lampeggia un segnale di avvertimento: Preparatevi a rimpicciolire. Proprio come una banca usa confezioni di coloranti sulle banconote per sventare i rapinatori, l'allegro cetriolo marino colora un predatore molto tempo dopo un attacco, rendendoli a loro volta bersagli facili.
AT 3 Pre-Editing-Methode 1	ZT 3.2
Touch me and I'll slime you When this pelagic sea cucumber is attacked, it flashes a warning sign: Prepare to get slimed. Just as a bank uses dye packs on banknotes to foil robbers, the sea cucumber's glowing	Toccami e ti rimpicciolisco Quando questo cetriolo pelagico viene attaccato, mostra un segnale di avvertimento: Preparatevi a rimpicciolire. Proprio come una banca usa confezioni di coloranti sulle banconote per sventare i rapinatori, l'allegro cetriolo marino colora un

goo stains a predator long after an attack, making them easy targets in turn.	predatore molto tempo dopo un attacco, rendendoli a loro volta bersagli facili.
AT 3 Pre-Editing-Method 4	ZT 3.3
Touch me and I'll slime you When attacked, this pelagic sea cucumber flashes a warning sign: Prepare to get slimed. Just as a bank uses dye packs on banknotes to foil robbers, the sea cucumber's glowing goo stains a predator long after an attack, which makes them easy targets in turn.	Toccamì e ti rimpicciolisco Quando viene attaccato, questo cetriolo di mare pelagico lampeggia un segnale di avvertimento: Preparatevi a rimpicciolire. Proprio come una banca utilizza confezioni di coloranti sulle banconote per sventare i ladri, il goo luccicante del cetriolo marino blocca un predatore molto dopo un attacco, il che a sua volta li rende facili bersagli.
AT 3 Pre-Editing-Method 5 (K)	ZT 3.4
Touch me and I'll slime you When attacked, this pelagic sea cucumber flashes a warning sign: Prepare to get slimed. Just as a bank uses dye packs on banknotes to foil robbers, the sea cucumber's glowing goo stains a predator long after an attack, making the predators easy targets in turn.	Toccamì e ti rimpicciolisco Quando viene attaccato, questo cetriolo di mare pelagico lampeggia un segnale di avvertimento: Preparatevi a rimpicciolire. Proprio come una banca usa confezioni di coloranti sulle banconote per sventare i ladri, il goo splendente del cetriolo marino blocca un predatore molto dopo un attacco, rendendo i predatori facili bersagli a loro volta.

AT 4 kein Pre-Editing	ZT 4.1
Looking out across the flat, treeless expanse of ice and fog , the barren landscape must have seemed wholly inhospitable. [...] From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule . Quelle: Long, Kat (2015). Life Bustles in the Long Arctic Night. <i>Hakai Magazine</i> 04.11.2015. https://www.hakaimagazine.com/news/life-bustles-long-arctic-night/ (Stand: 08.09.2020).	Guardando oltre la piatta, senza tregua di ghiaccio e nebbia , il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti, tranne il più difficile della vita, sono fuggiti per scalate più calde, o si sono accovacciati per aspettare il ritorno del Sole. [...] Le capesante islandesi, nel frattempo, possono rilevare una quantità sufficiente di luce ambientale dal debole bagliore del sole all'orizzonte per nuotare e nutrirsi con un programma tipo estivo .

AT 4 Pre-Editing-Method 4	ZT 4.2
<p>By looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule.</p>	<p>Guardando attraverso la piatta, senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti, tranne il più difficile della vita, sono fuggiti per scalate più calde, o si sono accovacciati per aspettare il ritorno del Sole. [...] Le capesante islandesi, nel frattempo, possono rilevare una quantità sufficiente di luce ambientale dal debole bagliore del sole all'orizzonte per nuotare e nutrirsi con un programma tipo estivo.</p>
AT 4 Pre-Editing-Method 1	ZT 4.3
<p>Looking out across the flat and treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule.</p>	<p>Guardando oltre la piatta e senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti, tranne il più difficile della vita, sono fuggiti per scalate più calde, o si sono accovacciati per aspettare il ritorno del Sole. [...] Le capesante islandesi, nel frattempo, possono rilevare una quantità sufficiente di luce ambientale dal debole bagliore del sole all'orizzonte per nuotare e nutrirsi con un programma tipo estivo.</p>
AT 4 Pre-Editing-Method 6 (K)	ZT 4.4
<p>Looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a schedule like that in summer.</p>	<p>Guardando oltre la piatta, senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti, tranne il più difficile della vita, sono fuggiti per scalate più calde, o si sono accovacciati per aspettare il ritorno del Sole. [...] Le capsule islandesi, nel frattempo, possono rilevare una quantità di luce ambientale sufficiente dal bagliore debole del sole</p>

	all'orizzonte per nuotare e nutrirsi con un programma come quello in estate.
--	---

AT 5 kein Pre-Editing	ZT 5.1
<p>The shark's bioluminescent stomach acts as an invisibility cloak that conceals its own shadow by matching the intensity of light shining from above—a trick called counterillumination. But a dark collar on its throat mimics the silhouette of a fish, luring bigger fish. When one of those hungry predators gets close enough to what it thinks is its next meal, it winds up as a meal itself: The shark takes a vicious chomp out of its deceived prey, leaving a cookie cutter-shaped wound.</p> <p>Quelle: Jabr, Ferris (2010). Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life. <i>Scientific American Inc.</i>, 05.08.2010. https://www.scientificamerican.com/article/edith-widder-bioluminescence/ (Stand: 08.09.2020).</p>	<p>Lo stomaco bioluminescente dello squalo agisce come un mantello di invisibilità che nasconde la propria ombra abbinando l'intensità della luce che brilla dall'alto — un trucco chiamato controilluminazione. Ma un collare scuro sulla gola imita la siluetta di un pesce, che attira pesci più grandi. Quando uno di questi predatori affamati si avvicina abbastanza a quello che pensa sia il suo prossimo pasto, finisce per diventare un pasto: Lo squalo estrae un fruscio vizioso dalla sua preda ingannata, lasciando una ferita a forma di taglio di biscotto.</p>
AT 5 Pre-Editing-Method 3	ZT 5.2
<p>The shark's bioluminescent stomach acts as an invisibility cloak that conceals its own shadow by matching the intensity of light shining from above—a trick called counterillumination. But a dark collar on its throat mimics the silhouette of a fish, luring bigger fish. When one of those hungry predators gets close enough to what it thinks is its next meal, it winds up as a meal itself: The shark chomps viciously its deceived prey, leaving a cookie cutter-shaped wound.</p>	<p>Lo stomaco bioluminescente dello squalo agisce come un mantello di invisibilità che nasconde la propria ombra abbinando l'intensità della luce che brilla dall'alto — un trucco chiamato controilluminazione. Ma un collare scuro sulla gola imita la siluetta di un pesce, che attira pesci più grandi. Quando uno di questi predatori affamati si avvicina abbastanza a quello che pensa sia il suo prossimo pasto, finisce per diventare un pasto: Lo squalo taglia brutalmente la sua preda ingannata, lasciando una ferita a forma di biscotto.</p>
AT 5 Pre-Editing-Method 4	ZT 5.3
<p>The shark's bioluminescent stomach acts as an invisibility cloak that conceals its own shadow by matching the intensity of light that shines from above—a trick called counterillumination. But a dark collar on its throat mimics the silhouette of a fish, which lures bigger fish. When one of those hungry predators gets close enough to what it thinks is its next meal, it winds up as a meal itself:</p>	<p>Lo stomaco bioluminescente dello squalo agisce come un mantello d'invisibilità che nasconde la propria ombra abbinando l'intensità della luce che risplende dall'alto - un trucco chiamato controilluminazione. Ma un collare scuro sulla gola imita la silhouette di un pesce, che attira pesci più grandi. Quando uno di questi predatori affamati si avvicina abbastanza a quello che pensa sia il</p>

The shark takes a vicious chomp out of its deceived prey, which leaves a cookie cutter–shaped wound.	suo prossimo pasto, finisce per diventare un pasto: Lo squalo estrae un fruscio vizioso dalla sua preda ingannata, che lascia una ferita a forma di taglio di biscotto.
AT 5 Pre-Editing-Method 6 (K)	ZT 5.4
The shark's bioluminescent stomach acts as an invisibility cloak that conceals its own shadow by matching the intensity of light shining from above—a trick called counterillumination. But a dark collar on its throat mimics the silhouette of a fish, luring bigger fish. When one of those hungry predators gets close enough to what it thinks is its next meal, it winds up as a meal itself: The shark takes a vicious chomp out of its deceived prey, leaving a wound in the shape of a cookie cutter.	Lo stomaco bioluminescente dello squalo agisce come un mantello di invisibilità che nasconde la propria ombra abbinando l'intensità della luce che brilla dall'alto — un trucco chiamato controilluminazione. Ma un collare scuro sulla gola imita la siluetta di un pesce, che attira pesci più grandi. Quando uno di questi predatori affamati si avvicina abbastanza a quello che pensa sia il suo prossimo pasto, finisce per diventare un pasto: Lo squalo estrae un fruscio vizioso dalla sua preda ingannata, lasciando una ferita a forma di taglierina.

AT 6 kein Pre-Editing	ZT 6.1
By lighting up, the dinoflagellates, which are sought after as food by small crustaceans known as copepods, may be trying to draw a copepod-eating fish to the scene. In lab experiments, copepods tend to scoot away when dinoflagellates luminesce, suggesting the light display does work as a deterrent. Quelle: Harder, Ben (2010). Many organisms rely on bioluminescence to breed, feed and evade predators. <i>The Washington Post</i> , 11.10.2010. https://www.washingtonpost.com/wp-dyn/content/article/2010/10/11/AR2010101104495.html (Stand: 08.09.2020).	Con l'illuminazione, i dinoflagellati, che vengono ricercati come cibo da piccoli crostacei noti come copepodi, potrebbero cercare di disegnare sulla scena un pesce che mangia copepodi . Negli esperimenti in laboratorio, i copepodi tendono a scovare quando i dinoflagellati emettono luminescenza, suggerendo che il display della luce funzioni come deterrente.
AT 6 Pre-Editing-Method 3	ZT 6.2
By lighting up, the dinoflagellates, which are sought after as food by small crustaceans known as copepods, may be trying to draw a copepod-eating fish to the scene. In lab experiments, copepods tend to flee when dinoflagellates luminesce, suggesting the light display does work as a deterrent.	Con l'illuminazione, i dinoflagellati, che vengono ricercati come cibo da piccoli crostacei noti come copepodi, potrebbero cercare di disegnare sulla scena un pesce che mangia copepodi. Negli esperimenti in laboratorio, i copepodi tendono a fuggire quando i dinoflagellati emettono luminescenza, suggerendo che il display della luce funzioni come deterrente.
AT 6 Pre-Editing-Method 4	ZT 6.3
By lighting up, the dinoflagellates, which are sought after as food by small crustaceans	Con l'illuminazione, i dinoflagellati, che vengono ricercati come cibo da piccoli

known as copepods, may be trying to draw a copepod-eating fish to the scene. In lab experiments, copepods tend to scoot away when dinoflagellates luminesce, which suggests that the light display does work as a deterrent.	crostacei noti come copepodi, potrebbero cercare di disegnare sulla scena un pesce che mangia copepodi. Negli esperimenti di laboratorio, i copepodi tendono a sgretolarsi quando i dinoflagellati si accendono, il che suggerisce che il display della luce funziona come deterrente.
AT 6 Pre-Editing-Methode 6 (K)	ZT 6.4
By lighting up, the dinoflagellates, which are sought after as food by small crustaceans known as copepods, may be trying to draw a fish that eats copepods to the scene. In lab experiments, copepods tend to scoot away when dinoflagellates luminesce, suggesting the display of light does work as a deterrent.	Con l'illuminazione, i dinoflagellati, che vengono ricercati come cibo da piccoli crostacei noti come copepodi, potrebbero cercare di disegnare un pesce che mangia copepodi sulla scena. Negli esperimenti di laboratorio, i copepodi tendono a sgretolarsi quando i dinoflagellati si accendono, suggerendo che la visualizzazione della luce funzioni come deterrente.

AT 7 kein Pre-Editing	ZT 7.1
Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life With help from a special Webcam, a marine biologist exposes undersea bioluminescence and helps illuminate the shady world of deep-sea life Like the deep sea itself, the reasons why many of these organisms flash, twinkle and gleam remain mysterious to science. Quelle: Jabr, Ferris (2010). Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life. <i>Scientific American Inc.</i> , 05.08.2010. https://www.scientificamerican.com/article/edith-widder-bioluminescence/ (Stand: 08.09.2020).	Pulire il Gleam: Una webcam in profondità fa luce sulla vita oceanica bioluminescente Con l'aiuto di una webcam speciale, un biologo marino espone la bioluminescenza sottomarina e aiuta ad illuminare il buio mondo della vita in profondità Come lo stesso mare profondo, le ragioni per cui molti di questi organismi lampeggiano, brillano e brillano rimangono misteriose per la scienza.

AT 7 Pre-Editing-Methode 2	ZT 7.2
Discovering the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life With help from a special Webcam, a marine biologist exposes undersea bioluminescence and helps illuminate the shady world of deep-sea life	Scoprire il Gleam: Una webcam in profondità fa luce sulla vita oceanica bioluminescente Con l'aiuto di una webcam speciale, un biologo marino espone la bioluminescenza sottomarina e aiuta ad illuminare il buio mondo della vita in profondità

Like the deep sea itself, the reasons why many of these organisms flash, twinkle and gleam remain mysterious to science.	Come lo stesso mare profondo, le ragioni per cui molti di questi organismi lampeggiano, brillano e brillano rimangono misteriose per la scienza.
AT 7 Pre-Editing-Method 1	ZT 7.3
<p>Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life With help from a special Webcam, a marine biologist exposes undersea bioluminescence and helps illuminate the shady world of deep-sea life</p> <p>The reasons why many of these organisms flash, twinkle and gleam remain mysterious to science like the deep sea itself.</p>	<p>Pulire il Gleam: Una webcam in profondità fa luce sulla vita oceanica bioluminescente Con l'aiuto di una webcam speciale, un biologo marino espone la bioluminescenza sottomarina e aiuta ad illuminare il buio mondo della vita in profondità</p> <p>Le ragioni per cui molti di questi organismi lampeggiano, brillano e brillano rimangono misteriose per la scienza come le profondità marine stesse.</p>
AT 7 Pre-Editing-Method 5 (K)	ZT 7.4
<p>Gleaning the Gleam: A Deep-Sea Webcam Sheds Light on Bioluminescent Ocean Life With help from a special Webcam, a marine biologist exposes undersea bioluminescence and helps illuminate the shady world of deep-sea life</p> <p>Like the deep sea itself, the reasons why many of these organisms flash, twinkle and gleam remain mysterious to science.</p>	<p>Pulire il Gleam: Una webcam in profondità fa luce sulla vita oceanica bioluminescente Con l'aiuto di una webcam speciale, un biologo marino espone la bioluminescenza sottomarina e aiuta ad illuminare il buio mondo della vita in profondità</p> <p>Come le acque profonde, le ragioni per cui molti di questi organismi lampeggiano, brillano e brillano rimangono misteriose per la scienza.</p>

AT 8 kein Pre-Editing	ZT 8.1
<p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glowin-the-dark plankton in a see-through kayak, spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and bottomless cookies.</p> <p>Quelle: Etherton, Caitlin (o.J.). See nature’s amazing glow-in-the-dark spectacles.</p>	<p>Tre ore a nord, BK Adventure offre tour eco-compatibili nel Merritt Island Wildlife Refuge, dove è possibile accarezzare il plancton glowin-the-dark in un kayak trasparente, spot manatebots e gelatine bioluminescenti su una paletta da picnic al tramonto, o scivolare attraverso il blu in un robusto gommone - una grande opzione per bambini o persone a mobilità limitata. Le guide suggeriscono di recarsi in una notte senza luna, portando maniche lunghe e spray di insetti a base vegetale, e prenotare la linea fluviale B&B Casa Coquina del Mar per rilassarsi nella vasca idromassaggio, colazione fatta in casa e biscotti imbottigliati.</p>

<p>National Geographic Society, 09.01.2019. https://www.nationalgeographic.com/travel/lists/bioluminescence-best-glow-in-dark-spots-around-world/ (Stand: 08.09.2020).</p>	
<p>AT 8 Pre-Editing-Methode 1</p>	<p>ZT 8.2</p>
<p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glowin-the-dark plankton in a see-through kayak and you can spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and bottomless cookies.</p>	<p>Tre ore a nord, BK Adventure offre dei tour eco-compatibili nel Merritt Island Wildlife Refuge, dove si può brillare attraverso il plancton glowin-the-dark in un kayak trasparente e si possono vedere manatebotti e gelatine bioluminescenti su una paletta da picnic al tramonto, o scivolare attraverso il blu in un robusto veicolo a leva, una grande opzione per i bambini o quelli con mobilità limitata. Le guide suggeriscono di recarsi in una notte senza luna, portando maniche lunghe e spray di insetti a base vegetale, e prenotare la linea fluviale B&B Casa Coquina del Mar per rilassarsi nella vasca idromassaggio, colazione fatta in casa e biscotti imbottigliati.</p>
<p>AT 8 Pre-Editing-Methode 6 (K)</p>	<p>ZT 8.3</p>
<p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through plankton that glows in the dark in a see-through kayak, spot manatees and bioluminescent comb jellies on a picnic paddle at sunset, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the riverside B&B Casa Coquina del Mar for lounging in a post-paddle hot tub, homemade breakfast, and bottomless cookies.</p>	<p>Tre ore a nord, BK Adventure offre visite eco-compatibili nel Merritt Island Wildlife Refuge, dove si può accarezzare il plancton che risplende all'oscuro in un kayak trasparente, manateette a posto e gelatine a pettine bioluminescenti in una pala da picnic al tramonto, o scivolare attraverso il blu in una navetta soffice - una grande opzione per i bambini o con quelli con mobilità limitata. Le guide suggeriscono di recarsi in una notte senza luna, portando maniche lunghe e spray di insetti a base vegetale, e prenotare la linea fluviale B&B Casa Coquina del Mar per rilassarsi in una vasca idromassaggio post-remo, colazione fatta in casa e biscotti imbottigliati.</p>
<p>AT 8 Pre-Editing-Methode 2</p>	<p>ZT 8.4</p>
<p>Three hours north, BK Adventure offers eco-friendly tours in Merritt Island Wildlife Refuge, where you can glitter through glowin-the-dark plankton in a see-through kayak, spot manatees and bioluminescent comb jellies on a sunset picnic paddle, or glide through the blue in a sturdy paddle raft—a great option for children or those with limited mobility. Guides suggest paddling on a moonless night, bringing long sleeves and plant-based bug spray, and booking the</p>	<p>Tre ore a nord, BK Adventure offre tour eco-compatibili nel Merritt Island Wildlife Refuge, dove è possibile accarezzare il plancton glowin-the-dark in un kayak trasparente, spot manatebots e gelatine bioluminescenti su una paletta da picnic al tramonto, o scivolare attraverso il blu in un robusto gommone - una grande opzione per bambini o persone a mobilità limitata. Le guide suggeriscono di remare per una notte senza luna, portare maniche lunghe e spray</p>

riverside B&B Casa Coquina del Mar for postpaddle hot tub lounging, homemade breakfast, and a bottomless pit of cookies.	per insetti a base di piante, e prenotare il lato fiume B&B Casa Coquina del Mar per una vasca da bagno calda dopo la pagaia, colazione fatta in casa e una buca senza fondo di biscotti.
---	--

AT 9 kein Pre-Editing	ZT 9.1
<p>The bright signal can startle and distract the predator and cause confusion about the whereabouts of its target. From small copepods to the larger vampire squid, this tactic can be very useful in the deep-sea. The "green bomber" worm (<i>Swima bombiviridis</i>) and four other similar worm species from the polychaete family release a bioluminescent "bomb" from their body when in harms way. These deep sea worms live close to the sea bottom and were only discovered in 2009. Some animals such as the deep-sea squid <i>Octopoteuthis deletron</i> even detach their bioluminescent arms, which stick to and probably distract their predators.</p> <p>Quelle: The Ocean Portal Team (2018). Bioluminescence. Smithsonian Institution. https://ocean.si.edu/ocean-life/fish/bioluminescence (Stand: 08.09.2020).</p>	<p>Il segnale luminoso può spaventare e distrarre il predatore e creare confusione sul luogo in cui si trova il bersaglio. Dai piccoli copepodi ai calamari più grandi, questa tattica può essere molto utile nelle profondità marine. Il verme "bombo verde" (<i>Swima bombiviridis</i>) e altre quattro specie di vermi simili della famiglia di poligonali rilasciano una "bomba" bioluminescente dal loro corpo quando sono in pericolo. Questi vermi di acque profonde vivono vicino al fondo marino e sono stati scoperti solo nel 2009. Alcuni animali, come il calamaro <i>Octopoteuthis deletron</i>, staccano persino le loro braccia bioluminescenti, che si attaccano e probabilmente distraggono i loro predatori.</p>

AT 9 Pre-Editing-Methode 5 (K)	ZT 9.2
<p>The bright signal can startle and distract the predator and cause confusion about the whereabouts of the target of the predator. From small copepods to the larger vampire squid, this tactic can be very useful in the deep-sea. The "green bomber" worm (<i>Swima bombiviridis</i>) and four other similar worm species from the polychaete family release a bioluminescent "bomb" from their body when in harms way. These deep sea worms live close to the sea bottom and were only discovered in 2009. Some animals such as the deep-sea squid <i>Octopoteuthis deletron</i> even detach their bioluminescent arms, which stick to and probably distract their predators.</p>	<p>Il segnale luminoso può spaventare e distrarre il predatore e creare confusione sul luogo in cui si trova l'obiettivo del predatore. Dai piccoli copepodi ai calamari più grandi, questa tattica può essere molto utile nelle profondità marine. Il verme "bombo verde" (<i>Swima bombiviridis</i>) e altre quattro specie di vermi simili della famiglia di poligonali rilasciano una "bomba" bioluminescente dal loro corpo quando sono in pericolo. Questi vermi di acque profonde vivono vicino al fondo marino e sono stati scoperti solo nel 2009. Alcuni animali, come il calamaro <i>Octopoteuthis deletron</i>, staccano persino le loro braccia bioluminescenti, che si attaccano e probabilmente distraggono i loro predatori.</p>

AT 9 Pre-Editing-Methode 2	ZT 9.3
<p>The bright signal can startle and distract the predator and cause confusion about the whereabouts of its target. From small copepods to the larger vampire squid, this tactic can be very useful in the deep-sea. The "green bomber" worm (<i>Swima bombiviridis</i>) and four other similar worm species from the polychaete family release a bioluminescent "bomb" from their body when in danger. These deep sea worms live close to the sea bottom and were only discovered in 2009. Some animals such as the deep-sea squid <i>Octopoteuthis deletron</i> even detach their bioluminescent arms, which stick to and probably distract their predators.</p>	<p>Il segnale luminoso può spaventare e distrarre il predatore e creare confusione sul luogo in cui si trova il bersaglio. Dai piccoli copepodi ai calamari più grandi, questa tattica può essere molto utile nelle profondità marine. Il verme "bombo verde" (<i>Swima bombiviridis</i>) e altre quattro specie di vermi simili della famiglia di poligonali rilasciano una "bomba" bioluminescente dal loro corpo quando sono in pericolo. Questi vermi di acque profonde vivono vicino al fondo marino e sono stati scoperti solo nel 2009. Alcuni animali, come il calamaro <i>Octopoteuthis deletron</i>, staccano persino le loro braccia bioluminescenti, che si attaccano e probabilmente distraggono i loro predatori.</p>
AT 9 Pre-Editing-Methode 3	ZT 9.4
<p>The bright signal can startle and distract the predator and cause confusion about the whereabouts of its target. From small copepods to the larger vampire squid, this tactic can be very useful in the deep-sea. The "green bomber" worm (<i>Swima bombiviridis</i>) and four other similar worm species from the polychaete family release a bioluminescent "bomb" from their body when in harms way. These deep sea worms live close to the sea bottom and were only discovered in 2009. Some animals such as the deep-sea squid <i>Octopoteuthis deletron</i> even detach their bioluminescent arms, which follow and probably distract their predators.</p>	<p>Il segnale luminoso può spaventare e distrarre il predatore e creare confusione sul luogo in cui si trova il bersaglio. Dai piccoli copepodi ai calamari più grandi, questa tattica può essere molto utile nelle profondità marine. Il verme "bombo verde" (<i>Swima bombiviridis</i>) e altre quattro specie di vermi simili della famiglia di poligonali rilasciano una "bomba" bioluminescente dal loro corpo quando sono in pericolo. Questi vermi di acque profonde vivono vicino al fondo marino e sono stati scoperti solo nel 2009. Alcuni animali, come il calamaro <i>Octopoteuthis deletron</i>, staccano persino le braccia bioluminescenti, che seguono e probabilmente distraggono i loro predatori.</p>

AT 10 kein Pre-Editing	ZT 10.1
<p>Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who led travelers astray at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic "ball and chain" relationship idiom, they</p>	<p>Come le mostruose versioni acquatiche della mitologia, la saggezza - figure luminescenti che hanno portato i viaggiatori fuori strada di notte - i pesci di angler di acque profonde tentano la loro preda con luccicanti palline che penzolano dalle bacchette di fronte alla bocca. I maschi di alcuni angloers sono miniscule rispetto alle femmine. Una volta trovata un'ondina con cui accoppiarsi, in un'inversione della misogina relazione</p>

<p>attach themselves to her body, which eventually absorbs all but their testicles.</p> <p>Quelle: Pallardy, Richard (o.J.). Everything's Illuminated: 6 Bioluminescent Organisms. <i>Encyclopaedia Britannica Inc.</i> https://www.britannica.com/list/6-bioluminescent-organisms (Stand: 08.09.2020)</p>	<p>"palla e catena", si attaccano al suo corpo, che alla fine assorbe tutti tranne i testicoli.</p>
<p>AT 10 Pre-Editing-Methode 5 (K)</p>	<p>ZT 10.2</p>
<p>Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who led travelers astray at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic "ball and chain" relationship idiom, they attach themselves to her body, which eventually absorbs all but the testicles of the anglers.</p>	<p>Come le mostruose versioni acquatiche della mitologia, la saggezza - figure luminescenti che hanno portato i viaggiatori fuori strada di notte - i pesci di angler di acque profonde tentano la loro preda con luccicanti palline che penzolano dalle bacchette di fronte alla bocca. I maschi di alcuni angloers sono miniscule rispetto alle femmine. Una volta trovata un'unità adatta con cui accoppiarsi, in un'inversione della misogina relazione "palla e catena", si attaccano al suo corpo, che alla fine assorbe tutti tranne i testicoli degli anglers.</p>
<p>AT 10 Pre-Editing-Methode 2</p>	<p>ZT 10.3</p>
<p>Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who led travelers astray at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic "dead weight" relationship idiom, they attach themselves to her body, which eventually absorbs all but their testicles.</p>	<p>Come le mostruose versioni acquatiche della mitologia, la saggezza - figure luminescenti che hanno portato i viaggiatori fuori strada di notte - i pesci di angler di acque profonde tentano la loro preda con luccicanti palline che penzolano dalle bacchette di fronte alla bocca. I maschi di alcuni angloers sono miniscule rispetto alle femmine. Una volta che trovano un'undine con la quale accoppiarsi, in un'inversione dell'idioma misogino di relazione "peso morto", si attaccano al suo corpo, che alla fine assorbe tutti i testicoli tranne i loro.</p>
<p>AT 10 Pre-Editing-Methode 3</p>	<p>ZT 10.4</p>
<p>Like monstrous aquatic versions of the mythological will o' the wisp---luminescent figures who misled travelers at night---deep-sea angler fishes tempt their prey with glowing baubles dangling from rods in front of their mouths. The males of some anglers are miniscule in comparison to the females. Once they find a suitable undine with whom to couple, in an inversion of the misogynistic "ball and chain" relationship idiom, they</p>	<p>Come le mostruose versioni acquatiche della mitologia, la saggezza - figure luminescenti che hanno fuorviato i viaggiatori di notte - i pesci angler di acque profonde tentano la loro preda di luccicanti panettoni che penzolano dalle bacchette di fronte alla bocca. I maschi di alcuni angloers sono miniscule rispetto alle femmine. Una volta trovata un'ondina con cui accoppiarsi, in un'inversione della misogina relazione "palla e catena", si attaccano al suo</p>

attach themselves to her body, which eventually absorbs all but their testicles.	corpo, che alla fine assorbe tutti tranne i testicoli.
--	--

AT 11 kein Pre-Editing	ZT 11.1
<p>In the special exhibition Unseen Oceans, visitors come upon a brilliant floor-to-ceiling spiral of models of underwater species biofluorescing just as they would in life. Among the catsharks, breams, and other fishes reflecting neon greens and reds is a glowing reptile: a hawksbill sea turtle. Only three years ago, this species, Eretmochelys imbricata, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have made the cut.</p> <p>Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, their main food source, from the reef's surface.</p> <p>Quelle: American Museum of Natural History (AMNH, 2019). The Glowing Hawksbill Sea Turtle. American Museum of Natural History (AMNH), 09.01.2019. https://www.amnh.org/explore/news-blogs/news-posts/the-glowing-hawksbill-sea-turtle (Stand: 08.09.2020).</p>	<p>Nell'esposizione speciale Invisibili Oceani, i visitatori si sono imbattuti in una brillante spirale "da terra a soffitto" di modelli di biofluorescenza sottomarina, proprio come farebbero nella vita. Tra gli squali, le razze e gli altri pesci che riflettono i verdi e i rossi di neon si trova un rettile splendente: una tartaruga marina di hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, che è attualmente elencata come fortemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non avrebbe fatto il taglio.</p> <p>Le tartarughe marine delle Hawksbill giocano un ruolo essenziale nella vita delle barriere coralline, estraendo dalla superficie della barriera le spugne, la loro principale fonte alimentare.</p>
AT 11 Pre-Editing-Methode 3	ZT 11.2
<p>In the special exhibition Unseen Oceans, visitors encounter a brilliant floor-to-ceiling spiral of models of underwater species biofluorescing just as they would in life. Among the catsharks, breams, and other fishes reflecting neon greens and reds is a glowing reptile: a hawksbill sea turtle. Only three years ago, this species, Eretmochelys imbricata, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have survived.</p> <p>Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, their main food source, from the reef's surface.</p>	<p>Nella mostra speciale Oceani invisibili, i visitatori incontrano una brillante spirale "dal pavimento al soffitto" di modelli di biofluorescenza subacquea proprio come nella vita. Tra gli squali, le razze e gli altri pesci che riflettono i verdi e i rossi di neon si trova un rettile splendente: una tartaruga marina di hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, attualmente elencata come fortemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non sarebbe sopravvissuta.</p> <p>Le tartarughe marine delle Hawksbill giocano un ruolo essenziale nella vita delle barriere coralline, estraendo dalla superficie</p>

	della barriera le spugne, la loro principale fonte alimentare.
AT 11 Pre-Editing-Methode 4	ZT 11.3
<p>In the special exhibition Unseen Oceans, visitors come upon a brilliant floor-to-ceiling spiral of models of underwater species which biofluoresce just as they would in life. Among the catsharks, breams, and other fishes which reflect neon greens and reds is a glowing reptile: a hawksbill sea turtle. Only three years ago, this species, Eretmochelys imbricata, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have made the cut.</p> <p>Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, their main food source, from the reef's surface.</p>	<p>Nell'esposizione speciale "Oceani invisibili", i visitatori arrivano su una brillante spirale da pavimento a soffitto di modelli di specie sottomarine che si biofluorescono proprio come farebbero nella vita. Tra gli squali, le razze e gli altri pesci che riflettono i verdi e i rossi di neon c'è un rettile splendente: una tartaruga marina di hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, che è attualmente elencata come fortemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non avrebbe fatto il taglio.</p> <p>Le tartarughe marine delle Hawksbill giocano un ruolo essenziale nella vita delle barriere coralline, estraendo dalla superficie della barriera le spugne, la loro principale fonte alimentare.</p>
AT 11 Pre-Editing-Methode 5 (K)	ZT 11.4
<p>In the special exhibition Unseen Oceans, visitors come upon a brilliant floor-to-ceiling spiral of models of underwater species biofluorescing just as they would in life. Among the catsharks, breams, and other fishes reflecting neon greens and reds is a glowing reptile: a hawksbill sea turtle. Only three years ago, this species, Eretmochelys imbricata, which is currently listed as critically endangered by the International Union for Conservation of Nature, wouldn't have made the cut.</p> <p>Hawksbill sea turtles play an essential role in the life of coral reefs, plucking out sponges, the main food source of the sea turtles, from the reef's surface.</p>	<p>Nell'esposizione speciale Invistati Oceani, i visitatori si sono imbattuti in una brillante spirale "da terra a soffitto" di modelli di biofluorescenza sottomarina, proprio come farebbero nella vita. Tra gli squali, le razze e gli altri pesci che riflettono i verdi e i rossi di neon si trova un rettile splendente: una tartaruga marina di hawksbill. Solo tre anni fa, questa specie, Eretmochelys imbricata, che è attualmente elencata come fortemente minacciata dall'Unione Internazionale per la Conservazione della Natura, non avrebbe fatto il taglio.</p> <p>Le tartarughe marine Hawksbill giocano un ruolo essenziale nella vita delle barriere coralline, strappando dalla superficie della barriera le spugne, la principale fonte alimentare delle tartarughe marine.</p>
AT 12 kein Pre-Editing	ZT 12.1
Looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as	Guardando oltre la piatta, senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una

<p>a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule.</p> <p>Quelle: Long, Kat (2015). Life Bustles in the Long Arctic Night. <i>Hakai Magazine</i>, 04.11.2015. https://www.hakaimagazine.com/news/life-bustles-long-arctic-night/ (Stand: 08.09.2020).</p>	<p>foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti, tranne il più difficile della vita, sono fuggiti per scalate più calde, o si sono accovacciati per aspettare il ritorno del Sole. [...] Le capesante islandesi, nel frattempo, possono rilevare una quantità sufficiente di luce ambientale dal debole bagliore del sole all'orizzonte per nuotare e nutrirsi con un programma tipo estivo.</p>
<p>AT 12 Pre-Editing-Methode 6 (K)</p>	<p>ZT 12.2</p>
<p>Looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry’s accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a schedule like that in summer.</p>	<p>Guardando oltre la piatta, senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti, tranne il più difficile della vita, sono fuggiti per scalate più calde, o si sono accovacciati per aspettare il ritorno del Sole. [...] Le capsule islandesi, nel frattempo, possono rilevare una quantità di luce ambientale sufficiente dal bagliore debole del sole all'orizzonte per nuotare e nutrirsi con un programma come quello in estate.</p>
<p>AT 12 Pre-Editing-Methode 3</p>	<p>ZT 12.3</p>
<p>Looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry’s accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all but the hardiest of life either fled for warmer climes, or settled to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule.</p>	<p>Guardando oltre la piatta, senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Secondo i racconti di Parry, e di altri viaggiatori, gli scienziati hanno costruito una foto della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo senza fine - quando tutti tranne il più difficile della vita sono fuggiti per climi più caldi, o si sono stabiliti per aspettare il ritorno del Sole. [...] Le capesante islandesi, nel frattempo, possono rilevare una quantità sufficiente di luce ambientale dal debole bagliore del sole all'orizzonte per nuotare e nutrirsi con un programma tipo estivo.</p>

AT 12 Pre-Editing-Methode 2	ZT 12.4
<p>Looking out across the flat, treeless expanse of ice and fog, the barren landscape must have seemed wholly inhospitable. [...] From Parry's accounts, and those of other travelers, scientists built a picture of the polar night as a period of Arctic dormancy, an extended slumber in the endless twilight—when all living beings left the severe cold weather and either fled for warmer climes, or hunkered down to wait for the Sun to return. [...] Iceland scallops, meanwhile, can detect enough ambient light from the weak glow of the sun over the horizon to swim and feed on a summer-like schedule.</p>	<p>Guardando oltre la piatta, senza tregua di ghiaccio e nebbia, il paesaggio sterile deve essere sembrato del tutto inospitale. [...] Dai resoconti di Parry, e da quelli di altri viaggiatori, gli scienziati hanno costruito un'immagine della notte polare come un periodo di dormienza Artica, un sonno prolungato nel crepuscolo infinito - quando tutti gli esseri viventi hanno lasciato il freddo e sono fuggiti per scalare, o sono fuggiti per aspettare il ritorno del Sole. [...] Le capesante islandesi, nel frattempo, possono rilevare una quantità sufficiente di luce ambientale dal debole bagliore del sole all'orizzonte per nuotare e nutrirsi con un programma tipo estivo.</p>

Abstract

Die Zielsetzung dieser Masterarbeit ist der Einsatz von Pre-Editing als Vorschlag zur Verbesserung der Resultate neuronaler maschineller Übersetzung für die Sprachkombination Englisch-Italienisch. Die Forschungsfrage dieser Arbeit lautet: Welche Pre-Editing-Methoden und Kombinationen aus Pre-Editing-Methoden verbessern das Resultat der maschinellen Übersetzung aus dem Englischen ins Italienische?

Um die Forschungsfrage zu evaluieren, wird eine Evaluierungsmethode angewandt, die noch nie im Bereich der Translationswissenschaft verwendet worden ist: das Best-Worst-Scaling. Für den empirischen Teil werden ausgewählte Textpassagen von Blogartikeln des Bereichs Biologie durch Pre-Editing-Methoden und Kombinationen aus Pre-Editing-Methoden präeditiert (maximal drei pro Ausgangstext). In einem folgenden Schritt werden die Textpassagen durch eine neuronale maschinelle Übersetzungssoftware übersetzt und die jeweiligen Übersetzungsvarianten einer Gruppe von Proband*innen zur Evaluierung vorgelegt. Die Proband*innen werden im Rahmen des Best-Worst-Scalings gefragt, ihre Präferenzen über die „beste“ und die „schlechteste“ Übersetzung aus einem Set anzugeben.

Die Ergebnisse zeigen, dass nur eine Pre-Editing-Methode aus sechs die am meisten positiven Bewertungen erzielt und deswegen potenziell zur Verbesserung der maschinellen Qualität führen kann. Diese Studie zeigt, dass die Tendenzen in Richtung Pre-Editing steigen könnten und sich diese Methode als sinnvoll für die Verbesserung maschineller Übersetzung erweisen könnte.