



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Aus dem Englischen übersetzt von Maschine und Mensch. Der Einfluss von Fehlerkategorien auf die Postedition einer literarischen maschinellen Übersetzung und ihre Qualität“

verfasst von / submitted by

Katharina Schuhmann, BA

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

Wien, 2021 / Vienna 2021

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

UA 070 331 342

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

Masterstudium Translation UG2002
Deutsch Englisch

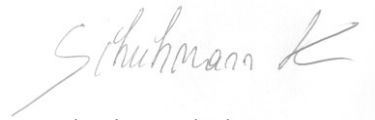
Betreut von / Supervisor:

Ass.-Prof. Mag. Dr. Waltraud Kolb

Selbstständigkeitserklärung

Ich versichere, die vorliegende Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Alle von mir für direkte und indirekte Zitate benutzten Quellen sind nach den Regeln des wissenschaftlichen Zitierens angegeben. Mir ist bekannt, dass beim Verstoß gegen diese Regeln eine positive Beurteilung der Arbeit nicht möglich ist. Ich habe die Arbeit bzw. Teile davon weder im In- noch im Ausland zur Begutachtung als Prüfungsarbeit vorgelegt.

Wien, im November 2021



Katharina Schuhmann, BA

Danksagung

*'First of all, Harry,
I want to thank you,'
said Dumbledore,
eyes twinkling again.*

(Rowling 2010/1998:356;
meine Hervorhebung)

Genau so, wie sich Dumbledore in einer der letzten Szenen des zweiten *Harry-Potter*-Bandes bei Harry für seine Loyalität bedankt, bevor er mit ihm über die ernste Seite der Geschehnisse spricht, möchte ich zu Beginn dieser Masterarbeit einigen Menschen meinen zutiefst empfundenen Dank aussprechen. Ohne sie wäre ich nicht so weit gekommen.

Ganz herzlich möchte ich mich bei meiner Betreuerin Frau Waltraud Kolb bedanken. Als ich sie in einer der ersten Vorlesungen meines Bachelorstudiums als Professorin hatte, hätte ich nie geglaubt, dass ich nicht nur meine Bachelorarbeit, sondern auch meine Masterarbeit bei ihr schreiben würde. Doch so war es, und sie hat mich bei beiden Arbeiten außerordentlich gut auf meinem Weg begleitet, mir hervorragendes Feedback gegeben und mich in allen Belangen unterstützt. Vielen, vielen Dank dafür!

Als nächstes gilt mein Dank den zwei anderen Professor_innen, die mir bei der Durchführung meiner Masterarbeitsstudie unter die Arme gegriffen haben, allen voran Frau Claudia Zecher, die mir nicht nur mögliche Posteditor_innen vorschlug und meinen Aufruf zur Teilnahme an meiner Studie an ihre Studierenden verschickte, sondern mich diese Studie auch während ihres Unterrichts abhalten ließ, als sich nicht genügend Teilnehmer_innen meldeten. Vielen Dank! Auch bei dem anderen Professor, der mir mögliche Posteditor_innen vorschlug und meinen Aufruf verschickte, hier allerdings anonym bleiben wollte, möchte ich mich für die Unterstützung bedanken.

Danken möchte ich auch Herrn Arda Tezcan von der Universität Ghent, Frau Ana Guerberof-Arenas von der Universität Surrey und Herrn Tim Van de Cruys von der KU Leuven: Sie erwiesen mir, einer fremden Studentin aus einem anderen Land, große Freundlichkeit und beantworteten meine Fragen bereitwillig und rasch.

Großer Dank gebührt auch meinen zwei fantastischen Posteditorinnen, die sich, obwohl sie selber mit ihrer eigenen Masterarbeit beschäftigt waren, bereit erklärten, jeweils eine Postedition für mich anzufertigen, und dabei wahrlich ihr Bestes gaben. Außerdem bedanke ich mich bei den zwei Teilnehmer_innen an meiner Studie, die sich auf meinen Aufruf hin meldeten, sowie bei den anderen Studierenden, die daran im Zuge von Frau Zechers Unterricht teilnahmen.

Ebenfalls bedanken möchte ich mich bei zwei meiner Studienkolleginnen: Denisa Drabantová besuchte mit mir das Masterkolloquium bei Frau Kolb und stand mir bei der Entwicklung meines Konzepts für die Masterarbeit stets mit hilfreichen Anmerkungen zur Seite. Vielen Dank dafür.

Sonja Hengsberger begleitete mich vom zweiten Semester an durch mein Bachelorstudium und durch das gesamte Masterstudium. Wir waren sogar gemeinsam ein Semester lang in Kanada. Ich bedanke mich bei ihr ganz herzlich für die vielen lustigen Stunden, die Motivation, die wir uns gegenseitig gaben, und die vielen Arbeiten, die sie für mich Korrektur gelesen hat, nicht zuletzt auch diese Masterarbeit. Ich hoffe, ich war für sie eine ebenso große Unterstützung wie sie für mich.

Zu guter Letzt gilt mein allergrößter Dank meiner Familie, die mir das Bachelor- und Masterstudium erst ermöglichte, die mir mit Verständnis begegnete, wenn ich aufgrund meines straffen Zeitplans manchmal nur an die Masterarbeit und an nichts anderes denken konnte, und die mich stets in all meinen Plänen unterstützte. Vielen Dank für alles.

Inhalt

Abkürzungsverzeichnis	7
Abbildungs- und Tabellenverzeichnis.....	8
Einleitung	9
1 Überblick über die (neuronale) maschinelle Übersetzung	12
2 Überblick über das Postediting.....	21
3 Forschungsüberblick: literarische maschinelle Übersetzung	30
4 Methode.....	44
4.1 Untersuchungsmaterial	44
4.2 Entwicklung der Fehlerkategorien.....	48
4.3 Posteditorinnen und Durchführung des Posteditings.....	58
4.4 Durchführung der Ratings	62
4.5 Datenauswertung	66
5 Datenanalyse.....	69
5.1 Fehlerkategorien	69
5.2 Ergebnisse der Datenauswertung.....	86
5.2.1 Posteditionen.....	86
5.2.2 Ratings	106
6 Diskussion	132
7 Zusammenfassung.....	150
Bibliographie.....	153
Abbildungen	153
Primärliteratur.....	153
Sekundärliteratur	154
Tertiärliteratur.....	174
Anhang	176
Anhang 1: Maschinelle Übersetzungen	176
Anhang 2: Posteditionen.....	181
Anhang 3: Anmerkungen der Rater_innen	187
Anhang 4: Unveröffentlichte Quellen	192
Anhang 5: Abstracts	193

Abkürzungsverzeichnis

Abkürzungen aus dem Forschungsbereich

BLEU	„bilingual evaluation understudy“ (Papineni et al. 2002:311)
CALT2021@Swansea	Computer-Assisted Literary Translation Conference in Swansea
CAT	computergestützte Übersetzung (<i>computer-aided translation</i>)
CNN	<i>convolutional neural network</i>
DQF	Dynamic Quality Framework
LMT	literarische maschinelle Übersetzung (<i>literary machine translation</i>)
MT	maschinelle Übersetzung (<i>machine translation</i>)
MQM	Multidimensional Quality Metric
NMT	neuronale maschinelle Übersetzung (<i>neural machine translation</i>)
PBSMT	phrasenbasierte statistische maschinelle Übersetzung (<i>phrase-based statistical machine translation</i>)
RBMT	regelbasierte maschinelle Übersetzung (<i>rule-based machine translation</i>)
RNN	rekurrentes neuronales Netz (<i>recurrent neural network</i>)
SCATE	„Smart Computer-aided Translation Environment“ (Tezcan et al. 2018:221)
SMT	statistische maschinelle Übersetzung (<i>statistical machine translation</i>)
TAUS	Translation Automation User Society
TM	Translation Memory
TTR	Type-Token-Verhältnis (<i>type-token ratio</i>)

Abkürzungen der vorliegenden Arbeit

FF	Forschungsfrage
HPAT	Rowling (2010/1997:233-239) = Ausgangstext (FF1)
HPHÜ	Rowling (2013/1998:235-241) = Humanübersetzung des HPAT (FF1)
HPMÜ	DeepL-Übersetzung des HPAT (FF1)
IUT	In-Universe-Terminologie/Terminus
L-NMT	literarische neuronale maschinelle Übersetzung
L-SMT	literarische statistische maschinelle Übersetzung
NZ	Nicht zuordenbar
PV	Persönliche Vorliebe
R	Rater_in
RRAT	Brown (2014a:[22-27]) = Ausgangstext (FF2, FF3)
RRHÜ	Brown (2014b:[24-29]) = Humanübersetzung des RRAT (FF2, FF3)
RRMÜ	DeepL-Übersetzung des RRAT (FF2, FF3)
RRPEm	Postedition mit Fehlerkategorien der RRMÜ (FF2, FF3)
RRPEo	Postedition ohne Fehlerkategorien der RRMÜ (FF2, FF3)

Abbildungs- und Tabellenverzeichnis

Abbildungen

Abb. 1: Beispielhafte Repräsentationen der Wörter „apple“, „pear“ und „car“, links als Liste von Zahlen, rechts graphisch (Van de Cruys 2021)	15
Abb. 2: Beispielhafte Darstellung der Funktionsweise des Beam-Search-Algorithmus (Koehn 2020:146)	19
Abb. 3: DQF-MQM-Fehlertypologie (Lommel, Görög et al. 2015:13; farblich adaptiert durch die Verfasserin)	53
Abb. 4: An L-NMT angepasste SCATE-Fehlertypologie (Tezcan et al. 2019:42)	56
Abb. 5: Paarweises Ranking RRPEo – RRPEm	110
Abb. 6: Paarweises Ranking RRPEo – RRHÜ	111
Abb. 7: Paarweises Ranking RRPEm – RRHÜ	111

Tabellen

Tabelle 1: Vergleichbarkeit der Ausgangstexte	46
Tabelle 2: Problematische Bereiche der L-NMT aus der Forschungsliteratur	48
Tabelle 3: Fehlerkategorien und Anzahl Fehler/Problemstellen in der HPMÜ	70
Tabelle 4: Anzahl Korrekturen pro Fehlerkategorie in der RRPEo und der RRPEm	87
Tabelle 5: Interrater_innen-Reliabilität	107
Tabelle 6: FF1 und Hypothese	132
Tabelle 7: Entwickelte Fehlerkategorien in der Forschungsliteratur	133
Tabelle 8: FF2 und FF3 und Hypothesen	139

Einleitung

Das Thema der vorliegenden Masterarbeit, nämlich der Einfluss von Fehlerkategorien auf die Postedition einer literarischen maschinellen Übersetzung und ihre Qualität, wurde von folgenden zwei Aussagen inspiriert:

NMT [neural machine translation] output can be deceptively fluent. Sometimes perfect target-language sentences are output, and less thorough translators and proofreaders may be seduced into accepting such translations, despite the fact that they may not be related to the source sentence at hand at all. (Way 2020:320)

There are many MT error classifications, the aim of which is [...] also to raise awareness among post-editors of the types of errors they will need to correct. If they know the types of errors frequently found in that output before performing this task, it is easier for post-editors to spot and correct them. (Guerberof-Arenas 2020:337)

Fehler in Übersetzungen von neuronalen maschinellen Übersetzungssystemen (NMT-Systemen) sind also schwer zu erkennen. Dies ist die Hauptaussage des ersten Zitats, während im zweiten behauptet wird, dass Fehlerkategorien bzw. -typologien die Arbeit von Posteditor_innen erleichtern. Werden diese zwei Aussagen kombiniert, stellt sich die Frage, ob Fehlerkategorien nicht nur eine Hilfestellung für Posteditor_innen sind, sondern ihre Anwendung während des Posteditings einer NMT-Übersetzung auch tatsächlich zu einer Postedition von höherer Qualität führt¹. Diese Frage, gekoppelt mit dem Interesse der Verfasserin der vorliegenden Arbeit an der literarischen maschinellen Übersetzung (*literary machine translation*, LMT), die das Thema ihrer Bachelorarbeit (Schuhmann 2020) darstellte, war der Ausgangspunkt für die im Rahmen dieser Masterarbeit durchgeführte Untersuchung. In dieser Untersuchung sollte herausgefunden werden, ob sich die Anwendung von Fehlerkategorien, die anhand einer von einem NMT-System erstellten literarischen Übersetzung entwickelt wurden, positiv auf die Qualität der Postedition einer NMT-Übersetzung eines ähnlichen literarischen Textes auswirkt. Die genauen Forschungsfragen (FF) für diese Untersuchung lauten wie folgt:

- FF1: Welche Fehlerkategorien lassen sich unter Berücksichtigung der Forschungsliteratur zur literarischen NMT anhand der von DeepL² erstellten Übersetzung eines Ausschnitts aus *Harry Potter and the Philosopher's Stone* ins Deutsche etablieren?
- FF2: Ist eine von DeepL erstellte Übersetzung eines Ausschnitts aus *Red Rising*, die unter Zuhilfenahme dieser Fehlerkategorien posteditiert wird, von höherer Qualität

¹ In Anlehnung an do Carmo & Moorkens (2021:36) bezeichnet „Postediting“ in der vorliegenden Arbeit den Prozess, während „Postedition“ die aus dem Postediting resultierende finale Übersetzung bezeichnet.

² <https://www.deepl.com/translator> (Stand: 14.08.2021).

als eine ohne diese Fehlerkategorien posteditierte DeepL-Übersetzung desselben Ausschnitts?

- FF3: In welchen Aspekten unterscheiden sich die beiden Posteditionen und welche Fehlerkategorien spielen hierbei eine besondere Rolle?

Die wissenschaftliche Relevanz dieser Forschungsfragen ergab sich zum Zeitpunkt der Verschriftlichung dieser Masterarbeit (Frühling bis Herbst 2021) aus drei Umständen: Erstens gab es zum Postediting von literarischen maschinellen Übersetzungen nur zwei Untersuchungen, bei denen DeepL zum Einsatz kam³, nämlich Kenny & Winters (2020a) und Kolb (2021). Zweitens waren diese beiden auch die einzigen Untersuchungen in diesem Forschungsgebiet (ob mit DeepL oder anderen maschinellen Übersetzungssystemen), die sich mit der Übersetzungsrichtung Englisch-Deutsch auseinandersetzten. Drittens existierten nach bestem Wissen der Verfasserin der vorliegenden Arbeit keine Studien zum Einfluss von Fehlerkategorien auf mithilfe solcher Kategorien hergestellte Posteditionen von literarischen maschinellen Übersetzungen und ihre Qualität.

Wie an der Formulierung der Forschungsfragen ersichtlich wird, wurde für ihre Beantwortung unterschiedliches Untersuchungsmaterial benötigt. Mithilfe eines Teils dieses Untersuchungsmaterials wurden Fehlerkategorien entwickelt (die die Antwort auf FF1 darstellen), die im zweiten Schritt von einer Studierenden beim Postediting der maschinellen Übersetzung des *Red-Rising*-Ausgangstextes verwendet wurden, während eine zweite Posteditorin dieselbe Übersetzung ohne diese Kategorien posteditierte. Die zwei resultierenden Posteditionen sowie die publizierte Humanübersetzung desselben Ausgangstextes wurden anschließend von anderen Studierenden einer Qualitätsbewertung durch mehrere Ratings (auch in Bezug auf die entwickelten Fehlerkategorien) unterzogen. Die so gesammelten Daten wurden dann ausgewertet, um FF2 und FF3 zu beantworten.

Diese Vorgehensweise wird in Kapitel 4 der vorliegenden Arbeit genauer beschrieben. In den Kapiteln 1-3 wird auf die (neuronale) maschinelle Übersetzung, das Postediting sowie die Forschung zur LMT eingegangen. In Kapitel 5 werden die Fehlerkategorien sowie die ausgewerteten Daten präsentiert, während in Kapitel 6 die Antworten auf die Forschungsfragen präzise ausformuliert und in Beziehung zur Forschungsliteratur gesetzt werden, auch im Hinblick auf weiterführende Studien. Danach folgt in Kapitel 7 eine knappe Zusammenfassung der vorliegenden Arbeit.

³ LMT-Studien mit DeepL wurden auch von Webster et al. (2020a) sowie von Schmidt (2021) durchgeführt, allerdings war dabei kein Postediting involviert.

Im Rahmen der Beantwortung der Forschungsfragen wird ebenfalls darauf eingegangen, ob sich die folgenden Hypothesen bewahrheitet haben, die sich auf die persönliche Erfahrung der Verfasserin der vorliegenden Arbeit mit der LMT sowie auf die Forschungsliteratur zum Thema stützen. In Bezug auf FF1 wurde erwartet, dass anhand der von DeepL erstellten Übersetzung des *Harry-Potter*-Ausgangstextes Fehlerkategorien etabliert werden können, die sich auf die folgenden Aspekte beziehen: Interpunktion (v. a. Anführungszeichen); Wörter und Mehrwortausdrücke, die für den betreffenden Roman erfunden wurden bzw. dort auf spezifische Art und Weise verwendet werden; Auslassungen; Übertragung der Bedeutung (falsch oder nicht vollständig); Auflösung von Anaphern über Satzgrenzen hinweg.

Unter Bezugnahme auf die oben angeführte Bemerkung von Guerberof-Arenas (2020:337) wurde bzgl. FF2 erwartet, dass die mit Fehlerkategorien erstellte Postedition insgesamt besser bewertet würde als die Postedition, die ohne die Fehlerkategorien angefertigt wurde, sowie dass erstere im Hinblick auf die Qualität auch häufiger mit der Humanübersetzung gleichgesetzt würde als letztere. Weiters wurde angenommen, dass sich diese bessere Bewertung der mit Fehlerkategorien erstellten Postedition auch bei einigen Ratings in Bezug auf die entwickelten Fehlerkategorien feststellen lassen würde. Wie viele und welche der Fehlerkategorien das allerdings genau sein würden, d. h., bzgl. welcher Fehlerkategorien sich die Posteditionen unterscheiden (FF3) und eine unterschiedliche Bewertung erhalten würden, ließ sich schwer vorhersagen, da die entsprechenden Fehlerkategorien zum Zeitpunkt der Aufstellung dieser Hypothesen erst entwickelt werden mussten. Unter der Annahme, dass sich die Hypothese für FF1 bewahrheiten würde, wurde jedoch erwartet, dass sich die mit Fehlerkategorien angefertigte Postedition in Bezug auf die Interpunktion, die Übersetzung romanspezifischer Wörter und Mehrwortausdrücke sowie die Übertragung der Bedeutung von der ohne Fehlerkategorien erstellten Postedition unterscheiden und in diesen Kategorien auch besser bewertet würde als letztere.

1 Überblick über die (neuronale) maschinelle Übersetzung

Wie der Name dieses Kapitels schon andeutet, wird hier ein kurzer Überblick über die maschinelle Übersetzung (*machine translation*, MT) gegeben: Nach einer grundlegenden Definition des Konzepts werden die verschiedenen Ansätze, mit denen MT realisiert werden kann, aufgezählt und kurz erklärt. Im Anschluss daran wird auf einen Ansatz, den der neuronalen MT, genauer eingegangen, da er in der Forschung seit seinem Aufkommen in den Jahren 2013/14 (vgl. Koehn 2017:6) und in der Industrie seit 2016/17 (vgl. Schmalz 2019:198f.) sehr populär geworden ist. Darüber hinaus liegt dieser Ansatz dem DeepL-Übersetzer zugrunde (vgl. DeepL 2018), der für die im Rahmen der vorliegenden Masterarbeit durchgeführte Studie verwendet wurde.

Die MT wird als die „translation from one natural language into another by means of a computerized system“ (Dorr et al. 1999:1) definiert. Dabei ist zu beachten, dass das System die Übersetzung (Output) anhand des Ausgangstextes (Input) ohne jegliches menschliches Eingreifen anfertigt und damit die MT von der computergestützten Übersetzung (*computer-aided translation*, CAT) zu unterscheiden ist (vgl. Werthmann & Witt 2014:82).

Ansätze für die Realisierung der MT gibt es mittlerweile sehr viele (vgl. Werthmann & Witt 2014:86). Neben hybriden Ansätzen, die unterschiedliche Strategien kombinieren, sowie wissensbasierten Ansätzen, bei denen die MT-Systeme nicht nur sprachliches, sondern auch allgemeines „Wissen“ zur Übersetzung nutzen sollen (vgl. Werthmann & Witt 2014:99), können die meisten anderen MT-Ansätze in zwei Gruppen eingeteilt werden, nämlich regelbasierte und korpusbasierte Ansätze (vgl. Tripathi & Sarkhel 2010; Costa-jussà 2012:414).

Die **regelbasierten Ansätze** (*rule-based MT*, RBMT) stützen sich auf Wörterbücher, Grammatiken und Übersetzungsregeln, die manuell erstellt werden müssen, was einen enormen Aufwand bedeutet und den hauptsächlichen Nachteil dieser Ansätze darstellt (vgl. Costa-jussà 2012:414; Werthmann & Witt 2014:86f., 91). Die RBMT-Ansätze können noch weiter unterteilt werden, für genauere Beschreibungen dazu wird allerdings auf Stein (2009:8f.) bzw. Werthmann & Witt (2014:87-91) verwiesen, um den Rahmen der vorliegenden Arbeit nicht zu sprengen und weil die regelbasierten (und anderen) von den korpusbasierten Ansätzen seit 1989 quasi verdrängt worden sind (vgl. Tripathi & Sarkhel 2010:390).

Diese Verdrängung geschah aufgrund des oben erwähnten Nachteils der RBMT, dem die **korpusbasierten Ansätze** entgegenwirken, indem sie sich nicht auf manuell zusammengestellte Regeln und dgl. stützen, sondern auf große Korpora (vgl. Werthmann & Witt

2014:91f.). Die korpusbasierten Ansätze können ebenfalls unterteilt werden, nämlich in kontextbasierte, beispielbasierte und statistische Ansätze (vgl. Tripathi & Sarkhel 2010:390f.).

Kontextbasierte Ansätze (nicht zu verwechseln mit „context-aware neural machine translation“ (Voita et al. 2018:1264; siehe unten)) arbeiten mit einsprachigen Korpora der Zielsprache (und evtl. auch der Ausgangssprache) sowie zweisprachigen Vollformenwörterbüchern, um Übersetzungen anzufertigen (vgl. Carbonell et al. 2006:19; Tripathi & Sarkhel 2010:391). Dabei wird der Ausgangstext in sogenannte N-gramme⁴ unterteilt, für die mithilfe des Vollformenwörterbuchs Übersetzungskandidaten generiert werden. Können diese Kandidaten im Kontext (d. h. den jeweils links und rechts angrenzenden N-grammen) bestätigt werden, werden sie als finale Übersetzung beibehalten (vgl. Carbonell et al. 2006:20f.).

Bei den **beispielbasierten Ansätzen** geschieht die Übersetzung mithilfe von parallelen Korpora, d. h. Ausgangstexten und ihren Übersetzungen, die (automatisch oder manuell) aligniert wurden (vgl. Werthmann & Witt 2014:91f.). Diese Parallelkorpora sind allerdings nur ein erster Schritt, denn aus ihnen werden Beispiele extrahiert und gespeichert, aus denen dann die Übersetzung, ähnlich wie in einem Translation-Memory-System, zusammengestellt wird (vgl. Kit et al. 2002:63f.).

Die **statistischen Ansätze** (*statistical MT*, SMT) arbeiten dahingegen direkt mit Parallelkorpora, aus denen die Systeme während des Trainings die Wahrscheinlichkeit berechnen, mit der ein Wort, eine Phrase oder ein Satz mit einem bzw. einer anderen übersetzt wird. In der Trainingsphase erstellt ein SMT-System also selbst und ohne jedwede linguistische Informationen ein Übersetzungsmodell (vgl. Werthmann & Witt 2014:92f.). Es gibt drei Unterarten dieser Art der SMT, je nachdem, wie groß die Segmente sind, für die das System die Wahrscheinlichkeiten berechnet: wortbasierte, syntaxbasierte und phrasenbasierte SMT (*phrase-based SMT*, PBSMT⁵) (vgl. Tripathi & Sarkhel 2010:391; Werthmann & Witt 2014:93). Letztere war bis ca. 2017/18 der Standard-MT-Ansatz und galt als „state-of-the-art“ in der MT-Forschung⁶ (vgl. Forcada 2017:292; Toral & Way 2018:264; van Genabith 2020:73), was auch der Grund dafür ist, warum im vorliegenden Kapitel nur auf PBSMT

⁴ „N-gramme sind Sequenzen von *N* Tokens, wobei ein Token aus einem Wort, einer Zahl oder Interpunktion besteht.“ (Brants 2007:239; Hervorhebung im Original)

⁵ Andere in der Forschungsliteratur verwendete Abkürzungen sind „PB-SMT“ (etwa in van Genabith 2020) und „PBMT“ (etwa in Schmalz 2019).

⁶ Tatsächlich wird die PBSMT oft als „SMT“ bezeichnet (vgl. Bessenyei 2017). Dies ist auch in vielen im Forschungsüberblick (Kapitel 3) zitierten Werken der Fall, weswegen im Rest der vorliegenden Arbeit, außer explizit anders angegeben, „PBSMT“ und „SMT“ als Synonyme zu verstehen sind.

genauer eingegangen wird. Für Erklärungen zur wortbasierten SMT wird auf Koehn (2010:81-118) bzw. Werthmann & Witt (2014:93f.) und zur syntaxbasierten auf Williams et al. (2016) verwiesen.

Wie der Name schon sagt, wird bei **PBSMT** nicht von Wörtern oder ganzen Sätzen ausgegangen, sondern von Phrasen. Diese Phrasen sind keine sprachlich motivierten Phrasen (wie Redewendungen oder dgl.), sondern einfache, für einen Menschen evtl. unübliche Wortsequenzen (z. B. „fun with the“) (vgl. Koehn 2010:127f.). Der Übersetzungsprozess funktioniert, einfach gesagt, wie folgt: Der Ausgangstext wird in Phrasen segmentiert, und für diese werden die wahrscheinlichsten Übersetzungen ausgewählt, die dann, falls nötig, noch in die für die Zielsprache richtige Reihenfolge gebracht werden (vgl. Koehn 2010:128).

Für diesen Übersetzungsprozess braucht es dementsprechend drei Modelle: das oben erwähnte Übersetzungsmodell, das Sprachmodell (das aus einsprachigen Korpora erzeugt wird (vgl. Werthmann & Witt 2014:92)) und das Umordnungsmodell (vgl. Koehn 2010:136). Für Erklärungen, wie genau diese Modelle funktionieren und wie sie verbessert und erweitert werden können, wird auf Koehn (2010:127-216) und van Genabith (2020:69-74) verwiesen.

Obwohl PBSMT-Systeme teilweise sehr gute Ergebnisse liefern, haben sie doch einige Nachteile: Sie funktionieren im Allgemeinen besser für verwandte Sprachen, die wenig morphologische Variation aufweisen und eine ähnliche Syntax haben. Außerdem treffen PBSMT-Systeme nur lokale Entscheidungen (vgl. van Genabith 2020:73), sodass sogenannte *long-distance dependencies*, also getrennte, aber grammatikalisch oder inhaltlich zusammenhängende Satzteile (vgl. Schmalz 2019:199), nur schwer richtig aufgelöst werden können, was zu zusammengestückelt anmutenden Outputs führen kann. Darüber hinaus sind die verwendeten Modelle sehr unterschiedlich, was aus der Perspektive des maschinellen Lernens nicht optimal ist (vgl. van Genabith 2020:73f.). Schlussendlich wird bei der PBSMT die Ähnlichkeit zwischen Wörtern und Phrasen nicht berücksichtigt, wie van Genabith (2020:74) anschaulich erklärt:

[PBSMT] will treat the two words ‘cat’ and ‘cats’ as completely different and unrelated, estimate their language, reordering and translation model probabilities completely independently, as it would for ‘zebra’ and ‘algebra’. In other words, for PB-SMT ‘cat’ and ‘cats’ are as different as ‘zebra’ and ‘algebra’.

Einigen dieser Nachteile von PBSMT-Systemen wirkt die seit 2016 durch Google Translate⁷ (vgl. Schmalz 2019:198) öffentlich verfügbare **NMT** entgegen, die auch zu den statistischen und damit den korpusbasierten Ansätzen gehört (vgl. van Genabith 2020:59), allerdings ganz

⁷ <https://translate.google.com/> (Stand: 07.08.2021).

anders aufgebaut ist (weswegen sie oben nicht genannt wurde). Dementsprechend hat die NMT, obwohl sie ebenfalls auf sehr großen Parallelkorpora beruht (vgl. Forcada 2017:292), eine andere Funktionsweise, die im Folgenden erläutert werden soll.

Zunächst muss beachtet werden, dass die NMT, wie ihr Name schon andeutet, mithilfe von **Netzwerken von künstlichen Neuronen** (*artificial neural networks*) arbeitet, die biologischen Neuronen ähneln (vgl. van Genabith 2020:74f.). Diese Netzwerke, die aus dem Bereich des Deep Learning stammen, ermöglichen es dem System, sich selbst, also ohne explizite Programmierung, Muster und Modelle anzueignen. Das einzige, was für diese Aneignung (das Training) benötigt wird, sind sehr große Mengen an Daten (vgl. Gupta 2018:23), im Fall von NMT-Systemen also, wie oben schon erwähnt, Parallelkorpora (und teilweise auch einsprachige Korpora (vgl. Stahlberg 2020:30f.)). Bei der Firma DeepL sind das die Ausgangstexte und Übersetzungen, die für die Übersetzungssuchmaschine Linguee⁸ gesammelt wurden (vgl. Merkert 2017).

Während des Trainings können NMT-Systeme an bestimmte Fachgebiete o. Ä. angepasst werden⁹: In einem ersten Schritt wird das System mit allg. Texten trainiert, in einem zweiten z. B. mit Texten aus dem Fachgebiet der Informationstechnologie (vgl. Koehn 2017:74). So können auch an Literatur angepasste MT-Systeme erstellt werden, wie das bei einigen in Kapitel 3 zitierten Studien der Fall war.

Die oben erwähnten Muster und Modelle, die sich das NMT-System während des Trainings aneignet, sind genauer gesagt verteilte **Repräsentationen** (*distributed representations* oder *embeddings*) von Wörtern (oder kleineren Einheiten) in ihrem Kontext. Diese Repräsentationen, auch „Vektoren“ genannt, sind im Grunde genommen Listen von Zahlen (vgl. Forcada 2017:293f.), wie in Abb. 1¹⁰ beispielhaft dargestellt.

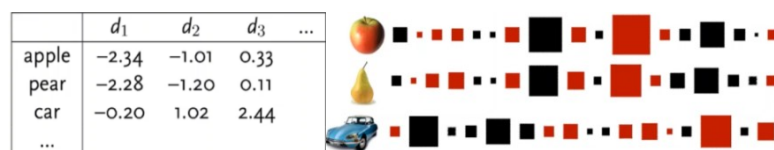


Abb. 1: Beispielhafte Repräsentationen der Wörter „apple“, „pear“ und „car“, links als Liste von Zahlen, rechts graphisch (Van de Cruys 2021)

⁸ <https://www.linguee.de/> (Stand: 07.08.2021)

⁹ Anpassung an Fachgebiete und dgl. ist selbstverständlich auch für PBSMT möglich, funktioniert aber auf andere Art und Weise (vgl. Koehn 2017:74).

¹⁰ Abdruck mit Einverständnis des Urhebers

Wie an Abb. 1¹¹ ebenfalls ersichtlich und von Forcada (2017:295) anschaulich mithilfe einer Analogie mit geographischen Koordinaten erklärt wird, haben Wörter, die ähnliche Konzepte bezeichnen (wie in der Abbildung „apple“ und „pear“ oder aus dem vorigen Beispiel von van Genabith (2020:74) „cat“ und „cats“), auch ähnliche Vektoren, während sich die Vektoren von Wörtern, deren Konzepte sich nicht ähneln („apple“ und „car“, „zebra“ und „algebra“), auch tatsächlich stark voneinander unterscheiden. NMT-Systeme sind sich, im Unterschied zu PBSMT-Systemen, folglich der Ähnlichkeit zwischen manchen Konzepten sozusagen „bewusst“.

Mit diesen Vektoren, die für die Ausgangssprache und die Zielsprache während des Trainings erstellt werden, kann nun neuer Input übersetzt werden. Dies erfolgt in den meisten Fällen in zwei Schritten: der Encodierung und der Decodierung (vgl. Forcada 2017:297). Wie die Übersetzung genau funktioniert, wird anhand des Beispielsatzes „The house is red.“ demonstriert, der ins Deutsche übersetzt werden soll.

Im ersten Schritt, der **Encodierung**, wird der entsprechende Satz in einen Vektor encodiert. Dazu wird ein schon existierender leerer Vektor $E(„“)$ (für Encoder bzw. Encodierung) mit dem Vektor des ersten Wortes des Satzes, also $e(„The“)$ (für *embedding*) kombiniert, sodass der Vektor $E(„The“)$ entsteht. Danach wird $E(„The“)$ mit $e(„house“)$ kombiniert, und es entsteht $E(„The house“)$. Dies wird fortgeführt, bis eine komplette Repräsentation des Satzes vorhanden ist: $E(„The house is red.“)$. Diese Encodierung kann zusätzlich auch noch rückwärts, d. h. von rechts nach links, erfolgen (vgl. Forcada 2017:297f.), was die Qualität der anhand dieses Vektors generierten Übersetzung verbessert (vgl. van Genabith 2020:90).

Die **Decodierung** dieses Vektors E ist der zweite Schritt der Übersetzung. Dazu erstellt der Decoder anhand von E einen ersten Decodierungsvektor D , der E enthält sowie eine „leere Stelle“ für den Zielsatz: $D(„The house is red.“, „“)$. Außerdem erstellt der Decoder einen zweiten Vektor, nämlich einen Wahrscheinlichkeitsvektor p (für *probability*), der die Wahrscheinlichkeiten für alle möglichen Wörter („x“) an der ersten Stelle des Zielsatzes angibt, also $p(x \mid „The house is red.“, „“)$. Für x wird dann das wahrscheinlichste Wort ausgewählt (vgl. Forcada 2017:298), was durch einen Vergleich des Vektors D mit den im Training gelernten zielsprachlichen Wortvektoren geschieht. Jener zielsprachliche Vektor, der dem Vektor D am ähnlichsten ist, wird ausgewählt (vgl. van Genabith 2020:88), sodass der neue

¹¹ Die drei Punkte in der Abbildung indizieren, dass die Vektoren nicht nur aus drei Zahlen bestehen, sondern aus vielen mehr, was auch notwendig ist, um die „richness observed in language“ abbilden zu können (vgl. Forcada 2017:295).

Decodierungsvektor im besten Fall folgendermaßen aussieht: $D(„The house is red.“, „Das“)$. Anschließend wird ein neuer Vektor p für die zweite Position im Zielsatz erstellt, $p(x \mid „The house is red.“, „Das“)$, und mit demselben Verfahren ein zielsprachlicher Vektor ausgewählt, um $D(„The house is red.“, „Das Haus“)$ zu erstellen. Dies wird so lange wiederholt, bis das System bei $D(„The house is red.“, „Das Haus ist rot.“)$ angelangt ist, aus dem der eigentliche Output generiert wird, der den Nutzer_innen des Übersetzungssystems angezeigt wird (vgl. Forcada 2017:298f.).

Anhand der obigen Beschreibung wird ersichtlich, dass sich solche NMT-Systeme bei der Übersetzung eines Wortes nicht nur auf dieses Wort „konzentrieren“, sondern immer auch den vorher erstellten Vektor in Betracht ziehen, der dementsprechend stets zum nächsten Schritt zurückgeführt wird (*fed back*). Darum wird diese Art von neuronalen Netzwerken **rekurrentes neuronales Netz** (*recurrent neural network*, RNN) genannt (vgl. Forcada 2017:297; van Genabith 2020:82). Die NMT-Architektur, bei der encodiert und dann decodiert wird, wird als Encoder-Decoder-Architektur oder nach Sutskever et al. (2014) als Sequence-to-Sequence-Modell (seq2seq-Modell) bezeichnet (vgl. Forcada 2017:299; van Genabith 2020:88).

RNN-basierte Encoder-Decoder-Architekturen werden zumeist mit einem sogenannten **Aufmerksamkeitsmechanismus** (*attention mechanism*; Bahdanau et al. (2014)) verfeinert. Dadurch zieht das NMT-System oder, genauer gesagt, der Decoder, nicht nur die vollständige Repräsentation des Ausgangssatzes ($E(„The house is red.“)$) bei der Decodierung in Betracht, sondern auch alle auf dem Weg zu diesem Vektor erstellten Repräsentationen, d. h. $E(„The“)$, $E(„The house“)$ etc. (vgl. Forcada 2017:299). So versucht der Decoder, herauszufinden, “which source words are the most important ones for the current task (generating the next output) of the decoder“ (van Genabith 2020:91f.), was die Qualität der Übersetzungen stark steigert (vgl. van Genabith 2020:92).

2017 führten Vaswani et al. (2017) eine neue NMT-Architektur ein, die sogenannte **Transformer-Architektur**, die ganz ohne RNNs auskommt und sich dafür nur auf den Aufmerksamkeitsmechanismus stützt, was die für das Training benötigten rechnerischen Ressourcen reduziert (vgl. van Genabith 2020:98f.). In einer Transformer-Architektur gibt es ebenfalls einen Encoder und einen Decoder, und beide werden durch den oben beschriebenen Aufmerksamkeitsmechanismus verbunden, der in diesem Kontext *cross-attention* genannt wird. Allerdings werden im Transformer auch noch zwei andere Aufmerksamkeitsmechanismen eingesetzt, nämlich Selbstaufmerksamkeit (*self-attention*) und maskierte Aufmerksamkeit (*masked attention*) (vgl. van Genabith 2020:99f.). Selbstaufmerksamkeit

wird im Encoder auf mehreren Ebenen verwendet und berechnet für jedes Input-Wort, wie es mit den anderen Input-Wörtern im gerade zu übersetzenden Satz zusammenhängt (vgl. van Genabith 2020:101). Die maskierte Aufmerksamkeit im Decoder funktioniert auf dieselbe Art und Weise (auch auf mehreren Ebenen), wobei jedoch in jedem Schritt nur auf die schon generierten Output-Wörter geachtet wird und jene, die noch generiert werden müssen, versteckt (*masked*) werden (vgl. van Genabith 2020:103). Für genauere Erklärungen zur Transformer-Architektur und dazu, wie die verschiedenen Aufmerksamkeitsmechanismen darin eingesetzt werden, wird auf van Genabith (2020:98-104) verwiesen.

Abgesehen von RNN-basierten Systemen und Transformer-Systemen (die in Google Translate kombiniert werden (vgl. Caswell & Liang 2020)) gibt es noch eine dritte Art von Encoder-Decoder-Architekturen, nämlich solche, die auf ***convolutional neural networks*** (CNN) basieren, von Gehring et al. (2017) eingeführt wurden und im DeepL-Übersetzer verwendet werden (vgl. Merkert 2017). Bei diesen Netzwerken funktioniert die Encodierung und Decodierung anders als bei den RNNs, was im Folgenden am selben Beispielsatz („The house is red.“) gezeigt wird.

Bei der **Encodierung** wird keine vollständige Repräsentation des Ausgangssatzes erzeugt, sondern Vektoren mit Kontextinformationen für jedes einzelne Wort, indem mehrere Wörter (z. B. zwei) links und rechts neben dem zu repräsentierenden Wort ebenfalls in Betracht gezogen werden. Der Ausgangssatz „The house is red.“ wird demnach vom System als „NULL NULL The house is red. NULL NULL“ betrachtet, wobei links vom Satzanfang und rechts vom Satzende Leerstellen („NULL“) eingefügt werden, da es an diesen Stellen keine Kontextwörter mehr gibt (der Satz ist eine isolierte Einheit). Für das Wort „The“ würde der Vektor dementsprechend $R(\text{„NULL NULL The house is“})$ lauten, für „house“ $R(\text{„NULL The house is red“})$ usw. Mit diesen Vektoren, die parallel erstellt werden können, werden weitere Vektoren berechnet, und der ganze Prozess wird zusätzlich dazu noch einige Male wiederholt (*convoluted*) (vgl. Forcada 2017:299f.).

Auch die **Decodierung** erfolgt mittels Leerstellen: In einen z. B. dreistelligen Leerstellenvektor $D(\text{„NULL NULL NULL“})$ wird das erste Wort des Zielsatzes, das mittels der während der Encodierung erstellten Vektoren berechnet wird, eingesetzt, um $D(\text{„NULL NULL Das“})$ zu ergeben. Anhand dieses D-Vektors sowie der Encodierungsvektoren wird dann der Satz nach und nach vervollständigt, bis den Nutzer_innen eines solchen Systems die Übersetzung „Das Haus ist rot.“ angezeigt werden kann (vgl. Forcada 2017:300).

Eine solche CNN-basierte Architektur kann ebenfalls durch einen Aufmerksamkeitsmechanismus auf mehreren Ebenen (wie beim Transformer) verfeinert werden, was beim

DeepL-Übersetzer der Fall ist. Darüber hinaus wird bei DeepL auch ein sogenannter Beam-Search-Algorithmus verwendet (vgl. Merkert 2017), der dem **Garden-Path-Problem** entgegenwirken soll. Dieses Problem entsteht dadurch, dass bei der Decodierung, wie sie oben beschrieben wurde (und zwar im Falle von RNNs und CNNs), immer das wahrscheinlichste nächste Wort in den Decodierungsvektor eingesetzt wird. Dieses wahrscheinlichste Wort wird dann bei der Berechnung des nächsten Wortes miteinbezogen usw., sodass das NMT-System einem bestimmten Pfad (*path*) in der Decodierung folgt. Dieser Pfad kann sich allerdings an einer späteren Stelle im Satz als falsch herausstellen, was z. B. bei fixen Wendungen der Fall sein kann (vgl. Koehn 2020:144).

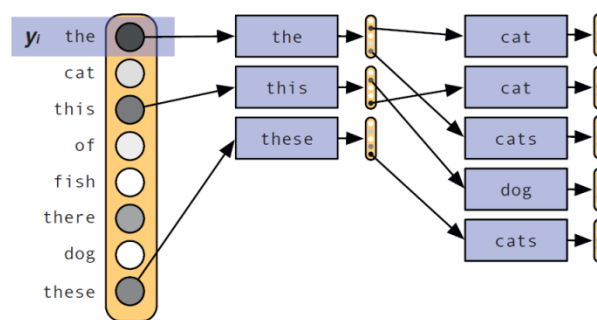


Abb. 2: Beispielhafte Darstellung der Funktionsweise des Beam-Search-Algorithmus (Koehn 2020:146)

Um dies zu umgehen, wird der **Beam-Search-Algorithmus** eingesetzt: Bei der Berechnung des wahrscheinlichsten Wortes an der ersten Position im Zielsatz wird ein Beam der n (z. B. drei) wahrscheinlichsten Wörter beibehalten, wie beispielsweise in Abb. 2 mit „the“, „this“ und „these“ gezeigt. Diese werden nach ihrer Wahrscheinlichkeit sortiert. Danach wird mit jedem dieser Wörter ein Decodierungsvektor erstellt, um das nächste wahrscheinlichste Wort zu berechnen, das sich je nach verwendetem Wort im Decodierungsvektor unterscheiden wird (für „the“ sind in der Abbildung „cat“ und „cats“ am wahrscheinlichsten, aber für „this“ „cat“ und „dog“). Die Wahrscheinlichkeit des jeweils ersten Wortes wird mit den Wahrscheinlichkeiten der von diesem Wort abhängigen Wörter in der zweiten Satzposition kombiniert, und der nächste Beam wird mit den „besten“ Wortpaaren erstellt (in der Abbildung etwa „the cat“ etc.). Dies wird solange wiederholt, bis alle so generierten Übersetzungshypothesen am Satzende angelangt sind (was für unterschiedliche Hypothesen unterschiedlich lange dauern kann). Die beste Hypothese, die durch Multiplikation der Wortwahrscheinlichkeiten determiniert wird, wird als Übersetzung ausgewählt (vgl. Koehn 2020:145f.).

Ein auf CNN basierendes NMT-System mit Aufmerksamkeitsmechanismus auf mehreren Ebenen und Beam-Search-Algorithmus war laut Merkert (2017) der Stand der

Technik bei **DeepL** im Jahr 2017, obwohl dies nicht mit Sicherheit gesagt werden kann, da „DeepL sein System nicht so freigiebig erklärt wie die Konkurrenz“ (Merkert 2017). Dennoch kann angenommen werden, dass DeepL seither entscheidende Veränderungen am System vorgenommen hat, da 2020 im Blog der Firma ein „[e]rneuter Durchbruch bei der KI-Übersetzungsqualität“ (DeepL 2020a) angekündigt wurde. Dabei könnte es sich um die Einführung von „context-aware neural machine translation“ nach Voita et al. (2018) in den DeepL-Übersetzer handeln, wie von Khalilov (2020) behauptet wird.

Bei der *context-aware* NMT wird, im Gegensatz zu den oben beschriebenen Prozessen, ein Satz nicht als isolierte Einheit betrachtet und übersetzt, sondern der Encoder erstellt neben der Repräsentation des zu übersetzenden Satzes auch eine Repräsentation für einen Kontextsatz (bzw. die Wörter dieses Kontextsatzes). Diese zwei Repräsentationen werden durch einen hier angesetzten Aufmerksamkeitsmechanismus kombiniert, um eine *context-aware* Repräsentation des Ausgangssatzes zu schaffen. Dadurch wird v. a. die Übersetzung von Pronomen verbessert, und anaphorische Relationen (siehe Kapitel 5.1) können leichter korrekt aufgelöst werden (vgl. Voita et al. 2018:1264f.).

Es gibt selbstverständlich auch noch **andere Möglichkeiten**, wie die Firma DeepL ihr System verbessert haben könnte: So vermutet der Forscher Andy Way in einem Sator-Artikel (Wyndham 2021), dass DeepL wahrscheinlich nicht mehr CNNs für das Übersetzungssystem, sondern die oben beschriebene Transformer-Architektur verwendet, die schließlich mittlerweile aufgrund der besseren Output-Qualität der dominante NMT-Ansatz ist. Laut Way wäre es sogar „überraschend“, würde DeepL immer noch CNNs verwenden (vgl. Wyndham 2021).

Ob der DeepL-Übersetzer nun tatsächlich *context-aware* ist oder statt CNNs die Transformer-Architektur einsetzt oder doch ganz andere Verbesserungen am System vorgenommen wurden, konnte aufgrund der spärlichen Informationen dazu nicht klar eruiert werden. Jedenfalls wird aber bei DeepL behauptet, dass die „neuen neuronalen Netze [...] in der Lage [sind], die Bedeutung der übersetzten Sätze in der Zielsprache wesentlich präziser darzustellen und gleichzeitig häufig eine professionellere Formulierung zu finden“ (DeepL 2020a). Außerdem wurden auch Blindtests mit dem neuen System durchgeführt, in denen professionelle Übersetzer_innen laut firmeneigenen Angaben die DeepL-Outputs viermal häufiger als beste Übersetzungen wählten als jene der Konkurrenz (vgl. DeepL 2020a).

Trotzdem wird wohl auch bei DeepL-Übersetzungen in den meisten Fällen ein sogenanntes Postediting notwendig sein. Worum es sich dabei genau handelt, wie dieses Konzept in der Industrie angewandt und in der Forschung untersucht wird, wird im nächsten Kapitel erläutert.

2 Überblick über das Postediting

In diesem Kapitel wird ein kurzer Überblick über das Konzept des Posteditings geliefert. Nach einer Erklärung, worum es sich beim Postediting überhaupt handelt und warum es bei Übersetzer_innen einen eher schlechten Ruf hat, werden verschiedene Arten des Posteditings unterschieden und Postediting-Richtlinien besprochen. Danach verschiebt sich der Fokus auf die Forschung zum Postediting, wobei etablierte Forschungsgebiete und -techniken angeschnitten und neue Entwicklungslinien präsentiert werden. Dabei muss beachtet werden, dass die in diesem Kapitel erwähnten Artikel und Studien nur beispielhaft angeführt sind und keineswegs den gesamten Forschungsstand in diesem Bereich repräsentieren. Außerdem beziehen sich diese Forschungsarbeiten nur auf nicht-literarische Texte, da Studien zum Postediting von maschinell übersetzten literarischen Texten erst im nächsten Kapitel behandelt werden.

Postediting wird laut Allen (2003:297) in vielen verschiedenen Feldern der natürlichen Sprachverarbeitung verwendet, z. B. in Bezug auf die automatische Fehlerkorrektur, die optische Zeichenerkennung, Translation Memorys und Kontrollierte Sprache, das Konzept wird aber zumeist mit der MT in Verbindung gebracht. Deswegen definiert Allen (2003:297) die Aufgabe von Posteditor_innen und damit auch das Postediting folgendermaßen: „to edit, modify and/or correct pre-translated text that has been processed by an MT system from a source language into (a) target language(s)“. Guerberof-Arenas (2020:334) fügt hinzu, dass das Postediting unter Konsultation des Ausgangstextes (wobei das nicht immer der Fall sein muss; siehe unten) und im Hinblick auf spezifische, vorab zu bestimmende Qualitätskriterien erfolgt, wobei der Zweck der Nutzung von MT und Postediting v. a. darin besteht, die Produktivität zu erhöhen und den Übersetzungsprozess zu beschleunigen. Außerdem werden beim Postediting MT-Fehlertypologien eingesetzt, um die MT-Systeme durch Rückführung der korrigierten Texte zu verbessern und um, wie in der Einleitung schon erwähnt, die Posteditor_innen über die Fehlerarten zu informieren, die häufig im MT-Output vorkommen, denn so können entsprechende Fehler leichter gefunden und korrigiert werden (vgl. Guerberof-Arenas 2020:337).

Das Postediting existiert seit den Anfängen von MT (vgl. Vieira 2020:319) und ist somit die älteste Form der **Kooperation zwischen Mensch und Maschine im Bereich der Übersetzung** (vgl. O'Brien et al. 2014:vii). Allerdings wurde das Postediting damals eher als unerwünschter letzter Schritt angesehen (vgl. Reifler 1954:30), also als passive Tätigkeit, um die Lücke zwischen fehlerhaften MT-Outputs und dem ultimativen Ziel der MT(-Forschung)

zu schließen, nämlich der voll automatischen Generierung von hochqualitativen Übersetzungen. Im Endeffekt halfen demnach die Menschen der Maschine, und diese maschinenzentrierte Sichtweise, gekoppelt mit den aus heutiger Sicht sehr schlechten MT-Outputs, mit denen die Posteditor_innen damals arbeiten mussten, führte zum schlechten Ruf des Posteditings (vgl. Vieira 2020:319). Dazu kommt heutzutage noch der Druck vonseiten der Industrie in Form von Normen wie der ISO 18587 (2017), in der als Vorteile der Nutzung von MT und Postediting die erhöhte Produktivität und Wettbewerbsfähigkeit angegeben werden – eine Botschaft, die laut do Carmo & Moorkens (2021:38) durchaus als „post-edit or perish“ interpretiert werden kann. Zusätzlich dazu werden der Arbeit von Übersetzer_innen als Posteditor_innen durch Richtlinien wie jene der Translation Automation User Society (TAUS) (Massardo et al. 2016) und daran angelehnte Richtlinien vielerlei Beschränkungen auferlegt (vgl. do Carmo & Moorkens 2021:38; siehe unten).

Allerdings ist die Situation dabei, sich zu verändern: MT ist mittlerweile in Tools zum computergestützten Übersetzen (CAT-Tools) verfügbar, d. h., aus einer maschinenzentrierten wird eine menschenzentrierte Sichtweise, in der MT als eine Ressource für Übersetzer_innen gesehen wird, mit der sie ihre Produktivität durch Postediting von MT-Output neben der Benutzung von Translation Memorys (TM) erhöhen können, wenn sie das wollen (vgl. Vieira 2020:319f.). Angesichts dieser konfligierenden Tendenzen (schlechter Ruf, Druck durch Normen, Beschränkungen durch Richtlinien vs. MT als Hilfestellung für Übersetzer_innen) scheint es nicht verwunderlich, dass Studien zur Einstellung von Übersetzer_innen zur MT und zum Postediting sowohl eher negative als auch eher positive Auffassungen zu Tage fördern (vgl. z. B. auf der eher negativen Seite Alvarez Vidal et al. (2020) und Gaspari et al. (2014), auf der eher positiven Seite Daems et al. (2017) und Lorena (2020) und für eher neutrale oder flexible Ergebnisse Guerberof-Arenas (2013) und Moorkens & O’Brien (2015)).

Die Integration von MT in CAT-Tools bedeutet laut Vieira (2020:320) auch, dass sich das Konzept des Posteditings „in a state of terminological flux“ befindet, weswegen es auch mehrere **Arten des Posteditings** gibt, die nach unterschiedlichen Merkmalen unterschieden werden: Beim interaktiven Postediting interagieren die Übersetzer_innen mit dem MT-System, während die Übersetzung angefertigt wird (z. B. wenn MT in ein CAT-Tool integriert ist oder das MT-System für die Sätze, die gerade geschrieben werden, Vervollständigungen vorschlägt). Dahingegen wird beim statischen Postediting der komplette MT-Output erst nach dessen Generierung in einem separaten Schritt posteditiert (vgl. Vieira 2020:322). Eine weitere Art des Posteditings, die laut Vieira (2020:328) in diesem Kontinuum zu verorten ist, ist das automatische Postediting: Dieses funktioniert im Grunde wie MT,

jedoch werden für das Training dieser Systeme unveränderte MT-Outputs als „Ausgangstexte“ und die von Menschen posteditierten MT-Outputs als „Zieltexte“ verwendet (vgl. Simard et al. 2007:509).

Neben interaktivem, statischem und automatischem Postediting wird, wie oben schon angedeutet, auch zwischen ein- und zweisprachigem Postediting unterschieden (vgl. Guerberof-Arenas 2020:347; Vieira 2020:322). Meistens wird unter Postediting zweisprachiges Postediting verstanden, was auch der Situation in der Sprachindustrie sowie in kommerziellen Übersetzungstools entspricht, in denen der Ausgangstext und der Zieltext Seite an Seite präsentiert werden (vgl. Vieira 2020:320f.). Jedoch wird auch zu einsprachigem Postediting geforscht, bei dem der Ausgangstext nicht berücksichtigt wird bzw. nicht berücksichtigt werden kann, weil die Posteditor_innen die Ausgangssprache nicht beherrschen und daher den Ausgangstext nicht verstehen (Guerberof-Arenas 2020:347f.; Vieira 2020:320f.). Der Grund für die Forschung in diese Richtung ist die Überlegung, dass durch einsprachiges Postediting auf zweisprachige Personen verzichtet und damit Kosten gespart werden könnten (vgl. Vieira 2020:231).

Ein letztes Unterscheidungsmerkmal, mit dem das Postediting in mehrere Arten oder Levels in diesem Fall unterteilt werden kann, ist das Ausmaß des Posteditings, d. h. in anderen Worten, wie viel am MT-Output verändert wird bzw. welche Endqualität die Postedition erreichen soll (vgl. Allen 2003:301ff.; Guerberof-Arenas 2020:335; ISO 2017; Massardo et al. 2016:16; Vieira 2020:324). Die Anzahl der in dieser Hinsicht unterschiedenen **Postediting-Levels** sowie die Beschreibungen dieser können je nach Unternehmen oder Forschungsprojekt variieren (vgl. Vieira 2020:324), jedoch ist die Unterteilung in einfaches und vollständiges Postediting am weitesten verbreitet (vgl. Hu & Cadwell 2016:347). Obwohl laut Way (2013:[2]) diese Dichotomie nicht mehr zeitgemäß ist und das Ausmaß des Posteditings stattdessen von „the purpose, value and shelf-life of the content“ abhängt und obwohl auch in den TAUS-Richtlinien darauf hingewiesen wird, dass das Ausmaß des Posteditings von der Qualität des unveränderten MT-Outputs abhängt (vgl. Massardo et al. 2016:16), wird aufgrund der weiten Verbreitung der Unterteilung in einfaches und vollständiges Postediting dennoch im Rahmen der vorliegenden Masterarbeit nur auf diese beiden Levels eingegangen.

Beim **einfachen Postediting** besteht das Ziel darin, „to obtain a merely comprehensible text without any attempt to produce a product comparable to a product obtained by **human translation**“ (ISO 2017; Hervorhebung im Original). Texte, die so posteditiert werden, dienen folglich der Assimilierung bzw. dem Verständnis (vgl. auch Allen 2003:301), d. h., die

Postedition soll v. a. verständlich und inhaltlich genau sein, andere Aspekte einer Übersetzung (wie z. B. ihr Stil) sind jedoch weniger relevant, und teilweise sind sogar orthographische oder grammatikalische Fehler in der Postedition erlaubt (vgl. Guerberof-Arenas 2020:335).

Im Gegensatz dazu ist das Ziel des **vollständigen Posteditings** die Erreichung der Qualität einer Humanübersetzung¹² (vgl. Guerberof-Arenas 2020:335; ISO 2017). Die Posteditionen werden im Falle des vollständigen Posteditings verbreitet bzw. publiziert (vgl. auch Allen 2003:301), weswegen nicht nur inhaltliche Genauigkeit und Verständlichkeit gefordert sind, sondern eine in allen Belangen (nahezu) fehlerfreie Übersetzung (vgl. Guerberof-Arenas 2020:335).

Genauere Anweisungen, wie diese beiden (oder andere) Postediting-Levels umgesetzt werden sollen, finden sich in **Postediting-Richtlinien**, die entweder generell ausfallen oder auch sprach- und/oder inhaltsspezifisch sein können (vgl. Hu & Cadwell 2016:346). Es gibt dementsprechend keine weithin akzeptierten und angewendeten Postediting-Richtlinien, da auch die Bedürfnisse der Organisationen, Sprachdienstleister oder Forscher_innen, die solche Richtlinien zusammenstellen, unterschiedlich sind (vgl. Hu & Cadwell 2016:346; Massardo et al. 2016:14). Um den Rahmen dieser Masterarbeit mit einem Vergleich von verschiedenen Postediting-Richtlinien (von denen ohnehin nur wenige öffentlich zugänglich sind (vgl. Hu & Cadwell 2016:346)) nicht zu sprengen, wird hier nur auf die TAUS-Richtlinien (Massardo et al. 2016) eingegangen, da diese grundlegende Richtlinien darstellen, die je nach Bedürfnis angepasst werden können (vgl. Massardo et al. 2016:14).¹³

Die TAUS-Richtlinien sind, wie schon erwähnt, nicht auf einfaches und vollständiges Postediting zugeschnitten, sondern auf zwei von der Postedition zu erwartende Qualitätsstufen, nämlich „good enough“ oder „fit for purpose“ und „publishable quality“ (Massardo et al. 2016:16). Die Argumentation dahinter ist, dass für einen qualitativ sehr hochwertigen MT-Output womöglich ein eher oberflächliches Postediting ausreicht, um die Qualität einer Humanübersetzung zu erreichen, während für einen MT-Output von sehr geringer Qualität evtl. ein sehr tiefgreifendes Postediting notwendig ist, um überhaupt eine verständliche, aber im Hinblick auf die Qualität nicht mit einer Humanübersetzung vergleichbare Postedition zu erreichen (vgl. Massardo et al. 2016:16). Dennoch können die Richtlinien für *good enough*

¹² Unter „Humanübersetzung“ wird in der vorliegenden Arbeit jedwede Übersetzung verstanden, die von einem Menschen **ohne** die Hilfe von MT durchgeführt wird bzw. angefertigt wurde, wobei andere Tools wie CAT-Tools und dgl. durchaus verwendet werden können (aber nicht müssen). Ob der Prozess des Übersetzens oder die fertige Übersetzung gemeint ist, ist aus dem Kontext ersichtlich.

¹³ Für einen Vergleich verschiedener Postediting-Richtlinien wird auf Hu & Cadwell (2016) verwiesen.

und *publishable quality* auch auf einfaches und vollständiges Postediting übertragen werden (vgl. Hu & Cadwell 2016:347).

Jedenfalls besagen die Richtlinien, dass für *good enough* (bzw. beim einfachen Postediting) auf eine semantisch korrekte Postedition abgezielt werden soll, in der es weder Auslassungen noch Hinzufügungen gibt und die weder beleidigende noch (kulturell) inakzeptable Inhalte enthält. Weiters sollen die grundlegenden orthographischen Regeln der jeweiligen Zielsprache zwar beachtet, jedoch Korrekturen, die den Stil oder den Lesefluss des Textes betreffen, nicht vorgenommen werden (vgl. Massardo et al. 2016:17). Um jedoch die Qualität einer Humanübersetzung zu erreichen (bzw. beim vollständigen Postediting), soll nicht nur auf Semantik, Auslassungen, Hinzufügungen, (kulturell) adäquate Inhalte und Orthographie geachtet werden, sondern auch auf korrekte Grammatik, Syntax, Interpunktion, Silbentrennung und Formatierung sowie auf den korrekten Einsatz von wesentlicher Terminologie (inkl. Termini, die nicht übersetzt werden sollen) (vgl. Massardo et al. 2016:18).

Bei beiden Arten des Posteditings ist in den TAUS-Richtlinien auch Folgendes angeführt: „Use as much of the raw MT output as possible.“ (Massardo et al. 2016:17 und 18) Es ist v. a. diese Empfehlung, durch die laut do Carmo & Moorkens (2021:38) den Posteditor_innen in ihrer Arbeit sehr viele Beschränkungen auferlegt werden, was, wie oben schon erwähnt, zum eher schlechten Ruf des Posteditings beiträgt.

Wie zuvor angedeutet, ist diese Einstellung von Übersetzer_innen zum Postediting eines der Themen, die die Forschung beschäftigen. Die **Hauptinteressen in diesem Forschungsgebiet** liegen allerdings beim Aufwand des Posteditings bzw. der Steigerung der Produktivität von Übersetzer_innen durch MT und Postediting sowie bei der Qualität von zu posteditierenden MT-Outputs bzw. von fertigen Posteditionen. Die beiden Konzepte (Aufwand bzw. Produktivität und Qualität) sind eng miteinander verbunden: Eine potentielle Produktivitätssteigerung durch MT würde zunichtegemacht, wenn die Qualität von Posteditionen als minderwertig eingeschätzt würde, gleichzeitig würde sich das Postediting gegenüber einer Humanübersetzung nicht lohnen, wenn schon die unveränderten MT-Outputs von sehr geringer Qualität wären (vgl. Vieira 2020:320).

In Bezug auf den **Postediting-Aufwand** sind in der Forschung v. a. die folgenden drei von Krings (2001:178) eingeführten Dimensionen relevant: zeitlicher, kognitiver und technischer Aufwand. Der zeitliche Aufwand bezieht sich auf die Zeit, die für das Postediting benötigt wird, und ist die in wirtschaftlicher Hinsicht wichtigste Dimension. Der kognitive Aufwand bezeichnet die Art und das Ausmaß der kognitiven Prozesse, die bei der Erkennung

von Fehlern und der Entwicklung einer Korrekturstrategie beim Postediting aktiviert werden müssen, während die technische Dimension den Aufwand beschreibt, der aufgebracht werden muss, um die fehlerhaften Passagen im MT-Output zu korrigieren, d. h. Operationen wie Löschen, Hinzufügen, Umstellen etc. (vgl. Krings 2001:178f.).

Oft werden eine (v. a. die zeitliche, da diese am einfachsten zu messen ist (vgl. Krings 2001:178)), zwei oder alle drei dieser Aufwandsdimensionen während des Posteditings sowie während einer Humanübersetzung erhoben, um die beiden Übersetzungsmodalitäten bzgl. ihres Aufwands miteinander vergleichen zu können (vgl. Vieira 2020:321): Mithilfe von Techniken wie etwa Schreib-Logging, Screen-Recording, Eye-Tracking, Think-Aloud-Protokollen und Fragebögen (vgl. Guerberof-Arenas 2020:339ff.) werden Daten gesammelt, die dann teilweise auch zur Berechnung von Messwerten für den (kognitiven) Aufwand verwendet werden, wie z. B. dem Verhältnis der Anzahl der Pausen zur Anzahl der Wörter (*pause-to-word ratio*) (vgl. Lacruz & Shreve 2014:257)¹⁴. Meist werden für die Erhebung der Daten spezielle Tools verwendet (vgl. Guerberof-Arenas 2020:339-342), etwa PET¹⁵ (Aziz et al. 2012) oder Translog¹⁶ (Translog o. J.).

Bei solchen Vergleichen des Aufwands für das Postediting und das Humanübersetzen werden in vielen Studien, darunter Daems et al. (2017:254f.), Green et al. (2013:439), Guerberof-Arenas (2014:173), Läubli et al. (2013:86), Zhechev (2014:9) etc., beim Postediting ein geringerer (zeitlicher) Aufwand und damit Produktivitätssteigerungen gegenüber dem Humanübersetzen verzeichnet (vgl. Guerberof-Arenas 2020:345; Vieira 2020:321). Das Potential zur Produktivitätssteigerung durch MT und Postediting ist demnach wissenschaftlich etabliert¹⁷ (vgl. Vieira 2020:322), allerdings gibt Guerberof-Arenas (2020:345f.) zu bedenken, dass die Daten, die zu diesen Ergebnissen führten, teilweise in sehr spezifischen Settings (im Sinne von Sprachkombination, MT-System, Ausgangstext etc.) erhoben wurden und dass allgemein eine sehr hohe Variabilität bei den Teilnehmer_innen in solchen Postediting-Studien besteht.

Tatsächlich wird auch an der Vorhersage des Postediting-Aufwands anhand der zu übersetzenden Ausgangstexte bzw. des zu posteditierenden MT-Outputs geforscht, z. B. in Daems et al. (2015), O'Brien (2011), Tatsumi (2009) oder Vieira (2014). Dabei werden

¹⁴ Für einen Überblick und einen Vergleich verschiedener Messwerte zum kognitiven Aufwand wird auf Vieira (2016) verwiesen.

¹⁵ <http://www.clg.wlv.ac.uk/projects/PET/> (Stand: 08.05.2021).

¹⁶ <http://www.translog.dk/> (Stand: 10.05.2021).

¹⁷ Es gibt aber durchaus auch Studien, die in einer oder mehreren der Aufwandsdimensionen beim Postediting gleich viel oder sogar mehr Aufwand als beim Humanübersetzen verzeichnen, z. B. Carl et al. (2011:137f.), Sekino (2015:150) oder Screen (2017:334), und das Postediting wird generell als kognitiv sehr anstrengend angesehen (vgl. Guerberof-Arenas 2020:347).

sowohl Charakteristika wie Satzlänge, Syntax sowie Anzahl und Verteilung von Wortarten in Texten als auch automatische MT-Evaluierungsmetriken (vgl. Vieira 2020:323) und Fehlerannotationen in MT-Outputs (vgl. Daems et al. 2015:32) für die Vorhersage des Aufwands herangezogen.¹⁸

Wie oben schon erwähnt, bedeutet höhere Produktivität nicht unbedingt auch höhere **Qualität**, weswegen letzteres Konzept, genauer gesagt die Qualität der durch MT und Postediting produzierten Übersetzungen, ebenfalls im Fokus der Postediting-Forschung steht. Dabei wird die Qualität meist von menschlichen Evaluator_innen evaluiert, und zwar auf verschiedene Arten und Weisen: Es können Fehler in den fertigen Posteditionen und Humanübersetzungen markiert bzw. annotiert werden, wie etwa in Daems et al. (2017:258), Guerberof-Arenas (2014:177) oder Plitt & Masselot (2010:14). Die Posteditionen und Humanübersetzungen können aber auch in paarweisen Rankings bewertet werden, wie in Green et al. (2013:442f.) oder Läubli et al. (2013:87f.). Schließlich kann die Bewertung auch mithilfe von Likert-Skalen oder Fragebögen in Bezug auf verschiedene Kategorien durchgeführt werden, etwa inhaltliche Genauigkeit, Klarheit, Stil, Grammatik, Syntax, Übersetzungsstrategien etc. (vgl. Fiederer & O'Brien 2009:56; Läubli et al. 2013:87; Screen 2017:332). Etwas andere Kategorien kamen hingegen in Screen (2019:142) und van Egdom & Pluymaekers (2019:164f.) zum Einsatz, wie z. B. Lesbarkeit und Verständlichkeit bzw. Nutzbarkeit, Einstellung gegenüber dem in der Übersetzung beschriebenen Produkt, Kaufabsicht etc., da hier, im Gegensatz zu den oben erwähnten Studien, die Evaluator_innen keine (erfahrenen oder sich in Ausbildung befindlichen) Übersetzer_innen oder Reviewer_innen waren, sondern Endnutzer_innen der Übersetzungen (vgl. Screen 2019:133; van Egdom & Pluymaekers 2019:160¹⁹).

Wie bei der Forschung zum Postediting-Aufwand deuten auch die Ergebnisse der Forschung zur Qualität, u. a. jene der oben genannten Studien, darauf hin, dass, „contrary to a more ‘popular belief’ amongst translation practitioners“ (Guerberof-Arenas 2014:167), die Qualität der Übersetzungen unter der Anwendung von Postediting von MT-Output nicht leidet (vgl. Guerberof-Arenas 2020:346; Vieira 2020:321)²⁰. Allerdings gibt Guerberof-

¹⁸ Diese Forschungsrichtung ist eng verbunden mit dem Konzept des Preeditings, d. h. die Vorbereitung eines Ausgangstextes für die MT, um eine höhere Qualität der Outputs zu erreichen (vgl. Guerberof-Arenas 2020:334).

¹⁹ Im Falle von van Egdom & Pluymaekers (2019) wurden die Posteditionen nicht mit Humanübersetzungen verglichen, sondern untereinander, da unterschiedliche Postediting-Levels angewendet wurden (vgl. van Egdom & Pluymaekers 2019:160).

²⁰ Natürlich gibt es auch hier Ausnahmen, z. B. verzeichnen O'Brien & Fiederer (2009:62) und Screen (2017:333) eine schlechtere Qualität der Posteditionen in Bezug auf die Kategorie „Stil“, und in Skadiņš et al. (2011:39) wurde den Posteditionen nach Fehlerannotationen ein schlechterer Qualitätswert gegeben.

Arenas (2020:346) auch hier die hohe Variabilität zwischen den Posteditor_innen zu bedenken, weswegen sowohl in Bezug auf die Produktivität beim Postediting als auch auf die Qualität der daraus resultierenden Posteditionen v. a. die Erfahrung der Übersetzer_innen, die als Posteditor_innen fungieren, als wichtiger Faktor genannt wird.

Jedoch gibt es in Bezug auf den **Einfluss der Erfahrung** auf die Produktivität und die Qualität keine eindeutigen Ergebnisse: In manchen Studien, z. B. de Almeida & O'Brien (2010) oder Moorkens & O'Brien (2015), waren die erfahrenen Übersetzer_innen schneller beim Postediting als die unerfahrenen (vgl. de Almeida & O'Brien 2010:[4]; Moorkens & O'Brien 2015:78) und produzierten auch qualitativ hochwertigere Posteditionen (vgl. Moorkens & O'Brien 2015:78f.). Dahingegen gab es in Daems et al. (2017:254) und Sekino (2015:150) keinen (statistisch relevanten) Unterschied zwischen erfahrenen Übersetzer_innen und Studierenden bzgl. des zeitlichen und in Daems et al. (2017:255) auch bzgl. des kognitiven Aufwands. Außerdem schließen Daems et al. (2017:258) aus ihren Ergebnissen, dass es zumindest bei der Übersetzung bzw. dem Postediting von MT-Output von allgemein-sprachlichen Texten keinen Qualitätsunterschied zwischen der Arbeit von Studierenden und jener von erfahrenen Übersetzer_innen gibt. In Aranberri et al. (2014:25f.) waren die Ergebnisse bzgl. der Produktivität geteilt: Bei einem Text war die Produktivitätssteigerung durch Postediting bei den unerfahrenen Übersetzer_innen größer, beim anderen Text jedoch bei den erfahrenen Übersetzer_innen, was darauf schließen lässt, dass auch andere Faktoren in dieser Hinsicht von Bedeutung sind, z. B. die Postediting-Erfahrung und die Einstellung gegenüber MT, die Schwierigkeit des Ausgangstextes und die Qualität des MT-Outputs (vgl. Aranberri et al. 2014:27).

Abgesehen von diesen Faktoren scheint auch der Ansatz des MT-Systems, das für die Generierung des MT-Outputs verwendet wird, Auswirkungen auf das Postediting und die daraus resultierenden Posteditionen zu haben. Die Forschung zu diesem Aspekt des Posteditings wird v. a. seit dem Aufkommen von NMT-Systemen intensiv betrieben und stellt damit eine eher neuere Entwicklungslinie in der Postediting-Forschung dar (vgl. Vieira 2020:325f.). Jedenfalls hat die **Nutzung von NMT** Auswirkungen auf das Postediting, denn der Output dieser Systeme ist flüssiger (*fluency*), aber inhaltlich (*adequacy*)²¹ weniger genau als jener von PBSMT-Systemen (vgl. Castilho et al. 2017:116; siehe auch Way (2020:320)),

²¹ „Fluency refers to the degree to which the target is well formed according to the rules of [...] [the target language]“, während „[a]dequacy refers to the degree to which information present in the original is also communicated in the translation“ (LDC 2002:[3]). Für *adequacy* wird auch die Bezeichnung *accuracy* im Kontext der MT-Forschung benutzt (vgl. Way 2018:164). Die beiden Konzepte werden für die Evaluierung von MT-Output verwendet (vgl. O'Brien 2012:66) und sind weit verbreitet (es gibt z. B. eigens von TAUS zusammengestellte Hinweise für die Verwendung dieser Konzepte (vgl. TAUS 2017)).

welche in den in diesem Kapitel bereits erwähnten Studien (außer Screen (2019)) verwendet wurden. Laut den Teilnehmenden in Castilho et al. (2017:117) seien Fehler im NMT-Output deswegen schwieriger zu erkennen, doch die Ergebnisse von Koponen et al. (2019:82) bestätigen diese Aussage nicht. Auch in Yamada (2019:102) führte die Verwendung von NMT zu einer höheren Endqualität der Posteditionen. In Bezug auf den Aufwand verzeichnen Castilho et al. (2017:117), Koponen et al. (2019:83) und Jia et al. (2019a:26) beim NMT-Postediting eine Reduktion der zeitlichen und technischen Dimension bzw. aller drei Aufwandsdimensionen im Vergleich zum PBSMT-Postediting, während Yamada (2019:102) feststellt, dass der kognitive Aufwand beim NMT-Postediting nicht geringer ist.

Postediting von NMT-Output wird auch mit der Humanübersetzung unter Einsatz eines TM sowie mit der reinen Humanübersetzung²² verglichen. Eine Kombination von Vorschlägen aus einem TM sowie von Vorschlägen von einem NMT-System für Segmente, für die es keine hochprozentigen Matches im TM gibt, scheint die Arbeit von Übersetzer_innen zu beschleunigen (vgl. Läubli et al. 2019:271), obwohl das Postediting von NMT-Segmenten laut Sánchez-Gijón et al. (2019:56) länger dauert als das Adaptieren von Vorschlägen aus dem TM. In Bezug auf die reine Humanübersetzung scheinen der zeitliche und kognitive Aufwand für das NMT-Postediting geringer zu sein (vgl. Jia et al. 2019b:79; Stasimioti & Sosoni 2020:120). Jia et al. (2019b:79) fanden ebenfalls heraus, dass beide Modalitäten eine vergleichbare Endqualität hervorbringen, während Stasimioti und Sosoni (2020:120) auch eine Reduktion im technischen Aufwand feststellten.

Als zweite Entwicklungslinie in der Postediting-Forschung, die ebenfalls durch das Auftauchen von NMT intensiviert wurde, ist die Erforschung des **Posteditings von maschinell übersetzten literarischen Texten** zu nennen (vgl. Vieira 2020:326). Auch in diesem Bereich spielen die oben besprochenen Konzepte des Aufwands, der Qualität, der Einstellung der Übersetzer_innen zum Postediting, der Arten des Posteditings etc. eine Rolle. Darüber hinaus gibt Vieira (2020:327) zu bedenken, dass in der Evaluierung von posteditiertem LMT-Output auch andere Konzepte berücksichtigt werden müssen, wie etwa das der Kreativität oder des Leseerlebnisses. Auf diese neue Entwicklungslinie der Postediting-Forschung wird im Rahmen des nächsten Kapitels – dem Forschungsüberblick über die LMT – näher eingegangen.

²² Unter „reiner Humanübersetzung“ als Unterbegriff zur in Fußnote 12 definierten „Humanübersetzung“ wird in der vorliegenden Arbeit jedwede Übersetzung verstanden, die von einem Menschen ohne Hilfe durch CAT-Tools oder MT durchgeführt wird bzw. angefertigt wurde (in englischsprachigen Artikeln oft „from-scratch translation“ oder „translation/translate from scratch“ genannt, z. B. in Kenny & Winters (2020a:134) oder Toral et al. (2018:1)). Ob der Prozess des Übersetzens oder die fertige Übersetzung gemeint ist, ist aus dem Kontext ersichtlich.

3 Forschungsüberblick: literarische maschinelle Übersetzung

Obwohl an MT seit der Existenz von Computern geforscht wird (vgl. Poibeau 2017²³), überrascht es nicht, dass die Forschung zur LMT erst im letzten Jahrzehnt ihren Anfang nahm, ist doch Literatur „considered to be the greatest challenge for machine translation“ (Toral & Way 2018:264). Dennoch gibt es, wie schon erwähnt, Forschung in diesem Bereich, die in diesem Kapitel mit einem Fokus auf Studien, die Fehlerkategorien und/oder Postediting involvieren, dargelegt wird.

Genzel et al. (2010) und **Greene et al. (2010)** gaben den Startschuss für die Forschung zur LMT: Sie benutzten SMT-Systeme dazu, Gedichte zu übersetzen, wobei die Systeme bei der Übersetzung vorprogrammierten Regeln in Bezug auf Reime, Metrum und Länge (vgl. Genzel et al. 2010:163) bzw. Rhythmus (vgl. Greene et al. 2010:254) folgen mussten.

Nach **Lee (2011)**, der eine Fallstudie über ein englisch-chinesisches Lyrikprojekt durchführte, in dem MT benutzt wurde, verschob sich ab dem Jahr 2012 der Fokus zunehmend von Lyrik hin zu Prosa. Dennoch wurde durchaus vereinzelt noch LMT-Forschung mit Gedichten betrieben (siehe unten).

Voigt & Jurafsky (2012) forschten als erste zur LMT von Prosa und etablierten die Wichtigkeit von referentieller Kohäsion, definiert als „the relation between co-referring entities in a narrative“ (Voigt & Jurafsky 2012:19, nach Halliday & Hasan 1976:31f.), bei der LMT. Daraus schlossen sie, dass die Berücksichtigung von größeren Kontexten als dem eines einzigen Satzes, also eines „intersentential context“ anstatt eines „intrasentential context“ (Moorkens et al. 2018:256), für die LMT wünschenswert wäre (vgl. Voigt & Jurafsky 2012:18).

Nach **Jones & Irvine (2013)**, die französische Lyrik und Prosa mithilfe eines SMT-Systems ins Englische übersetzten, den Output mit den jeweiligen Humanübersetzungen verglichen und ihre Ergebnisse im Kontext von Konzepten von Schleiermacher (1813, zit. nach Rössler 2002) und Venuti (²2008) diskutierten, erfolgte im Jahr 2014 die erste Postediting-Studie im Rahmen der LMT-Forschung: In **Besacier (2014)** wurde die französische, durch ein PBSMT-System erstellte Übersetzung eines englischsprachigen literarischen Textes einem Postediting sowie einer Revision unterzogen (vgl. Besacier 2014:390f.). Der resultierende Text wurde von menschlichen Evaluator_innen als „globalement lisible“, „compréhensible“ und „contenant peu de fautes“²⁴ (Besacier 2014:393) eingeschätzt, wobei der Humanübersetzer des Ausgangstextautors jedoch anmerkte, dass eine

²³ Für dieses Werk können keine Seitenzahlen angegeben werden, da es ein reines E-Book ist.

²⁴ „überwiegend lesbar“, „verständlich“, „mit wenigen Fehlern“ (übers. S.K.)

reine Humanübersetzung einen natürlicheren Text ergeben hätte, in dem der Einfluss des Englischen weniger spürbar gewesen wäre (vgl. Besacier 2014:393).

Ebenfalls im Jahr **2014** führten **Toral & Way** ein kleines Experiment durch, mit dem sie zeigten, dass die SMT zwar nicht für Gedichte und für nicht verwandte Sprachen nützlich zu sein scheint, wohl aber für die Übersetzung von Romanen zwischen verwandten Sprachen (vgl. Toral & Way 2014:177). Dementsprechend trainierten die beiden Forscher ein Jahr später erstmals MT-Systeme u. a. mit Romanen für die Übersetzungsrichtung Spanisch-Katalanisch (vgl. **Toral & Way 2015a**:127). Dabei stellte sich heraus, dass diese mit literarischen Texten trainierten Systeme stets bessere Ergebnisse bei der automatischen Evaluierung erzielten als die Vergleichssysteme, für die beim Training keine literarischen Texte verwendet wurden (vgl. Toral & Way 2015a:128). Die guten Ergebnisse bestätigten sich beim Vergleich der Übersetzung des besten dieser an Literatur angepassten Systeme mit der Humanübersetzung (vgl. Toral & Way 2015a:129f.). Genauere Untersuchungen dieses LMT-Outputs in einer weiterführenden Studie (**Toral & Way 2015b**) förderten v. a. Probleme mit Pronomen und Zeitenbildung sowie bei der Übersetzung von unbekannten Wörtern zutage (vgl. Toral & Way 2015b:258-262). Darüber hinaus konnten die beiden Forscher mittels weiterer Untersuchungen ihre oben angeführte Einschätzung, SMT für Literatur sei besser geeignet für die Übersetzung zwischen verwandten Sprachen, nicht unbedingt wiederholen; stattdessen habe die Art der angestrebten Übersetzung (freiere Übersetzung oder näher am Ausgangstext) mehr Einfluss auf die Eignung von MT zur Literaturübersetzung als die Sprachkombination (vgl. Toral & Way 2015b:254f.).

2016 verglichen **Bentivogli et al.** die Qualität der Outputs von PBSMT- und NMT-Systemen. Dabei konzentrierten sich die Forscher_innen zwar nicht auf literarische Texte, ihre Ergebnisse sind aber dennoch für die LMT von großer Wichtigkeit, denn sie fanden u. a. heraus, dass sich NMT offenbar besonders gut für lexikalisch vielfältige Texte eignet (vgl. Bentivogli et al. 2016:265), also z. B. für literarische Texte (vgl. Toral & Way 2018:264).

Zwei Jahre später publizierten Toral et al. (2018) und Moorkens et al. (2018) zwei miteinander zusammenhängende Studien zum Postediting von LMT. **Toral et al. (2018)** fokussierten auf den Aufwand des Posteditings von PBSMT- und NMT-Output: Ein Ausschnitt eines englischsprachigen Romans wurde von einem mit Romanen trainierten PBSMT- und einem ebenso trainierten NMT-System ins Katalanische übersetzt. Sechs professionelle Übersetzer_innen mit Erfahrung im Literaturübersetzen posteditierten einmal den PBSMT- und einmal den NMT-Output und fertigten eine reine Humanübersetzung eines anderen Ausschnitts des Romans an (vgl. Toral et al. 2018:2f.). Die Messungen ergaben, dass

die Übersetzer_innen beim Postediting des NMT-Outputs am schnellsten waren und die Tastatur am wenigsten benutzten, gefolgt vom PBSMT-Postediting und der reinen Humanübersetzung. Je länger die Ausgangssätze jedoch waren, desto kleiner wurde die Differenz zwischen dem zeitlichen und technischen Aufwand für das NMT-Postediting und dem Aufwand für die reine Humanübersetzung (vgl. Toral et al. 2018:10), was damit zusammenhängt, dass NMT bei längeren Sätzen schlechtere Ergebnisse liefert (vgl. Toral & Sánchez-Cartagena 2017:1069). In Bezug auf den kognitiven Aufwand fanden Toral et al. (2018:10) heraus, dass die Übersetzer_innen beim Postediting zwar weniger, aber dafür längere Pausen machten.

Moorkens et al. (2018) konzentrierten sich dahingegen auf die Einstellung der Übersetzer_innen zum Postediting, das diese im Rahmen der Studie durchgeführt hatten: Die Forscher_innen fragten die Übersetzer_innen, ob sie den MT-Output nützlich fanden und in Erwägung ziehen würden, auch künftig LMT-Output zu posteditieren. Außerdem verglichen sie die Aussagen der Übersetzer_innen mit ihren in Toral et al. (2018) beschriebenen Messungen zum Aufwand (vgl. Moorkens et al. 2018:241f.). Die Auffassungen der Übersetzer_innen bzgl. der Geschwindigkeit stimmten (meist) mit den Messungen überein, bzgl. der Tastaturanschläge gab es allerdings Diskrepanzen. Die Übersetzer_innen gaben weiters an, sie würden aufgrund der höheren Geschwindigkeit zum Postediting von LMT-Output greifen, wenn Produktivität ihre Hauptsorge wäre. Ginge es allerdings um Kreativität und eine hohe Endqualität, würden die Übersetzer_innen trotzdem die reine Humanübersetzung bevorzugen, da sie beim Postediting weniger kreativ seien (vgl. Moorkens et al. 2018:250ff.). Außerdem kritisierten sie die Segmentierung in Sätze beim Postediting, die ihnen aufgrund des fehlenden Kontexts das Gefühl gab, „in the darkness“ zu posteditieren/zu übersetzen (vgl. Moorkens et al. 2018:253). Müssten sich die Übersetzer_innen dennoch für ein MT-System entscheiden, würde ihre Entscheidung auf das NMT-System fallen, da dieses natürlichere Übersetzungen hervorbrächte (vgl. Moorkens et al. 2018:254). In Bezug auf Mehrdeutigkeiten kamen Moorkens et al. (2018:256) mit ihrer Aussage, dass Kontext, der über einen Satz hinausgeht, der nächste Schritt für literarische NMT (L-NMT) wäre, zum selben Schluss wie Voigt und Jurafsky (2012:18) in Bezug auf literarische SMT (L-SMT).

Ebenfalls im Jahr **2018** führten **Toral & Way** eine groß angelegte Studie zur LMT durch, in der sie zwölf Romane, darunter auch den siebten *Harry-Potter*-Band, von zuvor an Literatur angepassten PBSMT- und NMT-Systemen vom Englischen ins Katalanische übersetzen ließen und die Outputs einer Evaluierung unterzogen (vgl. Toral & Way 2018:267ff.).

Sowohl bei der automatischen Evaluierung durch BLEU²⁵ (vgl. Papineni et al. 2002) als auch bei der menschlichen Evaluierung durch ein Ranking schnitt das NMT- besser als das PBSMT-System ab (vgl. Toral & Way 2018:274; 281ff.), weswegen Toral & Way (2018:284) der Meinung sind, dass ausgehend von PBSMT mit NMT schon etwa 20 % des Weges zur Erreichung der Qualität einer Humanübersetzung von Literatur zurückgelegt worden sei.

2018 wurde auch für Gedichte das erste Mal NMT benutzt: **Ghazvininejad et al. (2018)** berichten über drei von ihnen konzipierte NMT-Modelle für die Übersetzung von französischen Gedichten ins Englische, wobei die Nutzer_innen der Systeme Rhythmus- und Reimmuster ähnlich wie in Genzel et al. (2010) und Greene et al. (2010) selbst bestimmen können (vgl. Ghazvininejad et al. 2018:67). In einer menschlichen Evaluierung stuften ca. 78 % der Teilnehmenden die Outputs als „okay“ bis „sehr gut“ ein (vgl. Ghazvininejad et al. 2018:69).

Auch **Humblé** konzentriert sich in seinem **2019** erschienenen Artikel auf maschinell übersetzte Gedichte, genauer gesagt auf die Google-Translate-Übersetzungen von Gedichten aus dem Englischen ins Portugiesische (vgl. Humblé 2019:44). Es zeigten sich u. a. Probleme bei Mehrdeutigkeit, bei Fällen, in denen die Zielsprache expliziter sein muss als die Ausgangssprache, und bei literarischen Mitteln, die in Gedichten häufig verwendet werden, wie z. B. Alliterationen. Allerdings waren die Google-Translate-Übersetzungen öfter als erwartet richtig (vgl. Humblé 2019:48), und der Autor schätzt, dass sie als „a kind of safeguard for the human translator“ (Humblé 2019:48) funktionieren könnten, was die Auffassung eines_einer Teilnehmenden in Moorkens et al. (2018) widerzuspiegeln scheint, der_die angab, dass MT „gives you an interpretation of the source sentence“ (Moorkens et al. 2018:252).

Im Gegensatz zu den oben beschriebenen Untersuchungen führte **Taivalkoski-Shilov (2019a)** eine ethische Komponente in die Forschung rund um LMT bzw. computergestütztes Literaturübersetzen ein. In ihrem Artikel wird u. a. das Konzept von Stimme nach Folkart (1996) im Kontext der (L)MT diskutiert, das als „faisceau de traits textuels qui donne l'impression de pouvoir être reconduit à un seul foyer d'énonciation“²⁶ (Folkart 1996²⁷) definiert und von Kenny & Winters (2020a und b) wieder aufgegriffen wird (siehe unten).

²⁵ BLEU steht für „bilingual evaluation understudy“ (Papineni et al. 2002:311), es wird aber meist nur die Abkürzung benutzt.

²⁶ „Ansammlung von textuellen Eigenschaften, die den Eindruck erweckt, auf eine einzelne Quelle rückführbar zu sein“ (übers. S.K.)

²⁷ Für dieses Werk können keine Seitenzahlen angegeben werden, da es ein Artikel aus einem reinen Online-Journal ist, der nicht als PDF verfügbar ist.

Außerdem fand 2019 im Rahmen des *Machine Translation Summit 2019* ein Workshop mit dem Titel „The Qualities of Literary Machine Translation“ statt (vgl. MT Summit 2019). Im Zuge dieses Workshops wurden sechs Artikel zur LMT eingereicht (siehe Inhaltsverzeichnis in Hadley et al. (2019:vi)), die im Folgenden kurz beschrieben werden.

In **Kuzman et al. (2019)** wurden die Outputs von zwei mit Literatur trainierten NMT-Systemen sowie von Google Translate in der Übersetzungsrichtung Englisch-Slowenisch miteinander verglichen, der zeitliche Postediting-Aufwand gemessen und eine Fehleranalyse durchgeführt (vgl. Kuzman et al. 2019:3f.). Der zeitliche Aufwand für das Postediting war bei allen drei Systemen kleiner als der Aufwand für eine reine Humanübersetzung (vgl. Kuzman et al. 2019:6), während die Fehleranalyse semantische Fehler, Auslassungen, Hinzufügungen, Interpunktionsfehler, falsche Übersetzungen von Präpositionen und Konjunktionen, Veränderungen des Modus, falsche Substantivformen sowie Koreferenzveränderungen in den maschinellen Übersetzungen zu Tage förderte (vgl. Kuzman et al. 2019:7f.). In der automatischen Evaluierung sowie in der Evaluierung von *fluency* und *adequacy* schnitt die Google-Translate-Übersetzung besser ab als die beiden Übersetzungen der an Literatur angepassten NMT-Systeme (vgl. Kuzman et al. 2019:5f.), was im Gegensatz zu den Ergebnissen bzgl. L-SMT von Toral & Way (2015a:128) steht.

Matusov (2019) adaptierte zwei NMT-Systeme in den Übersetzungsrichtungen Englisch-Russisch und Deutsch-Englisch an literarische Texte und verglich deren LMT-Outputs mit jenen von Google Translate. Außerdem entwickelte er eine neue Fehlertypologie²⁸, die speziell an die L-NMT angepasst ist und mit der die Outputs im Rahmen des Experiments evaluiert wurden (vgl. Matusov 2019:10). Für beide Übersetzungsrichtungen ergaben sich bei allen drei Systemen v. a. bzgl. der Semantik, der Konsistenz, der Auflösung von Pronomen, Redewendungen (nur bei Englisch-Russisch sehr problematisch) sowie der Syntax viele Fehler, und es wurden auch häufig Auslassungen festgestellt (vgl. Matusov 2019:16f.). Wie Voigt und Jurafsky (2012:18) und Moorkens et al. (2018:256) ist auch Matusov (2019:17) der Ansicht, dass LMT durch die Berücksichtigung von Mehr-Satz-Kontexten verbessert werden könnte.

In **Ó Murchú (2019)** Artikel wird über ein hybrides MT-System namens *Intergaelic* berichtet, das benutzt wurde, um eine maschinelle Übersetzung eines Romans vom Schottisch-Gälischen ins Irische anzufertigen, die anschließend noch posteditiert wurde (vgl. Ó Murchú 2019:20). Dahingegen wird in **Taivalkoski-Shilov (2019b)** die freie indirekte

²⁸ Auf diese Fehlertypologie wird in Kapitel 4.2 genauer eingegangen.

Rede aus der Perspektive der Narratologie als Beispiel für die Schwierigkeiten diskutiert, die LMT meistern müsste.

Şahin & Gürses (2019) ließen einige Übersetzungs- und Dolmetschstudierende einen Ausschnitt eines Romans vom Englischen ins Türkische übersetzen bzw. eine von Google Translate erstellte Übersetzung eines anderen Ausschnitts desselben Romans posteditieren (vgl. Şahin & Gürses 2019:26). Ein Vergleich der resultierenden Übersetzungen zeigte, dass der MT-Output näher bei der Struktur des Ausgangstextes war als die Humanübersetzungen und dass die Studierenden in den meisten Fällen den MT-Output übernahmen (vgl. Şahin & Gürses 2019:29). Wie in Moorkens et al. (2018:255) bevorzugte die Mehrheit der Teilnehmenden die reine Humanübersetzung gegenüber dem Postediting von LMT-Output, v. a. aufgrund der Schwierigkeiten von Google Translate mit komplexen Sätzen (vgl. Şahin & Gürses 2019:29), die auch schon von Toral & Sánchez-Cartagena (2017:1069) und Toral et al. (2018:10) angemerkt wurden.

Tezcan et al. (2019) verwendeten ebenfalls Google Translate, um einen Ausschnitt eines Romans von Agatha Christie vom Englischen ins Niederländische zu übersetzen. Den daraus resultierenden Output unterzogen sie einer Fehleranalyse anhand der von ihnen an die L-NMT angepassten SCATE-Fehlertypologie (nach Tezcan et al. 2018), die sich in die Bereiche *FLUENCY* und *ACCURACY*²⁹ einteilen lässt, und verglichen ihn mit der Humanübersetzung des Romans (vgl. Tezcan et al. 2019:40). Die Fehleranalyse förderte im Bereich *FLUENCY* v. a. Kohärenzprobleme zu Tage, im Bereich *ACCURACY* war die Fehlübersetzung die am häufigsten vertretene Fehlerart. Insgesamt gab es mehr *FLUENCY*- als *ACCURACY*-Fehler (vgl. Tezcan et al. 2019:45). Die Messung der lexikalischen Vielfalt ergab in drei von fünf Fällen eine höhere lexikalische Vielfalt für den NMT-Output als für die Humanübersetzung und den Ausgangstext, was allerdings an der Zielsprache (Niederländisch) sowie an den Fehlern im NMT-Output liegen könnte (vgl. Tezcan et al. 2019:46). Schließlich zeigte der Vergleich mit der Humanübersetzung wie bei Şahin & Gürses (2019:29), dass das NMT-System die syntaktische Struktur des Ausgangstextes nahezu beibehielt, was zu potentiell nicht natürlicher Sprachverwendung führen könnte (vgl. Tezcan et al. 2019:47f.), wie auch schon der in Besacier (2014:393) zitierte Humanübersetzer anmerkte.

²⁹ Die beiden Bereiche *FLUENCY* und *ACCURACY* der adaptierten SCATE-Fehlertypologie (SCATE steht für „Smart Computer-aided Translation Environment“ (Tezcan et al. 2018:221)) wurden aus Tezcan et al. (2019) direkt übernommen, deswegen sind sie kursiv und unterstrichen (siehe Kapitel 4.2 für weitere Erklärungen diesbezüglich und generell über die Typologie). Sie ähneln den Kategorien *fluency* und *adequacy*, so wie sie in Fußnote 21 definiert wurden, allerdings liegt der Fokus nicht auf Sprache und Inhalt, sondern darauf, ob Fehler ohne (*FLUENCY*) oder nur mit dem Ausgangstext (*ACCURACY*) in der Übersetzung erkennbar sind (vgl. Tezcan et al. 2019:41).

In einer an Tezcan et al. (2019) anschließenden Studie ließen **Fonteyne et al. (2020)** die maschinelle Übersetzung des kompletten Romans von Agatha Christie (wieder in der Übersetzungsrichtung Englisch-Niederländisch) mithilfe der in Tezcan et al. (2019) verwendeten Fehlertypologie annotieren. 44 % der Sätze enthielten keine Fehler (vgl. Fonteyne et al. 2020:3790), was im Vergleich zu etwa 20 % für die Übersetzungsrichtung Englisch-Russisch und ca. 29 % für Deutsch-Englisch (vgl. Matusov 2019:15f.) ein deutlich höherer Wert ist. Wie auch in Tezcan et al. (2019) zeigte die Fehleranalyse insgesamt mehr *FLUENCY*- als *ACCURACY*-Fehler auf (vgl. Fonteyne et al. 2020:3794), wobei Fehlübersetzungen (*ACCURACY*) und Kohärenzfehler sowie Fehler im Bereich Stil und Register (*FLUENCY*) am häufigsten vertreten waren (vgl. Fonteyne et al. 2020:3790). Fehlübersetzungen hingen meist mit der Semantik und mit Mehrwortwendungen zusammen, Kohärenzfehler betrafen v. a. Logikprobleme sowie das Tempus von Verben, bei Stil und Register wurden überwiegend holprige Konstruktionen als Fehler markiert (vgl. Fonteyne et al. 2020:3795). Ein Vergleich der Fehleranalyse mit Ergebnissen von Fehleranalysen, die an NMT-Output von allgemeinsprachlichen Texten (also nicht von Literatur) durchgeführt wurden, zeigte außerdem, dass bei der L-NMT die Wortreihenfolge ein größeres Problem als bei der MT von nicht-literarischen Texten zu sein scheint (vgl. Fonteyne et al. 2020:3796).

Ebenfalls in der Übersetzungsrichtung Englisch-Niederländisch wurde von **Webster et al. (2020a)** erstmals DeepL für die LMT-Forschung verwendet. In dieser Studie wurde auch die in Tezcan et al. (2019) an L-NMT angepasste SCATE-Fehlertypologie zur Fehleranalyse verwendet, und die NMT-Outputs (von DeepL und Google Translate) wurden zusätzlich dazu mit den Ausgangstexten und den Humanübersetzungen in Bezug auf ihre lexikalische Vielfalt, ihre lokale Kohäsion, ihre Syntax und ihren Stil verglichen (vgl. Webster et al. 2020a:1f.). Insgesamt waren in dieser Studie nur ca. 22 % aller Sätze in den Outputs von beiden NMT-Systemen fehlerfrei, wobei DeepL weniger Fehler machte als Google Translate. Bei beiden Systemen war allerdings eine erhöhte Fehleranzahl zu bemerken, je länger die Ausgangssätze waren (vgl. Webster et al. 2020a:10), was u. a. die Ergebnisse von Toral & Sánchez-Cartagena (2017:1069) und Toral et al. (2018:10) widerspiegelt. Die Fehlerarten, die am häufigsten vorkamen, waren wieder aus den Bereichen Fehlübersetzung, Kohärenz sowie Stil und Register (vgl. Webster et al. 2020a:10). Die lexikalische Vielfalt ebenso wie die lokale Kohäsion der NMT-Übersetzungen war niedriger als jene der Humanübersetzungen (vgl. Webster et al. 2020a:12ff.), und wie in Şahin & Gürses (2019:29) und Tezcan et al. (2019:47f.) ähnelte auch hier die Syntax der maschinellen Übersetzungen jener der Ausgangstexte (vgl. Webster et al. 2020a:15). Außerdem unterschied

sich der Stil der maschinellen Übersetzungen stark vom Stil der Humanübersetzungen, die beiden NMT-Outputs ähnelten sich aber untereinander (vgl. Webster et al. 2020a:17).

Um den Stil oder, genauer gesagt, um die von Taivalkoski-Shilov (2019a) in die Diskussion rund um die LMT eingeführte Stimme der Übersetzer_innen ging es auch in der zweiten Untersuchung aus dem Jahr 2020, in der DeepL verwendet wurde: **Kenny & Winters (2020a)** führten eine Studie mit einem bekannten Humanübersetzer durch, bei der der Übersetzer eine von DeepL produzierte Übersetzung vom Englischen ins Deutsche eines Ausgangstextes, den er einige Jahre zuvor schon rein human übersetzt hatte, posteditierte und seine Postedition auch kommentierte (vgl. Kenny & Winters 2020a:124). DeepL wurde u. a. deswegen ausgewählt, weil der Humanübersetzer der Meinung war, dieses NMT-System habe eine ausreichend gute Qualität erreicht, um für Literaturübersetzer_innen interessant zu sein (vgl. Kenny & Winters 2020a:133). Durch einen genauen Vergleich der fertigen Postedition mitsamt den Kommentaren des Übersetzers mit der reinen Humanübersetzung stellten Kenny & Winters (2020a:138) fest, dass die Stimme des Übersetzers in der Postedition durchaus bemerkbar war, allerdings nicht unbedingt dieselben Wörter oder Satzteile wie in der Humanübersetzung betraf. Insgesamt aber war die Stimme des Übersetzers in der Postedition weniger sichtbar (vgl. Kenny & Winters 2020a:144).

Da Silva & Costa (2020) beschäftigten sich in ihrem Artikel ebenfalls mit den ethischen Aspekten der LMT, allerdings führten sie kein Experiment oder dgl. durch. Stattdessen warfen sie, ausgehend von der Prämisse, dass das Literaturübersetzen eine Kunstform ist, eine Reihe von Fragen auf, um auf die möglichen ethischen Konsequenzen einer Benutzung von MT durch Literaturübersetzer_innen aufmerksam zu machen.

Toral, Oliver González et al. (2020) gingen des Thema anders an und präsentierten, basierend auf Toral & Way (2018), die Ergebnisse ihrer Untersuchungen zur LMT mit verschiedenen MT-Systemen, nämlich u. a. einem NMT-System mit einem RNN und einem anderen mit einer Transformer-Architektur. Dabei wurden für das Training und für die darauffolgende Übersetzung und Evaluierung dieselben Datensets benutzt wie in Toral & Way (2018) (vgl. Toral, Oliver González et al. 2020:282f.). Die Evaluierung mit BLEU sowie die menschliche Evaluierung durch ein Ranking ergaben deutlich bessere Ergebnisse für das Transformer-System als für die anderen MT-Systeme (vgl. Toral, Oliver González et al. 2020:283, 286). Bei einer anschließenden Evaluierung von einzelnen Sätzen im Hinblick auf ein Postediting, bei dem allerdings nur grammatikalische Richtigkeit und nicht Literarizität das Ziel war, wurden ca. 47 % der Sätze des Transformer-Systems als fehlerlos eingestuft, verglichen mit ca. 27 % für das RNN-System und ca. 14 % für Google Translate, was im

Gegensatz zu den Ergebnisse von Kuzman et al. (2019:5f.) zeigt, dass an Literatur angepasste Systeme, seien sie RNN- oder Transformer-Systeme, besser abschneiden als nicht angepasste (vgl. Toral, Oliver González et al. 2020:288).

Constantine (2020) beschrieb in seinem Artikel die kleinen Experimente, die er mit Texten von Voltaire und Google Translate in der Übersetzungsrichtung Französisch-Englisch durchgeführt hatte. Er schätzte die maschinellen Übersetzungen als „surprisingly competent“ (Constantine 2020:473) ein, fand aber dennoch Fehler, die sich v. a. auf den Stil, die Natürlichkeit der Sprachverwendung und den Umstand bezogen, dass das NMT-System den Ausgangstext sowie die Übersetzung, die es produziert, nicht „verstehen“ und daher z. B. Pronomen teilweise nicht korrekt auflösen konnte (vgl. Constantine 2020:474f.). Auch archaische Wörter oder Phrasen, die heutzutage unüblich sind, stellten das NMT-System vor Probleme (vgl. Constantine 2020:476).

In der Übersetzungsrichtung Englisch-Arabisch benutzten **Omar und Gomaa (2020)** Google Translate und ein weiteres, frei verfügbares MT-System, um Übersetzungen des ersten *Harry-Potter*-Bandes und einer Kurzgeschichte von Edgar Allen Poe anzufertigen, die dann mit den Humanübersetzungen verglichen und einer manuellen Evaluierung unterzogen wurden. Fehler ergaben sich v. a. in Bezug auf kontext- und nutzungsbezogene Formulierungen, sprachabhängige Kodierung von Informationen, semantische und syntaktische Ambiguität, lexikalische Lücken, Diskontinuität, Pragmatik und unbekannte Wörter. Außerdem gab es in beiden maschinellen Übersetzungen mehrere Sätze, die nicht übersetzt wurden (vgl. Omar & Gomaa 2020:231).

Mohar et al. (2020) verglichen maschinelle Übersetzungen eines Textes von Margaret Atwood aus dem Englischen ins Slowenische, angefertigt von Google Translate und Translator.eu³⁰, mit Studierendenübersetzungen desselben Textes (vgl. Mohar et al. 2020:126). Die NMT-Systeme hatten Schwierigkeiten mit unkonventionellen oder elliptischen (vgl. Mohar et al. 2020:131) sowie langen Sätzen (vgl. Mohar et al. 2020:134), was teilweise Constantines (2020:476) Ergebnisse widerspiegelt, aber v. a. jene von Toral & Sánchez-Cartagena (2017:1069) und Toral et al. (2018:10). Im Gegensatz zu den Studierenden hatten die NMT-Systeme auch Probleme mit Onomatopöie, mit der Auflösung von pronominalen Anaphern über Satzgrenzen hinweg (vgl. auch Constantine 2020:474f.), mit *present participles*, mit *phrasal verbs* und anderen fixen Wendungen sowie bei der Erkennung von Wortarten (vgl. Mohar et al. 2020:133-138). Die Neologismen der Autorin

³⁰ <https://www.translator.eu/> (Stand: 27.04.2021).

des Ausgangstextes und ihr persönlicher Stil stellten sowohl für die Studierenden als auch für Google Translate und Translator.eu eine Herausforderung dar (vgl. Mohar et al. 2020:138f.).

Im Jahr 2020 gab es außerdem ein virtuelles Expert_innentreffen zur LMT, nämlich das von Guerberof-Arenas & Asimakoulas (2020) organisierte *Creative Translation and Technologies Expert Meeting*³¹. Die Teilnehmenden präsentierten im ersten Teil dieses Treffens ihre Forschungsprojekte, u. a. auch das oben angeführte Werk von Kenny & Winters (2020a) (vgl. Kenny & Winters 2020b). **Toral, Castilho et al. (2020)** präsentierten ebenfalls Werke, die im Rahmen des vorliegenden Forschungsüberblicks schon genannt wurden, nämlich Toral & Way (2018), Moorkens et al. (2018) und Toral et al. (2018), und fassten zusammen, dass Posteditionen von LMT-Output einfacher und normalisierter sind und v. a. in Bezug auf grammatikalische Strukturen mehr ausgangssprachliche Interferenzen aufweisen als reine Humanübersetzungen (vgl. Toral, Castilho et al. 2020:1°44'). Hier noch nicht erwähnte und für die LMT relevante Forschungsprojekte sind jene von Castro (2020) und Oliver González et al. (2020) sowie v. a. von Vieira et al. (2020), die im Folgenden kurz beschrieben werden.

Castro (2020) sprach über Möglichkeiten, wie LMT für die Gesellschaft nützlich sein könnte. Ihr Forschungsprojekt fokussiert genauer gesagt darauf, wie MT die Übersetzung in unterrepräsentierten Sprachkombinationen fördern könnte und neue bzw. unbekannte Autor_innen in andere Sprach- und Kulturkreise (konkret in den britischen Buchmarkt) einbringen könnte (vgl. Castro 2020:0°42', 0°44', 0°48').

Oliver González et al. (2020) berichteten über ein Projekt, das sie auch schon beim *Machine Translation Summit 2019* präsentierten (vgl. Oliver González et al. 2019:vii), nämlich InLéctor³², eine Datenbank, die zweisprachige E-Books (Ausgangstexte und ihre Übersetzungen) anbietet. Dabei werden nicht nur Übersetzungen aus der Public Domain³³ verwendet, sondern auch solche, die von einem an Literatur angepassten NMT-System erstellt wurden. Da diese zweisprachigen E-Books eine Unterstützung dafür sein sollen, Bücher in der Originalsprache zu lesen, wird u. a. erforscht, ob die NMT-Outputs dabei helfen können, komplexe Sätze zu verstehen (vgl. Oliver González et al. 2020:2°39', 2°42').

Schließlich präsentierten **Vieira et al. (2020)** bei dem Expert_innentreffen einen Vergleich von reinen Humanübersetzungen und Posteditionen von LMT-Output vom

³¹ Die Videoaufnahme dieses virtuellen Expert_innentreffens war unter dem in der Bibliographie dieser Arbeit angeführten Link zum Zeitpunkt der Abfassung dieses Kapitels (April und Mai 2021) verfügbar, kurz danach wurde sie allerdings vom Netz genommen.

³² <https://inlector.wordpress.com/> (Stand: 28.05.2021).

³³ Die Übersetzungen aus der Public Domain sind nicht urheberrechtlich geschützt (vgl. Boyle 2008:39).

Englischen ins vereinfachte Chinesisch in Bezug auf die Kreativität der Übersetzungen (vgl. Vieira et al. 2020:2°06'30'', 2°08'). Die Beurteilung der Kreativität wurde von fünf Expert_innen durchgeführt (vgl. Vieira et al. 2020:2°08'), die sich in ihrer Beurteilung nur wenig bis gar nicht einig waren. Aus den Durchschnittswerten ihrer Beurteilungen ergab sich jedoch kein (statistisch relevanter) Unterschied zwischen der Kreativität von reinen Humanübersetzungen und Posteditionen von LMT-Output (vgl. Vieira et al. 2020:2°09'). Folglich schlossen Vieira et al. (2020:2°14'), dass der MT-Output die Übersetzer_innen, die an der Studie teilnahmen, nicht in ihrer Kreativität beschränkte, was im Gegensatz zu den Aussagen der Teilnehmenden in Moorkens et al. (2018:256) steht. Was allerdings mit den Ergebnissen von Moorkens et al. (2018:253) übereinstimmt, war der Umstand, dass die Segmentierung in Absätze die Produktivität der Übersetzer_innen im Vergleich zur Segmentierung in Sätze erhöhte (vgl. Vieira et al. 2020:2°16'30''). In Übereinstimmung mit Toral et al. (2018:10) ergaben die Messungen von Vieira et al. (2020:2°11'33'') weiters, dass die Übersetzer_innen beim Postediting (ob nun mit Absatz- oder Satzsegmentierung) weniger Pausen und Anschläge brauchten als beim reinen Humanübersetzen. Auch der zeitliche Aufwand war beim Postediting geringer, allerdings war hier der Unterschied, im Gegensatz zu Toral et al. (2018:10), statistisch nicht relevant.

Im Anschluss an dieses Expert_innentreffen veröffentlichten **Guerberof-Arenas & Toral (2020)** einen Artikel über Kreativität in verschiedenen Übersetzungsmodalitäten, nämlich LMT, Postediting von LMT-Output sowie reine Humanübersetzung, und die Auswirkungen dieser Modalitäten auf das Leseerlebnis. Dieser Aspekt der LMT wurde laut Guerberof-Arenas & Toral (2020:256) noch nicht ausreichend erforscht, obwohl bei der literarischen Übersetzung in den meisten Fällen das Leseerlebnis des Originals beibehalten werden soll. Das in der Studie benutzte MT-System war an Literatur angepasst, und das Postediting erfolgte in einem speziellen Tool mit Absatzsegmentierung, um den negativen Effekt der Satzsegmentierung auszugleichen (vgl. Guerberof-Arenas & Toral 2020:260f.), der von Moorkens et al. (2018:253) und Vieira et al. (2020:2°16'30'') dokumentiert wurde. Eine Fehleranalyse der Postedition und der Humanübersetzung mithilfe der DQF-MQM-Fehlertypologie (TAUS o. J.) ergab mehr Accuracy-Fehler in letzterer als in ersterer, während die Situation in Bezug auf Fluency genau anders herum und in Bezug auf Style³⁴ die Anzahl

³⁴ Die zwei DQF-MQM-Fehlerkategorien Accuracy und Fluency (übernommen aus TAUS (o. J.), deswegen kursiv und unterstrichen; siehe Kapitel 4.2) sind den Konzepten *adequacy* und *fluency*, wie sie in Fußnote 21 definiert wurden, sehr ähnlich. Die Kategorie Style beschreibt einen natürlichen, evtl. auch einem Styleguide oder ähnlichen Richtlinien entsprechenden Sprachgebrauch (vgl. TAUS o. J.). Genauer wird auf die DQF-MQM-Fehlertypologie in Kapitel 4.2 eingegangen.

der Fehler in beiden Übersetzungen gleich hoch war (vgl. Guerberof-Arenas & Toral 2020:268). Die Messung der Kreativität anhand der Fehleranzahl und der Anzahl von *creative shifts* (nach Bayer-Hohenwarter 2011) ergab im Gegensatz zu den Ergebnissen von Vieira et al. (2020:2°09') eine höhere Kreativität für die Humanübersetzung als für die Postedition und den nicht posteditierten Output (vgl. Guerberof-Arenas & Toral 2020:270). Auch das Leseerlebnis wurde von der Übersetzungsmodalität beeinflusst, allerdings gab es für keine der Variablen, mit denen das Leseerlebnis gemessen wurde, statistisch relevante Unterschiede zwischen der Humanübersetzung und der Postedition, während der nicht posteditierte LMT-Output eindeutig die schlechtesten Ergebnisse diesbezüglich erzielte (vgl. Guerberof-Arenas & Toral 2020:276).

Im Jahr 2021 fand die *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*)³⁵ statt, bei der ebenfalls über den oben erwähnten Artikel von Guerberof-Arenas & Toral (2020) gesprochen wurde. Darüber hinaus wurden die vorläufigen Ergebnisse einer weiterführenden Studie präsentiert, die mit derselben Methode zu demselben Thema, allerdings mit einem kreativeren und daher schwierigeren Ausgangstext aus einem anderen Genre (Science-Fiction) und zwei Zielsprachen (Katalanisch und Niederländisch) anstatt nur ersterer durchgeführt wurde (vgl. **Guerberof-Arenas & Toral 2021**). Die vorläufigen Ergebnisse unterschieden sich signifikant von jenen, die in Guerberof-Arenas & Toral (2020) präsentiert wurden, insofern als in beiden Zielsprachen v. a. die Anzahl der Fehler sowie der als sehr positiv bewerteten Stellen und der *creative shifts* in der reinen Humanübersetzung respektive sehr viel niedriger bzw. (sehr viel) höher als in der Postedition war (vgl. Guerberof-Arenas & Toral 2021).

Bei der *CALT2021@Swansea* wurde außerdem von **Shuyin (2021)** ein Forschungsprojekt zum Vergleich von Humanübersetzungen sowie Posteditionen von maschinell erstellten Übersetzungen von Xianxia-Romanen präsentiert. Diese speziellen Web-Romane aus dem Fantasy-Genre eignen sich v. a. deswegen gut für MT, weil spezifische, immer wiederkehrende Terminologie verwendet wird. Die Ergebnisse des Projekts zeigten zwar eine höhere allgemeine Qualität bei den Humanübersetzungen, aber auch eine aus verschiedenen Gründen bestehende Toleranz bei den Leser_innen solcher Romane für einige Fehler in den Posteditionen (vgl. Shuyin 2021).

³⁵ Die Abstracts zu den im Rahmen der Konferenz gehaltenen Präsentationen finden sich unter folgendem Link: <https://docs.google.com/document/d/e/2PACX-1vRWAAgOpMCWqGniXJAULzCra1e4We2U7IQj53e7qbKNIGHvjLGmHrwlsgbqCJCQIwJ11tztjQVmXMu6/pub> (Stand: 15.11.2021).

Zur MT von Gedichten gab es zwei Beiträge: **Hadley et al. (2021)** sprachen u. a. über die maschinelle Übersetzung von arabischen Gedichten durch ein speziell trainiertes MT-System. Die Analyse der Outputs zeigte eine sehr niedrige Qualität der Übersetzungen auf, was v. a. an der eher archaischen Sprache (wie in Constantine (2020:476)) sowie dem ständigen Wechsel zwischen Dialekt- und Standardsprache in den Ausgangstexten lag (vgl. Hadley et al. 2021).

Van de Cruys (2021) berichtete über sein System zur automatischen Generierung und Übersetzung von Gedichten, bei dem wie in Genzel et al. (2010), Greene et al. (2010) und Ghazvininejad et al. (2018) Beschränkungen bzgl. Reim und dgl. eingestellt werden können. Er schloss aus den für ein MT-System relativ freien Übersetzungen, dass diese „[a]dditional constraints can trigger a certain creative inclination“ (Van de Cruys 2021) beim verwendeten NMT-System, da es eine Übersetzung produzieren muss, die sich inhaltlich nicht übermäßig weit vom Ausgangstext entfernt, in der aber gleichzeitig auch die oben erwähnten Einschränkungen beachtet werden müssen (vgl. Van de Cruys 2021).

Schließlich wurde bei der *CALT2021@Swansea* auch über Experimente mit DeepL berichtet, zunächst von **Schmidt (2021)**. Dieser führte einige Beispiele aus Fallstudien mit DeepL in der Übersetzungsrichtung Deutsch-Englisch an, in denen sich v. a. Dialoge, Verben und onomatopoetische Wörter als problematisch für das MT-System erwiesen (vgl. Schmidt 2021), wobei die letzten zwei problematischen Bereiche (teilweise) mit den Ergebnissen von Mohar et al. (2020:133, 136f.) übereinstimmen. Trotzdem seien die Humanübersetzungen nicht leicht vom MT-Output zu unterscheiden gewesen, obwohl dieser nicht posteditiert worden war (vgl. Schmidt 2021).

Auch in **Kolb (2021)** wurde DeepL als NMT-System verwendet, um eine zu posteditierende Übersetzung ins Deutsche einer englischsprachigen Kurzgeschichte zu produzieren, womit diese Studie nach Kenny & Winters (2020a) erst die zweite ist, die sich mit dieser Übersetzungsrichtung beschäftigt. Kolb (2021) analysierte zunächst den NMT-Output, wobei v. a. Polysemie, ähnlich aussehende Wörter, Register, Redewendungen, Referenz über Satzgrenzen hinaus und Verben problematisch waren, was u. a. mit den Ergebnissen von Matusov (2019:16), Fonteyne et al. (2020:3790, 3795), Mohar et al. (2020:134-137), Webster et al. (2020a:10) und Schmidt (2021) in anderen Sprachen bzw. Übersetzungsrichtungen übereinstimmt. Der Hauptteil von Kolbs Studie fokussiert aber auf den technischen, zeitlichen und kognitiven Aufwand der Übersetzer_innen bzw.

Posteditor_innen sowie auf die *priming effects*³⁶ des MT-Outputs, d. h., inwiefern die Posteditor_innen dazu verleitet werden, den MT-Output einfach zu übernehmen, ohne weiter darüber nachzudenken. Bzgl. des Aufwands stellte sich durchschnittlich ein geringerer technischer und zeitlicher Aufwand bei den Posteditor_innen als bei den Übersetzer_innen heraus, wobei aber eine hohe Variabilität zwischen den Teilnehmenden herrschte. Der kognitive Aufwand schien für die Posteditor_innen allerdings höher zu sein, da sie mit drei Texten umgehen mussten (dem Ausgangstext, der maschinellen Übersetzung sowie ihrer finalen Übersetzung) anstatt nur mit zwei. Außerdem ließen sich beim Postediting deutliche *priming effects* beobachten, d. h., die Posteditionen waren lexikalisch weniger divers und einheitlicher als die Humanübersetzungen (vgl. Kolb 2021), was die Auffassungen der Teilnehmenden in Moorkens et al. (2018:252) sowie die Ergebnisse von Şahin & Gürses (2019:29) und Toral, Castilho et al. (2020:1°44') widerspiegelt.

³⁶ Der Begriff *priming effect* oder *priming* kommt ursprünglich aus der Psychologie und bedeutet dort „an improvement in performance in a perceptual or cognitive task [...] produced by context or prior experience“ (McNamara 2005:3). So z. B. können Wortpaare, die semantisch zusammenhängen, durch sogenanntes *semantic priming* schneller als Wörter (im Gegensatz zu Nicht-Wörtern, also einfachen Buchstabenfolgen) erkannt werden, als semantisch nicht zusammenhängende Wortpaare (vgl. Meyer & Schvaneveldt 1976:229f.; McNamara 2005:3). Auch in Bezug auf das Übersetzen werden *priming effects* untersucht: So fanden etwa Bangalore et al. (2016:235) heraus, dass Übersetzer_innen durch *syntactic priming* dazu verleitet werden, die syntaktische Struktur des Ausgangstextes in der Übersetzung zu reproduzieren. In einem ähnlichen Sinne sind dementsprechend auch die in Kolb (2021) untersuchten *priming effects* zu verstehen.

4 Methode

In diesem Kapitel werden die Methoden erklärt, die für die in der Einleitung skizzierte Untersuchung verwendet wurden. Dazu wird auf das Untersuchungsmaterial eingegangen und auf die Vorgehensweise bei der Entwicklung der Fehlerkategorien. Anschließend werden die Posteditorinnen und das Postediting beschrieben sowie die Durchführung der Ratings. Zum Schluss werden die Methoden behandelt, die für die Auswertung der im Laufe der Studie gesammelten Daten verwendeten wurden.

4.1 Untersuchungsmaterial

Zur Beantwortung der Forschungsfragen wurde verschiedenartiges Untersuchungsmaterial benutzt, das sich aus folgenden Texten zusammensetzt³⁷:

- HPAT: Ausschnitt aus *Harry Potter and the Philosopher's Stone* (Rowling 2010/1997)
- HPMÜ: DeepL-Übersetzung ins Deutsche des HPAT
- HPHÜ: korrespondierender Ausschnitt aus der publizierten Humanübersetzung *Harry Potter und der Stein der Weisen* (Rowling 2013/1998)
- RRAT: Ausschnitt aus *Red Rising* (Brown 2014a)
- RRMÜ: DeepL-Übersetzung ins Deutsche des RRAT
- RRPEo: ohne Fehlerkategorien angefertigte Postedition der RRMÜ
- RRPEm: mit Fehlerkategorien angefertigte Postedition der RRMÜ³⁸
- RRHÜ: korrespondierender Ausschnitt aus der publizierten Humanübersetzung *Red Rising* (Brown 2014b)

Als Ausgangstext für die erste Forschungsfrage dienten die Seiten 233-239 des ersten *Harry-Potter*-Bandes (Auflage aus dem Jahr 2010) der Autorin Joanne K. Rowling. *Harry Potter* wurde für diese Forschungsfrage ausgewählt, da die Verfasserin der vorliegenden Arbeit die Bücher sowohl in englischer als auch in deutscher Sprache sehr gut kennt, was für die Analyse der HPMÜ und der darauf folgenden Entwicklung der Fehlerkategorien hilfreich war. Der entsprechende Ausschnitt wurde gewählt, da er der Verfasserin dieser Masterarbeit

³⁷ Die in der Liste genannten Abkürzungen „HPAT“, „HPHÜ“, „RRAT“ und „RRHÜ“ dienen auch als Quellenangaben, um den Leser_innen bei Zitaten aus diesen Ausschnitten die ständige Wiederholung von „Rowling 2010/1997:xx“ und den anderen vollständigen Angaben zu ersparen. Die gekürzten Quellenangaben nehmen beispielsweise folgende Form an: „(HPAT:xx)“. Werden Informationen aus den Romanen zitiert, die nicht aus diesen Ausschnitten stammen, wird allerdings die vollständige Quellenangabe verwendet.

³⁸ Die zwei DeepL-Übersetzungen (HPMÜ und RRMÜ) befinden sich in Anhang 1, die zwei Posteditionen (RRPEo und RRPEm – „o“ steht für „ohne“, „m“ für „mit Fehlerkategorien“) in Anhang 2 dieser Arbeit.

mit seiner Mischung aus Dialogen, Beschreibungen, einer Passage aus einem fiktiven Sachbuch und der großen Anzahl an In-Universe-Termini³⁹ (IUT) als repräsentativ für die *Harry-Potter*-Bücher erschien. Außerdem war derselbe Ausschnitt schon in einer ähnlichen Analyse im Rahmen der Bachelorarbeit der Verfasserin verwendet worden (vgl. Schuhmann 2020:1), was die Untersuchung zusätzlich erleichterte.

Für die Forschungsfragen 2 und 3 wurde ein Ausschnitt aus einem Roman als Ausgangstext gewählt, der nicht den großen Bekanntheitsgrad von *Harry Potter* erreicht hat, obwohl er in Rezensionen manchmal damit verglichen wird (vgl. z. B. Goodreads 2013a; Yasin 2016; Evershed 2021): *Red Rising* des Autors Pierce Brown. Der Grund dafür bestand in der Annahme, dass sich für einen weniger bekannten Ausgangstext einfacher Teilnehmende finden würden, die nichts von diesem Text wissen und daher in ihrer Anfertigung der Posteditionen (siehe Kapitel 4.3) bzw. in ihrer Qualitätsbewertung (siehe Kapitel 4.4) nicht voreingenommen sind. Ein weiterer Grund für die Wahl von *Red Rising* war die große Zahl an IUT (vgl. z. B. Goodreads 2013b), die laut Auffassung der Verfasserin der vorliegenden Arbeit an die sprachliche Kreativität der *Harry-Potter*-Bücher erinnert, obwohl sie in Goodreads (2013b) als negativ empfunden wird.

Um die Vergleichbarkeit der maschinellen Übersetzung des Ausschnitts aus *Red Rising* und jener des HPAT und damit zumindest die teilweise Anwendbarkeit der mithilfe der HPMÜ entwickelten Fehlerkategorien auf die RRMÜ sicherzustellen, wurde die Auswahl des RRAT mithilfe von drei (statistischen) Kriterien getroffen: der lexikalischen Vielfalt, der durchschnittlichen Satzlänge sowie der Anzahl an IUT. Die Werte dieser Kriterien sind für die zwei Ausgangstexte in Tabelle 1 auf der nächsten Seite ersichtlich.

Das erste Kriterium wurde gewählt, weil eine hohe lexikalische Vielfalt charakteristisch für literarische Texte ist (vgl. Toral & Way 2018:264). Angegeben wird sie üblicherweise mit dem Type-Token-Verhältnis (*type-token ratio*, TTR), das durch die Division der Anzahl an Worttypes durch die Anzahl an Wörtern (Tokens) berechnet wird. Die Worttypes ergeben sich durch eine spezielle Zählweise, bei der Wörter, wenn sie in derselben Form im Text vorkommen, nur einmal gezählt werden. Ein höherer TTR-Wert deutet auf eine höhere lexikalische Vielfalt in einem Text hin, wobei zu beachten ist, dass die Länge des Textes Auswirkungen auf den TTR-Wert hat (vgl. Perkuhn et al. 2012:[2f.]). Für einen Vergleich der TTR-Werte von zwei (oder mehr Texten) sollten diese deshalb annähernd dieselbe Länge

³⁹ „In-Universe-Termini“ bzw. „In-Universe-Terminologie“ ist eine von der Verfasserin der vorliegenden Arbeit kreierte Bezeichnung für Wörter und Mehrwortausdrücke, die von Autor_innen für ihre literarischen Werke erfunden bzw. dort auf ganz spezifische Art und Weise verwendet werden. Für mehr Erklärungen und Beispiele diesbezüglich wird auf Kapitel 5.1 dieser Masterarbeit verwiesen.

haben, weshalb auch die Anzahl an Wörtern der Ausgangstexte (gemessen mit demselben Open-Source-Tool, mit dem auch der TTR-Wert gemessen wurde⁴⁰) in untenstehender Tabelle angegeben ist.

Die durchschnittliche Satzlänge wurde als zweites Kriterium definiert, da sie sich auf die Qualität von NMT-Übersetzungen auswirkt: Diese nimmt ab, je länger die Sätze des jeweiligen Ausgangstextes sind (vgl. Toral & Sánchez-Cartagena 2017:1071). Für die Berechnung der durchschnittlichen Satzlänge in Wörtern wurde die Lesbarkeitsstatistik⁴¹ von *Microsoft Word* 2010 verwendet.

Das dritte Kriterium schließlich ist die Anzahl an IUT, die sich im Ausgangstext befinden. Es wurde gewählt, da IUT, die sehr spezifische, oft auch nicht näher erklärte Konzepte aus der Welt des Ausgangstextes bezeichnen können und häufig autor_innenspezifischen Wortbildungsmustern folgen⁴², Humanübersetzer_innen durchaus Probleme bereiten (vgl. Jentsch 2006:190) und deswegen auch angenommen werden kann, dass sich eine große Anzahl an IUT auf die Qualität einer maschinellen Übersetzung auswirkt. Die IUT wurden in den beiden Ausgangstexten von der Verfasserin der vorliegenden Arbeit identifiziert und gezählt, wobei zu beachten ist, dass die Namen der verschiedenen Gesellschaftsklassen in *Red Rising*, die im RRAT vorkommen (nämlich „Red“, „Orange“, „Gray“, „Obsidian“, „Yellow“ und „Gold“ (RRAT:[26f.]⁴³)), als ein IUT gezählt wurden, da davon ausgegangen werden kann, dass diese nach dem gleichen Schema übersetzt werden (was übrigens auch in der Humanübersetzung der Fall ist (vgl. RRHÜ:[29]⁴⁴)).

Tabelle 1: Vergleichbarkeit der Ausgangstexte

	HPAT	RRAT
Anzahl an Wörtern	1 477	1 459
lexikalische Vielfalt (TTR)	39,13 %	44,35 %
durchschn. Satzlänge in Wörtern	12,8	12
Anzahl an IUT	17	18

⁴⁰ https://github.com/StevenCHowell/type_token_ratio (Stand: 20.03.2021).

⁴¹ In der *Microsoft-Word*-Lesbarkeitsstatistik werden auch noch andere Werte angezeigt, die zwar für die Art von Literatur, die hier behandelt wird (nämlich Unterhaltungsliteratur, Kinder- und Jugendliteratur etc.), grundsätzlich durchaus relevant sein können, aber für die vorliegende Untersuchung nicht von Belang sind. Die Lesbarkeitsstatistik wurde dementsprechend nur zur Ermittlung der durchschnittlichen Satzlänge verwendet.

⁴² Z. B. verwendet Joanne K. Rowling oft Alliterationen und Wortspiele in ihren Kreationen (vgl. Brøndsted & Dollerup 2004:67; Jentsch 2006:197), während Pierce Brown häufig Wörter lateinischen Ursprungs verwendet und ihnen romanspezifische Bedeutungen gibt (etwa „Aureate“ (RRAT:[26])) bzw. auch zusammengesetzte Wörter mit Großschreibung im Wort kreiert (etwa „holoCan“ (RRAT:[22])).

⁴³ Die englische Version von *Red Rising* liegt bei der Verfasserin nur als E-Book (PDF-Datei) vor, daher sind die ungefähren Seitenzahlen laut dem Programm *Adobe Acrobat Reader DC* in eckigen Klammern angegeben.

⁴⁴ Die deutsche Version von *Red Rising* liegt bei der Verfasserin nur als E-Book (EPUB-Datei) vor, daher sind die ungefähren Seitenzahlen laut dem Programm *Adobe Digital Editions* in eckigen Klammern angegeben.

Wie an der Tabelle ersichtlich wird, haben die beiden Ausgangstexte eine annähernd gleiche Wortanzahl und Anzahl an IUT. Die lexikalische Vielfalt ist im HPAT ein wenig niedriger als im RRAT, was durch die geringere durchschnittliche Satzlänge in letzterem ausgeglichen wird. Selbstverständlich bestehen allerdings trotz dieser Ähnlichkeiten auch zahlreiche Unterschiede zwischen den beiden Ausgangstexten: So ist der HPAT in *British English*, der RRAT in *American English* geschrieben. Des Weiteren ist die Autorin des HPAT eine Frau, der Autor des RRAT ein Mann, und die ursprünglichen Publikationsjahre der beiden Bücher (1997 und 2014) liegen 17 Jahre auseinander. Zudem unterscheiden sich die Altersempfehlungen für die beiden Romane deutlich: Thalia (o. J.a) empfiehlt den ersten *Harry-Potter*-Band für neun- bis elfjährige Kinder, Common Sense Media (o. J.a) für achtjährige und ältere, während die Altersempfehlungen für *Red Rising* bei 14-17 Jahren (vgl. Thalia o. J.b) bzw. bei 15 Jahren und älter liegen (vgl. Common Sense Media o. J.b). Dies hängt wahrscheinlich einerseits mit dem Plot zusammen (jener von *Red Rising* unterscheidet sich doch deutlich von jenem des ersten *Harry-Potter*-Bandes), andererseits aber vermutlich auch mit den in *Red Rising* häufig verwendeten, teilweise recht expliziten Schimpfwörtern. Da Schimpfwörter für Humanübersetzer_innen (und damit wahrscheinlich auch für MT) eine Herausforderung darstellen (vgl. Hjort 2009⁴⁵; Azura et al. 2019:43) und im HPAT keinerlei Schimpfwörter enthalten sind, wurde in Bezug auf *Red Rising* neben den oben angeführten Kriterien ebenfalls darauf geachtet, einen Ausschnitt zu wählen, der solche Ausdrücke nicht enthält. Die anderen oben genannten Unterschiede zwischen den Romanen waren bei der Auswahl des RRAT allerdings nicht relevant, da DeepL nicht zwischen *British* oder *American English* als Ausgangssprache unterscheidet⁴⁶ und alle anderen Unterschiede nicht per se sprach- oder textbezogen sind, sondern sich auf den äußeren Kontext der Romane beziehen und der DeepL-Übersetzer deswegen nicht dazwischen differenziert bzw. differenzieren kann.

Die oben genannten Überlegungen führten schließlich zur definitiven Auswahl der Seiten [22-27] der in der Bibliographie angeführten Version von *Red Rising* als Ausgangstext. Von diesem sowie vom HPAT wurden die maschinellen Übersetzungen am 22.03.2021 mithilfe der Dokumentenübersetzung von DeepL angefertigt. An diesen Übersetzungen wurden danach keine Änderungen mehr vorgenommen.

Für die erste Forschungsfrage wurde noch der entsprechende Ausschnitt (Seiten 235-241) der publizierten Humanübersetzung des ersten *Harry-Potter*-Bandes herangezogen,

⁴⁵ Für dieses Werk können keine Seitenzahlen angegeben werden, da es ein Artikel aus einem reinen Online-Journal ist, der nicht als PDF verfügbar ist.

⁴⁶ Anders wäre dies, wenn Englisch die Zielsprache der Übersetzung wäre, denn bei DeepL kann in diesem Fall seit August 2020 zwischen *British* und *American English* gewählt werden (vgl. DeepL 2020b).

während für die zweite und dritte Forschungsfrage neben dem korrespondierenden Ausschnitt der offiziellen Humanübersetzung von *Red Rising* (Seiten [24-29]) noch zwei unterschiedliche Posteditionen der RRMÜ benötigt wurden. Wie diese Posteditionen angefertigt wurden, wird in Kapitel 4.3 erklärt.

4.2 Entwicklung der Fehlerkategorien

Um die Fehlerkategorien zu entwickeln, die im Zuge des Posteditings verwendet (siehe Kapitel 4.3) und bzgl. derer mehrere Ratings durchgeführt (siehe Kapitel 4.4) wurden, waren mehrere Schritte notwendig: Zunächst wurde die HPMÜ einem genauen Vergleich mit dem HPAT und der HPHÜ unterzogen, um die Fehler in der HPMÜ zu finden. Zur Erleichterung dieses Prozesses wurde auch auf die in der Forschungsliteratur erhobenen problematischen Bereiche für L-NMT zurückgegriffen, die im Forschungsüberblick (Kapitel 3) schon genannt und in untenstehender Tabelle noch einmal überblicksartig angeführt sind. Die Reihenfolge der problematischen Bereiche, die aus den jeweiligen Werken übernommen wurde, spielt in der Tabelle keine Rolle. Außerdem muss angemerkt werden, dass sich die aufgelisteten problematischen Bereiche sowohl auf Eigenschaften des Ausgangstextes beziehen können (z. B. Mehrdeutigkeit), die für das MT-System problematisch waren, oder auf tatsächliche Fehler in der maschinellen Übersetzung (z. B. Interpunktion).

Tabelle 2: Problematische Bereiche der L-NMT aus der Forschungsliteratur

Werk(e)	Übersetzungsrichtung(en)	Problematische Bereiche
Moorkens et al. (2018:254)	Englisch-Katalanisch	<ul style="list-style-type: none"> – Höflichkeit – kontextadäquates Vokabular – figurative Sprache – kulturelle Elemente
Humblé (2019:48)	Englisch-Portugiesisch	<ul style="list-style-type: none"> – Polysemie – Genus und Numerus – Fälle, in denen die Zielsprache expliziter als die Ausgangssprache sein muss – Charakteristika von Gedichten (z. B. Alliterationen⁴⁷)

⁴⁷ Obwohl es sich beim HPAT nicht um ein Gedicht handelt, ist dieser problematische Bereich dennoch von Bedeutung, denn J. K. Rowling verwendet, wie oben erwähnt, v. a. Alliterationen sehr häufig (vgl. Brøndsted & Dollerup 2004:67).

Kuzman et al. (2019:7)	Englisch-Slowenisch	<ul style="list-style-type: none"> – Interpunktion – Präpositionen und Konjunktionen – Modus – Substantive – Koreferenz – Semantik (v. a. bzgl. fixer Wendungen und Ambiguität) – Auslassungen und Hinzufügungen
Matusov (2019:16f.)	<ul style="list-style-type: none"> – Englisch-Russisch – Deutsch-Englisch 	<ul style="list-style-type: none"> – Semantik – Konsistenz – Auflösung von Pronomen – fixe Wendungen – Auslassungen – Syntax
Tezcan et al. (2019:45) Fonteyne et al. (2020:3794f.) Webster et al. (2020a:10f.)	Englisch-Niederländisch	<ul style="list-style-type: none"> – lange Sätze – Kohärenz (v. a. Logik, Tempus) – Fehlübersetzungen (v. a. bzgl. Mehrwortwendungen und der Bedeutung von Wörtern) – Stil und Register (v. a. kein flüssiger Stil)
Constantine (2020:474-477)	Französisch-Englisch	<ul style="list-style-type: none"> – Stil – natürlicher und flüssiger Sprachgebrauch – Auflösung von Pronomen und dgl. – archaisches Vokabular, unübliche Wendungen – Mehrdeutigkeit
Omar & Gomaa (2020:231)	Englisch-Arabisch	<ul style="list-style-type: none"> – kontext- und nutzungsbezogene Formulierungen – sprachabhängige Kodierung von Informationen – Ambiguität (semantisch, syntaktisch) – lexikalische Lücken – Diskontinuität – Pragmatik – unbekannte Wörter – Auslassungen

Mohar et al. (2020:131-139)	Englisch-Slowenisch	<ul style="list-style-type: none"> – Syntax (unkonventionell, elliptisch, komplex) – Tempus – Onomatopöie – Auflösung von pronominalen Anaphern über Satzgrenzen hinweg – <i>present participles</i> – <i>phrasal verbs</i> und andere fixe Wendungen – Wörter, deren Wortart nur aus dem Kontext ersichtlich ist, nicht an ihrer Form – Neologismen – persönlicher Stil des _der Autors _Autorin des Ausgangstextes
Hadley et al. (2021)	Arabisch-Englisch	<ul style="list-style-type: none"> – archaische Sprache – Wechsel zwischen Dialekt- und Standardsprache
Schmidt (2021)	Deutsch-Englisch	<ul style="list-style-type: none"> – Dialoge – Verben – Onomatopöie
Kolb (2021)	Englisch-Deutsch	<ul style="list-style-type: none"> – Polysemie – ähnlich aussehende Wörter – Register – fixe Wendungen – Referenz über Satzgrenzen hinaus – Verben

Alle durch den Vergleich mit dem HPAT und der HPHÜ und mithilfe der oben genannten Problembereiche in der HPMÜ gefundenen Fehler wurden in einem separaten Dokument gemeinsam mit den entsprechenden Stellen aus dem HPAT und der HPHÜ und einer Erklärung des jeweiligen Fehlers aufgelistet. Die Erklärungen wurden unter Rückgriff auf die in der Bibliographie unter „Tertiärliteratur“ angeführten Werke (v. a. einsprachige Wörterbücher wie *Duden* (2021) und *Lexico.com* (2021a) und andere Werke zur deutschen und englischen Sprache) und, für inhaltliche Argumentationen, auf die *Harry-Potter*-Bücher selbst verfasst und dienten als Zwischenschritt für die nächste Phase in der Entwicklung der Fehlerkategorien. Wie solche Erklärungen aussehen, wird in Kapitel 5.1 ersichtlich, da sie dort verwendet werden, um die entwickelten Fehlerkategorien anhand von Beispielen aus der HPMÜ zu erläutern.

Wie an obigem Absatz ersichtlich wird, ist „Fehler“ in der vorliegenden Arbeit als problematische Stelle einer Übersetzung zu verstehen, deren Problemcharakter durch eine möglichst objektive und durch anerkannte Quellen belegte bzw. durch inhaltliche Argumentationen gestützte Erklärung bestätigt werden kann. Selbstverständlich wurden bei der Analyse der HPMÜ aber auch Problemstellen identifiziert, die nicht auf diese Art und Weise als Fehler bestätigt werden konnten. In solchen Fällen wurde der Problemcharakter durch Erklärungen erläutert, die nach dem eigenen Ermessen der Verfasserin der vorliegenden Arbeit verfasst wurden. Dementsprechend spielte bei der Erklärung dieser Problemstellen (die demnach weiterhin als „Problemstellen“ und nicht als „Fehler“ bezeichnet werden⁴⁸) der Subjektivitätsfaktor eine große Rolle (vgl. z. B. auch Stymne & Ahrenberg (2012) über die Subjektivität bei Fehleranalysen von MT-Output), was bei der Präsentation und Anwendung der aus ihnen entwickelten Fehlerkategorien stets im Hinterkopf behalten werden muss.

Jedenfalls wurden im nächsten Schritt mithilfe der in der ersten Phase verfassten Erklärungen der Problemstellen und Fehler eigene Fehlerkategorien entwickelt. Dazu wurde in den Erklärungen nach Gemeinsamkeiten Ausschau gehalten, durch die sich die Problemstellen und Fehler gruppieren ließen: So wurden z. B. Fehler, die sich auf das Tempus von Verben beziehen, in eine andere Gruppe eingeteilt als Interpunktionsfehler. Bei sehr großen Fehlergruppen wurde außerdem überprüft, ob Unterkategorisierungen möglich waren.

Auf diese Art und Weise wurden Schritt für Schritt speziell auf dieses Forschungsprojekt zugeschnittene Fehlerkategorien entwickelt, wobei die unten aufgezählten und im Folgenden beschriebenen Fehlertypologien, die in Forschungsprojekten zur L-NMT verwendet wurden, als Hilfestellung dienten:

- Fehlertypologie für L-NMT (Matusov 2019)
- DQF-MQM-Fehlertypologie (TAUS o. J.; in Guerberof-Arenas & Toral (2020) verwendet)
- SCATE-Fehlertypologie für L-NMT (Tezcan et al. 2019; außerdem in Fonteyne et al. (2020) und Webster et al. (2020a) verwendet)

Für seinen im Rahmen des *Machine Translation Summit 2019* eingereichten Artikel entwickelte Matusov (2019) eine eigene Fehlertypologie für L-NMT-Output, der im Gegensatz zu den anderen beiden Fehlertypologien keine Hierarchie zugrunde liegt und die aus zwölf Fehlerkategorien besteht (vgl. Matusov 2019:14f.). Die ersten beiden Kategorien

⁴⁸ Diese Unterscheidung in Problemstellen und Fehler ähnelt jener der Multidimensional Quality Metric (MQM; siehe unten), in der zunächst ein „issue“ in einer Übersetzung bzw. einem Text gefunden wird, das danach erst als „error“ bestätigt werden muss/kann (vgl. Lommel, Görög et al. 2015:7).

beziehen sich auf die Semantik (wobei es bzgl. Redewendungen eine eigene Kategorie gibt; siehe unten): Bei einem severe meaning error⁴⁹ hat der jeweilige Satzteil, in dem der Fehler vorkommt, eine Bedeutung, die im gegebenen Kontext falsch und somit für die Leser_innen der Übersetzung ohne den Ausgangstext nicht verständlich wäre, während bei einem minor meaning error die Bedeutung des Ausgangstextes zwar grundsätzlich in der Übersetzung wiedergegeben wird, aber leichte (z. B. stilistische oder andersartige) Abweichungen vorliegen, wobei die Übersetzung trotzdem ohne den Ausgangstext verstanden werden kann (vgl. Matusov 2019:14).

In die Kategorie unknown word or segmentation error fallen Fehler bei der Übersetzung von unbekannten oder seltenen Wörtern, welche meist einfach zu erkennen sind, z. B. weil nur ein Teil des unbekannten oder seltenen Wortes übersetzt wurde. Dies passiert, weil die meisten NMT-Systeme solche Wörter in bekannte Subwörter (nach Sennrich et al. 2016) einteilen, um sie zu übersetzen, was aber nicht automatisch zu einer richtigen Übersetzung führt, denn die Subwörter können im Kontext des unbekannten oder seltenen Wortes eine andere Bedeutung haben als bei einer alleinigen Verwendung. Außerdem bezieht sich diese Kategorie auf Segmentierungsfehler im Falle von Ausgangssprachen ohne Wortsegmentierung (z. B. Chinesisch) (vgl. Matusov 2019:14).

Consistency/term translation error bezeichnet eine inkonsistente Übersetzung von Namen, Fachtermini und anderen Wörtern, die eine einheitliche Übersetzung erfordert hätten, während sich ein pronoun resolution error auf die falsche Auflösung von Pronomen in der Übersetzung bezieht. Letzteres kommt relativ häufig vor, denn Pronomen können oft nur durch Berücksichtigung des vorigen Satzes oder gar eines größeren Kontextes richtig aufgelöst und damit übersetzt werden. Zur nächsten Kategorie, dem locution error, zählen wortwörtliche Übersetzungen von fixen Wendungen, die eine andere Bedeutung haben als die Wendungen im Ausgangstext, oder Übersetzungen durch zielsprachliche Wendungen, die eine andere oder zwar dieselbe/eine ähnliche Bedeutung haben, aber falsch formuliert sind (vgl. Matusov 2019:14).

Die folgenden drei Kategorien werden häufig in der MT-Forschung verwendet: Ein omission error bezeichnet Auslassungen und ein insertion error Hinzufügungen in der Übersetzung, während sich ein repetition error auf die Wiederholung von einzelnen Wörtern oder

⁴⁹ Um den Leser_innen die ständige Wiederholung der Klammerausdrücke für die oben genannten Werke zu ersparen, indiziert die gleichzeitige Unterstreichung und Kursivsetzung der Namen der Fehlerkategorien, dass diese direkt aus dem jeweiligen Werk, aus dem die Fehlerkategorien stammen, übernommen wurden.

Satzteilen in der Übersetzung bezieht, wobei auch Wiederholungen in verschiedenen Wortarten unter diese Kategorie fallen (z. B. „hölzernes Holz“) (vgl. Matusov 2019:14f.).

Bei den Syntaxfehlern gibt es wie bei den *meaning error[s]* zwei Gruppen, nämlich *severe syntax error[s]* und *minor syntax error[s]*. Bei den schwerwiegenden Syntaxfehlern ist die Struktur des betreffenden Satzes in der Übersetzung nicht korrekt, sodass der Satz unverständlich ist. Im Gegensatz zum *severe meaning error* kann ein *severe syntax error* nicht leicht einem einzelnen Wort oder Satzteil zugeordnet werden. Ein *minor syntax error*, die vorletzte Fehlerart, wäre hingegen ein syntaktischer oder morphologischer Fehler, der einfach und ohne große Veränderungen am übersetzten Satz korrigiert werden könnte. Die letzte Kategorie schließlich heißt *tone/register error* und beschreibt die falsche Verwendung von Höflichkeitsformen und Anreden sowie den Gebrauch von stilistisch nicht angemessenen Wörtern oder Formulierungen und dgl. (vgl. Matusov 2019:15).

Die zweite Fehlertypologie, die bei der Entwicklung der Fehlerkategorien für die vorliegende Masterarbeit half, ist die DQF-MQM-Fehlertypologie, die aus einer Harmonisierung von zwei anderen Fehlertypologien entstand, nämlich des ursprünglichen Dynamic Quality Framework (DQF) von TAUS und der Multidimensional Quality Metric (MQM) des QTLaunchPad-Projekts (vgl. Lommel, Görög et al. 2015). Wie oben schon angedeutet und in Abb. 3 ersichtlich, ist die DQF-MQM-Typologie hierarchisch aufgebaut.

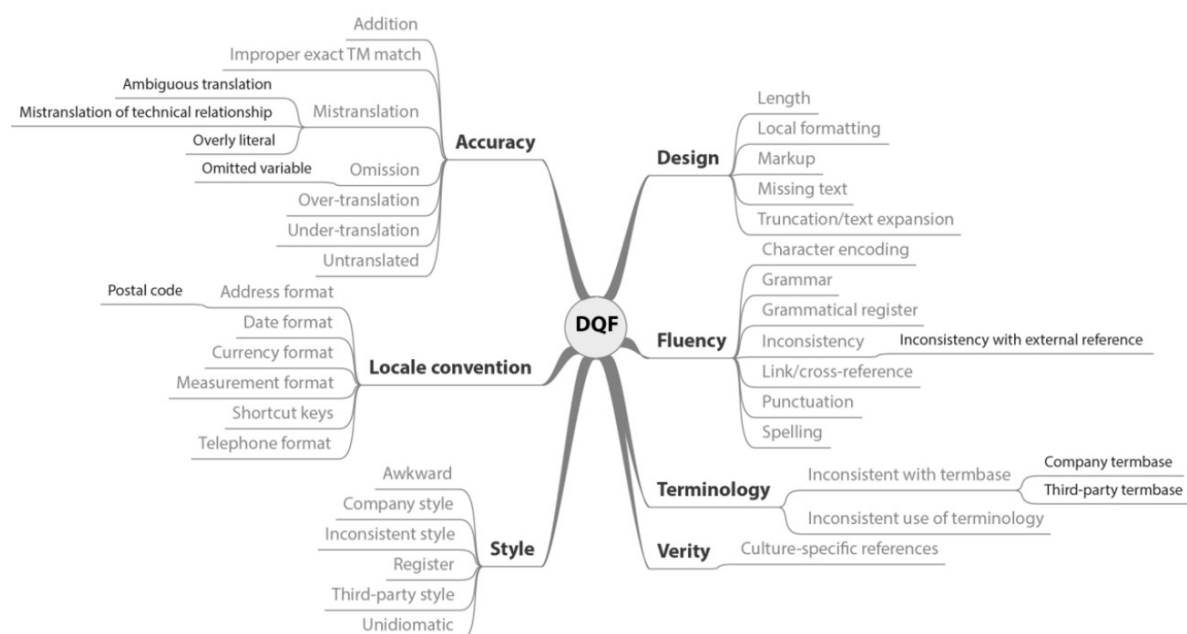


Abb. 3: DQF-MQM-Fehlertypologie (Lommel, Görög et al. 2015:13; farblich adaptiert durch die Verfasserin)⁵⁰

⁵⁰ In der Abbildung lautet der Name der Typologie nur DQF, da sie von MQM aus betrachtet als Untergruppe von MQM fungiert (vgl. Lommel, Görög et al. 2015:12).

In Guerberof-Arenas & Toral (2020:264) wurde diese Fehlertypologie verwendet, um die Akzeptabilität der maschinellen Übersetzung zu eruieren, allerdings wurde dabei auf die (in der Abbildung nicht gezeigten) Schweregrade, die den einzelnen Fehlern im Rahmen der DQF-MQM-Typologie zugeordnet werden können (vgl. TAUS o. J.), nicht eingegangen. Deshalb und weil Schweregrade in den im Rahmen dieser Mastarbeit entwickelten Fehlerkategorien keine Rolle spielen, werden die Schweregrade der DQF-MQM-Fehlertypologie auch hier nicht behandelt. Außerdem beschränkt sich die Beschreibung auf die ersten zwei Niveaus der hierarchisch aufgebauten Fehlertypologie⁵¹ (in Abb. 3 in schwarzer und fett hervorgehobener sowie grauer Schrift angeführt), um den Rahmen der vorliegenden Arbeit nicht zu sprengen.

Die DQF-MQM-Kategorie Accuracy stimmt relativ genau mit *adequacy* überein, wie sie in Kapitel 2 definiert wurde, und beschreibt Fehler in der Übersetzung, die dazu führen, dass diese den Ausgangstext nicht genau wiedergibt. Darunter fallen neben der auch in Matusovs (2019) Fehlertypologie präsenten Hinzufügung (Addition) und Auslassung (Omission) auch Untranslated (in der Übersetzung befinden sich Teile des ausgangssprachlichen Materials, für die eine Übersetzung notwendig gewesen wäre) und Mistranslation (der Inhalt des Ausgangstextes wird nicht genau wiedergegeben). Außerdem gibt es die Unterkategorien Over-translation und Under-translation, die sich auf Fälle beziehen, in denen die Übersetzung spezifischer oder weniger spezifisch ist als der Ausgangstext (vgl. TAUS o. J.), wobei dies natürlich auch den Eigenheiten der betroffenen Sprachen geschuldet sein kann und dann nicht als Fehler gewertet werden sollte (vgl. Lommel, Burchardt et al. 2015). Die letzte Unterkategorie auf dieser Ebene heißt Improper exact TM match und betrifft aus einem TM übernommene Exact Matches, die im gegebenen Kontext nicht korrekt sind (vgl. TAUS o. J.).

Im Gegensatz zu Accuracy ist die DQF-MQM-Kategorie Fluency nicht nur auf Übersetzungen, sondern auch auf Originaltexte anwendbar. Zudem ist sie weiter gefasst als ihr Pendant in Kapitel 2 (*fluency*), sie umfasst nämlich nicht nur sprachliche Probleme, sondern auch Probleme bzgl. der Form des Textes, nämlich interne Inkonsistenz (Inconsistency), inkonsistente Links (Link/cross-reference) oder falsche Zeichensätze (Character encoding). Zu den sprachlichen Fehlerarten in dieser Kategorie zählen Punctuation (falsch gesetzte oder fehlende Interpunktionszeichen), Spelling (Rechtschreibfehler) und Grammar, also Fehler, die sich auf Grammatik oder Syntax eines Textes beziehen und nicht unter Punctuation oder

⁵¹ Für Erklärungen der Kategorien auf dem dritten Niveau (in Abb. 3 in schwarzer, nicht hervorgehobener Schrift angeführt) wird auf Lommel, Burchardt et al. (2015) verwiesen.

Spelling fallen. Die letzte Kategorie unter Fluency heißt Grammatical register und beschreibt Fehler bzgl. des Sprachregisters im Text (vgl. TAUS o. J.).

Die dritte DQF-MQM-Kategorie ist Terminology. Sie bezieht sich darauf, dass Fachtermini anders als im Fachgebiet erwartet oder nicht den Vorgaben entsprechend übersetzt werden. Solche Vorgaben können z. B. mittels einer Terminologiedatenbank erfolgen – wird diese nicht beachtet, fällt der Fehler unter die Unterkategorie Inconsistent with termbase. Bei inkonsistenter Verwendung von einzelnen Termini im Text wird die Kategorie Inconsistent use of terminology benutzt (vgl. TAUS o. J.).

Die Kategorie Style bezieht sich, wie ihr Name schon sagt, auf stilistische Probleme. Darunter fallen die Kategorien Awkward und Unidiomatic, die für plumpe bzw. nicht natürlich anmutende Textstellen benutzt werden. Neben der Kategorie Inconsistent style (inkonsistenter Stil) gibt es außerdem noch die Fehlerarten Company style und Third-party style, die sich auf die Nichteinhaltung von unternehmensspezifischen Styleguides bzw. Styleguides von Dritten beziehen, die hätten beachtet werden sollen (vgl. TAUS o. J.).

Bei der nächsten Kategorie, nämlich Design, geht es um Fehler, die sich nicht auf den Text per se beziehen, sondern auf das Aussehen dieses Textes: Bei der Unterkategorie Length weicht die Länge des Ausgangstextes stark von jener des Zieltextes ab oder umgekehrt, bei Local formatting gibt es örtlich begrenzte Formatierungsprobleme (die sich also nicht auf den gesamten Text beziehen). Missing text ist nicht mit der Kategorie Omission (siehe oben) zu verwechseln, denn bei ersterer ist die Übersetzung komplett, doch in der finalen, fertig formatierten Version fehlen Teile dieser Übersetzung oder sind nicht sichtbar. Die Kategorie Markup schließlich bezieht sich auf inkorrekt gesetzte Tags, während Truncation/text expansion problematische Textexpansionen beschreibt, die z. B. dazu führen können, dass übersetzter Text über die Ränder von Schaltflächen einer Bedienungsoberfläche steht (vgl. TAUS o. J.).

Unter die Kategorie Locale convention fallen Fehler in der Übersetzung, in denen gegen die formalen Zwänge oder Konventionen des Zielmarktes bzgl. Adressen (Address format), Datumsangaben (Date format), Währungen (Currency format), Maßeinheiten (Measurement format) oder Telefonnummern (Telephone format) verstoßen wird. Außerdem gibt es die Unterkategorie Shortcut key, die für den Zielmarkt unlogisch vergebene Tastenkombinationen beschreibt (vgl. TAUS o. J.).

Die vorletzte Kategorie heißt Verity und bezieht sich auf „statements that contradict the world of the text“ (TAUS o. J.), d. h. solche, die mit der realen Welt nicht vereinbar sind. Culture-specific reference ist eine Unterkategorie davon und beschreibt kulturspezifische

Referenzen im Text, die für das Zielpublikum unverständlich sind. Schließlich gibt es noch die Kategorie *Other* (in Abb. 3 nicht zu sehen), die für alle Fehler zu verwenden ist, die nicht in eine der oben beschriebenen Kategorien passen (vgl. TAUS o. J.).

Abschließend wird noch die dritte der für die Entwicklung der Fehlerkategorien relevanten Fehlertypologien beschrieben, nämlich die an die L-NMT angepasste SCATE-Fehlertypologie von Tezcan et al. (2019). Die ursprüngliche SCATE-Fehlertypologie wurde nicht für LMT entwickelt, sondern für die Erstellung eines MT-Fehlerkorpus (vgl. Tezcan et al. 2018:221). Nach einer kleinen Anpassung an NMT durch Van Brussel et al. (2018:3800) adaptierten Tezcan et al. (2019:41f.) diese Typologie für die L-NMT, indem sie neben anderen kleineren Veränderungen die Unterkategorien *coherence* und *style & register* zur Hauptkategorie *FLUENCY* hinzufügten. Dadurch entstand die hierarchisch aufgebaute Fehlertypologie, die überblicksartig in Abb. 4 zu sehen ist und im Folgenden beschrieben wird, wobei allerdings nur auf die Hauptkategorien und ihre direkten Unterkategorien eingegangen wird.⁵²

FLUENCY	ACCURACY
<ul style="list-style-type: none"> • coherence <ul style="list-style-type: none"> ○ logical problem ○ non-existing word ○ cultural reference ○ discourse marker ○ co-reference ○ inconsistency ○ verb tense • lexicon <ul style="list-style-type: none"> ○ lexical choice ○ wrong preposition • grammar & syntax <ul style="list-style-type: none"> ○ agreement ○ verb form ○ word order ○ extra word(s) ○ missing word(s) • style & register <ul style="list-style-type: none"> ○ disfluency ○ repetition ○ register ○ untranslated • spelling • other 	<ul style="list-style-type: none"> • mistranslation <ul style="list-style-type: none"> ○ multiword ○ word sense ○ semantically unrelated ○ part-of-speech ○ partially translated ○ other • do not translate • untranslated • addition • omission • capitalisation & punctuation • other

Abb. 4: An L-NMT angepasste SCATE-Fehlertypologie (Tezcan et al. 2019:42)

Unter die Hauptkategorie *FLUENCY* fallen Fehler, die in der Übersetzung ohne Rückgriff auf den Ausgangstext entdeckt werden können (vgl. Tezcan et al. 2019:41). Zur Kategorie *coherence*, die aufgrund der Ergebnisse von Voigt & Jurafsky (2012:18) und Moorkens et al. (2018:256) bzgl. Mehr-Satz-Kontexten bei der LMT hinzugefügt wurde, zählen Fehler, durch die die Übersetzung (in einem größeren Kontext) unlogisch oder unverständlich wird (vgl.

⁵² Für Erklärungen der Unter-Unterkategorien wird auf Webster et al. (2020b) verwiesen.

Tezcan et al. 2019:41f.). Die Kategorie lexicon bezieht sich auf die Wortwahl und grammar & syntax auf die grammatikalischen und strukturellen Regeln der Zielsprache (vgl. Webster et al. 2020b:[3f.]). Die ebenfalls neu hinzugefügte Kategorie style & register beschreibt, wie der Name schon sagt, Probleme bzgl. des in der Übersetzung verwendeten Stils und Registers sowie der Lesbarkeit des Textes (vgl. Tezcan et al. 2019:42). Unter die Kategorie spelling fallen Fehler bzgl. der Orthographie (vgl. Webster et al. 2020b:[5]), während other für FLUENCY-Fehler verwendet wird, die in keine der unter FLUENCY angeführten Kategorien passen (vgl. Tezcan et al. 2018:228).

Zu der Hauptkategorie ACCURACY zählen Fehler, die nur durch einen Vergleich von Ausgangstext und Übersetzung in Letzterer entdeckt werden können (vgl. Tezcan et al. 2019:41). Unter die Subkategorie mistranslation fallen schlichtweg falsche Übersetzungen, z. B. weil die Bedeutung des Ausgangstextes nicht wiedergegeben wird oder Wörter nur teilweise übersetzt wurden (vgl. Webster et al. 2020b:[6-9]). Die nächsten beiden Unterkategorien – do not translate und untranslated – sind jeweils das Gegenteil der anderen: Sie beziehen sich auf Fälle, in denen Ausgangstextmaterial, das unverändert in die Übersetzung hätte übernommen werden sollen, übersetzt wurde bzw. in denen Ausgangstextmaterial, das hätte übersetzt werden sollen, unverändert in die Übersetzung übernommen wurde (vgl. Webster et al. 2020b:[9]). Die Kategorien addition und omission bezeichnen wie in der DQF-MQM-Typologie Hinzufügungen und Auslassungen, während sich die Kategorie capitalisation & punctuation⁵³ auf Fehler bei der Groß-/Kleinschreibung und der Interpunktion bezieht (vgl. Webster et al. 2020b:[10]). Die Kategorie other nimmt in Bezug auf ACCURACY dieselbe Funktion ein wie other in FLUENCY (vgl. Tezcan et al. 2018:228).

Mithilfe dieser drei Fehlertypologien wurden also aus den in der HPMÜ gefundenen Fehlern gemeinsam mit ihren Erklärungen die Fehlerkategorien entwickelt, die von einer Posteditorin während des Posteditings verwendet und bzgl. derer mehrere Ratings durchgeführt wurden. Wie das Postediting genau abläuft, wird im nächsten Kapitel erläutert, der Ablauf der Ratings in Kapitel 4.4. Die Fehlerkategorien selbst werden hingegen in Kapitel 5.1 als Antwort auf FF1 vorgestellt und erklärt.

⁵³ Die Kategorie capitalisation & punctuation ist eine Unterkategorie von ACCURACY und nicht FLUENCY aufgrund des folgenden Umstandes, der der Verfasserin dieser Arbeit von Tezcan (2021) erklärt wurde: Es gibt Interpunktions- und Groß-/Kleinschreibungsfehler, die nur mithilfe des Ausgangstextes entdeckt werden können (ACCURACY) (z. B. ein in der Übersetzung normal geschriebenes Wort, das im Ausgangstext in Dauergroß geschrieben ist, würde ohne diesen Ausgangstext nicht als Fehler erkannt werden). Außerdem gibt es Interpunktions- und Groß-/Kleinschreibungsfehler, denen Verstöße gegen die Regeln der Zielsprache zugrunde liegen und die daher ohne den Ausgangstext entdeckt werden können (FLUENCY). Letztere sind allerdings (immer noch laut Tezcan (2021)) bei NMT-Output selten. Deswegen gehört capitalisation & punctuation in der an die L-NMT angepassten SCATE-Fehlertypologie zu ACCURACY (vgl. Tezcan 2021).

4.3 Posteditorinnen und Durchführung des Posteditings

In diesem Kapitel wird der Ablauf des für die vorliegende Studie notwendigen Posteditings beschrieben. Dazu wird zunächst auf die Posteditorinnen selbst eingegangen, danach auf die Bedingungen, unter denen das Postediting stattfand.

Die beiden zu analysierenden Posteditionen der RRMÜ wurden von zwei Studierenden des Masterstudiums Translation der Universität Wien angefertigt. Die Entscheidung, Studierende als Posteditor_innen zu beauftragen, wurde getroffen, da diese für die Verfasserin der vorliegenden Arbeit leichter zu kontaktieren waren als professionelle, berufstätige Literaturübersetzer_innen. Selbstverständlich haben Studierende weniger Erfahrung als berufstätige Übersetzer_innen, jedoch gibt es, wie in Kapitel 2 erläutert, „no conclusive data when it comes to experience“ (Guerberof-Arenas 2020:346) beim Postediting und der Qualität der resultierenden Posteditionen. Darüber hinaus stellten Moorkens & O’Brien (2015:80) sogar fest, dass sich Studierende intensiver als professionelle Übersetzer_innen mit dem Zweck der jeweiligen Studie, an der sie teilnehmen, auseinandersetzen und der Anwendung von Technologie positiver gegenüber stehen. Aus diesen Gründen wurde der Umstand, dass das für die vorliegende Masterarbeit notwendige Postediting von Studierenden durchgeführt wird, als akzeptabel erachtet.

Die konkrete Auswahl der Posteditor_innen wurde mithilfe von zwei Professor_innen aus dem Schwerpunkt „Übersetzen in Literatur – Medien – Kunst“ des Masterstudiums Translation der Universität Wien getroffen. Jene Professor_innen schlugen der Verfasserin der vorliegenden Arbeit auf Anfrage Studierende vor, die für die Teilnahme an der Studie in Frage kämen, nämlich solche mit hohen sprachlichen Fähigkeiten (um zu vermeiden, dass durch mangelnde sprachliche Kompetenzen die Qualität der Posteditionen nachteilig beeinträchtigt wird und die Ergebnisse der Ratings so verfälscht werden).

Die vier vorgeschlagenen Studierenden wurden per E-Mail kontaktiert, um ihre Bereitschaft für eine Teilnahme festzustellen. Waren sie an einer Teilnahme interessiert (was bei allen vier der Fall war), wurden sie gebeten, einige Fragen zu beantworten, um herauszufinden, ob sie folgende drei Kriterien erfüllten:

1. Sie sind Studierende des Masterstudiums Translation der Universität Wien, und zwar im Schwerpunkt „Übersetzen in Literatur – Medien – Kunst“. (Da es in der Studie um eine literarische maschinelle Übersetzung geht, wurden Studierende dieses Schwerpunkts als am geeignetsten erachtet, eine Postedition eines solchen Textes anzufertigen.)

2. Ihre Muttersprache ist Deutsch (die Zielsprache der Übersetzung), und eine ihrer Arbeitssprachen im Studium ist Englisch (die Ausgangssprache der Übersetzung). (Dieses Kriterium wurde bestimmt, um die übliche Situation einer Literaturübersetzung aus der Fremd- in die Muttersprache (vgl. Kolb 2016:147f.) abzubilden und sicherzugehen, dass die Posteditor_innen zumindest im Rahmen ihres Studiums schon einmal einen literarischen Text vom Englischen ins Deutsche übersetzt hatten.)
3. Sie kennen weder das englische Original noch die deutsche Übersetzung von *Red Rising*. (Dieses Kriterium sollte die Unvoreingenommenheit der Posteditor_innen während des Posteditings sicherstellen.)

Eine_r der von den Professor_innen der Universität Wien vorgeschlagenen möglichen Posteditor_innen schied durch seine_ihre Antworten auf die in oben erwähneter E-Mail gestellten Fragen aus, da er_sie nicht im gewünschten Schwerpunkt studierte. Die anderen drei möglichen Posteditor_innen erfüllten alle drei Kriterien und beantworteten eine Reihe von weiteren Fragen, die sich auf ihren Studienfortschritt, ihre Erfahrung im Literaturübersetzen außerhalb der Universität sowie ihre Erfahrung mit DeepL und Postediting bezogen. Auf Grundlage der von den möglichen Posteditor_innen zur Verfügung gestellten Informationen wurden schließlich die zwei (weiblichen) Posteditorinnen ausgewählt, die die ähnlichsten Angaben gemacht hatten: Beide waren in der Schlussphase ihres Masterstudiums und hatten alle vier im gegebenen Schwerpunkt vorgesehenen Lehrveranstaltungen zum Thema des literarischen und medialen Übersetzens (vgl. Universität Wien 2017:9f.) zum Zeitpunkt des Posteditings absolviert. Erfahrung im Literaturübersetzen außerhalb des Universitätsbereiches hatten beide keine, allerdings hat Posteditorin 1 an einer Fortbildung zum Literaturübersetzen aus dem Norwegischen teilgenommen. Beide Posteditorinnen hatten Erfahrung in anderen Bereichen der Sprachindustrie: Posteditorin 1 hat zweieinhalb Jahre lang Videospielübersetzungen aus dem Japanischen und Englischen editiert, während Posteditorin 2 als Trainerin für Deutsch als Fremdsprache gearbeitet sowie ein Praktikum für Videountertitelung absolviert hat. Keine der beiden hatte professionelle Erfahrung im Posteditieren oder im Umgang mit DeepL, aber beide gaben an, auf informelle Art und Weise bzw. in universitären Lehrveranstaltungen den Übersetzungsdienst verwendet und die übersetzten Passagen danach bearbeitet zu haben.

Die Ähnlichkeit der Posteditorinnen und damit ihre Vergleichbarkeit waren ein wichtiger Faktor bei der Auswahl der Teilnehmenden, da dadurch auch von einer Vergleichbarkeit der von ihnen angefertigten Posteditionen ausgegangen werden kann. Selbstverständlich wird aber der persönliche Stil der Posteditorinnen (vgl. Saldanha

2005:178ff.), der auch in Posteditionen sichtbar ist (vgl. Kenny & Winters 2020a:144), trotzdem unterschiedlich sein, was die Bewertung in den anschließenden Ratings beeinflussen könnte. Um dies in ihrer eigenen Studie zu verhindern, wendeten Guerberof-Arenas & Toral (2020:260) die Methode des Aggregierens von Übersetzungen an, d. h., die finalen, zu bewertenden Texte stammten aus der Feder von zwei Übersetzer_innen (die erste Hälfte von Übersetzer_in A, die zweite von Übersetzer_in B und umgekehrt). Dies war allerdings bei der vorliegenden Studie nicht möglich, da der Einfluss der Benutzung von Fehlerkategorien während des Posteditings im Zentrum stand, d. h., es wurde eine Postedition benötigt, die komplett ohne Fehlerkategorien angefertigt wurde, und eine andere mit. Wäre die Methode des Aggregierens angewandt worden, hätte dementsprechend eine Posteditorin die erste Hälfte der RRMÜ ohne Fehlerkategorien posteditieren müssen, dann die zweite Hälfte mit. Dies wäre machbar gewesen, jedoch hätte die zweite Posteditorin zuerst die zweite Hälfte der RRMÜ ohne Fehlerkategorien posteditieren müssen und dann erst die erste Hälfte mit, was vermutlich sehr verwirrend für die betreffende Posteditorin gewesen wäre. Darüber hinaus hätte auch durch die Methode des Aggregierens der Posteditionen der persönliche Stil von Übersetzer_innen nicht aus der Gleichung genommen werden können, da auch die offizielle Humanübersetzung des RRAT bei den Ratings bewertet wurde. Aus diesen Gründen wurde die Methode des Aggregierens in der vorliegenden Studie nicht angewandt. Folglich musste bei der Interpretation und Diskussion der Ergebnisse der Ratings (siehe Kapitel 5.2.2 und 6) beachtet werden, dass der persönliche Stil des Humanübersetzers und der zwei Posteditorinnen die Bewertungen beeinflusst haben könnte.

Bevor allerdings die Ratings stattfinden konnten, musste das Postediting durchgeführt werden, wobei nach der Methode von Kenny & Winters (2020a:133) vorgegangen wurde: Es wurde ein statisches Postediting, also ein Postediting nach Anfertigung der RRMÜ (siehe Kapitel 2), durchgeführt. An Richtlinien wie jene von TAUS (Massardo et al. 2016) mussten sich die Posteditorinnen dabei nicht halten. Der fiktive Auftrag lautete, ebenfalls in Anlehnung an Kenny & Winters (2020a:133), dass die RRMÜ so posteditiert werden soll, dass die fertige Postedition den üblichen Qualitätsstandards der Posteditorinnen entspricht (vollständiges Postediting) und ihrer Ansicht nach bei einem Verlag eingereicht werden könnte.

Wie in Kenny & Winters (2020a:133) bekamen die Posteditorinnen den RRAT sowie die RRMÜ im *Microsoft-Word*-Format, damit sie in ihrer natürlichen Arbeitsumgebung arbeiten konnten. Während des Posteditings sollten sie den „Änderungen nachverfolgen“-Modus verwenden, und auch Kommentare (z. B. zur Erklärung ihrer Übersetzungslösungen)

konnten eingesetzt werden, waren aber nicht notwendig. Es wurde dementsprechend weder ein spezielles Postediting-Tool noch eine andere Software zur Aufzeichnung der Tastatur- und Mausbenutzung o. Ä. verwendet, da der Zweck der vorliegenden Studie nicht darin bestand, den Postediting-Aufwand zu messen oder den Prozess des Posteditings zu untersuchen.

Neben dem RRAT und der RRMÜ wurde den Posteditorinnen als zusätzliche Ressource der gesamte (englischsprachige) Roman *Red Rising* im PDF-Format zur Verfügung gestellt, damit sie den zu posteditierenden Ausschnitt in Kontext setzen (vgl. Kenny & Winters 2020a:133) und evtl. unbekannte Wörter und Konzepte aus *Red Rising* während des Posteditings im Buch nachschlagen konnten. Darüber hinaus durften die Posteditorinnen alle möglichen Wörterbücher und sonstigen Ressourcen benutzen, die sie für einen realen Auftrag ebenfalls verwenden würden, allerdings sollten sie weder auf die existierende Humanübersetzung von *Red Rising* noch auf darauf basierende deutschsprachige romanspezifische (Fan-)Websites oder dgl. zurückgreifen. Der Grund für diese Einschränkung bestand darin, dass die Posteditionen im Rating auch mit der Humanübersetzung verglichen wurden und es demnach kontraproduktiv gewesen wäre, wenn sich die Posteditorinnen daran orientiert hätten.

All diese Grundinformationen wurden den Posteditorinnen im Rahmen einer von der Verfasserin der vorliegenden Arbeit organisierten Videokonferenz mündlich übermittelt sowie schriftlich zur Verfügung gestellt. Während dieser Videokonferenz wurde mit den Posteditorinnen auch eine Deadline vereinbart, die für alle drei Beteiligten akzeptabel war. Außerdem unterschrieben die Posteditorinnen eine Einverständniserklärung zur Bestätigung, dass sie freiwillig an diesem Teil der Studie teilnahmen, sich an die Vorgaben halten würden und über ihre Rechte als Teilnehmende informiert waren. Danach konnte Posteditorin 1 die Videokonferenz verlassen und mit dem Postediting beginnen, während Posteditorin 2 (die zuvor per Zufallsverfahren durch Urbaniak & Plous (o. J.) ausgewählt worden war) erklärt wurde, dass sie für ihre Arbeit eine zusätzliche Hilfestellung bekommen würde, nämlich die Fehlerkategorien, die im Rahmen der vorliegenden Masterarbeit entwickelt wurden. Diese wurden der entsprechenden Posteditorin mithilfe einer gekürzten und adaptierten Version des Kapitels 5.1 erklärt und präsentiert; diese Version des Kapitels wurde der Posteditorin auch für spätere Referenz zur Verfügung gestellt.

Die Fehlerkategorien sollten der Posteditorin insofern als Hilfestellung dienen, als sie sie darauf aufmerksam machten, welche Fehler in der RRMÜ zu erwarten waren und wie damit umgegangen werden könnte (vgl. Guerberof-Arenas 2020:337). Sie konnten

systematisch verwendet werden (z. B. konnten zuerst nur Fehler/Problemstellen der ersten Fehlerkategorie in der RRMÜ korrigiert werden, danach Fehler/Problemstellen der zweiten Kategorie usw.) oder auch auf andere Art und Weise – dies stand der Posteditorin frei. Allerdings wurde sie darauf hingewiesen, dass in der RRMÜ durchaus auch Fehler/Problemstellen auftauchen könnten, die nicht in eine der Kategorien passten, und sie sich daher nicht nur auf die Fehlerkategorien konzentrieren, sondern sie lediglich als zusätzliche Ressource betrachten sollte. Durch diesen Hinweis sollte vermieden werden, dass die Posteditorin durch ihren Fokus auf die Fehlerkategorien womöglich andere Fehler übersieht.

Nach dieser Einführung wurde die Videokonferenz beendet, und auch Posteditorin 2 konnte mit ihrer Arbeit beginnen. Bei Fragen konnten sich beide Posteditorinnen jederzeit an die Verfasserin der vorliegenden Arbeit wenden, es wurden allerdings keine übersetzungstechnischen Fragen beantwortet, sondern nur solche, die mit den Fehlerkategorien per se oder mit dem (technischen) Ablauf des Posteditings und den einzuhaltenden Vorgaben zusammenhingen. Nach Abgabe ihrer Posteditionen war die Teilnahme der Posteditorinnen an der Studie beendet. Die fertigen Übersetzungen sowie die Humanübersetzung wurden dann, wie schon öfters erwähnt, einer Qualitätsbewertung durch andere Studierende unterzogen, was im folgenden Kapitel genauer erklärt wird.

4.4 Durchführung der Ratings

Die Ratings wurden wiederum von Studierenden im Schwerpunkt „Übersetzen in Literatur – Medien – Kunst“ des Masterstudiums Translation der Universität Wien durchgeführt. Im Gegensatz zu den Posteditorinnen (siehe Kapitel 4.3) wurden die Teilnehmenden an den Ratings allerdings nicht von der Verfasserin der vorliegenden Masterarbeit ausgewählt. Stattdessen schickten drei Professor_innen, die Lehrveranstaltungen im oben genannten Schwerpunkt betreuten, jeweils eine von der Verfasserin dieser Arbeit formulierte Einladung zur Teilnahme an die Studierenden ihrer Lehrveranstaltungen aus. All jene Studierenden, die die Teilnahme Kriterien zum Zeitpunkt der Studie erfüllten, konnten an den Ratings teilnehmen. Diese Kriterien waren dieselben, die auch für die Posteditorinnen definiert wurden, nämlich ein Studium im oben genannten Schwerpunkt, Deutsch als Mutter- und Englisch als Arbeitssprache im Studium sowie der Umstand, dass *Red Rising* den Teilnehmenden weder auf Englisch noch auf Deutsch bekannt war. Für eine ausführlichere Beschreibung der Kriterien sowie der Gründe, warum sie gewählt wurden, wird auf Kapitel 4.3 verwiesen.

Auf die ausgesandte Einladung zur Teilnahme hin meldeten sich lediglich zwei Studierende. Da dies zu wenige Teilnehmende gewesen wären, um aussagekräftige Ergebnisse zu erhalten, wurde dieser Teil der vorliegenden Masterarbeitsstudie zweimal abgehalten: einmal im Rahmen einer Videokonferenz mit den zwei Studierenden, die sich auf die Aussendung hin meldeten, und einmal im Rahmen einer Einheit der Lehrveranstaltung „Literarisches Schreiben und Lektorieren“, die im oben genannten Schwerpunkt zu absolvieren ist (vgl. Universität Wien 2017:9). Alle Studierenden, die an der Lehrveranstaltung teilnahmen, mussten die Ratings als Teil ihrer Verpflichtungen im Rahmen dieser Lehrveranstaltung durchführen, es wurden aber nur jene Ergebnisse ausgewertet, die auch von Studierenden stammten, die die oben genannten Kriterien erfüllten. Zwischen der Vorgehensweise im Rahmen der Videokonferenz und jener im Rahmen der Lehrveranstaltung (die im Folgenden beschrieben wird) bestand kein methodischer Unterschied, deswegen wurden die im jeweiligen Rahmen durchgeführten Ratings als gleichwertig behandelt und ausgewertet.

Im Rahmen der Konferenz bzw. der Lehrveranstaltungseinheit wurde den Teilnehmenden (im Folgenden „Rater_innen“ genannt) zunächst das Ziel der Studie erklärt, nämlich die Bewertung der Qualität von Posteditionen von LMT-Output. Das Ziel wurde dabei absichtlich nicht genauer formuliert, um die Rater_innen in ihrer Bewertung nicht zu beeinflussen. Aus demselben Grund und um zu vermeiden, dass die Rater_innen während der Ratings versuchen, die Humanübersetzung zu identifizieren (was bei einer geglückten Identifizierung aufgrund von Vorurteilen gegenüber MT, wie sie z. B. von Toral & Way (2015a:129) erwähnt werden, automatisch zu einer besseren Bewertung der Humanübersetzung und damit einer Verfälschung der Ergebnisse hätte führen können), wurde eine zweite Vorsichtsmaßnahme getroffen: In Anlehnung an Toral & Way (2015a:129; 2018:279) wussten die Rater_innen nicht, dass eine der Übersetzungen keine Postedition einer maschinellen Übersetzung, sondern eine Humanübersetzung war. Die Übersetzungen wurden lediglich in zufälliger Reihenfolge (die zuvor durch Mohr (2004) determiniert worden war) unter den Namen „Übersetzung A/B/C“ präsentiert, wobei Übersetzung A der RRPE_m, Übersetzung B der RRHÜ und Übersetzung C der RRPE_o entsprach⁵⁴.

Nach der Präsentation des Ziels der Studie wurden den Rater_innen der unten beschriebene Ablauf sowie weitere Informationen (z. B. über ihre Rechte im Rahmen der

⁵⁴ In den beiden zu bewertenden Posteditionen waren die Änderungen der Posteditorinnen selbstverständlich für die Rater_innen nicht sichtbar, d. h., die Verfasserin dieser Arbeit nahm vor den Ratings alle Änderungen der Posteditorinnen in ihren *Word*-Dokumenten an.

Studie, die Verarbeitung ihrer Daten, die zu benutzenden Ressourcen etc.) dargelegt. Die Rater_innen wurden im Anschluss gebeten, eine Einverständniserklärung zu unterschreiben, in der sie ihr Einverständnis zur Verarbeitung ihrer Daten gaben und bestätigten, alle genannten Informationen verstanden zu haben, (freiwillig) an der Studie teilzunehmen und sich an die Vorgaben zu halten.

Nach Abschluss dieser administrativen Einführung wurde den Rater_innen die Vorgehensweise bei den Ratings erklärt, die sich nach den in Toral & Way (2018:279) angewandten Regeln richtete: An die drei Übersetzungen konnten die Ränge 1-3 vergeben werden, wobei 1 der beste und 3 der schlechteste Rang war. Wurde die Qualität von mehreren Texten als gleichwertig eingeschätzt, wurde derselbe Rang vergeben, wobei der höchste zur Verfügung stehende Rang gewählt werden sollte (wurden also zwei der drei Texte als gleichwertig und besser eingeschätzt als der dritte Text, bekamen beide Rang 1, der dritte Rang 2). Die in Toral & Way (2018:279) ebenfalls verwendeten Tipps zur Einschätzung der Qualität der Übersetzungen wurden allerdings nicht angewandt, da die Ratings im Rahmen dieser Studie nicht nur allgemein stattfanden, sondern auch in Bezug auf die entwickelten Fehlerkategorien (siehe unten), wodurch die in den Tipps genannten Aspekte abgedeckt wurden.⁵⁵

Vor den eigentlichen Ratings wurde noch der Roman *Red Rising* kurz vorgestellt, wobei der Autor (Pierce Brown), das Publikationsjahr (2014) sowie das Genre des Buches („Science Fiction & Fantasy“ laut Penguin Random House (o. J.)) genannt wurden. Außerdem wurde den Rater_innen eine kurze (englischsprachige) Inhaltsangabe (entnommen aus Penguin Random House (o. J.)) zur Verfügung gestellt, damit sie wussten, worum es im Roman geht, und den zu bewertenden Ausschnitt kontextualisieren konnten.

Mit diesen Informationen wurde nun das erste allgemeine Rating von den Rater_innen durchgeführt. Dazu sollten sie die vier Texte (den RRAT und die drei Übersetzungen) lesen und danach an die Übersetzungen die Ränge 1-3 unter Befolgung der oben genannten Regeln vergeben. Sie konnten sich dabei auf ihr eigenes Qualitätsverständnis stützen. Nach Abschluss dieses ersten Ratings (für das sie 25 Minuten Zeit hatten, um im Rahmen der eineinhalb Stunden zu bleiben, die für die Lehrveranstaltung anberaumt waren) wurden die Rater_innen gebeten, ihre Bewertung nicht mehr zu verändern.

⁵⁵ Toral & Way (2018:279) nennen die oben beschriebene Vorgehensweise „ranking“ nach dem englischen Verb „to rank“. Dies ist auch der typische Name für Qualitätsbewertungen dieser Art (vgl. Castilho et al. 2018:21). Trotzdem wird in der vorliegenden Arbeit die Qualitätsbewertung mit „Rating“ statt „Ranking“ bezeichnet, erstens, um terminologische Konsistenz mit den anderen Bezeichnungen zu schaffen, die sich auf die Ratings beziehen („Rater_innen“, „Interrater_innen-Reliabilität“), und zweitens, um die Ratings klar von den paarweisen Rankings abzugrenzen, die zur Auswertung der Ergebnisse der Ratings durchgeführt wurden (siehe Kapitel 4.5).

Im zweiten Teil der Videokonferenz bzw. der Lehrveranstaltungseinheit wurden den Rater_innen die Fehlerkategorien aus Kapitel 5.1 der vorliegenden Arbeit erklärt, und zwar mithilfe des Dokuments, das Posteditorin 2 für ihre Arbeit benutzt hatte. Das Dokument, das für die Ratings nur leicht angepasst wurde, wurde den Rater_innen auch zur Verfügung gestellt. Nach der Präsentation der Fehlerkategorien sollten die Rater_innen ein Rating der drei Übersetzungen in Bezug auf jede einzelne Fehlerkategorie durchführen: Im Falle der Fehlerkategorie Interpunktion z. B. sollten die Übersetzungen nur im Hinblick auf die Qualität und Richtigkeit der Zeichensetzung bewertet werden (und nicht im Hinblick auf die allg. Qualität oder andere Fehlerkategorien). Die Ergebnisse der Fehlerkategorie-Ratings können sich daher durchaus von den Ergebnissen des allgemeinen Ratings unterscheiden. Es ist außerdem zu beachten, dass für die Fehlerkategorie Bedeutung (siehe Kapitel 5.1) nur ein Rating durchzuführen war, d. h., es wurde nur die Oberkategorie abgefragt und nicht auch die einzelnen Unterkategorien. Der Grund dafür lag einerseits darin, dass die Unterkategorien ohnehin in der Oberkategorie enthalten sind und somit beim Rating eher als zusätzliche Hinweise für die Bedeutungsevaluierung verstanden werden konnten, und andererseits, dass eine Abfrage der Unterkategorien das Rating für die Rater_innen erheblich verkompliziert hätte.

Eine weitere Besonderheit lag beim Rating zur Fehlerkategorie IUT (siehe Kapitel 5.1) vor: Um den Rater_innen die Durchführung dieses spezifischen Ratings zu erleichtern, wurde ihnen eine Liste mit allen im RRAT vorkommenden IUT sowie ihren Definitionen zur Verfügung gestellt. Dies wurde als notwendig erachtet, da den Rater_innen (im Gegensatz zu den Posteditorinnen) nicht der ganze Roman zur Verfügung stand und sie so die notwendigen Informationen zur Bewertung der Übersetzung der IUT (wie ihre genaue Bedeutung im Kontext von *Red Rising*) nicht bei der Hand hatten. Die Definitionen wurden dem Red Rising Wiki⁵⁶ entnommen oder ggf. von der Verfasserin der vorliegenden Arbeit anhand des Romans verfasst.

Da es zehn Fehlerkategorien gibt, gab es auch insgesamt zehn Fehlerkategorie-Ratings. Ein erstes dieser Ratings wurde noch im Rahmen der Videokonferenz bzw. der Lehrveranstaltungseinheit abgehalten, damit die Rater_innen, falls nötig, spezifische Fragen zu diesen Ratings stellen konnten. Waren das erste Fehlerkategorie-Rating beendet und alle etwaigen Fragen geklärt, wurde auch die Videokonferenz bzw. die Lehrveranstaltungseinheit abgeschlossen. Die restlichen neun Fehlerkategorie-Ratings sowie das darauf folgende zweite allgemeine Rating, das Hinweise darauf geben sollte, ob die Fehlerkategorien die

⁵⁶ https://red-rising.fandom.com/wiki/Red_Rising_Wiki (Stand: 17.08.2021).

Einschätzung der Rater_innen in Bezug auf die allgemeine Qualität veränderten, wurden von den Rater_innen eigenständig erledigt. Dafür hatten sie eine Woche Zeit.

Bei sämtlichen Ratings durften die Rater_innen alle möglichen Ressourcen verwenden, die sie sonst ebenfalls zum Übersetzen bzw. zum Redigieren oder Korrekturlesen von Texten benutzen würden, mit Ausnahme der Humanübersetzung von *Red Rising* (die im Internet auffindbar ist), Exzerpten davon oder darauf basierenden deutschsprachigen (Fan-)Websites oder dgl. zum Roman. Die Rater_innen hätten nämlich durch diese Ressourcen die Humanübersetzung in den drei zu bewertenden Texten identifizieren können, wodurch es, wie oben schon erwähnt, automatisch zu einer besseren Bewertung der Humanübersetzung hätte kommen können. Dies sollte durch die Einschränkung bei den erlaubten Ressourcen verhindert werden. Außerdem sollten sich die Rater_innen nicht untereinander oder mit jemand anderem über ihre Bewertungen austauschen oder diese aufeinander abstimmen, denn dies hätte eine korrekte Interpretation der Ergebnisse erschwert und das Fleiss' Kappa (Fleiss 1971) verzerrt (siehe Kapitel 4.5).

Ihre Qualitätsbewertungen (die Ränge 1-3) trugen die Rater_innen in ein speziell für diesen Zweck von der Verfasserin der vorliegenden Arbeit erstelltes *Microsoft-Word*-Dokument ein. In diesem Dokument gab es für jedes Rating auch ein eigenes Feld für Anmerkungen, das von den Rater_innen zur Erklärung ihrer Bewertung oder für Notizen genutzt werden konnte; dies war allerdings nicht verpflichtend. Das ausgefüllte Dokument wurde von den Rater_innen bis Ende der Abgabefrist an die Verfasserin der vorliegenden Arbeit per E-Mail geschickt. Damit war die Teilnahme an der Studie für die Rater_innen beendet.

4.5 Datenauswertung

Die während der Studie gesammelten Daten, d. h. einerseits die Posteditionen an sich und andererseits die Bewertungen der Rater_innen, wurden nach Beendigung der Studie ausgewertet. Wie dabei genau vorgegangen wurde, wird im Folgenden beschrieben.

Die „Auswertung“ der Posteditionen erfolgte durch eine Ermittlung der Anzahl der Korrekturen pro Fehlerkategorie (siehe Kapitel 5.1); die Korrekturen waren durch die Nutzung der „Änderungen nachverfolgen“-Funktion für die Verfasserin der vorliegenden Arbeit sichtbar und konnten somit leicht identifiziert werden. Es wurde dementsprechend z. B. ermittelt, bei wie vielen Korrekturen Fehler der Fehlerkategorie Interpunktion der Grund waren, bei wie vielen Problemstellen der Kategorie Stil usw. Dabei wurden dieselben

Ressourcen aus dem Bereich „Tertiärliteratur“ in der Bibliographie dieser Arbeit, d. h. Wörterbücher und dgl., wie für die Entwicklung der Fehlerkategorien verwendet (siehe Kapitel 4.2). War eine Zuordnung zu einer der Fehlerkategorien nicht möglich (etwa weil ihr eine Problemstelle/ein Fehler zugrunde lag, die/der zu keiner der Fehlerkategorien passte), wurde die entsprechende Korrektur der für die Auswertung kreierten Kategorie „Nicht zuordenbar“ zugeteilt. Korrekturen, die an Stellen der RRMÜ durchgeführt wurden, die von der Verfasserin der vorliegenden Arbeit weder als Problemstelle argumentiert noch als Fehler bestätigt werden konnten, die also ihrer Ansicht nach korrekt waren und keinerlei Änderung bedurft hätten⁵⁷, wurden einer zweiten für die Auswertung dazugekommenen Kategorie zugeteilt, nämlich „Persönliche Vorliebe“ (wobei die persönliche Vorliebe der Posteditorinnen gemeint ist). Bei Korrekturen, die direkt andere Korrekturen nach sich ziehen (wie z. B. eine Kasusänderung, die durch den Austausch einer Präposition durch eine andere bedingt wird), wurde nur die erste Änderung gezählt, die durch sie bedingte nicht. Die so ermittelte Anzahl der Korrekturen pro Fehlerkategorie wird für beide Posteditionen in Tabelle 4 in Kapitel 5.2.1 dargestellt.

Nach diesem ersten Schritt wurden die Bewertungen der Rater_innen nach der Methode von Toral & Way (2018:280-283) ausgewertet. Dafür wurde zuerst die Interrater_innen-Reliabilität festgestellt, und zwar durch die Berechnung des Fleiss' Kappa nach Fleiss (1971) und seine Interpretation nach Landis & Koch (1977:165). Die Berechnung erfolgte in *Microsoft Excel* 2010 mithilfe von Walther (2019) und wurde für jedes der insgesamt zwölf Ratings (zwei allgemeine und zehn für die Fehlerkategorien) durchgeführt. Die Ergebnisse dieser Berechnungen sind in Tabelle 5 in Kapitel 5.2.2 ersichtlich.

Im zweiten Schritt wurden die Ergebnisse der Ratings durch drei paarweise Rankings ausgewertet: Die Ränge der RRPEo wurden mit den Rängen der RRPEm pro Rating verglichen, danach wurden die Ränge der zwei Posteditionen jeweils mit jenen der Humanübersetzung verglichen. Dabei galt, wie in Toral & Way (2018:280), Übersetzung X > Übersetzung Y, wenn X einen besseren Rang als Y im Rating erhielt, X < Y, wenn X einen schlechteren Rang bekam, und X = Y, wenn beiden der gleiche Rang gegeben wurde.

Die Ergebnisse der paarweisen Rankings, die pro Rater_in für jedes der zwölf Ratings durchgeführt wurden, wurden zusammengeführt und, wiederum in Anlehnung an Toral & Way (2018:281f.), in Prozentwerten mithilfe von drei in *Microsoft Excel* 2010 erstellten

⁵⁷ Selbstverständlich ist auch in diesem Fall der schon in Kapitel 4.2 erwähnte Subjektivitätsfaktor nicht außer Acht zu lassen, denn es ist durchaus möglich, dass andere Evaluater_innen einige dieser Stellen doch zumindest als problematisch eingestuft hätten.

Balkendiagrammen (Abb. 5-7 in Kapitel 5.2.2) dargestellt. Dementsprechend kann aus einem dieser Balkendiagramme z. B. abgelesen werden, wie hoch der Anteil der Rater_innen ist, die die RRPEo im ersten allgemeinen Rating als besser bewerteten als die RRPEm, wie hoch er im umgekehrten Fall ist und wie viel Prozent der Rater_innen den beiden Posteditionen den gleichen Rang gaben.

Durch die Auswertung der Posteditionen und der Ratingergebnisse sollte dementsprechend festgestellt werden, inwiefern die Verwendung bzw. Nicht-Verwendung der Fehlerkategorien die Posteditionen und ihre Qualität beeinflusste (FF2 und FF3). Die Ergebnisse dieser Auswertung werden in Kapitel 5.2 der vorliegenden Arbeit präsentiert.

5 Datenanalyse

In diesem Kapitel werden nun die durch die im vorigen Kapitel beschriebenen Methoden gesammelten und ausgewerteten Daten präsentiert. Zuerst werden die entwickelten Fehlerkategorien und danach die Ergebnisse der Auswertung der Posteditionen und der Ratings durch die befragten Studierenden vorgestellt. Diese Daten stellen die Antworten auf die Forschungsfragen dar, welche dann im nächsten Kapitel präzise und knapp ausformuliert und im Rahmen der schon existierenden Forschung in diesem Bereich auch im Hinblick auf zukünftige Studien diskutiert werden.

5.1 Fehlerkategorien

Aus dem Vergleich der HPMÜ mit dem HPAT und der HPHÜ konnten **zehn Fehlerkategorien** herausgefiltert werden. Diese werden in Tabelle 3 auf der nächsten Seite überblicksartig dargestellt und danach jede einzeln erklärt und mit einem oder mehreren Beispielen aus den drei oben genannten Texten veranschaulicht. In Anlehnung an die DQF-MQM-Typologie (vgl. TAUS o. J.) sowie die an L-NMT angepasste SCATE-Fehlertypologie (vgl. Tezcan et al. 2019) wurden die Namen der Fehlerkategorien positiv formuliert (es wurde also kein Kompositum mit „-fehler“ wie in Matusov (2019) verwendet). Allerdings wurde wie in Matusov (2019) den Fehlerkategorien bis auf eine Unterkategorisierung keine Hierarchie zugrunde gelegt, da im Hinblick auf ihre Verwendung durch eine Posteditorin sowie durch Rater_innen, die damit keine Erfahrung besitzen, Einfachheit und Verständlichkeit im Vordergrund standen. Darüber hinaus muss noch angemerkt werden, dass es bei jeder Fehlerkategorie Problemstellen gab, deren Problemcharakter nicht durch anerkannte Quellen oder inhaltliche Argumentationen bestätigt, sondern nur mittels selbst verfasster Erklärungen erläutert werden konnte (siehe Kapitel 4.2). Dadurch spielt selbstverständlich der Subjektivitätsfaktor bei allen Fehlerkategorien eine Rolle, jedoch bei manchen erheblich mehr als bei anderen. Diese Fehlerkategorien sind demnach, obwohl sie im Sinne der Einheitlichkeit trotzdem teilweise als solche bezeichnet werden, eher als Problembereiche aufzufassen, die Hinweise darauf geben, wo in der maschinellen Übersetzung evtl. Verbesserungspotential besteht. Um welche der Kategorien es sich dabei genau handelt, wird im Folgenden ersichtlich.

Tabelle 3: Fehlerkategorien und Anzahl Fehler/Problemstellen in der HPMÜ

Fehlerkategorie	Anzahl Fehler/Problemstellen
Auflösung von Anaphern	3
Auslassungen und Hinzufügungen	5
Bedeutung <ul style="list-style-type: none"> – Fixe Wendungen und <i>phrasal verbs</i> – Polysemie – Silbentrennung 	27 ⁵⁸ <ul style="list-style-type: none"> – 7 (von 27) – 6 (von 27) – 2 (von 27)
Formatierung	1
Grammatik und Orthographie	2
Interpunktion	7 (+ 97) ⁵⁹
In-Universe-Terminologie (IUT)	8
Stil	9
Tempus und Modus von Verben	7
Tonfall und Register	16

Die erste Fehlerkategorie ist die **Auflösung von Anaphern**, ein problematischer Bereich, der auch in Kuzman et al. (2019:7), Matusov (2019:16f.), Constantine (2020:474f.), Mohar et al. (2020:134) und Kolb (2021) genannt wird, allerdings teilweise unter anderen Namen und mit einem Fokus auf Pronomen. In den in Kapitel 4.2 behandelten Fehlertypologien gibt es eine sich auf Anaphern beziehende Kategorie in Matusov (2019:14) und in Tezcan et al. (2019:42). Wie in diesen Werken bezeichnet auch die Fehlerkategorie **Auflösung von Anaphern** die inkorrekte Interpretation „[of a] relationship between two elements in a given discourse instance“ (Rocha 2000:82), was dazu führt, dass die Anapher, also das Element, das sich auf das andere (das Antezedens) bezieht (vgl. Ehrhardt & Heringer 2011:25), nicht korrekt übersetzt wird⁶⁰. Typische Anaphern sind Pronomen (vgl. Rocha 2000:82), doch in der HPMÜ kommen auch Verb- und Artikelanaphern vor. Für eine korrekte Interpretation und damit Übersetzung von Anaphern ist oft ein größerer Kontext als nur ein Satz vonnöten, denn nicht immer befindet sich das Antezedens im selben Satz wie die zu übersetzende Anapher oder es ist überhaupt nur impliziert (vgl. Rocha 2000:82f.).

⁵⁸ Nicht alle Fehler der Oberkategorie Bedeutung wurden auch in Unterkategorien eingeteilt, daher ergibt sich bei einer Summierung der Anzahl der Fehler jeder Unterkategorie nicht die Gesamtanzahl der Bedeutungsfehler (27). Für weitere Erklärungen diesbezüglich wird auf den entsprechenden Absatz weiter unten im Kapitel verwiesen.

⁵⁹ Um die Statistik in der Tabelle nicht zu verfälschen, wurden die 97 Anführungszeichenfehler gesondert angeführt. Für weitere Erklärungen diesbezüglich wird auf den entsprechenden Absatz weiter unten im Kapitel verwiesen.

⁶⁰ Es handelt sich dabei also nicht um das rhetorische Stilmittel der Anapher, bei dem mehrere aufeinanderfolgende Sätze gleich begonnen werden (vgl. Baumgarten ²2011:5).

Beispiel 1: Auflösung von Anaphern⁶¹

HPAT	‘ Stay there!’ she said, and she sprinted up the stairs to the girls’ dormitories. Harry and Ron bare-ly had time to exchange mystified looks [...].	(HPAT:237)
HPMÜ	Bleib da!“, sagte sie und sprintete die Treppe zu den Mädchenschlafsälen hinauf. Harry und Ron hatten gerade noch Zeit, verwunderte Blicke auszutauschen [...].	
HPHÜ	» Wartet hier!«, sagte sie und rannte die Stufen zu den Mädchenschlafsälen hoch. Harry und Ron hatten kaum Zeit, sich ratlose Blicke zuzuwerfen [...].	(HPHÜ:239)

In diesem Beispiel geht es um eine Verbanapher, nämlich „Stay“. Es stellt sich bei der Interpretation und Übersetzung dieser Anapher die Frage, ob Hermine („she“) mit einer oder mehreren Personen spricht, denn das verändert die Form des Imperativs im Deutschen (vgl. Fabricius-Hansen ⁹2016:554). Diese Information wird nie direkt im Text genannt, das Antezedens ist also implizit, es kann aber aus dem Kontext abgeleitet werden: Da im nächsten Satz Harry und Ron genannt werden, ist es logisch anzunehmen, dass Hermine mit beiden spricht und nicht nur mit einem von ihnen. Dementsprechend ist „Bleib“ in der HPMÜ ein Fehler und müsste bei korrekter Interpretation der Anapher „Bleibt“ heißen.

Die zweite Fehlerkategorie bezieht sich auf sehr typische Fehler von NMT-Systemen, nämlich **Auslassungen und Hinzufügungen** (vgl. Castilho et al. 2017:118). Dieser problematische Bereich wurde auch im Rahmen von Kuzman et al. (2019:7), Matusov (2019:16f.) und Omar & Gomaa (2020:231) identifiziert, und die beiden Konzepte sind auch Fehlerkategorien in allen in Kapitel 4.2 näher besprochenen Fehlertypologien (vgl. Matusov 2019:14; TAUS o. J.; Tezcan et al. 2019:42). Sie wurden hier zu einer Fehlerkategorie zusammengefasst, da es nur wenige Fehler dieser Art in der HPMÜ gibt und sich die Konzepte durch ihre Komplementarität für eine solche Zusammenfassung eignen, im Gegensatz zur Auflösung von Anaphern oder der Formatierung (siehe unten).

Bei **Auslassungen** wurden bestimmte Teile des Ausgangstextes nicht in die maschinelle Übersetzung übernommen, während **Hinzufügungen** Stellen in der maschinellen Übersetzung sind, die im Ausgangstext nicht zu finden sind. Damit sind selbstverständlich keine sprachlich motivierten **Auslassungen und Hinzufügungen** gemeint (z. B. wäre die inhaltlich korrekte Übersetzung eines zweiteiligen *phrasal verb* durch ein Ein-Wort-Verb keine Auslassung), sondern nur solche, durch die eine inhaltliche bzw. bedeutungstechnische Verzerrung entsteht, wie in den Beispielen ersichtlich wird.

⁶¹ **Fette** Hervorhebungen in den Beispielskästen wurden von der Verfasserin der vorliegenden Arbeit vorgenommen, um die jeweilige Fehlerkategorie besser zu veranschaulichen.

Beispiel 2a: Auslassungen und Hinzufügungen

HPAT	'I got this out of the library weeks ago for a bit of light reading.'	(HPAT:237; Kursivsetzung im Original)
HPMÜ	'Das habe ich mir vor Wochen aus der Bibliothek geholt, um ein bisschen zu lesen.'	(Kursivsetzung in der HPMÜ)
HPHÜ	»Das hab ich schon vor Wo-chen aus der Bibliothek ausgeliehen, leichte Lektüre.« » <i>Leicht?</i> «, sagte Ron [...].	(HPHÜ:239; Kursivsetzung im Original)

Wie an der nicht vorhandenen fetten Markierung im Ausschnitt der HPMÜ zu sehen ist, wurde das Adjektiv „light“ aus dem Ausgangstext nicht in die Übersetzung übernommen, weswegen dieser Fehler eine **Auslassung** ist. Dies ist insofern gravierend, als nicht nur Hermines Aussage (der erste Satz) inhaltlich verändert wird, sondern auch Rons ungläubige Frage ohne dieses Adjektiv nicht funktioniert (davon abgesehen, dass „Light“ in Rons Frage von der Bedeutung her falsch übersetzt wurde). An diesem Beispiel wird allerdings ebenfalls ersichtlich, dass auch Humanübersetzer_innen nicht immer fehlerfrei arbeiten, denn in der HPHÜ wurde „a bit of“ ausgelassen.

Beispiel 2b: Auslassungen und Hinzufügungen

HPAT	'Don't talk to me for a moment,' said Ron when Harry sat down next to him. 'I need to concen–' He caught sight of Harry's face. 'What's the matter with you? You look terrible.'	(HPAT:235)
HPMÜ	'Reden Sie einen Moment nicht mit mir', sagte Ron, als Harry sich neben ihn setzte. Ich muss mich konzentrieren ...' Er bemerkte Harrys Gesicht. 'Was ist denn los mit dir? Du siehst furchtbar aus.'	
HPHÜ	»Sei mal einen Augenblick ruhig«, sagte Ron, als Harry sich neben ihn setzte. »Ich muss mich konzen–« Dann sah er Harrys Gesicht. »Was ist denn mit dir los? Du siehst ja furchtbar aus.«	(HPHÜ:237)

Bei diesem Beispiel gibt es im Ausschnitt der HPMÜ eine fette Hervorhebung, allerdings nicht im Ausgangstext oder der Humanübersetzung. Es handelt sich also um eine **Hinzufügung** in der HPMÜ, nämlich des letzten Teils des Wortes „concentrate“ bzw. „konzentrieren“. Im Ausgangstext ist dieses Wort absichtlich unvollständig (der fehlende Teil wurde durch einen Gedankenstrich ersetzt), um Rons Entsetzen über Harrys Gesichtsausdruck besser auszudrücken: Er ist so entsetzt, dass er sich mitten im Wort unterbricht. In die Humanübersetzung wurde dies übernommen, in die maschinelle Übersetzung allerdings nicht. Stattdessen wurde der letzte Teil des Wortes hinzugefügt, was Rons starkes Gefühl abschwächt, zumal auch der Gedankenstrich durch drei Punkte ersetzt wurde.

Bedeutung ist die dritte Fehlerkategorie und auch die einzige, die Unterkategorien aufweist, weil es sehr viele Fehler in der HPMÜ gibt, die sich auf die Semantik von Wörtern, fixen Mehrwortausdrücken und nicht fixierten Wortsequenzen (ähnlich der Phrasen bei der PBSMT; siehe Kapitel 1) beziehen. Dieser problematische Bereich (inkl. Unterkategorien) wird auch in Humblé (2019:48), Kuzman et al. (2019:7), Matusov (2019:16f.), Tezcan et al. (2019:45), Fonteyne et al. (2020:3795), Webster et al. (2020a:11), Constantine (2020:476f.), Omar & Gomaa (2020:231), Mohar et al. (2020:137) und Kolb (2021) angeführt, oft in Bezug auf Mehrdeutigkeit. Fehlerkategorien, die sich auf die **Bedeutung** von Wörtern, Mehrwortausdrücken und Wortsequenzen beziehen, gibt es ebenfalls in Matusov (2019), TAUS (o. J.) und Tezcan et al (2019), nämlich severe/minor meaning error und M-/mistranslation.

Fehler in dieser Kategorie sind also Wörter, Mehrwortausdrücke oder Wortsequenzen in der maschinellen Übersetzung, welche die Bedeutung der korrespondierenden Wörter, Ausdrücke oder Sequenzen des Ausgangstextes nicht oder nicht genau wiedergeben. Die Kategorie ist abgegrenzt von Auslassungen und Hinzufügungen, weil bei **Bedeutungsfehlern** in der Übersetzung keine Wörter/Wortteile fehlen oder zusätzlich eingefügt wurden. Außerdem ist sie von der Kategorie Tempus und Modus von Verben zu unterscheiden, die sich auf die genannten Aspekte von Verben bezieht (die selbstverständlich auch Bedeutung transportieren), aber nicht auf die Bedeutung der Verben selbst (siehe unten).

Beispiel 3a: Bedeutung

HPAT	Once term had started, they were back to skimming through books for ten minutes during their breaks.	(HPAT:233)
HPMÜ	Als das Schuljahr begonnen hatte, blättern sie in den Pausen wieder zehn Minuten lang in Büchern.	
HPHÜ	Nach dem Ende der Ferien fingen sie wieder an zu suchen und in den Zehn-Minuten-Pausen die Bücher durchzublätern.	(HPHÜ:235)

In diesem Kontext bedeutet „term“ „[e]ach of the periods in the year, alternating with holiday [...], during which instruction is given in a school, college, or university [...]“ (Lexico.com 2021a), allerdings bedeutet „Schuljahr“ „Zeitraum eines Jahres für die Arbeit an der Schule [...], in dem nach einem Lehrplan bestimmte Unterrichtspensen zu bewältigen sind“ (Duden 2021; Hervorhebung im Original). Ein Schuljahr ist dementsprechend ein ganzes Jahr, ein „term“ allerdings nicht. Selbstverständlich kann der Beginn eines „term“ mit dem Beginn eines Schuljahres übereinstimmen (im Falle von *Harry Potter* wäre das am 1. September (vgl. Rowling 2010/1997:61)), aber dies ist hier nicht der Fall, denn drei Absätze vor diesem Ausschnitt werden Weihnachtsferien erwähnt (vgl. HPAT:233), d. h., dieser „term“ fängt während des Schuljahres an. Da also „term“ von der im Kontext gegebenen **Bedeutung** her

falsch übersetzt wurde, wird obiger Fehler, der in der Humanübersetzung durch einen Perspektivenwechsel (nicht Beginn eines „term“, sondern „Ende der Ferien“) umschifft wurde, zur Kategorie **Bedeutung** gezählt.

Diese Kategorie umfasst allerdings nicht nur generelle Bedeutungsfehler, sondern auch Fehler aus folgenden drei Unterkategorien: **fixe Wendungen und phrasal verbs**, **Polysemie** sowie **Silbentrennung**. Passte ein Fehler in der HPMÜ nicht in eine dieser Unterkategorien (wie der oben besprochene Fehler bei der Übersetzung von „term“), so wurde er für die Zählung der Fehler nur der Oberkategorie Bedeutung zugeteilt, passte er allerdings in eine der Unterkategorien, wurde er bei der jeweiligen Unterkategorie dazugezählt und bei der Oberkategorie. So ergab sich ein Wert von 27 Bedeutungsfehlern, von denen 15 noch in ihre jeweiligen Unterkategorien eingeteilt werden können.

Die erste dieser Unterkategorien bezieht sich auf die Bedeutung von **fixen Wendungen und phrasal verbs**, die in der Übersetzung nicht oder nicht genau transportiert wird. Im Gegensatz zu der Oberkategorie Bedeutung sind hier folglich stets Mehrwortausdrücke involviert. Bei der zweiten Unterkategorie ist der Grund für den Bedeutungsfehler die **Polysemie** von ausgangssprachlichen Wörtern oder Ausdrücken, d. h. „das Auftreten eines Zeichens mit mehreren untereinander verwandten Sememen (= Lesarten/Inhaltseinheiten)“ (Zöfgen 1989:780). Der falsch übersetzte Teil des Ausgangstextes hat dementsprechend zwei oder mehr mögliche, miteinander verwandte (einander ähnelnde) Bedeutungen bzw. Bedeutungsebenen (was bei der Oberkategorie Bedeutung nicht der Fall sein muss – hier kann auch Homonymie herrschen), doch der DeepL-Übersetzer wählte die falsche dieser Möglichkeiten aus. Die dritte Unterkategorie schließlich beschreibt Bedeutungsfehler, die aufgrund der **Silbentrennung** im Ausgangstext passierten. Dass die **Silbentrennung** der Grund für diese Fehler war, wurde festgestellt, indem die entsprechenden Sätze des Ausgangstextes noch einmal ohne die **Silbentrennung** vom DeepL-Übersetzer übersetzt wurden. War die Übersetzung in diesem Fall richtig, wurde der Fehler dieser Unterkategorie zugeordnet.

Beispiel 3b: Bedeutung – fixe Wendungen und *phrasal verbs*

HPAT	‘I never thought to look in here!’ she whispered excitedly.	(HPAT:237)
HPMÜ	‘Ich hätte nie gedacht , dass ich hier mal reinschaue!’, flüsterte sie aufgeregt.	
HPHÜ	»Ich hab einfach nicht daran gedacht , hier drin nachzu-schauen«, flüsterte sie erregt.	(HPHÜ:239)

Die **fixe Wendung**, die hier falsch übersetzt wurde, ist „never“ bzw. „not think to do something“, in diesem Fall „to look in here“. Die Wendung bedeutet „to not consider doing

something, especially when you later wish you had done it“ (*LDOCE Online* o. J.), was auch mit dem Kontext übereinstimmt: Nach wochenlangem Suchen nach einer bestimmten Information kommt Hermine durch eine Aussage von Harry plötzlich auf die Idee, in einem Buch nachzusehen, das sie schon vor langer Zeit aus der Bibliothek ausgeliehen hatte. In diesem Buch finden die drei Freunde dann die gewünschte Information (vgl. HPAT:237f.). Dies wird auch in der Humanübersetzung so transportiert, allerdings nicht in der HPMÜ: Hier lässt Hermine's Aussage die Leser_innen eher glauben, sie wäre überrascht, überhaupt in dieses Buch zu schauen (weil sie nicht gedacht hätte, dies jemals in ihrem Leben tun zu können/dürfen), und damit wird die Bedeutung der **fixen Wendung** in der HPMÜ nicht richtig wiedergegeben.

Beispiel 3c: Bedeutung – Polysemie

HPAT	<i>It also produces the Elixir of Life, which will make the drinker immortal.</i>	(HPAT:238; Kursivsetzung im Original)
HPMÜ	<i>Er produziert auch das Elixier des Lebens, das den Trinker unsterblich macht.</i>	(Kursivsetzung in der HPMÜ)
HPHÜ	Auch zeugt er das Elixier des Lebens, wel-ches den, der es trinkt , unsterblich macht.	(HPHÜ:240)

„drinker“ ist hier **polysemisch** – die zwei miteinander verwandten Bedeutungen sind „[a] person who drinks a particular drink“ und „[a] person who drinks alcohol, especially to excess“ (*Lexico.com* 2021a). Allerdings ist die erste Bedeutung gemeint, da im Ausgangstext weder angegeben wird, dass das „Elixir of Life“ ein alkoholisches Getränk ist, noch in diesem Kontext ein_e Alkoholiker_in erwähnt wird (vgl. HPAT:238). Die HPMÜ jedoch gibt mit „Trinker“, das jemanden bezeichnet, der „gewöhnheitsmäßig alkoholische Getränke in zu großer Menge zu sich [...] [nimmt]“ (*Duden* 2021), die zweite Bedeutung von „drinker“ wieder. Folglich entschied sich der DeepL-Übersetzer im Gegensatz zum Humanübersetzer für die falsche Bedeutungsebene von „drinker“.

Beispiel 3d: Bedeutung – Silbentrennung

HPAT	At that moment Neville toppled into the com-mon room .	(HPAT:235)
HPMÜ	In diesem Moment stürzte Neville in den Kommunikationsraum .	
HPHÜ	In diesem Moment stürzte Neville kopfüber in den Gemeinschaftsraum .	(HPHÜ:237)

In diesem Beispiel wird aufgrund der **Silbentrennung** in „com-mon“ die Bedeutung des Ausdrucks „common room“ nicht richtig wiedergegeben. Dies wurde, wie oben schon angemerkt, durch eine Probe mit DeepL ohne diese **Silbentrennung** überprüft. Ein weiteres Indiz dafür, dass die **Silbentrennung** der Grund für diesen Fehler war, ist außerdem, dass in

der HPMÜ ein paar Absätze vor diesem Ausschnitt „common room“ (ohne **Silbentrennung**) richtig als „Gemeinschaftsraum“ wiedergegeben wird. Es gibt noch einen weiteren Fehler dieser Art, der ebenfalls mit der Probe durch DeepL bestätigt wurde.

Die nächste Kategorie, die auf derselben Ebene wie die Kategorie Bedeutung liegt, ist die **Formatierung**. Als problematischer Bereich wird diese Kategorie in keinem der in Kapitel 4.2 angeführten Werke genannt, und sie kommt auch nur in einer der behandelten Fehlertypologien vor, nämlich der DQF-MQM-Typologie, die nicht an Literatur angepasst ist (vgl. TAUS o. J.). Die **Formatierungskategorie** hier ähnelt der DQF-MQM-Kategorie *Local formatting* insofern, als es um falsche Übertragungen von lokalen **Formatierungen** geht, die eine inhaltliche Bedeutung haben (mit denen z. B. ein bestimmtes Wort betont oder angezeigt wird, dass der betreffende Ausschnitt aus einem Buch stammt). Es gibt nur einen solchen Fehler in der HPMÜ (dafür mehrere in der HPHÜ), aber da diese Kategorie die einzige ist, die sich auf das Aussehen des Textes bezieht, wäre es nicht sinnvoll, sie mit einer anderen zusammenzufassen (wie das z. B. bei den Auslassungen und Hinzufügungen der Fall ist).

Beispiel 4: Formatierung

HPAT	I <i>told</i> you I'd read the name somewhere before [...].	(HPAT:237; Kursivsetzung im Original)
HPMÜ	Ich <i>sagte</i> doch, <i>dass</i> ich den Namen schon einmal irgendwo gelesen hatte [...].	(Kursivsetzung in der HPMÜ)
HPHÜ	Hab euch doch gesagt, dass ich den Namen schon mal irgendwo gelesen hab.	(HPHÜ:239)

Im HPAT wurde „*told*“ kursiv gesetzt, weil Harry damit betonen will, dass er seinen Freunden schon mehrmals gesagt hat, den erwähnten Namen schon einmal irgendwo gelesen zu haben, was auch am Anfang dieses Ausschnitts aus dem ersten *Harry-Potter*-Band erwähnt wird. In der HPMÜ wurde jedoch, aus Gründen, die nicht eruiert werden konnten, die Kursivsetzung von „*sagte*“ auf „*dass*“ ausgedehnt, was v. a. beim lauten Lesen die Lesart und Betonung dieses Satzes verändert. Darüber hinaus wurde in der HPMÜ auch die Partikel „*doch*“ eingefügt, die ebenfalls eine betonende Wirkung hat (vgl. *Duden* 2021), was die Lesart weiter verändert. Aus demselben Grund ist allerdings die fehlende Kursivsetzung in der HPHÜ (im Gegensatz zu jener in den Beispielen 3c und 6) nicht unbedingt als Fehler zu werten: Die im Ausgangstext ausgedrückte Betonung wird durch „*doch*“ realisiert, deswegen ist eine Kursivsetzung nicht notwendig.

Auch in der nächsten Kategorie – **Grammatik und Orthographie** – gibt es nur wenige Fehler: einen aus dem Bereich Grammatik und einen aus dem Bereich Orthographie. Diese

Fehler wurden zu einer Kategorie zusammengefasst, weil es nur wenige solcher Fehler in der HPMÜ gibt und sie beide auch ohne Hilfe des Ausgangstextes erkannt und korrigiert werden können. Unter diese Kategorie würden also Fehler bzgl. deutscher Fälle, Groß-/Kleinschreibung (siehe Beispiel 5), andere Rechtschreibfehler, falsche Bildung von Tempori und Modi etc. fallen. Als problematischer Bereich werden solche Fehler nur in Humblé (2019:48) genannt (und zwar nur auf Grammatik bezogen), doch es gibt diese oder ähnliche Fehlerkategorien auch in der DQF-MQM-Typologie (*Grammar*, *Spelling*) (vgl. TAUS o. J.) sowie in jener von Tezcan et al. (2019:42) (*grammar & syntax*, *spelling*).

Beispiel 5: Grammatik und Orthographie

HPAT	As Neville walked away Harry looked at the Famous Wizard card. 'Dumbledore again,' he said. 'He was the first one I ever –'	(HPAT:236)
HPMÜ	Als Neville wegging, schaute Harry auf die Karte "Berühmter Zauberer". 'Wieder Dumbledore', sagte er. 'Er war der erste , den ich je -'	
HPHÜ	Neville ging hinaus und Harry sah sich die Sammelkarte der berühmten Zauberer an. »Schon wieder Dumbledore«, sagte er. »Er war der Erste , den ich –«	(HPHÜ:238f.)

In diesem Beispiel wurde „erste“ in der HPMÜ fälschlicherweise klein geschrieben, obwohl eine Substantivierung vorliegt und die Ordnungszahl nicht attributiv-elliptisch gebraucht wird (vgl. Bibliographisches Institut o. J.a): Es kann im Satz nach der Ordnungszahl weder „Karte“ (sonst müsste es im Ausgangstext „It/This was the first one ...“ und dann in der Übersetzung „Das war die erste ...“ lauten) noch „Dumbledore“ (der eine Person ist) eingefügt werden. Dementsprechend ist die Großschreibung von „Erste“ wie in der HPHÜ richtig.

Wie der Name der nächsten Kategorie vermuten lässt, beschreibt **Interpunktion** alle möglichen Zeichensetzungsfehler, sowohl Verstöße gegen die Regeln der Zielsprache als auch falsche Übertragungen aus dem Ausgangstext (siehe unten). Als problematischer Bereich wird die **Interpunktion** nur in Kuzman et al. (2019:7) genannt, als Fehlerkategorie existiert sie in der DQF-MQM-Typologie (vgl. TAUS o. J.) sowie in jener von Tezcan et al. (2019:42), in letzterer zusammengefasst mit Groß-/Kleinschreibung.

Interpunktionsfehler gibt es in der HPMÜ nur in Bezug auf die Zeichen Komma, Bindestrich bzw. Gedankenstrich und Anführungszeichen. Bei den Kommatafehlern handelt es sich in erster Linie um fehlende Konsistenz bei deren Setzung, während Gedankenstriche aus dem Ausgangstext meist in Bindestriche (siehe Beispiele 5 und 6) und einmal in drei Auslassungspunkte umgewandelt wurden (siehe Beispiel 2b). Die überwiegende Mehrheit der **Interpunktionsfehler** in der HPMÜ, nämlich 97 von insgesamt 104, bezieht sich jedoch auf die Setzung von Anführungszeichen. Allerdings wurden, um die Statistik in Tabelle 3 nicht zu

verfälschen, diese 97 Fehler gesondert angeführt. Anführungszeichenfehler sind so zahlreich, da der DeepL-Übersetzer diese Zeichen systematisch falsch setzte: Bei fast allen direkten Reden wurden die einfachen Anführungszeichen aus dem Englischen im Deutschen beibehalten (siehe Beispiele 2a, 2b, 3b, 5 und 6), obwohl sie in der Zielsprache doppelt sowie unten und oben (anstatt nur oben gesetzt) sein müssten (vgl. Staaden ²2016:122f.). Manchmal wurde das erste einfache Anführungszeichen auch ausgelassen (siehe Beispiel 2b). In den seltenen Fällen, in denen doppelte Anführungszeichen verwendet wurden, wurden diese nur oben gesetzt (siehe Beispiel 5: Karte "Berühmter Zauberer") bzw. das erste Anführungszeichen überhaupt weggelassen (siehe Beispiele 1 und 2a). Darüber hinaus ist über alle Anführungszeichen in der HPMÜ zu sagen, dass sie im Schreibmaschinensatz, d. h. gerade (' oder "), gesetzt wurden anstatt typographisch („“,) wie allgemein bevorzugt (z. B. werden in der amtlichen Regelung der deutschen Rechtschreibung stets typographische Anführungszeichen verwendet; vgl. WAHRIG ⁸2011:79ff.). Im folgenden Beispiel schließlich ist nicht nur die Setzung der Anführungszeichen inkorrekt, sondern auch jene von anderen Satzzeichen:

Beispiel 6: Interpunktion

HPAT	'I've found Flamel! I <i>told</i> you I'd read the name somewhere before, I read it on the train coming here – listen to this: "Professor Dumbledore is particularly famous [...] <i>his work on alchemy with his partner, Nicolas Flamel</i> "!'	(HPAT:237; Kursivsetzung im Original)
HPMÜ	'Ich habe Flamel gefunden! Ich <i>sagte</i> doch, <i>dass</i> ich den Namen schon einmal irgendwo gelesen hatte, ich habe ihn im Zug auf dem Weg hierher gelesen - hören Sie sich das an: "Professor Dumbledore ist besonders berühmt für [...] <i>seine Arbeit an der Alchemie mit seinem Partner Nicolas Flamel!</i> "'	(Kursivsetzung in der HPMÜ)
HPHÜ	»Ich hab Flamel ge-funden! Hab euch doch gesagt, dass ich den Namen schon mal irgendwo gelesen hab. Es war im Zug hierher. Hört mal: »Professor Dumbledores Ruhm beruht vor allem auf [...] seinem Werk über Alchemie, ver-fasst zusammen mit seinem Partner Nicolas Flamel.«!	(HPHÜ:239)

Abgesehen davon, dass, wie oben schon erwähnt, in der HPMÜ nur einfache, immer oben gesetzte, nicht typographische Anführungszeichen für die direkte Rede verwendet wurden, sind auch die Anführungszeichen für das Zitat innerhalb der direkten Rede falsch gesetzt: Diese müssten im Deutschen einfache, aber trotzdem unten und oben gesetzte typographische Anführungszeichen sein (vgl. Staaden ²2016:124), stattdessen sind es doppelte im Schreibmaschinensatz, die sich nur oben befinden. Außerdem wurde in der HPMÜ der Gedankenstrich im zweiten Satz fälschlicherweise zu einem Bindestrich und das Ausrufezeichen am Ende in das Zitat inkludiert, obwohl im Ausgangstext genau das Gegenteil der Fall ist. Die **Interpunktion** am Ende des Zitats steht auch in der HPHÜ im Gegensatz zu jener im

Ausgangstext: Es wurde ein Punkt in das Zitat eingefügt, der zwar vom Zitat her grundsätzlich korrekt ist⁶², allerdings die Leser_innen aufgrund des danach gesetzten Ausrufezeichens vermutlich doch etwas stutzig macht.

Die siebte der entwickelten Fehlerkategorien heißt **In-Universe-Terminologie (IUT⁶³)** und bezieht sich auf Wörter und Mehrwortausdrücke, die für die Welt, in der der literarische Ausgangstext spielt, erfunden wurden bzw. dort auf spezifische Art und Weise verwendet werden (z. B. „Gryffindor“, „Quidditch“, „Leg-Locker Curse“ etc.). Diese IUT bereiten dem DeepL-Übersetzer wie vermutet Schwierigkeiten, und auch für Humanübersetzer_innen sind sie nicht leicht zu übersetzen (vgl. Jentsch 2006:190), z. B. weil sie Wortspiele beinhalten oder Humor evozieren (vgl. Jentsch 2006:196f.) oder, im Falle von Mehrwortausdrücken, alliteriert sind (vgl. Brøndsted & Dollerup 2004:67). Da „**In-Universe-Terminologie**“ eine von der Verfasserin der vorliegenden Arbeit entwickelte Bezeichnung ist, findet sie sich nicht direkt in einem der in Kapitel 4.2 angeführten Werke, allerdings nennen z. B. Moorkens et al. (2018:254) kulturelle Elemente als problematischen Bereich (die durch die IUT benannten Gegenstände oder Konzepte können in diesem Fall als kulturelle Elemente der Welt von *Harry Potter* aufgefasst werden), Omar & Gomaa (2020:231) führen unbekannte Wörter als problematischen Bereich an und Mohar et al. (2020:138f.) Neologismen. In den behandelten Fehlertypologien gibt es ebenfalls Fehlerkategorien, die der IUT ähneln oder Aspekte von dieser Kategorie beinhalten, nämlich unknown word or segmentation error sowie consistency/term translation error in Matusov (2019:14), Terminology in DQF-MQM (vgl. TAUS o. J.) und inconsistency in Tezcan et al. (2019:42).

In der HPMÜ werden frei erfundene Einzelwörter, wie in der HPHÜ, einfach in die Zielsprache übernommen, allerdings entstehen durch diese übernommenen Neologismen in der HPMÜ teilweise holprige Konstruktionen, die der IUT-Kategorie zugeordnet wurden und nicht der Kategorie Stil. Außerdem werden mit diesen Neologismen gebildete *compound nouns* teilweise uneinheitlich übersetzt (z. B. wurde bei drei von vier *compound nouns* mit dem erfundenen Wort „Quidditch“ im Deutschen ein Bindestrich im Kompositum verwendet, bei dem vierten nicht).

Mehr IUT-Fehler kommen allerdings in Bezug auf Mehrwortausdrücke vor, die aus existierenden und frei erfundenen bzw. nur aus existierenden Wörtern bestehen. Hierbei

⁶² Harry liest hier einen Text vor, der vorher im ersten *Harry-Potter*-Band schon einmal auftaucht und in dem der betreffende Satz nach „Flamel“ tatsächlich mit einem Punkt endet (vgl. Rowling 1997/2010:114 bzw. Rowling 1998/2013:114).

⁶³ Diese Abkürzung kann sowohl für „In-Universe-Terminologie“ als Konzept stehen als auch für einen einzigen oder mehrere solcher Termini. Ob das Konzept oder einzelne Termini gemeint sind, ist aus dem jeweiligen Kontext ersichtlich.

kommt es in der HPMÜ zu Übersetzungen, die sich aufgrund von verschiedenen Mängeln (wie z. B. Nicht-Beibehaltung der Alliteration, des Spiels mit Sprache, der von der Autorin verwendeten sprachlichen Struktur des **IUT** oder schlicht und einfach der Bedeutung der einzelnen Konstituenten des Ausdrucks) nicht reibungslos in das Leseerlebnis bzw. in die von der Autorin konstruierte Welt einfügen. Selbstverständlich kommen solche Probleme auch in der HPHÜ vor, da es nicht immer möglich ist, alle Aspekte eines **IUT** in der Übersetzung beizubehalten.

Darüber hinaus haben unterschiedliche Übersetzer_innen auch unterschiedliche Herangehensweisen an **IUT**, seien sie erfundene Einzelwörter oder Mehrwortausdrücke, wie sich gut an der Tabelle in Jentsch (2006:198f.) abzeichnet: So z. B. hat der französische Übersetzer der *Harry-Potter*-Bücher die Namen der vier Häuser von Hogwarts (etwa „Gryffindor“) an die Zielsprache angepasst, während dies der deutsche Übersetzer und die spanische Übersetzerin nicht getan haben. Folglich werden **IUT** von unterschiedlichen Übersetzer_innen unterschiedlich behandelt, d. h., die Strategien diesbezüglich sind subjektiv, weswegen auch im Falle der Bewertung dieser Übersetzungen als problematisch oder nicht-problematisch bzw. als Fehler oder Nicht-Fehler die Subjektivität eine große Rolle spielt. Dementsprechend konnten auch die in Bezug auf die **IUT** problematischen Stellen der HPMÜ (oder auch der HPHÜ) nur in Einzelfällen anhand von anerkannten Quellen als Fehler bestätigt werden (wie etwa an untenstehender Erklärung von Beispiel 7 ersichtlich wird), was der Grund dafür ist, warum diese Fehlerkategorie die erste ist, die eher als Problembereich einzustufen ist.

Beispiel 7: IUT

HPAT	Dumbledore had convinced Harry not to go look-ing for the Mirror of Erised again and for the rest of the Christmas holidays the Invisibility Cloak stayed folded at the bottom of his trunk.	(HPAT:233)
HPMÜ	Dumbledore hatte Harry davon überzeugt, nicht noch einmal nach dem Spiegel von Erised zu suchen, und für den Rest der Weihnachtsferien blieb der Unsichtbarkeitsumhang zusammengefaltet unten in seiner Truhe.	
HPHÜ	Dumbledore hatte Harry davon überzeugt, besser nicht mehr nach dem Spiegel Nerhegeb zu suchen, und die rest-lichen Tage der Weihnachtsferien blieb der Tarnumhang zusammengefaltet auf dem Boden seines großen Koffers.	(HPHÜ:235)

In diesem Beispiel kommen zwei **IUT** vor: „Mirror of Erised“ und „Invisibility Cloak“, gekennzeichnet durch die Großschreibung. Der erste **Terminus** besteht aus dem alltäglichen Wort „mirror“, der Präposition „of“ und dem auf den ersten Blick erfunden scheinenden Wort „erised“, das allerdings nur „desire“ rückwärts geschrieben ist. Das ist kein Zufall, denn der Spiegel zeigt kein Spiegelbild, sondern „the [...] desire of our hearts“ (Rowling

2010/1997:231). Dementsprechend gibt der **Terminus** einen versteckten Hinweis auf die Funktion des Gegenstandes, den er benennt. In der HPMÜ allerdings geht dieser Hinweis verloren, denn das aus „erised“ gebildete Wort „desire“ existiert im Deutschen nicht. Darüber hinaus ist die Konstruktion mit „von“ holprig und behindert den Lesefluss. In der HPHÜ wurde der Hinweis aus dem Ausgangsterminus beibehalten und „Erised“ durch „Nerhegeb“ ersetzt, das rückwärts geschrieben „Begehren“ heißt. Außerdem wurde die Präposition weggelassen, was zu einem natürlicher klingenden Terminus im Deutschen führt.

Der zweite **IUT** bezeichnet einen Umhang, der den_die Träger_in unsichtbar macht (vgl. Rowling 2010/1997:218). Es wird also wieder die Funktion des Gegenstandes durch den **Terminus** beschrieben, diesmal allerdings sehr eindeutig. Mit der wörtlichen Übersetzung der beiden Teilelemente und ihrer Verbindung mit einem Fugenelement löst der DeepL-Übersetzer dieses Problem auf flüssige und natürliche Art und Weise. Folglich liegt hier keine Problemstelle/kein Fehler in der HPMÜ vor, was zeigt, dass NMT-Systeme (oder zumindest DeepL) durchaus passende Vorschläge für die Übersetzung von **IUT** bieten können. Der Humanübersetzer entschied sich dahingegen für „Tarnumhang“ bzw. wird an anderer Stelle auch „Umhang, der unsichtbar macht“ (Rowling 1998/2013:220) benutzt. Erstere Lösung erscheint ungenauer als jene in der HPMÜ, da „tarnen“ nur „jemanden, etwas vor dem Erkennt-, Gesehenwerden schützen, indem man ihn, es verhüllt oder der Umgebung angleicht“ (Duden 2021) bedeutet, während der „Invisibility Cloak“ tatsächlich unsichtbar macht, sodass es unmöglich ist, den_die Träger_in zu sehen (vgl. Rowling 2010/1997:218).

Unter die nächste Kategorie – **Stil** – fallen Fehler wie z. B. eine direkte Übernahme des ausgangssprachlichen Satzbaus, was zu einer holprigen Konstruktion in der Zielsprache führt, zwar verständliche, aber im jeweiligen Kontext ungewöhnliche Präpositionen oder Adjektive oder auch Wortwiederholungen, die im Ausgangstext nicht auftauchen. Folglich ist **Stil** hier als für den Kontext angemessener und v. a. natürlicher und flüssiger Sprachgebrauch (*idiomatic* im Englischen) zu verstehen und kommt den DQF-MQM-Kategorien *Awkward* und *Unidiomatic* (vgl. TAUS o. J.) sowie *disfluency* und *repetition* aus Tezcan et al. (2019:42) nahe. In Matusov (2019:14f.) existiert diese Kategorie nicht, Elemente davon werden aber bei den Syntaxfehlern sowie den *tone/register error[s]* und *repetition error[s]* angeführt. Als problematischer Bereich wird der **Stil** in der hier präsentierten Auffassung zudem in Tezcan et al. (2019:45), Fonteyne et al. (2020:3975), Webster et al. (2020a:10f.) und Constantine (2020:474) angeführt.

Da es im vorliegenden Forschungsprojekt um literarische MT geht und um Verwirrung mit anderen Auffassungen von Stil zu vermeiden, erscheint an diesem Punkt eine kurze

Diskussion dieses Begriffs vonnöten, der abseits seiner vielschichtigen allgemeinsprachlichen Verwendung (vgl. *Duden* 2021) v. a. in den Disziplinen der Kunst-, Sprach- und Literaturwissenschaft Gebrauch findet (vgl. [*MLL*] ³2007:732; Sowinski 1991:1). In letzterer wird der Begriff Stil, auf Sprache bezogen, ebenfalls auf unterschiedliche Art und Weise verwendet (z. B. zur Beschreibung des sprachlichen Stils von bestimmten Texten oder der Wirkung einer Textsorte etc.), und es herrscht keine Einigkeit darüber, was in dem Begriff enthalten ist und was nicht bzw. worauf er anwendbar ist und worauf nicht (vgl. Sowinski 1991:1f.). Dies macht eine Definition des Begriffs innerhalb der Literaturwissenschaft naturgemäß schwierig, doch ist seit dem 18. Jahrhundert in dieser Disziplin die „Auffassung von Stil als dem Ausdruck der künstlerischen Individualität“ (Sowinski 1991:11) oder, anders formuliert, als die „in hohem Grade unverwechselbare[n] Grundmuster, die das Kunstschaffen einzelner Autoren [...], die Ausprägungsformen einzelner Kunstwerke [...] und bestimmter Werktypen [...] kennzeichnen“ ([*MLL*] ³2007:732), vorherrschend (vgl. Sowinski 1991:11).

In diesem Sinne ist auch „Stil“ in dem in Kapitel 3 zitierten Werk von Kenny & Winters (2020a) zu verstehen: Der an ihrer Studie beteiligte Übersetzer hat einen eigenen Übersetzungsstil, d. h., seine Übersetzungen sind durch bestimmte, ihn auszeichnende Merkmale gekennzeichnet. Dies ist allerdings nicht die Auffassung von Stil, die in der MT-Forschung hauptsächlich zum Tragen kommt: Wie an den Kategorien *Style* und *style & register* der DQF-MQM-Typologie bzw. der an die L-NMT angepassten SCATE-Fehlertypologie ersichtlich wird, wird **Stil** in diesem Forschungsbereich eher als für den Kontext angemessener (evtl. sogar in einem Styleguide vorgegebener), natürlicher und flüssiger Sprachgebrauch verstanden (vgl. TAUS o. J.; Tezcan et al. 2019:42). Folglich entspricht die hier vorliegende Fehlerkategorie **Stil**, die im folgenden Beispiel veranschaulicht wird, der in der MT-Forschung vorherrschenden Auffassung dieses Begriffs und nicht jener, die in der Literaturwissenschaft dominant ist.

Beispiel 8: Stil

HPAT	Chess was the only thing Hermione ever lost at, something Harry and Ron thought was very good for her.	(HPAT:236)
HPMÜ	Schach war das Einzige, bei dem Hermine jemals verloren hatte, etwas, von dem Harry und Ron dachten, dass es sehr gut für sie war.	
HPHÜ	Schach war das Einzige, bei dem Hermine immer verlor, und Harry und Ron waren der Meinung, das könne ihr nur guttun.	(HPHÜ:237)

Der in der HPMÜ fett markierte Teil dieses Satzes ist korrekt und verständlich, doch recht holprig und unnatürlich, da die ausgangssprachliche Konstruktion wortwörtlich (bis auf die

für die Zielsprache notwendigen Veränderungen) übernommen wurde, weswegen diese Problemstelle in der HPMÜ der Kategorie **Stil** zuzuordnen ist. Dahingegen ist die etwas freiere Konstruktion in der HPHÜ durch die Aneinanderreihung von zwei Hauptsätzen und die Vermeidung eines Nebensatzes mit „dass“ deutlich flüssiger zu lesen.

Die oben angeführte Erklärung der Problemstelle ist exemplarisch für nahezu alle Problemstellen dieser Fehlerkategorie, die demnach nicht als Fehler bestätigt werden konnten. Folglich hängt es stark vom Ermessen des_der Evaluierenden (in diesem Fall der Verfasserin der vorliegenden Arbeit) ab, welche Stellen einer Übersetzung der Fehlerkategorie **Stil** zugeordnet werden und welche für angemessen, natürlich und flüssig (und damit nicht als problematisch) befunden werden, denn Angemessenheit, Natürlichkeit und Flüssigkeit im Sprachgebrauch liegen selbstverständlich zumindest teilweise im Auge des_der Betrachtenden. Dementsprechend ist die Kategorie **Stil** eine weitere der Kategorien, die nicht als Fehlerkategorien per se, sondern als Problembereiche aufgefasst werden sollen.

Die vorletzte Fehlerkategorie heißt **Tempus und Modus von Verben**. Verben im Allgemeinen bzw. **Tempus** oder **Modus** tauchen als problematische Bereiche ebenfalls in Kuzman et al. (2019:7), Mohar et al. (2020:132), Schmidt (2021) und Kolb (2021) auf. In Matusov (2019:15) wird das **Tempus** von Verben bei den Fehlerkategorien bzgl. der Syntax erwähnt, in Tezcan et al. (2019:42) ist das **Tempus** von Verben eine Unterkategorie von *coherence*. Es ist zu beachten, dass sich die hier entwickelte Kategorie wie die korrespondierende in Tezcan et al. (2019:42) nicht auf Fehler bei der Bildung der **Tempi und Modi** bezieht (solche wären der Kategorie Grammatik und Orthographie zuzuordnen), sondern auf Fehler bei der Übertragung der im Ausgangstext ausgedrückten Aspekte in die Zielsprache (siehe auch Beispiel 8):

Beispiel 9: Tempus und Modus von Verben

HPAT	How he had managed to climb through the portrait hole was anyone's guess, because his legs had been stuck together with what they recognised at once as the Leg-Locker Curse. He must have had to bunny hop all the way up to Gryffindor Tower.	(HPAT:235)
HPMÜ	Wie er es geschafft hatte, durch das Porträtloch zu klettern, konnte man nur raten, denn seine Beine waren mit etwas zusammengeklebt, das sie sofort als den Beinschließerfluch erkannten. Er musste den ganzen Weg bis zum Gryffindor-Turm hoppeln .	
HPHÜ	Wie er es geschafft hatte, durch das Porträtloch zu klettern, war ihnen schleierhaft, denn seine Beine waren zusammengeklemmt, und sie erkannten sofort, dass es der Beinklammer-Fluch sein musste. Offenbar war er den ganzen Weg hoch in den Gryffindor-Turm gehoppelt wie ein Hase.	(HPHÜ:237)

Der im Ausgangstext fett markierte Ausdruck „must have had to bunny hop“ drückt aus, dass Neville („he“) aufgrund des Beinklammer-Fluches den Weg zum Gryffindor-Turm offenbar hatte hoppeln müssen. Es ist folglich eine Vorzeitigkeit gegeben, die aber in der HPMÜ nicht transportiert wird: Durch „musste“ als das Präteritum von „müssen“ wird eine Gleichzeitigkeit zum vorigen Satz ausgedrückt (sie erkannten den Fluch, und gleichzeitig musste Neville zum Turm hoppeln), was allerdings nicht logisch ist, denn in dem Moment, als die anderen Schüler_innen den Fluch erkennen, ist Neville schon im Turm, d. h., er musste vorher dorthin hoppeln. Folglich ist dies ein Fehler der Kategorie **Tempus und Modus von Verben**, der in der HPHÜ umgangen wurde: Die im Ausgangstext durch „must“ indizierte logische Folgerung (vgl. *Lexico.com* 2021a) wird durch „Offenbar“ ausgedrückt (vgl. *Duden* 2021), die Vorzeitigkeit durch das Plusquamperfekt von „hoppeln“.

Tonfall und Register ist die zehnte und letzte Fehlerkategorie, die mithilfe der HPMÜ entwickelt wurde. Als problematischer Bereich wird sie in Moorkens et al. (2018:254), Tezcan et al. (2019:45), Fonteyne et al. (2020:3795), Webster et al. (2020a:10f.), Kolb (2021) und teilweise in Schmidt (2021) genannt. Außerdem gibt es in allen drei in Kapitel 4.2 besprochenen Fehlertypologien korrespondierende Fehlerkategorien (vgl. Matusov 2019:15; TAUS o. J.; Tezcan et al. 2019:42). Die hier entwickelte Fehlerkategorie bezieht sich, wie in den oben genannten Werken, auf Fehler in der Wahl des **Registers** (z. B. Höflichkeitsform „Sie“ statt „du“; siehe Beispiel 2b) oder des **Tonfalls**. Hier ist v. a. die fehlende Anpassung in der HPMÜ an die mündliche Sprachverwendung bzw. an die informelle Sprechweise von Jugendlichen in den direkten Reden zu nennen (vgl. auch Schmidt 2021), die in der HPHÜ z. B. durch verkürzte Formen (besonders von „haben“; siehe Beispiele 2a, 3b, 4 und 6) oder unvollständige Sätze (siehe Beispiel 4) realisiert wird. Wurde in der HPHÜ in den Dialogen eine solche Anpassung vorgenommen, an der betreffenden Stelle der HPMÜ allerdings nicht (und es gab dort auch keinen anderen Fehler), wurde der Kategorie **Tonfall und Register** ein Fehler hinzugerechnet. Es ist v. a. dieser Umstand, durch den sich die relativ hohe Fehleranzahl in dieser Kategorie ergibt, allerdings existiert auch der umgekehrte Fall, in dem die HPMÜ zu informell für den Kontext ist:

Beispiel 10: Tonfall und Register

HPAT	<i>It also produces the Elixir of Life, which will make the drinker immortal.</i>	(HPAT:238; Kursivsetzung im Original)
HPMÜ	<i>Er produziert auch das Elixier des Lebens, das den Trinker unsterblich macht.</i>	(Kursivsetzung in der HPMÜ)
HPHÜ	Auch zeugt er das Elixier des Lebens, welches den, der es trinkt, unsterblich macht.	(HPHÜ:240)

Das fett markierte Wort „produces“ heißt in diesem Kontext „to make something or to bring something into existence“ (*Cambridge Dictionary* 2021). Das Verb „produzieren“ hat laut dem *Duden* (2021) ebenfalls diese Bedeutungsebene, allerdings ist sie als umgangssprachlich eingestuft und daher in diesem Ausschnitt unangebracht, da er offenbar aus einem Fachtext über Magie stammt. (Dies wird indiziert durch die Kursivsetzung, die gestalterische Hervorhebung des Ausschnitts durch die Einrückung und die Leerzeilen sowie durch den Umstand, dass Hermine sich das Buch, aus dem der Ausschnitt stammt, aus der Bibliothek ausgeliehen hat und der Inhalt von Ron offenkundig als schwierig eingestuft wird (vgl. HPAT:237f.)) Die Lösung in der Humanübersetzung ist allerdings auch nicht optimal, denn obwohl das **Register** passt, besteht ein Bedeutungsfehler, weil „zeugen“ laut dem *Duden* (2021) die oben genannte Bedeutung von „produce“ nicht wiedergibt.

Wie die obige Erklärung zeigt, konnten in dieser Kategorie doch einige Problemstellen eindeutig als Fehler bestätigt werden, bei anderen (wie der Verwendung von Höflichkeitsformen oder der Anpassung an die informelle Sprechweise) war dies jedoch nicht möglich, da auch die Herangehensweise an diese Aspekte einer Übersetzung wie bei der IUT subjektiv ist: Z. B. haben sich wiederum der deutsche Übersetzer und die spanische Übersetzerin der *Harry-Potter*-Bücher dazu entschieden, die enge Beziehung der drei Jugendlichen Harry, Ron und Hermine zu ihrem erwachsenen Freund Hagrid (der später sogar ihr Lehrer wird) mittels der Verwendung des informellen Pronomens „du“ zu verdeutlichen, während sich der französische Übersetzer für das formelle „vous“ entschied (vgl. Jentsch 2006:193). Jentschs (2006:193-196) weitere Ausführungen zeigen ebenfalls deutlich die Subjektivität der Herangehensweise an **Tonfall und Register**, aufgrund derer auch die Bewertung dieser Aspekte in einer Übersetzung als korrekt, problematisch oder als richtiger Fehler zumindest zum Teil subjektiv ist. Deswegen muss auch die Kategorie **Tonfall und Register** als Problembereich aufgefasst werden, jedoch in einem weitaus geringeren Ausmaß als die Kategorien Stil und IUT.

Diese zehn Fehlerkategorien bzw. Problembereiche stellen die Antwort auf FF1 dar. Sie werden nach den im Folgenden präsentierten Ergebnissen der Datenauswertung in Kapitel 6 dieser Arbeit noch in Beziehung zu den in der Einleitung genannten Hypothesen sowie zur Forschungsliteratur gesetzt.

5.2 Ergebnisse der Datenauswertung

In diesem Kapitel werden die Ergebnisse der in Kapitel 4.5 beschriebenen Datenauswertung präsentiert. Es erfolgt eine Analyse der beiden Posteditionen anhand der im vorigen Kapitel beschriebenen Fehlerkategorien und danach die Präsentation der Ergebnisse der Ratings.

5.2.1 Posteditionen

Die Gegenüberstellung der beiden Posteditionen im Hinblick auf die durchgeführten Korrekturen und die gerade vorgestellten Fehlerkategorien erfolgt überblicksartig in **Tabelle 4** auf der nächsten Seite. In der ersten Zeile nach den Überschriften ist die Gesamtzahl der Korrekturen angegeben, danach folgen die Korrekturen pro Fehlerkategorie (aufgelistet in derselben Reihenfolge wie in Kapitel 5.1). In der vorletzten Zeile sind unter „Nicht zuordenbar (NZ)“ Korrekturen von Problemstellen bzw. Fehlern aufgelistet, welche keiner der entwickelten Fehlerkategorien zugeordnet werden konnten, während in der letzten Zeile die Anzahl der Korrekturen angeführt ist, denen nach Ansicht der Verfasserin der vorliegenden Arbeit überhaupt keine Problemstellen oder Fehler zugrunde liegen und die demnach der persönlichen Vorliebe (PV) der Posteditorinnen entsprechen (siehe auch Kapitel 4.5 zu NZ und PV).

Zur **Interpretation der Tabelle** per se ist zu sagen, dass der höchste Wert bei jeder Fehlerkategorie (sei dieser nun bei der RRPEo oder der RRPem) nicht die Gesamtzahl der Problemstellen/Fehler dieser Kategorie in der RRMÜ angibt, sondern nur, wie viele solcher Problemstellen/Fehler in der entsprechenden Postedition korrigiert wurden. Die Gesamtzahl der Problemstellen/Fehler dieser Art in der RRMÜ kann also durchaus höher sein und entspricht auch nicht der Summe der Korrekturen in beiden Posteditionen zusammen. Dementsprechend ist ebenfalls zu beachten, dass eine gleiche oder ähnliche Anzahl an korrigierten Problemstellen/Fehlern in einer Kategorie nicht bedeutet, dass beide Posteditorinnen dieselben Problemstellen/Fehler in der RRMÜ korrigierten, sondern nur, dass sie insgesamt für (fast) gleich viele Problemstellen/Fehler dieser Art Korrekturen durchführten.

Es ist außerdem darauf hinzuweisen, dass, wie schon in einigen anderen Kapiteln erwähnt, die Subjektivität in der Auswertung eine große Rolle spielte: Obwohl selbstverständlich wie bei der Entwicklung der Fehlerkategorien (siehe Kapitel 4.2) durch den Rückgriff auf möglichst anerkannte Quellen und inhaltliche Argumentationen versucht wurde, so objektiv wie möglich vorzugehen, kann der Subjektivitätsfaktor nicht komplett aus der Gleichung genommen werden. Dies betrifft v. a. die Kategorie PV sowie die

Fehlerkategorien, die eher als Problembereiche aufzufassen sind, nämlich IUT, Stil sowie Tonfall und Register (siehe Kapitel 5.1 für Einzelheiten).

Tabelle 4: Anzahl Korrekturen pro Fehlerkategorie in der RRPEo und der RRPEm⁶⁴

Fehlerkategorie	Korrekturen RRPEo	Korrekturen RRPEm
Insgesamt	301	242
Auflösung von Anaphern	4	4
Auslassungen und Hinzufügungen	0	0
Bedeutung <ul style="list-style-type: none"> – Fixe Wendungen und <i>phrasal verbs</i> – Polysemie – Silbentrennung⁶⁵ 	39 – 2 (von 39) – 16 (von 39) – /	32 – 2 (von 32) – 15 (von 32) – /
Formatierung	0	0
Grammatik und Orthographie	3	3
Interpunktion	2 (+ 51)	9 (+ 52)
IUT	22	24
Stil	24	20
Tempus und Modus von Verben	1	1
Tonfall und Register	15	19
Nicht zuordenbar (NZ)	6	7
Persönliche Vorliebe (PV)	134	71

Wie an der Tabelle ersichtlich wird, gibt es in der von Posteditorin 1 angefertigten RRPEo 59 Korrekturen mehr als in der RRPEm von Posteditorin 2. Wird die Anzahl der PV-Korrekturen, also jener, die nach Ansicht der Verfasserin nicht unbedingt notwendig gewesen wären, von der Gesamtzahl der Korrekturen subtrahiert, ergibt sich für die RRPEo ein Wert von 167 und für die RRPEm einer von 171. Obwohl also Posteditorin 1 nicht über die Fehlerkategorien und damit die in literarischen DeepL-Übersetzungen vom Englischen ins Deutsche häufig vorkommenden Problemstellen und Fehlern informiert worden war, fand sie annähernd gleich viele Problemstellen und Fehler in der RRMÜ wie Posteditorin 2, die über diese Informationen verfügte.

In Bezug auf die einzelnen Fehlerkategorien präsentiert sich die Situation etwas differenzierter: Beide Posteditorinnen korrigierten gleich viele Problemstellen und Fehler bzgl. fünf von zehn Fehlerkategorien, bei zwei Kategorien korrigierte Posteditorin 1 mehr als

⁶⁴ Die Zählweise für die Kategorien Bedeutung und Interpunktion entspricht jener, die in Kapitel 5.1 angewendet wurde: Fehler bzw. Korrekturen der Unterkategorien fixe Wendungen und *phrasal verbs* sowie Polysemie wurden sowohl der jeweiligen Unterkategorie als auch der Oberkategorie Bedeutung zugeteilt; Korrekturen, die nicht in eine der Unterkategorien passten, wurden nur bei der Oberkategorie gezählt. Korrekturen von Anführungszeichenfehlern sind wiederum gesondert angeführt, um die Statistik nicht zu verfälschen.

⁶⁵ Im RRAT gibt es keine Silbentrennung, deswegen ist diese Unterkategorie hier nicht anwendbar.

Posteditorin 2, bei den restlichen drei ist es umgekehrt. Darüber hinaus gibt es zwar in der RRPEm von Posteditorin 2 mehr NZ-Korrekturen als in der RRPEo von Posteditorin 1, dafür aber fast doppelt so viele PV-Korrekturen in letzterer als in ersterer. Im Folgenden wird auf die Korrekturen der einzelnen Kategorien genauer eingegangen, und zwar in der Reihenfolge, in der die Kategorien in Tabelle 4 aufgelistet sind.

Die erste Kategorie ist die **Auflösung von Anaphern**. Diesbezüglich fanden beide Posteditorinnen dieselben 4 Fehler⁶⁶ in der RRMÜ, die sie auch auf exakt dieselbe Art und Weise korrigierten. Darunter fällt z. B. die Korrektur von „you = dich“ zu „euch“ in einer Rede der Herrscherin aus *Red Rising* an einige ihrer Untertanen. Die Korrekturen dieser Fehlerkategorie sind dementsprechend nicht auffällig, und das Wissen oder Nicht-Wissen um diese Art von Fehler scheint folglich keinen Einfluss auf die fertigen Posteditionen gehabt zu haben.

Dasselbe kann auch über die nächste Fehlerkategorie, **Auslassungen und Hinzufügungen**, gesagt werden: Beide Posteditorinnen identifizierten weder eine **Auslassung** noch eine **Hinzufügung**, durch die eine Verzerrung des Inhalts oder der zu übertragenden Bedeutung entstanden wäre.

In Bezug auf die Fehlerkategorie **Bedeutung** verhält sich die Situation deutlich anders, denn diese ist die erste von zwei Fehlerkategorien, in denen Posteditorin 1 zumindest bzgl. der Oberkategorie (die auch alle Korrekturen der beiden anwendbaren Unterkategorien enthält; siehe Fußnote 64 bzw. Kapitel 5.1) mehr Korrekturen durchführte als Posteditorin 2. Von diesen 39 bzw. 32 Korrekturen sind 8 identisch, d. h., sie beziehen sich auf denselben Fehler, der von beiden Posteditorinnen auf dieselbe Art und Weise korrigiert wurde, etwa die Änderung der falschen Übersetzung „miss = sich erinnern“ in „vermissen“. Weitere 14 Korrekturen der Oberkategorie **Bedeutung** beziehen sich in beiden Posteditionen ebenfalls auf dieselben Fehler, die Korrekturen dieser Fehler sind allerdings unterschiedlich, wie folgendes Beispiel zeigt:

⁶⁶ Um eine gute Lesbarkeit der Ausführungen zu garantieren und eine mögliche, durch nebeneinander stehende Zahlenwörter und numerische Werte (z. B. „vier und 17“) ausgelöste Verwirrung zu vermeiden, wird in diesem Kapitel entgegen der im Deutschen üblichen Regel, Zahlen bis zwölf auszuschreiben, bei der Angabe der Anzahl der Korrekturen der numerische Wert (der hier ohnehin im Vordergrund steht) verwendet (außer bei „eins“ und seinen Flexionsformen).

Beispiel 11a⁶⁷: Bedeutung – unterschiedliche Korrekturen desselben Fehlers

RRAT	We sing the Laureltide song as we go—a swooping melody of a man finding his bride in a field of gold.	(RRAT:[26])
RRMÜ	Wir singen das Laureltide-Lied, während wir gehen - eine schwingende Melodie über einen Mann, der seine Braut in einem Feld voller Gold findet.	
RRPEo	Unterwegs singen wir das Lied des Laurels - eine ausgelassene Melodie über einen Mann, der seine Braut in einem Feld voller Gold findet.	
RRPEm	Wir singen während des Gehens das Lorbeerflut-Lied – eine schwungvolle Melodie über einen Mann, der seine Braut in einem Feld voller Gold findet.	
RRHÜ	Unterwegs singen wir das Lorbeerzeit-Lied – eine mitreißende Melodie über einen Mann, der seine Braut in einem goldenen Feld findet.	(RRHÜ:[28])

Das Verb „swoop“ bedeutet „to move very quickly and easily through the air“ (*Cambridge Dictionary* 2021), im RRAT wird es in der *ing*-Form allerdings adjektivisch und in übertragener Bedeutung gebraucht. Es ist dementsprechend eine leichte, schnelle und vermutlich fröhliche Melodie gemeint, passend zur Gelegenheit der Preisverleihung, von der im RRAT in diesem Kontext die Rede ist. Keine der im *Duden* (2021) angeführten Definitionen des Verbs „schwingen“ (und damit des in der RRMÜ verwendeten Partizips I „schwingend“) stimmen mit der Bedeutung von „swoop“ überein, weswegen das Partizip auch in übertragener Bedeutung nicht unbedingt Fröhlichkeit oder Schnelligkeit evoziert. Zumindest waren offenbar die beiden Posteditorinnen dieser Ansicht, denn sie ersetzten das Partizip einerseits durch „ausgelassen“ (was Fröhlichkeit zum Ausdruck bringt) und andererseits durch „schwungvoll“ (was eher in Richtung Kraft und Schnelligkeit geht) (vgl. *Duden* 2021).

Neben solchen unterschiedlichen Korrekturen desselben Fehlers führten beide Posteditorinnen auch Korrekturen von **Bedeutungsfehlern** durch, die von der jeweils anderen Posteditorin nicht entdeckt (oder ignoriert) und dementsprechend auch nicht korrigiert wurden. In der RRPEo kommen 17 solcher Korrekturen bzgl. der **Bedeutung** vor, in der RRPEm 11. Auffällig ist hierbei, dass, während sich in der RRPEm alle 11 dieser Korrekturen

⁶⁷ Die Nummerierung der Beispiele wird aus dem vorangehenden Kapitel fortgesetzt, um bei Verweisen auf diese Beispiele Verwirrung bei den Leser_innen zu vermeiden. Wie im vorigen Kapitel wurden **fette** Hervorhebungen in den Beispielen von der Verfasserin der vorliegenden Arbeit vorgenommen, um die jeweilige Korrektur besser zu veranschaulichen. Die anderen im Beispiel befindlichen Korrekturen wurden aus der Postedition übernommen, sind aber nicht hervorgehoben. Für etwaige Verweise auf diese Beispiele aus dem nächsten Kapitel wird zwar auch die Humanübersetzung in den Beispielen angeführt, darauf eingegangen wird aber im Rahmen dieses Kapitels nicht (deswegen ist die Humanübersetzung in grauer Schrift ohne fette Hervorhebungen gehalten).

auf Einzelwörter und fixe Mehrwortausdrücke beziehen (siehe etwa die Übersetzung des polysemischen Verbs „drone (on)“ in Beispiel 12), dies bei der RRPEo nicht der Fall ist: 4 der 17 Korrekturen, die nur Posteditorin 1 durchführte, beziehen sich nämlich auf nicht fixierte Wortsequenzen/Phrasen, wie im folgenden Beispiel (und in Beispiel 16 in Kapitel 5.2.2):

Beispiel 11b: Bedeutung – Korrektur nur in der RRPEo

RRAT	Two soldiers of the Obsidian Color, colossal men and women nearly twice my size, are shown along with nimble Yellow doctors carrying several victims from the blast.	(RRAT:[26])
RRMÜ	Zwei Soldaten der Obsidianfarbe, kolossale Männer und Frauen, fast doppelt so groß wie ich, werden zusammen mit flinken gelben Ärzten gezeigt, die mehrere Opfer aus der Explosion tragen.	
RRPEo	Zwei Soldaten der Obsidianfarbe, kolossale Männer und Frauen, fast doppelt so groß wie ich, tragen zusammen mit agilen Gelben Ärzten mehrere Opfer der Explosion.	
RRPEm	Zwei Soldaten der Obsidian-Farbe, riesige Männer und Frauen, fast doppelt so groß wie ich, werden zusammen mit flinken Gelben Ärzten gezeigt, die mehrere Opfer aus der Explosion tragen.	
RRHÜ	Zwei Soldaten der Obsidian-Farbe, riesige Männer und Frauen, die fast doppelt so groß sind wie ich, bewegen sich zwischen flinken Ärzten der Gelben, die mehrere Opfer vom Schauplatz des Bombenanschlags wegbringen.	(RRHÜ:[28])

Im vorliegenden Beispiel wird vom Hauptcharakter eine Art Sendung bzw. Fernsehübertragung beschrieben, in der die Auswirkungen eines Bombenanschlags gezeigt werden. Der fett markierte Teil der RRMÜ bedeutet, dass Opfer aus der Explosion getragen werden, während diese stattfindet, was selbstverständlich schwer möglich ist und dementsprechend auch nicht der Bedeutung von „carrying victims from the blast“ entspricht. Damit ist eher gemeint, dass Opfer der Explosion getragen bzw. vom Explosionsort weggetragen werden, jedenfalls ist aber die Explosion zu diesem Zeitpunkt schon vorbei. Posteditorin 1 erkannte diesen Fehler in der RRMÜ und besserte ihn aus, im Gegensatz zu Posteditorin 2, wobei allerdings auch anzumerken ist, dass die Korrektur in der RRPEo nicht optimal ist, da die Information, die durch das Verb „shown“ ausgedrückt wird, verloren geht. Andererseits könnte es auch die Absicht der Posteditorin gewesen sein, „shown“ auszulassen, und zwar mit dem Argument, dass die Leser_innen sowieso wissen (sollten), dass von einer Sendung die Rede ist, da der ganze Absatz des RRAT, aus dem das Beispiel stammt, von dieser Sendung handelt (vgl. RRAT:[26]).

Jedenfalls korrigierte Posteditorin 1, wie oben erwähnt, 4 solche **Bedeutungsfehler**, Posteditorin 2 allerdings gar keinen. Ein möglicher Grund dafür, abgesehen von den Unter-

schieden zwischen den Posteditorinnen an sich (Erfahrungshintergrund, Arbeitsweise, subjektive Einschätzung der Korrektheit einer Übersetzung etc.), könnte in der Beschreibung der Fehlerkategorie **Bedeutung** liegen, die Posteditorin 2 zur Verfügung gestellt wurde: Die Fehler dieser Kategorie wurden nur anhand von Einzelwörtern und fixen Mehrwortausdrücken exemplarisch dargestellt (siehe Beispiele 3a-3d in Kapitel 5.1), Wortsequenzen wie die oben angeführte waren in den Beispielen nicht enthalten, was Posteditorin 2 dazu verleitet haben könnte, nur nach **Bedeutungsfehlern**, die jenen in den Beispielen ähnelten, Ausschau zu halten. Damit könnte von einem wie in Kolb (2021) diskutierten *priming effect* gesprochen werden (siehe Kapitel 3), nur dass dieser Effekt hier nicht von der maschinellen Übersetzung ausgeht, sondern von der Beschreibung der Fehlerkategorien.

Die obige Beschreibung der Korrekturen der Fehlerkategorie Bedeutung geht von der Oberkategorie aus, d. h., sie umfasst nicht nur alle Korrekturen, die nur der Oberkategorie zugeordnet wurden, sondern auch all jene der zwei anwendbaren Unterkategorien. Werden nur die Unterkategorien in Betracht gezogen, ist die Situation zwischen den beiden Posteditionen ausgeglichener. In Bezug auf **fixe Wendungen und phrasal verbs** gibt es in beiden Texten jeweils 2 Korrekturen: Eine davon bezieht sich auf einen Fehler, der von beiden Posteditorinnen erkannt und auf dieselbe Art und Weise korrigiert wurde („buck off = ignorieren“ statt „abwenden“). Außerdem fand und korrigierte jede der Posteditorinnen einen Fehler dieser Kategorie, für den die jeweils andere Posteditorin keine Korrektur durchführte (siehe Beispiel 11c).

Dahingegen fanden die beiden Posteditorinnen sehr viel mehr Fehler in der RRMÜ bzgl. **polysemischer** englischer Wörter und Ausdrücke, wobei Posteditorin 1 einen mehr korrigierte als Posteditorin 2. Es gibt in beiden Posteditionen Korrekturen von **Polysemie-Fehlern**, die in der jeweils anderen Postedition nicht korrigiert wurden (5 in der RRPEo und 4 in der RRPEm), wie die oben erwähnte Korrektur in Beispiel 12 von „drone (on) = leiern“ statt „dröhnen“ in der RRPEm. Darüber hinaus wurden 11 **Polysemie-Fehler** von beiden Posteditorinnen gefunden, 2 davon wurden auch auf dieselbe Art und Weise korrigiert (etwa „drill = bohren“ statt „trainieren“), für die anderen 9 führten die Posteditorinnen unterschiedliche Korrekturen durch, wie im folgenden Beispiel:

Beispiel 11c: Bedeutung – Unterkategorien – unterschiedliche Korrekturen + Korrektur nur in der RRPE_m

RRAT	The holoCan above me shows images of mankind's struggles, which are then followed by soaring music as the Society's triumphs flash past .	(RRAT:[22f.])
RRMÜ	Die HoloCan über mir zeigt Bilder von den Kämpfen der Menschheit, die dann von steigender Musik gefolgt werden, während die Triumphe der Gesellschaft vorbeiziehen .	
RRPE _o	Die HoloCan über mir zeigt Bilder von den Kämpfen der Menschheit, gefolgt von triumphaler Musik, als die Erfolge der Gesellschaft vorbeiziehen .	
RRPE _m	Die HoloBüchse über mir zeigt Bilder von Menschheitskämpfen, die dann von tosender Musik übertönt werden, als die Triumphe der Gesellschaft im Schnelldurchlauf gezeigt werden .	
RRHÜ	Über mir zeigt die HoloBox Szenen aus den Kämpfen der Menschheit und die Triumphe der Weltengesellschaft. Dazu spielt erhebende Musik.	(RRHÜ:[24])

In diesem Beispiel gibt es einerseits einen **Polysemie**-Fehler (bzgl. „soaring“) und einen, der sich auf ein *phrasal verb*, nämlich „flash past“, bezieht. Der **Polysemie**-Fehler besteht darin, dass das Verb „soar“ zwar „steigen“ bedeuten kann, in diesem Fall aber wohl eher „to rise to a majestic stature“ ([MWD] 2021) gemeint ist, was der Bedeutungsinhalt von „steigen“ nicht abdeckt (vgl. *Duden* 2021). Dies wurde von beiden Posteditorinnen erkannt, aber auf unterschiedliche Art und Weise korrigiert: Posteditorin 1 wählte das Adjektiv „triumphal“, Posteditorin 2 „tosend“. Beide Wörter erscheinen von ihrer Bedeutung her ebenfalls nicht angemessen, allerdings passt „triumphal“ deutlich besser zu dem durch „soaring“ und „triumphs“ entstehenden Eindruck des Ruhms als „tosend“, das eher an einen Sturm erinnert.

Das *phrasal verb*, dessen Übersetzung von Posteditorin 2 im Gegensatz zu Posteditorin 1 korrigiert wurde, bedeutet „to move somewhere fast“ (*Cambridge Dictionary* 2021) und bezieht sich hier in übertragener Bedeutung auf Bilder oder Videoausschnitte, die auf einer Art Bildschirm (der im Beispiel erwähnten „holoCan“) gezeigt werden. Das von DeepL verwendete Verb „vorbeiziehen“ als Übersetzung für „flash past“ ist problematisch, weil es nicht unbedingt mit Schnelligkeit in Verbindung gebracht werden muss, wie die Definitionen im *Duden* (2021) zeigen. Deswegen entschied sich vermutlich Posteditorin 2 dazu, den entsprechenden Teil des Satzes umzuformulieren und mit dem Wort „Schnelldurchlauf“ die Schnelligkeit, mit der die Bilder oder Ausschnitte aufeinanderfolgen, mehr zu betonen.

Bei der Fehlerkategorie **Formatierung** waren sich die Posteditorinnen wiederum einig, denn beide führten diesbezüglich überhaupt keine Korrekturen durch, was einfach zu erklären

ist: Es gibt zwar kursiv formatierte Textstellen im Ausgangstext (vgl. RRAT:[23]), diese wurden allerdings alle unverändert in die RRMÜ übertragen (siehe etwa Beispiel 17 in Kapitel 5.2.2) und bedurften somit nach Ansicht der Posteditorinnen keiner Korrekturen. Diesbezüglich ist zu bemerken, dass die kursiven Textstellen im RRAT ganze Absätze sind und keine einzelnen Wörter, bei deren **Formatierungsübertragung** der DeepL-Übersetzer teilweise mehr Probleme zu haben scheint, wie Beispiel 4 in Kapitel 5.1 zeigt.

Überwiegend einig waren sich die beiden Posteditorinnen auch bzgl. der Kategorie **Grammatik und Orthographie**, in der sie dieselben 3 Fehler fanden, 2 davon aber unterschiedlich korrigierten. Eine dieser unterschiedlichen Korrekturen ist in Beispiel 11c ersichtlich und bezieht sich auf die Phrase „die dann von steigender Musik gefolgt werden“. Das Verb „folgen“ in der RRMÜ ist inhaltlich korrekt, allerdings wird es in dieser Satzkonstruktion falsch verwendet: Es ist **grammatikalisch** inkorrekt zu sagen, dass jemand oder etwas gefolgt wird, da das Passiv des intransitiven Verbs „folgen“ kein Subjekt haben kann (vgl. Fabricius-Hansen ⁹2016:558). Posteditorin 1 korrigierte diesen Fehler mittels einer Umstellung des Satzbaus, sodass „gefolgt“ eine attributive Funktion hat. Dahingegen ersetzte Posteditorin 2 „gefolgt“ durch „übertönt“, was zwar einen **grammatikalisch** korrekten Satz ergibt, allerdings inhaltlich nicht dem RRAT entspricht. Der Fehler, den beide Posteditorinnen auf die gleiche Art und Weise korrigierten, bezieht sich auf die **Orthographie** des Planetennamens „Mars“: Bei diesem fehlt in der RRMÜ der letzte Buchstabe, der von beiden Posteditorinnen richtigerweise in den Posteditionen angehängt wurde.

Die Fehlerkategorie **Interpunktion** ist die erste der drei Kategorien, in denen Posteditorin 2 mehr Fehler korrigierte als Posteditorin 1, und zwar in Bezug auf Anführungszeichenfehler sowie jene **Interpunktionsfehler**, die andere Zeichen betreffen. Die Anzahl der Korrekturen von Anführungszeichenfehlern ist in beiden Posteditionen sehr hoch, da der DeepL-Übersetzer auch hier die Anführungszeichen systematisch falsch setzte: Es wurden zwar doppelte Anführungszeichen verwendet, aber diese waren stets oben (statt unten und oben) gesetzt und außerdem gerade (") statt typographisch („“) (siehe Beispiel 13 in diesem Kapitel und Beispiel 18 in Kapitel 5.2.2). Beide Posteditorinnen korrigierten diese Fehler, wobei Posteditorin 1 einen davon übersehen zu haben scheint, sodass Posteditorin 2 eine Korrektur mehr in diesem Bereich durchführte. Es kann angenommen werden, dass Posteditorin 2, die aufgrund der Beschreibung der Fehlerkategorien über die Häufigkeit und Systematik der Anführungszeichenfehler in DeepL-Übersetzungen Bescheid wusste, in der RRMÜ gezielt nach diesen Fehlern suchte und sie ausbesserte, was bei Posteditorin 1 (zumindest zu Beginn ihres Posteditings) wahrscheinlich nicht der Fall war.

Von den 51 bzw. 52 Korrekturen von Anführungszeichenfehlern sind allerdings 2 bemerkenswert, da sie in den Posteditionen unterschiedlich ausfielen: Bei der betreffenden direkten Rede handelt es sich um die oben erwähnte Rede der Herrscherin aus *Red Rising*, die über drei Absätze hinweg durchgehend wiedergegeben wird. Um dies zu kennzeichnen, ist im RRAT am Anfang eines jeden Absatzes ein Anführungszeichen gesetzt und am Ende des letzten Absatzes (vgl. RRAT:[23]). Der DeepL-Übersetzer hat diese Setzung der Anführungszeichen übernommen, genauso wie Posteditorin 2 (die nur die üblichen Verbesserungen von oben zu unten gesetzten und von geraden zu typographischen Anführungszeichen durchführte). Diese Vorgehensweise ist allerdings als inkorrekt einzustufen, denn in Band 9 der *Duden*-Reihe (Hennig et al. ⁸2016:69) wird ohne Ausnahmen angeführt, dass Anführungszeichen „bei der direkten Rede [...] am Anfang und am Ende der Aussage“ stehen, was Posteditorin 1 richtig erkannte: Sie löschte die Anführungszeichen am Anfang des zweiten und des dritten Absatzes komplett und behielt damit nur jene am Anfang des ersten und am Ende des letzten Absatzes bei.

Wie oben schon erwähnt, führte Posteditorin 2 auch mehr Korrekturen anderer **Interpunktions**fehler durch als Posteditorin 1, darunter fehlende Kommata nach direkten Reden (die von Posteditorin 1 übersehen wurden) sowie fälschlicherweise gesetzte Bindestriche, die korrekterweise Gedankenstriche sein sollten. Die letztgenannten Fehler, die vom DeepL-Übersetzer ebenfalls systematisch begangen und in der Beschreibung der Fehlerkategorien durch Beispiele veranschaulicht wurden, wurden alle von Posteditorin 2 (die demnach von diesen Fehlern wusste) korrigiert (siehe Beispiel 11a), während Posteditorin 1 nur einen (von 6) dieser Fehler ausbesserte. Es ist allerdings anzumerken, dass auch diese eine Änderung eines Bindestrichs in ein Komma von Posteditorin 1 vermutlich nicht einem Bewusstsein für diese Fehler geschuldet ist (denn sonst hätte sie höchstwahrscheinlich alle diese Fehler ausgebessert), sondern eher einer anderen Korrektur, die ebenfalls im betreffenden Satz von dieser Posteditorin durchgeführt wurde. Demnach könnte auch hier von einem *priming effect* der Kategorie Interpunktion auf Posteditorin 2 gesprochen werden, der allerdings nicht nur positiv einzustufen ist (siehe Kapitel 5.2.2).

Auch in Bezug auf die **IUT** führte Posteditorin 2 mehr Korrekturen durch als Posteditorin 1. Von den 22 bzw. 24 **IUT**-Korrekturen in den Posteditionen sind 9 identisch und betreffen die Großschreibung der Farben (die Gesellschaftsklassen in *Red Rising* bezeichnen), wie z. B. „Yellow = Gelb“ statt „gelb“ in Beispiel 11b (siehe auch Beispiel 15 in Kapitel 5.2.2). Weitere 7 **IUT**-Problemstellen wurden von beiden Posteditorinnen erkannt, allerdings unterschiedlich ausgebessert, etwa „Laureltide song = Lied des Laurels“ (RRPEo)

bzw. „Lorbeerflut-Lied“ (RRPEm) in Beispiel 11a. An diesen beiden unterschiedlichen Übersetzungen wird aufgrund der Herangehensweise an „Laurel“ (eine Art Preis im Rahmen von *Red Rising*) ebenfalls ersichtlich, dass es auch beim Problembereich der **IUT** Korrekturen gab, die eine Posteditorin durchführte, die andere jedoch nicht und umgekehrt: Im Falle von Posteditorin 1 sind das 6, bei Posteditorin 2 sind es 8, darunter die Änderung in der RRPEm von „Laurel“, das der DeepL-Übersetzer aus dem Englischen übernahm, in „Lorbeere“.

Was diese eher statistische und nüchterne Auswertung der **IUT**-Korrekturen nicht bzw. nur ansatzweise zeigt, ist die schon in Kapitel 5.1 erwähnte unterschiedliche Herangehensweise an die Übersetzung dieser Termini: Während beide Posteditorinnen die von DeepL ihrer Meinung nach richtig übersetzten **IUT** wie „Color = Farbe“ (siehe Beispiel 11b) oder „Society = Gesellschaft“ (siehe Beispiel 11c) übernahmen und bei den Namen der einzelnen Farben nur geringfügige Verbesserungen vornahmen (wie die oben erwähnte Groß- statt Kleinschreibung oder die Hinzufügung eines Bindestriches; siehe „Obsidian-Farbe“ in der RRPEm in Beispiel 11b), gingen sie mit den 12 vom DeepL-Übersetzer aus dem Englischen ins Deutsche übernommenen **IUT** unterschiedlich um: Posteditorin 1

- veränderte 7 dieser Übernahmen nicht (siehe „HoloCan“ in Beispiel 11c, „Frysuit“ in Beispiel 14),
- fand für 2 eine eigene Übersetzung (siehe „Webbery = Weberei“ siehe Beispiel 18 in Kapitel 5.2.2 und die oben erwähnte Übersetzung „Laureltide song = Lied des Laurels“),
- deutschte 2 Übernahmen durch geringfügige Änderungen ein (das vom DeepL-Übersetzer übernommene *collective noun* „Aureate“ bekam eine deutsche Pluralendung (siehe Beispiel 12), und das ebenfalls übernommene „Obsidians“ (im Englischen als Plural zu „Obsidian“, das einen Menschen bezeichnet, der dieser Farbe angehört) wurde zu „Mitglieder der Obsidiankaste“)
- und ersetzte einen **IUT** durch einen anderen (nämlich „Goldbrows“ durch „Goldene“ in Beispiel 12).

In der RRPEo mangelt es folglich dem Umgang mit den **IUT** etwas an Konsistenz, v. a. auch, da diese Mischung von englischen und deutschen Termini im Text eher willkürlich zu sein scheint. Dies könnte sich störend auf den Lesefluss auswirken und v. a. bei Leser_innen mit weniger guten Englischkenntnissen das Verständnis behindern. Im Gegensatz dazu ging Posteditorin 2 merklich anders mit den **IUT** um: Sie

- überlegte sich für 10 der vom DeepL-Übersetzer übernommenen **IUT** eine eigene Übersetzung (siehe etwa „holoCan = HoloBüchse“ in Beispiel 11c, „Goldbrows = Goldrüben“ in Beispiel 12 und „frysuit = Schmoranzug“ in Beispiel 14)
- und beließ nur 2 der Übernahmen unverändert (nämlich „Aureate“ in Beispiel 12 und „Obsidians“).

Diese unterschiedliche Herangehensweise ist, wie auch oben schon erwähnt, selbstverständlich vom persönlichen Übersetzungsstil der beiden Posteditorinnen und ihren Vorlieben bzgl. Übernahme oder Übersetzung solcher **IUT** geprägt. Es ist allerdings anzunehmen, dass Posteditorin 2, obwohl sie laut eigener Aussage „mit der In-Universe-Terminologie ziemlich gekämpft [hat]“ (Posteditorin 2 2021)⁶⁸, durch die ihr zur Verfügung gestellte Erklärung des Problembereichs **IUT** noch zusätzlich dazu angeregt wurde, für fast alle **IUT** deutsche Entsprechungen zu finden, da auch in den Beispielen in dieser Erklärung die betreffenden **IUT** vom Humanübersetzer ins Deutsche übertragen worden waren (siehe Beispiel 7 in Kapitel 5.1). Dies würde wiederum für einen von der Beschreibung der Fehlerkategorien ausgehenden *priming effect* sprechen.

Darüber hinaus ist zur **IUT**-Übersetzung noch zu sagen, dass es in beiden Posteditionen diesbezüglich Unstimmigkeiten gibt, die verdeutlichen, wie schwierig dieser Bereich einer Übersetzung ist und wie viel dabei beachtet werden muss (siehe auch Kapitel 5.1 dazu). Je 4 solche Unstimmigkeiten in beiden Posteditionen werden im Folgenden behandelt, jeweils beginnend mit den **IUT** in Beispiel 12:

Beispiel 12: IUT: unterschiedliche Korrekturen derselben Problemstelle + Korrektur nur in der RRPEo

RRAT	Lune drones on above us on the HC, high above, as the Goldbrows (Aureate to be technic) ought to be.	(RRAT:[26])
RRMÜ	Lune dröhnt über uns auf dem HC, hoch oben, wie es sich für die Goldbrows (Aureate , um technisch zu sein) gehört.	
RRPEo	Lune dröhnt aus der HC, hoch über uns, wie es sich für die Goldenen (Aureaten , um genau zu sein) gehört.	
RRPEm	Lune leiert über uns auf dem HC, hoch oben, wie es sich für die Goldrüben (Aureate , um es technisch auszudrücken) gehört.	
RRHÜ	Hoch oben, wie es sich für die Goldenen gehört, plappert Lune in der HB weiter.	(RRHÜ:[28])

⁶⁸ Die Aussagen der Posteditorinnen zu ihrem Postediting wurden im Rahmen der vorliegenden Masterarbeit nicht verlangt, d. h. ohne Aufforderung dazu an die Verfasserin dieser Arbeit übermittelt, bzw. sind Antworten auf Fragen, die die Verfasserin noch vor dem Postediting gestellt hat, um die Ähnlichkeit zwischen den möglichen Posteditor_innen festzustellen (siehe Kapitel 4.3). Deswegen und weil nur wenige sich direkt auf das Postediting beziehende Aussagen vorliegen, wurden diese Aussagen nicht separat ausgewertet, sondern sind nur an geeigneten Stellen angeführt.

Dieses Beispiel zeigt zunächst eine dieser Unstimmigkeiten in der RRPEo, nämlich „Goldbrows = Goldene“. Die englische Bezeichnung ist ein Slang-**Terminus**, der von den Roten, um die es in der entsprechenden Szene geht, etwas abwertend für die herrschende Klasse der Gesellschaft von *Red Rising* verwendet wird (vgl. [RRW] o. J.a). Die offizielle Bezeichnung dieser herrschenden Klasse heißt „Golds = Goldene“, d. h., für die zwei **Termini** wird in der RRPEo dieselbe Übersetzung benutzt. Diese Gleichsetzung der beiden **Termini** ist allerdings nach Ansicht der Verfasserin als normalisierend und der Szene abträglich zu werten, da dadurch in der RRPEo (im Gegensatz zur RRPEm) die Charakterisierung von „Goldbrows“ als Slang-Ausdruck und damit ein Aspekt der komplexen Beziehung der Roten zu den Goldenen verloren geht.

Außerdem wurde in der RRPEo, wie oben erwähnt, der IUT „Webbery“, der den Arbeitsplatz der Roten Frauen beschreibt und wo ein Stoff namens „bioSilk“ geerntet wird (vgl. RRAT:[24]), mit „Weberei“ übersetzt (siehe Beispiel 18 in Kapitel 5.2.2). Dies mag auf den ersten Blick aufgrund der Ähnlichkeit der beiden Wörter wie eine angemessene Übersetzung erscheinen, aber der Umstand, dass die in der „Webbery“ geerntete „bioSilk“ von sogenannten „Spiderworms“ produziert wird (vgl. [RRW] o. J.b), legt nahe, dass der Wortstamm des Terminus, nämlich „web“, tatsächlich in etwa „[a] filmy network spun by some insect larvae“ (*Lexico.com* 2021a) bedeutet und damit eher dem Deutschen „Netz“ entspricht. Deswegen und weil im RRAT eindeutig von „ernten“ („harvest“), also keiner produzierenden (wie „weben“), sondern einer sammelnden Tätigkeit die Rede ist (vgl. RRAT:[24]), scheint die Übersetzung von Posteditorin 2, nämlich „Netzerei“, angemessener zu sein, zumal dieses Wort, im Gegensatz zu „Weberei“, erfunden ist, genauso wie „Webbery“ im Ausgangstext.

Eine weitere Unstimmigkeit in der RRPEo bezieht sich auf „nervenucleic“, eine Art Medikament, um übermäßige Narbenbildung zu verhindern (vgl. RRAT:[24]; Brown 2014a:[48]). Dieser IUT wurde vom DeepL-Übersetzer mit „Nervennahrung“ übersetzt, allerdings änderte Posteditorin 1 diese Übersetzung auf „Nervennucleic“. Diese eingedeutschte Version von „nervenucleic“ ist problematisch, da die Eindeutschung nur unvollständig vorgenommen wurde: „nerve“ wurde übersetzt mit „Nerven“, allerdings hätte für eine vollständige Eindeutschung „nucleic“ evtl. ein „k“ statt eines „c“, aber zumindest eine deutsche Endung, beispielsweise „-in“, bekommen müssen (etwa in Anlehnung an das Wort „Nukleinsäure“). Posteditorin 2 nahm ebenfalls eine Eindeutschung vor, vergab aber dabei die Endung „-id“. Somit lautet der Terminus lautet bei ihr „Nervennucleid“, was etwas angemessener erscheint als „Nervennucleic“ in der RRPEo.

Schließlich wurde in der RRPEo unterschiedlich mit „Obsidian Color“ und „Obsidians“ umgegangen: Die Übersetzung für „Obsidian Color“ lautet nämlich „Obsidianfarbe“, jene für „Obsidians“, wie oben erwähnt, „Mitglieder der Obsidiankaste“. Beide Übersetzungen sind an sich korrekt (erstere wirkt allerdings mit einem Bindestrich („Obsidian-Farbe“) wie in der RRPEm klarer; siehe Beispiel 11b), aber die Verwendung von „-kaste“ erscheint fragwürdig, da es dadurch der Postedition an Konsistenz mangelt und „-kaste“ ohnehin nicht notwendig wäre, da schließlich der Terminus „Color = Farbe“ genau dafür steht.

Wie oben erklärt, gibt es auch in der RRPEm solche Unstimmigkeiten bei den **IUT**-Übersetzungen, darunter die unveränderte Übernahme von „Aureate“, wie in Beispiel 12 oben ersichtlich. Dieses englische Wort lässt sich ohne eine deutsche Pluralendung wie in der RRPEo nur schwer auf flüssige und natürliche Art und Weise im Deutschen verwenden. Dasselbe gilt für die zweite unveränderte Übernahme („Obsidians“): Dieser **Terminus** bezeichnet, wie oben erklärt, spezifische Mitglieder der Obsidian-Farbe, ist also im Englischen die Pluralform von „Obsidian“. Die ins deutsche übernommene Pluralbildung mit „s“ ist fragwürdig, da im Deutschen nur Fremd- und Kurzwörter, Eigennamen sowie auf Vokal endende Substantive oder solche aus dem Niederdeutschen das Plural mit „s“ bilden, wenn von umgangssprachlicher Pluralbildung abgesehen wird (vgl. Hennig et al. ⁸2016:729) und „Obsidian“ als grundsätzlich existierendes deutsches Wort in keine dieser Gruppen fällt. Dementsprechend erscheint die Umschreibung von Posteditorin 1 angemessener, aber es wäre auch eine Pluralbildung wie „die Obsidianen“ denkbar.

Außerdem ist in der RRPEm der Umgang mit dem **IUT** „holoCan“ und seiner Abkürzung „HC“ als Unstimmigkeit zu werten. Posteditorin 2 übersetzte nämlich „holoCan“ mit „HoloBüchse“, versäumte es aber, „HC“ in der Übersetzung entsprechend auf „HB“ zu ändern. Bei Leser_innen, die mit der englischen Version des **Terminus** nicht vertraut sind, würde dies vermutlich Verwirrung stiften. Posteditorin 1 hatte dieses Problem nicht, da sie „HoloCan“ in der RRPEo beließ.

Das vierte und letzte zu besprechende Beispiel einer Unstimmigkeit in der RRPEm ist die Übersetzung „symbol of Gold = Symbol des Goldes“: Das Symbol ziert die Handrücken der Herrscherin aus *Red Rising* und zeigt ihre Zugehörigkeit zur herrschenden Klasse, also zu den oben erwähnten Goldenen, an. Durch die Übersetzung „des Goldes“ wird allerdings suggeriert, dass auf ihren Händen ein Symbol prangt, das für das Metall oder auch das chemische Element Gold steht und nicht für die Farbe der Goldenen. Dies erkannte

Posteditorin 1 im Gegensatz zu Posteditorin 2 und änderte die RRMÜ entsprechend auf „Symbol der Goldenen“.⁶⁹

Damit kann zur Besprechung der Korrekturen in der nächsten Kategorie übergegangen werden, dem **Stil**. Dieser ist die zweite der beiden Kategorien, in denen Posteditorin 1 mehr Korrekturen durchführte als Posteditorin 2. Da die Entscheidung, ob Korrekturen diesem Problembereich zugeordnet werden können oder nicht, noch subjektiver ist als jene bzgl. der anderen Fehlerkategorien, wurden Korrekturen nur als **Stil**-Korrekturen gewertet, wenn auch eine möglichst fundierte Begründung dafür geliefert werden konnte (siehe Kapitel 5.1 zum Subjektivitätsfaktor bei der Kategorie **Stil** und für ein Beispiel einer solchen Begründung) und/oder wenn beide Posteditorinnen die RRMÜ an derselben Stelle korrigierten und die entsprechenden Korrekturen nicht einer der anderen Kategorien (inkl. NZ und PV) zugeordnet werden konnten. Dementsprechend ist auch der Großteil, nämlich 18 von 24 (RRPEo) bzw. 20 (RRPEm), der korrigierten **Stil**-Problemstellen von beiden Posteditorinnen entdeckt worden. Von diesen 18 Problemstellen wurden 3 auf genau dieselbe Art und Weise korrigiert, für die restlichen 15 fielen die Korrekturen unterschiedlich aus (siehe auch Beispiel 12 bei den Übersetzungen für „to be technic“):

Beispiel 13: Stil – identische Korrekturen + unterschiedliche Korrekturen derselben Problemstelle

RRAT	I find the kitchen room of the home empty, but I hear Eo in the bedroom. “ Stop right where you are! ” she commands through the door.	(RRAT:[23])
RRMÜ	Ich finde den Küchenraum des Hauses leer, aber ich höre Eo im Schlafzimmer. “ Bleiben Sie stehen, wo Sie sind! “, befiehlt sie durch die Tür.	
RRPEo	Die Küche des Hauses ist leer, aber ich höre Eo aus dem Schlafzimmer. „ Bleib sofort stehen! “, hallt ihr Befehl durch die Tür.	
RRPEm	Die Küche des Hauses ist leer, aber ich höre Eo im Schlafzimmer. „ Bleib, wo du bist! “, befiehlt sie mir durch die Tür.	
RRHÜ	Ich sehe niemanden in der Küche unserer Wohnung, aber ich höre Eo im Schlafzimmer. »Bleib, wo du bist!«, befiehlt sie durch die Tür.	(RRHÜ:[25])

⁶⁹ Die Tatsache, dass es diese gerade besprochenen Unstimmigkeiten in den Posteditionen gibt, bedeutet selbstverständlich nicht, dass es nicht auch sehr gut gelungene **IUT**-Übersetzungen in den Posteditionen gibt. Darunter fällt nach Ansicht der Verfasserin dieser Arbeit etwa „frysuit = Schmoranzug“ in der RRPEm: Ein „frysuit“ ist eine Art Anzug aus Plastik, in dem es sehr heiß wird und der von speziellen Minenarbeitern (nur Männern), im RRAT bzw. in der RRPEm „Helldivers = Höllentaucher“ genannt, zum Schutz bei ihrer Arbeit getragen wird (vgl. Brown 2014a:[11]). Für dieses Konzept ist „Schmoranzug“ eine gelungene Übersetzung, denn „schmören“ passt erstens von der Bedeutung und vom Register her sehr gut zu „fry“ (vgl. *Duden* 2021; *Lexico.com* 2021a) und zweitens evozieren die zwei **IUT** „Schmoranzug“ und „Höllentaucher“ in Verbindung im Deutschen auch die Redewendung „in der Hölle schmören“, die die gefährlichen und unangenehmen Arbeitsbedingungen dieser Höllentaucher treffend beschreibt.

Die erste Problemstelle in diesem Beispiel ist der „Küchenraum“, die wortwörtliche Übersetzung des englischen „kitchen room“. „Küchenraum“ ist zwar ein korrektes deutsches Wort, aber eine *DWDS*-Korpussuche in den Referenz- und Zeitungskorpora⁷⁰ nach „Küchenraum“ und „Küche“ ergab, dass ersteres Wort weitaus weniger oft verwendet wird als letzteres (96 Treffer für „Küchenraum“⁷¹, 39 017 Treffer für „Küche“⁷²). Folglich kann argumentiert werden, dass „Küche“ natürlicher und angemessener ist als „Küchenraum“, zumal sich beide Posteditorinnen diesbezüglich einig zu sein scheinen.

Die Übersetzung von „Stop right where you are!“ ist die zweite Problemstelle in diesem Beispiel. Wiederum herrschte Einigkeit zwischen beiden Posteditorinnen bzgl. des Problemcharakters dieser Übersetzung, der (abgesehen von der fälschlich verwendeten Höflichkeitsform; siehe unten) darin besteht, dass die wörtliche Übersetzung des englischen Ausgangssatzes im Deutschen seltsam und tautologisch anmutet: Es würde reichen, „Bleib stehen!“ oder „Bleib, wo du bist!“ zu sagen, um das Gewünschte zu erreichen – eine Kombination der beiden Varianten wie in der RRMÜ ist dementsprechend überflüssig und entspricht nicht einem natürlichen oder flüssigen Sprachgebrauch. Die Posteditorinnen entschieden sich jeweils für eine der oben angeführten Varianten, wobei Posteditorin 1 durch Hinzufügung des „sofort“ noch deutlicher machte, dass der von Eo Angesprochene auch tatsächlich im selben Augenblick und genau dort, wo er sich gerade befindet, stehen bleiben soll (wie es im Englischen durch „right“ ausgedrückt wird).

Der Großteil der **Stil**-Korrekturen in beiden Posteditionen betrifft wie im Beispiel oben eine für den Kontext unübliche Wortwahl oder wörtliche Übersetzungen, die teilweise zu einem unnatürlichen Satzbau führen. Darüber hinaus korrigierte Posteditorin 1 auch Wortwiederholungen und achtete darauf, Tempi in einem Absatz durchgängig zu verwenden (z. B. gibt es in der RRMÜ einen richtigerweise im Präteritum verfassten Absatz, in dem sich ein Satz im Perfekt befindet, welcher von Posteditorin 1 ebenfalls auf das Präteritum geändert wurde). Posteditorin 2 führte keine solchen Korrekturen durch, dafür erkannte sie im Gegensatz zu Posteditorin 1 (siehe unten), dass die Konstruktion „auf + Artikel + Planetenname“ (z. B. „auf dem Mars“) üblicher ist als jene ohne Artikel („auf Mars“), wie eine erneute Korpussuche mittels oben erwähnter Korpora zeigte⁷³, und fügte dementsprechend in alle ohne Artikel vorkommenden Konstruktionen den Artikel ein.

⁷⁰ <https://www.dwds.de/d/korpora/public> (Stand: 18.09.2021).

⁷¹ <https://tinyurl.com/fx72tdjb> (Stand: 18.09.2021).

⁷² <https://tinyurl.com/27kwhpwk> (Stand: 18.09.2021).

⁷³ <https://tinyurl.com/ptcz4te6> (Stand: 19.09.2021) und <https://tinyurl.com/wum4k7hr> (Stand: 19.09.2021).

Trotzdem führte Posteditorin 1 insgesamt, wie oben schon erwähnt, mehr **Stil**-Korrekturen durch als Posteditorin 2. Abgesehen vom persönlichen Erfahrungshintergrund der Posteditorinnen, ihrer Arbeitsweise usw. könnte ein weiterer Grund für diesen Umstand sein, dass Posteditorin 1 sich „nur“ auf drei Texte, nämlich den Ausgangstext, die RRMÜ und ihre Postedition, konzentrieren musste und damit mehr kognitive Kapazitäten für die zielsprachlichen Versionen zur Verfügung hatte als Posteditorin 2, die sich schließlich zusätzlich dazu noch mit den Fehlerkategorien und ihrer Beschreibung auseinandersetzen musste. Dies hängt insofern auch mit der Kategorie **Stil** zusammen, als diese in der vorliegenden Fehlerkategorietypologie einen angemessenen, natürlichen und flüssigen Sprachgebrauch beschreibt (siehe Kapitel 5.1). Verstöße gegen diese natürliche Sprachverwendung können vermutlich bei größeren Kapazitäten für die zielsprachlichen Versionen und damit vermehrter Konzentration darauf leichter erkannt und korrigiert werden als bei geringeren Kapazitäten dafür⁷⁴. Außerdem liegt aufgrund der großen Zahl an PV-Korrekturen in der RRPEo, von denen mehrere auch den Inhalt verzerren (siehe unten), die Vermutung nahe, dass sich Posteditorin 1, unabhängig von ihren größeren Kapazitäten dafür, tatsächlich mehr auf die Übersetzungen als auf den Ausgangstext konzentrierte.

Tempus und Modus von Verben ist wiederum eine Fehlerkategorie, bzgl. derer sich die Posteditorinnen weitgehend einig waren: Beide fanden hier nur einen Fehler, nämlich die Übersetzung der *ing*-Form „killing“, die im gegebenen Kontext etwas Vergangenes beschreibt, in „tötet“, d. h. die Präsensform des inhaltlich richtigen Verbs „töten“ (siehe Beispiel 15 in Kapitel 5.2.2). Posteditorin 1 änderte das Präsens zum Präteritum, Posteditorin 2 zum Perfekt, wobei aus Gründen der Konsistenz mit einem vorhergehenden Absatz, in dem ebenfalls im Präteritum über Vergangenes berichtet wird (siehe Anmerkung dazu bei der Besprechung der Stil-Korrekturen weiter oben), die Lösung von Posteditorin 1 angemessener erscheint.

Der Problembereich **Tonfall und Register** ist die dritte der Kategorien, in denen Posteditorin 2 mehr Korrekturen durchführte als Posteditorin 1. Konkret stimmten 10 der 15 bzw. 19 Korrekturen dieser Kategorie in den beiden Posteditionen exakt überein, wobei dies hauptsächlich fälschlich verwendete Höflichkeitsformen betrifft, die von den Posteditorinnen

⁷⁴ Deswegen sind z. B. in der an die L-NMT angepassten SCATE-Fehlertypologie von Tezcan et al. (2019; siehe Kapitel 4.2) auch die Unterkategorien *disfluency* und *repetition*, die gemeinsam etwa der hier entwickelten Fehlerkategorie **Stil** entsprechen, der Hauptkategorie *FLUENCY* zugeordnet, deren Fehler ohne den Ausgangstext entdeckt werden können (vgl. Tezcan et al. 2019:41f.). Bei einer Fehlerannotation mithilfe dieser Typologie werden *FLUENCY*-Fehler dann auch tatsächlich ohne den Ausgangstext in der maschinellen Übersetzung annotiert, damit alle zur Verfügung stehenden Kapazitäten auf die Übersetzung angewandt und so ausgangssprachliche Interferenzen vermieden werden (können) (vgl. Fonteyne et al. 2020:3792).

in ein Duzen umgesetzt wurden, das wohl für ein Gespräch zwischen einem (noch dazu erst 16 Jahre alten) Ehepaar, wie es im Text vorkommt, üblicher ist (siehe Beispiel 13). Weitere 3 Korrekturen bzgl. **Tonfall und Register** beziehen sich auf dieselben Fehler, sind aber in beiden Posteditionen unterschiedlich: Z. B. wurde, um einer direkten Rede einen stärkeren mündlichen Charakter zu geben und gleichzeitig die Betonung zu erhöhen, von beiden Posteditorinnen die Partikel „denn“ eingesetzt, allerdings an unterschiedlichen Positionen im Satz. Die restlichen 2 bzw. 6 Korrekturen wurden nur von jeweils einer Posteditorin durchgeführt und betrafen entweder wiederum die (abgesehen von den Höflichkeitsformen) in der RRMÜ zu formelle Sprechweise der vorkommenden Jugendlichen oder Wörter, deren Register für den spezifischen Kontext nicht passend war:

Beispiel 14: Tonfall und Register – Korrektur nur in der RRPEm

RRAT	She picks the melted pieces of my frysuit out of the wound [...].	(RRAT:[24])
RRMÜ	Sie pickt die geschmolzenen Stücke meines Frysuits aus der Wunde [...].	
RRPEo	Sie pickt die geschmolzenen Teile meines Frysuits aus der Wunde [...].	
RRPEm	Sie entfernt die geschmolzenen Stücke meines Schmoranzugs aus der Wunde [...].	
RRHÜ	Sie zupft die geschmolzenen Reste meines Kochanzugs aus der Wunde [...].	(RRHÜ:[26])

Das Verb „pick“ bedeutet im Kontext „to remove separate things or small pieces from something [...]“ (*Cambridge Dictionary* 2021); das deutsche Verb „picken“ kann laut dem *Duden* (2021) ebenfalls diese Bedeutung haben, allerdings nur in der Umgangssprache. Da das Beispiel keine direkte Rede darstellt, ist dementsprechend das **Register** in der Übersetzung nicht unbedingt passend, was Posteditorin 2 durch ein standardsprachliches Verb korrigierte.

Zur Anpassung der direkten Reden an die informellere Sprechweise von Jugendlichen ist noch anzumerken, dass sich auch in diesem Bereich evtl. ein *priming effect* durch die Beschreibung der Fehlerkategorien bemerkbar macht: Während Posteditorin 1 solche Anpassungen nur durch Einfügungen von Partikeln (wie oben beschrieben) oder einmal durch die Verkürzung von „etwas“ auf „was“ vornahm, verwendete Posteditorin 2 abgesehen von solchen Strategien auch dreimal die Strategie der Verbverkürzung, also z. B. „hab“ statt „habe“. Diese Strategie wird im Dokument zu den Fehlerkategorien, das Posteditorin 2 zur Verfügung gestellt wurde, erwähnt und durch Beispiele erläutert (siehe Kapitel 5.1), was, wie gesagt, auf einen *priming effect* hindeutet, obwohl selbstverständlich auch der persönliche Stil

der beiden Posteditorinnen und ihr Erfahrungshintergrund diesbezüglich eine Rolle gespielt haben werden.

Damit ist die Besprechung der Korrekturen bzgl. der zehn Fehlerkategorien bzw. Problembereiche abgeschlossen. Es gibt jedoch auch, wie oben schon erläutert, Korrekturen in beiden Posteditionen, die keiner Kategorie zugeordnet werden konnten, entweder weil die ihnen zugrunde liegenden Fehler und Problemstellen in keine der Fehlerkategorien passten (NZ) oder weil sie nach Erachten der Verfasserin der vorliegenden Arbeit nicht notwendig gewesen wären und somit der persönlichen Vorliebe der Posteditorinnen entsprechen (PV). Die Korrekturen dieser beiden Kategorien werden im Folgenden besprochen.

NZ-Korrekturen gibt es in beiden Posteditionen nur sehr wenige (6 bzw. 7), was darauf hindeutet, dass die entwickelten Fehlerkategorien für diese Art von LMT-Output den Großteil der vorkommenden Fehler und Problemstellen abdecken, v. a. da die **NZ**-Korrekturen, wie gleich ersichtlich wird, eher anderen Korrekturen in den Posteditionen geschuldet sind als Fehlern oder Problemstellen in der ursprünglichen RRMÜ. Tatsächlich sind nur 2 **NZ**-Korrekturen auf eine Problemstelle in der ursprünglichen RRMÜ zurückzuführen: Dabei handelt es sich einerseits um die Übersetzung „backs = Rücken“, die im gegebenen Kontext zwar inhaltlich, grammatikalisch und in allen anderen Aspekten korrekt ist, aber in der vorliegenden Satzkonstruktion zu einer möglichen semantischen Ambiguität führt, die dadurch herrührt, dass das Wort „Rücken“ im Deutschen im Singular und im Plural dieselbe Form hat (vgl. *Duden* 2021) und deswegen nicht klar sein könnte, ob „Rücken“ den hinteren Teil des Rumpfes der betreffenden Person bezeichnet oder ihre Handrücken. Beide Posteditorinnen erkannten diese mögliche Ambiguität und machten den Satz durch ihre Korrektur („backs = Rücken ihrer Hände“ in der RRPEo bzw. „Handrücken“ in der RRPEm) semantisch eindeutiger.

Die zweite dieser **NZ**-Korrekturen, die auf die RRMÜ zurückzuführen sind, betrifft die Übersetzung des Wortes „township“, das einmal im Singular und einmal im Plural im RRAT vorkommt. Vom DeepL-Übersetzer wurde „township“ im Singular als „Gemeinde“ übersetzt, im Plural allerdings in die deutsche Übersetzung übernommen („Townships“). Beide Posteditorinnen besserten diese Inkonsistenz mit den jeweils von ihnen gewählten Übersetzungen für „township“ aus („Sektor“ in der RRPEo und „Gemeinde“ in der RRPEm).

Der Grund für die anderen 4 bzw. 5 **NZ**-Korrekturen liegt nicht in der RRMÜ per se, sondern in vorher durchgeführten anderen Änderungen der beiden Posteditorinnen. Diese **NZ**-Korrekturen sollten dementsprechend etwa Wort(-stamm-)wiederholungen ausbessern, die durch diese Änderungen entstanden wären, und Konsistenz in die Übersetzung bringen. So ist

etwa in Beispiel 11c ersichtlich, dass Posteditorin 1 nach ihrer Änderung von „soaring = steigend“ auf „triumphal“ auch „triumphs = Triumphe“ auf „Erfolge“ ausbesserte, um eine Wortstammwiederholung zu vermeiden, während Posteditorin 2 einmal in der oben erwähnten Rede der Herrscherin aus *Red Rising* die Übersetzung „obedience = Gehorsam“ auf „Gehorsamkeit“ ausbesserte und dies dann bei den anderen Stellen, an denen „obedience“ in der Rede vorkam, ebenfalls tat.

Schlussendlich sollen im Folgenden die **PV**-Korrekturen erläutert werden, also jene Korrekturen, die nach Ansicht der Verfasserin der vorliegenden Arbeit grundlos durchgeführt wurden. Sie können in zwei Gruppen eingeteilt werden: erstens solche, durch die die Übersetzung korrekt blieb, und zweitens solche, durch die in die Übersetzung Fehler und Problemstellen eingebaut wurden, die vorher noch nicht existiert hatten.

Zur ersten Gruppe gehören z. B. Änderungen, durch die ein Wort aus der RRMÜ, das inhaltlich, stilistisch und in allen anderen Belangen korrekt war, durch ein Synonym ersetzt wurde, das in den genannten Aspekten keinen Unterschied zu machen scheint, oder Umformulierungen von Sätzen, die ebenfalls die Übersetzung nicht oder nur in Bezug auf die Stilebene (nicht zu verwechseln mit der Kategorie Stil, die nur einen natürlichen Sprachgebrauch beschreibt) zu verändern scheinen. In diese Gruppe fallen etwa die in Beispiel 11b ersichtliche Änderung von Posteditorin 1 von „flink“ auf „agil“ als Übersetzung für „nimble“ sowie die Korrektur von Posteditorin 2 von „kolossal“ auf „riesig“ als Übersetzung für „colossal“, ebenfalls in diesem Beispiel. Außerdem wäre in dieser Gruppe für die RRPEo beispielsweise die Umstellung des Satzes „Ich war damals elf.“ auf „Damals war ich elf.“ (Ausgangssatz: „I was eleven then.“ (RRAT:[27])) in der RRPEo zu nennen oder die Umformulierung von „Sie [...] schimpft nicht mit mir“ auf „Sie [...] schimpft mich nicht“ (Ausgangssatz: „She [...] doesn’t berate me“ (RRAT:[24])) in der RRPEm.

Von den insgesamt 134 **PV**-Korrekturen in der RRPEo fallen 99 in diese erste Gruppe, in der RRPEm sind es 58 von 71. Das bedeutet, dass in der RRPEo 35 und in der RRPEm 13 **PV**-Korrekturen der zweiten Gruppe zugeteilt werden können, d. h. jener Gruppe von Änderungen, durch die die Übersetzung problematischer oder sogar fehlerhafter wird als vor der Änderung. Die Bandbreite dieser durch die Posteditorinnen eingeführten Problemstellen und Fehlern reicht von nur minimalen Inkorrektheiten bis zu tatsächlich inhaltsverzerrenden Fehlern.

Zu den minimalen Inkorrektheiten zählt beispielsweise in der RRPEo der Umgang mit der oben schon erwähnten Konstruktion „auf + Artikel + Planetenname“: Verwendete der DeepL-Übersetzer diese Konstruktion (was nicht immer der Fall war), löschte Posteditorin 1

den Artikel (so wurde aus „auf der Venus“ „auf Venus“), obwohl die RRMÜ eigentlich üblicher gewesen wäre. In der RRPEm ist die oben schon angeführte Änderung von „obedience = Gehorsam“ auf „Gehorsamkeit“ zu den minimalen Inkorrektheiten zu zählen: „Gehorsamkeit“ ist laut dem *Duden* (2021) v. a. auf Kinder anzuwenden, während das allgemeinere „Gehorsam“ generell die „Unterordnung unter den Willen einer Autorität“ bedeutet und damit in der Rede einer Herrscherin an ihre Untertanen angemessener gewesen wäre. Zu den inhaltsverzerrenden und damit schwerwiegenden Korrekturen gehören z. B. die Änderung der im Kontext korrekten Übersetzung „spider = Spinne“ auf „Insekt“ in der RRPEo oder jene von „webs = Spinnweben“ auf „Weben“ in der RRPEm. Posteditorin 1 führte insgesamt 6 solcher unerklärlichen und nach Ansicht der Autorin der vorliegenden Arbeit schwerwiegenden fehlerhaften Korrekturen durch, Posteditorin 2 nur 2.

Dies sowie der Umstand, dass in der RRPEo die **PV**-Korrekturen 44,5 % der Gesamtzahl der Korrekturen ausmachen, in der RRPEm allerdings nur 29,3 %, legen die Vermutung nahe, dass sich Posteditorin 1 weniger intensiv mit dem RRAT auseinandersetzte als Posteditorin 2 bzw. sich nicht so viele Gedanken darüber machte, ob sie in der Übersetzung „recht viel ändern soll oder nicht“⁷⁵. Dies sind die Worte von Posteditorin 2 (2021), die sich dementsprechend mit dieser Frage und „ob das jetzt einfach der Stil des Autors ist“ (Posteditorin 2 2021) vermutlich mehr auseinandersetzte.

Die Tatsache, dass Posteditorin 1 so viel ungezwungener posteditierte als Posteditorin 2, könnte an der beruflichen Erfahrung ersterer als Editorin von Videospielübersetzungen liegen: Bei ihrer Beschreibung dieser Tätigkeit gab sie gegenüber der Verfasserin der vorliegenden Arbeit an, sie hätte dabei „recht große Freiheiten“ gehabt und „mit klarem Fokus auf dem deutschen Text“ (Posteditorin 1 2021) (d. h. der Übersetzung) gearbeitet, da sie eine der Ausgangssprachen nicht beherrschte. Dies könnte sie dazu verleitet haben, auch beim Postediting freier zu arbeiten und sich nicht so sehr auf den RRAT zu konzentrieren, obwohl ihre andere Bemerkung, das Editieren von Videospielübersetzungen habe „mit Posteditieren oder literarischem Lektorieren [...] eher wenig zu tun“ (Posteditorin 1 2021) gehabt, auch andeutet, dass sie zwischen dem Postediting und ihrer beruflichen Tätigkeit doch einen Unterschied sah.

Zusammenfassend ist über die beiden Posteditionen zu sagen, dass die Verwendung oder Nicht-Verwendung von Fehlerkategorien beim Postediting in Bezug auf einige dieser

⁷⁵ Wie in Kapitel 4.3 erwähnt, hatten beide Posteditorinnen den Auftrag, die RRMÜ so zu posteditieren, dass die fertige Übersetzung ihren üblichen Qualitätsstandards entspricht, dementsprechend lag die unterschiedliche Herangehensweise der Posteditorinnen höchstwahrscheinlich nicht am Auftrag.

Kategorien keinen oder nur einen geringen Unterschied gemacht zu haben scheint (Auflösung von Anaphern, Auslassungen und Hinzufügungen, Formatierung, Grammatik und Orthographie, Stil, Tempus und Modus von Verben), während bzgl. anderer Fehlerkategorien ein mehr oder weniger deutlicher *priming effect* auf Posteditorin 2 zu bemerken ist (Bedeutung, Interpunktion, IUT, Tonfall und Register). Was diese Analyse außerdem gezeigt hat, ist die große Rolle, die der Subjektivitätsfaktor beim Postediting spielt: Der Großteil der Korrekturen in beiden Posteditionen entspricht nach Ansicht der Verfasserin der vorliegenden Arbeit der persönlichen Vorliebe der jeweiligen Posteditorinnen, die durchaus auch von ihrem Erfahrungshintergrund geprägt sein dürfte, und auch beim Problembereich Stil (der vermutlich der subjektivste der drei Problembereiche ist) sind Unterschiede in der Herangehensweise zu erkennen, die sich nicht unbedingt durch die Verwendung bzw. Nicht-Verwendung der Fehlerkategorien erklären lassen.

5.2.2 Ratings

An den Ratings nahmen insgesamt, d. h. auf die Einladung zur Teilnahme hin sowie im Rahmen der Lehrveranstaltung „Literarisches Schreiben und Lektorieren“ im Wintersemester 2021 (siehe Kapitel 4.4), 30 Studierende teil. Die Ratings von 21 **Teilnehmenden** konnten nicht ausgewertet werden, weil diese Studierenden einem oder mehreren der in den Kapiteln 4.3 und 4.4 genannten Kriterien (Masterstudium Translation mit Schwerpunkt „Übersetzen in Literatur – Medien – Kunst“, Deutsch als Muttersprache, Englisch als Arbeitssprache im Studium, *Red Rising* weder auf Englisch noch auf Deutsch bekannt) nicht entsprachen bzw., in einem Fall, ihr Einverständnis zur Verarbeitung ihrer Daten nicht erteilten. Dementsprechend wurden die Ratings von insgesamt neun Rater_innen ausgewertet. Die Präsentation der Ergebnisse dieser Auswertung erfolgt in zwei Schritten: Zunächst wird die Interrater_innen-Reliabilität (Fleiss' Kappa) angegeben und interpretiert, danach werden die eigentlichen Rating-Ergebnisse mittels paarweiser Rankings (siehe Kapitel 4.5) in Balkendiagrammen ausgewertet und diskutiert.

Die Tabelle auf der nächsten Seite zeigt die **Interrater_innen-Reliabilität**, d. h., inwiefern sich die Rater_innen in ihren Bewertungen einig waren, mittels des Fleiss' Kappa nach Fleiss (1971) für jedes der zwölf Ratings (zwei allgemeine Ratings und je eines pro Fehlerkategorie; siehe Kapitel 4.4). Die Werte für die Fehlerkategorie-Ratings sind wiederum in derselben Reihenfolge wie schon in den beiden vorangegangenen Kapiteln angeführt. In der letzten Zeile wird schließlich der Durchschnittswert für alle zwölf Ratings angegeben.

Tabelle 5: Interrater_innen-Reliabilität

Rating	Fleiss' Kappa
Allg. Rating #1	0,11
Auflösung von Anaphern	0,07
Auslassungen und Hinzufügungen	-0,01
Bedeutung (inkl. Unterkategorien)	0,25
Formatierung	0,36
Grammatik und Orthographie	-0,11
Interpunktion	-0,08
IUT	-0,06
Stil	0,18
Tempus und Modus von Verben	0,28
Tonfall und Register	-0,05
Allg. Rating #2	0,03
Durchschnitt	0,08

Die Interpretation der in der Tabelle angegebenen Kappa-Werte erfolgt anhand von Landis & Koch (1977:165). Wird die von den beiden Forschern aufgestellte Einteilung der Werte in unterschiedliche **Grade der Übereinstimmung** zwischen den Rater_innen angewendet, ergibt sich bei fünf Ratings eine schlechte (*poor*), bei vier eine niedrige (*slight*) und bei den restlichen drei eine angemessene Übereinstimmung (*fair agreement*). Die Berechnung des Durchschnitts dieser Werte ergibt einen Wert von 0,08, d. h., insgesamt gesehen stimmten die Rater_innen in ihren Bewertungen nur wenig überein. Im Folgenden wird noch genauer auf die einzelnen Werte eingegangen.

Bei fünf von zwölf Ratings bestand nur eine **schlechte Übereinstimmung**, nämlich bei jenen für die Fehlerkategorien Auslassungen und Hinzufügungen, Grammatik und Orthographie, Interpunktion, IUT sowie Tonfall und Register. Da die beiden letztgenannten Kategorien ohnehin eher als Problembereiche denn als wirkliche Fehlerkategorien aufzufassen sind, weil bei ihnen die Subjektivität eine mehr oder weniger große Rolle spielt (siehe Kapitel 5.1), ist die schlechte Übereinstimmung zwischen den Rater_innen bei diesen Ratings nicht verwunderlich. Im Gegenteil: Dieses Ergebnis bestärkt sogar die in Kapitel 5.1 getroffenen Feststellungen zur Subjektivität der beiden Kategorien.

Dies ist allerdings nicht der Fall bei den restlichen drei Kategorien, die eine schlechte Übereinstimmung aufweisen. Auslassungen und Hinzufügungen, Grammatik und Orthographie sowie Interpunktion sind tatsächliche Fehlerkategorien, dementsprechend wäre ein höherer Grad der Übereinstimmung erwartbar gewesen. Bei den Auslassungen und Hinzufügungen könnte der Grund für die schlechte Übereinstimmung darin liegen, dass diese

Fehlerkategorie überwiegend auf inhaltlichen Argumentationen beruht, wie Kapitel 5.1 zeigt, die sich doch zwischen verschiedenen Personen unterscheiden können (dazu unten mehr). Im Hinblick auf Grammatik und Orthographie sowie Interpunktion, die schließlich in Regelwerken nachschlagbare Eigenschaften der jeweiligen Zielsprache beschreiben, kann dies allerdings nur schwer der Grund für die schlechte Übereinstimmung sein. Bei diesen zwei Kategorien könnten aber mangelnde Konzentration auf die jeweiligen Ratings, mangelnde Bereitschaft, Zeit und Arbeit dafür zu investieren, oder dgl. eine Rolle gespielt haben; auch auf mangelnde Sprachkompetenz in der Zielsprache gibt es Hinweise (siehe unten), was aufgrund des Umstandes, dass die Zielsprache die Muttersprache der Rater_innen war, nicht unbedingt erwartet wurde. Darüber hinaus deuten manche Anmerkungen zum Interpunktionsrating darauf hin, dass die entsprechende Kategorie nicht von allen Rater_innen richtig verstanden wurde (siehe unten), was ebenfalls ein Grund für die schlechte Übereinstimmung diesbezüglich sein könnte.

Im Gegensatz zu diesen fünf Kategorien ist der Grad der Übereinstimmung bei den zwei allgemeinen Ratings sowie bei der Auflösung von Anaphern und beim Stil höher, befindet sich aber trotzdem nur im Bereich *slight agreement*. Bei den zwei allgemeinen Ratings ist dies nicht verwunderlich, da sich die Rater_innen hierbei auf ihr eigenes Qualitätsverständnis stützen durften und keinerlei Vorgaben dazu erhielten, was eine gute Übersetzungsqualität ausmacht und was nicht (siehe Kapitel 4.4). Der im Vergleich zum Kappa-Wert für das erste allgemeine Rating geringere Wert für das zweite dieser Ratings scheint darauf hinzudeuten, dass die Bearbeitung der drei Übersetzungen anhand der Fehlerkategorien die allgemeine Bewertung zumindest einiger Rater_innen verändert hat. Bei der Auflösung von Anaphern könnte der Grund für die niedrige Übereinstimmung abgesehen von mangelnder Konzentration und dgl. darin liegen, dass das Konzept von Anaphern vermutlich nicht allen Rater_innen vor der Erklärung dieser Fehlerkategorie bekannt gewesen war. Dies könnte zu Unsicherheiten geführt haben, was denn nun Anaphern im Ausgangstext sind und was nicht, was wiederum auch Unsicherheit bei der Bewertung der Auflösung dieser Anaphern in den Übersetzungen ausgelöst haben könnte. Zum Kappa-Wert beim Stil-Rating ist zu sagen, dass die überdurchschnittliche Übereinstimmung, die er ausdrückt und die schon fast an den Bereich *fair agreement* grenzt, aufgrund der großen Rolle der Subjektivität bei diesem Problembereich eine Überraschung ist, v. a. da bei den anderen beiden Problembereichen, wie oben erwähnt, nur schlechte Übereinstimmung herrschte. Dafür könnte es viele Gründe geben; einer davon ist womöglich die Tendenz von Posteditorin 2, wie bei den Teilnehmenden in Şahin & Gürses (2019:29) sowie Kolb (2021), die RRMÜ in weiten Teilen zu

übernehmen. Dies, gekoppelt mit dem Umstand, dass NMT-Systeme oft die syntaktischen Strukturen der zu übersetzenden Ausgangstexte beibehalten (vgl. Tezcan et al. 2019:47f.; Webster et al. 2020a:15), was auch hier der Fall war, wie eine kurze Analyse der RRMÜ bestätigte, könnte dazu geführt haben, dass die Rater_innen instinktiv einen eher unnatürlich anmutenden Sprachgebrauch v. a. bei der RRPEm wahrnahmen, was die relativ hohe Übereinstimmung im Stil-Rating nach sich gezogen haben könnte. Die Anmerkungen der Rater_innen (siehe unten) unterstützen diese Vermutung.

Bei den Ratings zu den noch nicht genannten Kategorien, nämlich Bedeutung, Formatierung und Tempus und Modus von Verben, liegt die Übereinstimmung im Bereich *fair agreement*, was den höchsten unter den Rater_innen vorkommenden Grad der Übereinstimmung darstellt. Bzgl. der Kategorie Bedeutung liegt dies evtl. daran, dass die Rater_innen explizit auch auf die Möglichkeit der Verwendung von einsprachigen Wörterbüchern für dieses Rating hingewiesen wurden und sich somit auf anerkannte Quellen für ihre Bewertung stützen konnten (obwohl dies womöglich nicht bei allen der Fall war; siehe unten). Beim Rating der Tempi und Modi der Verben in den Übersetzungen dürfte dahingegen der Umstand, dass der Großteil des RRAT in der *present tense* verfasst ist, d. h. nicht viele verschiedene Tempi zum Einsatz kommen, die Bewertung vereinfacht haben, was vermutlich zur höheren Übereinstimmung geführt hat. Die vergleichsweise sehr hohe Übereinstimmung beim Formatierungsrating liegt wahrscheinlich einerseits daran, dass nur drei Absätze im RRAT überhaupt formatiert waren (somit mussten auch in den Übersetzungen nur drei Absätze diesbezüglich bewertet werden), und andererseits an dem Umstand, dass die Kategorie nichts mit Sprache per se, sondern mit dem Aussehen von Texten zu tun hat, sodass diese Kategorie wohl die am wenigsten subjektive der entwickelten Fehlerkategorien ist. Insgesamt zeigen diese Ausführungen und auch der niedrige Durchschnittswert der Kappa-Werte aber doch, dass nicht nur das Postediting von literarischen maschinellen Übersetzungen eine sehr **subjektive Angelegenheit** ist, sondern auch die Qualitätsbewertung von solchen Posteditionen und Humanübersetzungen, selbst wenn dieser durch vorher erklärte Fehlerkategorien ein Rahmen gesetzt wird. Dieser Umstand muss bei der Interpretation der Rating-Ergebnisse, die im Folgenden vorgestellt werden, stets im Hinterkopf behalten werden. Ebenfalls zu beachten ist, dass es einige nicht nachvollziehbare Bewertungen gibt (siehe unten), die darauf hindeuten, dass einige Rater_innen, wie schon vermutet, weniger konzentriert arbeiteten bzw. sich weniger intensiv mit den Übersetzungen beschäftigten als andere und auch ein bloßes Vergeben der Ränge ohne jedwede fundierte Überlegungen durch

einzelne Rater_innen nicht auszuschließen ist. Weitere Studien wären also notwendig, um die Ergebnisse zu validieren.

Jedenfalls wurden die Ratings, wie oben erwähnt und in Kapitel 4.5 ausführlich erklärt, mittels drei **paarweiser Rankings** ausgewertet, bei denen zuerst die Ränge der RRPEo mit jenen der RRPEm, dann die Ränge der RRPEo mit jenen der RRHÜ und zuletzt die Ränge der RRPEm mit jenen der RRHÜ in jedem Rating verglichen wurden. Die Ergebnisse dieser paarweisen Rankings sind in drei Balkendiagrammen (Abb. 5-7) dargestellt.

Wie in Abb. 5 auf den ersten Blick ersichtlich wird, wurde die von Posteditorin 1 erstellte RRPEo häufiger besser als ($RRPEo > RRPEm$) oder zumindest gleich gut wie die RRPEm von Posteditorin 2 ($RRPEo = RRPEm$) bewertet. Eine schlechtere Bewertung ($RRPEo < RRPEm$) erhielt die RRPEo nur vergleichsweise selten. Im zweiten paarweisen Ranking (Abb. 6) ist die Situation etwas ausgeglichener, trotzdem erhielt die RRPEo auch im Vergleich mit der RRHÜ überraschenderweise etwas häufiger eine bessere Bewertung ($RRPEo > RRHÜ$) als umgekehrt. Dieser umgekehrte Fall ($RRPEo < RRHÜ$) kam aber etwas öfter vor als eine gleich gute Bewertung der beiden Übersetzungen ($RRPEo = RRHÜ$). Dahingegen überwogen im dritten paarweisen Ranking (Abb. 7) eindeutig die besseren Bewertungen der RRHÜ im Gegensatz zur RRPEm ($RRPEm < RRHÜ$), während sich der umgekehrte Fall ($RRPEm > RRHÜ$) sowie gleich gute Bewertungen ($RRPEm = RRHÜ$) in etwa die Waage hielten.

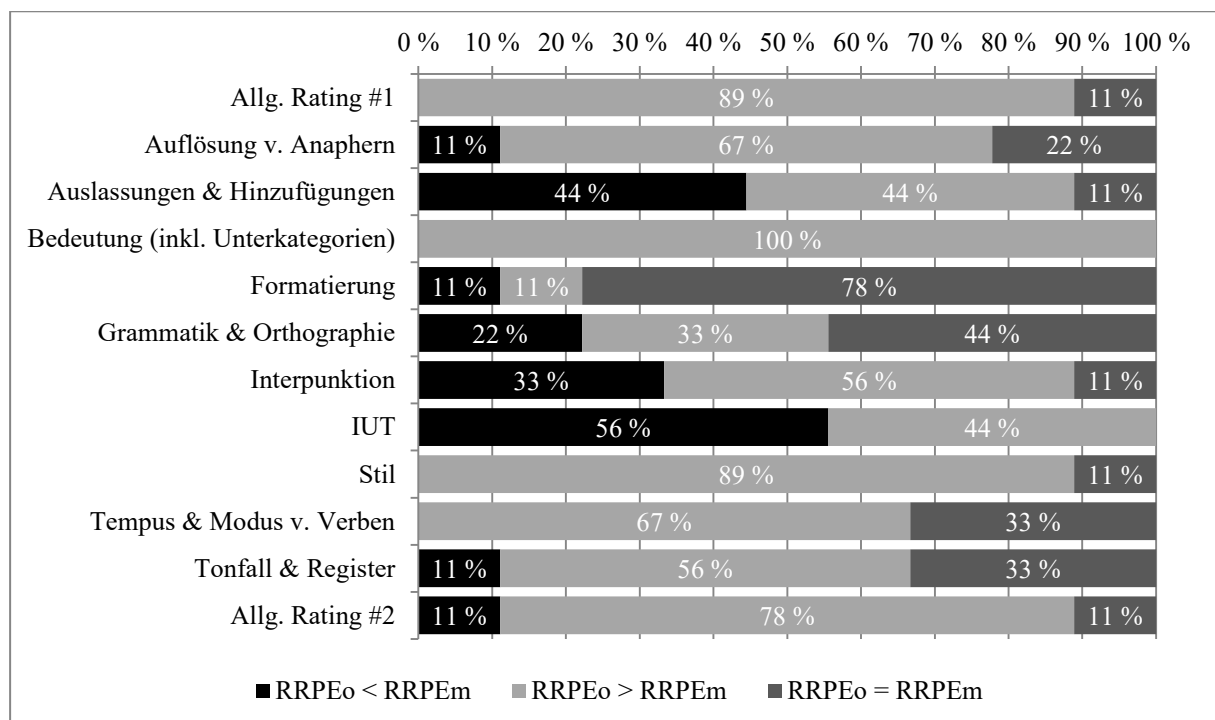


Abb. 5: Paarweises Ranking RRPEo – RRPEm

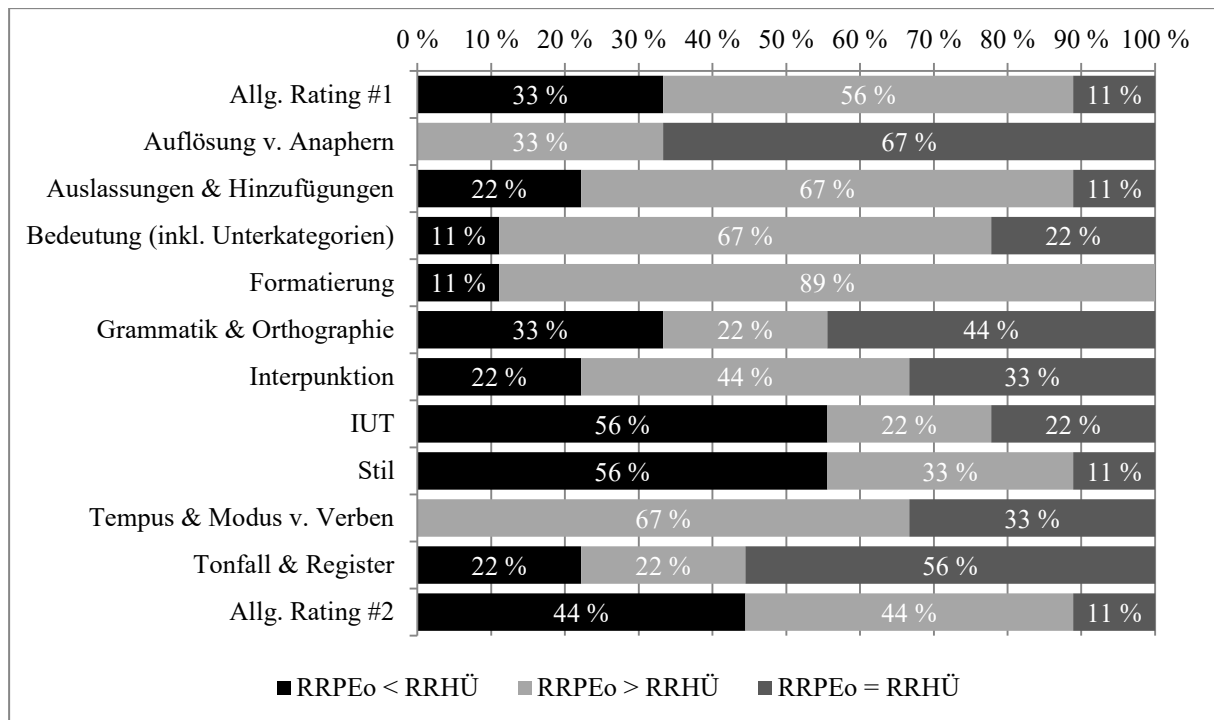


Abb. 6: Paarweises Ranking RRPEo – RRHÜ

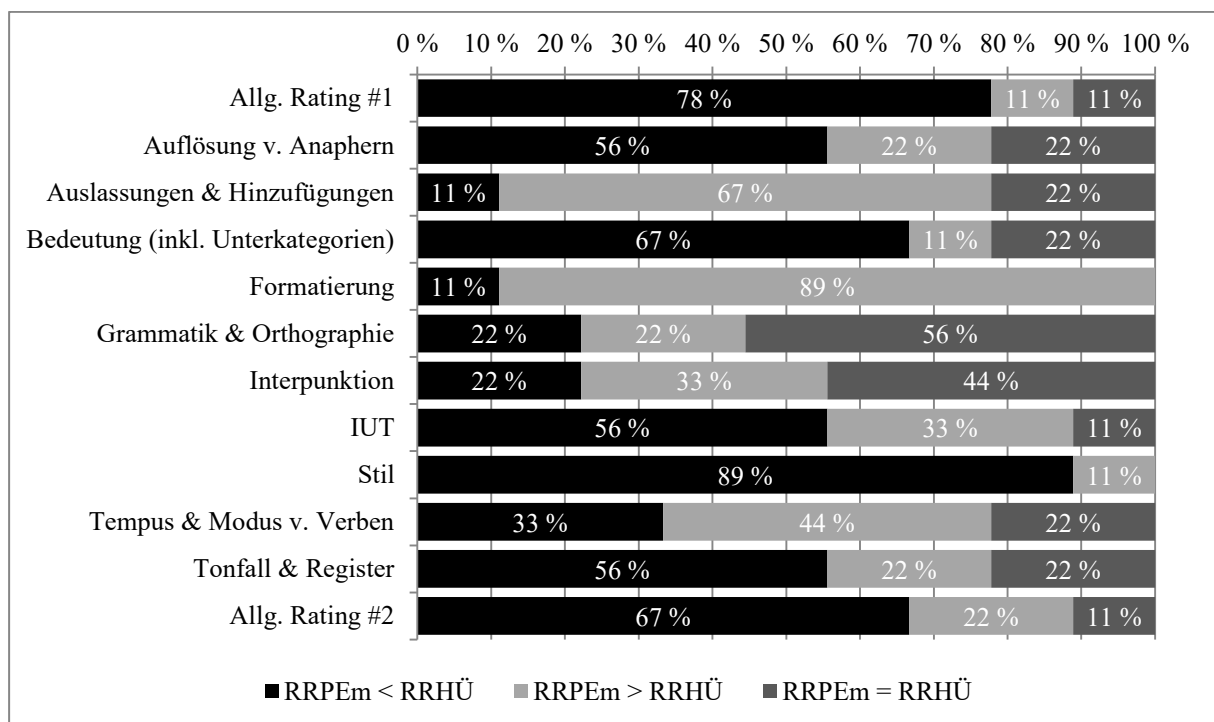


Abb. 7: Paarweises Ranking RRPEm – RRHÜ

Nach dieser kurzen allgemeinen Besprechung der Ergebnisse der paarweisen Rankings wird im Folgenden noch genauer auf die einzelnen Ratings eingegangen. Dabei werden stets zuerst die Ergebnisse der drei paarweisen Rankings angeführt und die Anmerkungen der

Rater_innen⁷⁶ (sofern es welche gibt) interpretiert, bevor daraus allgemeine Schlussfolgerungen für jedes Rating gezogen werden.

Im **ersten allgemeinen Rating** wurde die RRPEo überhaupt nie schlechter als die RRPEm eingestuft, sondern von der überwiegenden Mehrheit besser bewertet. Nur ein_e Rater_in gab den beiden Posteditionen denselben Rang. Der Vergleich der RRPEo mit der RRHÜ ist etwas ausgeglichener: Drei Rater_innen bewerteten die RRHÜ besser als die RRPEo, bei fünf war es umgekehrt, und eine_r vergab an beide denselben Rang. Das dritte paarweise Ranking schließlich ergab sieben Mal $RRPEm < RRHÜ$ und jeweils einmal $RRPEm > RRHÜ$ und $RRPEm = RRHÜ$.

Aus den Anmerkungen der Rater_innen ist zu schließen, dass in der RRPEo von Posteditorin 1 v. a. die IUT, von denen diese Posteditorin viele im Englischen beließ (siehe Kapitel 5.2.1), negativ auffielen (R1, R2), aber auch der Aspekt Tonfall und Register (R2). Auch bei der RRHÜ fielen Tonfall und Register negativ auf (R2), wohingegen die natürliche Sprachverwendung und die gute Lesbarkeit sowie die Syntax in dieser Übersetzung gelobt wurden (R1, R2, R4). Auch habe es diese Übersetzung geschafft, den Stil des Ausgangstextes am besten wiederzugeben („Stil“ ist also in dieser Anmerkung von R4 wie „Stil“ in der Literaturwissenschaft zu verstehen; siehe Kapitel 5.1). In Bezug auf die RRPEm waren laut R2 v. a. die Bedeutungsübertragung, die Unnatürlichkeit der Sprachverwendung (also die Kategorie Stil) und Tempus (und Modus) von Verben Kritikpunkte.

Die drei Rater_innen, die Anmerkungen zu diesem Rating machten, sahen die RRHÜ also positiver als die RRPEo (und bewerteten sie auch dementsprechend), aber insgesamt scheint sich doch für das **erste allgemeine Rating** die Rangfolge (1) RRPEo, (2) RRHÜ, (3) RRPEm abzuzeichnen. Die Benutzung der Fehlerkategorien durch Posteditorin 2, die die RRPEm anfertigte, dürfte folglich auf den ersten Blick keine positiven Auswirkungen auf die Qualität dieser Postedition gehabt haben, eher im Gegenteil, wobei aber auch nicht alles den Fehlerkategorien zugeschrieben werden kann oder sollte, denn der persönliche Übersetzungs- und Posteditingstil von Posteditorin 2 wird auch zur schlechteren Bewertung beigetragen haben. Die vielen PV-Korrekturen, die Posteditorin 1 durchführte und die nach Ansicht der Verfasserin dieser Arbeit nicht notwendig gewesen wären (siehe Kapitel 5.2.1), dürften im Gegensatz dazu zur Qualität der Postedition beigetragen haben. Darüber hinaus zeigt dieses

⁷⁶ Werden die Anmerkungen einzelner Rater_innen zitiert, wird auf sie mittels des Buchstabens „R“ (für „Rater_in“) sowie mit einer zufällig von der Verfasserin dieser Arbeit an diese Rater_innen vergebenen Nummer verwiesen, z. B. „R1“ für „Rater_in 1“. So können die Anmerkungen auch in Anhang 3 dieser Arbeit, in dem sie zur Gänze abgebildet sind, gefunden werden.

Ergebnis auch das Potential der Kombination L-NMT+Postediting, da es die RRPEo schaffte, besser abzuschneiden als die RRHÜ.

Bei der **Auflösung von Anaphern** waren sich die Rater_innen weniger einig als beim ersten allgemeinen Rating, trotzdem erhielt die RRPEo mehrheitlich die besseren Bewertungen verglichen mit der RRPEm: Gleich sechs Rater_innen stufen die RRPEo besser ein als die RRPEm, bei nur einem_einer Rater_in war es umgekehrt, und die restlichen zwei gaben beiden Posteditionen denselben Rang. Beim Vergleich RRPEo und RRHÜ ergab sich dreimal $RRPEo > RRHÜ$ und sechsmal $RRPEo = RRHÜ$, während fünf Rater_innen in diesem Rating die RRPEm schlechter als die RRHÜ bewerteten, bei zwei das Umgekehrte der Fall war und wiederum zwei die beiden Übersetzungen als gleichwertig einstufen.

Eine Interpretation dieser Ergebnisse wird durch den Umstand erschwert, dass es nur zwei, recht kurz ausfallende Anmerkungen zu diesem Rating gab: Laut R1 ist die **Auflösung von Anaphern** in allen Übersetzungen „sehr gut“ gelungen, „aber in A gab es 2 Fehler“⁷⁷. Welche Fehler das seien, führte R1 nicht an. Dahingegen gab R5 an, für eine von ihm_ihr identifizierte **Anapher** („they show“) habe es in der RRPEm eine „unpräzise Auflösung“ gegeben, was zur schlechteren Bewertung der RRPEm führte:

Beispiel 15⁷⁸: Auflösung von Anaphern

RRAT	They show the horrors of a terrorist bomb killing a Red mining crew and an Orange technician group.	(RRAT:[26])
RRMÜ	Sie zeigen die Schrecken einer terroristischen Bombe, die eine rote Minenmannschaft und eine orangefarbene Technikergruppe tötet.	
RRPEo	Die Schrecken eines Terrorangriffs mit einer Bombe wird gezeigt , die eine Rote Minencrew und eine Orange Technikergruppe tötete.	
RRPEm	Sie zeigen die Schrecken einer Terrorbombe, die eine Rote Minenmannschaft und eine Orange Technikergruppe getötet hat.	
RRHÜ	Sie zeigen die Schrecken eines terroristischen Bombenanschlags, durch den ein Bergbauteam der Roten und eine Technikergruppe der Orangen ums Leben gekommen sind.	(RRHÜ:[28])

⁷⁷ Wie in Kapitel 4.4 schon angeführt, wurden die Übersetzungen den Rater_innen unter den Namen „Übersetzung A/B/C“ präsentiert. In den Anmerkungen der Rater_innen gilt daher: Übersetzung A = RRPEm, Übersetzung B = RRHÜ und Übersetzung C = RRPEo.

⁷⁸ Die Nummerierung der Beispiele wird aus dem vorigen Kapitel fortgesetzt, um bei Verweisen auf diese Beispiele Verwirrung bei den Leser_innen zu vermeiden. Die drei von den Rater_innen bewerteten Übersetzungen sind in den Beispielkästen in diesem Kapitel in derselben Reihenfolge wie schon im vorigen Kapitel (und nicht in der, in der sie den Rater_innen präsentiert wurden) angeführt, wiederum zur Vermeidung von Verwirrung bei den Leser_innen. Für etwaige Verweise auf diese Beispiele aus dem vorigen Kapitel wird zwar auch die RRMÜ in den Beispielen angeführt, darauf eingegangen wird aber im Rahmen dieses Kapitels bei den spezifischen Beispielen nicht (deswegen ist die RRMÜ in grauer Schrift ohne fette Hervorhebungen gehalten), da die Rater_innen diese RRMÜ bei ihrer Bewertung auch **nicht** sahen. Die **fetten** Hervorhebungen in den Beispielkästen wurden wiederum von der Verfasserin der vorliegenden Arbeit platziert, um die Anmerkungen der Rater_innen besser zu veranschaulichen.

Es ist Streitbar, ob „they show“ tatsächlich eine **Anapher** ist, doch die Bewertung von R5 diesbezüglich ist ohnehin nicht unbedingt als überlegt einzustufen: Die in der RRPEm angekreidete Übersetzung („Sie zeigen“) wurde nämlich in der RRHÜ auf dieselbe Art und Weise verwendet, von R5 aber in diesem Fall nicht als negativ bewertet, was eine Inkonsistenz in der Bewertung darstellt. Dies ist die erste dieser nicht nachvollziehbaren Bewertungen, die oben erwähnt wurden und die darauf hindeuten, dass bei einzelnen Rater_innen wohl mangelnde Konzentration und/oder fehlende Bereitschaft, sich mit dem Thema intensiv auseinanderzusetzen, geherrscht haben dürfte. Außerdem ist beim Rating bzgl. dieser Fehlerkategorie nicht auszuschließen, dass, wie ebenfalls oben schon vermutet, die Rater_innen unsicher bzgl. des Konzepts von **Anaphern** waren, was die Ergebnisse diesbezüglich weiter verschwimmen lassen würde.

Werden trotz der gerade genannten Gründe alle Ergebnisse des Ratings zur **Auflösung von Anaphern** zusammengefasst, kann die folgende Rangfolge der Übersetzungen abgeleitet werden: (1) RRPEo und RRHÜ und (2) RRPEm. Dies ist insofern überraschend, als beide Posteditorinnen in diesem Bereich exakt dieselben Korrekturen durchführten (die Benutzung von Fehlerkategorien hier also keinen Unterschied machte; siehe Kapitel 5.2.1) und somit angenommen werden könnte, ihre Posteditionen würden auch gleich bewertet werden, was allerdings, wie oben angeführt, nur selten der Fall war. Die RRHÜ landet zwar ebenfalls auf Platz 1, doch der Umstand, dass sie manchmal doch schlechter als beide Posteditionen eingestuft wurde, ist überraschend und veranlasste eine schnelle Untersuchung der RRHÜ auf Fehler bei der **Auflösung von Anaphern** durch die Verfasserin der vorliegenden Arbeit, die allerdings keine solchen Fehler in der RRHÜ ergab. Dies entspricht auch der allgemeinen Auffassung, dass die **Auflösung von Anaphern** nur für MT-Systeme (und damit indirekt auch für Posteditor_innen, die Outputs dieser Systeme verwenden und korrigieren müssen), nicht aber für Humanübersetzer_innen ein Problem darstellt (vgl. z. B. Mohar et al. 2020:134). Dies müsste selbstverständlich noch durch eine genauere Analyse der RRHÜ bestätigt werden, denn dass auch Humanübersetzer_innen zuweilen Fehler passieren, wurde schon in Kapitel 5.1 deutlich. Die vorliegenden Daten scheinen aber doch, wie das Beispiel oben, darauf hinzudeuten, dass manche Rater_innen bei diesem Rating offenbar nicht immer ganz bei der Sache waren und/oder mit dem Konzept von **Anaphern** nicht ausreichend vertraut waren, um es auf nachvollziehbare Art und Weise bewerten zu können.

Die paarweisen Rankings der Rating-Ergebnisse zur Kategorie **Auslassungen und Hinzufügungen** ergab folgende, etwas ausgeglichene Situation: Vier Rater_innen bewerteten die RRPEo hier schlechter als die RRPEm, bei weiteren vier war es umgekehrt,

und eine_r bewertete die beiden gleich gut. Beim Vergleich zwischen RRPEo und RRHÜ ergab sich zweimal $RRPEo < RRHÜ$, sechsmal $RRPEo > RRHÜ$ und einmal $RRPEo = RRHÜ$. Das paarweise Ranking zwischen der RRPEm und der RRHÜ ist ähnlich und ergab einmal eine schlechtere Bewertung für die RRPEm, sechsmal eine bessere und zweimal eine gleich gute.

Zwei Rater_innen gaben in den Anmerkungen Beispiele für **Auslassungen und Hinzufügungen** in den Übersetzungen: R5 führte zwei tatsächliche **Auslassungen** in der RRHÜ an, R6 eine weitere (so wurde z. B. von R6 richtig erkannt, dass in der RRHÜ die im RRAT in Klammern gesetzte Erklärung des IUT „Goldbrows“ fehlt; siehe dazu Beispiel 12 in Kapitel 5.2.1). Außerdem machte R5 auch auf die Übersetzung „delays the progress of making [...] (RRAT:[26]) = verzögert unsere Aufgabe, [...] zu machen“ aufmerksam, die seiner_ihrer Meinung nach eine Hinzufügung in der RRPEo darstellt. Da sich im RRAT „progress“ allerdings tatsächlich auf eine in diesem Satz nicht genannte Aufgabe bezieht (nämlich die Aufgabe der Roten, den Planeten Mars für die anderen Farben vorzubereiten; vgl. RRAT:[26]), würde diese Übersetzung von der Verfasserin der vorliegenden Arbeit nicht unbedingt als **Hinzufügung**, sondern eher als Bedeutungsproblemstelle eingestuft werden.

Diese Ausführungen unterstreichen deutlich den oben bei der Besprechung der Interrater_innen-Reliabilität schon vermuteten Grund für den geringen Kappa-Wert bei diesem Rating (und damit auch für die ausgeglichene Situation bei den Bewertungen): Die Kategorie **Auslassungen und Hinzufügungen** basiert vorwiegend auf inhaltlichen Argumentationen, die zwar evtl. nicht so subjektiv sind wie Einschätzungen der Natürlichkeit der Sprachverwendung, aber doch mehr von den jeweiligen Evaluater_innen abhängen als z. B. die Argumentation von Grammatikfehlern, die durch Rückgriff auf Regelwerke zur Zielsprache unterstützt werden kann. Darüber hinaus zeigen die Ausführungen auch, dass die Unterscheidung zwischen einer **Auslassung/Hinzufügung** und einem Bedeutungsfehler zumindest teilweise von der evaluierenden Person abhängt. Diese Beobachtung wird unterstützt durch die Anmerkung von R6 zur RRPEm: In dieser Anmerkung führte R6 die Übersetzung der Wortsequenz „old now at thirty-five“ (siehe Beispiel 16) als **Auslassung** in der RRPEm an. Diese Übersetzung wurde allerdings von der Verfasserin der vorliegenden Arbeit als Bedeutungsfehler der RRMÜ eingestuft, der von Posteditorin 2 nicht korrigiert wurde (siehe Kapitel 5.2.1). Dieser Ansicht ist offenbar auch R5, der_die in seiner_ihrer Anmerkung zum Bedeutungsrating ebenfalls die Übersetzung dieser Wortsequenz anführte und sie somit als Bedeutungsfehler in der RRPEm klassifizierte.

Beispiel 16: Auslassungen und Hinzufügungen / Bedeutung

RRAT	In the distance, drunk Uncle Narol, old now at thirty-five , plays his zither for the children who dance about our legs [...].	(RRAT:[26f.])
RRMÜ	In der Ferne spielt der betrunkene Onkel Narol, inzwischen fünfunddreißig Jahre alt, auf seiner Zither für die Kinder, die um unsere Beine tanzen [...].	
RRPEo	Weiter vorne spielt mein betrunkenener Onkel Narol, alt mit seinen fünfunddreißig Jahren , auf seiner Zither für die Kinder, die um unsere Beine tanzen.	
RRPEm	In der Ferne spielt der betrunkene Onkel Narol, inzwischen fünfunddreißig Jahre alt , auf seiner Zither für die Kinder, die um unsere Beine tanzen [...].	
RRHÜ	In der Ferne sehe ich Onkel Narol. Er ist betrunken und mit fünfunddreißig schon recht alt . Er spielt auf der Zither für die Kinder, die um unsere Füße herumtanzen.	(RRHÜ:[28])

Wie bei der Auflösung von Anaphern scheint also auch beim Rating zu den **Auslassungen und Hinzufügungen** evtl. etwas Verwirrung oder Unsicherheit bzgl. der zu bewertenden Aspekte geherrscht zu haben, aber eine Rangfolge der Übersetzungen kann trotzdem aufgestellt werden und sieht folgendermaßen aus: (1) RRPEo und RRPEm und (2) RRHÜ. Wie die Auswertung der Korrekturen der Posteditorinnen zeigt auch dieses Ergebnis, dass die Benutzung der Fehlerkategorien durch Posteditorin 2 keinen wesentlichen Unterschied bei der Kategorie **Auslassungen und Hinzufügungen** gemacht zu haben scheint. Interessant ist aber, dass in dieser Fehlerkategorie die RRHÜ insgesamt schlechter als die beiden Posteditionen abschneidet, was darauf hindeutet, dass L-NMT+Postediting in der Lage ist, die inhaltlichen Aspekte eines Ausgangstextes zumindest genauso vollständig wie und zuweilen sogar vollständiger als eine reine Humanübersetzung in die Zielsprache zu übertragen. Ob diese Aspekte auch bedeutungstechnisch richtig bzw. angemessen übertragen wurden, geht aus der Besprechung der Ergebnisse des nächsten Ratings, des **Bedeutungsratings**, hervor.

Bei diesem herrschte über alle drei paarweisen Rankings recht große Übereinstimmung⁷⁹, aber nirgends so viel wie im Vergleich der Ränge der beiden Posteditionen: Sämtliche Rater_innen waren hier der Ansicht, dass in der RRPEo die **Bedeutung** des

⁷⁹ Dass trotz dieser Einigkeit kein noch höherer Kappa-Wert beim Bedeutungsrating herauskam, liegt daran, dass beim paarweisen Ranking nur die Variablen „höherer Rang“ oder „niedrigerer Rang“ in Betracht gezogen wurden, während bei der Interrater_innen-Reliabilität tatsächlich die einzelnen Ränge 1, 2 und 3 in die Berechnung einfließen. D. h., dass die RRPEo zwar in diesem Fall von allen Rater_innen besser bewertet wurde als die RRPEm, aber nicht, dass beide Posteditionen von allen Rater_innen auch jeweils denselben Rang erhielten (z. B. erhielt die RRPEo von einem_einer Rater_in den Rang 1, die RRPEm den Rang 2, doch von einem_einer anderen wurden jeweils die Ränge 2 und 3 vergeben. In beiden Fällen erhielt die RRPEo also einen höheren Rang als die RRPEm, aber bzgl. des Ranges selbst waren sich die beiden Rater_innen nicht einig.).

Ausgangstextes besser wiedergegeben wird als in der RRPEm. Verglichen mit der RRHÜ wurde die RRPEo einmal schlechter als die RRHÜ, sechsmal besser und zweimal gleich gut bewertet. Im dritten paarweisen Ranking schließlich ergab sich sechsmal $RRPEm < RRHÜ$, einmal $RRPEm > RRHÜ$ und zweimal $RRPEm = RRHÜ$.

Die **Bedeutung** ist eine der Fehlerkategorien, die beim ersten allgemeinen Rating in den Anmerkungen erwähnt wurde, das Rating diesbezüglich dürfte also von besonderer Wichtigkeit sein. Tatsächlich fielen auch die Anmerkungen zu diesem Rating meist länger aus als die Anmerkungen zu den schon besprochenen. R1 stellte in seiner_ihrer eher allgemeinen Anmerkung fest, es gäbe v a. in der RRHÜ „einige[...] Bedeutungsfehler[...]“. Als Grund nannte R1 die Art der Übersetzung („eher frei[...]“), Beispiele gab er_sie aber nicht an. Andere Rater_innen gaben dahingegen schon Beispiele für **Bedeutungsfehler** in allen drei Übersetzungen: Etwa nannte R5 die oben angeführte Übersetzung für „old now at 35“ in der RRPEm, von R5 und R6 wurde z. B. die Übersetzung „benefits = Prämien“ in der RRHÜ ([26]) genannt und wiederum von R6 die Übersetzung „salute = salutieren“ in der RRPEo. Dieses letzte Beispiel (das übrigens das einzige für die RRPEo ist) zeigt, dass bei der Evaluierung dieser Fehlerkategorie nicht unbedingt, wie bei der Präsentation dieser Kategorie empfohlen, von allen Rater_innen einsprachige Wörterbücher benutzt wurden (sondern vielleicht manche Rater_innen mehr nach Gefühl vorgingen), denn wird „salutieren“ im übertragenen Sinn verstanden, stimmt die **Bedeutung** dieses Wortes mit der im RRAT gemeinten Aktion „[s]how or express admiration and respect for“ (*Lexico.com* 2021a) genauer überein als jene des in der RRPEm und der RRHÜ verwendeten Wortes „(be-)grüßen“ (vgl. *Duden* 2021), das nicht negativ angemerkt wurde.

R7 fasste die Anmerkungen der Rater_innen mit seiner_ihrer Aussage, „alle drei Übersetzungen [hätten] ihre Probleme“, zusammen (vgl. auch R6), merkte zusätzlich dazu aber auch einen für ihn_sie großen Kritikpunkt an der RRPEm an, aufgrund dessen er_sie sogar aufhören würde, die Übersetzung zu lesen: In dieser Postedition wurden nämlich „human breed“ und „human race“ mit „Rasse“ übersetzt, was im Deutschen nicht geht“ (R7; siehe Beispiel 17 für mehr Kontext diesbezüglich). Dies ist etwa laut dem Biologen Kattmann (1999) tatsächlich problematisch, allerdings ist R7 der_die einzige Rater_in, der_die dies anmerkte, und auch der Verfasserin der vorliegenden Arbeit sind deutsche Humanübersetzungen englischsprachiger Fantasy- und Science-Fiction-Werke bekannt, in denen von Menschen als „Rasse“ (meist im Gegensatz zu erfundenen „Rassen“ wie Feen oder Kobolden) gesprochen wird, wenn in den Ausgangstexten, so wie hier im RRAT, von der Gesamtheit der Menschen, also der ganzen Menschheit und nicht einzelnen Gruppen von Menschen, die Rede

ist. Dieser Ansicht schließt sich etwa auch der deutsche Fantasy-Autor de Grandpair (2013:35) an.

Ob nun dieses Beispiel von R7 tatsächlich negativ eingeschätzt wird oder nicht, liegt im Auge des_der Betrachtenden, doch Tatsache ist, dass es für die RRPEm in den Anmerkungen der Rater_innen mehr negative Beispiele zur **Bedeutungsübertragung** gab als für die RRHÜ und viel mehr als für die RRPEo. Dies erklärt auch, warum eine Zusammenfassung der Ergebnisse folgende Rangfolge der Übersetzungen ergibt: (1) RRPEo, (2) RRHÜ und (3) RRPEm. Es scheint also, als habe das Wissen von Posteditorin 2 um die in L-NMT-Übersetzungen häufig vorkommenden Probleme bei **Bedeutungsübertragungen** ihr nicht geholfen, eine in dieser Hinsicht qualitativ hochwertigere Postedition als ihre Kollegin anzufertigen. Der mögliche *priming effect* der Fehlerkategorien auf Posteditorin 2 (siehe Kapitel 5.2.1) in diesem Bereich ist somit als negativer Effekt einzustufen, wobei aber selbstverständlich auch andere Aspekte wie die geringere Erfahrung im Editieren von Texten von Posteditorin 2 o. Ä. diesbezüglich eine Rolle gespielt haben werden. In Bezug auf die RRPEo dürften sich dahingegen die Korrekturen von fälschlich übersetzten Wortsequenzen, die Posteditorin 1 durchführte, Posteditorin 2 aber nicht (wie in Beispiel 11b in Kapitel 5.2.1 oder in Beispiel 16 in diesem Kapitel), positiv auf die Qualität der RRPEo ausgewirkt haben, v. a. im Vergleich zur RRPEm. Der Umstand, dass die RRHÜ wiederum insgesamt schlechter abschnitt als die RRPEo, aber besser als die RRPEm, scheint aber doch zu zeigen, dass L-NMT+Postediting nicht nur alle inhaltlichen Aspekte des Ausgangstextes in die Übersetzung bringen, sondern diese in den meisten Fällen auch **bedeutungstechnisch** korrekt übertragen kann, wobei aber viel von dem_der jeweiligen Posteditor_in abhängt.

Beim Rating zur nächsten Kategorie, der **Formatierung**, waren sich die Rater_innen relativ einig, wie auch der Kappa-Wert dazu zeigt. Im paarweisen Ranking RRPEo vs. RRPEm bewertete nur jeweils ein_e Rater_in die beiden Posteditionen mit unterschiedlichen Rängen (einmal erstere besser als letztere, einmal umgekehrt), während die anderen sieben die Posteditionen mit dem gleichen Rang bedachten. Im Vergleich mit der RRHÜ ergab sich in beiden paarweisen Rankings nur einmal $RRPEo < RRHÜ$ bzw. $RRPEm < RRHÜ$, ansonsten nur $RRPEo > RRHÜ$ bzw. $RRPEm > RRHÜ$.

Die möglichen Gründe für die hohe Übereinstimmung in diesem Rating wurden bei der Besprechung der Interrater_innen-Reliabilität schon genannt. Darüber hinaus dürfte auch der Umstand, dass die **Formatierung** in den beiden Postedition identisch ist und genau mit jener des RRAT übereinstimmt, das Rating diesbezüglich erleichtert haben. R6 merkte dazu sehr anschaulich Folgendes an: Die „Kursivsetzung hebt hervor, dass die Stimme von der

Einspielung am Bildschirm kommt und die Sprecherin also nicht ‚physically present‘ ist; dies fehlt bei B“ (siehe Beispiel 17). Dieser inhaltlichen Begründung der schlechteren Bewertung der RRHÜ im Gegensatz zu beiden Posteditionen dürften sich wohl die meisten Rater_innen angeschlossen haben, allerdings ist nicht auszuschließen, dass manche Rater_innen die Posteditionen tatsächlich nur aufgrund des Umstandes, dass die Kursivsetzung in ihnen wie im RRAT vorhanden ist, besser bewerteten als die RRHÜ und sich im Gegensatz zu R6 keine Gedanken über die Hintergründe dieser Formatierung machten (siehe z. B. R5: „keine Kursivsetzung“). Dies würde für einen *priming effect* der Beschreibung dieser Fehlerkategorie auf die Rater_innen sprechen: Sie könnten trotz des Beispiels 4 aus Kapitel 5.1, das ihnen zur Verfügung stand, geglaubt haben, die **Formatierung** aus dem RRAT müsste unbedingt und in jedem Fall in die Übersetzungen übertragen werden, und alles andere sei ein Fehler, obwohl es sicher in einzelnen Fällen auch Argumente gegen eine solche Übertragung gibt (wie eben dieses Beispiel in Kapitel 5.1 zeigt).

Jedenfalls wurde die RRHÜ von den meisten Rater_innen schlechter in dieser Kategorie bewertet und die Posteditionen mehrheitlich besser. Warum zwei Rater_innen die Posteditionen unterschiedlich bewerteten, obwohl ihre **Formatierung** identisch ist, konnte nicht eruiert werden, da diese speziellen Rater_innen keine Anmerkungen machten. Es kann aber sein, dass außer der Kursivsetzung bei diesen Rater_innen und auch bei anderen noch andere Aspekte in das **Formatierungs**rating einfließen: So merkte etwa R5 an, dass in der RRHÜ eine „andere Art der Anführungszeichen“ (nämlich „» «“; siehe Beispiel 13 in Kapitel 5.2.1 und Beispiel 18 in diesem Kapitel) verwendet wird, obwohl Anführungszeichen in der Präsentation dieser Fehlerkategorie und auch in der schriftlichen Erklärung dazu nie erwähnt wurden. Dies zeigt wiederum, dass die Rating-Ergebnisse stets mit Vorsicht interpretiert werden sollten und immer im Hinblick darauf, dass womöglich nicht alle Rater_innen ihrer Aufgabe die volle Aufmerksamkeit schenken.

Trotzdem kann selbstverständlich auch für dieses Rating eine Rangfolge der Übersetzungen etabliert werden, nämlich (1) RRPEo und RRPEm und (2) RRHÜ. Die Benutzung der Fehlerkategorien durch Posteditorin 2 scheint also wiederum keinen großen Einfluss auf die Qualität der resultierenden Postedition gehabt zu haben, dies könnte aber auch am RRAT liegen: Wie in Kapitel 5.2.1 erklärt, waren in diesem Text nur ganze Absätze kursiv gesetzt. Diese **Formatierung** übertrug der DeepL-Übersetzer unverändert in die RRMÜ. Dahingegen waren im HPAT auch Einzelwörter so **formatiert**, und deren Übertragung bereitete dem DeepL-Übersetzer teilweise mehr Probleme. Dementsprechend könnte mit einem anderen Ausgangstext auch das Wissen um diese Fehlerkategorie mehr Einfluss auf die resultierende

Postedition haben als in diesem Fall. Das schlechtere Abschneiden der RRHÜ in diesem Bereich könnte einerseits, wie oben erwähnt, auf einen *priming effect* der Fehlerkategorie auf die Rater_innen hindeuten, aber andererseits auch darauf, dass die **Formatierung** von den meisten Rater_innen als inhaltlich bedeutsam empfunden wird und eine angemessene Übertragung in die Übersetzungen ihnen deshalb wichtig erscheint.

Beim Rating zur nächsten Fehlerkategorie, **Grammatik und Orthographie**, waren sich die Rater_innen überraschenderweise nicht sehr einig, obwohl, wie oben schon angemerkt, diese Kategorie neben der Formatierung und der Interpunktion wohl eine der objektivsten ist, da die Richtigkeit der in der Übersetzung verwendeten **Grammatik und Orthographie** in Regelwerken zur deutschen Sprache nachgeschlagen werden kann. Jedenfalls ergab das paarweise Ranking zwischen den beiden Posteditionen zweimal $RRPE_o < RRPE_m$, dreimal $RRPE_o > RRPE_m$ und viermal $RRPE_o = RRPE_m$. Im Vergleich mit der RRHÜ schnitt die $RRPE_o$ dreimal schlechter, zweimal besser und viermal gleich gut ab, während die $RRPE_m$ zweimal schlechter als, zweimal besser als und fünfmal gleich gut wie die RRHÜ bewertet wurde.

Es ergibt sich also eine recht ausgeglichene Situation, was auch aus den Anmerkungen der Rater_innen geschlossen werden kann: So merkte etwa R7 an, es gäbe hier „[k]ein[en] besonders große[n] Unterschied zwischen den drei Übersetzungen“, und R1, dass die „tatsächlichen Zahlen an Fehlern in allen Übersetzungen sehr niedrig“ seien. Beispiele für solche Fehler gab nur R5, nämlich u. a. die Übersetzung „Your lives = Eure Leben“ in der $RRPE_o$:

Beispiel 17: Grammatik und Orthographie

RRAT	<i>Your lives, your blood, are a down payment for the immortality of the human race as we move beyond Earth and Moon.</i>	(RRAT:[23]; Kursivsetzung im Original)
RRMÜ	<i>Eure Leben, euer Blut, sind eine Anzahlung für die Unsterblichkeit der menschlichen Rasse, während wir uns über Erde und Mond hinaus bewegen.</i>	(Kursivsetzung in der RRMÜ)
RRPE _o	Eure Leben, euer Blut, sind eine Anzahlung für die Unsterblichkeit der Menschheit, die jenseits von Erde und Mond hinausreisen wird.	(Kursivsetzung in der RRPE _o)
RRPE _m	Eure Leben, euer Blut, sind eine Vorleistung für die Unsterblichkeit der menschlichen Rasse, während wir uns über Erde und Mond hinausbewegen.	(Kursivsetzung in der RRPE _m)
RRHÜ	Euer Leben, euer Blut sind Anzahlungen für die Unsterblichkeit der Menschheit, die die Erde und den Mond hinter sich gelassen hat.	(RRHÜ:[25])

Das Plural des Wortes „Leben“ wird vom *Duden* (2021) nicht als falsch, aber doch als „selten“ eingestuft; die Übersetzung kann somit als Problemstelle gerechnet werden. Allerdings ist die Bewertung von R5 diesbezüglich inkonsistent: Wie der Beispielkasten gut zeigt, wird diese problematische Übersetzung nicht nur in der RRPEo, sondern auch in der RRPem verwendet, R5 führt dieses Beispiel aber nur bzgl. der RRPEo an. R5 nannte zwar noch ein zweites Beispiel eines Grammatikfehlers, der tatsächlich nur in der RRPEo zu finden ist („Die Schrecken [...] wird gezeigt“ statt „werden“; siehe Beispiel 15), aber trotzdem zeigen diese Ausführungen wiederum, dass die Bewertungen der Rater_innen nicht immer nachvollziehbar sind, und einige von ihnen wohl weniger konzentriert arbeiteten bzw. sich weniger intensiv mit den Übersetzungen beschäftigten als andere, wie auch schon in der Besprechung zur Interrater_innen-Reliabilität vermutet.

Soll trotz der teilweise nicht nachvollziehbaren Bewertungen wie bei den anderen Ratings eine Rangfolge der Übersetzungen etabliert werden, so ergibt sich aus den paarweisen Rankings eine, in der alle drei Übersetzungen auf Platz 1 landen. Es könnte allerdings auch die Rangfolge (1) RRPem und RRHÜ und (2) RRPEo argumentiert werden, da das zweite Beispiel von R5 doch ein recht auffälliger **Grammatik**fehler ist. Solche kommen in den anderen beiden Übersetzungen nach Ansicht der Verfasserin dieser Arbeit nicht vor. Es muss hier allerdings angemerkt werden, dass dieser Fehler höchstwahrscheinlich nicht mangelnder Sprachkompetenz von Posteditorin 1 geschuldet ist, sondern eher dem Umstand, dass die Posteditorinnen im Zuge ihres Posteditings die Funktion „Änderungen nachverfolgen“ in *Microsoft Word* verwenden mussten (siehe Kapitel 4.3). Da beide Posteditorinnen (aber v. a. Posteditorin 1) sehr viele Korrekturen durchführten, sahen sie dementsprechend auch viele Änderungen in ihrem Dokument, was das Finden von Fehlern, die durch ihre eigenen Satzumstellungen und andere Korrekturen entstanden (wie es bei diesem Fehler in der RRPEo der Fall war), vermutlich nicht gerade erleichterte. Es ist anzunehmen, dass Posteditorin 1 dieser Fehler nicht passiert wäre, hätte sie die „Änderungen nachverfolgen“-Funktion nicht verwenden müssen.

Da dieser **Grammatik**fehler in der RRPEo nicht vom DeepL-Übersetzer stammt, kann weiters anhand der Ergebnisse festgestellt werden, dass auch hier die Benutzung der Fehlerkategorien durch Posteditorin 2 keinen positiven Einfluss auf die Qualität der RRPem hatte, aber auch keinen negativen wie bei der Kategorie Bedeutung. Darüber hinaus ist der Umstand, dass manche Rater_innen offenbar auch in der RRHÜ **Grammatik- und Orthographie**fehler fanden und sie dementsprechend manchmal schlechter als die Posteditionen bewerteten, bemerkenswert. Dies veranlasste wiederum eine kurze Analyse der RRHÜ durch

die Verfasserin der vorliegenden Arbeit, bei der allerdings keine solchen Fehler in der RRHÜ festgestellt werden konnten, was selbstverständlich noch durch eine genauere Analyse bestätigt werden müsste. Trotzdem scheinen die teilweise schlechteren Bewertungen der RRHÜ wiederum wie das obige Beispiel von mangelnder Konzentration bei manchen Rater_innen zu zeugen, aber auch evtl. von mangelnder Sprachkompetenz in ihrer Muttersprache.

Auf Grammatik und Orthographie folgte die Fehlerkategorie **Interpunktion**, bei der sich die Rater_innen ebenfalls nicht unbedingt einig waren. Im ersten paarweisen Ranking bewerteten drei Rater_innen die RRPEo schlechter als die RRPEm, bei fünf war es umgekehrt, und ein_e Rater_in bewertete die beiden Posteditionen gleich gut. Verglichen mit der RRHÜ wurde der RRPEo in dieser Kategorie zweimal ein schlechterer Rang gegeben, viermal ein besserer und dreimal der gleiche. Das dritte paarweise Ranking ergab zweimal $RRPEm < RRHÜ$, dreimal $RRPEm > RRHÜ$ und viermal $RRPEm = RRHÜ$.

Den Anmerkungen nach zu urteilen, können die Rater_innen in Bezug auf dieses Rating in zwei Gruppen eingeteilt werden: eine, die die Fehlerkategorie richtig verstand, und eine, bei der dies nicht der Fall war. Die erste Gruppe, zu der R5 und R8 gehören, interpretierte die **Interpunktions**kategorie richtig und bewertete die Übersetzungen danach, ob die Zeichensetzung in ihnen den Regeln der deutschen Sprache entspricht. Interessant ist diesbezüglich, dass R5 neben der Anführung von anderen Fehlern die Zeichensetzung in der RRPEm als „zu „englisch““ bewertete⁸⁰. Im Gegensatz zu R5 gab R8 an, ihm_ihr sei bis auf die öffnenden Anführungszeichen inmitten der Rede der Herrscherin der Gesellschaft, die auch im vorigen Kapitel schon erwähnt wurden und von Posteditorin 1 richtigerweise gelöscht, in der RRPEm und auch in der RRHÜ aber belassen wurden, „in den Übersetzungen tatsächlich so gut wie nichts aufgefallen“. Dies zeugt bei R8 zwar von einem richtigen Verständnis der Fehlerkategorie, aber auch entweder von mangelnder Kompetenz bei der Zeichensetzung oder von

⁸⁰ Diese Kritik, die auf eine Übernahme der englischen Zeichensetzung durch den DeepL-Übersetzer (und dann auch Posteditorin 2) hindeutet, veranlasste eine kurze Untersuchung der RRMÜ durch die Verfasserin dieser Arbeit. Diese ergab tatsächlich eine weitgehende Übernahme der englischen Zeichensetzung, aber meist nur, wenn diese auch im Deutschen korrekt war. War eine übernommene Zeichensetzung in der Zielsprache nicht korrekt, wurde sie so verändert, dass sie den Regeln des Deutschen entsprach, allerdings ohne Syntaxumstellungen vorzunehmen. Dies war aber in einigen Ausnahmen nicht der Fall, darunter regelmäßig bei direkten Reden, die auf ein Fragezeichen enden. Bei solchen ist im Englischen vor dem Begleitsatz kein Komma notwendig (vgl. Lexico.com 2021b), im Deutschen aber schon (vgl. Bibliographisches Institut o. J.b). Der DeepL-Übersetzer fügte dieses Komma allerdings nicht ein (dies tat dann Posteditorin 2). Auffällig ist an der Zeichensetzung in der RRMÜ ebenfalls, dass die Semikola aus dem Englischen übernommen wurden. Ob diese in den betreffenden Sätzen angemessen sind oder nicht, liegt im Auge des Betrachtenden (vgl. Bibliographisches Institut o. J.c), aber Semikola sind nach Ansicht der Verfasserin dieser Arbeit im Deutschen doch viel seltener in Gebrauch als im Englischen. Es könnte sein, dass R5 diese Auffassung teilte und deswegen sowie aufgrund der oft übernommenen Syntax (die die **Interpunktion** selbstverständlich beeinflusst) die Zeichensetzung in der RRPEm, in der diese Semikola nicht geändert wurden, als „zu „englisch““ einstufte.

mangelnder Konzentration beim Rating, gibt es doch in beiden Posteditionen Fehler diesbezüglich (etwa Bindestriche, die Gedankenstriche sein sollten (siehe Beispiel 11a in Kapitel 5.2.1), oder fehlende Kommata nach direkten Reden oder um Infinitivgruppen).

Die zweite Gruppe interpretierte im Gegensatz dazu offenbar die Fehlerkategorie so, dass die **Interpunktion** in den Übersetzungen stets exakt der Zeichensetzung im Ausgangstext entsprechen sollte, um korrekt zu sein⁸¹. Deutliche Hinweise auf diese fälschliche Auffassung der Kategorie **Interpunktion** gibt es etwa in den Anmerkungen von R1 („in C gibt es Interpunktion, die nicht mit der im AT übereinstimmt“) und von R7 („In Übersetzung B ist so gut wie alles falsch“). Diese letzte Behauptung stimmt nach Ansicht der Verfasserin dieser Arbeit nicht, da die Zeichensetzung in der RRHÜ durchaus den Regeln der deutschen Sprache entspricht, bis auf die oben schon erwähnten Anführungszeichen inmitten der direkten Rede. Trotzdem ist die Zugehörigkeit von R7 zur zweiten Gruppen nicht ganz eindeutig, denn diesem_dieser Rater_in fielen auch die Kommata in der RRPEm, die „einfach 1:1 übertragen [wurden]“, negativ auf (ähnlich wie R5). Auch bei R6 ist es nicht ganz klar, zu welcher Gruppe er_sie gehört, doch da diese_r Rater_in nur vermeintliche Fehler aufzählte, bei denen die Zeichensetzung aus dem Ausgangstext verändert wurde, und keine tatsächlichen **Interpunktionsfehler**, kann auch hier auf eine fälschliche Interpretation der entsprechenden Fehlerkategorie geschlossen werden.

Diese unterschiedlichen Auffassungen der Fehlerkategorie **Interpunktion** seitens der Rater_innen machen die weitere Interpretation dieser Rating-Ergebnisse schwierig (wenn nicht sogar unmöglich), denn auch wenn die Bewertungen von R1, R6 und R7 ausgeklammert werden, kann nicht ausgeschlossen werden, dass auch noch andere Rater_innen diese Fehlerkategorie falsch interpretierten. Darüber hinaus besteht die Möglichkeit, dass auch bei anderen Rater_innen (nicht nur bei R8) mangelnde Konzentration beim **Interpunktionsrating** herrschte bzw. die Zeichensetzungsregeln im Deutschen nicht allen Rater_innen hinreichend bekannt waren. Die Etablierung einer Rangfolge der Übersetzungen, wie dies bei den bisher besprochenen Ratings geschah, erscheint somit nicht ratsam. Müsste die Verfasserin dieser Arbeit ihre eigene Rangfolge aufstellen, würde sie (1) RRHÜ und (2) RRPEo und RRPEm lauten, doch diese (wenn auch fundierte) persönliche Meinung ist selbstverständlich wenig

⁸¹ Dies ist durch die Fehlerkategorie tatsächlich aber nur in Ausnahmefällen wie jenem in Beispiel 6 in Kapitel 5.1 gemeint (und auch dann nur, wenn eine Übertragung nicht gegen die Zeichensetzungsregeln der Zielsprache verstößt), da eine durchgängig exakte Übertragung der Zeichen aus dem Ausgangstext in die Übersetzungen aufgrund der unterschiedlichen **Interpunktionsregeln** in der Ausgangs- und Zielsprache sprachlich nicht korrekt ist. Den Anmerkungen der Rater_innen nach zu urteilen, hätte aber die Prävalenz der Regeln der Zielsprache über die **Interpunktionsübertragung** aus dem Ausgangstext in der Erklärung der Fehlerkategorie deutlicher hervorgehoben werden müssen und nicht als selbstverständlich angenommen werden sollen.

geeignet, um daraus Schlüsse über den Einfluss der Benutzung der Fehlerkategorien durch Posteditorin 2 auf die Qualität ihrer Postedition bzgl. der **Interpunktion** zu ziehen. Darüber hinaus besteht die Möglichkeit, dass auch Posteditorin 2 diese Fehlerkategorie falsch verstand und deshalb bis auf die in der Beschreibung dieser Fehlerkategorie genannten Fehler (Anführungszeichen, Binde- und Gedankenstriche etc.) nur wenig ausbesserte. Aus diesen Gründen wird von einer weiteren Interpretation der Rating-Ergebnisse zu dieser Kategorie gänzlich abgesehen.

Beim nächsten Fehlerkategorie-Rating ging es um die Kategorie **IUT**. Das erste paarweise Ranking ergab diesbezüglich fünfmal eine schlechtere und viermal eine bessere Bewertung für die RRPEo verglichen mit der RRPEm. Darüber hinaus wurde die RRPEo auch fünfmal schlechter als die RRHÜ bewertet, zweimal besser und zweimal gleich gut. Für die RRPEm ergab sich ebenfalls fünfmal eine schlechtere Bewertung als für die RRHÜ, aber auch dreimal eine bessere und einmal eine gleich gute.

Diese Kategorie wurde von zwei Rater_innen in Bezug auf die RRPEo im ersten allgemeinen Rating in ihren Anmerkungen genannt, sie dürfte also, genauso wie die Bedeutung, eine wichtige Kategorie sein. Dies zeigt sich auch an den vielen Anmerkungen, die es zu diesem Rating gab. Die meisten der Rater_innen, die Anmerkungen machten, fanden wie schon beim allgemeinen Rating die Übernahme vieler englischer **IUT** in die deutsche RRPEo mehr oder weniger störend, nämlich R1, R4, R6 und R8. Letztere_r argumentierte seine_ihre Kritik so, dass deutsche Übersetzungen „für ein deutschsprachiges Publikum leichter verständlich“ gewesen wären, „da auch nicht jede*r Leser*in Englisch kann bzw. jedes englische Wort versteht“. Zumindest diese vier Rater_innen stimmten dementsprechend der in Kapitel 5.2.1 angestellten Vermutung zu, die Mischung von englischen und deutschen **Termini** würde sich störend auf den Lesefluss auswirken und für einige Leser_innen eher hinderlich sein.

Im Gegensatz dazu wurde die RRPEo von R7 zwar wegen der Übersetzung „Goldbrows = Goldene“ kritisiert (siehe Beispiel 12 in Kapitel 5.2.1), aber ansonsten fand diese_r Rater_in wie auch R5 die Behandlung der **IUT** in der RRPEo „gelungener“, weil die „wörtliche“ Übersetzung der **IUT** in der RRPEm „öfters zu stockigen Formulierungen führ[e]“. In Bezug auf die **IUT** in der RRPEm wurde außerdem die Nicht-Übereinstimmung von „HoloBüchse“ und der dazugehörigen Abkürzung „HC“ (siehe Kapitel 5.2.1) von R6 kritisiert, während die Übersetzung „Goldbrows = Goldrüben“ gelobt wurde (R6; siehe wiederum Beispiel 12 in Kapitel 5.2.1).

Auch zur RRHÜ waren die Rater_innen geteilter Meinung: R5 stufte manche der IUT-Übersetzungen darin als „gänzlich inkorrekt“ oder „normalisierend“ ein (etwa „frysuit = Kochanzug“ (siehe Beispiel 14 in Kapitel 5.2.1) oder „Flush = Dusche“ (siehe Beispiel 18)). R6 schloss sich dieser Kritik an, aber es gäbe laut R6 „vereinzelt [...] doch gute Ansätze“, wie etwa „holoCan = HoloBox“ (siehe Beispiel 11c in Kapitel 5.2.1). R7 kritisierte außerdem die Streichung des IUT „Aureate“ in der RRHÜ (siehe Beispiel 12 in Kapitel 5.2.1), bewertete aber die IUT-Übersetzungen in der RRHÜ und in der RRPEo als gleich gut, während R1 die RRHÜ und die RRPEm in dieser Hinsicht gleich gut einstufte.⁸²

Aus den Bewertungen der Rater_innen und ihren Anmerkungen kann folgende Rangfolge etabliert werden: (1) RRHÜ, (2) RRPEm und (3) RRPEo. Auch die Rangfolge (1) RRHÜ und RRPEm und (2) RRPEo könnte evtl. angedacht werden, es scheint aber klar, dass die RRPEo diesbezüglich insgesamt auf dem letzten Platz landet, da nicht nur in den Anmerkungen zum IUT-Rating relativ viel Kritik an der Behandlung der IUT darin geübt wurde, sondern auch zusätzlich dazu von R1 und R2 im ersten allgemeinen Rating. Aufgrund dieser Erwähnungen schon beim ersten Rating ist es auch unwahrscheinlich, dass die Fehlerkategorie IUT einen *priming effect* auf alle Rater_innen wie auf Posteditorin 2 (siehe Kapitel 5.2.1) ausübte, auf einige (v. a. jene, die keine Anmerkungen machten) ist aber ein solcher Effekt nicht auszuschließen. Trotzdem kann ausgehend von den Ergebnissen vorsichtig behauptet werden, dass der *priming effect* der Kategorie IUT auf Posteditorin 2 durchaus positiv war. Die Tatsache, dass sich die Rater_innen überhaupt so viele Gedanken bei diesem Rating machten (und auch das Bedürfnis empfanden, dazu Anmerkungen zu machen, obwohl das nicht verlangt war), deutet auch darauf hin, dass der Übersetzung von IUT in Science-Fiction- und Fantasy-Werken (besonders, wenn sie so reich an IUT sind wie *Harry Potter* und *Red Rising*) eine große Bedeutung beigemessen wird. Ebenfalls deutlich wurde die schon oft erwähnte Schwierigkeit der Übersetzung von IUT (es gab schließlich fundierte Kritik an allen drei Übersetzungen diesbezüglich), aber auch, dass diese Kategorie tatsächlich sehr subjektiv ist und zu Recht als Problembereich eingeschätzt werden sollte.

Bei der nächsten Kategorie bzw., genauer gesagt, dem nächsten Problembereich, **Stil**, spielt zwar grundsätzlich auch die Subjektivität eine große Rolle, aber die Rater_innen waren sich diesbezüglich weitaus einiger als beim IUT-Rating: Das erste paarweise Ranking ergab

⁸² Um die Kritik/das Lob an den Übersetzungen dieser IUT zu verstehen, ist wohl eine Erklärung für jene IUT notwendig, die nicht schon in Kapitel 5.2.1 dieser Masterarbeit erklärt wurden: „Flush“ bezeichnet ein Gebäude, in dem sich die Minenarbeiter und Roten Frauen nach ihrer Arbeitsschicht in speziellen Röhren waschen. Dabei wird kein Wasser eingesetzt, sondern ein Strom von Luft (vgl. Brown 2014a:[20f.]), weswegen die Übersetzung „Dusche“ vermutlich als „normalisierend“ empfunden wurde. Eine „holoCan“ ist eine Art riesiger Bildschirm in Würfelform (vgl. Brown 2014a:[22]).

achtmal RRPEo > RRPEm, aber nur einmal RRPEo = RRPEm. Verglichen mit der RRHÜ wurde die RRPEo fünfmal schlechter, dreimal besser und einmal gleich gut bewertet, während die RRPEm gleich achtmal schlechter und nur einmal besser als die RRHÜ eingestuft wurde.

Wie IUT wurde auch der **Stil** in den Anmerkungen zum ersten allgemeinen Rating genannt, die Kategorie scheint also ebenfalls recht auffällig und wichtig zu sein. In den Anmerkungen zum **Stil**-Rating selbst wurde wiederum v. a. die RRPEm kritisiert: So schrieb R1, dass die RRPEm „hier mit Abstand am schlechtesten ab[schneidet]“. In seinen_ihren weiteren Ausführungen stimmte R1 mit R5 überein und merkte an, die RRPEm sei „sehr oft nahe an der englischen Konstruktion geblieben, wodurch die Übersetzung teilweise recht holprig ist“. R6 merkte zusätzlich dazu noch an, dass manche Passagen „nicht ideal gelöst“ oder „nicht flüssig“ sind, etwa „während des Gehens“ (siehe Beispiel 11a in Kapitel 5.2.1).

Auch zu den anderen beiden Übersetzungen gab es bisweilen ähnliche Kritik, die allerdings in beiden Fällen nicht so heftig ausfiel: R5 entdeckte bei der RRPEo „hin und wieder Übernahmen des AT“ (etwa die Übersetzung „bringers of chaos = Bringer des Chaos“), während R6 „etwas holprige Formulierungen“ bei der RRHÜ feststellte (etwa „You didn’t marry me because I tell you everything, remember. (RRAT:[24]) = Vergiss nicht, dass du mich nicht geheiratet hast, weil ich dir alles sage.“ (RRHÜ:[26])). R6 lobte darüber hinaus die RRPEo mit der Anmerkung, sie sei „insgesamt am angenehmsten zu lesen“ und es herrsche eine „gute Dynamik“.

Die insgesamt hohe Übereinstimmung der Rater_innen in ihren Ergebnissen und ihren Anmerkungen ist, wie oben bei der Interrater_innen-Reliabilität schon erwähnt, bemerkenswert, spielt doch die Subjektivität bei diesem Problembereich eine sehr große Rolle. Diese hohe Übereinstimmung erlaubt es auch, eine eindeutige Rangfolge der Übersetzungen bzgl. des **Stils** zu etablieren, nämlich (1) RRHÜ, (2) RRPEo und (3) RRPEm. Das Wissen um die Fehlerkategorie **Stil** bei Posteditorin 2 hat dementsprechend keine positiven Auswirkungen auf die diesbezügliche Qualität ihrer Postedition gehabt. Tatsächlich scheinen sich ihre geringeren kognitiven Kapazitäten für die zielsprachlichen Versionen sowie ihre Bedenken bzgl. des Ausmaßes ihrer Änderungen (siehe Kapitel 5.2.1) negativ auf ihre Postedition ausgewirkt zu haben, weil dadurch in ihr mehr unveränderte Teile der RRMÜ übrigblieben, in der, wie oben schon erklärt, der Satzbau aus dem RRAT oft übernommen wurde. Dahingegen hatten die zusätzlichen **Stil**-Korrekturen von Posteditorin 1 und ihre freiere Art des Posteditings, das sich bei ihr nicht so sehr auf den RRAT konzentrierte, positive Effekte auf die Natürlichkeit, Flüssigkeit und Angemessenheit der Sprachverwendung in ihrer

Postedition, auch wenn die RRPEo gesamt gesehen nicht ganz an die RRHÜ herankam. Die Tatsache, dass sie aber überhaupt dreimal besser als die reine Humanübersetzung bewertet wurde, ist bemerkenswert und zeigt doch, dass auch in diesem Bereich durchaus Potential bei L-NMT+Postediting besteht. Dieses Potential scheint aber sehr stark von den Posteditor_innen und ihrem persönlichen Übersetzungs- und Posteditingstil abzuhängen.

Bei der nächsten Fehlerkategorie, **Tempus und Modus von Verben**, herrschte laut den Kappa-Werten die zweithöchste Übereinstimmung zwischen den Rater_innen. Die paarweisen Rankings ergaben ein etwas differenzierteres Bild: Im Vergleich mit der RRPEm erzielte die RRPEo sechsmal einen besseren Rang und dreimal den gleichen, und auch die paarweisen Rankings zwischen RRPEo und RRHÜ fielen so aus (sechsmal RRPEo > RRHÜ, dreimal RRPEo = RRHÜ). Beim Vergleich zwischen RRPEm und RRHÜ ergab sich dahingegen dreimal RRPEm < RRHÜ, viermal RRPEm > RRHÜ und zweimal RRPEm = RRHÜ.

Obwohl R2 in den Anmerkungen zum ersten allgemeinen Rating den Einsatz von **Tempora** in der RRPEm kritisierte, wurde in den Anmerkungen zu diesem spezifischen Fehlerkategorie-Rating überraschenderweise nur die RRHÜ kritisiert, so z. B. von R5, dem_der „einige schwerwiegende Zeitenfehler“ auffielen. Welche das sind, gab R5 nicht an, aber dafür merkte R1 an, in der RRHÜ seien „vor allem einige Fehler in den HC-Durchsagen gemacht [worden], die, anders als das Original, andeuten, dass die Menschheit bereits auf anderen Planeten lebt“. Konkreter wurde R1 nicht, aber höchstwahrscheinlich meinte er_sie damit etwa die in Beispiel 17 ersichtliche Übersetzung „as we move beyond Earth and Moon = Menschheit, die die Erde und den Mond hinter sich gelassen hat“ in der RRHÜ ([25]). Das in der RRHÜ verwendete Perfekt suggeriert im Gegensatz zur *present tense* im RRAT ([23]), dass sich die Menschheit schon über Erde und Mond hinweg ausgebreitet hat (und dies nicht erst zum Zeitpunkt des Gesprochenen geschieht), d. h. schon auf anderen Planeten des Sonnensystems lebt. Damit wird das sorgsam aufgebaute Lügenkonstrukt der herrschenden Klasse gegenüber der Roten Arbeiterklasse geschwächt, die im Glauben gehalten wird, sie müsse im Untergrund der Planeten schuften, um das Terraforming dieser Planeten voranzutreiben, damit auch die anderen Farben, d. h. die ganze Menschheit, dort leben können. Durch das Perfekt in der RRHÜ wird also der Inhalt verzerrt, was laut R1 und auch R6 nicht nur einmal passiert.

In Bezug auf die RRPEo oder die RRPEm wurden solche Fehler von den Rater_innen nicht erhoben, doch die RRPEm erhielt mehrheitlich schlechtere Bewertungen in dieser Kategorie. Ein möglicher Grund dafür könnte sein, dass Posteditorin 2 in ihrer RRPEm manchmal das Perfekt verwendete bzw. die RRMÜ diesbezüglich nicht auf das Präteritum

änderte, während Posteditorin 1 solche Änderungen schon durchführte. Diese Änderungen wurden zwar in Kapitel 5.2.1 der Kategorie Stil zugeordnet (weil es tatsächlich angemessener ist, in einem Absatz durchgängig das Präteritum oder das Perfekt zu verwenden und nicht zu mischen, die beiden Zeitformen aber grundsätzlich dasselbe oder zumindest Ähnliches ausdrücken), aber das heißt nicht, dass die Rater_innen diese Verwendung der **Tempora** nicht (auch) bei der Kategorie **Tempus und Modus von Verben** evaluierten. Selbstverständlich kann die unterschiedliche Bewertung der beiden Posteditionen bei manchen Rater_innen aber auch an mangelnder Konzentration o. Ä. liegen.

Jedenfalls ergibt eine Zusammenfassung dieser Ergebnisse und Anmerkungen folgende Rangfolge der Übersetzungen bzgl. der in ihnen verwendeten **Tempi und Modi**: (1) RRPEo, (2) RRPEm und (3) RRHÜ. Wiederum dürften also die Fehlerkategorien Posteditorin 2 nicht geholfen haben, in diesem Bereich eine qualitativ hochwertigere Übersetzung zu erstellen, doch der Umstand, dass beide Posteditionen insgesamt besser abschnitten als die RRHÜ, zeigt auch, dass L-NMT+Postediting in Bezug auf diese Fehlerkategorie qualitätsvolle Übersetzungen liefern kann.

Das letzte Fehlerkategorie-Rating war jenes zu **Tonfall und Register**. Das paarweise Ranking zwischen der RRPEo und der RRPEm ergab folgende Ergebnisse: einmal RRPEo < RRPEm, fünfmal RRPEo > RRPEm und dreimal RRPEo = RRPEm. Im Vergleich zur RRHÜ wurde die RRPEo zweimal schlechter, zweimal besser und fünfmal gleich gut bewertet, während die RRPEm fünfmal eine schlechtere Bewertung, zweimal eine bessere und zweimal eine gleich gute im Vergleich zur RRHÜ erzielte.

Diese Fehlerkategorie wurde wiederum indirekt im ersten allgemeinen Rating genannt, nämlich im eher negativen Sinne von R2 in Bezug auf die RRPEo und die RRHÜ. Auch R6 fand bei diesen beiden Übersetzungen einen Kritikpunkt zu **Tonfall und Register** und schrieb, der „Dialog zwischen Darrow & Eo [sei] manchmal etwas zu formell“. R5 erwähnte in seiner_ihrer Anmerkung die Dialoge nicht, aber sein_ihr Beispiel für ein „höheres Register [...] [in der RRHÜ] als im AT“ bezieht sich ebenfalls auf eine direkte Rede (siehe Beispiel 18 auf der nächsten Seite).

R1 kritisierte dahingegen die RRPEm: Darin gebe es „immer wieder zu hochgestochene Ausdrücke“. Ein Beispiel dafür gab R1 allerdings nicht an. Im Gegensatz dazu merkte R6 positiv zur RRPEm an, dass in dieser Übersetzung der „Unterschied zwischen formeller & informeller mündlicher Sprachverwendung klar erkennbar“ sei.

Beispiel 18: Tonfall und Register

RRAT	“You didn’t go in the Flush,” I say, smiling. “Didn’t have time. Had to skirt out of the Webbery to pick something up.”	(RRAT:[24])
RRMÜ	"Du bist nicht in den Flush gegangen", sage ich und lächle. "Hatte keine Zeit. Musste aus dem Webbery raus, um etwas abzuholen."	
RRPEo	„Du warst nicht im Flush“, sage ich und lächle. „Ich hatte keine Zeit. Musste aus der Weberei raus , um was abzuholen.“	
RRPEm	„Du warst nicht in der Spülung“, sage ich und lächle. „Hatte keine Zeit. Musste schnell aus der Netzerei raus , um etwas abzuholen.“	
RRHÜ	»Du warst nicht in der Dusche«, sage ich lächelnd. »Keine Zeit. Musste schnell die Weberei verlassen , um noch etwas zu besorgen.«	(RRHÜ:[25])

Aus diesen Ausführungen wird klar ersichtlich, dass sich zumindest die drei Rater_innen, die Anmerkungen machten, nicht unbedingt einig in ihrer Kritik waren, was die Rolle der Subjektivität zeigt, die auch bei dieser eher als Problembereich einzuschätzenden Fehlerkategorie zu beachten ist. Trotzdem kann aus den Ergebnissen eine Rangfolge etabliert werden, nämlich (1) RRPEo und RRHÜ und (2) RRPEm. Wiederum kann dementsprechend der *priming effect* der Fehlerkategorien auf Posteditorin 2 in Bezug auf **Tonfall und Register** insgesamt als eher negativ eingestuft werden. Andererseits zeigt sich auch bei dieser Kategorie durch die Gleichwertigkeit von RRPEo und RRHÜ, dass L-NMT+Postediting durchaus mit einer reinen literarischen Humanübersetzung in diesem Bereich mithalten kann.

Damit bleibt noch die Besprechung der Ergebnisse des **zweiten allgemeinen Ratings**. Die paarweisen Rankings dazu ergaben beim Vergleich RRPEo – RRPEm einmal eine schlechtere Bewertung für die RRPEo, siebenmal eine bessere und einmal eine gleich gute. Im zweiten paarweisen Ranking ergab sich je viermal RRPEo < RRHÜ und RRPEo > RRHÜ sowie einmal RRPEo = RRHÜ. Verglichen mit der RRHÜ wurde die RRPEm sechsmal schlechter, zweimal besser und einmal gleich gut bewertet.

Die Anmerkungen der Rater_innen zu diesem Rating fielen unterschiedlich aus: R1 und R5 stimmten miteinander überein, dass die RRPEm am schlechtesten sei; R1 lieferte dafür als Begründung, dass sich die RRPEm „zu nah am Englischen hält“. Auch R6 war der Meinung, dass die RRPEo „flüssiger zu lesen“ sei. Damit wurde in zwei Anmerkungen zum zweiten allgemeinen Rating die Kategorie Stil noch einmal erwähnt und sogar von R6 als „besonders wichtige Kategorie“ hervorgehoben.

Auch die IUT wurde in den Anmerkungen angeführt, nämlich zweimal im Hinblick auf die RRPEo: R1 erwähnte die IUT hier in negativer Hinsicht, und auch R6 schrieb, die RRPEo sei „etwas komplizierter[...]“ bei den IUT. Immer noch laut R6 könnte allerdings diese Kompliziertheit vernachlässigt werden, da sie sich „aus dem Kontext des Buches“ ergäbe. Zur Kategorie IUT kann dementsprechend gesagt werden, dass sie zwar auch als sehr wichtig empfunden wird, aber dem Stil nachgereiht ist.

Schließlich wurde auch Kritik an der RRHÜ geübt, diese fiel aber eher unspezifisch aus: So schrieb R1, es gäbe in dieser Übersetzung „einige Baustellen [...]“, die ich beim ersten schnellen Lesen nicht bemerkt habe“, während R5 die RRHÜ mit der RRPEo gleichsetzte, aber anführte, beide hätten „an jeweils unterschiedlichen Stellen gewisse Schwächen“. In starkem Kontrast dazu kritisierte R6 die RRHÜ mit den Worten „B fällt insgesamt eindeutig zurück“.

Aus den paarweisen Rankings und diesen Anmerkungen kann auch für das **zweite allgemeine Rating** eine Rangfolge der Übersetzungen aufgestellt werden, die folgendermaßen aussieht: (1) RRPEo und RRHÜ, (2) RRPEm. Die Gesamtsituation hat sich damit verglichen mit dem ersten allgemeinen Rating nicht wesentlich verändert: Beide Posteditionen bleiben auf dem ihnen schon zuvor zugewiesenen ersten bzw. letzten Platz respektive, nur die RRHÜ rückt einen Platz auf. Allerdings könnte aufgrund der heftigen Kritik von R6 an der RRHÜ auch die Rangfolge (1) RRPEo, (2) RRHÜ und (3) RRPEm argumentiert werden – in diesem Fall hätte sich die Gesamtsituation überhaupt nicht verändert, obwohl an den einzelnen Rängen, die die Rater_innen vergaben, doch erkennbar ist, dass bei gleich sieben von ihnen die Bewertungen nach ihrer Beschäftigung mit den Fehlerkategorien etwas anders ausfielen.

Der Umstand, dass die Gesamtsituation vor und nach den Fehlerkategorie-Ratings mehr oder weniger gleich blieb, sich die einzelnen Bewertungen aber doch änderten, deutet darauf hin, dass die Fehlerkategorien durchaus einen Einfluss auf die Bewertungen hatten, dieser sich aber je nach Rater_in anders auswirkte: Unterschiedliche Rater_innen maßen wohl unterschiedlichen Fehlerkategorien beim zweiten allgemeinen Rating Wichtigkeit zu und vergaben ihre Bewertungen dementsprechend, sodass sich insgesamt gesehen in etwa wieder die Rangfolge des ersten allgemeinen Ratings ergab.

Darüber hinaus kann über zumindest zwei Fehlerkategorien, Stil und IUT, behauptet werden, dass sie den Rater_innen bedeutsam erschienen, denn sie wurden beide im ersten sowie im zweiten allgemeinen Rating mehrmals erwähnt. Auch die Kategorien Bedeutung sowie Tonfall und Register scheinen den Rater_innen relativ wichtig gewesen zu sein, bei

Tempus und Modus von Verben ist das schwer zu sagen, denn diese Kategorie könnte teilweise, wie oben erläutert, auch mit der Stil-Kategorie zusammengefallen sein. Jedenfalls erscheint es interessant, dass u. a. genau den drei Fehlerkategorien bei den Ratings Bedeutung zugemessen wurde, bei denen die Subjektivität eine große Rolle spielt und die demnach eher als Problembereiche zu sehen sind.

Aus all diesen Ausführungen kann **zusammenfassend** abgeleitet werden, dass für die Rater_innen die RRPEo die beste Übersetzung war, gefolgt von der RRHÜ, während die RRPEm etwas abgeschlagen auf Platz 3 landete. Es ergaben sich mehr oder weniger große qualitative Unterschiede zwischen den beiden Posteditionen in Bezug auf die Fehlerkategorien Auflösung von Anaphern, Bedeutung, IUT, Stil, Tempus und Modus von Verben sowie Tonfall und Register. Bei den Kategorien Auslassungen und Hinzufügungen, Formatierung sowie Grammatik und Orthographie schnitten die Posteditionen dahingegen insgesamt gleich gut ab. Zur Kategorie Interpunktion ist es nicht möglich, eine solche Feststellung zu treffen, da diese von einigen Rater_innen offenbar falsch verstanden wurde.

Insgesamt und auch im Hinblick auf eine Mehrheit der Fehlerkategorien scheint sich dementsprechend die Benutzung dieser Fehlerkategorien gar nicht oder eher negativ auf die Qualität der RRPEm ausgewirkt zu haben (nur bei den IUT war dies merklich anders), aber auch der Posteditingstil von Posteditorin 2 sowie ihre geringere Erfahrung beim Editieren von Übersetzungen haben (höchstwahrscheinlich) zur schlechteren Bewertung ihrer Postedition beigetragen. Trotzdem zeigen die Ergebnisse auf der anderen Seite, dass Potential für die Kombination L-NMT+Postediting gegeben ist (immerhin landete in fast allen zusammenfassenden Rangfolgen zumindest eine Postedition, hauptsächlich die RRPEo, auf Platz 1), wobei den Posteditor_innen in dieser Kombination eindeutig sehr viel Bedeutung beigemessen werden sollte.

Abschließend ist auch noch einmal darauf hinzuweisen, dass diese Schlussfolgerungen nur vorläufiger Natur sind, und zwar aufgrund der geringen Zahl an Rater_innen, der insgesamt niedrigen Interrater_innen-Reliabilität sowie des Umstandes, dass offenbar manche Rater_innen weniger konzentriert und/oder überlegt arbeiteten als andere und/oder noch dazu manche Fehlerkategorien nicht richtig verstanden. Weiterführende Studien über den Einfluss von Fehlerkategorien auf die Qualität der Postedition einer literarischen maschinellen Übersetzung sind demnach wünschenswert. Um Möglichkeiten für solche Studien sowie um eine konkrete Beantwortung der Forschungsfragen geht es auch im nächsten Kapitel dieser Masterarbeit.

6 Diskussion

In diesem Kapitel werden die schon ausführlich besprochenen Ergebnisse der durchgeführten Untersuchungen und Analysen knapp als Antworten auf die Forschungsfragen (die gemeinsam mit den dazugehörigen Hypothesen noch einmal angeführt sind) präsentiert. Diese Antworten werden außerdem in Beziehung zu den Hypothesen gesetzt sowie in Bezug auf die Forschungsliteratur diskutiert, wobei auch Möglichkeiten für zukünftige Studien in Betracht gezogen werden.

Tabelle 6: FF1 und Hypothese

FF1	Welche Fehlerkategorien lassen sich unter Berücksichtigung der Forschungsliteratur zur literarischen NMT anhand der von DeepL erstellten Übersetzung eines Ausschnitts aus <i>Harry Potter and the Philosopher's Stone</i> ins Deutsche etablieren?
Hyp.	<ul style="list-style-type: none"> – Interpunktion (v. a. Anführungszeichen) – Wörter und Mehrwortausdrücke, die für den betreffenden Roman erfunden wurden bzw. dort auf spezifische Art und Weise verwendet werden – Auslassungen – Übertragung der Bedeutung (falsch oder nicht vollständig) – Auflösung von Anaphern über Satzgrenzen hinweg

Die **Antwort auf FF1** sind die in Kapitel 5.1 ausführlich präsentierten zehn Fehlerkategorien, nämlich 1) Auflösung von Anaphern, 2) Auslassungen und Hinzufügungen, 3) Bedeutung (mit den Unterkategorien fixe Wendungen und *phrasal verbs*, Polysemie sowie Silbentrennung), 4) Formatierung, 5) Grammatik und Orthographie, 6) Interpunktion, 7) IUT, 8) Stil, 9) Tempus und Modus von Verben sowie 10) Tonfall und Register. Wie durch einen Vergleich der Auflistung in der Tabelle sowie dieser Antwort gut ersichtlich wird, haben sich alle Hypothesen bzgl. FF1 bewahrheitet, wobei die Auflösung von Anaphern und die Auslassungen in der HPMÜ weniger gravierend waren als Bedeutung, Interpunktion und IUT, d. h., es wurden bzgl. ersterer deutlich weniger Fehler gefunden als bzgl. letzterer. Auch die Korrekturen der beiden Posteditorinnen an der RRMÜ spiegeln diese Beobachtung wider. Die der Kategorie Bedeutung untergeordneten Kategorien wurden zwar nicht in dieser Form antizipiert, aber sie gliedern sich dennoch in die Hypothese ein, da sie ebenfalls eine falsche oder nicht vollständige Übertragung der Bedeutung beschreiben.

Wie in Kapitel 5.1 schon dargestellt, werden die meisten der entwickelten Fehlerkategorien in der **Forschungsliteratur** zur L-NMT als problematische Bereiche erwähnt und kommen auch in den drei während der Entwicklung als Hilfestellung verwendeten Fehlertypologien (Matusov (2019); TAUS (o. J.); Tezcan et al. (2019); siehe Kapitel 4.2) in gleicher oder zumindest ähnlicher Form vor. Dies zeigt die untenstehende Tabelle in Kurzform, wobei

„tlw.“ heißt, dass die entsprechende Fehlerkategorie nicht in genau der Form, die sie im Rahmen dieser Masterarbeit annimmt, erwähnt wird bzw. enthalten ist, sondern nur ansatzweise oder mit einem anderen Fokus. Ist weder „ja“ noch „tlw.“ angekreuzt, bedeutet das, dass die entsprechende Fehlerkategorie überhaupt nicht erwähnt wird bzw. enthalten ist. Die Angabe in Klammern hinter den „x“ im zweiten Bereich der Tabelle führt an, in wie vielen Werken zur L-NMT aus dem Forschungsüberblick (Kapitel 3 bzw. siehe auch Tabelle 2 in Kapitel 4.2) die jeweilige Fehlerkategorie als problematischer Bereich erwähnt wird. Welche Werke das genau sind, wurde in Kapitel 5.1 schon angeführt.

Tabelle 7: Entwickelte Fehlerkategorien in der Forschungsliteratur

Fehlerkategorie	Als probl. Bereich erwähnt		In Matusov (2019) enth.		In TAUS (o. J.) enth.		In Tezcan et al. (2019) enth.	
	ja	tlw.	ja	tlw.	ja	tlw.	ja	tlw.
Auflösung von Anaphern	x (2)	x (3)		x			x	
Auslassungen & Hinzufügungen	x (1)	x (2)	x		x		x	
Bedeutung (inkl. Unterkategorien)	x (10)		x		x		x	
Formatierung					x			
Grammatik & Orthographie		x (1)			x		x	
Interpunktion	x (1)				x		x	
IUT		x (3)		x		x		x
Stil	x (4)			x	x		x	
Tempus & Modus v. Verben	x (4)			x				x
Tonfall & Register	x (5)	x (1)	x		x		x	

Wie an der Tabelle ersichtlich wird, werden sieben der entwickelten Fehlerkategorien in der **Forschungsliteratur** als problematischer Bereich erwähnt, zwei nur teilweise und eine gar nicht. In Bezug auf die Fehlertypologien sind in Matusov (2019) drei der Fehlerkategorien ganz, vier teilweise und wiederum drei gar nicht enthalten, während in der DQF-MQM-Fehlertypologie (TAUS o. J.) und in jener von Tezcan et al. (2019) jeweils sieben ganz, zwei teilweise und eine gar nicht inkludiert ist. In den dieser Tabelle zugrunde liegenden Werken werden meist andere Zielsprachen als Deutsch und, in einem geringeren Ausmaß, auch andere Ausgangssprachen als Englisch verwendet (siehe Tabelle 2 in Kapitel 4.2), und die Fehlertypologien sind ebenfalls nicht auf eine Übersetzungsrichtung oder Sprachkombination

beschränkt. Folglich erweisen sich die meisten der entwickelten Fehlerkategorien auf den ersten Blick auch als sprachübergreifend (wobei diejenigen, die öfter erwähnt werden, sich auch auf mehr Sprachen ausdehnen als die anderen), doch selbstverständlich müsste diese Beobachtung noch mittels einer expliziten Anwendung der Fehlerkategorien auf L-NMT-Outputs in anderen Zielsprachen und aus anderen Ausgangssprachen überprüft werden.

Darüber hinaus ist zu den zehn Fehlerkategorien außerdem noch zu sagen, dass sie **keinesfalls** den Anspruch auf den Status einer Fehlertypologie erheben (können), die **für alle Arten von LMT-Output** anwendbar wäre. Zwar kann aus der geringen Anzahl an NZ-Korrekturen in den Posteditionen (von denen die meisten ohnehin anderen Änderungen der Posteditorinnen geschuldet waren und nicht Fehlern oder Problemstellen in der RRMÜ an sich) geschlossen werden, dass die Fehlerkategorien relativ umfassend sind und den Großteil der möglichen Fehler und Problemstellen abdecken, aber dies beschränkt sich streng genommen nur auf die Übersetzungsrichtung Englisch-Deutsch (auch wenn die Fehlerkategorien, wie oben erwähnt, auch sprachübergreifend relativ umfassend erscheinen), die beiden hier verwendeten Ausgangstexte sowie den DeepL-Übersetzer. Außerdem wurden die Fehlerkategorien nur anhand eines vergleichsweise kurzen Textausschnitts entwickelt, d. h., selbst bei ihrer Anwendung auf andere Ausschnitte des ersten *Harry-Potter*-Bandes und aus *Red Rising* könnten noch Fehler und Problemstellen auftauchen, die durch die Fehlerkategorien nicht abgedeckt werden. Folglich ist eine generalisierte Anwendung dieser Fehlerkategorien nur schwer vertretbar. Ihre Anwendbarkeit könnte aber in weiterführenden Studien mit längeren Textausschnitten aus diesen Genres und auch anderen NMT-Systemen überprüft werden, dann evtl. ebenso in Bezug auf andere Genres und in anderen Sprachkombinationen.

Im Gegensatz zur obigen allgemeinen Diskussion der Fehlerkategorien werden im Folgenden einige der Kategorien einzeln noch genauer diskutiert⁸³, beginnend mit der Kategorie **Auflösung von Anaphern**: Obwohl sie in der L-NMT-Forschungsliteratur relativ

⁸³ Von Vergleichen der Anzahl der gefundenen Fehler/Problemstellen in der HPMÜ und der Anzahl der Korrekturen pro Fehlerkategorie an der RRMÜ wird weitgehend abgesehen, da diese Werte auf unterschiedliche Art und Weise erreicht wurden (die HPMÜ wurde von der Verfasserin der vorliegenden Arbeit genauestens analysiert (siehe Kapitel 4.2), während die RRMÜ erst von zwei Posteditorinnen posteditiert wurde, bevor die durchgeführten Korrekturen von der Verfasserin den jeweiligen Fehlerkategorien zugeteilt wurden, sofern dies möglich war (siehe Kapitel 4.5)). Folglich wäre ein Vergleich in den meisten Fällen nur wenig aussagekräftig, v. a. bei den Fehlerkategorien, die als Problembereiche einzuschätzen sind, da bei diesen zusätzlich noch der Faktor Subjektivität einen allfälligen Vergleich und seine Interpretation erschwert. Darüber hinaus können einige Diskrepanzen in der Anzahl der gefundenen Fehler, die bei einem Vergleich sichtbar werden, einfach und ohne Bezug auf Sprache per se erklärt werden, wodurch sie eher weniger relevant für das Thema der vorliegenden Masterarbeit und die L-NMT-Forschung sind (z. B. gibt es in den Posteditionen weniger Korrekturen von Anführungszeichenfehlern als gefundene Anführungszeichenfehler in der HPMÜ, da es im RRAT schlicht und einfach weniger direkte Reden und damit auch weniger Anführungszeichen gibt als im HPAT). Bei einer Fehlerkategorie wird dennoch ein vorsichtiger Vergleich gewagt, da ein sprachlicher Aspekt der Ausgangstexte der Grund für die Diskrepanz in der Anzahl der gefundenen Fehler in dieser Kategorie sein könnte (siehe unten).

oft als problematischer Bereich (mit einem Fokus auf Pronomen) genannt wird, stellte sie für den DeepL-Übersetzer bei den verwendeten Ausgangstexten eine vergleichsweise geringe Hürde dar. Tatsächlich finden sich in der HPMÜ Fehler dieser Kategorie lediglich bei solchen **Anaphern**, bei denen das Antezedens nicht im selben Satz und auch nicht im Satz davor, sondern noch weiter davor oder nur impliziert ist. Dies scheint darauf hinzudeuten, dass die in Khalilov (2020) aufgestellte Behauptung, der DeepL-Übersetzer wäre *context-aware*, zutrifft, auch wenn sich der Kontext auf den Satz vor dem gerade zu übersetzenden Satz beschränkt. Selbstverständlich wäre dies noch in weiterführenden Studien zu überprüfen, denn es könnten auch andere Änderungen am DeepL-Übersetzer vorgenommen worden sein, aber dass eine große Qualitätsverbesserung in Bezug auf die **Auflösung von Anaphern** stattgefunden hat, kann von der Verfasserin dieser Arbeit zumindest in der Übersetzungsrichtung Englisch-Deutsch bestätigt werden: In ähnlichen Analysen der Verfasserin im Rahmen ihrer Bachelorarbeit wurden in einer DeepL-Übersetzung aus dem Jahr 2019 desselben Ausgangstextes (HPAT) elf **Anaphern**-Fehler gefunden (vgl. Schuhmann 2020:13), im Gegensatz zu drei im hier analysierten Output (HPMÜ) aus dem Jahr 2021.

Die zweite Fehlerkategorie, für die eine ausführlichere Diskussion erforderlich ist, ist jene der **Bedeutung**. Wie Tabelle 7 zeigt, wird diese in zehn Werken zur L-NMT als Problembereich erwähnt, entweder generell oder im Hinblick auf die Unterkategorien fixe Wendungen und *phrasal verbs* sowie Polysemie oder Mehrdeutigkeit (dieser Begriff wird meist als Oberbegriff verstanden, der Polysemie und Homonymie umfasst). Interessanterweise wird die Silbentrennung nie erwähnt, deswegen kann vermutet werden, dass in den Studien zur L-NMT ausschließlich Texte verwendet wurden, die keine Silbentrennung enthielten. Trotzdem erscheint es sinnvoll, die Silbentrennung in Ausgangstexten als Quelle von (unnötigen) Fehlern im MT-Output deutlich hervorzuheben, denn somit können auch in weiteren Studien in diesem Bereich solche Fehler vermieden werden. Außerdem ist diese Fehlerkategorie ein erster Schritt in Richtung eines Preeditings von maschinell zu übersetzenden literarischen Texten, das für Verlage, die sich evtl. an eine maschinelle (Vor-)Übersetzung ihrer Werke wagen wollen, von Bedeutung sein kann.

Auch die Kategorie **Formatierung** ist auffällig, da sie, wie an Tabelle 7 ersichtlich wird, in den Werken zur L-NMT überhaupt nicht erwähnt wird und auch nur in der nicht auf die LMT zugeschnittenen DQF-MQM-Typologie als Fehlerkategorie enthalten ist. Ein Grund dafür könnte selbstverständlich sein, dass die in den L-NMT-Studien verwendeten NMT-Systeme keinerlei Probleme mit der korrekten **Formatierungsübertragung** hatten (auch der DeepL-Übersetzer machte schließlich in der HPMÜ nur einen Fehler diesbezüglich).

Andererseits könnte es aber auch sein, dass in vielen L-NMT-Studien solche Fehler gar nicht auftauchen können, da die verwendeten Ausgangstexte entweder keinerlei **Formatierung** enthalten oder grundsätzlich schon **formatierte** Textstellen beinhalten würden, diese aber nicht in Betracht gezogen werden können, da ein in einem reinen Textformat (z. B. „.txt“) gespeichertes Dokument übersetzt wird und kein wie hier verwendetes *Microsoft-Word*-Dokument o. Ä. Jedenfalls erweckt der Umstand, dass die **Formatierung** weder im Positiven noch im Negativen im Rahmen der L-NMT-Forschung genannt wird, den Eindruck, als sei sie in diesem Forschungsbereich und damit bei der Übersetzung von literarischen Werken von geringer Bedeutung, obwohl genau das Gegenteil der Fall ist: V. a. die Kursivsetzung hat, wie der HPAT und der RRAT zeigen, vielfältige Funktionen, darunter den Ausdruck einer Betonung in direkten Reden sowie die Signalisierung, dass ein Text in der jeweiligen Szene eines Romans aus einem Buch stammt oder die Quelle einer direkter Rede nicht eine der in der Szene anwesenden Personen ist, sondern z. B. eine Fernsehübertragung. Dementsprechend sollte in weiterführenden Studien zur L-NMT auch die Übertragung der **Formatierung** durch die MT-Systeme in Betracht gezogen werden, um herauszufinden, ob diese Systeme tatsächlich eher wenige bis gar keine Probleme damit haben oder, im anderen Fall, woran diese Probleme liegen und ob sie evtl. eher auf **formatierte** Einzelwörter als ganze Absätze eingegrenzt werden können, wie in Kapitel 5.2.1 vermutet wird.

Die Fehlerkategorie **Interpunktion** (und auch darin inkludiert die Problematik der Anführungszeichen) war in der Hypothese zu FF1 enthalten, d. h., für die Verfasserin der vorliegenden Arbeit war die große Zahl an Fehlern dieser Kategorie in beiden maschinellen Übersetzungen keine Überraschung. In der bisherigen L-NMT-Forschung jedoch scheint die **Interpunktion** nicht als großes Problem für NMT-Systeme aufgefasst worden zu sein, wird sie doch nur einmal in den Werken aus dem Forschungsüberblick als problematischer Bereich erwähnt. Dies könnte evtl. daran liegen, dass DeepL bisher nur selten in Studien in diesem Forschungsbereich zum Einsatz kam (siehe Einleitung und Kapitel 3). Ob die **Interpunktion** bzw., in einem engeren Rahmen, die Setzung von Anführungszeichen tatsächlich ein für den DeepL-Übersetzer spezifisches Problem ist, könnte in weiterführenden Studien z. B. durch einen Vergleich von mehreren NMT-Systemen überprüft werden. Außerdem könnte untersucht werden, ob die **Interpunktion** nur bei der im Rahmen dieser Arbeit verwendeten Dokumentenübersetzung problematisch ist oder auch bei der Übersetzung von direkt in das Textfeld auf der Webseite von DeepL kopierten Textausschnitten. Schließlich stellt sich auch die Frage, ob die **Interpunktion** nur in der Übersetzungsrichtung Englisch-Deutsch (die

ebenfalls bis jetzt wenig untersucht worden ist) ein Problem darstellt oder auch in der anderen Richtung oder überhaupt in anderen Sprachkombinationen.

Wie aufgrund der Schwierigkeiten von Humanübersetzer_innen bei der Übersetzung von **IUT** erwartet, stellte sich in den Analysen auch die maschinelle Übersetzung solcher **Termini** als ein eindeutiger Problembereich heraus. Es muss allerdings ebenfalls angemerkt werden, dass der DeepL-Übersetzer nach Einschätzung der Verfasserin der vorliegenden Arbeit bei manchen **Termini** auch durchaus brauchbare Übersetzungsvorschläge lieferte, auf eine „logical, but unquestioning“ Art und Weise, um es mit Constantines (2020:477) Worten auszudrücken. Folglich ist offenbar Potential für die maschinelle **IUT**-Übersetzung gegeben, weshalb auch in weiterführenden Studien mehr darauf eingegangen werden könnte, als das bisher der Fall war: Mehrheitlich wird nämlich die **Terminologie** von literarischen Werken, wie Tabelle 7 zeigt, gar nicht oder nur teilweise als Problembereich genannt, und sie ist auch in den oben angeführten Fehlertypologien nur ansatzweise enthalten, wie in Kapitel 5.1 genauer ausgeführt wird. Dies könnte auch damit zusammenhängen, dass in den meisten L-NMT-Studien nicht unbedingt auf Werke aus dem Science-Fiction- und Fantasy-Bereich eingegangen wird, der nach Erfahrung der Verfasserin dieser Arbeit das Vorkommen von **IUT** begünstigt bzw. teilweise sogar notwendig macht. Werden doch Werke aus diesen Genres als Ausgangstexte verwendet, wie in Toral et al. (2018), Moorkens et al. (2018), Toral & Way (2018), Toral, Oliver González et al. (2020), Omar & Gomaa (2020), Guerberof-Arenas & Toral (2021) und Shuyin (2021), wird nur selten genauer auf die MT-Outputs eingegangen, nämlich lediglich in Moorkens et al. (2018) durch Aussagen von Posteditor_innen und in Omar & Gomaa (2020) durch makrotextuelle Analysen. Tatsächlich sind es auch diese beiden Werke, in denen zwei der drei teilweisen Erwähnungen von **IUT** erfolgen, die dritte kommt in Mohar et al. (2020) vor (siehe Kapitel 5.1 für mehr Details diesbezüglich). Jedenfalls könnte in weiterführenden Studien auch ein reiner Fokus auf die **IUT**-Übersetzung gelegt und z. B. untersucht werden, wie ein mit literarischen Werken (aus dem Science-Fiction- und Fantasy-Bereich) trainiertes NMT-System mit dieser Art von **Terminologie** umgeht. Außerdem könnten evtl. einem solchen NMT-System bzgl. der **IUT**-Übersetzung ähnliche Beschränkungen auferlegt werden, wie sie Van De Cruys (2021) zur MT-Übersetzung von Gedichten beschreibt, um dem System eine gewisse „creative inclination“ einzugeben und sie zu einer für Literaturübersetzer_innen „attractive source of inspiration“ (Van De Cruys 2021) zu machen.

Die Kategorie **Tempus und Modus von Verben**, die in der Forschungsliteratur relativ oft erwähnt wird, aber in keiner der drei Fehlertypologien tatsächlich in dieser Form

vorkommt, ist jene, die in Fußnote 83 erwähnt wird und bei der trotz der in dieser Fußnote genannten Gründe dagegen die Anzahl der gefundenen Fehler in der HPMÜ jener der Korrekturen dieser Kategorie in den Posteditionen gegenüber gestellt wird: In der HPMÜ wurden sieben solcher Fehler gefunden, während in beiden Posteditionen nur eine (und zwar dieselbe) Korrektur dieser Kategorie zugeordnet werden konnte. Dies könnte an den Ausgangstexten oder, genauer gesagt, an dem **Tempus** liegen, in dem die Geschichte in den beiden Romanen erzählt wird: In *Harry Potter* wird das *past simple* verwendet, in *Red Rising* die *present tense*. Es liegt daher die Vermutung nahe, dass das *past simple* dem DeepL-Übersetzer in literarischen Texten dieser Art mehr Schwierigkeiten bereitet als die *present tense*. Selbstverständlich könnte aber die Diskrepanz in der Anzahl der gefundenen Fehler dieser Art auch dem Zufall oder anderen Faktoren geschuldet sein (z. B. könnten die Posteditorinnen auch Fehler dieser Art übersehen haben), weswegen diese Vermutung in weiterführenden Studien mit längeren Textausschnitten und einer einheitlichen Auswertungsmethode überprüft werden müsste.

Abgesehen von diesen Auffälligkeiten bzgl. einiger einzelner Fehlerkategorien ist noch anzumerken, dass bei drei Fehlerkategorien (IUT, Stil und Tonfall und Register) die **Subjektivität** bei der Entscheidung, was nun ein Fehler ist und was nicht, was nur problematisch oder doch korrekt ist, in beiden Analysen sehr deutlich geworden ist. Allerdings ist von dieser großen Rolle des Subjektivitätsfaktors fast keine Rede bei der Präsentation von Fehleranalyseergebnissen in der LMT-Forschungsliteratur oder bei der Erklärung von Fehlertypologien, auch nicht im Hinblick auf Kategorien wie *Awkward* und *Unidiomatic* (DQF-MQM) oder *disfluency* (angepasste SCATE-Typologie). Lediglich in Lommel, Görög et al. (2015:18-22), wo die Harmonisierung der vormals eigenständigen DQF- und MQM-Typologien beschrieben wird, und in der Beschreibung der ursprünglichen, d. h. nicht an L-NMT angepassten SCATE-Fehlertypologie (vgl. Tezcan et al. 2018:235ff.) gibt es Analysen zur Übereinstimmung zwischen Evaluat_innen bei der Benutzung dieser Typologien. Auch die Unterscheidung zwischen „Fehler“ und „Problemstelle“ (siehe Kapitel 4.2), die im Rahmen dieser Arbeit getroffen wird und aufgrund des Subjektivitätsfaktors als so wichtig erscheint, gibt es in vergleichbarer Art und Weise nur in der MQM-Typologie (vgl. Lommel, Görög et al. 2015:7). In zukünftigen ähnlichen Studien zur Evaluierung der Qualität von L-NMT-Outputs und Posteditionen solcher Outputs könnte der Subjektivität, die wohl v. a. bei der Bewertung von Literatur und damit auch von Literaturübersetzungen zum Tragen kommt, durch die Anwendung der Unterscheidung

zwischen „Fehler“ und „Problemstelle“ und damit einhergehend zwischen „Fehlerkategorie“ und „Problembereich“ Rechnung getragen werden.

Damit wird dieser erste Teil der Diskussion beendet. Im zweiten Teil geht es um die Forschungsfragen 2 und 3, um ihre jeweiligen Hypothesen und, wie im ersten Teil, um weiterführende Studien, die sich aus den Antworten zu diesen Fragen und ihrer Verbindung zur Forschungslektüre ergeben könnten.

Tabelle 8: FF2 und FF3 und Hypothesen

FF2	Ist eine von DeepL erstellte Übersetzung eines Ausschnitts aus <i>Red Rising</i> , die unter Zuhilfenahme der entwickelten Fehlerkategorien posteditiert wird, von höherer Qualität als eine ohne diese Fehlerkategorien posteditierte DeepL-Übersetzung desselben Ausschnitts?
Hyp.	<ul style="list-style-type: none"> – ja, RRPEm insgesamt von höherer Qualität als RRPEo – RRPEm öfter mit Humanübersetzung gleichgesetzt als RRPEo – RRPEm auch bzgl. Interpunktion, IUT und Bedeutung von höherer Qualität
FF3	In welchen Aspekten unterscheiden sich die beiden Posteditionen und welche Fehlerkategorien spielen hierbei eine besondere Rolle?
Hyp.	<ul style="list-style-type: none"> – Interpunktion – IUT – Bedeutung

Vor der konkreten Beantwortung dieser Forschungsfragen soll noch kurz erklärt werden, warum sie **zusammen** präsentiert werden. Abgesehen davon, dass sie sich beide (im Gegensatz zu FF1) auf die Posteditionen beziehen, liegt der Grund darin, dass sie und ihre Hypothesen miteinander zusammenhängen und sich bei ihrer Entwicklung und Aufstellung eine aus der anderen ergab: Zunächst wurden nämlich FF2 und die ersten zwei Hypothesen dazu aufgestellt. Aus der angenommenen besseren Bewertung der RRPEm gegenüber der RRPEo ergab sich die Frage, in Bezug auf welche Fehlerkategorien sich die Posteditionen denn überhaupt unterscheiden würden, also FF3. Dazu wurden die drei in der Tabelle genannten Fehlerkategorien als hypothetische Antwort angenommen. Als logische Folgerung ergab sich, dass, wenn sich die Hypothese zu FF3 bewahrheitet, sich auch die Qualitätsbewertungen in diesen drei Kategorien für die zwei Posteditionen unterscheiden müssten. Da weiters, wie oben erwähnt, allgemein angenommen wurde, dass sich die Benutzung der Fehlerkategorien positiv auf die Qualität der Postedition auswirken würde (= erste Hypothese zu FF2), wurde dementsprechend auch angenommen, dass dies bei den drei Kategorien ebenfalls der Fall sein müsste, woraus sich die dritte Hypothese zu FF2 ergab.

Da dies nun geklärt ist, können auch die **Antworten** auf diese Forschungsfragen angeführt werden: **FF2** lässt sich zunächst gesamt gesehen eindeutig mit „nein“ beantworten, denn in einer Zusammenfassung der Rating-Ergebnisse ergab sich der erste Platz für die RRPEo, die ohne Fehlerkategorien erstellt worden war, und der letzte für die RRPEm, bei deren Anfertigung die Fehlerkategorien zum Einsatz kamen. Auch der direkte Vergleich der beiden Posteditionen (siehe Abb. 5 in Kapitel 5.2.2) lässt diesen Schluss zu. Damit hat sich die erste Hypothese zu FF1, nach der die RRPEm eine höhere Qualität aufweisen müsste, nicht bewahrheitet.

Aus den paarweisen Rankings der Posteditionen mit der RRHÜ geht außerdem hervor, dass sich auch die **zweite Hypothese** nicht bewahrheitet hat: Die RRPEm wurde insgesamt nur 18-mal mit der RRHÜ gleichgesetzt, die RRPEo dagegen 26-mal.⁸⁴ Zu dieser Hypothese ist anzumerken, dass bei ihrer Aufstellung implizit angenommen wurde, dass die RRHÜ als reine und publizierte Humanübersetzung in den Ratings insgesamt gesehen alleine auf Platz 1 landen würde, sodass eine Gleichstellung mit ihr eine sehr hohe Qualität bedeutet hätte. Da dies allerdings nicht der Fall war (tatsächlich konnte eine Gesamt-Rangfolge aufgestellt werden, bei der die RRHÜ hinter der RRPEo auf Platz 2 zu finden ist; siehe Kapitel 5.2.2), kann aus dem Umstand, dass sich diese zweite Hypothese als falsch herausstellte, eher geschlossen werden, dass sich die RRPEo und die RRHÜ mehr in ihrer Qualität ähneln als die RRPEm und die RRHÜ.

Schließlich wurde in der **dritten Hypothese zu FF2** angenommen, die RRPEm würde in Bezug auf drei Fehlerkategorien besser eingestuft als die RRPEo. Bei der erstgenannten dieser drei Kategorien, der Interpunktion, kann diese Annahme weder bestätigt noch dementiert werden, da offenbar einige Rater_innen und evtl. auch Posteditorin 2 diese Kategorie nicht richtig verstanden hatten und somit von einer weiteren Auswertung abgesehen wurde (siehe Kapitel 5.2.2). Bei der IUT als zweiter der genannten Fehlerkategorien ist dies anders: Tatsächlich wurde diesbezüglich die RRPEm insgesamt besser bewertet als die RRPEo; dieser Teil der Hypothese hat sich dementsprechend bewahrheitet. In Bezug auf den dritten und letzten Teil der Hypothese kann dies nicht behauptet werden, denn bei der Bedeutung schnitt die RRPEm, wie schon über alle Ratings hinweg gesehen, wieder schlechter als die RRPEo und auch schlechter als die RRHÜ ab.

⁸⁴ Dabei wurden aus den in Kapitel 5.2.2 erläuterten Gründen (falsches Verständnis der Kategorie Interpunktion bei manchen Rater_innen und evtl. auch bei Posteditorin 2) die Ergebnisse der paarweisen Rankings zur Interpunktionskategorie nicht mitgezählt.

Als **Antwort auf FF3** sind die fünf Fehlerkategorien Bedeutung, Interpunktion, IUT, Stil sowie Tonfall und Register, aber v a. auch die zusätzlichen, nicht auf Fehlern oder Problemstellen beruhenden Korrekturen (PV-Korrekturen) der Posteditorinnen zu nennen. Schon an dieser kurzen Auflistung der Aspekte, in denen sich die Posteditionen unterscheiden, zeigt sich, dass sich die genannte Hypothese in allen Teilen bewahrheitet hat, aber die angegebene Antwort bedarf wohl trotzdem noch einer genaueren Erörterung, da es sich schließlich bei FF3 auch nicht um eine einfache Ja/Nein-Frage handelt: Bzgl. der **Bedeutung** unterscheiden sich die beiden Posteditionen dahingehend, dass Posteditorin 1 (also die Posteditorin, die ohne die Fehlerkategorien arbeitete) zusätzlich zu Korrekturen von Einzelwörtern und fixen Einheiten (wie Redewendungen) auch **bedeutungstechnisch** falsch übersetzte, nicht fixierte Phrasen korrigierte, was bei Posteditorin 2 (also der Posteditorin, die mit den Fehlerkategorien arbeitete) nicht der Fall war. Dementsprechend kann angenommen werden, dass diese Kategorie, bei deren Beschreibung keine solchen Wortsequenzen als Beispiele für Fehler genannt wurden, einen (für die Qualität negativen) *priming effect* auf die zweite Posteditorin ausübte. Ein *priming effect* war ebenfalls bei der Kategorie **Interpunktion** auszumachen: Posteditorin 2 korrigierte alle in den Beispielen zu dieser Kategorie explizit genannten Fehler, während Posteditorin 1 dies nicht durchgehend tat. Darüber hinaus setzte Posteditorin 2 auch fehlende Beistriche nach direkten Reden, aber ansonsten übernahm sie, im Gegensatz zu ihrer Kollegin, weitgehend die **Interpunktion** der RRMÜ, die wiederum in großen Teilen vom RRAT übernommen worden war. Dies zeugt zusätzlich zum *priming effect* von einem evtl. Missverständnis dieser Kategorie, wie es auch bei einigen Rater_innen der Fall war (siehe Kapitel 5.2.2). In Bezug auf die Übersetzung der **IUT** unterscheiden sich die Posteditionen deutlich, insofern als Posteditorin 1 die Tendenz hatte, die von DeepL aus dem Englischen ins Deutsche übernommenen **Termini** zu belassen, während sich Posteditorin 2 für diese **Termini** deutsche Entsprechungen überlegte, was von einem (für die Qualität positiven) *priming effect* dieser Fehlerkategorie auf Posteditorin 2 zeugt. Bei der Kategorie **Stil** war kein solcher Effekt beobachtbar, die Posteditionen unterscheiden sich aber trotzdem in mehrerlei Hinsicht: Unübliche Wortwahl oder wörtlich übersetzte Konstruktionen wurden grundsätzlich von beiden korrigiert, aber Posteditorin 1 achtete darüber hinaus noch auf eine einheitliche Nutzung von Tempora und korrigierte Wortwiederholungen, sodass sie insgesamt mehr **Stil**korrekturen durchführte als ihre Kollegin. Als möglicher Grund dafür wurden geringere kognitive Kapazitäten bei Posteditorin 2 für die zielsprachlichen Versionen vermutet, da sie sich zusätzlich zu den drei Texten auch noch mit den Fehlerkategorien beschäftigen musste, sowie generell eine größere Konzentration auf den RRAT als bei

Posteditorin 1. Bei der Kategorie **Tonfall und Register** war ebenfalls ein (für die Qualität eher negativer) *priming effect* auszumachen, insofern als Posteditorin 2, abgesehen von anderen Strategien, die beide einsetzten, die in den Beispielen zu dieser Fehlerkategorie genannte Strategie der Verbverkürzung in Dialogen verwendete, Posteditorin 1 aber nicht. Zu diesen Ausführungen ist zusätzlich zu sagen, dass sich selbstverständlich nicht nur die Verwendung oder Nicht-Verwendung der Fehlerkategorien auf die Posteditionen ausgewirkt hat, sondern auch und v. a. der persönliche Posteditingstil der Posteditorinnen, d. h. ihre subjektiven Entscheidungen während des Posteditings. Dies zeigt sich selbstverständlich an allen Korrekturen, die unterschiedlich oder gar nur von einer Posteditorin durchgeführt wurden, aber noch deutlicher an den **PV-Korrekturen**: Von diesen nicht unbedingt notwendigen Korrekturen führte Posteditorin 1 viel mehr durch als Posteditorin 2, sodass die RRPEo insgesamt freier ist und sich sowohl vom RRAT als auch von der RRMÜ deutlicher unterscheidet als die RRPEm. Aufgrund dieser Ähnlichkeit der RRPEm mit der RRMÜ ist darüber hinaus nicht auszuschließen, dass auch die maschinelle Übersetzung (und nicht nur manche Fehlerkategorien) einen (für die Qualität negativen) *priming effect* auf Posteditorin 2 ausübte.

Zu all diesen Antworten ist zu sagen, dass sie aufgrund der geringen Zahl an Rater_innen, mangelnder Konzentration, falschen Verständnisses zumindest einer Fehlerkategorie (evtl. zwei) bei manchen Rater_innen und v. a. der großen Rolle der Subjektivität, die sich sowohl beim Postediting (siehe Antwort zu FF3) an sich als auch bei den Ratings deutlich gezeigt hat, nur **vorläufig** und keineswegs als absolut oder bestätigt anzusehen sind. Außerdem beschränken sie sich nur auf einen sehr eingegrenzten Bereich der L-NMT (DeepL, Übersetzungsrichtung Englisch-Deutsch, Fantasy- und Science-Fiction-Literatur für Kinder und Jugendliche) und können daher nicht auf andere Bereiche übertragen werden. Darüber hinaus ist nach bestem Wissen der Verfasserin dieser Masterarbeit der Einfluss von speziell für die L-NMT entwickelten Fehlerkategorien auf die Postedition eines solchen L-NMT-Outputs und ihre Qualität noch nie zuvor untersucht worden. Da dies einen Vergleich der Ergebnisse mit der existierenden Forschungsliteratur schwierig macht, wird im Folgenden eher auf Möglichkeiten für weiterführende Studien eingegangen (die aufgrund der Neuartigkeit der vorliegenden Untersuchung sehr wichtig sind), und die Ergebnisse werden nur, wo das möglich und sinnvoll ist, in Verbindung mit der Forschungsliteratur gesetzt.

So eine Verbindung ist etwa bei der Antwort auf FF2 möglich, da diese Forschungsfrage auf **zwei Befunden aus der Forschungsliteratur** beruht. Diese zwei Befunde von Way (2020) und Guerberof-Arenas (2020) wurden schon in der Einleitung angeführt: In Way

(2020:320) wird behauptet, Fehler in NMT-Outputs seien schwer zu erkennen, während Guerberof-Arenas (2020:337) angibt, Fehlertypologien würden Posteditor_innen ihre Arbeit erleichtern, weil sie dadurch über häufig vorkommende Fehler in MT-Outputs informiert seien und diese dann leichter finden würden. In beiden Behauptungen haben die Forscher_innen wohl nicht die L-NMT und Postediting von L-NMT-Output im Sinn (es wird in ihren Artikeln nie „literary (machine) translation“ o. Ä. erwähnt), trotzdem ergaben sich aus diesen Behauptungen FF2 und die Hypothesen dazu. Letztere konnten allerdings, wie oben klar dargestellt, nicht (bzw. bei der dritten Hypothese nicht ganz) bestätigt werden, was Zweifel an der Anwendbarkeit der Behauptungen der beiden Forscher_innen auf den oben spezifizierten Bereich der L-NMT aufkommen lässt.

In **weiterführenden Studien** könnte demnach untersucht werden, ob diese Zweifel berechtigt sind, zunächst nur in Bezug auf das hier verwendete NMT-System, die Übersetzungsrichtung und die Art von literarischem Text, aber später auch im Hinblick auf einen weiter gefassten Bereich, der andere NMT-Systeme, andere Sprachkombinationen und literarische Texte aus anderen Genres einschließt. Zu diesem Zweck könnten neben Studierenden auch professionelle Literaturübersetzer_innen sowohl als Posteditor_innen als auch als Rater_innen dienen. Als Rater_innen könnten auch Endnutzer_innen, also Leser_innen von literarischen Übersetzungen dienen, die keinen übersetzungswissenschaftlichen oder -praktischen Hintergrund haben, um herauszufinden, ob, ähnlich wie in Shuyin (2021), bei solchen Leser_innen eine gewisse Toleranz für Fehler in Posteditionen von L-NMT-Output besteht. Darüber hinaus könnten auch zusätzliche Daten (nicht nur die Posteditionen an sich und die Qualitätsbewertungen) bei weiterführenden Studien erhoben werden, wie etwa Daten zum kognitiven, zeitlichen und technischen Aufwand beim Postediting mit und ohne Fehlerkategorien. V. a. Messungen zum kognitiven Aufwand wären in dieser Hinsicht notwendig, um die oben und in Kapitel 5.2.1 angestellte Vermutung, Posteditorin 2 hätte durch ihre Verwendung der Fehlerkategorien weniger kognitive Kapazitäten für die zielsprachlichen Versionen und damit einen insgesamt höheren kognitiven Aufwand gehabt, zu bestätigen oder zu widerlegen. Mit und ohne Fehlerkategorien erstellte Posteditionen könnten außerdem auch auf die in ihnen an den Tag gelegte Kreativität untersucht werden, wie das in Guerberof-Arenas & Toral (2020) gemacht wurde. In dieser Hinsicht erscheint auf den ersten Blick v. a. die Kategorie IUT vielversprechend, deren *priming effect* auf Posteditorin 2 diese wahrscheinlich zur Entwicklung von deutschen Entsprechungen für die englischen IUT anregte, bei der wohl viel Kreativität gefordert war.

Selbstverständlich könnten bei all diesen Möglichkeiten für weiterführende Studien auch andere Fehlerkategorien als die hier entwickelten zum Einsatz kommen, entweder solche, die genau wie hier ausgehend von einem spezifischen Text entwickelt werden, oder solche, die schon „vorgefertigt“ sind, wie jene, die in Kapitel 4.2 besprochen wurden. Jedenfalls sind viele weiterführende Studien nötig, um ein umfassenderes Bild zum Einfluss der Benutzung von Fehlerkategorien während des Posteditings von L-NMT-Output auf die resultierende Postedition und ihre Qualität zu erstellen und die logische Folgerung aus der Kombination der Behauptungen von Way (2020:320) und Guerbero-Arenas (2020:337) (also die erste Hypothese zu FF2) im Bereich der L-NMT tatsächlich zu dementieren (oder auch, entgegen der Ergebnisse der vorliegenden Studie, zu bestätigen).

Obwohl sich also die Benutzung von Fehlerkategorien in dieser Studie insgesamt gesehen nicht positiv auf die Qualität einer Postedition von L-NMT-Output auswirkte, zeigen die bessere Gesamtbewertung einer Postedition, der RRPEo, gegenüber der RRHÜ und auch die einzelnen Rating-Ergebnisse doch das grundsätzliche **Potential von Posteditionen von literarischem DeepL-Output** (und das trotz der Syntaxübernahmen durch das NMT-System von DeepL, die etwa auch von Webster et al. (2020a:15) erhoben wurden⁸⁵). Dies scheint die Meinung des in Kenny & Winters (2020a) teilnehmenden Humanübersetzers zu bestätigen, der behauptete, dieses MT-System sei für Literaturübersetzer_innen aufgrund der guten Qualität der Outputs interessant (vgl. Kenny & Winters 2020a:133). Trotzdem müsste dieses augenscheinliche Potential von literarischen DeepL-Übersetzungen, die anschließend posteditiert werden, noch in vielen weiterführenden Studien überprüft werden, denn es nahm, wie oben angemerkt, nur eine geringe Zahl an Rater_innen an der Studie teil und zwischen ihnen herrschte auch nur wenig Übereinstimmung (siehe Kapitel 5.2.2). In dieser Hinsicht wäre es in weiterführenden Studien v. a. interessant zu untersuchen, wie gut DeepL-Posteditionen im literarischen Bereich verglichen mit Posteditionen von Outputs abschneiden, die von mit Literatur trainierten NMT-Systemen erstellt wurden, da letzteren schließlich eine höhere Qualität zugeschrieben wird (vgl. Toral, Oliver González et al. 2020:288).

Außerdem könnte noch mehr ins Detail gegangen und untersucht werden, in welchen Bereichen oder Fehlerkategorien die Kombination L-NMT+Postediting schon viel **Potential** aufweist und in welchen noch Verbesserungsbedarf besteht. Die hier erhobenen Ergebnisse deuten auf großes Potential dieser Kombination in den Bereichen Auslassungen und Hinzufügungen, Formatierung, Grammatik und Orthographie sowie Tempus und Modus von

⁸⁵ In Tezcan et al. (2019:47f.) ist ebenfalls von Syntaxübernahmen die Rede, aber hier wurde nicht DeepL, sondern Google Translate verwendet (vgl. Tezcan et al. 2019:41).

Verben hin – in den Ratings zu diesen Fehlerkategorien landeten insgesamt beide Posteditionen vor der RRHÜ bzw. gemeinsam mit der RRHÜ auf dem ersten Platz. Potential zeigte sich auch in den Kategorien Bedeutung sowie Tonfall und Register, bei denen die RRPEo auf Platz 1 vor der RRHÜ landete, aber etwas weniger als bei den zuvor genannten Kategorien, was bedeutet, dass hier die Posteditor_innen in der Kombination viel ausmachen. Am wenigsten Potential zeigte L-NMT+Postediting laut den Rater_innen in den Bereichen Stil (vermutlich aufgrund der Syntaxübernahmen von DeepL, von denen selbstverständlich auch Posteditorin 1 nicht alle veränderte) und IUT (wahrscheinlich, weil hier, wie oben angeführt, die Kreativität eine große Rolle spielt), in denen beide Posteditionen hinter der RRHÜ landeten. Diese nur vorläufigen Feststellungen zum Potential von L-NMT+Postediting könnten in weiterführenden Studien überprüft werden, wobei darauf geachtet werden müsste, dass sämtliche Fehlerkategorien (v. a. aber die Interpunktion und in geringerem Ausmaß auch die Auflösung von Anaphern; siehe Kapitel 5.2.2) von allen Teilnehmenden richtig verstanden werden, um über alle Kategorien Schlüsse zum Potential ziehen zu können.

Im Gegensatz zur Antwort auf FF2, die gerade diskutiert wurde und die die Aussagen von Way (2020:320) und Guerberof-Arenas (2020:337) nicht unbedingt bestätigen konnte, scheint die Antwort auf FF3 eine andere Aussage von Guerberof-Arenas (2020:346) durchaus zu bestätigen, nämlich jene zur **Variabilität der Posteditor_innen** (im Rahmen dieser Arbeit eher unter „Subjektivität“ oder „Posteditingstil“ behandelt⁸⁶): Obwohl bei der Auswahl der Posteditorinnen darauf geachtet wurde, dass sich diese so ähnlich wie möglich sind, sowohl, was ihren Studiumsfortschritt zum Zeitpunkt des Posteditings betrifft, als auch in Bezug auf ihre Erfahrung im Bereich Literaturübersetzen, Postediting und generell in der Sprachindustrie (siehe Kapitel 4.3), stellten sich die von ihnen durchgeführten Korrekturen und damit auch ihre Posteditionen insgesamt gesehen als sehr verschieden heraus, wie obige Antwort zu FF3 und auch die Bewertungen der Rater_innen zeigen. Die Fehlerkategorien haben zwar höchstwahrscheinlich v. a. durch ihre möglichen *priming effects* auf Posteditorin 2 zu den Unterschieden zwischen den Posteditionen beigetragen, aber trotzdem liegen diese Unterschiede wohl zu einem Großteil an den Posteditorinnen selbst und ihren subjektiven Entscheidungen während des Posteditings, d. h. ihrem Posteditingstil (schließlich machen die PV-Korrekturen in beiden Posteditionen den jeweils größten Anteil an Korrekturen aus (siehe Tabelle 4 in Kapitel 5.2.1), und auch viele der Korrekturen, die

⁸⁶ Auch das Konzept der „Stimme“ von Übersetzer_innen/Posteditor_innen (vgl. Taivalkoski-Shilov 2019a und v. a. Kenny & Winters 2020a und 2020b) fällt in diesen Bereich.

Fehlerkategorien zugeteilt werden konnten, fielen bei den Posteditor_innen unterschiedlich aus).

Da also in der vorliegenden Studie der Einfluss der Fehlerkategorien auf die Posteditionen vom Einfluss des Posteditingstils der Posteditorinnen nur schwer getrennt werden konnte, sind weiterführende Studien, in denen dies besser gelingt, wünschenswert. In diesem Sinne könnte wohl doch ein **Aggregieren von Übersetzungen** angedacht werden, wie es in Guerberof-Arenas & Toral (2020:260) angewandt wurde, aber in modifizierter Form, um eine mögliche Verwirrung bei den Posteditor_innen (siehe Kapitel 4.3) gering zu halten: Es könnten statt nur eines Ausschnitts zwei Ausschnitte aus einem literarischen Werk als Ausgangstexte gewählt werden. Diese Ausschnitte sollten sich selbstverständlich möglichst ähnlich sein – um diese Ähnlichkeit festzustellen, könnten die in Kapitel 4.1 beschriebenen, zur Auswahl der hier verwendeten Ausgangstexte HPAT und RRAT verwendeten Kriterien und zusätzliche (wie etwa die Anzahl an Dialogen o. Ä.) benutzt werden. Beide Ausschnitte (AT1 und AT2) könnten dann maschinell übersetzt (MT1 und MT2) und anschließend in zwei Hälften geteilt werden, sodass insgesamt vier Ausgangstexte (AT1a, AT1b und AT2a, AT2b) und vier MT-Outputs (MT1a, MT1b und MT2a, MT2b) vorliegen. MT1a und MT2a könnten von Posteditor_in A posteditiert werden, zunächst MT1a ohne Fehlerkategorien, anschließend MT2a mit, während sich Posteditor_in B MT1b und MT2b in gleicher Art und Weise widmet. Daraus würden vier Posteditionen entstehen, PE1a, PE2a, PE1b und PE2b, welche zu zwei Gesamtübersetzungen aggregiert werden könnten (also PE1a+PE1b und PE2a+PE2b), die dann im Anschluss einigen Rater_innen zur Qualitätsbewertung präsentiert werden. Beide Posteditionen wären somit von zwei Posteditor_innen erstellt worden (was die in den Posteditionen sichtbare Variabilität zwischen den Posteditor_innen bzw. die Auswirkungen ihrer subjektiven Postediting-Entscheidungen zumindest etwas ausgleichen sollte), aber trotzdem würde eine Postedition insgesamt ohne und eine insgesamt mit Fehlerkategorien erstellt worden sein, was eine Untersuchung des Einflusses der Fehlerkategorien auf die Qualität der resultierenden Posteditionen ermöglicht. Dadurch könnte evtl. dieser Einfluss genauer herausgefiltert werden, als das in der vorliegenden Studie der Fall war, ohne für große Verwirrung bei den Posteditor_innen zu sorgen.

Zur Variabilität der Posteditor_innen gehört auch ihre **Erfahrung**, sowohl, was das Postediting und das Literaturübersetzen selbst betrifft, als auch im Hinblick auf andere Bereiche der Sprachindustrie. Wie in Kapitel 2 dieser Arbeit erklärt, wird die Erfahrung im Bereich des Posteditings von maschinell übersetzten Fachtexten mehrheitlich als wichtig beim Postediting angesehen, allerdings gibt es zum Einfluss der Erfahrung (auf Qualität und

Produktivität) unterschiedliche Ergebnisse. Moorkens & O'Brien (2015:78f.) schlossen z. B. in ihrer Untersuchung, dass erfahrenere Übersetzer_innen auch qualitativ hochwertigere Posteditionen anfertigen als unerfahrenere. Die hier vorgestellten Ergebnisse lassen auf Ähnliches beim Postediting von maschinell übersetzten literarischen Texten schließen: Posteditorin 1 hatte im Gegensatz zu Posteditorin 2 (die andere Erfahrungswerte in der Sprachindustrie aufwies; siehe Kapitel 4.3) schon eine Fortbildung zum Thema des Literaturübersetzens absolviert, aber v. a. hatte sie Erfahrung als Editorin von Videospielübersetzungen. Es gibt Hinweise darauf, dass diese Tätigkeit ihre Art des Posteditings stark beeinflusst hat (sie posteditierte weitaus ungezwungener als ihre Kollegin; siehe obige Antwort zu FF3 und Kapitel 5.2.1), und aus den Rating-Ergebnissen, d. h. der insgesamt besseren Bewertung der RRPEo von Posteditorin 1, kann geschlussfolgert werden, dass sich dieser Einfluss positiv auf die Qualität ihrer Postedition auswirkte. Dies müsste aber selbstverständlich in weiterführenden Studien zum Einfluss der Erfahrung beim L-NMT-Postediting noch mehr untersucht werden. So könnten beispielsweise die hier erstellten Posteditionen im Hinblick auf ihre Qualität auch mit zwei Posteditionen von erfahrenen Literaturübersetzer_innen verglichen werden. Dabei müsste allerdings derselbe MT-Output wie in der vorliegenden Masterarbeit benutzt werden (siehe Anhang 1), denn an DeepL werden ständig Verbesserungen durchgeführt (siehe dazu u. a. den ersten Teil dieser Diskussion), sodass ein Output von einer späteren Version des NMT-Systems den erfahrenen Literaturübersetzer_innen einen Vorteil geben könnte und einen Vergleich der resultierenden Posteditionen erschweren würde.

Im Hinblick auf die Erfahrung wäre es außerdem interessant zu untersuchen, ob während des Posteditings zu benutzende Fehlerkategorien ebenso starke und mehrheitlich für die Qualität der Postedition negative *priming effects* auf **erfahrene Literaturübersetzer_innen** wie auf Posteditorin 2 haben (können) oder ob der Übersetzungs- und damit vermutlich auch der Posteditingstil ersterer schon so sehr gefestigt ist, dass er sich nicht oder nur wenig durch die Fehlerkategorien verändert. Mögliche Hinweise dazu geben die Ergebnisse von de Almeida & O'Brien (2010:[4f.]), in deren Studie zum Postediting eines maschinell übersetzten Fachtextes die erfahrenen Übersetzer_innen die Postediting-Guidelines eher missachteten als die unerfahrenen und auch mehr Änderungen aufgrund von persönlichen Vorlieben durchführten als letztere (genauso wie Posteditorin 1 im Vergleich zu Posteditorin 2 in dieser Studie). Deshalb wird vermutet, dass sich erfahrene Literaturübersetzer_innen nur wenig bis gar nicht von den Fehlerkategorien beeinflussen lassen würden. Sollten sich doch auch bei erfahrenen Literaturübersetzer_innen *priming effects*

beobachten lassen, wäre es interessant zu sehen, bei welchen Fehlerkategorien dies der Fall ist (bei ähnlichen wie bei Posteditorin 2 oder ganz anderen?) und auch, wie sie den persönlichen Stil der Übersetzer_innen verändern. Diesbezüglich erscheint wiederum speziell die Kategorie IUT interessant, da sie als einzige einen *priming effect* auf Posteditorin 2 ausübte, der sich positiv auf die Qualität ihrer Postedition auswirkte.

Ein *priming effect* wurde aber nicht nur durch die Fehlerkategorien, sondern wohl auch **durch den L-NMT-Output** an sich ausgeübt. Da dieser Effekt den Rater_innen allerdings nur bei Posteditorin 2 in verstärktem Ausmaß auffiel, ist auch hier anzunehmen, dass bei Posteditor_innen mit mehr Erfahrung dieser *priming effect* schwächer ausfällt. Diese Annahme müsste aber ebenfalls noch in weiterführenden Studien überprüft werden, denn sie kann mithilfe der Ergebnisse aus anderen Forschungsarbeiten, in denen ein solcher *priming effect* durch L-NMT-Outputs erhoben wurde (etwa Şahin & Gürses (2019:29), Toral, Castilho et al. (2020:1°44') und Kolb (2021)⁸⁷), nicht bestätigt werden. In diesen Studien nahmen nämlich stets nur entweder Studierende (also wenig erfahrene) oder professionelle (also erfahrene) Übersetzer_innen (keine gemischten Gruppen) teil. Das Ausmaß des beobachteten *priming effect* wurde dementsprechend nicht zwischen Teilnehmenden mit unterschiedlicher Erfahrung verglichen.

Bislang wurde nur die Erfahrung bzw., allgemeiner gesagt, die Variabilität der Posteditor_innen in dieser und in weiterführenden Studien diskutiert, aber schließlich gibt es bei den **Rater_innen** genauso viel Variabilität bzw. Subjektivität (siehe die geringe Interrater_innen-Reliabilität), sodass auch diesbezüglich eine kurze Diskussion angebracht erscheint. In dieser Hinsicht wäre es in weiterführenden Studien interessant, die hier vorgestellten Ratings mit erfahrenen Literaturübersetzer_innen als Rater_innen zu wiederholen und die Ergebnisse dann zu vergleichen, um wiederum den Einfluss der Erfahrung in der Qualitätsbewertung herauszufiltern. Diesbezüglich könnte auch untersucht werden, ob den erfahrenen Literaturübersetzer_innen die gleichen Kategorien wichtig sind wie den Studierenden in diesen Ratings, zu denen angenommen werden kann, dass ihnen v. a. Stil und IUT, aber auch Bedeutung sowie Tonfall und Register bedeutsam erschienen. Darüber hinaus könnte auch mit den im Rahmen dieser Studie zwar gesammelten, aber nicht ausgewerteten Daten (siehe Kapitel 5.2.2) weitergearbeitet und untersucht werden, inwiefern sich die schon ausgewerteten Ergebnisse von jenen unterscheiden, die von Teilnehmenden erstellt wurden,

⁸⁷ Nur in Kolb (2021) ist tatsächlich von einem *priming effect* die Rede, in den anderen oben angeführten Werken wird diese Bezeichnung nicht verwendet und eher von „rel[iance] on MT solutions“ (Şahin & Gürses 2019:29) und von „simpler“, „more normalised“ Posteditionen mit „more interference from the source language“ (vgl. Toral, Castilho et al. 2020:1°44') gesprochen.

die nicht den für die Ratings festgelegten Kriterien (siehe Kapitel 4.4) entsprachen. V. a. das Kriterium „Deutsch als Muttersprache“ erscheint diesbezüglich interessant, aber auch der Einfluss des Kriteriums „Englisch als Arbeitssprache“ oder des Schwerpunkts bzw. des Masterstudiums an sich könnte untersucht werden. Für Untersuchungen zum Kriterium „*Red Rising* in beiden Sprachen nicht bekannt“ müssten Teilnehmende gesucht werden, die den Roman z. B. in der Ausgangssprache kennen, um herauszufinden, ob ein aufgrund dieser Kenntnis tieferes Verständnis des Romans zu anderen Bewertungen (und auch anderen Argumentationen dafür) führt.

Sicherlich könnten noch viele weitere Möglichkeiten für zukünftige Studien aufgezählt werden, aber um den Rahmen dieses Kapitels nicht zu sprengen, wird hier ein Schlussstrich gezogen. Die Ausführungen zeigen wohl trotzdem, dass sowohl im Bereich der Fehlerkategorien zur L-NMT als auch im Bereich des Posteditings von L-NMT-Output noch sehr viel unerforscht bleibt. Es besteht also enorm viel Potential für weiterführende Studien in diesen Bereichen, die evtl. durch Ergebnisse dieser Masterarbeit inspiriert werden könnten.

7 Zusammenfassung

Die vorliegende Masterarbeit beschäftigt sich mit dem Thema der literarischen maschinellen Übersetzung und dem Postediting solcher Übersetzungen. Es wurde das Ziel verfolgt, Fehlerkategorien anhand einer DeepL-Übersetzung vom Englischen ins Deutsche eines Ausschnitts des ersten *Harry-Potter*-Bandes zu entwickeln und danach zu untersuchen, wie sich die Verwendung dieser Kategorien während des Posteditings einer DeepL-Übersetzung vom Englischen ins Deutsche eines Ausschnitts aus *Red Rising* auf die resultierende Postedition und ihre Qualität auswirkt.

Nach einer kurzen Einleitung wurde zunächst ein theoretischer Überblick über die (neuronalen) maschinelle Übersetzung gegeben, wobei speziell auch auf den DeepL-Übersetzer eingegangen wurde. Es folgte ein Überblick über das Postediting vonseiten der Industrie sowie vonseiten der Forschung. Als neue Entwicklungslinie in der Postediting-Forschung wurde das Postediting von literarischen maschinellen Übersetzungen identifiziert, auf das im nächsten Kapitel, dem Forschungsüberblick, eingegangen wurde, gemeinsam mit einem zweiten Fokus auf Fehleranalysen von solchen Übersetzungen und dabei zum Einsatz kommenden Fehlertypologien.

Im vierten Kapitel wurden die verwendeten Methoden erklärt: Die Auswahl der Ausgangstexte erfolgte auf Basis von drei Kriterien, nämlich lexikalische Vielfalt, durchschnittliche Satzlänge und Anzahl an für den jeweiligen Roman spezifischen Termini. Bei der anschließenden Entwicklung der Fehlerkategorien wurde eine genaue Analyse der maschinellen Übersetzung des *Harry-Potter*-Ausgangstextes vollzogen, durch die Fehler und Problemstellen in dieser Übersetzung identifiziert und mithilfe von bereits existierenden und in diesem Forschungsgebiet verwendeten Fehlertypologien in Kategorien eingeteilt wurden. Die studentischen Posteditorinnen wurden auf Vorschlag von Professor_innen des Studienganges Translation der Universität Wien sowie nach Feststellung ihrer Ähnlichkeit ausgewählt und posteditierten selbstständig die maschinelle Übersetzung des *Red-Rising*-Ausgangstextes. Dabei benutzte Posteditorin 2 im Gegensatz zu Posteditorin 1 die entwickelten Fehlerkategorien, für die ihr eine mündliche und schriftliche Beschreibung zur Verfügung gestellt wurde. Die Korrekturen der beiden Posteditorinnen wurden von der Verfasserin dieser Arbeit analysiert, und die fertigen Posteditionen sowie die publizierte Humanübersetzung des *Red-Rising*-Ausgangstextes wurden von anderen Studierenden des oben genannten Studienganges in mehreren Ratings auf ihre Qualität hin bewertet, wobei auch Ratings in Bezug auf die entwickelten Fehlerkategorien durchgeführt wurden. Die

Ergebnisse dieser Ratings wurden nach Ermittlung der Interrater_innen-Reliabilität mittels paarweiser Rankings ausgewertet.

Im Anschluss wurden die Ergebnisse der Untersuchungen und damit die Antworten auf die Forschungsfragen präsentiert: Zunächst konnten anhand der maschinellen Übersetzung des Ausschnitts aus dem ersten *Harry-Potter*-Band zehn Fehlerkategorien entwickelt werden, nämlich 1) Auflösung von Anaphern, 2) Auslassungen und Hinzufügungen, 3) Bedeutung (samt der Unterkategorien fixe Wendungen und *phrasal verbs*, Polysemie und Silbentrennung), 4) Formatierung, 5) Grammatik und Orthographie, 6) Interpunktion, 7) IUT (im Sinne einer Terminologie, die dem jeweiligen literarischen Werk eigen ist), 8) Stil (im Sinne eines natürlichen, flüssigen und für den Kontext angemessenen Sprachgebrauchs), 9) Tempus und Modus von Verben sowie 10) Tonfall und Register. Bei den Kategorien IUT, Stil sowie Tonfall und Register erwies sich die Subjektivität als zentraler Faktor, weswegen sie eher als Problembereiche denn als tatsächliche Fehlerkategorien präsentiert wurden.

Die darauf folgende Analyse der Korrekturen der beiden Posteditorinnen im Hinblick auf die entwickelten Fehlerkategorien ergab auffällige Unterschiede in der Behandlung der folgenden Arten von Fehlern bzw. Problemstellen: Bedeutung, Interpunktion, IUT, Stil und Tonfall und Register. Bei allen außer Stil konnte ein möglicher *priming effect* dieser Kategorien auf Posteditorin 2 festgestellt werden, d. h., der Einfluss der ihr zur Verfügung gestellten Beschreibung der Fehlerkategorien war an ihren Korrekturen bzw. an den Fehlern/Problemstellen, die sie im Gegensatz zu Posteditorin 1 nicht korrigierte, sichtbar. Bei der Kategorie Stil wurden dahingegen geringere kognitive Kapazitäten bei Posteditorin 2 für die zielsprachlichen Versionen und mehr Konzentration auf den Ausgangstext als Erklärung für die geringere Anzahl an Korrekturen in dieser Kategorie vermutet. Am deutlichsten unterschieden sich die Posteditionen aber bzgl. der nicht unbedingt notwendigen, d. h. der persönlichen Vorliebe entsprechenden Korrekturen, die beide Posteditorinnen durchführten, Posteditorin 1 allerdings in viel größerem Ausmaß als Posteditorin 2. Die Rolle der Subjektivität beim Postediting machte sich durch diese Korrekturen deutlich bemerkbar und war wohl insgesamt größer als der Einfluss der Fehlerkategorien.

In den Ratings stuften die neun Rater_innen die von Posteditorin 1 ohne Fehlerkategorien angefertigte Postedition insgesamt als die beste Übersetzung ein, gefolgt von der publizierten Humanübersetzung und danach der Postedition, die Posteditorin 2 mithilfe der Fehlerkategorien erstellt hatte. Kritisiert wurde an dieser zweiten Postedition v. a. der Bereich Stil, aber sie fiel den Rater_innen auch bzgl. der Kategorien Bedeutung, Tempus und Modus von Verben sowie Tonfall und Register (eher) negativ auf. Einzig und allein in Bezug auf die

IUT wurde die zweite Postedition besser bewertet als die erste. Obwohl also die Benutzung der Fehlerkategorien während des Posteditings insgesamt gesehen einen eher negativen Einfluss auf die Qualität der Postedition von Posteditorin 2 zu haben schien, zeigten die Ergebnisse grundsätzlich schon das (stark von den jeweiligen Posteditor_innen abhängige) Potential von posteditierten literarischen DeepL-Übersetzungen, nämlich anhand der überwiegend sehr guten Bewertung der ersten Postedition, die schließlich insgesamt besser eingestuft wurde als die Humanübersetzung.

Nach der Präsentation dieser Ergebnisse wurden sie im Hinblick auf die Forschungsliteratur diskutiert, und es wurden auch mehrere Themen für mögliche weiterführende Studien genannt, zunächst zu den Fehlerkategorien: Abgesehen von einer Überprüfung der breiteren Anwendbarkeit der hier entwickelten Fehlerkategorien (etwa bezogen auf andere Genres, andere Sprachkombinationen, andere maschinelle Übersetzungssysteme etc.) könnten auch DeepL-spezifische Untersuchungen durchgeführt werden, etwa zur vermuteten *context awareness* dieses Systems oder zur Problematik der Interpunktion. Außerdem wurde vorgeschlagen, die Bereiche Silbentrennung und Formatierung sowie v. a. das in dieser Masterarbeit neu eingeführte Konzept der IUT in weiterführenden Studien zur literarischen maschinellen Übersetzung mehr zu beachten und die Rolle der Subjektivität bei der Übersetzungsevaluierung mittels Fehlerkategorien stärker zu berücksichtigen.

Im Hinblick auf das Postediting von literarischen maschinellen Übersetzungen und die resultierenden Posteditionen wurden aufgrund der Neuartigkeit der im Rahmen dieser Masterarbeit durchgeführten Untersuchung v. a. Wiederholungen dieser vorgeschlagen, auch unter teilweise anderen Bedingungen wie etwa professionellen Literaturübersetzer_innen statt Studierenden als Teilnehmer_innen o. Ä. So könnte auch der Einfluss der Erfahrung auf die Qualität der Posteditionen und auf mögliche *priming effects* untersucht werden. Bei diesen weiterführenden Untersuchungen könnten darüber hinaus zusätzliche Daten, etwa zum Postediting-Aufwand, erhoben werden. Auch eine andere Methode zur Erstellung der Posteditionen wurde erdacht, die es ermöglichen sollte, den Einfluss der Fehlerkategorien klarer vom Einfluss der subjektiven Entscheidungen der Posteditor_innen zu trennen. Darüber hinaus wurde vorgeschlagen, das Potential von DeepL und generell von neuronalen maschinellen Übersetzungssystemen zur Literaturübersetzung sowie den Einfluss der Eigenschaften der Rater_innen auf die Qualitätsbewertungen noch genauer zu untersuchen.

Bibliographie

Abbildungen

- Abb. 1: Van de Cruys, Tim (2021). Up and About, or Betwixt and Between? The Poetry of a Translation Machine. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (CALT2021@Swansea) on 12 May 2021. Abdruck mit Einverständnis des Urhebers.
- Abb. 2: Koehn, Philipp (2020). *Neural Machine Translation*. New York: Cambridge University Press, 146.
- Abb. 3: Lommel, Arle; Görög, Attila; Melby, Alan; Uszkoreit, Hans; Burchardt, Aljoscha & Popović, Maja (2015). QT21 – Deliverable 3.1: Harmonised Metric. <http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf> (Stand: 07.06.2021), 13.
- Abb. 4: Tezcan, Arda; Daems, Joke & Macken, Lieve (2019). When a ‘sport’ is a person and other issues for NMT of novels. In: Hadley et al. (eds.), 42.
- Abb. 5: Schuhmann, Katharina (2021). *Aus dem Englischen übersetzt von Maschine und Mensch. Der Einfluss von Fehlerkategorien auf die Postedition einer literarischen maschinellen Übersetzung und ihre Qualität*. Masterarbeit, Universität Wien, 110.
- Abb. 6: Schuhmann, Katharina (2021). *Aus dem Englischen übersetzt von Maschine und Mensch. Der Einfluss von Fehlerkategorien auf die Postedition einer literarischen maschinellen Übersetzung und ihre Qualität*. Masterarbeit, Universität Wien, 111.
- Abb. 7: Schuhmann, Katharina (2021). *Aus dem Englischen übersetzt von Maschine und Mensch. Der Einfluss von Fehlerkategorien auf die Postedition einer literarischen maschinellen Übersetzung und ihre Qualität*. Masterarbeit, Universität Wien, 111.

Primärliteratur

- Brown, Pierce (2014a). *Red Rising* [E-Book]. New York: Del Rey.
- Brown, Pierce (2014b). *Red Rising* [E-Book]. Übersetzt aus dem Amerikanischen ins Deutsche von Bernhard Kempen. München: Wilhelm Heyne Verlag.
- Rowling, Joanne K. (2010/1997). *Harry Potter and the Philosopher’s Stone*. London/Berlin/New York: Bloomsbury Publishing.
- Rowling, Joanne K. (2013/1998). *Harry Potter und der Stein der Weisen*. Übersetzt aus dem Englischen ins Deutsche von Klaus Fritz. Hamburg: Carlsen Verlag.

Sekundärliteratur

- Allen, Jeffrey (2003). Postediting. In: Somers, Harold (ed.) *Computers and Translation. A translator's guide*. Amsterdam/Philadelphia: John Benjamins, 297-318.
- Alvarez Vidal, Sergi; Oliver González, Antoni & Badia, Toni (2020). Post-editing for Professional Translators: Cheer or Fear? *Revista Tradumàtica. Tecnologies de la Traducció* 18, 49-69. <https://doi.org/10.5565/rev/tradumatica.275> (Stand: 31.05.2021).
- Angelone, Erik; Ehrensberger-Dow, Maureen & Massey, Gary (eds.) (2020). *The Bloomsbury Companion to Language Industry Studies*. London/New York/Oxford/New Delhi/Sydney: Bloomsbury Academic.
- Aranberri Nora; Labaka, Gorka; Diaz de Ilarraza, Arantza & Sarasola, Kepa (2014). Comparison of Postediting productivity between professional translators and lay users. In: O'Brien, Simard et al. (eds.), 20-33.
- Aziz, Wilker; Castilho, Sheila & Specia, Lucia (2012). PET: a Tool for Postediting and Assessing Machine Translation. In: Calzolari et al. (eds.), 3982-3987.
- Azura, Fachrina; Dewi, Haru Deliana & Hidayat, Rahayu Surtiati (2019). Profanity and Characterization: A Study of Translation Strategies and their Effects on *The Catcher in the Rye*. *Journal of Language and Literature* 19 (2), 43-54. <https://doi.org/10.24071/joll.v19i2.2119> (Stand: 02.08.2021).
- Bahdanau, Dzmitry; Cho, KyungHyun & Bengio, Yoshua (2014). Neural Machine Translation by Jointly Learning to Align and Translate. <https://arxiv.org/pdf/1409.0473v1.pdf> (Stand: 07.08.2021).
- Bangalore, Srinivas; Behrens, Bergljot; Carl, Michael; Ghankot, Maheshwar; Heilmann, Arndt; Nitzke, Jean; Schaeffer, Moritz & Sturm, Annegret (2016). Syntactic Variance and Priming Effects in Translation. In: Carl, Michael; Bangalore, Srinivas & Schaeffer, Moritz (eds.) *New Directions in Empirical Translation Process Research. Exploring the CRITT TPR-DB* (New Frontiers in Translation Studies). Cham/Heidelberg/New York/Dordrecht/London: Springer International Publishing, 211-238.
- Baumgarten, Hans (²2011). *Compendium RHETORICUM. Die wichtigsten Stilmittel. Eine Auswahl*. Göttingen: Vandenhoeck & Ruprecht.
- Bayer-Hohenwarter, Gerrit (2011). "Creative Shifts" as a Means of Measuring and Promoting Translational Creativity. *META – Journal des traducteurs/META – Translators' Journal* 56 (3), 663-692. <https://doi.org/10.7202/1008339ar> (Stand: 28.04.2021).

- Bentivogli, Luisa; Bisazza, Arianna; Cettolo, Mauro & Federico, Marcello (2016). Neural versus Phrase-Based Machine Translation Quality: a Case Study. In: Su, Jian; Duh, Kevin & Carreras, Xavier (eds.) *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin: Association for Computational Linguistics, 257-267. <https://www.aclweb.org/anthology/D16-1025.pdf> (Stand: 25.04.2021).
- Besacier, Laurent (2014). Traduction automatisée d'une oeuvre littéraire: une étude pilote. In: Blache, Philippe; Béchet, Frédéric & Bigi, Brigitte (eds.) *Proceedings of TALN 2014* (Volume 2: Short Papers). Marseille: Association pour le Traitement Automatique des Langues, 389-394. <https://www.aclweb.org/anthology/F14-2001.pdf> (Stand: 13.03.2021).
- Bessenyei, Gábor (2017). Neural Machine Translation: the rising star. <https://www.memsource.com/blog/neural-machine-translation-the-rising-star/> (Stand: 07.08.2021).
- Boyle, James (2008). *The Public Domain: Enclosing the Commons of the Mind*. New Haven/London: Yale University Press.
- Brants, Thorsten (2007). Statistisch basierte Sprachmodelle und maschinelle Übersetzung. In: Kallmeyer, Werner & Zifonun, Gisela (Hg.) *Sprachkorpora – Datenmengen und Erkenntnis* (Jahrbuch des Instituts für Deutsche Sprache 2006). Berlin/New York: Walter de Gruyter, 235-248.
- Brøndsted, Katrine & Dollerup, Cay (2004). The names in Harry Potter. *Perspectives: Studies in Translatology* 12 (1), 56-72. <https://doi.org/10.1080/0907676X.2004.9961490> (Stand: 06.06.2021).
- Calzolari, Nicoletta; Choukri, Khalid; Declerck, Thierry; Doğan, Mehmet U.; Maegaard, Bente; Mariani, Joseph; Moreno, Asuncion; Odijk, Jan & Piperidis, Stelios (eds.) (2012). *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*. Istanbul: European Language Resources Association. <https://aclanthology.org/volumes/L12-1/> (Stand: 08.09.2021).
- Carbonell, Jaime; Klein, Steve; Miller, David; Steinbaum, Michael; Grassian, Tomer & Frey, Jochen (2006). Context-Based Machine Translation. In: Gerber, Laurie; Habash, Nizar & Lavie, Alon (eds.) *Proceedings of the 7th Conference of the Association for Machine Translation of the Americas: Technical Papers*. Cambridge: Association for Machine Translation of the Americas, 19-28. <https://aclanthology.org/2006.amta-papers.3.pdf> (Stand: 06.08.2021).

- Carl, Michael; Dragsted, Barbara; Elming, Jakob; Hardt, Daniel & Jakobsen, Arnt L. (2011). The Process of Postediting: a Pilot Study. In: Sharp, Bernadette (ed.) *Human-machine interaction in translation: Proceedings of the 8th International Natural Language Processing and Cognitive Science Workshop* (Copenhagen Studies in Language 41). Copenhagen: Samfunslitteratur Press, 131-142. <https://www.hlt.inesc-id.pt/~fmmb/wiki/uploads/Work/dict.ref7.pdf> (Stand: 13.05.2021).
- Castilho, Sheila; Moorkens, Joss; Gaspari, Federico; Calixto, Iacer; Tinsley, John & Way, Andy (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics* 108, 109-120. <https://doi.org/10.1515/pralin-2017-0013> (Stand: 29.05.2021).
- Castilho, Sheila; Doherty, Stephen; Gaspari, Federico & Moorkens, Joss (2018). Approaches to Human and Machine Translation Quality Assessment. In: Moorkens et al. (eds.), 9-38.
- Castro, Olga (2020). Transnational Feminism, Women Writers in Translation, Stateless Cultures/Literatures in Translation. Presented at: Guerberof-Arenas & Asimakoulas (organisers), 0°42'00''-0°53'05''.
- Caswell, Isaac & Liang, Bowen (2020). Recent Advances in Google Translate. <https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html> (Stand: 12.08.2021).
- Common Sense Media (o. J.a). Harry Potter and the Sorcerer's Stone. <https://www.commonsensemedia.org/book-reviews/harry-potter-and-the-sorcerers-stone> (Stand: 02.08.2021).
- Common Sense Media (o. J.b). Red Rising. <https://www.commonsensemedia.org/book-reviews/red-rising> (Stand: 02.08.2021).
- Constantine, Peter (2020). *Google Translate Gets Voltaire: Literary Translation and the Age of Artificial Intelligence*. *Contemporary French and Francophone Studies* 23 (4), 471-479. <https://doi.org/10.1080/17409292.2019.1694798> (Stand: 14.03.2021).
- Costa-jussà, Marta R. (2012). An overview of the phrase-based statistical machine translation techniques. *The Knowledge Engineering Review* 27 (4), 413-431. <https://doi.org/10.1017/S026988891200029X> (Stand: 04.08.2021).

- Daems, Joke; Vandepitte, Sonia; Hartsuiker, Robert J. & Macken, Lieve (2015). The Impact of Machine Translation Error Types on Postediting Effort Indicators. In: O'Brien, Sharon & Simard, Michel (eds.) *Proceedings of the 4th Workshop on Postediting Technology and Practice (WPTP4)*. Miami: Association for Machine Translation in the Americas, 31-45. https://amtaweb.org/wp-content/uploads/2015/10/MTSummitXV_WPTP4Proceedings.pdf (Stand: 14.05.2021).
- Daems, Joke; Vandepitte, Sonia; Hartsuiker, Robert J. & Macken, Lieve (2017). Translation Methods and Experience: A Comparative Analysis of Human Translation and Postediting with Students and Professional Translators. *META – Journal des traducteurs/META – Translators' Journal* 62 (2), 245-270. <https://doi.org/10.7202/1041023ar> (Stand: 15.05.2021).
- da Silva, Igor A. L. & Costa, Cynthia B. (2020). On the Translation of Literature as a Human Activity par Excellence: Ethical Implications for Literary Machine Translation. *Aletria: Revista De Estudos De Literatura* 30 (4), 225-248. <https://doi.org/10.35699/2317-2096.2020.22047> (Stand: 08.04.2021).
- de Almeida, Giselle & O'Brien, Sharon (2010). Analysing Postediting Performance: Correlations with Years of Translation Experience. In: Yvon, François & Hansen, Viggo (eds.) *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*. Saint Raphaël: European Association for Machine Translation. <https://www.aclweb.org/anthology/2010.eamt-1.19.pdf> (Stand: 15.05.2021).
- DeepL (2018). Willkommen zum DeepL Blog! <https://www.deepl.com/de/blog/20180215> (Stand: 04.08.2021).
- DeepL (2020a). Erneuter Durchbruch bei der KI-Übersetzungsqualität. <https://www.deepl.com/de/blog/20200206> (Stand: 12.08.2021).
- DeepL (2020b). Amerikanisches und britisches Englisch ab sofort im DeepL Übersetzer verfügbar. <https://www.deepl.com/de/blog/20200824> (Stand: 02.08.2021).
- de Grandpair, Holger (2013). *Wie man einen Fantasy-Bestseller schreibt, Teil 2: Am Beispiel der erfolgreichsten Fantasy-Romane*. O. O.: Arthilien Fantasy-Verlag.
- do Carmo, Félix & Moorkens, Joss (2021). Differentiating editing, Postediting and revision. In: Koponen, Maarit; Mossop, Brian; Robert, Isabelle S. & Scocchera, Giovanna (eds.) *Translation Revision and Postediting. Industry Practices and Cognitive Processes*. London/New York: Routledge, 35-49.

- Dorr, Bonnie J.; Jordan, Pamela W. & Benoit, John W. (1999). A Survey of Current Paradigms in Machine Translation. In: Zerkowicz, Marvin (ed.) *Advances in Computers* (Volume 49). San Diego/San Francisco/New York: Academic Press, 1-70. https://www.researchgate.net/publication/222507296_A_Survey_of_Current_Paradigms_in_Machine_Translation (Stand: 04.08.2021).
- Ehrhardt, Claus & Heringer, Hans J. (2011). *Pragmatik* (Linguistik für Bachelor 7). Paderborn: Wilhelm Fink.
- Evershed, Nick (2021). Red Rising by Pierce Brown. <https://www.theguardian.com/books/2021/jul/01/ive-never-turned-pages-so-quickly-22-books-we-couldnt-put-down> (Stand: 01.08.2021).
- Fiederer, Rebecca & O'Brien, Sharon (2009). Quality and Machine Translation: A realistic objective? *The Journal of Specialised Translation* 11, 52-74. https://jostrans.org/issue11/art_fiederer_obrien.pdf (Stand: 13.05.2021).
- Fleiss, Joseph L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76 (5), 378-382.
- Folkart, Barbara (1996). Polyologie et registres de traduction: le cas d'*Ulysses*. *Palimpsestes* [online] 10. <https://journals.openedition.org/palimpsestes/1513> (Stand: 13.04.2021).
- Fonteyne, Margot; Tezcan, Arda & Macken, Lieve (2020). Machine Translation Under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level. In: Calzolari, Nicoletta; Béchet, Frédéric; Blache, Philippe; Choukri, Khalid; Cieri, Christopher; Declerck, Thierry; Goggi, Sara; Isahara, Hitoshi; Maegaard, Bente; Mariani, Joseph; Mazo, Hélène; Moreno, Asuncion; Odijk, Jan & Piperidis, Stelios (eds.) *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Marseille: The European Language Resources Association, 3790-3798. <http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.468.pdf> (Stand: 17.03.2021).
- Forcada, Mikel (2017). Making sense of neural machine translation. *Translation Spaces* 6 (2), 291-309. <https://doi.org/10.1075/ts.6.2.06for> (Stand: 05.08.2021).
- Gaspari, Federico; Toral, Antonio; Naskar, Sudip K.; Groves, Declan & Way, Andy (2014). Perception vs Reality: Measuring Machine Translation Post-Editing Productivity. In: O'Brien, Simard et al. (eds.), 60-72.
- Gehring, Jonas; Auli, Michael; Grangier, David; Yarats, Denis & Dauphin, Yann N. (2017). Convolutional Sequence to Sequence Learning. <https://arxiv.org/abs/1705.03122> (Stand: 12.08.2021).

- Genzel, Dimitriy; Uszkoreit, Jakob & Och, Franz (2010). “Poetic” Statistical Machine Translation: Rhyme and Meter. In: Li & Màrquez (eds.), 158-166.
- Ghazvininejad, Marjan; Choi, Yejin & Knight, Kevin (2018). Neural Poetry Translation. In: Walker, Marilyn; Ji, Heng & Stent, Amanda (eds.) *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Volume 2: Short Papers). New Orleans: Association for Computational Linguistics, 67-71. <https://www.aclweb.org/anthology/N18-2011.pdf> (Stand: 28.04.2021).
- Goodreads (2013a). Terry Brooks’s Reviews > Red Rising. <https://www.goodreads.com/review/show/745501796> (Stand: 01.08.2021).
- Goodreads (2013b). Khanh, first of her name, mother of bunnies’s Reviews > Red Rising. <https://www.goodreads.com/review/show/672290380> (Stand: 01.08.2021).
- Green, Spence; Heer, Jeffrey & Manning, Christopher D. (2013). The Efficacy of Human Postediting for Language Translation. In: Bødker, Susanne; Brewster, Steven; Baudisch, Patrick; Beaudouin-Lafon, Michel & Mackay, Wendy E. (eds.) *Proceedings of the 31st Annual CHI Conference on Human Factors in Computing Systems*. Paris: Association for Computing Machinery, 439-448. <https://doi.org/10.1145/2470654.2470718> (Stand: 09.05.2021).
- Greene, Erica; Bodrumlu, Tugba & Knight, Kevin (2010). Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation. In: Li & Màrquez (eds.), 524-533.
- Guerberof-Arenas, Ana (2013). What do professional translators think about Postediting? *The Journal of Specialised Translation* 19, 75-95. https://www.jostrans.org/issue19/art_guerberof.pdf (Stand: 29.04.2021).
- Guerberof-Arenas, Ana (2014). Correlations between productivity and quality when Postediting in a professional context. *Machine Translation* 28 (3-4), 165-186. <https://doi.org/10.1007/s10590-014-9155-y> (Stand: 09.05.2021).
- Guerberof-Arenas, Ana (2020). Pre-editing and Postediting. In: Angelone et al. (eds.), 333-360.

- Guerberof-Arenas, Ana & Asimakoulas, Dimitris (organisers) (2020). *Creative Translation and Technologies Expert Meeting*. Surrey, 29 May 2020. [Videoclip 5°49'36'']
https://surrey-ac.zoom.us/rec/play/uJcodr2u-zM3HtSRsgSDCv5xW420eqOs1CQb_KFfxEm8BIDNVaIMLYTM7FO1h7mhKk4iDA5fD0Ac0Ah (Stand: 15.03.2021).⁸⁸
- Guerberof-Arenas, Ana & Toral, Antonio (2020). The Impact of Postediting and Machine Translation on Creativity and Reading Experience. *Translation Spaces* 9 (2), 255-282.
<https://doi.org/10.1075/ts.20035.gue> (Stand: 29.04.2021).
- Guerberof-Arenas, Ana & Toral, Antonio (2021). Readers' engagement and reception in literary texts translated in different modalities. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*) on 11 May 2021.
- Gupta, Amit (2018). Introduction to Deep Learning: Part 1. *Chemical Engineering Progress* 114 (6), 22-29.
- Hadley, James; Popović, Maja; Afli, Haithem & Way, Andy (eds.) (2019). *Proceedings of the Qualities of Literary Machine Translation*. Dublin: European Association for Machine Translation. <https://www.aclweb.org/anthology/W19-73.pdf> (Stand: 14.03.2021).
- Hadley, James; Way, Andy; Aboomar, Mohammad & Poncelas, Alberto (2021). Pivot Translation Reconceptualised: A Computer-Augmented Approach to the Translation of Poetry. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*) on 12 May 2021.
- Halliday, Michael A. K. & Hasan, Ruqaiya (1976). *Cohesion in English* (English Language Series 9). London: Longman.
- Hjort, Minna (2009). Swearwords in Subtitles. *inTRAlinea* [online] Special Issue „The Translation of Dialects in Multimedia“. <http://www.intralinea.org/specials/article/1718> (Stand: 02.08.2021).
- Hu, Ke & Cadwell, Patrick (2016). A Comparative Study of Postediting Guidelines. *Baltic Journal of Modern Computing* 4 (2), 346-353. https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/4_2_23_Hu.pdf (Stand: 01.05.2021).
- Humblé, Philippe (2019). Machine translation and poetry. The case of English and Portuguese. *Literatura e Tradução* 72 (2), 41-56. <https://doi.org/10.5007/2175-8026.2019v72n2p41> (Stand: 09.04.2021).

⁸⁸ Die Videoaufnahme des virtuellen Expert_innentreffens war unter diesem Link zum Zeitpunkt der Abfassung von Kapitel 3 (April und Mai 2021) verfügbar, kurz danach wurde sie allerdings vom Netz genommen.

- ISO – International Organization for Standardization (2017). *ISO 18587:2017 Translation services – Postediting of machine translation output – Requirements*. [Preview available under] <https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:en> (Stand: 29.04.2021).
- Jentsch, Nancy K. (2006). Harry Potter and the Tower of Babel: Translating the Magic. In: Lathey, Gillian (ed.) *The Translation of Children's Literature. A Reader*. Clevedon/Buffalo/Toronto: Multilingual Matters, 190-208.
- Jia, Yanfang; Carl, Michael & Wang, Xiangling (2019a). Post-editing neural machine translation versus phrase-based machine translation for English-Chinese. *Machine Translation* 33 (1-2), 9-29. <https://doi.org/10.1007/s10590-019-09229-6> (Stand: 29.05.2021).
- Jia, Yanfang; Carl, Michael & Wang, Xiangling (2019b). How does post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation* 31, 60-86. https://jostrans.org/issue31/art_jia.pdf (Stand: 29.05.2021).
- Jones, Ruth & Irvine, Ann (2013). The (Un)Faithful Machine Translator. In: Lendvai, Piroska & Zervanou, Kalliopi (eds.) *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Sofia: Association for Computational Linguistics, 96-101. <https://www.aclweb.org/anthology/W13-2713.pdf> (Stand: 13.03.2021).
- Kattmann, Ulrich (1999). Menschenrassen. In: *Lexikon der Biologie*. Heidelberg: Spektrum Akademischer Verlag. <https://www.spektrum.de/lexikon/biologie/menschenrassen/42123> (Stand: 28.10.2021).
- Kenny, Dorothy & Winters, Marion (2020a). Machine translation, ethics and the literary translator's voice. *Translation Spaces* 9 (1), 123-149. <https://doi.org/10.1075/ts.00024.ken> (Stand: 17.03.2021).
- Kenny, Dorothy & Winters, Marion (2020b). Another Way of Looking at Machine Translation and Literary Translation. Presented at: Guerberof-Arenas & Asimakoulas (organisers), 0°26'33''-0°41'08''.
- Khalilov, Maxim (2020). Seven Machine Translation Trends in 2020. <https://blog.taus.net/seven-machine-translation-trends-in-2020> (Stand: 12.08.2021).
- Kit, Chunyu; Pan, Haihua & Webster, Jonathan J. (2002). Example-Based Machine Translation: A New Paradigm. In: Sin-wai, Chan (ed.) *Translation and Information Technology*. Hong Kong: Chinese University Press, 57-78.

- Koehn, Philipp (2010). *Statistical Machine Translation*. New York: Cambridge University Press.
- Koehn, Philipp (2017). Statistical Machine Translation. Draft of Chapter 13: Neural Machine Translation. <https://arxiv.org/pdf/1709.07809.pdf> (Stand: 07.08.2021).
- Koehn, Philipp (2020). *Neural Machine Translation*. New York: Cambridge University Press.
- Kolb, Waltraud (2016). Literaturübersetzen. In: Kadrić, Mira & Kaindl, Klaus (Hg.) *Berufsziel Übersetzen und Dolmetschen. Grundlagen, Ausbildung, Arbeitsfelder*. Tübingen: A. Francke Verlag, 137-157.
- Kolb, Waltraud (2021). Literary translation and post-editing: Priming effects and engagement with the text. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*) on 13 May 2021.
- Koponen, Maarit; Salmi, Leena & Nikulin, Markku (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation* 33 (1-2), 61-90. <https://doi.org/10.1007/s10590-019-09228-7> (Stand: 30.05.2021).
- Krings, Hans P. (2001). *Repairing Texts. Empirical Investigations of Machine Translation Postediting Processes* (Translation Studies 5). Translated from German to English by Koby, Geoffrey S.; Shreve, Gregory M.; Mischerikow, Katja & Litzer, Sarah. Kent/London: The Kent State University Press.
- Kuzman, Taja; Vintar, Špela & Arčan, Mihael (2019). Neural Machine Translation of Literary Texts from English to Slovene. In: Hadley et al. (eds.), 1-9.
- Lacruz, Isabel & Shreve, Gregory M. (2014). Pauses and Cognitive Effort in Postediting. In: O'Brien, Winther Balling et al. (eds.), 246-272.
- Landis, Richard T. & Koch, Gary G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33 (1), 159-174. <https://www.jstor.org/stable/2529310> (Stand: 13.04.2021).
- Läubli, Samuel; Fishel, Mark; Massey, Gary; Ehrensberger-Dow, Maureen & Volk, Martin (2013). Assessing Postediting Efficiency in a Realistic Translation Environment. In: O'Brien, Sharon; Simard, Michel & Specia, Lucia (eds.) *Proceedings of MT Summit XIV Workshop on Postediting Technology and Practice (WPTP-2)*. Nice: European Association for Machine Translation, 83-91. https://www.academia.edu/4778251/Second_Workshop_on_Post_Editing_Technologies_and_Practice_WPTP_2 (Stand: 09.05.2021).

- Läubli, Samuel; Amrhein Chantal; Düggin, Patrick; Gonzalez, Beatriz; Zwahlen, Alena & Volk, Martin (2019). Post-editing Productivity with Neural Machine Translation: An Empirical Assessment of Speed and Quality in the Banking and Finance Domain. In: Forcada, Mikel; Way, Andy; Haddow, Barry & Sennrich, Rico (eds.) *Proceedings of Machine Translation Summit XVII* (Volume 1: Research Track). Dublin: European Association for Machine Translation, 267-272. https://fc233576-d4db-4de9-b637-e339437e9442.filesusr.com/ugd/705d57_e6b5a5c517fc41769bdd57b67e57bdc9.pdf (Stand: 29.05.2021).
- LDC – Linguistic Data Consortium (2002). Linguistic Data Annotation Specification: Assessment of Fluency and Adequacy in Arabic-English and Chinese-English Translations. <https://catalog.ldc.upenn.edu/docs/LDC2003T17/TransAssess02.pdf> (Stand: 05.06.2021).
- Lee, Tong K. (2011). The death of the translator in machine translation: A bilingual poetry project. *Target. International Journal of Translation Studies* 23 (1), 92-112. <https://doi.org/10.1075/target.23.1.06lee> (Stand: 19.03.2021).
- Li, Hang & Màrquez, Lluís (eds.) (2010). *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Cambridge: Association for Computational Linguistics. <https://www.aclweb.org/anthology/D10-1.pdf> (Stand: 13.03.2021).
- Lommel, Arle; Burchardt, Aljoscha; Görög, Attila; Uszkoreit, Hans & Melby, Alan K. (eds.) (2015). Multidimensional Quality Metrics (MQM) Issue Types. <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> (Stand: 10.06.2021).
- Lommel, Arle; Görög, Attila; Melby, Alan; Uszkoreit, Hans; Burchardt, Aljoscha & Popović, Maja (2015). QT21 – Deliverable 3.1: Harmonised Metric. <http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf> (Stand: 07.06.2021).
- Lorena, Pérez Macías (2020). What Do Translators Think About Postediting? A Mixed-Methods Study Of Translators' Fears, Worries and Preferences On Machine Translation Postediting. *Revista Tradumática. Technologies de la Traducció* 18, 11-32. <https://doi.org/10.5565/rev/tradumatica.227> (Stand: 29.04.2021).
- Massardo, Isabella; van der Meer, Jaap; O'Brien, Sharon; Hollowood, Fred; Aranberri, Nora & Drescher, Katrin (2016). MT Postediting Guidelines. Amsterdam: TAUS Signature Editions. <https://info.taus.net/mt-Postediting-guidelines> (Stand: 08.05.2021).
- Matusov, Evgeny (2019). The Challenges of Using Neural Machine Translation for Literature. In: Hadley et al. (eds.), 10-19.

- McNamara, Timothy P. (2005). *semantic priming. perspectives from memory and word recognition* (Essays in Cognitive Psychology). New York/Hove: Psychology Press.
- Merkert, Pina (2017). Maschinelle Übersetzer: DeepL macht Google Translate Konkurrenz. <https://www.heise.de/newsticker/meldung/Maschinelle-Uebersetzer-DeepL-macht-Google-Translate-Konkurrenz-3813882.html> (Stand: 07.08.2021).
- Meyer, David E. & Schvaneveldt, Roger W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90 (2), 227-234.
- [MLL] – *Metzler Lexikon Literatur* (³2007). Burdorf, Dieter; Fasbender, Christoph & Moennighoff, Burkhard (Hg.). Stuttgart/Weimar: Verlag J.B. Metzler.
- Mohar, Tjaša; Orthaber, Sara & Onič, Tomaž (2020). Machine Translated Atwood: Utopia or Dystopia? *Elope: English Language Overseas Perspectives and Enquiries* 17 (1), 125-141. <https://doi.org/10.4312/elope.17.1.125-141> (Stand: 17.03.2021).
- Mohr, Jonathan (2004). Randomly Scrambled Lists of Numbers. <http://www.augustana.ualberta.ca/~mohrj/algorithms/scramble.html> (Stand: 23.07.2021).
- Moorkens, Joss & O'Brien, Sharon (2015). Postediting Evaluations: Trade-offs between Novice and Professional Participants. In: Durgar El-Kahlout, Ilknur; Özkan, Mehmed; Sánchez-Martínez, Felipe; Ramírez-Sánchez, Gema; Hollowood, Fred & Way, Andy (eds.) *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*. Antalya: European Association for Machine Translation, 75-81. <https://www.aclweb.org/anthology/W15-4910.pdf> (Stand: 29.04.2021).
- Moorkens, Joss; Castilho, Sheila; Gaspari, Federico & Doherty, Stephen (eds.) (2018). *Translation Quality Assessment. From Principles to Practice* (Machine Translation: Technologies and Applications 1). Cham: Springer International Publishing.
- Moorkens, Joss; Toral, Antonio; Castilho, Sheila & Way, Andy (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces* 7 (2), 240-262. <https://doi.org/10.1075/ts.18014.moo> (Stand: 14.03.2021).
- MT Summit 2019 (2019). Workshops. <https://www.mtsummit2019.com/workshops> (Stand: 12.04.2021).
- O'Brien, Sharon (2011). Towards predicting Postediting productivity. *Machine Translation* 25 (3), 197-215. <https://doi.org/10.1007/s10590-011-9096-7> (Stand: 10.05.2021).

- O'Brien, Sharon (2012). Towards a Dynamic Quality Evaluation Model for Translation. *The Journal of Specialised Translation* 17, 55-77. https://www.jostrans.org/issue17/art_obrien.pdf (Stand: 05.06.2021).
- O'Brien, Sharon; Simard, Michel & Specia, Lucia (eds.) (2014). *Proceedings of the Third Workshop on Post-Editing Technology and Practice (WPTP-3)*. Vancouver: Association for Machine Translation in the Americas. https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf (Stand: 31.05.2021).
- O'Brien, Sharon; Winther Balling, Laura; Carl, Michael; Simard, Michel & Specia, Lucia (eds.) (2014). *Post-Editing of Machine Translation. Processes and Applications*. Newcastle upon Tyne: Chambridge Scholars Publishing.
- O'Brien, Sharon; Winther Balling, Laura; Carl, Michael; Simard, Michel & Specia, Lucia (2014). Foreword. In: O'Brien, Winther Balling et al. (eds.), vii-viii.
- Oliver González, Antoni; Toral, Antonio & Guerberof, Ana (2019). InLéctor. Neural Machine Translation for the creation of bilingual ebooks. In: Hadley et al. (eds.), vii.
- Oliver González, Antoni; Toral, Antonio & Guerberof, Ana (2020). InLéctor. Neural Machine Translation for the creation of bilingual ebooks. Presented at: Guerberof-Arenas & Asimakoulas (organisers), 2°38':55''-2°46'50''.
- Omar, Abdulfattah & Gomaa, Yasser A. (2020). The Machine Translation of Literature: Implications for Translation Pedagogy. *International Journal of Emerging Technologies in Learning* 15 (11), 228-235. <https://doi.org/10.3991/ijet.v15i11.13275> (Stand: 10.04.2021).
- Ó Murchú, Eoin P. (2019). Using Intergaelic to pre-translate and subsequently post-edit a sci-fi novel from Scottish Gaelic to Irish. In: Hadley et al. (eds.), 20-25.
- Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing (2002). Bleu: a Method for Automatic Evaluation of Machine Translation. In: Isabelle, Pierre; Charniak, Eugene & Lin, Dekang (eds.) *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia: Association for Computational Linguistics, 311-318. <https://www.aclweb.org/anthology/P02-1040.pdf> (Stand: 26.04.2021).
- Penguin Random House (o. J.). Red Rising. By Pierce Brown. <https://tinyurl.com/fk5e9c4d> (Stand: 21.07.2021).
- Perkuhn, Rainer; Keibel, Holger & Kupietz, Marc (2012). – Ergänzungen zu – Korpuslinguistik. <https://tinyurl.com/wnnx8enb> (Stand: 01.08.2021).

- Plitt, Mirko & Masselot, François (2010). A Productivity Test of Statistical Machine Translation Postediting in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics* 93, 7-16. <https://ufal.mff.cuni.cz/pbml/93/art-plitt-masselot.pdf> (Stand: 13.05.2021).
- Poibeau, Thierry (2017). *Machine Translation* (The MIT Press Essential Knowledge Series) [E-Book]. Cambridge/London: The MIT Press.
- Porsiel, Jörg (Hg.) (2020). *Maschinelle Übersetzung für Übersetzungsprofis*. Berlin: BDÜ Weiterbildungs- und Fachverlagsgesellschaft.
- Posteditorin 1 (2021). E-Mail der Autorin an die Verfasserin der vorliegenden Arbeit am 21.04.2021. Unveröffentlicht.⁸⁹
- Posteditorin 2 (2021). E-Mail der Autorin an die Verfasserin der vorliegenden Arbeit am 17.08.2021. Unveröffentlicht.⁹⁰
- Reifler, Erwin (1954). The First Conference on Mechanical Translation. *Mechanical Translation* 1 (2), 23-32.
- Rocha, Marco (2000). A corpus-based study of anaphora in English and Portuguese. In: Botley, Simon & McEnery, Anthony M. (eds.) *Corpus-based and Computational Approaches to Discourse Anaphora*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 81-94.
- Rössler, Martin (Hg.) (2002). *Friedrich Daniel Ernst Schleiermacher. Akademievorträge* (Friedrich Daniel Ernst Schleiermacher. Kritische Gesamtausgabe. Schriften und Entwürfe 11). Berlin/New York: Walter de Gruyter.
- [RRW] – Red Rising Wiki (o. J.a). Slang. <https://red-rising.fandom.com/wiki/Slang> (Stand: 29.08.2021).
- [RRW] – Red Rising Wiki (o. J.b). BioSilk. <https://red-rising.fandom.com/wiki/BioSilk> (Stand: 24.09.2021).
- Rowling, Joanne K. (2010/1998). *Harry Potter and the Chamber of Secrets*. London/Berlin/New York/Sydney: Bloomsbury.
- Şahin, Mehmet & Gürses, Sabri (2019). Would MT kill creativity in literary retranslation? In: Hadley et al. (eds.), 26-34.

⁸⁹ Siehe Anhang 4.

⁹⁰ Siehe Anhang 4.

- Saldanha, Gabriela (2005). *Style of Translation: An exploration of stylistic patterns in the translations of Margaret Jull Costa and Peter Bush*. Dissertation, Dublin City University. http://doras.dcu.ie/17624/1/Gabriela_Saldanha_20121129124548.pdf (Stand: 15.07.2021).
- Sánchez-Gijón, Pilar; Moorkens, Joss & Way, Andy (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation* 33 (1-2), 31-59. <https://doi.org/10.1007/s10590-019-09232-x> (Stand: 29.05.2021).
- Schleiermacher, Friedrich D. E. (1813). Über die verschiedenen Methoden des Übersetzens. Vorgetragen in der Königlichen Akademie der Wissenschaften in Berlin.
- Schmalz, Antonia (2019). Maschinelle Übersetzung. In: Wittpahl, Volker (Hg.) *Künstliche Intelligenz. Technologie – Anwendung – Gesellschaft*. Wiesbaden: Springer Vieweg, 194-211.
- Schmidt, Bradley (2021). Ghosts in the Machine. Case studies of DeepL's proficiency and how machines still need help. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*) on 12 May 2021.
- Schuhmann, Katharina (2020). *Aus dem Englischen übersetzt ... von einer Maschine. Vergleich einer von einem Menschen erstellten und einer maschinellen literarischen Übersetzung*. Bachelorarbeit, Universität Wien.
- Screen, Benjamin (2017). Machine Translation and Welsh: Analysing free Statistical Machine Translation for the professional translation of an under-researched language pair. *The Journal of Specialised Translation* 28, 317-344. https://jostrans.org/issue28/art_screen.pdf (Stand: 13.05.2021).
- Screen, Benjamin (2019). What effect does Postediting have on the translation product from an end-user's perspective? *The Journal of Specialised Translation* 31, 133-157. https://jostrans.org/issue31/art_screen.pdf (Stand: 13.05.2021).
- Sekino, Kyoko (2015). An investigation of the relevance-theoretic approach to cognitive effort in translation and the Postediting process. *Translation & Interpreting* 7 (1), 142-154.
- Sennrich, Rico; Haddow, Barry & Birch, Alexandra (2016). Neural Machine Translation of Rare Words with Subword Units. In: Erk, Katrin/Smith, Noah A. (eds.) *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers). Berlin: Association for Computational Linguistics, 1715-1725. <https://www.aclweb.org/anthology/P16-1162.pdf> (Stand: 07.06.2021).

- Shuyin, Zhang (2021). On discussion of the machine translation of Chinese Xianxia novels. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*) on 12 May 2021.
- Simard, Michel; Goutte, Cyril & Isabelle, Pierre (2007). Statistical Phrase-Based Postediting. In: Sidner, Candace; Schultz, Tanja; Stone, Matthew & Zhai, ChengZiang (eds.) *Proceedings of the Main Conference of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*. Rochester: Association for Computational Linguistics, 508-515. <https://www.aclweb.org/anthology/N07-1064.pdf> (Stand: 02.05.2021).
- Skadiņš, Raivis; Puriņš, Maris; Skadiņa, Inguna & Vasiļjevs, Andrejs (2011). Evaluation of SMT in localization to under-resourced inflected language. In: Forcada, Mikel L.; Depraetere, Heidi & Vandeghinste, Vincent (eds.) *Proceedings of the 15th Annual Conference of the European Association for Machine Translation*. Leuven: European Association for Machine Translation, 35-40. <https://www.aclweb.org/anthology/2011.eamt-1.7.pdf> (Stand: 14.05.2021).
- Sowinski, Bernhard (1991). *Stilistik: Stiltheorien und Stilanalysen* (Sammlung Metzler. Realien zur Literatur 263). Stuttgart: J.B. Metzlersche Verlagsbuchhandlung.
- Stahlberg, Felix (2020). Neural Machine Translation: A Review and Survey. <https://arxiv.org/abs/1912.02047> (Stand: 05.08.2021).
- Stasimioti, Maria & Sosoni, Vilelmini (2020). Translation vs Post-editing of NMT Output: Measuring effort in the English-Greek language pair. In: Ortega, John E.; Federico, Marcello; Orasan, Constanin & Popovic, Maja (eds.) *Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation (14th Conference of the Association for Machine Translation in the Americas)*. Online: Association for Machine Translation in the Americas, 109-124. <https://aclanthology.org/2020.amta-pemdt.8.pdf> (Stand: 29.05.2021).
- Stein, Daniel (2009). Maschinelle Übersetzung – ein Überblick. *Journal for Language Technology and Computational Linguistics* 24 (3), 5-18. <https://tinyurl.com/mj82963b> (Stand: 05.08.2021).
- Stymne, Sara & Ahrenberg, Lars (2012). On the practice of error analysis for machine translation evaluation. In: Calzolari et al. (eds.), 1785-1790.

- Sutskever, Ilya; Vinyals, Oriol & Le, Quoc V. (2014). Sequence to Sequence Learning with Neural Networks. In: Ghahramani, Zoubin; Welling, Max; Cortes, Corinna; Lawrence, Neil D. & Weinberger, Kilian Q. (eds.) *Advances in Neural Information Processing Systems 27 (NIPS 2014)*. <https://proceedings.neurips.cc/paper/2014> (Stand: 08.08.2021), 3104-3112.
- Taivalkoski-Shilov, Kristiina (2019a). Ethical issues regarding machine(-assisted) translation of literary texts. *Perspectives. Studies in Translation Theory and Practice* 27 (5), 689-703. <https://doi.org/10.1080/0907676X.2018.1520907> (Stand: 14.03.2021).
- Taivalkoski-Shilov, Kristiina (2019b). Free indirect discourse: an insurmountable challenge for literary MT systems? In: Hadley et al. (eds.), 35-39.
- Tatsumi, Midori (2009). Correlation between Automatic Evaluation Metric Scores, Postediting Speed, and Some Other Factors. https://www.researchgate.net/publication/228805918_Correlation_between_automatic_evaluation_metric_scores_Postediting_speed_and_some_other_factors (Stand: 10.05.2021).
- TAUS – Translation Automation User Society (2017). Quality Evaluation using Adequacy and Fluency Approaches. <https://cdn2.hubspot.net/hubfs/2734675/Reports,%20ebooks/Adequacy-Fluency%20Best%20Practice%20Guidelines.pdf> (Stand: 05.06.2021).
- TAUS – Translation Automation User Society (o. J.). Harmonized DQF-MQM Error Typology. <https://www.taus.net/qt21-project#harmonized-error-typology> (Stand: 07.06.2021).
- Tezcan, Arda (2021). Antwort des Autors auf eine E-Mail der Verfasserin der vorliegenden Arbeit am 22.06.2021. Unveröffentlicht.⁹¹
- Tezcan, Arda; Hoste, Véronique & Macken, Lieve (2018). SCATE Taxonomy and Corpus of Machine Translation Errors. In: Corpas Pastor, Gloria & Durán-Muñoz (eds.) *Trends in E-Tools and Resources for Translators and Interpreters* (Approaches to Translation Studies 45). Leiden/Boston: Brill Rodopi, 219-244.
- Tezcan, Arda; Daems, Joke & Macken, Lieve (2019). When a ‘sport’ is a person and other issues for NMT of novels. In: Hadley et al. (eds.), 40-49.
- Thalia (o. J.a). Harry Potter 1 and the Philosopher’s Stone. <https://www.thalia.at/shop/home/artikeldetails/ID39151343.html> (Stand: 02.08.2021).
- Thalia (o. J.b). Red Rising. <https://www.thalia.at/shop/home/artikeldetails/ID42435712.html> (Stand: 02.08.2021).

⁹¹ Siehe Anhang 4.

- Toral, Antonio & Sánchez-Cartagena, Víctor M. (2017). A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions. In: Lapata, Mirella; Blunsom, Phil & Koller, Alexander (eds.) *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (Volume 1: Long Papers). Valencia: Association for Computational Linguistics, 1063-1073. <https://www.aclweb.org/anthology/E17-1100.pdf> (Stand: 25.04.2021).
- Toral, Antonio & Way, Andy (2014). Is Machine Translation Ready for Literature? In: Esteves-Ferreira João; Macan, Juliet; Mitkov, Ruslan & Stefanov, Olaf-Michael (eds.) *Proceedings of Translating and the Computer* 36. London: The International Association for the Advancement in Language Technology, 174-176. <http://www.tradulex.com/varia/TC36-london2014.pdf> (Stand: 09.04.2021).
- Toral, Antonio & Way, Andy (2015a). Translating Literary Text between Related Languages using SMT. In: Feldman, Anna; Kazantseva, Anna; Szpakowicz, Stan & Koolen, Corina (eds.) *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. Denver: Association for Computational Linguistics, 123-132. <https://www.aclweb.org/anthology/W15-0714.pdf> (Stand: 13.03.2021).
- Toral, Antonio & Way, Andy (2015b). Machine-assisted translation of literary text: A case study. *Translation Spaces* 4 (2), 240-267. <https://doi.org/10.1075/ts.4.2.04tor> (Stand: 13.03.2021).
- Toral, Antonio & Way, Andy (2018). What Level of Quality Can Neural Machine Translation Attain on Literary Text? In: Moorkens et al. (eds.), 263-287.
- Toral, Antonio; Wieling, Martijn & Way, Andy (2018). Postediting Effort of a Novel With Statistical and Neural Machine Translation. *Frontiers in Digital Humanities* 5, 1-11. <https://doi.org/10.3389/fdigh.2018.00009> (Stand: 14.03.2021).
- Toral, Antonio; Castilho, Sheila; Guerberof-Arenas, Ana; Hu, Ke; Moorkens, Joss; Oliver González, Antoni & Way, Andy (2020). Literary Machine(-Assisted) Translation. Presented at: Guerberof-Arenas & Asimakoulas (organisers), 1°36'00''-1°48'05''.
- Toral, Antonio; Oliver González, Antoni & Ribas Ballestin, Pau (2020). Machine Translation of Novels in the Age of Transformer. In: Porsiel (Hg.), 276-296.
- Translog (o. J.). Translog. <http://www.translog.dk/> (Stand: 10.05.2021).
- Tripathi, Sneha & Sarkhel, Juran K. (2010). Approaches to machine translation. *Annals of Library and Information Studies* 57 (4), 388-393. https://www.researchgate.net/publication/228574546_Approaches_to_machine_translation (Stand: 04.08.2021).

- Universität Wien (2017). Curriculum für das Masterstudium Translation. https://transvienna.univie.ac.at/fileadmin/user_upload/z_translationswiss/Studium/Curricula/Curriculum_MA_Translation_Juni2018.pdf (Stand: 14.07.2021).
- Urbaniak, Geoffrey C. & Plous, Scott (o. J.). Research Randomizer (Version 4.0). <https://www.randomizer.org/> (Stand: 18.07.2021).
- Van Brussel, Laura; Tezcan, Arda & Macken, Lieve (2018). A Fine-grained Error Analysis of NMT, PBMT and RBMT Output for English-to-Dutch. In: Calzolari, Nicoletta; Choukri, Khalid; Cieri, Christopher; Declerck, Thierry; Goggi, Sara; Hasida, Koiti; Isahara, Hitoshi; Maegaard, Bente; Mariani, Joseph; Mazo, Hélène; Moreno, Asuncion; Odijk, Jan; Piperidis, Stelios; Tokunaga, Takenobu (eds.) *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: European Language Resources Association (ELRA), 3799-3804. <https://www.aclweb.org/anthology/L18-1600.pdf> (Stand: 12.06.2021).
- Van de Cruys, Tim (2021). Up and About, or Betwixt and Between? The Poetry of a Translation Machine. Presented at the *Computer-Assisted Literary Translation Conference* in Swansea (*CALT2021@Swansea*) on 12 May 2021.
- van Egdom, Gys-Walt & Pluymaekers, Mark (2019). Why go the extra mile? How different degrees of post-editing affect perceptions of texts, senders and products among end users. *The Journal of Specialised Translation* 31, 158-176. https://www.jostrans.org/issue31/art_vanegdom.pdf (Stand: 13.05.2021).
- van Genabith, Josef (2020). Neural Machine Translation. In: Porsiel (Hg.), 59-115.
- Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N.; Kaiser, Łukasz & Polosukhin, Illia (2017). Attention Is All You Need. <https://arxiv.org/abs/1706.03762> (Stand: 12.08.2021).
- Venuti, Lawrence (²2008). *The translator's invisibility: a history of translation*. London/New York: Routledge.
- Vieira, Lucas N. (2014). Indices of cognitive effort in machine translation Postediting. *Machine Translation* 28 (3-4), 187-216. <https://doi.org/10.1007/s10590-014-9156-x> (Stand: 10.05.2021).
- Vieira, Lucas N. (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of Postediting process data. *Machine Translation* 30 (1-2), 41-62. <https://doi.org/10.1007/s10590-016-9188-5> (Stand: 10.05.2021).

- Vieira, Lucas N. (2020). Postediting of machine translation. In: O'Hagan, Minako (ed.) *The Routledge Handbook of Translation and Technology*. London/New York: Routledge, 319-335.
- Vieira, Lucas N.; Carl, Michael; Youdale, Roy & Zhang, Xiaochun (2020). Machine Translation and Literary Texts. A Network of Possibilities. Presented at: Guerberof-Arenas & Asimakoulas (organisers), 2°01'55''-2°18'05''.
- Voigt, Rob & Jurafsky, Dan (2012). Towards a Literary Machine Translation: The Role of Referential Cohesion. In: Elson, David; Kazantseva, Anna; Mihalcea, Rada & Szpakowicz, Stan (eds.) *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. Montréal: Association for Computational Linguistics, 18-25. <https://www.aclweb.org/anthology/W12-2503.pdf> (Stand: 13.03.2021).
- Voita, Elena; Serdyukov, Pavel; Sennrich, Rico & Titov, Ivan (2018). Context-Aware Neural Machine Translation Learns Anaphora Resolution. In: Gurevych, Iryna & Miyao, Yusuke (eds.) *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers). Melbourne: Association for Computational Linguistics, 1264-1274. <https://aclanthology.org/P18-1117.pdf> (Stand: 05.08.2021).
- Walther, Björn (2019). Fleiss' Kappa in Excel berechnen – Daten analysieren in Excel (41). [Videoclip 00°10'38''] <https://www.youtube.com/watch?v=mynT0L2Xkvs> (Stand: 15.03.2021).
- Way, Andy (2013). Traditional and Emerging Use Cases for Machine Translation. In: *Translating and the Computer* 35. London: The International Association for Advancement in Language Technology. https://www.computing.dcu.ie/~away/PUBS/2013/Way_AS LIB_2013.pdf (Stand: 02.05.2021).
- Way, Andy (2018). Quality Expectations of Machine Translation. In: Moorkens et al. (eds.), 159-178.
- Way, Andy (2020). Machine translation: Where are we at today? In: Angelone et al. (eds.), 311-323.
- Webster, Rebecca; Fonteyne, Margot; Tezcan, Arda; Macken, Lieve & Daems, Joke (2020a). Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw Neural Machine Translation Outputs in a Corpus of English Literary Classics. *Informatics* 7 (3), 1-21. <https://doi.org/10.3390/informatics7030032> (Stand: 14.03.2021).

- Webster, Rebecca; Fonteyne, Margot; Tezcan, Arda; Macken, Lieve & Daems, Joke (2020b). Annotating translation errors in a machine translated novel. <https://github.com/margotfonteyne/StylesNMT/blob/master/AnnotationGuidelines.pdf> (Stand: 10.06.2021).
- Werthmann, Antonina & Witt, Andreas (2014). Maschinelle Übersetzung – Gegenwart und Perspektiven. In: Stickel, Gerhard (ed.) *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: Peter Lang, 79-103.
- Williams, Philip; Sennrich, Rico; Post, Matt & Koehn, Philipp (2016). *Syntax-based Statistical Machine Translation* (Synthesis Lectures on Human Language Technologies 33). O. O.: Morgan & Claypool.
- Wyndham, Anna (2021). Inside DeepL: The World's Fastest-Growing, Most Secretive Machine Translation Company. <https://slator.com/inside-deepl-the-worlds-fastest-growing-most-secretive-machine-translation-company/> (Stand: 13.10.2021).
- Yamada, Masaru (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation* 31, 87-106. https://jostrans.org/issue31/art_yamada.pdf (Stand: 29.05.2021).
- Yasin, Amirah (2016). Red Rising ROSE ABOVE ALL MY EXPECTATIONS. <https://www.silverpetticoatreview.com/2016/05/24/red-rising-book-review/> (Stand: 01.08.2021).
- Zhechev, Ventsislav (2014). Analysing the Postediting of Machine Translation at Autodesk. In: O'Brien, Winther Balling et al. (eds.), 2-23.
- Zöfgen, Ekkehard (1989). Homonymie und Polysemie im allgemeinen einsprachigen Wörterbuch. In: Hausmann, Franz J.; Reichmann, Oskar; Wiegand, Herbert E. & Zgusta, Ladislav (Hg.) *Wörterbücher. Ein internationales Handbuch zur Lexikographie – Dictionaries. An International Encyclopedia of Lexicography – Dictionnaires. Encyclopédie internationale de lexicographie* (Erster Teilband – First Volume – Tome Premier) (Handbücher zur Sprach- und Kommunikationswissenschaft 5.1 – Handbooks of Linguistics and Communication Science 5.1 – Manuels de linguistique et des sciences de communication 5.1). Berlin/New York: Walter de Gruyter, 779-789.

Tertiärliteratur⁹²

- Bibliographisches Institut (o. J.a). Groß- und Kleinschreibung. <https://www.duden.de/sprachwissen/rechtschreibregeln/Gro%C3%9F-%20und%20Kleinschreibung#D72> (Stand: 26.06.2021).
- Bibliographisches Institut (o. J.b). Anführungszeichen. <https://www.duden.de/sprachwissen/rechtschreibregeln/anfuhrungszeichen> (Stand: 01.11.2021).
- Bibliographisches Institut (o. J.c). Das Semikolon. <https://www.duden.de/sprachwissen/sprachratgeber/Das-Semikolon> (Stand: 01.11.2021).
- Cambridge Dictionary (2021). Cambridge University Press (ed.). <https://dictionary.cambridge.org/> (Stand: 22.03.2021).
- Duden (2021). Bibliographisches Institut (Hg.). <https://www.duden.de/> (Stand: 22.03.2021).
- Duden Redewendungen. Wörterbuch der deutschen Idiomatik (Duden 11) (⁴2013). Dudenredaktion (Hg.). Berlin: Dudenverlag.
- DWDS – Digitales Wörterbuch der deutschen Sprache (o. J.). Berlin-Brandenburgische Akademie der Wissenschaften (Hg.). <https://www.dwds.de/d/wb-dwdsweb> (Stand: 04.04.2021).
- Fabricius-Hansen, Cathrine (⁹2016). Das Verb. In: Wöllstein, Angelika & Dudenredaktion (Hg.) *Die Grammatik. Unentbehrlich für richtiges Deutsch* (Duden 4). Berlin: Dudenverlag, 395-576.
- [FISD] – Farlex Idioms & Slang Dictionary. The Most Complete Collection of Idioms and Slang in the English Language (2017). Farlex International (ed.). Dublin: Farlex International.
- Hennig, Mathilde; Schneider, Georg J.; Osterwinter, Ralf & Steinhauer, Anja (⁸2016). *Duden. Das Wörterbuch der sprachlichen Zweifelsfälle. Richtiges und gutes Deutsch* (Duden 9). Berlin: Dudenverlag.
- LDOCE Online (o. J.). Pearson ELT (ed.). <https://www.ldoceonline.com/> (Stand: 22.03.2021).
- Lexico.com (2021a). Oxford University Press (ed.). <https://www.lexico.com/> (Stand: 22.03.2021).
- Lexico.com (2021b). How to use a comma (,). <https://www.lexico.com/grammar/comma> (Stand: 01.11.2021).

⁹² Nicht alle der in dieser Rubrik genannten Quellen sind auch tatsächlich in dieser Masterarbeit angeführt, aber sie wurden zur Entwicklung der Fehlerkategorien sowie zur Analyse der Korrekturen der Posteditorinnen benutzt (siehe Kapitel 4.2 und 4.5).

[MWD] – *Merriam-Webster.com Dictionary* (2021). Merriam-Webster (ed.). <https://www.merriam-webster.com/> (Stand: 18.09.2021).

Staadten, Steffi (²2016). *Rechtschreibung und Zeichensetzung endlich beherrschen. Regeln und Übungen*. Paderborn: Ferdinand Schöningh.

WAHRIG. *Die deutsche Rechtschreibung* (⁸2011). WAHRIG-Redaktion (Hg.). Gütersloh/München: wissenmedia.

Anhang

Anhang 1: Maschinelle Übersetzungen

HPMÜ

Dumbledore hatte Harry davon überzeugt, nicht noch einmal nach dem Spiegel von Erised zu suchen, und für den Rest der Weihnachtsferien blieb der Unsichtbarkeitsumhang zusammengefasst unten in seiner Truhe. Harry wünschte, er könnte das, was er im Spiegel gesehen hatte, so einfach vergessen, aber er konnte es nicht. Er begann, Albträume zu haben. Immer wieder träumte er davon, wie seine Eltern in einem grünen Lichtblitz verschwanden, während eine hohe Stimme vor Lachen gackerte.

'Siehst du, Dumbledore hatte recht, dieser Spiegel kann dich verrückt machen', sagte Ron, als Harry ihm von diesen Träumen erzählte.

Hermine, die am Tag vor Schulbeginn zurückkam, sah die Dinge anders. Sie war hin- und hergerissen zwischen Entsetzen über die Vorstellung, dass Harry drei Nächte hintereinander nicht im Bett war und durch die Schule streifte ("Wenn Filch dich erwischte hätte!") und Enttäuschung darüber, dass er nicht wenigstens herausgefunden hatte, wer Nicolas Flamel war.

Sie hatten die Hoffnung fast aufgegeben, Flamel jemals in einem Buch der Bibliothek zu finden, obwohl Harry immer noch sicher war, dass er den Namen irgendwo gelesen hatte. Als das Schuljahr begonnen hatte, blätterten sie in den Pausen wieder zehn Minuten lang in Büchern. Harry hatte sogar noch weniger Zeit als die anderen beiden, denn das Quidditch-Training hatte wieder begonnen.

Wood bearbeitete das Team härter als je zuvor. Selbst der endlose Regen, der den Schnee ersetzt hatte, konnte seine Laune nicht trüben. Die Weasleys beschwerten sich, dass Wood ein Fanatiker geworden sei, aber Harry war auf Woods Seite. Wenn sie ihr nächstes Spiel gegen Hufflepuff gewannen, würden sie Slytherin zum ersten Mal seit sieben Jahren in der Hausmeisterschaft überholen. Ganz abgesehen davon, dass er gewinnen wollte, stellte Harry fest, dass er weniger Albträume hatte, wenn er nach dem Training müde war.

Dann, während einer besonders nassen und schlammigen Trainingseinheit, überbrachte Wood dem Team eine schlechte Nachricht. Er war gerade sehr wütend auf die Weasleys, die sich ständig gegenseitig mit Sturzbomben bewarfen und dazu neigten, von ihren Besen zu fallen.

'Hört auf mit dem Unsinn!', schrie er. 'Das ist genau die Art von Dingen, mit denen wir das Spiel verlieren werden! Snape ist diesmal Schiedsrichter und er wird nach jeder Ausrede suchen, um Gryffindor Punkte abzuziehen!'

George Weasley ist bei diesen Worten wirklich von seinem Besen gefallen.

'*Snape* ist Schiedsrichter?', stotterte er durch einen Mund voll Schlamm. 'Wann hat er jemals ein Quidditchspiel geschiedsrichtert? Er wird nicht fair sein, wenn wir Slytherin überholen könnten.'

Der Rest des Teams landete neben George, um sich ebenfalls zu beschweren.

'Es ist nicht *meine* Schuld', sagte Wood. 'Wir müssen nur sicherstellen, dass wir ein sauberes Spiel spielen, damit Snape keine Ausrede hat, um auf uns herumzuhacken.'

Was ja schön und gut war, dachte Harry, aber er hatte einen anderen Grund, warum er Snape nicht in seiner Nähe haben wollte, während er Quidditch spielte ...

Der Rest des Teams blieb zurück, um wie immer am Ende des Trainings miteinander zu reden, aber Harry ging direkt zurück in den Gryffindor-Gemeinschaftsraum, wo er Ron und Hermine beim Schachspielen fand. Schach war das Einzige, bei dem Hermine jemals verloren hatte, etwas, von dem Harry und Ron dachten, dass es sehr gut für sie war.

'Reden Sie einen Moment nicht mit mir', sagte Ron, als Harry sich neben ihn setzte. Ich muss mich konzentrieren...' Er bemerkte Harrys Gesicht. 'Was ist denn los mit dir? Du siehst furchtbar aus.'

Leise sprechend, damit es niemand sonst hörte, erzählte Harry den anderen beiden von Snapes plötzlichem, unheimlichen Wunsch, Quidditch-Schiedsrichter zu werden.

'Spielen Sie nicht', sagte Hermine sofort.

'Sag, dass du krank bist', sagte Ron.

'Tu so, als würdest du dir das Bein brechen', schlug Hermine vor.

'Brich dir *wirklich* das Bein', sagte Ron.

'Ich kann nicht', sagte Harry. 'Es gibt keinen Ersatzsucher. Wenn ich aussteige, kann Gryffindor gar nicht mehr spielen.'

In diesem Moment stürzte Neville in den Kommunikationsraum. Wie er es geschafft hatte, durch das Porträtloch zu klettern, konnte man nur raten, denn seine Beine waren mit etwas zusammengeklebt, das sie sofort als den Beinschließerfluch erkannten. Er musste den ganzen Weg bis zum Gryffindor-Turm hoppeln.

Alle fielen vor Lachen um, außer Hermine, die aufsprang und den Gegenfluch aussprach. Nevilles Beine sprangen auseinander und er kam zitternd auf die Beine.

'Was ist passiert?', fragte Hermine und führte ihn zu Harry und Ron hinüber.

'Malfoy', sagte Neville zittrig. 'Ich habe ihn vor der Bibliothek getroffen. Er sagte, er hätte jemanden gesucht, an dem er das üben könnte.'

'Geh zu Professor McGonagall!' Hermine drängte Neville. 'Melde ihn!'

Neville schüttelte den Kopf. 'Ich will nicht noch mehr Ärger', murmelte er.

'Du musst dich ihm entgegenstellen, Neville!', sagte Ron. 'Er ist es gewohnt, auf den Leuten herumzut trampeln, aber das ist kein Grund, sich vor ihn zu legen und es ihm leichter zu machen.'

'Es gibt keinen Grund, mir zu sagen, dass ich nicht mutig genug bin, um in Gryffindor zu sein, das hat Malfoy schon getan', würgte Neville.

Harry tastete in der Tasche seines Umhangs und zog einen Schokoladenfrosch heraus, den allerletzten aus der Schachtel, die Hermine ihm zu Weihnachten geschenkt hatte. Er gab ihn Neville, der aussah, als könnte er weinen.

'Du bist zwölf von Malfoy wert', sagte Harry. 'Der Sprechende Hut hat dich für Gryffindor ausgewählt, nicht wahr? Und wo ist Malfoy? In stinkendem Slytherin.'

Nevilles Lippen zuckten zu einem schwachen Lächeln, als er den Frosch auspackte.

Danke, Harry ... Ich glaube, ich gehe ins Bett ... Willst du die Karte, du sammelst sie doch, oder?'

Als Neville wegging, schaute Harry auf die Karte "Berühmter Zauberer".

'Wieder Dumbledore', sagte er. 'Er war der erste, den ich je -'

Er keuchte. Er starrte auf die Rückseite der Karte. Dann sah er zu Ron und Hermine auf.

'*Ich habe ihn gefunden!*', flüsterte er. 'Ich habe Flamel gefunden! Ich *sagte* doch, *dass* ich den Namen schon einmal irgendwo gelesen hatte, ich habe ihn im Zug auf dem Weg hierher gelesen - hören Sie sich das an: "Professor Dumbledore ist besonders berühmt für seinen Sieg über den dunklen Zauberer Grindelwald im Jahr 1945, für die Entdeckung der zwölf Verwendungsmöglichkeiten von Drachenblut *und für seine Arbeit an der Alchemie mit seinem Partner Nicolas Flamel!*"'

Hermine sprang auf ihre Füße. Sie hatte nicht mehr so aufgeregt ausgesehen, seit sie die Noten für ihre allererste Hausaufgabe zurückbekommen hatten.

Bleib da!'", sagte sie und sprintete die Treppe zu den Mädchenschlafsälen hinauf. Harry und Ron hatten gerade noch Zeit, verwunderte Blicke auszutauschen, bevor sie mit einem riesigen alten Buch im Arm zurückflitzte.

'Ich hätte nie gedacht, dass ich hier mal reinschaue!', flüsterte sie aufgeregt. 'Das habe ich mir vor Wochen aus der Bibliothek geholt, um ein bisschen zu lesen.'

Licht?", sagte Ron, aber Hermine sagte ihm, er solle still sein, bis sie etwas nachgeschlagen habe, und fing an, hektisch in den Seiten zu blättern und vor sich hin zu murmeln.

Endlich fand sie, was sie suchte.

'Ich wusste es! Ich *wusste* es!'

'Dürfen wir schon sprechen?', sagte Ron mürrisch. Hermine ignorierte ihn.

'Nicolas Flamel', flüsterte sie dramatisch, 'ist der *einzigste bekannte Hersteller des Steins der Weisen!*'

Das hatte nicht ganz den Effekt, den sie erwartet hatte.

'Die was?', sagten Harry und Ron.

'Oh, *ehrlich*, lest ihr zwei nicht? Schaut - lest das, dort.'

Sie schob ihnen das Buch zu, und Harry und Ron lasen:

Die uralte Lehre der Alchemie befasst sich mit der Herstellung des Steins der Weisen, einer legendären Substanz mit erstaunlichen Kräften. Der Stein verwandelt jedes Metall in reines Gold. Er produziert auch das Elixier des Lebens, das den Trinker unsterblich macht.

Im Laufe der Jahrhunderte gab es viele Berichte über den Stein der Weisen, aber der einzige Stein, der derzeit existiert, gehört Herrn Nicolas Flamel, dem bekannten Alchemisten und Opernliebhaber. Herr Flamel, der letztes Jahr seinen sechshundertfünfundsechzigsten Geburtstag feierte, genießt ein ruhiges Leben in Devon mit seiner Frau Perenelle (sechshundertachtundfünfzig).

'Siehst du?', sagte Hermine, als Harry und Ron fertig waren. 'Der Hund muss Flamels Stein der Weisen bewachen! Ich wette, er hat Dumbledore gebeten, ihn für ihn zu bewachen, weil sie Freunde sind und er wusste, dass jemand hinter ihm her ist. Deshalb wollte er auch, dass der Stein aus Gringotts verschwindet!'

'Ein Stein, der Gold macht und verhindert, dass man jemals stirbt!', sagte Harry. 'Kein Wunder, dass Snape hinter ihm her ist! *Jeder* würde ihn haben wollen.'

Und es ist kein Wunder, dass wir Flamel nicht in dieser *Studie über die neuesten Entwicklungen in der Zauberei* finden konnten", sagte Ron. 'Er ist nicht gerade neu, wenn er sechshundertfünfundsechzig ist, nicht wahr?'

*

Am nächsten Morgen in Verteidigung gegen die dunklen Künste diskutierten Harry und Ron immer noch darüber, was sie mit dem Stein der Weisen machen würden, wenn sie einen hätten, während sie verschiedene Methoden zur Behandlung von Werwolfbissen aufschrieben. Erst als Ron sagte, dass er sich eine eigene Quidditch-Mannschaft kaufen würde, erinnerte

sich Harry an Snape und das kommende Spiel.

'Ich werde spielen', sagte er zu Ron und Hermine. 'Wenn ich es nicht tue, werden alle Slytherins denken, ich hätte nur zu viel Angst, Snape gegenüberzutreten. Ich werde es ihnen zeigen ... es wird ihnen wirklich das Lächeln aus dem Gesicht wischen, wenn wir gewinnen.'

'Solange wir dich nicht vom Spielfeld wischen', sagte Hermine.

RRMÜ

Die HoloCan über mir zeigt Bilder von den Kämpfen der Menschheit, die dann von steigender Musik gefolgt werden, während die Triumphe der Gesellschaft vorbeiziehen. Das Siegel der Gesellschaft, eine goldene Pyramide mit drei parallelen Balken, die an den drei Seiten der Pyramide befestigt sind, ein Kreis, der alles umgibt, brennt sich in den Bildschirm ein. Die Stimme von Octavia au Lune, der gealterten Herrscherin der Gesellschaft, erzählt vom Kampf der Menschen bei der Kolonialisierung der Planeten und Monde des Systems.

"Seit Anbeginn der Menschheit ist unsere Saga als Spezies eine der Stammeskriege gewesen. Sie war eine der Prüfungen, eine der Opfer, eine der Kühnheit, den natürlichen Grenzen der Natur zu trotzen. Jetzt sind wir durch Pflicht und Gehorsam vereint, aber unser Kampf ist nicht anders. Söhne und Töchter aller Farben, wir sind aufgefordert, erneut Opfer zu bringen. Hier, in unserer besten Stunde, werfen wir unsere besten Samen zu den Sternen. Wo sollen wir zuerst gedeihen? Auf der Venus? Merkur? Mars? Die Monde von Neptun, Jupiter?"

Ihre Stimme wird feierlich, während ihr altersloses Gesicht mit seinem königlichen Ausdruck vom HC herabschaut. Ihre Hände schimmern mit dem Symbol des Goldes, das auf ihrem Rücken prangt - ein Punkt in der Mitte eines geflügelten Kreises - goldene Flügel markieren die Seiten ihrer Unterarme. Nur eine Unvollkommenheit trübt ihr goldenes Gesicht - eine lange, sichelförmige Narbe, die entlang ihres rechten Wangenknochens verläuft. Ihre Schönheit ist wie die eines grausamen Raubvogels.

"Ihr tapferen roten Pioniere des Mar - die Stärksten der menschlichen Rasse - opfert euch für den Fortschritt, opfert euch, um den Weg für die Zukunft zu ebnen. Eure Leben, euer Blut, sind eine Anzahlung für die Unsterblichkeit der menschlichen Rasse, während wir uns über Erde und Mond hinaus bewegen. Ihr geht dorthin, wo wir nicht hinkommen. Ihr leidet, damit andere nicht leiden müssen.

"Ich grüße dich. Ich liebe dich. Das Helium-3, das Sie abbauen, ist das Lebenselixier des Terraforming-Prozesses. Bald wird der rote Planet atembare Luft und bewohnbaren Boden haben. Und bald, wenn der Mars bewohnbar ist, wenn ihr tapferen Pioniere den roten Planeten für uns weichere Farben bereit gemacht habt, werden wir uns zu euch gesellen und ihr werdet unter dem Himmel, den eure Arbeit geschaffen hat, in höchstem Ansehen stehen. Euer Schweiß und Blut treibt das Terraforming an!

"Tapfere Pioniere, denkt immer daran, dass Gehorsam die höchste Tugend ist. Vor allem Gehorsam, Respekt, Aufopferung, Hierarchie ..."

Ich finde den Küchenraum des Hauses leer, aber ich höre Eo im Schlafzimmer.

"Bleiben Sie stehen, wo Sie sind!", befiehlt sie durch die Tür. "Schauen Sie unter keinen Umständen in diesen Raum."

"Okay." Ich halte an.

Sie kommt eine Minute später heraus, aufgeregt und errötet. Ihr Haar ist mit Staub und Spinnweben bedeckt. Ich streiche mit den Händen durch das Gewirr. Sie kommt direkt von der Webbery, wo sie die BioSilk ernten.

"Du bist nicht in den Flush gegangen", sage ich und lächle.

"Hatte keine Zeit. Musste aus dem Webbery raus, um etwas abzuholen."

"Was hast du aufgeschnappt?"

Sie lächelt süß. "Du hast mich nicht geheiratet, weil ich dir alles erzähle, denk dran. Und geh nicht in dieses Zimmer."

Ich mache einen Ausfallschritt zur Tür. Sie blockiert mich und zieht mein Schweißband über meine Augen. Ihre Stirn stößt gegen meine Brust. Ich lache, schiebe das Band weg und packe ihre Schultern, um sie weit genug zurückzudrücken, um ihr in die Augen zu sehen.

"Oder was?" frage ich mit einer hochgezogenen Augenbraue.

Sie lächelt mich nur an und legt den Kopf schief. Ich weiche von der Metalltür zurück. Ich tauche ohne mit der Wimper zu zucken in geschmolzene Minenschächte ein. Aber es gibt einige Warnungen, die man abwenden kann, und andere, die man nicht abwenden kann.

Sie stellt sich auf die Zehenspitzen und gibt mir einen guten Kuss auf die Nase. "Guter Junge; ich wusste, du würdest leicht zu trainieren sein", sagt sie. Dann rümpft sie die Nase, weil sie mein Brennen riecht. Sie verhätschelt mich nicht, schimpft nicht mit mir, spricht nicht einmal, außer dass sie sagt: "Ich liebe dich", mit einem Hauch von Sorge in der Stimme.

Sie pickt die geschmolzenen Stücke meines Frysuits aus der Wunde, die sich von den Fingerknöcheln bis zum Handgelenk erstreckt, und zieht einen Netzverband mit Antibiotikum und Nervennahrung an.

"Woher hast du das?" frage ich.

"Wenn ich Sie nicht belehre, fragen Sie mich auch nicht, was was ist."

Ich küsse sie auf die Nase und spiele mit dem dünnen Band aus geflochtenem Haar um ihren Ringfinger. Mein mit Seidenstücken umwickeltes Haar bildet ihren Ehering.

"Ich habe heute Abend eine Überraschung für dich", sagt sie.

"Und ich habe einen für dich", sage ich und denke an den Laurel. Ich setze ihr mein Schweißband wie eine Krone auf den Kopf. Sie rümpft die Nase über seine Nässe.

"Oh, nun, eigentlich habe ich zwei für dich, Darrow. Schade, dass Sie nicht vorausgedacht haben. Du hättest mir einen Würfelzucker oder ein Satinlaken oder ... vielleicht sogar Kaffee zum ersten Geschenk schenken können."

"Kaffee!" Ich lache. "Was für eine Farbe, dachten Sie, haben Sie geheiratet?"

Sie seufzt. "Keine Vorteile für einen Taucher, überhaupt keine. Verrückt, stur, unüberlegt ..."

"Geschickt?" sage ich mit einem verschmitzten Lächeln, während ich meine Hand an der Seite ihres Rocks hochgleiten lasse.

"Ich schätze, das hat seine Vorteile." Sie lächelt und schlägt meine Hand weg, als sei sie eine Spinne. "Jetzt zieh die Handschuhe an, wenn du nicht von den Frauen gequatscht werden willst. Deine Mutter ist schon vorausgegangen."

[...]

Wir laufen Hand in Hand mit den anderen aus unserer Gemeinde durch die Tunnelstraßen zum Common. Lune dröhnt über uns auf dem HC, hoch oben, wie es sich für die Goldbrows (Aureate, um technisch zu sein) gehört. Sie zeigen die Schrecken einer terroristischen Bombe, die eine rote Minenmannschaft und eine orangefarbene Technikergruppe tötet. Die Söhne des Ares werden dafür verantwortlich gemacht. Ihre seltsame Glyphe des Ares, ein grausamer Helm mit stacheligen Sonnenstrahlen, die aus der Krone explodieren, brennt über den Bildschirm; Blut tropft von den Stacheln. Kinder werden verstümmelt gezeigt. Die Söhne des Ares werden als Stammesmörder bezeichnet, als Bringer des Chaos. Sie werden verteufelt. Die graue Polizei und die Soldaten der Gesellschaft schieben Trümmer weg. Zwei Soldaten der Obsidianfarbe, kolossale Männer und Frauen, fast doppelt so groß wie ich, werden zusammen mit flinken gelben Ärzten gezeigt, die mehrere Opfer aus der Explosion tragen.

Es gibt keine Söhne des Ares in Lykos. Ihr vergeblicher Krieg berührt uns nicht; wieder einmal wird eine Belohnung für Informationen über Ares, den Terroristenkönig, ausgesetzt. Wir haben die Sendung schon tausendmal gehört, und doch kommt sie uns wie Fiktion vor. Die Sons denken, dass wir schlecht behandelt werden, also jagen sie Dinge in die Luft. Es ist ein sinnloser Wutanfall. Jeder Schaden, den sie anrichten, verzögert den Fortschritt, den Mars für die anderen Farben bereit zu machen. Es schadet der Menschheit.

In der Tunnelstraße, wo Jungen darum wetten, die Decke zu berühren, strömen die Menschen aus den Townships fröhlich zum Laureltide-Tanz. Wir singen das Laureltide-Lied, während wir gehen - eine schwingende Melodie über einen Mann, der seine Braut in einem Feld voller Gold findet. Es gibt Gelächter, wenn die Jungen versuchen, an den Wänden

entlang zu rennen oder reihenweise Saltos zu schlagen, nur um dann auf ihr Gesicht zu fallen oder von einem Mädchen übertroffen zu werden.

Lichter sind entlang des langen Korridors aufgereiht. In der Ferne spielt der betrunkene Onkel Narol, inzwischen fünfunddreißig Jahre alt, auf seiner Zither für die Kinder, die um unsere Beine tanzen; auch er kann nicht ewig finster dreinschauen. Er trägt das Instrument an Schultergurten aufgehängt, so dass es auf seinen Hüften ruht, mit dem Resonanzboden aus Plastik und den vielen gespannten Metallsaiten nach oben zur Decke. Der rechte Daumen schlägt die Saiten an, außer wenn der Zeigefinger herunterfällt oder der Daumen einzelne Saiten zupft, während die linke Hand Saite für Saite die Basslinie zupft. Es ist wahnsinnig schwierig, die Zither anders als schwermütig klingen zu lassen. Onkel Narols Finger sind der Aufgabe gewachsen, aber meine machen nur tragische Musik.

Er spielte mir vor und brachte mir die Tänze bei, die mein Vater mir nie beibringen konnte. Er brachte mir sogar den verbotenen Tanz bei, für den sie dich umbringen würden. Wir haben ihn in den alten Minen getanzt. Er schlug mit einer Gerte auf meine Knöchel, bis ich nahtlos durch die schwungvollen Bewegungen pirouettierte, ein Stück Metall in der Hand, wie ein Schwert. Und wenn ich es richtig machte, küsste er meine Stirn und sagte mir, ich sei der Sohn meines Vaters. Es waren seine Lektionen, die mich lehrten, mich zu bewegen, die mich die anderen Kinder schlagen ließen, als wir in den alten Tunneln Fangen und Gespenster spielten.

"Die Golds tanzen in Paaren, die Obsidians in Dreiergruppen, die Grays in Dutzendgruppen", erklärte er mir. "Wir tanzen allein, denn nur allein üben die Helldiver. Nur allein kann ein Junge zum Mann werden."

Ich erinnere mich an diese Tage, Tage, an denen ich jung genug war, um ihn nicht für den Gestank von Gesöff in seinem Atem zu verurteilen. Ich war damals elf. Das ist erst fünf Jahre her. Und doch fühlt es sich wie ein ganzes Leben an.

Anhang 2: Posteditionen

RRPEo

Die HoloCan über mir zeigt Bilder von den Kämpfen der Menschheit, gefolgt von triumphaler Musik, als die Erfolge der Gesellschaft vorbeiziehen. Das Siegel der Gesellschaft, eine goldene Pyramide mit drei parallelen Balken an den drei Seiten der Pyramide und einem Kreis, der sie umgibt, brennt sich in den Bildschirm ein. Die Stimme von Octavia au Lune, der betagten Herrscherin der Gesellschaft, erzählt vom kontinuierlichen Bestreben der Menschheit, die Planeten und Monde des Systems zu kolonisieren.

„Seit Anbeginn der Menschheit ist die Geschichte unserer Spezies gekennzeichnet von Stammeskriegen. Von Prüfungen, Verlusten, von der Kühnheit, den Grenzen der Natur zu trotzen. Nun sind wir vereint durch Pflicht und Gehorsam, aber unsere Strapazen sind noch immer dieselben. Söhne und Töchter aller Farben, wir müssen erneut Opfer bringen. Hier, in unserer besten Stunde, übergeben wir unsere besten Abkömmlinge den Sternen. Wo werden wir zuerst gedeihen? Auf Venus? Merkur? Mars? Den Monden Neptuns, Jupiter?“

Ihre Stimme wird feierlich, während ihr zeitloses Gesicht mit seinem königlichen Ausdruck von der HC herabschaut. Ihre Hände schimmern mit dem Symbol der Goldenen, das auf dem Rücken ihrer Hände prangt - ein Punkt in der Mitte eines geflügelten Kreises, goldene Flügel an den Seiten ihrer Unterarme. Nur eine Unvollkommenheit trübt ihr goldenes Gesicht - eine lange, sichelförmige Narbe entlang ihres rechten Wangenknochens. Ihre Schönheit gleicht der eines grausamen Raubvogels.

„Ihr tapferen Roten Pioniere von Mars - die Stärksten unter uns - opfert euch für den Fortschritt, opfert euch, um den Weg für die Zukunft zu ebnen. Eure Leben, euer Blut, sind eine Anzahlung für die Unsterblichkeit der Menschheit, die jenseits von Erde und Mond hinausreisen wird. Ihr geht dorthin, wo wir es nicht können. Ihr leidet, damit andere nicht leiden müssen.

Ich salutiere euch. Ich liebe euch. Das Helium-3, das ihr abbaut, ist das Lebenselixier des Terraformings. Bald wird der rote Planet atembare Luft und bewohnbaren Boden haben. Und bald, wenn Mars bewohnbar ist, und ihr tapferen Pioniere den roten Planeten für uns weichere Farben vorbereitet habt, werden wir zu euch stoßen und ihr werdet unter dem Himmel, den eure Arbeit geschaffen hat, in höchstem Ansehen stehen. Euer Schweiß und Blut treibt das Terraforming an!

Tapfere Pioniere, denkt immer daran, dass Gehorsam die höchste Tugend ist. Gehorsam, Respekt, Aufopferung, Hierarchie ... “

Die Küche des Hauses ist leer, aber ich höre Eo aus dem Schlafzimmer.

„Bleib sofort stehen!“, hallt ihr Befehl durch die Tür. „Schau unter keinen Umständen hier rein.“

„Okay.“ Ich bleibe stehen.

Kurz darauf kommt sie hervor, nervös und mit roten Wangen, die Haare voll mit Staub und Spinnweben. Ich streiche mit den Händen durch das Gewirr. Sie kommt direkt von der Weberei, in der BioSilk geerntet wird.

„Du warst nicht im Flush“, sage ich und lächle.

„Ich hatte keine Zeit. Musste aus der Weberei raus, um was abzuholen.“

„Was denn?“

Sie lächelt süß. „Du hast mich nicht geheiratet, damit ich dir immer alles erzähle, vergiss das nicht. Und geh nicht in dieses Zimmer.“

Ich mache einen Ausfallschritt zur Tür. Sie blockiert mich und zieht mein Schweißband über meine Augen. Ihre Stirn schiebt sich gegen meine Brust. Ich lache, schiebe das Band hoch und umfasse ihre Schultern und drücke sie weit genug zurück, um ihr in die Augen sehen zu können.

„Oder was?“ frage ich mit hochgezogener Augenbraue.

Sie lächelt nur und legt den Kopf schief. Ich weiche von der Metaltür zurück. Ich kann, ohne mit der Wimper zu zucken, in geschmolzene Minenschächte eintauchen. Aber es gibt einige Warnungen, die man ignorieren kann, und andere, bei denen das keine gute Idee ist.

Sie stellt sich auf die Zehenspitzen und drückt mir einen Kuss auf die Nase. „Sehr brav! Ich wusste, dass du einfach zu trainieren bist.“ Dann rümpft sie die Nase, weil sie meine Verbrennung riecht. Sie bemitleidet mich nicht, schimpft nicht mit mir, sagt nichts außer: „Ich liebe dich“, mit einem Hauch von Sorge in der Stimme.

Sie pickt die geschmolzenen Teile meines Frysuits aus der Wunde, die sich von meinen Fingerknöcheln bis zum Handgelenk erstreckt, und legt mir einen Webwickel mit Antibiotikum und Nervennucleic an.

„Woher hast du das?“ frage ich.

„Wenn ich dir keinen Vortrag halte, dann frag du mich auch nicht, woher das stammt.“

Ich küsse sie auf die Nase und spiele mit dem dünnen Band aus geflochtenem Haar an ihrem Ringfinger. Mein Haar, mit Seidenstücken verflochten, bildet ihren Ehering.

„Ich habe heute Abend eine Überraschung für dich“, sagt sie.

„Und ich habe eine für dich“, sage ich und denke an den Laurel. Ich setze ihr mein Schweißband wie eine Krone auf den Kopf. Sie rümpft die Nase, weil es völlig durchnässt ist.

„Oh, eigentlich habe ich zwei für dich, Darrow. Schade, dass du nicht vorausgedacht hast. Du hättest mir einen Zuckerwürfel oder ein Satintuch oder ... vielleicht sogar Kaffee als Beigabe zum ersten Geschenk schenken können.“

„Kaffee!“ Ich lache. „Was dachtest du denn, welche Farbe du geheiratet hast?“

Sie seufzt. „Einen Diver zu heiraten bringt wirklich gar keine Vorteile mit sich. Verrückt sind sie, stur, leichtsinnig ...“

„Fingerfertig?“ sage ich mit einem neckischen Lächeln, während ich meine Hand an der Seite ihres Rocks hochgleiten lasse.

„Das kann seine Vorteile haben, schätze ich.“ Sie lächelt und schlägt meine Hand weg, als sei sie ein Insekt. „Jetzt zieh dir diese Handschuhe über, wenn du nicht von allen Frauen beschwatzt werden willst. Deine Mutter ist schon vorausgegangen.“

[...]

Wir gehen Hand in Hand mit den anderen aus unserem Sektor durch die Tunnelstraßen zum Common. Lune dröhnt aus der HC, hoch über uns, wie es sich für die Goldenen (Aureaten, um genau zu sein) gehört. Die Schrecken eines Terrorangriffs mit einer Bombe wird gezeigt, die eine Rote Minencrew und eine Orange Technikergruppe tötete. Die Söhne des Ares werden dafür verantwortlich gemacht. Ihre seltsame Glyphe des Ares, ein grausamer Helm mit Sonnenstacheln, die daraus hervorschießen, brennt sich in den Bildschirm. Blut tropft von den Stacheln. Verstümmelte Kinder. Die Söhne des Ares werden als Stammesmörder bezeichnet, als Bringer des Chaos. Sie werden verteufelt. Die Graue Polizei und Soldaten räumen Trümmerhaufen weg. Zwei Soldaten der Obsidianfarbe, kolossale Männer und Frauen, fast doppelt so groß wie ich, tragen zusammen mit agilen Gelben Ärzten mehrere Opfer der Explosion.

Es gibt keine Söhne des Ares in Lykos. Ihr vergeblicher Krieg berührt uns nicht. Aber wieder wird eine Belohnung ausgesetzt für Informationen über Ares, den König des Terrors. Wir haben die Ausstrahlung schon tausende Male gehört, und doch klingt sie wie eine Fiktion. Die Söhne glauben, dass wir schlecht behandelt werden, also jagen sie Dinge in die Luft. Sinnloser Trotz. Jeder Schaden, den sie anrichten, verzögert unsere Aufgabe, Mars für die anderen Farben bereit zu machen. Sie schaden der Menschheit.

In der Tunnelstraße, wo Jungen darum wetten, die Decke zu berühren, strömen die Menschen aus den Sektoren fröhlich zum Tanz des Laurels. Unterwegs singen wir das Lied des Laurels - eine ausgelassene Melodie über einen Mann, der seine Braut in einem Feld voller Gold findet. Es gibt Gelächter, als die Jungen versuchen, an den Wänden entlang zu laufen oder reihenweise Saltos zu schlagen, nur um dann auf die Nase zu fallen oder von einem Mädchen übertroffen zu werden.

Entlang des Korridors sind Lampen aufgehängt. Weiter vorne spielt mein betrunkenener Onkel Narol, alt mit seinen fünfunddreißig Jahren, auf seiner Zither für die Kinder, die um unsere Beine tanzen. Nicht einmal er kann ewig finster dreinschauen. Er trägt das Instrument an Schultergurten, so dass es an seiner Hüfte anliegt, der Plastikboden und die vielen gespannten Metallsaiten zeigen nach oben. Sein rechter Daumen schlägt die Saiten an, wenn sein Zeigefinger sich nicht gerade herabsenkt oder der Daumen an einzelnen Saiten zupft, während die linke Hand Saite für Saite am Bass spielt. Es ist wahnsinnig schwierig, die Zither nicht schwermütig klingen zu lassen. Onkel Narols Finger sind der Aufgabe gewachsen, aber meinen gelingt nur tragische Musik.

Er spielte früher für mich und brachte mir die Tänze bei, die mein Vater mir nie beibringen konnte. Er zeigte mir sogar den verbotenen Tanz, für den sie Leute umbringen. Wir tanzten ihn in den alten Minen. Er schlug meine Knöchel mit einer Gerte, bis ich nahtlos durch die schwungvollen Bewegungen Pirouetten schlagen konnte, ein Stück Metall wie ein Schwert in meiner Hand. Und wenn ich alles richtig machte, küsste er meine Stirn und sagte mir, ich sei der Sohn meines Vaters. Es war sein Unterricht, der mich lehrte, mich zu bewegen, der mich die anderen Kinder besiegen ließ, wenn wir in den alten Tunneln Fangen und Gespenster spielten.

„Die Goldenen tanzen in Paaren, die Mitglieder der Obsidiankaste in Dreiergruppen, die Grauen zu Dutzenden“, erklärte er mir. „Wir tanzen allein, denn wir Helldiver bohren auf uns allein gestellt. Nur allein kann ein Junge zum Mann werden.“

Ich vermisse diese Tage. Tage, an denen ich noch jung genug war, um ihn nicht für die Fahne in seinem Atem zu verurteilen. Damals war ich elf. Das ist erst fünf Jahre her. Und doch fühlt es sich wie ein ganzes Leben an.

RRPEm

Die HoloBüchse über mir zeigt Bilder von Menschheitskämpfen, die dann von tosender Musik übertönt werden, als die Triumphe der Gesellschaft im Schnelldurchlauf gezeigt werden. Das Gesellschaftssiegel, eine goldene Pyramide mit drei parallelen Balken, die an den drei Seiten der Pyramide befestigt sind, inmitten eines allumschließenden Kreises, brennt sich in den Bildschirm ein. Die Stimme von Octavia au Lune, der betagten Herrscherin der Gesellschaft, erzählt vom Kampf der Menschen bei der Kolonialisierung der Planeten und Monde des Systems.

„Seit Anbeginn der Menschheit ist unsere Geschichte als Spezies eine der Stammeskriege gewesen. Sie ist eine der Prüfungen gewesen, eine der Opfer, eine der Kühnheit, den natürlichen Grenzen der Natur zu trotzen. Jetzt sind wir durch Pflicht und Gehorsamkeit vereint, aber unser Kampf ist kein anderer. Söhne und Töchter aller Farben, wir sind aufgefordert, erneut Opfer zu bringen. Jetzt, in unserer großen Stunde, werfen wir unsere besten Samen zu den Sternen. Wo sollen wir zuerst aufblühen? Auf der Venus? Dem Merkur? Dem Mars? Den Monden von Neptun, Jupiter?“

Ihre Stimme wird immer feierlicher, während ihr ewig junges Gesicht mit seinem majestätischen Ausdruck vom HC herabschaut. Ihre Hände schimmern durch das Symbol des Goldes, das ihre Handrücken schmückt – ein Punkt in der Mitte eines geflügelten Kreises – goldene Flügel zeichnen sich an den Seiten ihrer Unterarme ab. Nur eine Unvollkommenheit trübt ihr goldenes Gesicht – eine lange, sichelförmige Narbe, die entlang ihres rechten Wangenknochens verläuft. Ihre Schönheit ist wie die eines grausamen Raubvogels.

„Ihr tapferen Roten Pioniere des Mars – die Stärksten der menschlichen Rasse – opfert euch für den Fortschritt, opfert euch, um den Weg für die Zukunft zu ebnen. Eure Leben, euer Blut, sind eine Vorleistung für die Unsterblichkeit der menschlichen Rasse, während wir uns über Erde und Mond hinausbewegen. Ihr geht dorthin, wo wir nicht hinkönnen. Ihr leidet, damit andere nicht leiden.“

„Ich begrüße euch. Ich liebe euch. Das Helium-3, das ihr abbaut, ist der Lebensnerv des Terraforming-Prozesses. Bald wird der rote Planet atembare Luft und bewohnbaren Boden haben. Und bald, wenn der Mars bewohnbar ist, wenn ihr tapferen Pioniere den roten Planeten für uns schwächere Farben bereit gemacht habt, werden wir uns zu euch gesellen und ihr werdet unter dem Himmel, den eure mühevolle Arbeit geschaffen hat, in höchstem Ansehen stehen. Euer Schweiß und Blut treiben das Terraforming an!“

„Tapfere Pioniere, denkt immer daran, dass Gehorsamkeit die höchste Tugend ist. Besonders Gehorsamkeit, Respekt, Aufopferung, Hierarchie ...“

Die Küche des Hauses ist leer, aber ich höre Eo im Schlafzimmer.

„Bleib, wo du bist!“, befiehlt sie mir durch die Tür. „Schau auf keinen Fall in dieses Zimmer.“

„Okay.“ Ich halte inne.

Sie kommt eine Minute später aufgeregt und mit roten Wangen heraus. Ihr Haar ist mit Staub und Weben bedeckt. Ich streiche mit den Händen durch das Gewirr. Sie kommt direkt von der Netzerei, wo sie die BioSeide ernten.

„Du warst nicht in der Spülung“, sage ich und lächle.

„Hatte keine Zeit. Musste schnell aus der Netzerei raus, um etwas abzuholen.“

„Was hast du abgeholt?“

Sie lächelt süß. „Du hast mich nicht geheiratet, weil ich dir alles erzähle, vergiss das nicht. Und geh nicht in dieses Zimmer.“

Ich mache einen Ausfallschritt zur Tür. Sie blockiert mich und zieht mein Schweißband über meine Augen. Ihre Stirn drückt gegen meine Brust. Ich lache, schiebe das Band weg und drücke sie an ihren Schultern weit genug zurück, um ihr in die Augen sehen zu können.

„Oder was?“, frage ich mit einer hochgezogenen Augenbraue.

Sie lächelt mich nur an und legt den Kopf schief. Ich weiche von der Metalltür zurück. Ich tauche ohne mit der Wimper zu zucken in geschmolzene Minenschächte ein. Es gibt jedoch einige Warnungen, die man ignorieren kann, und andere, die man nicht ignorieren kann.

Sie stellt sich auf die Zehenspitzen und gibt mir ein dickes Küsschen auf die Nase. „Braver Junge; ich wusste, man könnte dich leicht abrichten“, sagt sie. Dann rümpft sie die Nase, weil sie meine Verbrennungen riecht. Sie verhätschelt mich nicht, schimpft mich nicht, sagt nicht einmal etwas, außer: „Ich liebe dich“, mit nur einem Hauch von Sorge in ihrer Stimme.

Sie entfernt die geschmolzenen Stücke meines Schmoranzugs aus der Wunde, die sich von den Fingerknöcheln bis zum Handgelenk erstreckt, und verbindet sie fest mit einer Mullbinde mit Antibiotikum und Nervennucleid.

„Woher hast du das?“, frage ich.

„Wenn ich dir keine Standpauke halte, fragst du mich auch nicht, was was ist.“

Ich küsse sie auf die Nase und spiele mit dem dünnen Band aus geflochtenem Haar um ihren Ringfinger. Mein mit einigen Seidenfäden verflechtes Haar bildet ihren Ehering.

„Ich hab heute Abend eine Überraschung für dich“, sagt sie.

„Und ich hab eine für dich“, sage ich und denke an die Lorbeere. Ich setze ihr mein Schweißband wie eine Krone auf den Kopf. Sie rümpft die Nase, weil es ganz nass ist.

„Oh, nun, eigentlich hab ich zwei für dich, Darrow. Ein Pech, dass du nicht vorausgedacht hast. Du hättest mir einen Würfelzucker oder ein Satinlaken oder ... vielleicht sogar Kaffee als erstes Geschenk geben können.“

„Kaffee!“ Ich lache. „Was dachtest denn du, welche Farbe du geheiratet hast?“

Sie seufzt. „Keinerlei Vorteile für einen Taucher, überhaupt keine. Verrückt, stur, voreilig ...“

„Geschickt?“, sage ich mit einem verschmitzten Lächeln, während ich meine Hand an der Seite ihres Rocks hochgleiten lasse.

„Ich schätze, das hat seine Vorteile.“ Sie lächelt und schlägt meine Hand weg, als sei sie eine Spinne. „Jetzt zieh die Handschuhe an, außer du willst dir das Geschnatter von den Frauen anhören. Deine Mutter ist schon vorausgegangen.“

[...]

Wir gehen Hand in Hand mit den anderen aus unserer Gemeinde durch die Tunnelstraßen zur Allmende. Lune leiert über uns auf dem HC, hoch oben, wie es sich für die Goldrüben (Aureate, um es technisch auszudrücken) gehört. Sie zeigen die Schrecken einer Terrorbombe, die eine Rote Minenmannschaft und eine Orange Technikergruppe getötet hat. Die Söhne des Ares werden dafür verantwortlich gemacht. Ihre seltsame Glyphe des Ares, ein grausamer Helm mit spitzen Sonnenstrahlen, die aus der Krone explodieren, brennt über den Bildschirm; Blut tropft von den Spitzen. Kinder werden verstümmelt gezeigt. Die Söhne des Ares werden als Stammesmörder bezeichnet, als Chaostifter. Sie werden geächtet. Die Graue Polizei und die Soldaten der Gesellschaft schieben die Trümmer weg. Zwei Soldaten der Obsidian-Farbe, riesige Männer und Frauen, fast doppelt so groß wie ich, werden zusammen mit flinken Gelben Ärzten gezeigt, die mehrere Opfer aus der Explosion tragen.

Es gibt keine Söhne des Ares in Lykos. Ihr vergeblicher Krieg berührt uns nicht; erneut wird eine Belohnung für Informationen über Ares, den Terroristenkönig, ausgesetzt. Wir haben die Übertragung schon tausendmal gehört, und doch kommt sie uns unreal vor. Die Söhne denken, dass wir schlecht behandelt werden, also jagen sie Dinge in die Luft. Es ist ein sinnloser Wutausbruch. Jeder Schaden, den sie anrichten, verzögert den Fortschritt, den Mars für die anderen Farben bereit zu machen. Es schadet der Menschheit.

In der Tunnelstraße, wo Jungs darum wetten, die Decke zu berühren, strömen die Menschen aus den Gemeinden fröhlich zum Lorbeerflut-Tanz. Wir singen während des Gehens das Lorbeerflut-Lied – eine schwungvolle Melodie über einen Mann, der seine Braut in einem Feld voller Gold findet. Gelächter bricht aus, als die jungen Burschen versuchen, an den Wänden entlang zu laufen oder reihenweise Saltos zu schlagen, nur um dann auf den Kopf zu fallen oder von einem Mädchen übertroffen zu werden.

Lichter befinden sich entlang des langen Gangs. In der Ferne spielt der betrunkene Onkel Narol, inzwischen fünfunddreißig Jahre alt, auf seiner Zither für die Kinder, die um unsere Beine tanzen; auch er kann nicht ewig mürrisch dreinschauen. Er trägt das Instrument an Schultergurten aufgehängt, so dass der Resonanzboden aus Plastik auf seinen Hüften aufliegt und die vielen gespannten Metallsaiten nach oben zur Decke schauen. Der rechte Daumen schlägt die Saiten an, außer wenn sich der Zeigefinger nach unten bewegt oder wenn der Daumen einzelne Saiten zupft, und all das während die linke Hand Saite für Saite die Basslinie zupft. Es ist wahnsinnig schwierig, die Zither nicht schwermütig klingen zu lassen. Onkel Narols Finger sind der Aufgabe gewachsen, während meine nur tragische Musik erzeugen.

Früher spielte er mir immer vor und brachte mir die Tänze bei, die mein Vater mir nie beibringen konnte. Er brachte mir sogar den verbotenen Tanz bei, für den sie einen umbringen würden. Wir haben in den alten Minen geübt. Er schlug mit einer Gerte auf meine Knöchel, bis ich nahtlos durch die schwungvollen Bewegungen pirouettierte, mit einem Metallstück in meiner Hand, das einem Schwert ähnelte. Und wenn ich es richtig machte, küsste er meine Stirn und sagte mir, ich sei der Sohn meines Vaters. Es war sein Unterricht, der mich das Tanzen lehrte, der mich die anderen Kinder schlagen ließ, wenn wir in den alten Tunneln Fangen und Gespenster spielten.

„Die Goldenen tanzen in Paaren, die Obsidians in Dreiergruppen, die Grauen zu Dutzend“, erklärte er mir. „Wir tanzen allein, denn nur allein bohren die Höllentaucher. Nur allein kann ein Junge zum Mann werden.“

Ich vermisste diese Tage, Tage, an denen ich jung genug war, um ihn nicht für seine Alkoholfahne zu verurteilen. Ich war damals elf. Das ist erst fünf Jahre her. Und doch kommt es mir wie eine Ewigkeit vor.

Anhang 3: Anmerkungen der Rater_innen

In den Anmerkungen gilt: Übersetzung A = RRPEm, Übersetzung B = RRHÜ und Übersetzung C = RRPEo. Leere Zeilen bedeuten, dass der_die jeweilige Rater_in keine Anmerkung machte.

Erstes allgemeines Rating

R1	Übersetzung B und C sind qualitativ sehr nach aneinander, während A deutlich dahinter liegt. Den Ausschlag für das Rating gab schließlich, dass in Übersetzung C für meinen Geschmack zu viele Wörter im Englischen belassen wurden und Übersetzung B sich getraut hat, sich vom englischen Satzbau zu entfernen, um ein besseres deutsches Ergebnis zu erzielen.
R2	Übersetzung A wirkt teilweise sehr unrund (Satzstruktur, Wortwahl, Tempora); kein angenehmer Lesefluss. Tonalität bei Übersetzung B wirkt teilweise inkonsistent; ansonsten gut lesbar/verständlich, schöne Übersetzung. Tonalität bei Übersetzung C wirkt teilweise inkonsistent; kein Fan der Verwendung von Anglizismen.
R3	
R4	Übersetzung B klingt viel natürlicher und ist angenehmer zu lesen, auch der Stil erscheint mir am passendsten im Vergleich zum Original.
R5	
R6	
R7	
R8	
R9	

Auflösung von Anaphern

R1	Alle Texte waren in diesem Punkt sehr gut, aber in A gab es 2 Fehler.
R2	
R3	
R4	
R5	A: unpräzise Auflösung > they show
R6	
R7	
R8	
R9	

Auslassungen und Hinzufügungen

R1	
R2	
R3	
R4	

R5	B: Auslassung > for US softer Colors, yet again C: Hinzufügung > unsere Aufgabe
R6	B: Aureate nicht erklärt bzw. von „nur“ Gold abgegrenzt (bei A & C in Klammer); A: Betonung, dass man mit 35 (wie im AT) schon „alt“ ist, fehlt
R7	
R8	
R9	

Bedeutung (inkl. Unterkategorien)

R1	Vor allem bei B kam es zu einigen Bedeutungsfehlern, was wohl der eher freien Übersetzung geschuldet ist.
R2	
R3	
R4	
R5	A: Wortwahl, die im Kontext nicht ideal ist > z. B. innehalten, Spülung, Lorbeere; missverständliche Übersetzung > old now at 35 B: inkorrekte bzw. inkohärente Ü > Stirnband, Prämien
R6	in allen Übersetzungen Ausdrücke, die nicht mit dem AT übereinstimmen und in den jeweils anderen besser gelungen sind, z.B. A: „abrichten“, „Kaffee <u>als</u> erstes Geschenk“, „der Onkel“ statt „mein Onkel“; B: „Prämien“, „getötet werden <u>kann</u> “ (weniger Nachdruck als bei A & C); C: „salutieren“
R7	Es haben alle drei Übersetzungen ihre Probleme, aber in Übersetzung A wird „human breed“ und „human race“ mit „Rasse“ übersetzt, was im Deutschen nicht geht – in einem Buch würde ich danach wahrscheinlich aufhören zu lesen.
R8	
R9	

Formatierung

R1	Übersetzungen A und C folgen der Formatierung des AT, während B die HC Durchsage nicht kursiv setzt.
R2	
R3	
R4	
R5	B: keine Kursivsetzung und andere Art der Anführungszeichen
R6	Kursivsetzung hebt hervor, dass die Stimme von der Einspielung am Bildschirm kommt und die Sprecherin also nicht „physically present“ ist; dies fehlt bei B
R7	
R8	
R9	

Grammatik und Orthographie

R1	Obwohl alle drei Ränke belegt sind, sind die tatsächlichen Zahlen an Fehlern in allen Übersetzungen sehr niedrig.
----	---

R2	
R3	
R4	
R5	C: kleinere Fehler > euer Leben; Die Schrecken ... werden gezeigt
R6	
R7	Kein besonders großer Unterschied zwischen den drei Übersetzungen.
R8	
R9	

Interpunktion

R1	Die Interpunktion ist grundsätzlich bei allen recht gut, aber in C gibt es Interpunktion, die nicht mit der im AT übereinstimmt, Gedankenstriche sind zu Bindestrichen geworden, und Beistriche nach direkten Reden fehlen.
R2	
R3	
R4	
R5	A und B: Fehlerhafte Anführungszeichen; C: Bindestriche statt Gedankenstriche; A: Zeichensetzung z. T. zu „englisch“
R6	B bei „ein Punkt ...“ mit Gedankenstrich, Beistrich überflüssig; bei C 3x Bindestrich & 1x Beistrich statt Gedankenstrich
R7	In Übersetzung B ist so gut wie alles falsch – Anführungszeichen, Gedankenstriche, Beistriche, Strichpunkte, etc. In Übersetzung A wurde die Interpunktion richtig ins Deutsche übertragen, aber so gut wie alle Beistriche wurden auch einfach 1:1 übertragen, während in Übersetzung C die Beistrichsetzung besser ist. Allerdings sind in Übersetzung C auch einige Gedankenstriche falsch (also es sind einfache Bindestriche) und es fehlen beispielsweise Strichpunkte. Abgesehen von den Gedankenstrichen könnte man die anderen Entscheidungen jedoch argumentieren, meiner Meinung nach.
R8	Hierzu ist mir in den Übersetzungen tatsächlich so gut wie nichts aufgefallen. Der einzige Grund warum A und B schlechter sind als C ist, dass diese beiden Übersetzungen bei der Rede von Lune nach einem der Absätze erneut ein „“ setzten, womit sozusagen ein (‘‘) zu wenig gesetzt wurde bzw. ein (,,) zu viel.
R9	

IUT

R1	C erhielt den letzten Rang, weil unnötig viele IUT unübersetzt blieben, auch wenn manche englische Begriffe gut beibehalten wurden. A und B sind gleichauf, da sich in beiden ebenso viele schlechte und sehr gute Übersetzungen finden.
R2	
R3	
R4	In C ist alles einfachübernommen worden, klingt sehr befremdlich

R5	A: Vieles wortwörtlich übersetzt B: viele wörtliche Übersetzung, z. T. gänzlich inkorrekt > Hochzeitsband, Kochanzug oder normalisierend > Dusche C: weitgehend auf Englisch belassen
R6	Bei B finde ich die Umsetzung manchmal nicht detailliert genug, vereinzelt gibt es aber doch gute Ansätze, z.B. „HoloBox“ hört sich im Deutschen am besten an, aber z.B. „Bioseide“ oder „Kochanzug“ wirken in der Schreibweise schon zu „normal“ für IUT; A lest sich etwas leichter als C, wo die meisten Ausdrücke original übernommen wurden, auch „Goldrüben“ sehr gut als „slang term“, aber zu A: „HC“ stimmt nicht mit „HoloBüchse“ überein; Fazit: es gibt überall gute und weniger gelungene Elemente
R7	In Übersetzung B und C wurden einige IUTs ausgelassen, zB „Aureate“ oder „Goldbrow“, weswegen ich sie niedriger bewertet habe, auch wenn ich viele der IUTs in Übersetzung B und C gelungener finde, als in Übersetzung A, wo fast alle IUTs einfach wörtlich übersetzt sind, was auch öfters zu stockigen Formulierungen führt.
R8	Obwohl C hier keinerlei Fehler gemacht hat, war es für mich tatsächlich viel unleserlicher, wenn für die IUT die englischen Begriffe verwendet wurden, anstatt einer deutschen Übersetzung der Wörter. Dazu kommt, dass es für ein deutschsprachiges Zielpublikum leichter verständlich ist, wenn sie bspw. Lorbeere lesen anstatt Laurel, da auch nicht jede*r Leser*in Englisch kann bzw. jedes englische Wort versteht.
R9	

Stil

R1	A schneidet hier mit Abstand am schlechtesten ab, weil A sehr oft nahe an der englischen Konstruktion geblieben ist, wodurch die Übersetzung teilweise recht holprig ist.
R2	
R3	
R4	
R5	A: sehr oft holprige Übernahmen der Syntax des ATs C: hin und wieder Übernahme des AT > Bringer des Chaos
R6	etwas holprige Formulierungen bei B: „Vergiss nicht, dass du mich nicht ...“; „ich sehe niemanden in der Küche“, „Was sonst?“; A: „die man ignorieren kann“ & „die man nicht ignorieren kann“ oder nicht ideal gelöst, „während des Gehens“ nicht flüssig; C insgesamt am angenehmsten zu lesen, gute Dynamik
R7	
R8	
R9	

Tempus und Modus von Verben

R1	B hat vor allem einige Fehler in den HC-Durchsagen gemacht, die, anders als das Original, andeuten, dass die Menschheit bereits auf anderen Planeten lebt. (Was aber, soweit ich verstanden habe, den Roten nicht mitgeteilt wird.)
R2	
R3	
R4	
R5	B: einige schwerwiegende Zeitenfehler
R6	B: „ausgesetzt <u>waren</u> “ (→Vergangenheit), im Original „the struggles man <u>faces</u> “ → Gegenwart, noch immer; „hatten wir uns [...] zurückgezogen“ → passt zeitlich nicht ganz in den Absatz
R7	
R8	
R9	

Tonfall und Register

R1	Vor allem A verwendet immer wieder zu hochgestochene Ausdrücke.
R2	
R3	
R4	
R5	B: z. T. höheres Register im ZT als im AT > verlassen
R6	A: Unterschied zwischen formeller & und informeller mündlicher Sprachverwendung klar erkennbar; B & C: Dialog zwischen Darrow & Eo manchmal etwas zu formell
R7	
R8	
R9	

Zweites allgemeines Rating

R1	Nach der sehr ausführlichen Beschäftigung mit den Übersetzungen, finde ich nun C am besten. Trotz der vielen unübersetzten Begriffe, liest sich C gut und flüssig, während es in B doch einige Baustellen gibt, die ich beim ersten schnellen Lesen nicht bemerkt habe. A ist immer noch auf Platz 3, da sich die Übersetzung zu nah am Englischen hält.
R2	
R3	
R4	
R5	A ist eindeutig am schlechtesten. B und C sind prinzipiell gut, aber haben an jeweils unterschiedlichen Stellen gewisse Schwächen.
R6	A und C sind einigermaßen ausgeglichen, beim Punkt IUT könnte über die etwas kompliziertere Version C hinwegsehen werden (ergibt sich aus dem Kontext des Buches); C ist insgesamt aber flüssiger zu lesen (würde ich als besonders wichtige Kategorie einstufen); B fällt insgesamt eindeutig zurück

R7	
R8	
R9	

Anhang 4: Unveröffentlichte Quellen

Die in der Bibliographie der vorliegenden Arbeit unter „Posteditorin 1 (2021)“ und „Posteditorin 2 (2021)“ vermerkten Quellen sind E-Mails der Posteditorinnen an die Verfasserin dieser Arbeit. Im Folgenden sind diese E-Mails wortgetreu abgebildet, jedoch wurde der Name der Posteditorinnen zur Wahrung ihrer Anonymität gestrichen.

Posteditorin 1 (2021):

Hallo nochmal!

Alles klar, danke für die Info! Ich bin auch gerade in der Anfangsphase meiner Masterarbeit, kann ich also gut verstehen. :)

Wegen deiner Fragen:

Das war eine Weiterbildung, in der es um die literarische Übersetzung aus dem Norwegischen ging (Skandinavische Übersetzerwerkstatt 2019). Wir mussten für den Kurs zwei Übersetzungen vorbereiten, jeweils ca 5-7 Normseiten würde ich schätzen, die wir dann im Rahmen des 4-tägigen Kurses besprochen haben.

Ja, genau. Eher lektoriert würde ich sagen. Ich hab bis kurz vor dem Beginn meines Masters bei Pokémon gearbeitet und da Übersetzungen aus dem Japanischen und Englischen editiert (mit klarem Fokus auf dem deutschen Text, ich spreche nämlich kein Japanisch). Die Arbeit ging recht weit über reines Korrekturlesen hinaus und wir hatten als Editor*innen recht große Freiheiten, mit Posteditieren oder literarischem Lektorieren hatte das eher wenig zu tun.

Ich würde sagen, ca. 1000 Wörter am Ende (wahrscheinlich ein bisschen weniger als das, waren meistens so ca. 200 Wörter pro Aufgabe). Das war bei Frau Ripplinger (kann sie wirklich sehr empfehlen, bei ihrem Kurs habe ich wahnsinnig viel gelernt :)).

Liebe Grüße!

Posteditorin 2 (2021):

Hallo Katharina,

anbei habe ich dir mein Postediting angehängt.

Ich muss sagen, mit der In-Universe-Terminologie habe ich ziemlich gekämpft, da gute Ausdrücke im Deutschen zu finden. Und allgemein beim Stil (besonders bei einigen Erklärungen/Beschreibungen von Vorgängen oder Dingen) war ich mir manchmal nicht sicher, ob das jetzt einfach der Stil des Autors ist bzw. ob ich im Deutschen recht viel ändern soll oder nicht. (Die Frage, wie viel man eingreift, stellt sich aber denke ich eh bei jedem Postediting.)

Ich habe jetzt mal alles nach bestem Wissen und Gewissen gemacht und ich hoffe, es passt so und du kannst genug Schlüsse für deine Studie daraus ziehen.

Ich hoffe, ich konnte dir helfen und wünsche dir weiterhin viel Erfolg bei der Masterarbeit.

Liebe Grüße

Im Zuge der Recherche für die vorliegende Arbeit wurde außerdem Herr Arda **Tezcan** kontaktiert, um eine Frage bzgl. der in Kapitel 4.2 beschriebenen Fehlertypologie zu stellen. Folgender Text ist die in der Bibliographie dieser Arbeit unter Tezcan (2021) vermerkte Antwort des Wissenschaftlers auf diese Frage, die Herr Tezcan der Verständlichkeit halber in seine Antwort kopierte (kursiv gesetzter Absatz).

Tezcan (2021):

Dear Katharina,
Thank you for your e-mail.
However, I have some trouble with the subcategory "capitalisation & punctuation", and I was wondering whether you could explain to me why it belongs to Accuracy in the taxonomy adapted to literature, but to Fluency in the original taxonomy (from the 2018 article "SCATE Taxonomy and Corpus of Machine Translation Errors" in Trends in E-Tools and Resources for Translators and Interpreters).
The main distinction between the accuracy and the fluency categories (according to this taxonomy) is that fluency errors are visible on the target text alone, and accuracy errors are only visible when the source text is additionally provided. In fact, in the original version of the taxonomy, we had "capitalisation and punctuation" errors both under accuracy ("mechanical") and under fluency ("orthography"). Again the reason that they appeared under both categories is the reason that some of these errors were already visible on the target text alone (fluency) and some of them were only visible when the source text is additional provided (accuracy). But this difference was especially relevant for statistical MT systems, which did not handle capitalisation and fluency as good as neural MT systems.
The main reason that this distinction is not apparent anymore in the more recent version of the taxonomy is mainly due to using a neural MT system in the study.
I hope this helps but again please let me know if you have further questions.
Good luck with your research and your thesis.
Best regards,
Arda

Anhang 5: Abstracts

Abstract auf Deutsch

Das Ziel der vorliegenden Masterarbeit zum Thema der literarischen maschinellen Übersetzung bestand erstens darin, anhand der DeepL-Übersetzung aus dem Englischen ins Deutsche eines literarischen Ausgangstextes Fehlerkategorien für diese Art von maschinellen Übersetzungen zu entwickeln. Zweitens sollte untersucht werden, inwiefern sich die Verwendung dieser Fehlerkategorien während des Posteditings der DeepL-Übersetzung aus dem Englischen ins Deutsche eines ähnlichen Ausgangstextes auf die resultierende Postedition und ihre Qualität auswirkt.

Die Fehlerkategorien wurden durch eine genaue Analyse der ersten DeepL-Übersetzung entwickelt und beim Postediting der zweiten DeepL-Übersetzung von einer Übersetzungsstudierenden eingesetzt, während eine zweite Studentin ohne diese Kategorien posteditierte. Die resultierenden Posteditionen wurden von der Verfasserin dieser Arbeit in Bezug auf die vorgenommenen Korrekturen analysiert und gemeinsam mit der publizierten Humanübersetzung des Ausgangstextes von anderen Übersetzungsstudierenden einer Qualitätsbewertung mittels mehrerer Ratings unterzogen.

Die Analyse der ersten DeepL-Übersetzung ergab zehn Fehlerkategorien. Der Einfluss von vier dieser Kategorien (Bedeutung, Interpunktion, romanspezifische Terminologie, Tonfall und Register) auf die Posteditorin, die sie benutzte, war in ihrer Postedition deutlich sichtbar. Außerdem unterschieden sich die Posteditionen bzgl. der Kategorie Stil und v. a. der vielen nicht notwendigen Korrekturen, die beide durchführten, was die Rolle des Subjektivitätsfaktors beim Postediting verdeutlichte. In den Ratings bewerteten die Studierenden gesamt gesehen die Postedition, die ohne Fehlerkategorien angefertigt wurde, am besten, gefolgt von der publizierten Humanübersetzung und der zweiten Postedition. Die Verwendung der Fehlerkategorien durch eine Posteditorin schien sich demnach eher negativ auf die Qualität ihrer Postedition auszuwirken, wobei aber auch ihr eigener Posteditingstil eine große Rolle gespielt haben dürfte.

Abstract in English

The aim of this Master's thesis on the topic of literary machine translation was twofold: first of all, using the DeepL translation from English to German of a literary source text, error categories for this kind of machine translation output were to be developed. Secondly, it was to be studied how the use of these error categories during post-editing of the DeepL translation from English to German of a similar literary source text would affect the resulting post-edition and its quality.

The error categories were developed by closely analysing the first DeepL translation. Then, a translation student used these error categories during post-editing of the second DeepL translation, while another student post-edited the translation without these categories. The corrections made in the post-editions were analysed by the author of this thesis, and other translation students rated the post-editions as well as the published human translation of the source text for quality.

The analysis of the first machine translation yielded ten error categories. The influence of four of these categories (meaning, punctuation, novel-specific terminology, tone and

register) on the post-editor who used them was clearly visible in her post-edition. Moreover, the post-editions differed with regard to the category style and, most importantly, the many unnecessary corrections both post-editors made, which underlined the role subjectivity plays in post-editing. From the ratings, it could be deducted that, overall, the post-edition that was made without the error categories was the best translation according to the students, followed by the published human translation and the second post-edition. Therefore, the use of the error categories during post-editing seemed to have a rather negative impact on the quality of the resulting post-edition, but the post-editing style of the respective post-editor probably also had a significant role to play in this.