



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Qualitätsvergleich zwischen maschinell und human erstellter
Übersetzungen im Fachgebiet Musik in der Sprachenkombination
Englisch-Deutsch“

verfasst von / submitted by

Miloš Milohanović, BA

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

Wien, 2022 / Vienna 2022

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

UA 070 363 331

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

Masterstudium Translation UG2002
Bosnisch/Kroatisch/Serbisch
Deutsch

Betreut von / Supervisor:

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

Inhaltsverzeichnis

Danksagung.....	4
1. Einleitung.....	5
1.1. Forschungsfrage und Hypothese.....	7
1.2. Motivation und Zielsetzung.....	8
2. Grundlagen.....	10
2.1. Geschichte der maschinellen Übersetzung.....	10
2.2. Arten der maschinellen Übersetzung.....	12
2.3. Neuronale maschinelle Übersetzung.....	15
2.4. Frei Verfügbare NMÜ-Systeme.....	16
2.5. Übersetzungsqualität.....	17
2.6. Evaluierung der Übersetzungsqualität von MÜ-Systemen.....	20
2.7. Best-Worst-Scaling.....	25
2.8. Fachsprache der Musik.....	26
3. Aktueller Stand der Forschung.....	31
3.1. Qualität von NMÜ allgemein und in der Sprachkombination Englisch-Deutsch.....	31
3.2. Übersetzen im Fachgebiet Musik.....	36
4. Methodik.....	38
4.1. Auswahl der Textproben.....	38
4.2. Erstellung der Übersetzungen.....	39
4.3. Auswahl der Zielgruppe.....	40
4.4. Anwendung der BWS-Methode.....	41
4.5. Erstellung des Leitfadens für das Interview.....	42
5. Ergebnisse der Umfrage und des Interviews.....	43
5.1. Demographische Angaben.....	43
5.2. Bewertung der Übersetzungen.....	44
5.3. Mensch oder Maschine?.....	47
5.4. Auswertung der Interviews.....	49
6. Diskussion.....	53
6.1. Zusammenfassung der Umfrage.....	53
6.2. Rückblick auf die BWS-Methode.....	54
6.3. Vergleich zu früheren Studien.....	55
7. Fazit.....	57

8. Literaturverzeichnis	59
9. Anhang 1	68
10. Anhang 2	71
11. Anhang 3	74
12. Anhang 4	82
Der Leitfaden für das Interview:	82
12.1. Interview Teilnehmer*in 1:	83
12.2. Interview Teilnehmer*in 2:	86
12.3. Interview Teilnehmer*in 3:	91
12.4. Interview Teilnehmer*in 4:	96
12.5. Interview Teilnehmer*in 5:	101
12.6. Interview Teilnehmer*in 6:	105
12.7. Interview Teilnehmer*in 7:	109
Abstract	114

Danksagung

An erster Stelle gilt ein besonderer Dank meiner Betreuerin Ass.-Prof. Mag. Dr. Dagmar Gromann, BSc, die mich während meiner Forschungsarbeit mit ihren fachlichen und kompetenten Ratschlägen unterstützt und begleitet hat.

Ich bedanke mich herzlich bei allen Lehrenden am Zentrum für Translationswissenschaft, die stets ihr Bestes taten, meinen Kolleg*innen und mir während des Bachelor- und Masterstudiums wertvolles Wissen zu vermitteln. Ein großer Dank gilt auch den bereits genannten Kolleg*innen, dessen Freundschaft und Hilfe mir das Studium deutlich erleichtert haben. Ebenso bedanke ich mich bei allen Mitarbeiter*innen der Universität Wien, die diese Bildungsstätte zu dem machen, was sie ist.

Nicht weniger bedanke ich mich bei meinen Eltern, ohne deren selbstlose Unterstützung ich an der Universität Wien nicht hätte studieren können.

1. Einleitung

In der Translationswissenschaft spielt die Qualitätsbewertung von Übersetzungen, Übersetzungsmethoden und maschineller Übersetzung eine wichtige Rolle, sowohl aus praxisbezogenen Gründen in der Gegenwart als auch um zukünftige Änderungen im Feld der Translation zu erkennen. Es ist daher nachvollziehbar, dass die Frage der Übersetzungsqualität schon sehr lange behandelt wird und dass verschiedene Theorien und Ansätze verwendet wurden, um die Qualität einer Übersetzung möglichst umfangreich beurteilen zu können (vgl. z.B. House 1997 oder Reiss 1986). Neben der Erforschung, wie die Qualität menschlicher Translation gemessen werden kann, wird auch die Qualität der maschinellen Übersetzung gemessen und mit der Qualität menschlicher Übersetzung verglichen. (vgl. z. B. Ahrenberg 2017).

Was der Begriff der Qualität für eine Übersetzung genau bedeutet, wird in der Praxis im Normalfall zwischen Auftraggeber*innen und Übersetzer*innen ausgehandelt und beinhaltet unter anderem eine genaue Wiedergabe des Ausgangstextes, die Erfüllung einer bestimmten Funktion in einer bestimmten Situation und in einer Kultur, und texttypologische Überlegungen. Übersetzungsqualität ist nicht objektiv messbar und daher gibt es keine einheitliche Methode zur Bewertung von Übersetzungen. So sind beispielsweise Wetterberichte, Bedienungsanleitungen oder Geschäftsberichte Beispiele für stark konventionalisierte Textsorten, während literarische Texte oder Essays keine stark konventionalisierte Textmerkmale aufweisen. Es kann daher behauptet werden, dass die Qualität einer Fachübersetzungen von konventionalisierten Texten in erster Linie sachliche Informationen bewahren und eine einheitliche Terminologie liefern muss, mit der Absicht, Mehrdeutigkeit zu vermeiden und für Standardisierung zu sorgen (vgl. z. B. Taylor 2006: 40). Für Texte ohne stark konventionalisierten Textmerkmale könnte die Bewahrung der stilistischen und tonalen Entscheidungen, die einen Text einzigartig machen, als das wichtigste Ziel für eine literarische Übersetzung angesehen werden (vgl. Schöffner 1998).

In Hinblick auf maschinell erstellte Übersetzungen sollen zeitgemäße Bewertungsverfahren bestimmte Mängel aufdecken und wichtige Hinweise darauf geben, wo Verbesserungen notwendig sind. Üblicherweise ist es das Ziel der Forschung und Entwicklung im Bereich der maschinellen Übersetzung, die Qualität der menschlichen Übersetzung zu erreichen oder zumindest zufriedenstellende Ergebnisse für Bereiche zu liefern, in denen eine „ideale“ oder „perfekte“ Übersetzung nicht unbedingt erforderlich ist. Rasche, kontinuierliche Verbesserungen in der Informatik, der Computerlinguistik und hauptsächlich auf dem Gebiet der künstlichen neuronalen Netze haben zu maschinellen Übersetzungssystemen geführt, die in relativ kurzer Zeit im Vergleich

zu den Systemen der Vergangenheit erheblich verbessert werden konnten. Moderne statistische maschinelle Übersetzungssysteme, neuronale Übersetzungssysteme und hybride maschinelle Übersetzungssysteme leisten einen enormen Beitrag zur Bewältigung immer größer werdender Mengen zu übersetzender Texte und übertreffen ältere statistische maschinelle Übersetzungssysteme und regelbasierte maschinelle Übersetzung (vgl. z. B. Alam et al. 2019, Costa-Jussà et al. 2012 oder Dwojak & Jassem 2019). Das Ziel, maschinelle Übersetzungen in der gleichen Qualität wie menschliche Übersetzungen zu erstellen, ist trotz enormer Fortschritte größtenteils noch nicht erreicht. Hochwertige maschinell erstellte Übersetzungen verlangen je nach Verwendungszweck noch ein gewisses Maß an Vorbereitung der Ausgangstexte und Nachbearbeitung der maschinell erstellten Übersetzung.

Derzeit sind neuronale maschinelle Übersetzungssysteme auf Basis künstlicher neuronaler Netze die fortgeschrittenste Lösung auf dem Gebiet der maschinellen Übersetzung (vgl. Kenny 2018:436), die jedoch oft dem gleichen Problem zum Opfer fallen wie die Systeme der Vergangenheit. Die Qualität der Übersetzung hängt von der gewählten Sprachkombination ab, wobei Sprachen mit größeren verfügbaren Datenmengen wie Englisch und Deutsch dem System ermöglichen, weit überlegene Ergebnisse zu liefern im Vergleich zu den Fällen, in denen Sprachen mit kleineren verfügbaren Textkorpora verwendet werden.

Ein Gebiet, in welchem maschinelle Übersetzung üblicherweise zufriedenstellende Resultate liefert, ist das Fachübersetzen. Einige Merkmale, die Fachtexte auszeichnen, sind Fachlichkeit, standardisierte Terminologie und Objektivität. Diese Merkmale erleichtern das maschinelle Übersetzen, weil Individualität und unübliche Formulierungen vermieden werden und daher ist es weniger wahrscheinlich, dass maschinelle Übersetzungssysteme Fehler machen können. Einige Fachgebiete sind jedoch von Natur aus an das Subjektive gebunden. So behandeln musikwissenschaftliche Texte die Kunstgattung Musik. Und obwohl die Musikwissenschaft eine Wissenschaft und ein Fachgebiet ist, kann nicht vollkommen ausgeschlossen werden, dass musikwissenschaftliche Texte einen gewissen Grad an Subjektivität enthalten, was das maschinelle Übersetzen erschweren könnte.

Diese Masterarbeit widmet sich der Bewertung maschineller Übersetzung von Texten im Fachgebiet Musikologie anhand vom Sprachenpaar Englisch-Deutsch. In der Masterarbeit werden grundlegende Konzepte vorgestellt, die für die Bewertung genutzt werden. Zunächst wird die Masterarbeit die verschiedenen Sichtweisen in Bezug auf den Begriff der Qualität vorstellen. Anschließend werden Formen der maschinellen Übersetzung vorgestellt, wobei ein Schwerpunkt auf der neuronalen maschinellen Übersetzung liegt. Im weiteren Verlauf der Masterarbeit wird der

aktuelle Forschungsstand zu den Themen maschinelles Übersetzen und Übersetzen im Bereich der Musik analysiert. Im darauffolgenden Kapitel wird die Forschungsmethode zur Bewertung der Qualität der Übersetzung – Best-Worst-Scaling (BWS) – detailliert beschrieben. Diese Methode wurde aufgrund der subjektiven Natur der Bewertung der Qualität einer Übersetzung und von Texten im Allgemeinen gewählt. Die Bewertung wird von Fachexpert*innen im Fachgebiet Musik durchgeführt. Nach der darauffolgenden Präsentation der Resultate wird eine Diskussion dargeboten. Die Arbeit endet mit einem Fazit und dem Literaturverzeichnis, sowie mit den Anhängen zur Arbeit.

1.1. Forschungsfrage und Hypothese

Die Forschungsfrage der Masterarbeit ist, ob Musiker*innen und Musikwissenschaftler*innen einen Unterschied in der Qualität zwischen maschinell und human übersetzten musikbezogenen Texten erkennen können. Das Ziel ist es festzustellen, ob maschinelle Übersetzung (MÜ) anhand von musikbezogenen Ausgangstexten solche Zieltex te erstellen kann, dass ihre Qualität mit der von Humanübersetzungen vergleichbar ist. Die Frage ist auf zwei Unterfragen gegliedert:

1) Können Musiker*innen und Musikwissenschaftler*innen erkennen, welche Übersetzungen von MÜ-Systemen und welche von Übersetzer*innen erstellt wurden? In der Informatik existiert die Methode mit dem Namen Turing-Test, mit welcher getestet wird, ob ein Computer, ein dem Menschen gleichwertiges Denkvermögen hätte (vgl. Akman et al. 2000). Diese erste Unterfrage der Masterarbeit ist deshalb wichtig, weil sie eine Art von Test für MÜ darstellt, welche einem Turing-Test in gewisser Hinsicht ähnelt.

2) Wird in der Regel die menschliche oder die maschinelle Übersetzung im Vergleich von jeweils zwei Beispielen als die bessere eingestuft?

Im Normallfall ist zu erwarten, dass menschliche Übersetzer*innen bessere Übersetzungen liefern als MÜ-Systeme. Jedoch kann es durchaus passieren, dass Übersetzungen der MÜ-Systeme in gewissen Fällen als die besseren Übersetzungen eingestuft werden, unabhängig davon, ob sie als MÜs erkannt werden oder ob sie irrtümlich für menschliche Übersetzungen gehalten werden.

Für die Evaluierung der Qualität von MÜ existieren zahlreiche manuelle, automatische und hybride Ansätze. Für eine erste Evaluierung der MÜ im Fachgebiet Musik eignet sich die Best-Worst-Scaling Methode (BWS-Methode) am besten, denn sie bietet die Möglichkeit festzustellen, ob MÜ-Systeme Zieltex te mit solcher Qualität erstellen können, dass Fachexpert*innen sie nicht von human-erstellten Übersetzungen unterscheiden können.

Diese Masterarbeit allein wird nicht genug Daten sammeln können, um alle möglichen Fälle auszuarbeiten. So wird beispielsweise nicht untersucht, ob sich die Bewertung ändert, wenn den

Evaluator*innen neben den Übersetzungen auch der Ausgangstext gezeigt wird. Die Arbeit wird jedoch gewisse Tendenzen aufzeigen können. Es wird angenommen, dass die Übersetzungsqualität der MÜ sehr nahe an den menschlichen Übersetzungen sein wird, jedoch am Ende doch seitens des Evaluator*innen schlechter bewertet wird als die menschlichen Übersetzungen. Der Grund für diese Annahme ist, dass es für Englisch und Deutsch sehr große Textkorpora gibt und dass sich diese auf die Resultate der MÜ sehr positiv auswirken; allerdings können auch neuronale Übersetzungssysteme nicht den Kontext auf die gleiche Weise verstehen, wie menschliche Übersetzer*innen, weshalb es durchaus zu Fehlern kommen wird, die Fachexpert*innen auffallen werden.

1.2. Motivation und Zielsetzung

Heute ist MÜ relevanter und zugänglicher als je zuvor. Zu den zahlreichen Einsatzbereichen für MÜ zählen u. a. die institutionalisierte Anwendung (Übersetzungsagenturen, Unternehmen und internationale Institutionen), private Anwendung (meistens in Form von kostenlosen Online-Dienstleistungen die genutzt werden, kurze Texte, Webseiten oder digitale Dokumente schnell zu übersetzen) und öffentliche Anwendung (z. B. bei mehrsprachigen Informations- und Auskunftssystemen) (vgl. z. B. Krenz & Ramlow 2008). Die wahrscheinlich wichtigsten Gründe für die intensiviertere Verwendung von MÜ sind die Zeit- und Kostenersparnis (vgl. z. B. Trujillo 1999), aber von Bedeutung sind auch sozial-politische, kommerzielle und wissenschaftliche Faktoren (vgl. z. B. Arnold 1994). MÜ ist heute ein sehr wichtiger Bestandteil vieler Diskurse und daher ist es wichtig, existierende MÜ-Systeme zu analysieren und zu bewerten.

Bislang wurde sehr viel Aufwand betrieben, um MÜ in den Sprachenkombinationen Deutsch-Englisch und Englisch-Deutsch zu verbessern (z. B. Alkhouli et al. 2017; Arun et al. 2008) und dies hat generell zu sehr guten Ergebnissen geführt, u. a. auch aufgrund der vergleichsweise großen Textkorpora. Allerdings wurde die Qualität der MÜ in Fachgebiet Musik noch nicht erforscht. Dies ist auch darauf zurückzuführen, dass das Gebiet der Musik in seiner Ganzheit nur sehr selten aus der Sicht der Translationswissenschaft betrachtet wurde. Das, was bisher erforscht wurde, sind in erster Linie die Ansätze bei der Übersetzung von Libretti und Liedertexten, oder auch die Adaptierung von Werken mit Gesang wie z. B. Opern (vgl. Susam-Sarajeva, 2008); trotz der Tatsache, dass sich die Musikterminologien unterschiedlicher Sprachen unterscheiden (vgl. Paola & Miller 2018) und dass es in der Musikwissenschaft auch innerhalb derselben Sprache regionale terminologische Unterschiede gibt (vgl. Curtiss 1972), bleibt der Fachbereich der Musikwissenschaft in der Translationswissenschaft fast vollkommen unerforscht. Diese Masterarbeit soll einen Blick auf die Herausforderungen des Übersetzens musikwissenschaftlicher Texte bieten, sowohl für MÜ-Systeme

als auch für menschliche Übersetzer*innen. Wie schon angesprochen ist es auch wichtig einzuschätzen, ob MÜ-Systeme zufriedenstellende Resultate im Fachgebiet Musik erzielen können, da es sich um eine Wissenschaft handelt, die eine Kunstgattung untersucht und daher unumgänglich mit dem Subjektiven in Verbindung steht.

Globalisierung und Internationalisierung in der Wissenschaft ermöglichen es und verlangen sogar, interlingual und interkulturell zu forschen. In diesem Sinne ermöglicht die MÜ Wissenschaftler*innen, fremdsprachliche Texte mit relativ wenig Aufwand zu nutzen – angenommen die Qualität der MÜ ist befriedigend. Für Musikwissenschaftler*innen kann dies besonders wichtig sein, denn die unterschiedlichen kulturellen Ansätze können zu neuen Einsichten führen. Mangel an zeitlichen und finanziellen Ressourcen machen es oft unpraktisch oder sogar unmöglich, Übersetzer*innen für jeden einzelnen Text oder kurze Textpassagen zu beauftragen. Daher ist die Frage der Qualität der MÜ für den Fachbereich Musik sehr relevant.

Das Ziel der Masterarbeit ist es, human-erstellte und maschinelle Übersetzungen musikwissenschaftlicher Textproben von Fachexpert*innen im Bereich Musik bewerten zu lassen. Somit soll ein Einblick in die Leistung maschineller Übersetzung im Vergleich zur Humanübersetzung im musikwissenschaftlichen Bereich geliefert werden. Die aus dieser Masterarbeit gewonnenen Einsichten könnten zukünftig für potenzielle Anwender*innen relevant sein. In diesem Fall wären es hauptsächlich Musikwissenschaftler*innen und Musiker*innen, welche für musikwissenschaftliche oder autodidaktische Zwecke Rechercharbeit betreiben und gegebenenfalls auch unübersetzte fremdsprachige Literatur konsultieren möchten.

2. Grundlagen

Wie bereits in der Einleitung beschrieben, sind zwei Grundlagenkonzepte besonders relevant für das Verständnis des Themas und für die Beantwortung der Forschungsfrage dieser Masterarbeit. Das erste Grundlagenkonzept bezieht sich auf die Bewertung der Übersetzungsqualität im Allgemeinen und genauer auf die Bewertung der Qualität maschineller Übersetzung. Das zweite Grundlagenkonzept bezieht sich auf das maschinelle Übersetzen und seine Formen, wobei für das Sprachenpaar Englisch-Deutsch momentan die neuronale maschinelle Übersetzung (NMÜ) am relevantesten ist. Das folgende Kapitel soll einen Überblick über die in den aufgeführten Themen vorhandenen Konzepte geben.

2.1. Geschichte der maschinellen Übersetzung

In diesem Unterkapitel wird die Geschichte der maschinellen Übersetzung beschrieben, um einen Überblick über die Entwicklung von regelbasierten Übersetzungssystemen, über statistische und beispielbasierte Übersetzungssysteme bis hin zu neuronalen Übersetzungssystemen darzustellen. Die genauere Funktionsweise dieser Systeme wird in einem eigenen Unterkapitel erklärt.

Die Idee der MÜ so wie wir sie heute verstehen ist fast so alt wie die Erfindung des Computers und hat dadurch einen relativ ähnlichen Entwicklungspfad (Koehe 2020:33). Erste Ideen zur Automatisierung von Übersetzungsprozessen sind aber noch älter und können bereits viel früher zurückverfolgt werden (vgl. Schäfer 2002:19). Als Vorläufer der maschinellen Übersetzung gelten Georges Artsrouni und Petr Trojanskij, welche Patente auf Übersetzungsmaschinen angemeldet haben; bei diesen Maschinen handelte es sich eher um mechanische Wörterbücher als um echte MÜ-Systeme (vgl. Hutchins 2000:5). Diese Ansätze fanden keinen Anklang, unter anderem auch deshalb, weil sie viel Vorbearbeitung benötigten und keine grammatische Analyse durchführen konnten (vgl. Krenz & Ramlow 2009: 55f.)

Die Grundlage für die Entwicklung von automatisierten maschinellen Übersetzungssystemen wurde mit der Entwicklung der ersten Computer im den 1940er Jahren gelegt. Die ersten computerbasierten Ansätze waren regelbasiert und kommen aus der Zeit des Zweiten Weltkrieges (vgl. z. B. Quah 2006:58ff.). Intensive Forschung auf diesem Gebiet dauerte bis ca. 1960, hauptsächlich in den USA und in der Sowjetunion. Diese frühen Ansätze leisteten ungenügende Resultate und nach dem Automatic Language Processing Advisory Committee Bericht (ALPAC 1966) kam es zum Konsens, dass MÜ-Systeme zu dieser Zeit zu langsam, zu teuer und im Allgemeinen nicht nützlich genug seien. Es ist aber gleichzeitig der Fall, dass die Erwartungen an die

MÜ zu dieser Zeit unrealistisch waren, denn es wurde unter der Annahme gearbeitet, dass MÜ perfekte Translation liefern soll (vgl. Quah 2006:60).

Der ALPAC-Bericht brachte die Forschung im Bereich der maschinellen Übersetzung fast zum Stillstand und infolgedessen war die erste Hälfte der 1970er Jahre nicht sehr produktiv für die maschinelle Übersetzung, vor allem in den USA. Außerhalb der USA war die Traduction Automatique à l'Université de Montréal (TAUM) in Kanada eine der wenigen Forschergruppen, die größere Forschungen zur maschinellen Übersetzung fortsetzten. Im Jahr 1976 wurde das maschinelle Übersetzungssystem TAUM-Météo entwickelt, welches auch heute noch in Betrieb ist und zur Übersetzung von meteorologischen Mitteilungen in der Sprachenkombination Englisch-Französisch eingesetzt wird (vgl. Quah 2006:61). Die Forschung wurde auch in Deutschland fortgesetzt, wo das Saarbrücker Übersetzungssystem (SUSY) für die Sprachen Deutsch, Englisch und Russisch entwickelt wurde (vgl. Quah 2006:61). Darüber hinaus wurde die Forschung in der Sowjetunion, Ungarn, der Tschechoslowakei, China und Japan fortgesetzt, wo ebenfalls einige weniger bekannte maschinelle Übersetzungssysteme entwickelt wurden (Quah 2006:62).

In der Zeit zwischen den späten 1970er und den frühen 1990er Jahren erlebte die Forschung im Bereich der maschinellen Übersetzung weltweit einen Aufschwung mit der Entwicklung mehrerer Systeme, darunter SPANAM, ENGSPAN, METAL, EUROTRA, PaTrans und Mu. In dieser Zeit begannen auch große japanische Elektronik- und Telekommunikationsunternehmen wie Fujitsu und Hitachi, maschinelle Übersetzungssysteme für den kommerziellen Einsatz zu entwickeln. In den 1980er Jahren wurden Fortschritte in der Computerlinguistik erzielt, die es der Forschung im Bereich der maschinellen Übersetzungssysteme ermöglichten, anspruchsvollere Ansätze für die Übersetzung zu entwickeln. Diese Fortschritte haben zur Entwicklung von fortschrittlicheren regelbasierten Systemen geführt, aber auch die Entwicklung von datengesteuerten Ansätzen ermöglicht (Quah 2006:62ff). Die ersten datengesteuerten Ansätze wurden seit den 1990er Jahren entwickelt. Gleichzeitig wurden sowohl statistische als auch beispielbasierte Systeme entwickelt. Keine der beiden Methoden hat sich auf syntaktische oder semantische Regeln gestützt, sondern nur auf große elektronische Textkorpora, um Äquivalenzmuster festzustellen. Die Forschung auf dem Bereich älterer Methoden wurde auch mit der Einführung neuer Methoden fortgeführt (Quah 2006:62ff).

Gleichzeitig kann eine Verschiebung in der maschinellen Übersetzungsforschung von „rein“ zu „angewandt“ beobachtet werden. Dieser Trend hat sich über die Jahre fortgesetzt, zusammen mit der Integration von maschinellen Übersetzungssystemen mit anderen Übersetzungswerkzeugen, insbesondere im Bereich der Softwarelokalisierung. Eine Wegbewegung von Stand-Alone-Systemen zu Netzwerksystemen und von offline zu online kann ebenfalls bemerkt werden. Der Einfluss, den

das Internet und die relativ kostengünstigen PCs seit Ende der 1990er und Anfang der 2000er Jahre auf die Forschung im Bereich der maschinellen Übersetzung hatten, hatte ebenfalls einen bedeutenden Einfluss auf das Übersetzen im Allgemeinen, da dadurch ein neues Medium für die Übersetzung geschaffen wurde (Quah 2006:64f).

Beispielbasierte und statistische MÜ-Systeme haben deutlich bessere Resultate geliefert als ihre Vorgänger, aber schnelle Entwicklungen in der Informatik und Computerlinguistik haben dazu geführt, dass ab ca. 2015 der Fokus auf neuronale maschinelle Übersetzungssysteme gesetzt wird (Koehn 2020:39). Im folgenden Unterkapitel werden die einzelnen Ansätze genauer beschrieben.

2.2. Arten der maschinellen Übersetzung

Am Anfang wurden regelbasierte maschinelle Übersetzungssysteme erstellt, welche bis zum Ende des 20. Jahrhunderts entwickelt und benutzt wurden. Wie der Name schon andeutet, funktionierten diese Systeme auf dem Prinzip des Kombinierens von grammatikalischen Regeln und großen Wörterbüchern. Die Qualität der Übersetzung kann variieren und hängt von zwei wesentlichen Faktoren ab. Der erste Faktor ist, welches Sprachenpaar verwendet wird. Der zweite Faktor ist, ob es sich um ein allgemeines Übersetzungssystem handelt oder ob es für ein bestimmtes Fachgebiet verwendet wird und dabei Fachterminologie in das System integriert wurde (vgl. Stein 2009:8). Diese Systeme nahmen keine Rücksicht auf Kontext und aus der heutigen Perspektive kann gesagt werden, dass sie sehr primitiv waren, weshalb sie Resultate geliefert haben, die verallgemeinert nicht zufriedenstellend waren (vgl. Quah 2006:60ff).

Prinzipiell existieren zwei verschiedene regelbasierte Ansätze zur MÜ: der direkte Ansatz und der indirekte Ansatz. Der indirekte Ansatz wird weiter in zwei Methoden unterteilt: die Transfermethode und die Interlingua-Methode (vgl. Schäfer 2002).

Im Falle des direkten Ansatzes versucht ein MÜ-System, eine Übersetzung direkt aus einer Ausgangssprache in einer Zielsprache zu erstellen, ohne eine zusätzliche semantische oder syntaktische Analyse durchzuführen. Diese Systeme sind in der Regel wörterbuchbasiert und produzieren eine Wort-für-Wort-Übersetzung. Anschließend wird versucht, durch mehr oder weniger komplexe Regeln die Zielwörter neu zu ordnen, um eine Wortreihenfolge zu erhalten, die den Gegebenheiten der Zielsprache so nahe wie möglich kommt. In diesem Rahmen besteht keine Notwendigkeit, den Ausgangstext zu analysieren und ein einfaches zweisprachiges Wörterbuch wird als ausreichend angesehen. Dieser Ansatz gilt nicht als optimal, da jede Sprache ihre eigenen Besonderheiten hat. Dennoch kann der Ansatz einige grobe Informationen über den Inhalt eines

Textes liefern und kann akzeptabel sein, wenn die beiden betreffenden Sprachen sehr nahe verwandt sind (gleiche Sprachfamilie, ähnliche Syntax, etc.) (vgl. z. B. Poibeau 2017:27f).

Indirekte Systeme sind komplexer als direkte Systeme, da sie eine Art von syntaktischer Analyse einbeziehen. Forscher*innen haben versucht, komplexere Strategien zu entwickeln, um die Struktur der an der Übersetzung beteiligten Sprachen zu berücksichtigen. Bei der Transfermethode ist es das Ziel, die Struktur des Ausgangssatzes, die durch die syntaktische Analyse erhalten wird, auszunutzen und die Wort-für-Wort-Beschränkung der direkten Übersetzung zu vermeiden. Dieses Konzept kann auch auf die semantische Ebene angewendet werden, um die richtige Bedeutung eines Wortes in Abhängigkeit vom Kontext zu wählen. Das Ergebnis soll also idiomatischer sein als bei der direkten regelbasierten Übersetzung, sofern die syntaktische Komponente genaue Informationen über die Ausgangs- und die Zielsprache liefert. In der Praxis stellt diese Methode ein Problem dar, weil es unmöglich ist, alle Verwendungskontexte eines bestimmten Wortes vorherzusagen (vgl. z. B. Poibeau 2017: 27ff).

Die anspruchsvollsten regelbasierten Systeme basieren auf einer Interlingua, die eine mehr oder weniger formale Darstellung des zu übersetzenden Inhalts darstellt. In diesem Zusammenhang versucht das System bei der Übersetzung von Sprache A nach Sprache B zunächst, den Inhalt von A in eine Interlingua zu übertragen, bevor es von der Interlingua in die Zielsprache B übersetzt. Übertragungsregeln beziehen sich immer auf zwei verschiedene Sprachen und müssen daher für jedes neue betrachtete Sprachenpaar angepasst werden. Die Interlingua-Methode soll dieses Problem lösen, indem sie eine sprachunabhängige Repräsentationsebene schafft. Im Vergleich zu Transfersystemen benötigt diese Methode zwar immer noch eine Analysekomponente, um vom Ausgangstext zur interlingualen Repräsentation zu gelangen, aber aus dieser Repräsentation können dann direkt Übersetzungen in mehrere Sprachen entstehen. Diese Art von maschineller Übersetzung ist äußerst komplex und herausfordernd denn das maschinelle „Verstehen“ eines Textes kann potenziell die Darstellung einer Unendlichkeit von ausgedrückten und abgeleiteten Informationen bedeuten, weshalb diese Methode als unerreichbares Ziel betrachtet wird (vgl. z. B. Poibeau 2017:27ff).

Die Systeme, die nach den regelbasierten Systemen entwickelt wurden, haben Resultate erzielt, die als wesentlich zufriedenstellender angesehen werden, da die Herausforderungen, mit denen die maschinelle Übersetzung konfrontiert ist, durch die Analyse enormer Datenmengen, d. h. der Art von Daten, mit denen Computer leichter umgehen als Menschen, genauer angegangen werden können (vgl. z. B. Poibeau 2017: 29f).

Modernere Systeme werden als datengesteuerte („data-driven“) oder korpusbasierte Systeme bezeichnet und basieren auf großen Sammlungen von (hauptsächlich parallelen) Textkorpora, welche

das System zum Vergleichen von früheren Übersetzungen und Erstellung neuer Übersetzungen nutzt. Diese Systeme wurden seit dem Anfang des 21. Jahrhundert genutzt und werden heute weiterhin entwickelt. Quah (2006) definiert drei Hauptformen von korpusbasierten Systemen: beispielbasierte, statistische und neuronale Systeme. Im Gegensatz zu den regelbasierten Ansätzen wurden bei den korpusbasierten Ansätzen alignierte Texte – Paare von Ausgangs- und Zielsprachentexten – verwendet, was bedeutet, dass im Normalfall die Ausgangs- und Zielsprachentexte strukturell auf Satzebene abgeglichen werden.

Die Idee hinter der beispielbasierten maschinellen Übersetzung ist, dass es praktischer wäre, direkt Übersetzungsfragmente zu verwenden, die in bestehenden zweisprachigen Korpora zu finden sind, statt zu versuchen, neue Wörterbücher und Analyse-/Transferregeln zwischen zwei Sprachen zu entwickeln. Die beispielbasierte maschinelle Übersetzung erfolgt in der Regel in drei Schritten, um einen bestimmten Satz zu übersetzen. Zunächst versucht das System, Fragmente des zu übersetzenden Satzes in den für die Ausgangssprache verfügbaren Korpora zu finden. Alle relevanten Fragmente werden gesammelt und gespeichert. Anschließend durchsucht das System die vorgegebenen Übersetzungen nach Übersetzungsäquivalenzen in der Zielsprache. Nachdem die relevanten Fragmente in der Zielsprache gesammelt wurden, wird anhand einer Reihe von Regeln oder statistischen Indizes versucht, aus den identifizierten Fragmenten einen vollständigen Satz zu erstellen. Dies ist eine schwierige Aufgabe, da diese Fragmente in der Regel unvollständig sind, sich überschneiden und nicht autonomen syntaktischen Phrasen entsprechen. Es wurde versucht, Systeme zu entwickeln, die nur relevante syntaktische Phrasen verwenden, z. B. vollständige Substantivphrasen oder Verbphrasen, aber dies führt aus verschiedenen Gründen nicht zu einer Verbesserung, hauptsächlich wegen der geringen Datenmenge. Es ist generell sehr schwierig, genügend relevante Beispiele auf Phrasenebene zu sammeln. Im Allgemeinen bietet die beispielbasierte maschinelle Übersetzung eine akzeptable Abdeckung des zu übersetzenden Textes, insbesondere in Fällen, in denen sich eine Sprache für diese Methode gut eignet, z. B. Japanisch. Trotz einiger Schwächen dieser Methode hat die Kombination der beispielbasierten Übersetzung mit einer statistischen Analyse von sehr großen Korpora zu vielversprechenden Ergebnissen geführt (vgl. z. B. Poibeau 2017:80ff).

Statistische maschinelle Übersetzungssysteme (SMÜ-Systeme) basieren auf Ideen aus der Kommunikationstheorie. Während die beispielbasierten MÜ-Systeme einen Übersetzungsspeicher mit häufig wiederkehrenden Sätzen oder Phrasen mit ihren jeweiligen Übersetzungen speichern und diese dann mittels Algorithmen für die Erstellung von Übersetzungen nutzen (vgl. Quah 2006:81), nutzen statistische Übersetzungssysteme Textkorpora in Verbindung mit

Wahrscheinlichkeitsprinzipien, um eine Übersetzung zu generieren (vgl. Quah 2006:80). Der statistische Ansatz, der auf der Alignierung auf Wortebene basiert, führte zu bemerkenswerten Fortschritten bei der maschinellen Übersetzung. Dieser Ansatz berücksichtigt die statistische Natur der Sprache, was bedeutet, dass sich der Ansatz auf die häufigsten Muster in einer Sprache konzentriert. Diese Alignierung auf Wortebene ist eine wesentlich komplexere Aufgabe als das Alignierung auf Satzebene, da es auf der Satzebene oft eine eins-zu-eins Zuordnungsmöglichkeit zwischen dem Ausgangstext und seiner Übersetzung gibt (ein Satz aus dem Ausgangstext entspricht im Allgemeinen einem Satz in der Übersetzung). Dies gilt jedoch nicht unbedingt auf Wortebene, da sich Sprachen in Bezug auf Grammatik und Syntax erheblich unterscheiden und zudem viele Wörter nicht direkt übersetzt werden können. In diesem Sinne werden die meisten Korrespondenzen als „asymmetrisch“ betrachtet. Trotz dieser Herausforderungen und anderer Einschränkungen ist der statistische Ansatz in der Lage, akzeptable Übersetzungen für eine beträchtliche Anzahl von einfachen Sätzen zu erzeugen. In bestimmten Fällen können statistische Modelle dank asymmetrischer Alignments (ein Wort aus der Ausgangssprache aligniert mit mehreren Wörtern aus der Zielsprache) auch Idiome identifizieren (vgl. z. B. Poibeau 2017:86ff).

Da sich diese Arbeit hauptsächlich mit neuronaler maschineller Übersetzung beschäftigt, wird diesem Ansatz ein eigenes Unterkapitel gewidmet.

2.3. Neuronale maschinelle Übersetzung

In relativ kurzer Zeit wurde die neuronale maschinelle Übersetzung zur wichtigsten technologischen Neuerung der maschinellen Übersetzung. Im Jahr 2015 wurde bei dem *Workshop on Machine Translation*, einem internationalen Wettbewerb für maschinelle Übersetzungssysteme, nur ein einziges rein neuronales maschinelles Übersetzungssystem vorgestellt. Es war konkurrenzfähig, aber unterlegen gegenüber traditionellen statistischen Systemen. Ein Jahr später gewann ein neuronales maschinelles Übersetzungssystem in fast jedem Sprachpaar. Im Jahr 2017 waren fast alle Einreichungen neuronale maschinelle Übersetzungssysteme (Koehn 2020:40).

Im Unterschied zu früheren Ansätzen basieren die Ansätze neuronaler MÜ auf neuronalen Netzen. Ein neuronales Netzwerk ist eine maschinelle Lerntechnik, die eine Reihe von Eingaben annimmt und Ausgaben vorhersagt (vgl. Koehn 2020:67). Das Design neuronaler Netze wurde durch Neuronen inspiriert. Von dieser aus der Biologie stammenden Inspiration übernehmen neuronale Netze die Idee der Verknüpfung von Eingängen (durch eine gewichtete Summe), einer Aktivierungsfunktion und einem Ausgangswert. Die genaue Funktionsweise unterschiedlicher neuronaler Netze unterscheidet sich jedoch stark von natürlichen Neuronen (vgl. Koehn 2020: 31).

Die Begriffe neuronale Netze und Deep Learning werden oft als Synonyme verwendet, wobei es bei beiden angegebenen Begriffen keine vollkommene Überschneidung gibt (vgl. Koehn 2020:31). Unabhängig von den genauen Begriffen ist das Ziel der NMÜ immer gleich: computergesteuerte Übersetzungssysteme sollen große Datenmengen verarbeiten können, aus diesen Datenmengen lernen und letztendlich hochwertige Übersetzungen produzieren können. Bei der Erstellung einer Übersetzung gehören zu den Herausforderungen der NMÜ die Nichtübereinstimmung von Domänen, die Menge der verfügbaren Trainingsdaten, selten verwendete Wörter, verrauschte Daten und die genaue Auswahl von zu verwendenden Architekturen und Trainingsparametern (vgl. Koehn 2020:293ff).

2.4. Frei Verfügbare NMÜ-Systeme

Heute existieren einige MÜ-Systeme, welche der Öffentlichkeit kostenlos über das Internet zugänglich sind. Die Verfügbarkeit von frei verfügbaren MÜ-Systemen im Internet seit etwa 1994 wird als eine der wichtigsten Entwicklungen in der Geschichte der maschinellen Übersetzung angesehen. Diese Systeme haben die Wahrnehmung der breiten Öffentlichkeit tiefgreifend beeinflusst und auch die Entwicklung der Technologie im Bereich der maschinellen Übersetzung geprägt (vgl. Gaspari et al. 2006). Mittlerweile existieren mehrere frei verfügbare MÜ-Systeme, welche auf der Basis von neuronalen Netzen funktionieren. Einige bekanntere von denen sollen in diesem Kapitel beispielhaft vorgestellt werden.

Das 1968 von Dr. Peter Toma gegründete Unternehmen SYSTRAN ist eines der ältesten Unternehmen für maschinelle Übersetzung (vgl. SYSTRAN 2013). SYSTRAN, das aus der Forschung an der Universität Georgetown hervorging, hat umfangreiche Arbeiten für den US-Geheimdienst und die Luftwaffe durchgeführt, Pionierarbeit für verschiedene zukunftsweisende Technologien der maschinellen Übersetzung geleistet (vgl. SYSTRAN 2018) und die Technologie für die kostenlos nutzbaren maschinellen Übersetzungswebsites Yahoo! Babel Fish und Google Translate bereitgestellt (vgl. Chitu 2007). In der Vergangenheit verwendeten SYSTRAN-Systeme regelbasierte maschinelle Übersetzung und statistische maschinelle Übersetzungstechnologie, die die erste ihrer Art auf dem Markt war (vgl. SYSTRAN 2018). Gegenwärtig verwendet SYSTRAN Translator neuronale Netze zum Übersetzen und die kostenlose Version von SYSTRAN Translator existiert in Form einer Website, ähnlich wie andere maschinelle Übersetzungsdienste (vgl. Brown 2016).

Google Translate, auf Deutsch auch bekannt als Google Übersetzer, ist ein Online NMÜ-System von Google LLC, welches Wörter, Texte und ganze Webseiten maschinell übersetzen kann.

Der Dienst wurde im Jahr 2006 eingeführt (vgl. Och 2006). Google Translate besteht aus mehreren verzahnt arbeitenden neuronalen Netzen (vgl. Merkert 2016) und ist momentan das meistgenutzte Online NMÜ-System. Google Translate bietet Nutzer*innen mehrere Methoden an, das System zu verbessern, z. B. indem sie inkorrekte Übersetzungen ändern, bessere Übersetzungen liefern und speichern oder bereits existierende Übersetzungen auswählen (vgl. Google Translate 2020). Die Resultate, welche Google Translate liefert, generell als zufriedenstellend beschrieben (vgl. z. B. LingoHub 2020).

DeepL ist ein Online NMÜ-System der DeepL GmbH, das im Jahr 2017 online gestellt wurde. DeepL verwendet neuronale Netze, die mit der Datenbank von Linguee trainiert wurden (vgl. Coldewey & Lardinois 2017). Linguee ist ein mehrsprachiges Online-Wörterbuch, das von der gleichen Firma betrieben wird (vgl. LingoHub 2020). DeepL unterscheidet sich von anderen Systemen daran, dass es den Nutzer*innen bietet, ihre Übersetzungen schnell zu bearbeiten; Nutzer*innen können auf die Übersetzung draufklicken und ein Wort oder die ganze Satzstruktur mit wenigen Klicks umändern. DeepL lernt aus dem Wörterbuch und den darin enthaltenen, hochwertigen, alignierten Textpaaren. Ähnlich wie Google Translate liefert auch DeepL Resultate, die im Allgemeinen als sehr zufriedenstellend bezeichnet werden (vgl. LingoHub 2020).

Baidu, die größte chinesische Suchmaschine, hat im Juni 2011 ein eigenes frei verfügbares Online MÜ-System zur Verfügung gestellt. Ähnlich wie andere Systeme, welche sich auf neuronale Netze stützen, hat Baidu Translate anfangs regelbasierte, statistische und beispielbasierte Methoden verwendet. Seit 2015 wird ein NMÜ-System verwendet. Baidu Translate unterstützt eine große Anzahl von Sprachen und liefert in erster Linie hochwertige Übersetzungen ins Chinesische bzw. aus dem Chinesischen (vgl. Smartcat 2021).

2.5. Übersetzungsqualität

Die Bewertung der Qualität von Übersetzungen, vor allem Fachübersetzungen, ist in mehreren Situationen erforderlich, u. a. im Zuge des Fremdsprachenerwerbs und der Übersetzer*innenausbildung, bei der Beurteilung von Bewerber*innen für eine Position als Übersetzer*in in einem Unternehmen, während der Abwicklung kommerzieller Übersetzungsaufträge, seitens der Auftraggeber*innen oder der Bedarfsträger*innen einer Übersetzung, und Gutachtertätigkeit in Rechtsstreitigkeiten, deren Gegenstand translatorische Leistungen sind (vgl. Peter A. Schmitt 1997). Daher ist es nachvollziehbar, dass das Bewerten von Übersetzungen für die Translationswissenschaft von großer Bedeutung ist und dass dieses Thema in vielen Werken behandelt wurde (vgl. z. B. Schäffner 1998).

Bei der Bewertung einer Übersetzung kommt es zu ähnlichen Problemen wie bei jeder Kritik eines Textes. Bislang war es nicht möglich, völlig objektive Kriterien für gute und schlechte Übersetzungen zu definieren und ein einheitliches Bewertungssystem zu entwickeln. Yu (2014) stellt fest, dass beim Rezipieren einer Übersetzung die Kultur der Rezipient*innen eine wesentliche Rolle spielt und beschreibt Kultur als eine Reihe von Werten und Normen, die von einer sozialen Gruppe geteilt werden. Diese Werte und Normen, und daher auch Kulturen in ihrer Gesamtheit, sind von Natur aus subjektiv (vgl. z. B. González Rey 2016). Spencer-Oatey (2008) vertritt die Sichtweise, dass die Kultur der Rezipient*innen ihr Interpretationsvermögen beeinflussen. Daher folgt, dass die Interpretation eines Textes an das Subjektive gebunden ist. Als die wichtigsten Faktoren, die die Interpretation eines Textes beeinflussen, nennt Lauscher (2006) u. a. die Weltanschauung der Rezipient*innen, ihren Bildungsgrad und das Niveau ihrer Fachkenntnisse zu einem bestimmten Thema. Daher muss bei der Bewertung von Übersetzungen Rücksicht auf unterschiedliche kulturelle und soziale Faktoren genommen werden, denn es sind gerade diese Faktoren, die darauf Einfluss haben, wie eine Übersetzung in einer Zielkultur rezipiert wird.

Lauscher (2006) beschreibt Bewertungssituationen, welche von den oben genannten Faktoren beeinflusst werden. Genauer nennt Lauscher (2006:58) das bewertende Subjekt, den Bewertungsgegenstand, Bewertungsrezipient*innen, das Bewertungsziel und den Bewertungsmaßstab. Lauscher erklärt, dass die Translationsqualität als ein Konsens zu betrachten ist, und zwar ein Konsens darüber, was eine Übersetzung überhaupt ist und welche Erwartungen sie erfüllen muss, wobei die oben angeführten Faktoren eine wichtige Rolle spielen (Lauscher 2006: 59ff.).

Ein wichtiger Faktor für die Messung der Qualität von Übersetzungen ist die Tatsache, dass Übersetzungen immer eine Funktion erfüllen. Dieses Primat des Zwecks ist die Basis der Skopostheorie:

Eine Handlung bezweckt die Erreichung eines Zieles und damit die Änderung eines bestehenden Zustandes. Die Motivation für eine Handlung besteht darin, daß das angestrebte Ziel höher eingeschätzt wird als der bestehende Zustand. (Reiss & Vermeer 1984: 103)

So stellt neben der Äquivalenz, welche ein zentraler Aspekt linguistischer Übersetzungstheorien ist (z. B. Mounin 1967; Nida 1964; House 1997), auch die Adäquatheit ein unumgängliches Kriterium bei der Auswertung von Übersetzungsqualität dar (vgl. z. B. Montella & Monti 2015). Ähnliche Sichtweisen werden von Nord (2016) präsentiert und das Konzept der Funktionsgerechtigkeit wird zum Zentralpunkt der Übersetzerischen Praxis. Laut Kadrić et al. (2009) haben u. a. folgende Faktoren einen Einfluss auf die Qualität einer Übersetzung: das Thema und die Länge des Textes, der

Liefertermin, die Form der Textübertragung und die Intentionen der Auftraggeber*innen, der Autor*innen und des Zielpublikums. Diese Faktoren sind auch für die MÜ relevant, weil MÜ schneller und günstiger sein muss als die Humanübersetzung, jedoch muss sie gleichzeitig immer noch zufriedenstellende Resultate liefern.

Die Antwort auf die Frage nach der Übersetzungsqualität maschinell erstellter Übersetzungen hängt auch davon ab, wer sie beantwortet. Die Qualität von MÜ wird von unterschiedlichen Akteur*innen bewertet: von Informatiker*innen und Systementwickler*innen, von menschlichen Übersetzer*innen und Sprachwissenschaftler*innen, aber auch von potenziellen Anwender*innen und Zieltextempfänger*innen (vgl. Hutchins & Sommers 1992). Für die Letzteren sind Kosten, Zeit und Qualität bezogene Faktoren von Bedeutung. Wie schon angesprochen, sind die Erwartungen der Rezipient*innen hoch: maschinelle Übersetzung soll schneller und günstiger sein als eine Humanübersetzung, dabei soll der Qualitätsverlust nicht beträchtlich sein. Wie gut die maschinell erstellte Übersetzung sein muss und wie viel Qualitätsverlust toleriert werden kann, hängt von dem Verwendungszweck ab (vgl. Hutchins & Sommers 1992). Für die Rezipient*innen der maschinellen Übersetzung sind die wichtigsten Kriterien einige jener, welche Van Slype (1979) als relevant für die Makroevaluierung von Übersetzungen aufzählt, und zwar: kognitive Kriterien (Verständlichkeit, Inhaltstreue, Kohärenz, Brauchbarkeit und Akzeptabilität), ökonomische Kriterien (Lesegeschwindigkeit, Korrekturgeschwindigkeit, Übersetzungsgeschwindigkeit, Übersetzungsprozess) und linguistische Kriterien (semantische, lexikalische, syntaktische und morphologische Eigenschaften der Übersetzung). Kurzgefasst, hängt der Qualitätsgrad einer Übersetzung aus der Perspektive der Rezipient*innen hauptsächlich davon ab, ob sie ohne größere Probleme den Zieltext rezipieren und gegebenenfalls bearbeiten können.

Was als gut genug gilt, hängt von der jeweiligen Aufgabenstellung ab. Eine in der Praxis prävalente Betrachtungsweise ist, dass Qualität das ist, was der/die Kund*innen als Qualität betrachten. Letztendlich wird der Erfolg der maschinellen Übersetzung daran festgemacht, ob sie hilft, eine bestimmte Aufgabe zu erfüllen und Menschen, die in diesem Aufgabenbereich arbeiten, werden als die besten Beurteiler*innen angesehen. In diesem Sinne ist die maschinelle Übersetzung in der realen Welt kein Endziel, sondern ein Verarbeitungsschritt in einer größeren Aufgabenstellung, wie z. B. Problemlösung oder Informationsbeschaffung (vgl. Koehn 2020: 42).

Um die höchste Übersetzungsqualität zu sichern, werden bei der maschinellen Übersetzung oft zwei Methoden eingesetzt, welche die Qualität steigern sollen. Diese Methoden sind Vorbearbeitung (Pre-Editing) und Nachbearbeitung (Post-Editing). Im Zusammenhang mit der maschinellen Übersetzung beinhaltet Vorbearbeitung die Verwendung einer Reihe von terminologischen und stilistischen

Richtlinien oder Regeln zur Vorbereitung des Originaltextes, um die Qualität der maschinellen Übersetzung zu verbessern. Personen, welche Vorbereitung durchführen, befolgen im Normalfall bestimmte Regeln, um typographische Fehler zu entfernen und mögliche inhaltliche Fehler zu korrigieren, um kürzere Sätze zu schreiben, bestimmte grammatikalische Strukturen oder semantische Entscheidungen zu verwenden und um bestimmte Begriffe zu markieren, die möglicherweise nicht übersetzt werden sollen (vgl. Guerberof Arenas, 2020:334). Im Kontext der maschinellen Übersetzung kann Nachbearbeitung als die Bearbeitung, Modifizierung oder Korrektur einer von einem MÜ-System erstellten Übersetzung definiert werden. Im Allgemeinen besteht das Ziel darin, eine maschinelle Übersetzung so zu bearbeiten, dass die nachbearbeitete Übersetzung in Bezug auf Stil, Grammatik, Rechtschreibung und Terminologie mit von menschlichen Übersetzer*innen erstellten Übersetzungen vergleichbar ist (vgl. Guerberof Arenas, 2020:334ff).

2.6. Evaluierung der Übersetzungsqualität von MÜ-Systemen

Im Bereich der maschinellen Übersetzung werden verschiedene Ansätze zur Evaluierung der Übersetzungsqualität eingesetzt. Diese können grob in automatische und manuelle Ansätze unterteilt werden (vgl. Popović 2018).

Automatische Vorgänge werden hauptsächlich für die Evaluierung maschineller Übersetzungen eingesetzt und können Zeit und Kosten ersparen (vgl. Popović 2018:142). Eine Übersetzung wird automatisch mit früheren Übersetzungen verglichen (vgl. Koehn 2020:53). Die meisten in der Praxis verwendeten Bewertungsmetriken für maschinelle Übersetzungen basieren auf dem Vergleich der MÜ-Systemausgabe mit menschlichen, als hochwertig eingestuften Referenzübersetzungen. Bei den meisten Systemen zur automatischen Evaluierung wird verglichen, wie viele der Wörter oder Wortgruppen in einer maschinell erstellten Übersetzung auch in einer Referenzübersetzung oder mehreren Referenzübersetzungen vorhanden sind. Gleichzeitig wird auch verglichen, ob die Reihenfolge der Wörter in der maschinell erstellten Übersetzung ebenfalls in der Referenzübersetzung vorzufinden ist. Um genauere Bewertungen zu erstellen, verwenden diese Evaluierungssysteme auch weitere Kriterien. So können beispielsweise Übersetzungen als zu kurz und daher als fehlerhaft eingestuft werden, um Auslassungen zu bestrafen. Evaluierungssysteme können auch eingestellt werden, um eine Übersetzung besser oder schlechter zu bewerten, abhängig davon, ob in der Übersetzung Wörter vorkommen, die Synonyme für in der Referenzübersetzung vorkommenden Wörter sind, anstatt dass das gleiche Wort verwendet wurde.

Ein Beispiel für ein solches System ist die Metrik Bilingual Evaluation Understudy (BLEU) nach Papineni et al. (2011). Es handelt sich um eine Metrik zur Bewertung der Qualität von

maschinell übersetzten Texten, wobei Qualität als die Übereinstimmung zwischen der Übersetzung eines MÜ-Systems und einer menschlich erstellten Übersetzung angesehen wird. Bewertungen werden für einzelne übersetzte Segmente berechnet, indem sie mit einer Reihe von qualitativ als hochwertig eingestuften Referenzübersetzungen verglichen werden. Diese Bewertungen werden dann über das gesamte Korpus gemittelt, um eine Einschätzung der Gesamtqualität der Übersetzung zu erhalten. Allgemeine Verständlichkeit oder grammatikalische Korrektheit werden jedoch nicht berücksichtigt (Papineni et al. 2001).

Metric for Evaluation of Translation with Explicit Ordering (METEOR) bewertet maschinelle Übersetzungshypothesen, indem es sie mit einer oder mehreren Referenzübersetzungen aligniert. Alignments basieren auf exakten, Stamm-, Synonym- und Paraphrasen-Matches zwischen Wörtern und Sätzen (vgl. CMU 2011). Meteor wurde entwickelt, um mehrere beobachtete Schwächen von BLEU zu beheben. So ist METEOR beispielsweise eine abstimmbare, Alignment-orientierte Metrik, während BLEU im Allgemeinen eine präzisionsorientierte Metrik ist. Im Gegensatz zu BLEU, die nur die Präzision berechnet, berechnet METEOR sowohl die Präzision als auch die Trefferquote (engl. Recall) und kombiniert beide, um das parametrisierte harmonische Mittel zu berechnen. Im Gegensatz zu BLEU werden bei METEOR längere Antworten nicht bestraft und es wird ein gewisses Maß an linguistischem Wissen in Form von Wortstamm- und Synonym-Matching berücksichtigt, was ermöglicht, Äquivalenzen zwischen der MÜ-Ausgabe und der Referenzübersetzung zu identifizieren. Im Gegensatz zu BLEU verwendet METEOR nicht direkt mehrere Referenztexte, sondern wählt bei mehreren Referenzen die Referenzübersetzung für jedes Segment aus, die die beste METEOR-Bewertung ergibt. Die Fähigkeit von METEOR, Variabilität durch Stemming und Synonyme zu verarbeiten, reduziert jedoch den erwarteten Gewinn aus dem gleichzeitigen Vergleich mit mehreren Referenzen und mildert damit die Auswirkungen dieser Einschränkung (Dorr et al. 2011:762f).

Eine weitere Möglichkeit, die Qualität maschineller Übersetzungen zu messen, ist eine Gruppe einigermaßen verwandter Methoden namens Word Error Rate (WER), Translation Edit Rate (TER) und Human-mediated Translation Error Rate (HTER). WER war eine der ersten automatischen Metriken, die zur Bewertung von MÜ-Systemen verwendet wurden. WER vergleicht das Produkt eines maschinellen Übersetzungssystems mit einer Referenzübersetzung und markiert alignierte Wörter aus der maschinellen Übersetzung und dem Referenztext als Matches; Wörter, für die keine Matches gefunden wurden, werden als Substitutionen, Löschungen oder Einfügungen gekennzeichnet, und WER ist die Summe dieser nicht übereinstimmenden Wörter geteilt durch die Anzahl der Wörter in der Referenzübersetzung (vgl. Dorr et al. 2011:760).

TER befasst sich mit dem Fehler der Satzumstellung bei WER, indem es die Verschiebung von Wörtern in Blöcken zulässt (vgl. Dorr et al. 2011:765), was zu einer Metrik führt, die den Bearbeitungsaufwand misst, den ein Mensch durchführen müsste, um eine maschinelle Übersetzung so zu ändern, dass sie genau mit der Referenzübersetzung übereinstimmt (vgl. Dorr et al. 2006).

HTER ist ein komplexeres und halbautomatisches Maß, bei dem Menschen die Übersetzungen nicht direkt bewerten, sondern eine neue Referenzübersetzung erstellen, die näher an der maschinellen Übersetzung liegt, aber den Sprachfluss und die Bedeutung der ursprünglichen Referenz beibehält. Diese neue Zielreferenz wird dann als Referenzübersetzung verwendet, wenn die MÜ-Ausgabe mit anderen Metriken wie TER, BLEU oder METEOR bewertet wird (vgl. Dorr et al 2009).

Ein potenzieller Nachteil automatisierter Ansätze ist es, dass im Normalfall mehrere Übersetzer*innen sehr unterschiedliche Übersetzungen erstellen können, die jedoch einen sehr ähnlichen Qualitätsgrad haben. Eine Lösung für dieses Problem ist, nicht nur eine, sondern mehrere Referenzübersetzungen zu verwenden, jedoch wird dieser Ansatz heute selten verwendet (vgl. Koehn 2020:55).

Ein weiterer Nachteil automatisierter Ansätze ist, dass es oft unklar ist, wieso eine konkrete Übersetzung eine konkrete Bewertung bekommen hat, da die Bewertung von vielen Faktoren abhängt, z. B. von der Anzahl der Referenzübersetzungen, dem Sprachenpaar, dem Fachbereich und sogar von dem Tokenisierungsschema, welches verwendet wird, um die maschinell erstellte Übersetzung und die Referenzübersetzung in Wörter/Wortgruppen aufzubrechen. Letztlich wird auch kritisiert, dass automatisierte Evaluierungssysteme menschliche Übersetzungen oft schlechter einstufen, als sie es eigentlich sind, vor allem im Vergleich zu MÜs (vgl. Koehn 2020: 53ff).

Trotz dieser Nachteile der automatischen Evaluierung werden automatisierte Evaluierungssysteme weiterhin verwendet, nicht zuletzt auch deshalb, weil in der Praxis gezeigt wurde, dass MÜs, die automatisch als schlecht bewertet werden, auch von Evaluatord*innen als schlecht eingestuft werden, und umgekehrt (vgl. Koehn 2020:60).

Ein Versuch, die Nachteile der automatischen Evaluierung von MÜ zu umgehen, ist der Einsatz von Systemen zur Qualitätseinschätzung maschineller Übersetzungen (engl. Quality Estimation for Machine Translation, abgekürzt MTQE). Im Unterschied zur automatischen Evaluierung ist der Leitgedanke hinter der Qualitätseinschätzung, auf menschlichen Rückmeldungen basierende Modelle zu erstellen, um die Qualität der von MÜ-Systemen erstellten Übersetzungen auch ohne Referenztexte vorherzusagen. Mit anderen Worten ist es das Ziel, auch ohne Zugang zu einwandfreien Referenzübersetzungen eine Einschätzung darüber zu geben, wie gut oder zuverlässig

die von einem MÜ-System produzierten Texte sein werden. Des Weiteren zielt dieser Ansatz darauf ab, die Leistung eines Systems anhand einzelner Datenpunkte abzuschätzen, anstatt nur die Gesamtleistung des Systems. Zwei Systeme, die ein Beispiel für MTQE darstellen, sind DeepQuest und MARMOT (Paetzold et al. 2018).

DeepQuest kann verwendet werden, um QE-Modelle entweder auf Wort-, Satz- oder Dokumentenebene zu trainieren. Dieses System verwendet bereits etablierte Ansätze zur Qualitätseinschätzung. DeepQuest benötigt außer den QE-Trainingsdaten keine zusätzlichen Ressourcen, aber parallele Korpora können für bessere Ergebnisse verwendet werden (vgl. Blain et al. 2018).

MARMOT wurde für die Merkmalsextraktion, das Modelltraining und die Vorhersage mit Schwerpunkt auf QE auf Wort- und Phrasenebene entwickelt. Es kann Merkmale in verschiedenen Formaten ausgeben, die für andere maschinelle Lernwerkzeuge geeignet sind. MARMOT wurde als Basissystem für die Aufgaben auf Wort- und Phrasenebene beim *Workshop on Machine Translation* verwendet (vgl. qe-team 2017).

Die manuelle Evaluation wird für die Evaluierung von menschlichen und maschinellen Übersetzungen genutzt. Manuelle Evaluierungsmethoden sind weiterhin die qualitativ hochwertigsten, denn alle halb- oder vollautomatischen Systeme haben dieselben Probleme, wie MÜ-Systeme an sich. Bei der manuellen existieren unterschiedliche Bewertungsmethoden und die Evaluation kann entweder mit oder ohne Vorlage von Ausgangstexten oder Referenzübersetzungen erfolgen.

Einzelne Sätze oder ganze Texte können als ein Ganzes bewertet werden, indem Evaluator*innen entscheiden sollen, ob die Gesamtqualität einer Übersetzung zufriedenstellend ist oder nicht. Einerseits kann die Evaluierung von bilingualen Personen durchgeführt, die sowohl die Ausgangssprache als auch die Zielsprache beherrschen. Typischerweise werden qualifizierten Evaluator*innen, die die Ausgangs- und Zielsprache beherrschen, mehrere Sätze zusammen mit den maschinellen Übersetzungen gezeigt. Evaluator*innen werden dann gebeten, zu bewerten, ob die Übersetzungen korrekt sind (vgl. Koehn 2020:45). Andererseits können auch Personen eingesetzt werden, die nur die Zielsprache beherrschen. Bei dieser Methode wird im Normalfall eine Referenzübersetzung genutzt. Die Evaluator*innen werden gebeten, zu bewerten, ob die MÜ und die Referenzübersetzung übereinstimmen (vgl. Popović 2018:130f).

Neben der Auswertung der Gesamtqualität ist es auch möglich, Übersetzungen in Hinblick auf bestimmte Qualitätsparameter auszuwerten. Zwei sehr oft verwendete Parameter sind Geläufigkeit und Adäquatheit (Inhaltstreue). Bei dieser Methode wird den Evaluator*innen eine

Übersetzung präsentiert, woraufhin sie auf einer Skala eine Qualitätsbewertung abgeben. Im Gegensatz zur Gesamtbewertung werden die einzelnen Parameter bei dieser Methode einzeln ausgewertet. Auch bei dieser Methode können sowohl bilinguale Evaluator*innen in Kombination mit einem Ausgangstext als auch monolinguale Evaluator*innen in Kombination mit Referenzübersetzungen eingesetzt werden (vgl. Popović 2018:130f). Diese Auswertung in Hinblick auf Qualitätsparameter kann auf zwei Weisen eingesetzt werden. Einerseits kann ein einzelnes System bewertet werden. Evaluator*innen werden gebeten, die Adäquatheit und die Geläufigkeit einer MÜ auf einer Skala zu bewerten. Andererseits ist es möglich, zwei oder mehrere Systeme miteinander zu vergleichen. Hier werden Evaluator*innen gebeten, um von zwei oder mehreren MÜs die beste und/oder die schlechteste in Hinblick auf Adäquatheit und Geläufigkeit auszuwählen (Koehn 2020:45ff).

Weiterhin ist es möglich, nicht nur zu bewerten, ob oder wie gut bzw. schlecht eine Übersetzung ist, sondern im Zieltext nach konkreten Fehlern zu suchen und diese Fehler zu kategorisieren. Im Falle dieser Fehlerklassifikation nach einer Typologie suchen Evaluator*innen nach verschiedenen Fehlern im Zieltext, welche danach nach unterschiedlichen Kriterien kategorisiert werden. Diese Kategorien beinhalten u. A.: morphologische Fehler, lexikalische Fehler, orthographische Fehler, syntaktische Fehler und semantische Fehler (vgl. Popović 2018:130ff.).

Neben den oben beschriebenen Ansätzen beschreibt Koehn (2020:41ff) auch die aufgabenbezogene Evaluierung. Diese Art der Evaluierung basiert darauf, ob ein MÜ-System Benutzer*innen bei der Lösung von unterschiedlichen Aufgaben behilflich sein kann. Als mögliche Aufgaben, welche als Metriken für die Qualität eines MÜ-Systems dienen können, nennt Koehn u. a. auch Informationsauffindung, grundsätzliches Verständnis fremdsprachiger Inhalte, das Lösen unterschiedlichster technischer Probleme und der Einsatz von maschineller Übersetzung seitens menschlicher Übersetzer*innen zur Steigerung ihrer Produktivität. Bei dieser Art von Evaluierung ist es wichtig eine größere Aufgabenstellung in sehr kleine Teile aufzubrechen und diese Teile als die Metrik zu definieren. Die aufgabenbezogene Evaluierung kann durchaus nützlich sein, aber sie ist gleichzeitig sehr komplex und kostenintensiv.

Weiterhin beschreibt Koehn (2020:51f) auch die Evaluierung nach dem benötigten Zeitaufwand für die Nachbearbeitung maschineller Übersetzungen. Beispielsweise können Auslassungen bei der Übersetzung für die Bedeutung eines Satzes sehr negative Folgen haben, wie etwa die Verneinungen, welche dazu führen, dass ein Satz völlig gegenteilige Bedeutung bekommt. Diese Fehler können aber meistens durch Nachbearbeitung sehr leicht und schnell behoben werden.

Die bisher beschriebenen Methoden können auch verwendet werden, um mehrere MÜ-Systeme miteinander zu vergleichen. Eine weitere Methode für den Vergleich mehrerer Systeme ist das Ranking. Maschinell übersetzte Sätze oder ganze Zieltexte werden im Normalfall auf einer Skala direkt von sehr gut bis schlecht gereiht. In diesem Fall wird unter Qualität normalerweise die Gesamtqualität der Übersetzungen verstanden. Der Vorteil dieser Methode ist der, dass MÜ-Systeme untereinander direkt und ihrer Gesamtheit ausgewertet werden, denn die Evaluators*innen entscheiden lediglich, welche Übersetzung sie bevorzugen (vgl. Dorr 2011:758).

Im nächsten Unterkapitel wird die Methode mit dem Namen Best-Worst-Scaling vorgestellt, welche für den praktischen Teil dieser Masterarbeit ausgewählt wurde.

2.7. Best-Worst-Scaling

Für Zwecke der für diese Masterarbeit erstellten Umfrage wurde die Best-Worst-Scaling Methode für die Bewertung der Übersetzungsqualität ausgewählt. Best-Worst Scaling (BWS) ist eine Datenerhebungsmethode, bei welcher die Befragten aufgefordert werden, die beste bzw. die schlechteste Option aus einer Reihe von Auswahlmöglichkeiten auszuwählen. BWS wird zunehmend verwendet, um mehr Auswahldaten von Personen zu erhalten und/oder um Auswahlprozesse zu verstehen (vgl. Louviere et al. 2015).

Diese Methode hat ihren Anfang im Jahr 1987 (vgl. Louviere et al. 2015:3). Einerseits entstand diese Methode aus dem Gedanken, dass es nützlich wäre, nicht nur herauszufinden, was in einer Befragung die Befragten am meisten bevorzugen, sondern auch was sie als am wenigsten attraktiv sehen. In dieser Methode wurde der Vorteil gesehen, dass die Befragten gezwungen sind, zwischen mehreren Optionen nur eine beste und nur eine schlechteste Option auszuwählen (vgl. Marley 2009). BWS wurde bereits in mehreren Fachgebieten verwendet, z. B. in der Gesundheitspflege, um herauszufinden, was sich Patient*innen von ihren Behandlungen erhoffen (vgl. z. B. Craig et al. 2017 oder Bridges et al. 2016), in der Marketingforschung (vgl. z. B. Gudergan et al. 2013), in der Soziologie, Geographie und Anthropologie (vgl. z. B. Abe et al. 2018).

Zu den Vorteilen von BWS gehört auch die Tatsache, dass diese Methode einige Einschränkungen von Ratingskala-Methoden aufhebt. Dazu gehören Inkonsistenzen in der Bewertung durch verschiedene Personen oder sogar durch dieselbe Person, Voreingenommenheit gegenüber einem Skalenbereich und die Möglichkeit, dass die Befragten durch die Anzahl der zur Auswahl stehenden Skalenabstufungen eingeschränkt oder überfordert sind (Kiritchenko & Mohammad 2017).

2.8. Fachsprache der Musik

Im Fokus dieser Masterarbeit stehen unterschiedliche Texte im Fachbereich Musik und insbesondere jene Texte, die Musiktheorie behandeln. Musiktheorie ist ein breites Themenfeld und beschäftigt sich u.a. mit Überlegungen zu Tonsystemen, Notations- und Stimmsystemen, der Lehre vom Kontrapunkt und Harmonie, Form und Instrumentation, Erklärungsmodellen von Rhythmus und Metrum, Methoden der Analyse, der Diskussion musikalischer Topoi oder Untersuchungen zur musikalischen Erwartung (vgl. Grimm & Wald-Fuhrmann 2017:9).

Diese Themen können in Texten behandelt werden, welche sehr unterschiedlichen Textsorten angehören können und in Form von sehr unterschiedlichen Medien auftreten, z. B. in Fachbüchern, in Fachzeitschriftenartikeln, in Online-Foren und Blogeinträgen. Der Grad an Formalität und Fachlichkeit dieser Texte kann dadurch sehr variieren. Unabhängig von der genauen Textsorte und dem Medium ist es das Ziel dieser Texte, eine Analyse und Interpretation einer Komposition, einer bestimmten Anordnung von Noten, einer Akkordfolge, eines bestimmten Rhythmus oder eines anderen musikbezogenen Phänomens durchzuführen. Die Durchführung dieser Analyse ist herausfordernd, denn eine Theorie der musikalischen Bedeutung kann nur dann erfolgreich sein, wenn sie versucht, das Mysteriöse zu entmystifizieren, ohne es des Mysteriösen zu berauben (vgl. Almén & Pearsall 2006 :2).

Trotz ihrer zahlreichen Unterschiede weisen Texte, die sich mit Musiktheorie oder anderen musikbezogenen Themen befassen, auch einige Gemeinsamkeiten auf. Da viele Texte, die sich mit Musiktheorie befassen, eine gewisse Art der Analyse beinhalten, gibt es viele Perspektiven, aus welchen Autor*innen ein bestimmtes Phänomen betrachten können. Einerseits kann ein ganzes Musikstück analysiert werden, wobei alle formalen Elemente und ihr Beitrag zur Komposition und ihrer Wirkung detailliert beschrieben werden. Andererseits kann auch nur ein bestimmter analytischer Punkt untersucht werden, z. B. die Beziehungen oder Muster, die in einer bestimmten Musiktradition oder einem bestimmten Genre vorhanden sind, Veränderungen der Tonart und der musikalischen Skalen in einem bestimmten Musikstück, theoretische Prinzipien oder historische Aufführungsstandards. In beiden Fällen führen die Autor*innen dieser Texte oft Beispiele an, meist in Form von Passagen aus einer Partitur, um ihre Analyse zu demonstrieren, zu erklären oder zu untermauern (vgl. Hunter RWC 2017).

Um ihre Analyse und Interpretation zu behandelnder musikalischer Phänomene so eindeutig wie möglich vorzustellen und Mehrdeutigkeit zu vermeiden, verwenden Autor*innen

musiktheoretischer bzw. musikwissenschaftlicher Texte die in der Musikwissenschaft vorhandene Fachterminologie. Einige vorzufindende Fachtermini sind:

- Melodie: eine bestimmte Abfolge von Noten;
- Tonumfang: der Abstand zwischen den tiefsten und höchsten Tönen;
- Motiv: ein Muster aus musikalischen Noten, das sich in einer Komposition wiederholt;
- Rhythmus: der Puls der Komposition;
- Takte: metrische Einheiten, die in der Musiknotation durch Linien getrennt sind;
- Metrum: Gruppen von Schlägen in einem wiederkehrenden Muster;
- nicht-metrisch oder unmetrisch: freier Rhythmus ohne erkennbare Zeit;
- Harmonie: eine Reihe von gleichzeitig erklingenden Noten;
- Akkorde: drei Noten, die normalerweise aus übereinanderliegenden Terzen bestehen, die auch Dreiklänge genannt werden;
- Konsonanz: eine harmonische Kombination, die stabil ist;
- Dissonanz: eine harmonische Kombination, die unstabil ist;
- Monophonie: bedeutet wörtlich "ein Klang", bezeichnet eine einzige Melodie, ohne Harmonie oder jegliche Begleitung;
- Homophonie: eine Melodie mit einer harmonischen Begleitung, welche die Melodie unterstützt;
- Polyphonie: zwei oder mehr Stimmen, die gleichzeitig gesungen oder gespielt werden;
- Akkordophone: Saiteninstrument;
- Aerophone: Blasinstrumente, unterteilt in Holz- und Blechblasinstrumente;
- Membranophon: Musikinstrument, das zur Klangerzeugung eine gespannte Membran besitzt (vgl. Hunter RWC 2017).

Da Musik jedoch eine Kunstgattung ist, ist oft unvermeidlich, dass musikbezogene Texte neben Fachtermini auch allgemeinsprachige Ausdrücke beinhalten. Als Beispiel soll ein Ausschnitt der Analyse der Beethovens 7. Symphonie von Engelbach (2017) dienen:

1.2 Sonatenhauptsatz (vivace) (T. 63-450)

Die Exposition beginnt mit vier Takten, die den Rhythmus einleiten und auch den Grundrhythmus des ganzen Satzes vorgeben. In T. 67 erklingt das Hauptthema. Das erste Auftreten des Themas in der Flöte erscheint tänzerisch und leicht und wird von den Streichern rhythmisch akzentuiert. In T. 89 erscheint das Thema erneut, diesmal jedoch im Rahmen eines orchestralen Tuttis in den Violinen, es wirkt hier mächtig und energisch, die Einleitung erfolgt über das rasant aufwärts steigende 16tel Motiv b. Begleitet wird dieser zweite Themeneinsatz durch 16tel Läufe in der 2. Violine und der Viola, die wiederum oft gleichmäßige 16tel-Repetitionen enthalten und die einen Kontrast zu dem ungleichmäßigen Rhythmus des Themas haben. Trompeten und Celli spielen den Grundrhythmus. (Engelbach 2017:6)

Obwohl dieser Ausschnitt hauptsächlich von klar definierten musiktheoretischen Fachtermini geprägt ist, verwendet die Autorin auch subjektive, in der Allgemeinsprache vorzufindende und in der Musiklehre nicht klar definierte Ausdrücke, um subjektive Eindrücke zu beschreiben. An manchen Stellen ist es ohne genaueren Kontext nicht völlig klar, ob ein Ausdruck im fachlichen oder im allgemeinen Sinne verwendet wird; ob mit „ungleichmäßigem Rhythmus“ die Abwesenheit von einem erkennbaren Grundschlag gemeint wird, oder ob sich die Taktart in kurzer Zeit schnell ändert, ist nicht vollkommen klar.

Viele Texte, die Musik zum Thema haben, sind in erster Linie didaktischer Natur. Dies hat einen Einfluss darauf, wie sie verfasst werden. Diese Texte erfüllen die Rolle einer künstlerischen Wissensvermittlung und werden oft an Musikschulen und Konservatorien verwendet, können aber auch zur Wissensvermittlung an Musiker*innen ohne formale Musikausbildung eingesetzt werden, z. B. in Form von Blogeinträgen. Die Eigenschaften musiktheoretischer und musikwissenschaftlicher Texte können daher auf Eigenschaften der Musiklehre zurückgeführt werden.

Fontelles Rodríguez (2021) beschreibt einige wichtige Eigenschaften der Musiklehre:

- Die Musiklehre ist Teil einer nicht-obligatorischen Ausbildung;
- Sie beinhaltet einen selektiven Unterricht, d. h. sie ist im Gegensatz zur allgemeinen Bildung nicht für eine breite Öffentlichkeit konzipiert;
- Das Ziel ist die Ausbildung von Amateur- oder Berufsmusiker*innen;
- Die Auszubildenden sind motiviert, sie mögen Musik und ihr Studium;
- Die Bandbreite der Begabungen und Einstellungsvielfalt wird als gering eingeschätzt;
- Sie beinhaltet ein didaktisches Format mit kleinen Gruppen oder individualisiertem Unterricht;
- Es existieren spezifische Submodalitäten für jeden der musikalischen Ausbildungsbereiche, im Gegensatz zum Universalwissen in der Allgemeinbildung.

Die übersetzerische Tätigkeit in Bezug auf das Fachgebiet Musik wurde bislang einigermaßen vernachlässigt und relativ wenig dokumentiert (Susam-Sarajeva 2008). Das, was bereits dokumentiert und worüber geschrieben wurde, waren hauptsächlich Übersetzungen von Song-Texten, Libretti und ähnliche Formen von Texten, die dafür gedacht sind, gesungen zu werden (Susam-Sarajeva 2008). In diesem Sinne hat sich die Übersetzungswissenschaft in Bezug auf Musik bisher hauptsächlich mit einem Phänomen beschäftigt, das der Übersetzung von Poesie gleicht und daher ähnliche Probleme aufweist (vgl. z. B. Raffel 1964). Des Weiteren überschneidet sich die Übersetzung von Musiktexten und Poesie, weil viele Werke der Poesie und Gedichte in Lieder umgewandelt und danach übersetzt werden (vgl. Low 2003). Frith (2004) ist der Ansicht, dass die Schwierigkeit zu erklären, wie Musik kulturell funktioniert, ein Grund dafür ist, dass sie nicht ausreichend ernst genommen wird. Daher ist es nicht verwunderlich, dass die übersetzerische Tätigkeit im Zusammenhang mit dem Fachbereich Musik weiterhin weitgehend undokumentiert bleibt und dass das Fachgebiet Musik im Vergleich zu anderen Fachgebieten seitens der Translationswissenschaft relativ wenig untersucht wurde.

Franzon et al. (2014) präsentieren eine allgemeine Literaturliste zum Thema Translation und Musik, welche um die 50 Einträge enthält, um eine Übersicht darüber zu erstellen, was bisher zum Thema geleistet wurde und zukünftige Forschungsmöglichkeiten zu erkunden. Alle diese Werke befassen sich hauptsächlich mit der Übersetzung oder Adaptierung von Musiktexten, jedoch wird Musikwissenschaft, Musiktheorie und die Terminologie in dem Fachgebiet Musik nach dem Kenntnisstand des Autors dieser Masterarbeit von keinem der genannten Werke genauer behandelt.

Es existieren auch Beiträge zu pragmatischen Problemen, die daraus entstehen, dass Musik und Sprache aufeinander angewiesen sind. Chanan (2013) beschreibt beispielsweise das Problem, dass viele Opern nur als Übersetzungen zu einem Teil des Repertoires eines Opernhauses werden können, denn Opernsänger*innen können nicht in allen Sprachen singen, sondern nur in einer Auswahl von in der Praxis üblichen Sprachen. Andererseits macht es in vielen Fällen keinen Sinn, eine Oper zu übersetzen: als Beispiel nennt Chanan (2013) die glagolitische Messe, weil die Übersetzung dieses Stücks unsinnig wäre, da der Zweck des Stücks die Wiederbelebung einer ausgestorbenen Sprache ist.

Einige Werke, welche sich mit der Übersetzung musikbezogener Texte befassen, werden noch in einem eigenen Unterkapitel innerhalb des Kapitels Aktueller Stand der Forschung vorgestellt.

3. Aktueller Stand der Forschung

3.1. Qualität von NMÜ allgemein und in der Sprachkombination Englisch-Deutsch

Wie bereits erwähnt, liefern NMÜ-Systeme Resultate, die im Allgemeinen als relativ zufriedenstellend bewertet werden, vor allem im Vergleich zu früheren Ansätzen. Für den Vergleich, der für die Zwecke der Masterarbeit durchgeführt wird, werden DeepL und Google Translate verwendet und aus diesem Grund wird in diesem Kapitel der Fokus auf Studien gelegt, die diese zwei Systeme untersucht haben.

Arratibel et al. (2018) haben eine Studie am Beispiel von 3 Systemen durchgeführt: einerseits mit NMÜ-Systemen OpenNMT und Google Translate, und andererseits mit dem SMÜ-System Moses. Der Fokus der Studie lag auf dem Sprachenpaar Englisch-Baskisch, welches ein ressourcenarmes Sprachenpaar darstellt. Zur Evaluierung wurde BLEU eingesetzt. Es wurde festgestellt, dass die statistische maschinelle Übersetzung nach wie vor ein zufriedenstellender Ansatz für ressourcenarme Sprachenpaare sein kann, es zeigte sich aber, dass NMÜ auch bei ressourcenarmen Sprachenpaaren wie Englisch-Baskisch bessere Resultate liefert als SMÜ-Systeme. Google Translate lieferte Resultate, die besser bewertet wurden als die Resultate von OpenNMT. Es ist jedoch zu beachten, dass alle drei im Experiment verwendeten Systeme im Vergleich zu Experimenten mit ressourcenreichen Sprachenpaaren Resultate lieferten, die als schlechter bewertet wurden (vgl. Arratibel et al. 2018).

Bouillon et al. (2018) stellen am Beispiel von DeepL fest, dass *nicht-spezialisierte* NMÜ-Systeme gegebenenfalls sogar besser abschneiden als *spezialisierte* SMÜ-Systeme. Für die Zwecke dieser Studie wurden Übersetzungen aus dem Deutschen ins Französische verglichen, welche von Microsoft Translator Hub (MTH) und DeepL erstellt wurden. Die übersetzten Texte haben sich mit dem Thema Jahresbericht befasst. Die Übersetzungen wurden zweifach evaluiert. Zuerst wurden die Übersetzungen ohne Nachbearbeitung automatisch evaluiert, indem jeder Übersetzung ein BLEU-Wert zugeteilt wurde. Anschließend wurden zwei Übersetzer*innen gebeten, die von den beiden Systemen produzierte Ausgabe (insgesamt 500 Segmente) nachzubearbeiten. Beide waren bei der Nachbearbeitung der von DeepL erzeugten Übersetzung viel schneller und ihre endgültigen Texte erhielten eine bessere HTER-Bewertung. Danach wurden Übersetzungen mit Nachbearbeitung manuell evaluiert. Die Evaluator*innen waren drei Masterstudent*innen des Übersetzens mit Französisch als Erstsprache. Sie wurden gebeten, die nachbearbeiteten Segmente von DeepL und

MTH zu vergleichen und anzugeben, welche der Übersetzungen die beste war, oder ob sie beide Segmente als qualitativ gleichwertig sahen. Die Ergebnisse automatischer und menschlicher Bewertungen zeigen, dass DeepL insgesamt als besser betrachtet wird als MTH, jedoch wird die Qualität von DeepL von dem automatischen Bewertungssystem BLEU unterschätzt (Bouillon et al. 2018).

In Bezug auf das Sprachenpaar Englisch-Deutsch kommen Bentivogli et al. (2016) zu einer ähnlichen Feststellung. Im Bereich der maschinellen Übersetzung hat sich der aktuelle neuronale Ansatz als die erste Technologie herauskristallisiert, die in der Lage ist, die Resultate der phrasenbasierten Ansätze zu übertreffen. In dieser im Jahr 2015 durchgeführten Untersuchung übertraf NMÜ in der Sprachkombination Englisch-Deutsch etablierte phrasenbasierte Systeme auf dem damals neuesten Stand der Technik. Die Resultate wurden 2016 veröffentlicht und es wurde festgestellt, dass NMÜ-Systeme unabhängig von der Satzlänge den Gesamtaufwand für die Nachbearbeitung deutlich verringern im Vergleich zu den besten phrasenbasierten Übersetzungssystemen. Die Analyse wies jedoch auch darauf hin, dass einige Aspekte der NMÜ weiterer Arbeit bedürfen, wie z. B. der Umgang mit langen Sätzen und die Neuordnung bestimmter linguistischer Konstituenten, die ein tiefes semantisches Verständnis des Textes erfordern (Bentivogli et al. 2016).

Lotz und Van Rensburg (2014) haben anhand einer empirischen Studie gezeigt, dass sich das Resultat von Google Translate (in der ehemaligen SMÜ-Form) für das Sprachenpaar Afrikaans-Englisch in der Zeit zwischen den Jahren 2010 und 2012 deutlich verbessert hat. In der Studie wurden die Übersetzungen von Google Translate mit Übersetzungen verglichen, welche von zwei menschlichen Übersetzer*innen verfasst wurden. Die Qualität aller Übersetzungen wurde von Prüfer*innen ausgewertet und die Resultate der Auswertung wurden verglichen, wobei gezeigt wurde, dass sich die Qualität der Übersetzungen, die von Google Translate verfasst werden, mit der Zeit steigert. Es wurde gezeigt, dass sich die Qualität maschineller Übersetzungssysteme in relativ kurzer Zeit deutlich verbessern kann. Es ist auch anzunehmen, dass heute Google Translate mit dem neuronalen Ansatz bessere Resultate liefert als früher mit dem statistischen Ansatz.

Für Zwecke dieser Arbeit sind auch Studien relevant, welche die von maschinellen MÜs erstellten Übersetzungen mit menschlichen Übersetzungen vergleichen. Läubli et al. (2018) erklären, dass neuere Forschungsergebnisse darauf hindeuten, dass die neuronale maschinelle Übersetzung der professionellen menschlichen Übersetzung gleichwertig sei. Sie untersuchten diese Behauptung empirisch, wobei sie sowohl die Bewertung einzelner Sätze als auch ganzer Dokumente verglichen. In ihrer Studie rekrutierten sie professionelle Übersetzer*innen als Evaluator*innen, die zwischen

zwei Übersetzungen die bessere auswählen sollten. Die Auswahl war zwischen einer professionellen menschlichen Übersetzung und einer maschinellen Übersetzung, die von einem von Microsoft entwickelten System auf dem neuesten Stand der Technik erstellt wurde. Die Übersetzungen erfolgten aus dem Chinesischen ins Englische. Die Bewerter*innen bewerteten nach zwei Kriterien: Adäquatheit, bei der die Bewerter*innen sowohl den Ausgangstext als auch die Übersetzung sehen (Beantwortung der Frage, welche Übersetzung die Bedeutung des Ausgangstextes angemessener wiedergibt), und Geläufigkeit, bei der die Bewerter*innen nur die Übersetzung sehen (Beantwortung der Frage, welcher Text die bessere Verwendung vom Englischen darstellt). Die Ergebnisse zeigen, dass sich maschinelle und menschliche Übersetzungen in Hinblick auf Adäquatheit auf Satzebene statistisch nicht bedeutsam unterscheiden, was laut Läubli et al. mit früheren Studien übereinstimmt. Bei der Bewertung ganzer Dokumente zeigen die Bewerter*innen jedoch eine statistisch bedeutsame Präferenz für menschliche Übersetzungen. In Hinblick auf Geläufigkeit bevorzugten die Bewerter*innen menschliche Übersetzungen sowohl auf Satz- als auch auf Dokumentenebene. Sie folgern daraus, dass einer der Gründe, warum die Bewertung auf Dokumentenebene die Qualitätslücke zwischen maschineller und menschlicher Übersetzung vergrößert, darin liegt, dass das von ihnen getestete maschinelle Übersetzungssystem immer noch auf Satzebene arbeitet und den weiteren Kontext nicht berücksichtigt (vgl. Läubli et al. 2018).

In einer vergleichenden Studie, in der das Google Translation MT System und die menschliche Übersetzung untersucht wurden, verglichen Cai et al. (2014) die Übersetzungen von gesprochenen und geschriebenen Textauszügen. Die Übersetzungen erfolgten vom Chinesischen ins Englische. Die maschinelle Übersetzung von Google Translate und die menschliche Übersetzung, die von einem Team von Fachexpert*innen für die englische als auch für die chinesische Sprache angefertigt wurde, wurden miteinander und auch mit dem chinesischen Original verglichen. Für den Vergleich wurden in dieser Studie mehrere automatische Evaluierungstools verwendet. Die Ergebnisse zeigten, dass die MÜ sowohl mit der menschlichen Übersetzung als auch mit den chinesischen Originaltexten in Bezug auf Formalität und Kohäsion hoch vergleichbar war. Die Autor*innen erklären, dass es möglich ist, die Schlussfolgerung zu ziehen, dass die Resultate von Google Translate der menschlichen Übersetzung auf semantischer und pragmatischer Ebene nahe kommen. Jedoch bedarf es einer Verbesserung auf der syntaktischen oder grammatikalischen Ebene. Dies bedeutet, dass Google Translate zwar eine verständliche und lesbare Übersetzung liefert, aber immer noch grammatikalische Fehler auftreten.

Bergman et al. (2014) führten eine Studie durch, in welcher die Qualität von MÜ mit Nachbearbeitung und menschlichen Übersetzungen im Bereich der öffentlichen Gesundheit

verglichen wurde. Die in dieser Studie verwendeten Dokumente wurden aus dem Englischen ins Spanische übersetzt. Die maschinellen Übersetzungen wurden von Google Translate erstellt und von Fachexpert*innen des öffentlichen Gesundheitswesens mit Erstsprache Spanisch nachbearbeitet, denen das englische Originaldokument und das maschinell übersetzte Dokument vorgelegt wurden; die Fachexpert*innen wurden angewiesen, die Grammatik und die inhaltliche Genauigkeit zu korrigieren, jedoch keine stilistischen Änderungen aufgrund persönlicher Vorlieben vorzunehmen. Die menschlichen Übersetzungen wurden von beauftragten Übersetzungsdienstleister*innen und internen zweisprachigen Mitarbeiter*innen angefertigt. Die Evaluierung der Übersetzungen wurde von einer separaten Gruppe spanischer Erstsprachler*innen aus dem Gesundheitswesen vorgenommen. Den Evaluators*innen wurde der Ausgangstext zusammen mit den beiden Übersetzungen vorgelegt, und sie wurden angewiesen, anzugeben, ob sie die eine Übersetzung der anderen vorziehen oder ob sie sie für gleichwertig halten. Die Analyse der Qualitätseinstufung umfasste insgesamt 25 gepaarte Dokumente. Die Ergebnisse dieser Studie zeigen, dass MÜ mit menschlicher Nachbearbeitung zu einer ähnlichen Qualität wie menschliche Übersetzungen führen kann. Die Autor*innen kamen zu dem Schluss, dass MÜ in Verbindung mit Nachbearbeitung die Verfügbarkeit von mehrsprachigem Material für das Gesundheitswesen erheblich verbessern könnte.

Besonders relevant sind zwei Studien, die als Teil von Masterarbeiten am Zentrum für Translationswissenschaft der Universität Wien durchgeführt wurden und die BWS-Methode verwenden. Da diese Studien Teile von Masterarbeiten sind, sind sie relativ klein im Umfang und weisen gewisse Einschränkungen auf. Dennoch stellen sie einen wesentlichen Beitrag für die Translationswissenschaft da, weil das nach dem Kenntnisstand des Autors die ersten wissenschaftlichen Arbeiten sind, in der BWS zur Schätzung der Translationsqualität eingesetzt wurden.

In ihrer Masterarbeit untersucht Finetti (2020) welche Methoden zur Vorbereitung das Resultat der maschinellen Übersetzung aus dem Englischen ins Italienische verbessern. Um festzustellen, ob eine gewisse Methode zur Vorbereitung die Qualität der Übersetzungen steigert, wendet Finetti die BWS-Methode an, um zu untersuchen, ob Proband*innen MÜ vorbereiteter Texte oder menschliche Übersetzungen bevorzugen. Im Zuge dieser Studie wurden die Zieldokumente von fünf Proband*innen – ausgebildeten Translatoren*innen – bewertet. Die Proband*innen bekamen neben den Ausgangstexten auch jeweils vier Übersetzungen zu sehen – eine MÜ der nicht-vorbereiteten Texte und drei MÜs vorbereiteter Texte wobei unterschiedliche Kombinationen von Methoden zur Vorbereitung verwendet wurden. Auf einer Skala von -4 bis +4 wurden zwei Übersetzungen bewertet – die, die als die schlechteste bzw. die beste empfunden waren. Finetti kommt zum Schluss,

dass unabhängig von der Vorbearbeitungsmethode die MÜs vorbearbeiteter Texte im Allgemeinen als besser bewertet werden als MÜs nicht-vorbearbeiteter Texte. Dabei bemerkt Finetti auch, dass alle Übersetzungen in Hinsicht auf Qualität als ungenügend im Vergleich zu den Erwartungen der Proband*innen waren. Finettis Arbeit ist zwar relativ klein im Umfang mit nur 5 Evaluat*innen, aber dennoch zeigen die Resultate Tendenzen auf, die mit früheren Studien im Einklang sind.

Eine weitere Masterarbeit, welche die BWS-Methode verwendet, ist die Arbeit von Etzlinger (2022), in welcher sie die Qualität von 4 NMÜs mit einer menschlichen Übersetzung vergleicht. Der untersuchte Fachbereich sind die Naturwissenschaften und das untersuchte Sprachenpaar ist Englisch-Deutsch. Die maschinellen Übersetzungen der Ausgangstextpassagen wurden mit vier verschiedenen NMÜ-Systemen erstellt, nämlich DeepL Translator, Google Translate, Microsoft Bing Translator und SYSTRAN Translate; die menschliche Übersetzung wurde von einem professionellen Übersetzungsbüro erstellt, welches sich auf Fachübersetzungen spezialisiert. Als Evaluat*innen haben insgesamt neun Übersetzer*innen und acht Naturwissenschaftler*innen an der Umfrage teilgenommen. Die Evaluat*innen bewerteten immer insgesamt 5 Sets mit jeweils 4 Übersetzungen (jeweils der einen Humanübersetzung und 3 von den insgesamt 4 MÜs). Die Evaluat*innen bewerteten die Rohausgabe von 3 der 4 untersuchten NMT-Systeme als schlechter als die menschliche Übersetzung, während DeepL deutlich besser abschnitt als die menschliche Übersetzung. Es ist auch das einzige NMÜ-System, das eine positive Gesamtbewertung erzielte, während die anderen drei Systeme alle eine negative Gesamtbewertung erhielten. Auch diese Studie war relativ klein im Umfang, zeigt jedoch deutlich, dass unter gewissen Bedingungen NMÜ-Systeme menschlichen Übersetzungen nahekommen oder diese sogar übertreffen.

Zusammengefasst kann gesagt werden, dass MÜ in der Sprachkombination Englisch-Deutsch schon sehr lange Resultate geliefert hat, die als sehr zufriedenstellend oder gut betrachtet werden. Eine Erklärung dafür ist die Tatsache, dass für dieses Sprachenpaar sehr viele einsprachige und zweisprachige Textkorpora existieren. Darüber hinaus kann verallgemeinert gesagt werden, dass NMÜ-Systeme zufriedenstellende Resultate liefern, die die von Menschen erstellten Übersetzungen zwar nicht übertreffen, aber ihnen nachweislich sehr ähnlich sind. Die oben beschriebenen Studien zeigen, dass auch NMÜ-Systeme nicht ideal und fehlerfrei sind, jedoch themenunabhängig Resultate liefern können, die als zufriedenstellend bewertet werden, vor allem für die Sprachkombination Englisch-Deutsch.

3.2. Übersetzen im Fachgebiet Musik

Es existieren relativ wenige Werke, die sich mit der Übersetzung von musikwissenschaftlichen Texten befassen oder unterschiedliche Terminologien in diesem Fachgebiet untersuchen. Eines davon ist der Beitrag von Granada (2018), welcher potenziell verwirrende Polysemie und falsche Freunde in der Musikterminologie unterschiedlicher Sprachen behandelt. Anhand von mehreren Beispielen zeigt Granada, dass in unterschiedlichen Sprachgemeinschaften musikbezogene Termini aufgrund von kulturellen Unterschieden unterschiedlich gedeutet werden, selbst wenn sie den gleichen Begriff benennen. Granada kommt zu dem Schluss, dass diese unterschiedlichen Sichten auch damit in Verbindung stehen, dass die in der Musik verwendeten Termini auch allgemeinsprachliche Bedeutungen haben, welche sich in unterschiedlichen Sprachen auf unterschiedliche Begriffe beziehen.

Kiš Žuvela (2018) zeigt am Beispiel der kroatischen Sprache, dass Musiktheorie ein sehr herausforderndes Fachgebiet für Translator*innen ist und das bisherige Übersetzungen vieler Werke nicht hochwertig genug sind. Kiš Žuvela vermerkt, dass unter anderem Polysemie, Homonymie, Fremdwörter, Fremdausdrücke und lexikalische Leerstellen die Teilnehmer*innen an Diskursen über Musik daran hindern, eindeutige und kompetente Kommunikation herzustellen, und stellt weiter fest, dass auch offiziell zugelassene Lehrbücher und Unterrichtsmaterialien eine starke Inkonsistenz im Terminologiegebrauch aufweisen.

Wiener (2002) diskutiert die Problematik des Übersetzens in der Musik am Beispiel von Übersetzungen musiktheoretischer Traktate aus dem 18. Jahrhundert. Wiener betrachtet diese Übersetzungen als Phänomen des Kulturtransfers und betont, dass komplexe kulturtransferielle Momente nicht einfach durch Begriffsaneignungen mittels substituier, scheinbar „äquivalenter“ Übersetzung zu bewältigen sind. Wiener bemerkt, dass es bei vielen Übersetzungen geschah, dass der Zieltext in Bezug auf Textart, Textlänge und Struktur vom Ausgangstext sehr abweicht. Diese Änderungen haben große Auswirkungen darauf, wie Zieltexte in der Zielkultur verstanden und rezipiert werden und erschweren einen direkten Vergleich zwischen Ausgangstexten und Zieltexten.

Giannini (2014) beschreibt den Entscheidungsprozess während des Übersetzens anhand von kurzen Fallstudien – Übersetzungen der musiktheoretischen Terminologie von Arnold Schönberg. Giannini beschreibt das „Verhandeln“ von Übersetzungen – den Prozess während welchen Übersetzer*innen bewusst übersetzerische Entscheidungen treffen und dabei u. A. zwischen den Intentionen der Ausgangstextautor*innen, den Wünschen der Verlage und den Erwartungen des Zielpublikums balancieren. Giannini bemerkt, dass sich durch unterschiedliche Epochen die

Präferenz zwischen Ausgangstext-Loyalität und Zielkultur-Loyalität ändern kann. Zumeist können Übersetzungen weder dem Ausgangstext noch der Zielkultur vollkommen loyal sein. Die Übersetzung musikwissenschaftlicher Texte ist immer von der musiktheoretischen Tradition der Zielkultur geprägt, daher haben Übersetzer*innen oft trotz Loyalität gegenüber dem Ausgangstext eine Anzahl der Termini zielkulturorientiert ausgewählt und verwendet. Dies kann gegebenenfalls problematisch sein, denn bestimmte Musikwissenschaftler*innen können einen in einer gegebenen musiktheoretischen Tradition vorhandenen Terminus übernehmen, diesen jedoch weiterentwickeln und mit ihm etwas konkreteres meinen. In solchen Fällen griffen Übersetzer*innen dazu, diese Besonderheiten in Paratexten zu Übersetzungsstrategien zu beschreiben, was, wie Giannini bemerkt, größtenteils aus der Mode gekommen ist. Das Fazit ist, dass keine Übersetzungsstrategie einwandfrei und immer korrekt ist, sondern dass übersetzerische Entscheidungen, auch und vielleicht insbesondere im Fachgebiet Musik, immer unter Beachtung kultureller und pragmatischer Gegebenheiten zu treffen sind. Im Fall von Musik handelt es sich dabei in erster Linie um Rezeptionsmechanismen, dem musiktheoretischen und kulturellen Hintergrund der Zielsprache und den Verhandlungen zwischen aller im Übersetzungsprozess involvierten Personen.

Wie schon am Anfang angedeutet, existieren nach dem Wissensstand des Autors dieser Masterarbeit derzeit relativ wenige Werke, die die Problematik des Übersetzens in der Musikwissenschaft behandeln. Neben den oben beschriebenen Werken existieren einige Werke, die sich mit dem Übersetzen von Lyrik beschäftigen, beispielsweise Minors (2020, 2021) und Desblache (2018). Diese Werke sind für den Zweck dieser Masterarbeit von geringer Bedeutung, dennoch wurde an dieser Stelle zwecks Vollkommenheit auch auf sie verwiesen.

4. Methodik

Im folgenden Kapitel wird die Methodik für die Datenerhebung und Auswertung beschrieben. Um die Qualität von NMÜ-Systemen mit der Qualität menschlicher Übersetzungen zu vergleichen, wurde eine Umfrage erstellt, in welcher befragte Fachexpert*innen im Gebiet Musik bewerten sollen, welche von vier angebotenen Übersetzungen sie für die Beste und welche für die Schlechteste halten. Des Weiteren sollten die Befragten auch angeben, welche Übersetzungen ihrer Meinung nach von Menschen und welche von einem MÜ-System erstellt wurden. Zusammen mit einem Interview sollen die Resultate dieser Umfrage Einsicht darüber geben, ob Fachexpert*innen diese zwei Übersetzungsansätze voneinander unterscheiden können. In folgenden Unterkapiteln wird die Methodik genauer beschrieben.

4.1. Auswahl der Textproben

Für Zwecke dieser Masterarbeit wurden zuerst Textproben aus dem Fachgebiet Musik ausgewählt. Die Textproben befassen sich hauptsächlich mit Musiktheorie, aus Passagen der ausgewählten Quellen, welche ein musikwissenschaftliches oder musikkritisches Phänomen oder Konzept behandeln. Gleichzeitig wurden Textproben auch aus anderen Arten von Texten ausgewählt, in welchen andere fachgebietsrelevante Inhalte behandelt werden. Bei der Auswahl der Textproben wurde darauf geachtet, dass ein breiterer Kontext deutlich erkennbar und leicht zu erschließen ist. Es wurden insgesamt sechs Textproben ausgewählt werden, welche aus jeweils zwei bis vier Sätzen bestehen, abhängig von der Länge der Sätze.

Die Textproben wurden manuell aus unterschiedlichen Quellen extrahiert. Das genauere Quellenverzeichnis und die ausgewählten Textpassagen sind in Anhang 1 vollständig angegeben. Folgende Werke wurden als Quellen ausgewählt:

- openmusictheory.com – ein Open-Source, interaktives Online-Lehrbuch für Musiktheoriekurse auf Universitätsniveau, vorwiegend basiert auf Ressourcen, die von Kris Shaffer, Bryn Hughes und Brian Moseley verfasst wurden;
- *Concise Introduction to Tonal Harmony* – die Autoren L. Poundie Burstein und Joseph N. Straus beschreiben grundlegende Konzepte der Musiktheorie;
- *The Complete Musician: An Integrated Approach To Tonal Theory, Analysis, And Listening* – beginnend mit den musikalischen Grundlagen, deckt Steven G. Laitz alle Themen ab, die für ein gründliches Verständnis der Musiktheorie im Grundstudium notwendig sind, indem er sich auf Musik im Kontext konzentriert;

- Baroque Music Today: Music As Speech, Ways to a New Understanding of Music – in diesen Aufsätzen fasst Nikolaus Harnoncourt seine Ansichten über die Barockmusik zusammen, die auf jahrelanger Beschäftigung mit der Aufführung der Musik der Vergangenheit beruhen und erörtert einige der Probleme der historischen Interpretation;
- The Synthesizer: A Comprehensive Guide To Understanding, Programming, Playing, And Recording The Ultimate Electronic Music Instrument – Mark Vail beschreibt die vollständige Geschichte des Synthesizers: die Ursprünge der vielen Formen des Instruments, entscheidende Fortschritte, für das Instrument relevante Kompositionen, die Grundlagen und Feinheiten der Akustik und des Syntheseklangs.
- Analysis of Jazz: A Comprehensive Approach – Laurent Cugny untersucht und verknüpft die theoretischen und methodischen Prozesse, die dem Jazz zugrunde liegen.

Bei der Auswahl der Quellen der Textproben war es wichtig, dass die Quellen relevant und qualitativ hochwertig sind. Daher wurde darauf geachtet, dass alle Quellen auf relevanten Webseiten gute Bewertungen haben und dass sie von Leser*innen als relevant und nützlich eingestuft wurden. Die Angemessenheit der Textproben wurde auch durch die Fachkenntnisse des Autors dieser Masterarbeit bestätigt.

4.2. Erstellung der Übersetzungen

Für den Zweck dieser Masterarbeit wurden insgesamt sechs Textproben aus dem Englischen ins Deutsche übersetzt. Für jede der Textproben wurden vier Übersetzungen erstellt: jeweils einmal von DeepL, Google Translate und zwei menschlichen Übersetzer*innen. Bei der Erstellung der maschinellen Übersetzungen wurden weder Vorbereitung noch Nachbearbeitung vorgenommen.

Für die menschliche Übersetzung wurden zwei Student*innen des Masterstudiengangs Translation der Universität Wien beauftragt, welche vor dem Übersetzen der Textproben bestätigt haben, dass sie bereits musiktheoretische Texte oder Texte mit ähnlicher Thematik übersetzt haben und/oder dass sie als Musiker*innen Erfahrung auf dem Themengebiet Musiktheorie gesammelt haben. Die beiden Humanübersetzer*innen wurden darüber informiert, dass die Übersetzungen für den Vergleich menschlicher und maschineller Übersetzung dienen sollen und dass sie aufgrund der Natur des Vergleichs keine MÜ-Systeme für die Erstellung ihrer Übersetzungen benutzen dürfen. Ob die Übersetzer*innen sich an diese Vorgabe gehalten haben, wurde nicht weiter überprüft, aber unter Betracht der in weiteren Kapiteln beschriebenen Resultate kann angenommen werden, dass sie den Anweisungen Folge geleistet haben. Den Übersetzer*innen wurden die Textproben einschließlich

eines Quellenverzeichnis zugesandt. Sie wurden weiterhin angewiesen, unter der Annahme zu arbeiten, dass sie für ein Zielpublikum von Fachexpert*innen im Bereich Musik übersetzen und dass das Format und Medium der Übersetzung identisch bleiben, wie das Format und das Medium der Ausgangstexte.

Für die Erstellung der maschinellen Übersetzungen wurden Google Translate und DeepL verwendet. Einer der Gründe für die Auswahl dieser Systeme ist der, dass diese Systeme kostenlos sind und daher die Durchführung der Auswertung erleichtern. Weiterhin erfreuen sich diese Systeme großer Beliebtheit. Ein weiterer Grund ist auch der, dass diese Systeme auf großen Textmengen trainiert wurden und somit eine höhere Übersetzungsgenauigkeit zu erhoffen ist. Der letzte Grund ist der, dass diese Systeme, wie schon im Kapitel 3 beschrieben wurde, im allgemeinen zufriedenstellende Resultate liefern. Die Übersetzungen können im Anhang 2 gefunden werden, in der Reihenfolge, wie sie auch in der Umfrage erscheinen.

4.3. Auswahl der Zielgruppe

Die Übersetzungen der Textproben wurden von Fachexpert*innen im Bereich Musik mit Deutsch als Erstsprache evaluiert. Bevor die Suche nach den Evaluator*innen stattfand, musste erstmal entschieden werden, was genau unter *Fachexpert*in* zu verstehen ist. Für den Zweck dieser Arbeit wurden sowohl formal-ausgebildete Musiker*innen und Musikwissenschaftler*innen als auch Autodidakt*innen in Betracht gezogen. Personen, die Interesse hatten, an der Umfrage teilzunehmen, wurden gegebenenfalls gebeten, kurz zu beschreiben, welche Art formaler Ausbildung sie haben bzw. welche Erfahrungen sie als Autodidakt*innen besitzen. Die Entscheidung, sowohl formal-ausgebildete Personen als auch Autodidakt*innen zu befragen, entstand aus dem Gedanken, dass sich unterschiedliche Erfahrungen und Perspektiven vertreten und dokumentiert werden sollten.

Die potenziellen Evaluator*innen wurden auf unterschiedliche Weisen kontaktiert. Einerseits wurden Lehrkräfte in formalen Bildungseinrichtungen per öffentlich zugänglichen E-Mail-Adressen angeschrieben. Andererseits wurden Musiker*innen und Musikwissenschaftler*innen mittels Suchanzeigen in auf das Fachgebiet Musik fokussierten Gruppen auf der Social Media Website Facebook gefunden. Zuletzt kamen weitere Evaluator*innen aus dem persönlichen Bekant*innenkreis des Autors dieser Arbeit hinzu. Somit wurden insgesamt sieben Evaluator*innen gefunden und befragt.

4.4. Anwendung der BWS-Methode

Für die Evaluierung der Übersetzungsqualität im Vergleich zur Humanübersetzung wurde für diese Arbeit die im Kapitel 2 beschriebene BWS-Methode ausgewählt. Diese Methode wurde deshalb ausgewählt, weil sich die in Kapitel 2 beschriebenen Vorteile dieser Methode sehr gut für die Evaluierung von Übersetzungen eignen. Die Übersetzungsqualität kann mit dieser Methode leicht bewertet werden, fast unabhängig von objektiven Metriken. Der Fokus liegt auf der subjektiven Wahrnehmung der Evaluator*innen und es soll eine Einsicht über den Gesamteindruck gewonnen werden, statt über einzelne Eigenschaften der Übersetzungen. Weiterhin wurde diese Methode ausgewählt, weil sie eine besondere Art der Ranking-Methode darstellt. Wie bereits erwähnt eignet sich Ranking gut, um die Resultate mehrerer MÜ-Systeme direkt zu vergleichen. Und obwohl es nicht möglich sein wird, die genauen Unterschiede zwischen menschlicher und maschineller Übersetzung zu identifizieren, wird dank dieser Methode ersichtlich, ob Evaluator*innen zwischen menschlicher und maschineller Übersetzung allgemein unterscheiden können oder nicht.

Die Datenerhebung wurde so konzipiert, dass den Evaluator*innen die Übersetzungen der Textproben und die Anweisungen für die Durchführung der Befragung in Form einer Umfrage übermittelt werden. Die Umfragen finden persönlich und vor Ort statt. Schon während der Kontaktaufnahme wird den potenziellen Evaluator*innen kurz erklärt, dass die Datenerhebung aus einer Umfrage und einem Interview besteht. Die erste Frage der Umfrage befasst sich mit der Qualität der 4 Übersetzungen.

Als eine Inspirationsquelle für die Erstellung der Umfrage diente Finetti (2020), aber die dieser Arbeit zu Grunde liegende Umfrage unterscheidet sich in mehrerer Hinsicht von jener in Finettis Arbeit. Für die Evaluierung sind insgesamt sechs Textproben ausgewählt worden, welche dann übersetzt wurden. Die Evaluator*innen wählen für jede der sechs Textproben eine beste und eine schlechteste Übersetzung aus und geben gleichzeitig an, wie gut die beste Übersetzung war (auf eine Skala von 0 bis 4) und wie schlecht die schlechteste Übersetzung war (auf einer Skala von 0 bis -4). Evaluator*innen geben weiterhin an, welche der vier Übersetzungen ihrer Meinung nach von Menschen erstellt wurden und welche von einem NMÜ-System. Nach dieser Evaluierung nehmen die Evaluator*innen an dem Interview teil, um ihre Antworten auf die Umfrage zu begründen und etwaige Kommentare oder Kritikpunkte anzugeben, damit ihre Antworten besser kontextualisiert werden können. Am Ende wird den Evaluator*innen gezeigt, ob ihre Einschätzungen bezüglich der Art der Übersetzung richtig waren oder nicht. Die Evaluator*innen haben keinen Zugang zu den

Textproben in der Ausgangssprache, um zu vermeiden, dass Fremdsprachenkenntnisse der Evaluator*innen ihre Entscheidung beeinflussen.

4.5. Erstellung des Leitfadens für das Interview

Um die Ergebnisse der Umfrage in einen genaueren Kontext zu setzen, das Profil der Evaluator*innen festzustellen und die Motivation hinter ihren Antworten besser zu verstehen, wurde entschieden, dass Evaluator*innen vor der Umfrage auch einige demographische Angaben machen sollen und dass mit den Evaluator*innen nach der Umfrage ein kurzes Interview geführt wird. Als eine erste Inspirationsquelle für die Auswahl der zu erhebenden demographischen Daten und für die Erstellung des Leitfadens für das Interview dienten Almaghout et al. (2015), Axinn et al. (2011) und Briggs (2018). Die demographischen Daten, die erhoben wurden, sind die Altersgruppe, das Geschlecht, ob Deutsch eine Erstsprache ist, ob weitere Erstsprachen vorhanden sind, Angaben zu Fremdsprachenkenntnissen und Angaben zur Fachkompetenz im Bereich Musik. Bei der Auswahl der Fragen bzw. ihrer Modifikation war es wichtig, zu entscheiden, welche Erfahrungen der Evaluator:innen für den Zweck dieser Masterarbeit relevant sind. In erster Linie ging es dabei um Erfahrung mit MÜ und Übersetzen allgemein, die Gründe für das Erstellen oder Beauftragen von Übersetzungen und die Zufriedenheit mit den Resultaten. Es war auch sehr wichtig, zu erfahren, ob die Evaluator:innen mit Übersetzungen im Fachgebiet Musikwissenschaft Kontakt hatten. Weiterhin war es von großer Bedeutung, genauere Informationen darüber zu bekommen, welche Art von Ausbildung die Evaluator:innen in den Gebieten Musik und Musikwissenschaft hatten (formale Ausbildung, Autodidaktik oder eine Kombination von beiden), um mögliche Unterschiede in den Antworten zu begründen. Letztlich wurden auch Fragen zur Umfrage erstellt, damit festgestellt werden kann, ob sich aus der Sicht der Evaluator:innen die Methode gut für die Datenerhebung eignet.

5. Ergebnisse der Umfrage und des Interviews

Im folgenden Kapitel werden die Ergebnisse der Umfrage und des Interviews vorgestellt. Im ersten Unterkapitel werden die demographischen Angaben der Teilnehmer*innen anonymisiert vorgestellt. Das zweite Unterkapitel beschreibt die Bewertung der Übersetzungen. Ein weiteres Unterkapitel wird den Antworten auf die Frage nach der Art der Übersetzung (menschlich oder maschinell) gewidmet. Im letzten Unterkapitel werden anhand des durchgeführten Interviews die Motivation und die Gründe hinter den Antworten der Evaluator*innen beschrieben.

5.1. Demographische Angaben

Die Evaluator*innen wurden gebeten, einige Angaben zur Person zu machen. Diese beinhalten die Altersgruppe, welcher sie angehören (18-30, 30-45, 45-60, 60+); das Geschlecht (freie Angabe), ob Deutsch eine Erstsprache ist, ob sie weitere Erstsprachen haben, ob sie Fremdsprachen beherrschen und Angaben zur Fachkompetenz im Bereich Musik.

Befragt wurden insgesamt zwei Evaluator*innen in der Altersgruppe zwischen 18 und 30, vier in der Altersgruppe zwischen 30 und 45 und eine*r Evaluator*in zwischen 45 und 60; es gab keine Evaluator*innen in der Altersgruppe über 60. Vier der Evaluator*innen haben zum Geschlecht männlich angegeben, während drei Evaluator*innen weiblich angegeben haben. Alle Evaluator*innen haben bestätigt, dass Deutsch eine Erstsprache ist. Jeweils eine*r Evaluator*in haben Italienisch, Slowakisch, Tschechisch, Ungarisch und Zimbrisch genannt. Alle Evaluator*innen nannten Englisch als eine Fremdsprache. Weitere Fremdsprachen, die genannt wurden, waren Französisch, Tschechisch, Spanisch, B/K/S und Latein.

In Hinblick auf die Bewertungen der Übersetzungsqualität gab es keine erwähnenswerten Tendenzen. Obwohl die einzelnen Evaluator*innen die Übersetzungen unterschiedlich bewertet haben, konnte kein Zusammenhang zwischen der Bewertung und einer bestimmten demographischen Gruppe oder ähnlichem festgestellt werden. Im Gegensatz dazu gab es gewisse Tendenzen in Hinblick auf die Frage zur Art der Übersetzung (menschlich oder maschinell). So haben die jüngsten Evaluator*innen in der Altersgruppe von 18 bis 30 die meisten Fehleinschätzungen abgegeben (jeweils 9); die Evaluator*innen in der Altersgruppe zwischen 30 und 45 hatten weniger Fehleinschätzungen (jeweils 4, 5, 7 und 8); eine Evaluator*in in der Altersgruppe zwischen 45 und 60 hatte die wenigsten Fehleinschätzungen (3). Eine weitere Tendenz ist auch in Bezug auf den formalen Bildungsgrad im Bereich Musik zu erkennen. So hat die Person mit dem höchsten Bildungsgrad eines Doktorats im Bereich Musikwissenschaft im Vergleich zu den anderen

Evaluator*innen die meisten Übersetzungen richtig zugeordnet. Eine Person, die ein Doktorat im Fachgebiet Musikwissenschaft anstrebt, hat ebenfalls eine deutlich geringere Anzahl an Fehleinschätzungen als die restlichen Evaluator*innen. Evaluator*innen mit einer formalen Ausbildung bis zum Universitätsgrad gekoppelt mit autodidaktischen Methoden, haben mehr Fehleinschätzungen abgegeben als Evaluator*innen mit höherem formalem Bildungsgrad im Bereich Musik, jedoch wurde die größte Anzahl an Fehlerschätzungen von jenen Evaluator*innen abgegeben, die entweder nur einen relativ geringen Bildungsgrad im Bereich Musik haben (z. B. Musikgrundschule oder Privatunterricht) oder die nur reine Autodidakt*innen sind.

Aufgrund der relativ geringen Evaluator*innenzahl war es zu erwarten, dass nicht alle demographischen Gruppen repräsentiert werden. So ist es dazu gekommen, dass es keine Vertreter*innen der ältesten Altersgruppe (60+) unter den Evaluator*innen gab. Des Weiteren gab es auch keine Evaluator*innen, die bei der Frage zum Geschlecht weder männlich noch weiblich angegeben haben. Die Fremdsprachenkenntnisse der Evaluator*innen hatten keinen bemerkenswerten Einfluss auf die Antworten.

Zusammenfassend kann gesagt werden, dass diese Umfrage darauf hindeutet, dass die meisten demographischen Eigenschaften von Musiker*innen relativ wenig Einfluss darauf haben, wie sie die Qualität von Übersetzungen einschätzen und bewerten. Ausschlaggebend sind vor allem der Bildungsgrad, insbesondere in Bezug auf das Fachgebiet Musik, und das Alter. Eine zukünftige Studie sollte nach Möglichkeit für eine höhere Anzahl an Evaluator*innen konzipiert werden, sodass auch die Chance höher ist, dass eine höhere Anzahl an Repräsentation unterschiedlicher demographischen Gruppen gewährleistet werden kann und die Ergebnisse aussagekräftiger sind.

5.2. Bewertung der Übersetzungen

Im Folgenden werden die Ergebnisse der Bewertung präsentiert (siehe Tabelle 1 und Tabelle 2, S. 45 und 46). Die Bewertung der 4 Übersetzungen zeigt, dass maschinelle Übersetzung im Regelfall als schlechter empfunden wird als Übersetzungen menschlicher Übersetzer*innen. Wie in vorherigen Kapiteln beschrieben, erfolgte die Bewertung der besten und schlechtesten Übersetzung jedes Sets auf einer Skala von -4 bis +4. Die sechs Sets der Übersetzungen wurden von insgesamt sieben Evaluator*innen bewertet. Somit ist das schlechtestmögliche Resultat, welches eines der MÜ-Systeme bzw. eine*r der Übersetzer*innen erzielen könnte, -168; dies wurde passieren, falls die Übersetzungen eines MÜ-Systems bzw. eine*r der Übersetzer*innen in jedem der sechs Sets von allen sieben Evaluator*innen mit -4 bewertet würden. Entsprechend ist das bestmögliche Resultat +168.

Tabelle 1: Bewertung der Übersetzungen als beste und schlechteste mit Punkteverteilung

		GoogleT	DeepL	H1	H2	G+D	H1+H2
T1	Set1		-1		+3	-1	+3
	Set2		-2		+4	-2	+4
	Set3	-3			+2	-3	+2
	Set4		-2	+3		-2	+3
	Set5			-3	+3	0	0
	Set6	+2		-2		+2	-2
	insgesamt	<u>-1</u>	<u>-5</u>	<u>-2</u>	<u>+12</u>	<u>-6</u>	<u>+10</u>
T2	Set1	-4			+4	-4	+4
	Set2	-4		+3		-4	+3
	Set3	-4	+3			-1	0
	Set4	-3		+4		-3	+4
	Set5	-4		+4		-4	+4
	Set6	-3			+3	-3	+3
	insgesamt	<u>-22</u>	<u>+3</u>	<u>+11</u>	<u>+7</u>	<u>-19</u>	<u>+18</u>
T3	Set1		-1		+2	-1	+2
	Set2	-2		+2		-2	+2
	Set3	-4			+1	-4	+1
	Set4	-3	+1			-2	0
	Set5	+2			0	+2	0
	Set6	-4			+3	-4	+3
	insgesamt	<u>-11</u>	<u>0</u>	<u>+2</u>	<u>+6</u>	<u>-11</u>	<u>+8</u>
T4	Set1	+3		-2		+3	-2
	Set2	-4			+4	-4	+4
	Set3	-4			+3	-4	+3
	Set4	-4	+3			-1	0
	Set5	-3			+3	-3	+3
	Set6	-2			+4	-2	+4
	insgesamt	<u>-14</u>	<u>+3</u>	<u>-2</u>	<u>+14</u>	<u>-11</u>	<u>+12</u>
T5	Set1		-3		+2	-3	+2
	Set2	-3			+3	-3	+3
	Set3	-4		+4		-4	+4
	Set4	-4		+3		-4	+3
	Set5	-1		+1		-1	+1
	Set6		-2		+3	-2	+3
	insgesamt	<u>-12</u>	<u>-5</u>	<u>+8</u>	<u>+8</u>	<u>-17</u>	<u>+16</u>
T6	Set1		-2	+4		-2	+4
	Set2	-3	+3			0	0
	Set3	-4		+3		-4	+3
	Set4	-4	+3			-1	0
	Set5			+4	-3	0	+1
	Set6	-2			+3	-2	+3
	insgesamt	<u>-13</u>	<u>+4</u>	<u>+11</u>	<u>0</u>	<u>-9</u>	<u>+11</u>
T7	Set1	+4			-3	+4	-3
	Set2	-4			+4	-4	+4
	Set3	-4			+4	-4	+4
	Set4	-4		+4		-4	+4
	Set5		+4	-4		+4	-4
	Set6	-4			+4	-4	+4
	insgesamt	<u>-12</u>	<u>+4</u>	<u>0</u>	<u>+9</u>	<u>-8</u>	<u>+9</u>

Tabelle 2: Gesamtbewertung und durchschnittliche Bewertung

		GoogleT	DeepL	H1	H2	G+D	H1+H2
T1...T7	Set1 insgesamt	+3	-7	+2	+8	-4	+10
	Set2 insgesamt	-20	+1	+5	+15	-19	+20
	Set3 insgesamt	-27	+3	+7	+10	-24	+17
	Set4 insgesamt	-22	+5	+14	0	-17	+14
	Set5 insgesamt	-6	+4	+2	+3	-2	+5
	Set6 insgesamt	-13	-2	-2	+20	-15	+18
	insgesamt	-85	+4	+28	+56	-81	+84

		GoogleT	DeepL	H1	H2	G+D	H1+H2
(T1...T7)/7	Set1 durchschnittlich	+0.43	-1.00	+0.29	+1.14	-0.57	+1.43
	Set2 durchschnittlich	-2.86	+0.14	+0.71	+2.14	-2.71	+2.86
	Set3 durchschnittlich	-3.86	+0.43	+1.00	+1.43	-3.43	+2.43
	Set4 durchschnittlich	-3.14	+0.71	+2.00	0.00	-2.43	+2.00
	Set5 durchschnittlich	-0.86	+0.57	+0.29	+0.43	-0.29	+0.71
	Set6 durchschnittlich	-1.86	-0.29	-0.29	+2.86	-2.14	+2.57
	insgesamt durchschnittlich	-12.14	+0.57	+4.00	+8.00	-11.57	+12.00

Übersetzungen beider MÜ-Systeme wurden im Regelfall von den Befragten als schlechter empfunden als die menschlichen Übersetzungen. Die von Google Translate erstellten Übersetzungen wurden im Vergleich zu den restlichen als besonders schlecht empfunden. Im Gegensatz zu DeepL und den zwei menschlichen Übersetzerinnen, hat Google Translate ein negatives Resultat erzielt und wurde auch im Vergleich zu DeepL als schlecht empfunden. Der Punkteunterschied zwischen Google Translate und DeepL ist dabei mehrfach größer als der Punkteunterschied zwischen DeepL und den menschlichen Übersetzer*innen. Der Punkteunterschied zwischen DeepL und H1 ist etwas kleiner, als der Punkteunterschied zwischen H1 und H2. Die perzipierte Übersetzungsqualität ist daher deutlich geringer als der Unterschied gegenüber Google Translate. Die von Google Translate erstellten Übersetzungen wurden in 28/42 Fällen (66,67%) als die schlechteste Übersetzung ausgewählt, im Regelfall mit einer Bewertung von -3 oder -4.

Besonders auffallend ist die Bewertung von Teilnehmer*in T2, welche*r in allen 6 Sets die von Google Translate ausgewählten Übersetzung als die schlechteste empfand mit einer Gesamtbewertung von -22 (die niedrigste Gesamtbewertung pro Evaluator*in ist dabei -24). Die von Google Translate erstellten Übersetzungen wurden nur in Ausnahmefällen als die beste Übersetzung ausgewählt, in insgesamt vier Fällen. Von diesen vier Fällen ist Set 1 das, welches zwei Mal die beste Übersetzung von Google Translate enthielt; Google Translates Übersetzung im Set 1 wurde nur einmal als die schlechteste Übersetzung ausgewählt (seitens T2), und von den 6 anderen

Evaluator*innen entweder als die beste Übersetzung bewertet oder als weder die beste noch die schlechteste Übersetzung eingestuft.

Es lässt sich auch bemerken, dass sich Tendenzen bei Punktevergabe je nach Set deutlich unterscheiden. So ist der Punkteunterschied zwischen der besten und der schlechtesten Übersetzung in den Sets 2, 3, 4 und 6 recht groß, während er in Sets 1 und 5 verhältnismäßig klein ist. Set 1 ist ebenfalls auffallend, weil es das einzige Set ist, in welchem die Übersetzung von DeepL als die insgesamt schlechteste betrachtet wurde. Ähnlich ist Set 5 auffallend, weil es das einzige Set ist, in welchem die Übersetzung von DeepL als die insgesamt beste betrachtet wurde. Letztens ist Set 4 auffallend, weil es das einzige Set ist, in welchem die Übersetzung von H1 die Übersetzung von H2 übertrifft. Eine einfache Erklärung für diese Unterschiede in den Bewertungstendenzen ist nicht klar ersichtlich. Die Textlänge scheint eine geringe Rolle zu spielen und das in den Passagen behandelte Themengebiet ebenso. Es dürfte sich um eine komplexe Interaktion zwischen den Trainingsdaten der MÜ-Systeme bzw. der Erfahrung und der übersetzerischen Entscheidungen auf der einen Seite und der verwendeten Satztermini und Satzstruktur auf der anderen Seite handeln, welche dazu geführt hat, dass die Übersetzungen genau auf diese Weise bewertet wurden.

5.3. Mensch oder Maschine?

Im Folgenden werden die Antworten auf die Frage nach der Art der Übersetzung (menschlich oder maschinell) vorgestellt (siehe Tabelle 3 und Tabelle 4, S. 48 und 49). Jede*r der Teilnehmer*innen hat bei der Umfrage jeweils sechs Übersetzungen erstellt von jeweils vier Übersetzer*innen als entweder menschlich oder maschinell eingestuft. Da es insgesamt sieben Teilnehmer*innen gab, wurde das Resultat der 4 Übersetzer*innen jeweils 42-mal bewertet.

In diesem Zusammenhang wurden die von Google Translate erstellten Übersetzungen in 35 von 42 Fällen (83,33%) richtig als maschinell erstellte Übersetzungen eingestuft, während DeepLs Übersetzungen nur in 28 von 42 Fällen (66,67%) richtig eingestuft wurden. Übersetzungen von H1 wurden in 27 von 42 Fällen (64,28%) korrekt als menschliche Übersetzungen erkannt, während Übersetzungen von H2 in 32 von 42 Fällen (76,19%) richtig kategorisiert wurden. Alle 4 Übersetzungen wurden öfter korrekt als falsch kategorisiert. Die von Google Translate erstellten Übersetzungen, welche zugleich als die allgemein am schlechtesten bewertet wurden, fallen am deutlichsten auf und werden sehr oft als maschinelle Übersetzungen erkannt. Ähnlich ergeht es Übersetzungen von H2, welche generell als die besten Übersetzungen bewertet wurden und relativ oft richtig als menschliche Übersetzungen erkannt wurden.

Tabelle 3: Schätzungen, ob eine Übersetzung eine Humanübersetzung oder eine maschinelle Übersetzung ist (grün: richtig geraten; rot: falsch geraten)

		GoogleT	DeepL	H1 ("Jana")	H2 ("Paula")
T1	Set1	human	maschinell	maschinell	human
	Set2	maschinell	maschinell	maschinell	human
	Set3	maschinell	human	maschinell	human
	Set4	maschinell	maschinell	human	human
	Set5	maschinell	maschinell	maschinell	human
	Set6	human	maschinell	maschinell	maschinell
T2	Set1	maschinell	human	maschinell	human
	Set2	maschinell	maschinell	human	human
	Set3	maschinell	human	human	human
	Set4	maschinell	maschinell	human	human
	Set5	maschinell	maschinell	human	human
	Set6	maschinell	human	human	human
T3	Set1	human	maschinell	human	human
	Set2	human	maschinell	human	maschinell
	Set3	maschinell	human	human	maschinell
	Set4	maschinell	maschinell	human	human
	Set5	human	maschinell	maschinell	human
	Set6	maschinell	maschinell	human	human
T4	Set1	human	maschinell	human	human
	Set2	maschinell	maschinell	human	human
	Set3	maschinell	maschinell	maschinell	human
	Set4	maschinell	human	maschinell	human
	Set5	maschinell	maschinell	human	human
	Set6	maschinell	human	human	human
T5	Set1	maschinell	maschinell	human	human
	Set2	maschinell	maschinell	maschinell	human
	Set3	maschinell	human	human	human
	Set4	maschinell	maschinell	human	human
	Set5	maschinell	maschinell	maschinell	maschinell
	Set6	maschinell	maschinell	human	human
T6	Set1	maschinell	maschinell	human	human
	Set2	maschinell	human	maschinell	maschinell
	Set3	maschinell	human	human	human
	Set4	maschinell	human	maschinell	maschinell
	Set5	maschinell	maschinell	human	maschinell
	Set6	maschinell	human	human	maschinell
T7	Set1	human	human	maschinell	maschinell
	Set2	maschinell	maschinell	human	human
	Set3	maschinell	maschinell	human	human
	Set4	maschinell	human	human	maschinell
	Set5	maschinell	human	maschinell	human
	Set6	maschinell	maschinell	human	human

Tabelle 4: Anzahl der richtig erkannten Art der Übersetzung

	GoogleT	DeepL	H1	H2
T1	4/6	5/6	1/6	5/6
T2	6/6	4/6	5/6	6/6
T3	3/6	5/6	5/6	4/6
T4	5/6	4/6	4/6	6/6
T5	6/6	5/6	4/6	5/6
T6	6/6	2/6	4/6	2/6
T7	5/6	3/6	4/6	4/6
Insgesamt	35/42 (83,33%)	28/42 (66,67%)	27/42 (64,29%)	32/42 (76,19%)

Die perzipierte Qualität der Übersetzungen von DeepL und H1 relativ nahe (im Kontext der Punktevergabe ist der Unterschied kleiner, als der Unterschied zwischen den beiden Humanübersetzungen), aber DeepLs Übersetzungen werden dennoch als die schlechteren der beiden betrachtet. Jedoch muss bemerkt werden, dass DeepLs Übersetzungen im Vergleich zu jenen von Google Translate deutlich öfter mit menschlichen Übersetzungen verwechselt werden. Gleichzeitig muss bemerkt werden, dass Übersetzungen von H1 im Vergleich zu jenen von H2 relativ oft für maschinelle Übersetzungen gehalten wurden.

Bei der Kategorisierung nach menschlicher oder maschineller Übersetzung können einige Unterschiede in den Tendenzen der Evaluators*innen bemerkt werden. T2, welche*r den höchsten Bildungsgrad in Bezug auf das Fachgebiet Musik angegeben hat, hat Übersetzungen von Google Translate und H2 in allen Fällen richtig kategorisiert. Ähnlich hat T4 Übersetzungen von H2 in allen Fällen richtig kategorisiert, während T5 und T6 Google Translates Übersetzungen immer richtig erkannt haben. Die restlichen Teilnehmer*innen haben bei jeder der 4 Übersetzer*innen mindestens einmal falsch geschätzt. Alle 7 Teilnehmer*innen haben in Bezug auf DeepLs und H1s Übersetzungen mindestens eine Fehlschätzung gegeben. Auch in diesem Fall ist es nicht möglich, diese Unterschiede in den Bewertungstendenzen auf einen bestimmten Faktor zurückzuführen und es lässt sich annehmen, dass es sich um mehrere, miteinander interagierende Faktoren handelt.

5.4. Auswertung der Interviews

Um die Motivation hinter den Antworten der Evaluators*innen besser zu verstehen, wurden nach den Umfragen auch Interviews durchgeführt. Das Interview wurde durchgeführt, bevor den Evaluators*innen gezeigt wurde, ob sie mir ihren Angaben zur Art der Übersetzung richtig lagen. Der Leitfaden für das Interview und die transkribierten Interviews sind in Anhang 4 zu finden.

Alle Evaluator*innen berichten, dass sie MÜ verwendet haben. Alle Evaluator*innen haben von Google Translate gehört und es auch bereits verwendet. Von DeepL haben nur T2 und T5 gehört und sie haben auch erwähnt, dass sie es relativ regelmäßig verwenden. Weitere Systeme, die erwähnt wurden, waren TransDict, welches von T2 verwendet wird, und Babel Fish, welches T7 in der Vergangenheit verwendet hat.

Alle Evaluator*innen waren in Bezug auf die Übersetzungsqualität skeptisch, insbesondere im Fall von Google Translate. Die Evaluator*innen waren sich einig, dass Google Translate genügt, wenn es darum geht, einen Text sehr schnell zu übersetzen und nur die grundlegende Bedeutung hinter dem Text zu verstehen. Für präzisere Formulierungen würden die Evaluator*innen jedoch zu menschlichen Übersetzer*innen greifen. T2 und T5 empfinden die Übersetzungsqualität von DeepL als höher als die von Google Translate und bevorzugen DeepL, jedoch immer noch nur als Orientierungshilfe und nicht als Ersatz für menschliche Übersetzungen. Nichtsdestotrotz sind sich die Evaluator*innen einig, dass ihnen MÜ geholfen hat, Texte zu verstehen, die sie sonst nicht verstehen hätten können.

Der Verwendungszweck war für die Evaluator*innen sehr unterschiedlich. T1 und T6 teilten mit, dass sie MÜ in der Schule benutzt hatten, als sie Übersetzungen als Hausaufgabe im Fremdsprachenunterricht erstellen mussten. T2, T3, T5 und T7 hatten MÜ beruflich verwendet, unter anderem auch im musikbezogenen Kontext. Als Beispiel wurden Customer Support und allgemeine E-Mail-Korrespondenz genannt, insbesondere, wenn es wichtig ist, die Korrespondenz schnell durchzuführen und es keine Zeit für professionelle Übersetzungen gibt oder wenn darum geht, einen Text in seiner Gesamtheit zu verstehen und Genauigkeit von geringerer Bedeutung ist. Alle Evaluator*innen hatten MÜ auch für private Zwecke genutzt. Des Weiteren hatten die Evaluator*innen auch Erfahrung mit Übersetzen und Dolmetschen ohne MÜs, wie z. B. T4 und T5, welche privat unter Freund*innen gedolmetscht haben. Keine*r der Evaluator*innen hatte eine Ausbildung als Übersetzer*in oder Dolmetscher*in.

Die Evaluator*innen haben sich alle als Befürworter von MÜ geäußert. Zu den Vorteilen von MÜ nannten sie in erster Linie die Geschwindigkeit, in der eine Übersetzung maschinell erstellt werden kann, aber auch den wirtschaftlichen Nutzen, der daraus entsteht, dass MÜ-Systeme oft kostenlos zur Verfügung gestellt werden. Zu den Nachteilen zählten sie in erster Linie die Übersetzungsqualität, die sie für geringer als die Übersetzungsqualität von menschlichem Übersetzer*innen hielten. T6 und T7 erwähnten auch, dass MÜ das Potential hat, die menschlichen Übersetzer*innen zu ersetzen. Genauer genommen erwähnten sie, dass MÜ als Hilfe beim Übersetzungsprozess genutzt werden kann und dadurch den menschlichen Übersetzer*innen die

Arbeit erleichtert werden kann, was sie als einen Vorteil einstufen; jedoch haben sie auch die Sorge geäußert, dass MÜs den menschlichen Übersetzer*innen die Arbeit vollkommen wegnehmen könnten, was sie als einen Nachteil sehen.

Keine*r der Evaluatord*innen hat bisher an einer Bewertung von Übersetzungsqualität teilgenommen. Die Evaluatord*innen waren sich einig, dass das Format dieser Bewertung angemessen war. Sie fanden es hilfreich, dass die Umfrage von Angesicht zu Angesicht stattgefunden hat und dass die Anweisungen für das Ausfüllen der Umfrage klar und deutlich waren; das Musterblatt mit einem im Voraus bewertetem Set 0 war hilfreich, um die Aufgabe zu verdeutlichen. Evaluatord*innen haben erwähnt, dass sowohl die Umfrage als auch das Interview unterhaltsam waren und haben sich deutlich gefreut zu sehen, ob sie mit ihren Vermutungen in Hinblick auf die zweite Frage recht hatten. Die Bewertung von Angesicht zu Angesicht zu organisieren ist zeitaufwendig, gewährleistet aber, dass alle Evaluatord*innen die Aufgaben richtig verstehen und ermöglicht die Durchführung von Interviews. Dadurch können zwar nur vergleichsmäßig wenige Personen befragt werden, als z. B. mittels einer Online-Umfrage, aber es kann angenommen werden, dass die Resultate maßgeblicher sind.

Die Evaluatord*innen waren nur bedingt sicher, ob sie mit ihren Annahmen in Hinblick auf die zweite Frage richtig lagen. Keine*r der Evaluatord*innen hat sich getraut zu denken, dass sie alle Übersetzungen richtig zugeordnet hatten, aber sie waren sich sicher, dass sie zumindest die Mehrheit richtig zugeordnet hatten. Die Evaluatord*innen haben alle angegeben, dass sie es bei einigen Übersetzungssets sehr leicht gefunden haben, zu entscheiden, welche der Übersetzungen menschlich und welche maschinell erstellt wurden, aber dass sie gleichzeitig mit anderen Sets ihre Entscheidung nicht so leicht treffen konnten. Als ausschlaggebend für die Entscheidung nannten Sie die verwendete Satzstruktur, die Wortwahl, insbesondere in Hinblick auf die Fachterminologie und ob die Übersetzung natürlich klingt. Ein Beispiel der fehlerhaften Übersetzung in Hinblick auf die Fachterminologie, welches erwähnt wurde war das englische Wort „Root“ aus der Textprobe Nummer 2, welches von beiden MÜ-Systemen als „Wurzel“ übersetzt wurde (von Google Translate durchgehend; von DeepL einmal); die korrekte Übersetzung in einem musikwissenschaftlichen Kontext wäre nämlich „Grundton“ gewesen. Google Translates ungeschickter Umgang mit dem Fachterminus *Cantus firmus* im Set 1 wurde von T2 als ein offensichtliches Beispiel für maschinelle Übersetzung genannt. Ein weiteres von T2 genanntes Beispiel ist die im Set 6 präsente Verwendung des Terminus *Triplet*, welcher eine direkte Übernahme aus des englischen Terminus *triplet* darstellt; während *Triplet* in informalem Kontext durchaus verwendet wird, ist *Triole* im formalen Kontext ein passenderer Terminus, welcher sowohl von

beiden Humanübersetzer*innen als auch von DeepL verwendet wurde. T4 nannte als Beispiel einer leichten Entscheidung das Set 3 aufgrund der Tatsache, dass die korrekte Übersetzung von *Minor 7th* und *Major 7th* aus dem Englischen nicht *Mol-Septime* und *Dur-Septime*, sondern *kleine Septime* und *große Septime* lautet; die falsche Übersetzung wurde von beiden MÜ-Systemen und von H1 verwendet, während nur H2 die richtige Übersetzung dieser Termini verwendet hat. All dies deutet darauf hin, dass keines der beiden MÜ-Systeme den Kontext wirklich erkennen und verstehen konnte, weshalb derartige Fehler gelegentlich passierten, jedoch aber auch, dass gleichartige Fehler gelegentlich auch von Humanübersetzer*innen gemacht werden können.

6. Diskussion

In diesem Kapitel werden die im vorangegangenen Kapitel vorgestellten Ergebnisse ausführlich analysiert. Zu Beginn werden die Schlussfolgerungen dieser Arbeit mit einem Rückblick auf die Forschungsfrage und auf die in der Arbeit formulierten Annahmen eingeleitet. Im Anschluss daran werden die während der Studie erkannten Vor- und Nachteile der BWS-Methode retrospektiv beschrieben. Danach werden die Ergebnisse der Studie mit anderen, in den vorangegangenen Kapiteln erwähnten Studien verglichen.

6.1. Zusammenfassung der Umfrage

Diese Arbeit analysiert den Vergleich der Einschätzung von Musikexpert*innen zwischen maschinell und human erstellter Übersetzung. An dieser Stelle muss darauf hingewiesen werden, dass eine solche, klein angelegte empirische Untersuchung keine allgemeingültige Antwort geben kann. Dies ist darauf zurückzuführen, dass die Evaluatord*innenanzahl relativ gering gehalten wurde und dass nur die Sprachenkombination Englisch-Deutsch untersucht wurde. Die Ergebnisse dieser Untersuchung sollen daher nur als eine Tendenz verstanden werden, die sich jedoch deutlich erkennen lässt.

Verallgemeinert kann festgestellt werden, dass beide menschliche Übersetzer*innen besser abgeschnitten haben als die zwei MÜ-Systeme. Genauer genommen wurden die von Google Translate erstellten Übersetzungen als die weitaus schlechtesten empfunden, während der Unterschied zwischen den menschlichen Übersetzungen und DeepL deutlich kleiner ist, jedoch trotzdem nachweisbar ist und fortwährend dazu geführt hat, dass DeepL schlechter bewertet wurde als die menschlichen Übersetzungen.

In Anbetracht der Ergebnisse der Umfrage und der Antworten während des Interviews zeichnet sich eine klare Tendenz in Hinblick auf die Forschungsfrage ab. Fachexpert*innen im Bereich Musik können in den *meisten* Fällen immer noch einen Unterschied zwischen menschlicher und maschineller Übersetzung erkennen, wobei die menschliche Übersetzung fast immer als die bessere betrachtet wird. In Ausnahmefällen wurde jedoch die maschinelle Übersetzung als die bessere Lösung ausgewählt und dabei in allen Fällen außer in einem (T3, Set 4) fälschlicherweise als menschliche Übersetzung eingestuft. Es lässt sich schließen, dass Qualität immer noch mit Menschlichkeit assoziiert wird, aber dass der Unterschied in der Qualität mit der Zeit immer geringer wird und dass heutige MÜs menschlicher und natürlicher klingen als je zuvor.

6.2. Rückblick auf die BWS-Methode

In Kapiteln 2 und 4 wurde die BWS-Methode bereits beschrieben und es wurden die Vor- und Nachteile dieser Methode argumentiert. Nun soll rückblickend beschrieben werden, wie sich die BWS-Methode in dieser Studie bewährt hat.

Der Autor ist der Meinung, dass die Erprobung der Best-Worst-Scaling-Methode im Rahmen dieser wissenschaftlichen Arbeit von großer Bedeutung war, da sie nach dem derzeitigen Kenntnisstand des Autors dieser Arbeit sowohl in der Übersetzungswissenschaft als auch in der Musikwissenschaft nur selten als Bewertungsmethode gewählt wurde. Genauer gesagt sind die einzigen wissenschaftlichen Arbeiten, die die BWS-Methode im Rahmen der Übersetzungswissenschaft verwendeten, die bereits beschriebenen Masterarbeiten von Finetti und Etzlinger. Insgesamt hat sich BWS als eine sehr gute Methode zur Ermittlung intuitiver Präferenzen erwiesen, insbesondere in Verbindung mit einer persönlichen Datenerhebung. Die Evaluat*innen haben die Idee hinter der Methode verstanden und waren in der Lage, Antworten zu geben, mit denen sie zufrieden waren und die ihre Wahrnehmung genau wiedergaben. Keine*r der Evaluat*innen hat sich darüber beschwert, dass die Wahl der Antwort in irgendeiner Weise erzwungen oder eingeschränkt wurde.

Im Gegensatz zu Online-Bewertung dauerte die persönliche Datenerhebung vor Ort etwas länger, was als ein Nachteil betrachtet werden kann, aber ein dazugehöriger Vorteil ist, dass die Evaluat*innen während der persönlichen Datenerhebung genauere Anweisungen bekamen. Das Interview hat gezeigt, dass die persönliche Datenerhebung dazu führt, dass Evaluat*innen die Anweisung zur Ausfüllung des Fragebogens korrekt verstehen, weshalb es keinen Bedarf gab, jegliche Evaluat*innen aus der Studie auszuschließen. Mehrere Evaluat*innen haben während des Interviews zum Ausdruck gebracht, dass sie die Evaluierung als spaßig empfunden haben und haben sich gefreut, zu sehen, ob sie mit ihren Angaben zur Art der Übersetzung richtig lagen.

Ein Vorteil der Methode bestand auch darin, dass die Evaluat*innen keine Vorkenntnisse oder Fachkenntnisse in den Bereichen Übersetzung oder Linguistik benötigten - die Evaluat*innen mussten keine vordefinierten Fehlerkategorien oder Aspekte der Übersetzungen kennen und verstehen, um die Übersetzungen zu bewerten, sondern konnten ihre Qualitätswahrnehmung durch eine Gesamtbewertung zum Ausdruck bringen.

6.3. Vergleich zu früheren Studien

In diesem Unterkapitel werden die Einsichten aus der Umfrage und des Interviews mit den Resultaten bereits beschriebener früherer Studien verglichen. Ein direkter Vergleich ist nicht einfach, weil die in den erwähnten Studien BWS nicht angewandt wurde, sondern andere Methoden. Des Weiteren wird ein direkter Vergleich auch dadurch erschwert, dass die untersuchten Sprachkombinationen sich oft unterscheiden. Es lassen sich jedoch einige von der Untersuchungsmethode und von der Sprachkombination unabhängige Tendenzen in Bezug auf die perzipierte Übersetzungsqualität feststellen.

Im Kapitel 3 wurden mehrere Studien zum Thema Übersetzungsqualität und Vergleich zwischen MÜ und Humanübersetzung beschrieben. Zusammenfassend konnte festgestellt werden, dass heutige NMÜ-Systeme Übersetzungen liefern, welche allgemein als zufriedenstellend bewertet werden, jedoch den menschlichen Übersetzer*innen immer noch unterlegen sind. Die Resultate dieser Studie sind Resultaten früherer Studien in dieser Hinsicht sehr ähnlich.

Die im Kapitel 3 beschriebenen Studien dokumentieren, dass MÜ allgemein als zufriedenstellend bewertet wird und dass sich die Qualität von MÜ mit der Zeit steigert. Dabei hatte der neuronale Ansatz zur von Menschen wahrgenommenen und der durch automatisierte Ansätze messbaren Steigerung der Qualität eine maßgebliche Rolle gespielt. Dies ist auch anhand der Antworten in den Interviews zu dieser Masterarbeit erkennbar. Die befragten Evaluator*innen waren sich einig, dass sie oft unsicher waren, welche der gezeigten Übersetzungen von Menschen und welche von Maschinen erstellt wurde. Sie waren sich auch einig, dass die Qualität der Übersetzungen im Allgemeinen angemessen und relativ hoch war. Die Evaluator*innen haben einige Fehler oder stilistische Mängel erwähnt, welche rückblickend hauptsächlich in den Übersetzungen von Google Translate vorhanden waren. Die Entscheidungen in der Umfrage deuten darauf hin, dass diese Fehler und Mängel oft genug vorkommen und auffallend sind, dass Google Translate als die schlechteste der 4 Übersetzungslösungen zu bewerten ist, zumindest in dem in dieser Masterarbeit untersuchten Themenbereich. Dennoch kann verallgemeinert gesagt werden, dass der perzipierte Qualitätsunterschied nur im Vergleich mit mehreren hochwertigen Übersetzungen zum Vorschein kommt. Die Interviews deuten darauf hin, dass beide untersuchte MÜ-Systeme verständliche Übersetzungen liefern, dass jedoch die von Google Translate gelieferten Übersetzungen Bedarf an Nachbearbeitung aufzeigen, während die von DeepL erstellten Übersetzungen den menschlichen deutlich ähnlicher sind und auch ohne Nachbearbeitung mit menschlichen Übersetzungen mithalten können.

Besonders relevant sind die zwei beschriebenen Studien von Finetti (2020) und Etzlinger (2022). In beiden Fällen hat sich BWS als eine gut geeignete Methode zur Messung der perzipierten Übersetzungsqualität erwiesen. Ein Unterschied zwischen diesen zwei Studien und der dieser Masterarbeit zugrunde liegenden Studie ist, dass die ersteren eine Online-Umfrage verwendet haben, während der dieser Masterarbeit zugrunde liegende Datenerhebung vor Ort stattfand. Es lässt sich schließen, dass BWS eine sehr flexible Methode darstellt, welche auf unterschiedliche Weisen eingesetzt werden kann. Eine weitere wichtige Bemerkung ist, dass DeepLs unbearbeitete Übersetzungen auch in Etzlingers Studie deutlich besser abgeschnitten haben als jene anderer NMÜ-Systeme. Etzlingers Studie zeigte, dass DeepL sogar in der Lage ist, die Übersetzungsqualität von menschlichem Übersetzer*innen zu übertreffen, was im Unterschied zu den dieser Masterarbeit zugrunde liegenden Studie steht. Dies könnte auf unterschiedliche Weisen erklärt werden. Einerseits könnte es sein, dass die Textproben Etzlingers für die ausgewählten menschlichen Übersetzer*innen zu herausfordernd waren. Andererseits könnte es sein, dass es für Musiktheorie weniger hochwertige Ressourcen gibt, als für naturwissenschaftliche Texte und dass daher DeepL bei der Übersetzung von musikbezogenen Texten schlechter abschneidet als bei der Übersetzung von naturwissenschaftlichen Texten. Jedoch kann nicht ausgeschlossen sein, dass unter unterschiedlichen Bedingungen (z. B. andere Textproben) das Resultat nicht anders gewesen wäre. In beiden Fällen wurden die menschlichen Übersetzungen und die von DeepL erstellten Übersetzungen qualitativ als sehr ähnlich eingestuft und zugleich als deutlich hochwertiger als jene von Google Translate.

In Bezug auf das Fachgebiete Musik und Musikwissenschaft ist es wichtig zu betonen, dass, wie aus den durchgeführten Interviews ersichtlich ist, die Verwendung korrekter bzw. inkorrektter musiktheoretischer Fachterminologie einer der wichtigsten Faktoren waren, ob eine Übersetzung als gut oder schlecht bzw. als menschlich oder maschinell bewertet wurde. Die Evaluator*innen waren sich einig, dass einige Übersetzungen die korrekte Terminologie nicht konsistent genug verwendet haben und dass sie diese daher als schlechter und als maschinell erstellt eingestuft haben. Dass das Übersetzen von musiktheoretischen Texten eine große Herausforderung für Menschen darstellt, wurde in Kapitel 3.2. beschriebenen Texten festgestellt; die dieser Masterarbeit zugrunde liegende Umfrage lässt schließen, dass dieses Fachgebiet auch für NMÜ-Systeme herausfordernd ist.

7. Fazit

Das Ziel dieser Masterarbeit war es, herauszufinden, wie Musiker*innen und Musikwissenschaftler*innen die unbearbeiteten deutschen Übersetzungen von hochspezialisierten, Fachgebiet-spezifischen englischen Texten aus dem Bereich der Musik erstellt von neuronalen maschinellen Übersetzungssystemen im Vergleich zur menschlichen Übersetzung bewerten und ob sie konsistent zwischen maschineller und menschlicher Übersetzung unterscheiden können.

Dies wurde untersucht, indem insgesamt sechs Textabschnitte jeweils viermal übersetzt wurden (jeweils von zwei menschlichen Übersetzer*innen, DeepL und Google Translate) und einer Gruppe von insgesamt sieben Fachexpert*innen im Fachgebiet Musik bzw. Musikwissenschaft zur Evaluierung gegeben wurden. Die Evaluierung erfolgte mittels der Best-Worst-Scaling Methode, bei der den Evaluatord*innen Sets mit den vier Übersetzungen gezeigt wurden. Die Evaluatord*innen hatten die Aufgabe, die beste und die schlechteste Übersetzung auszuwählen, wobei sie der besten Übersetzung eine positive und der schlechtesten eine negative Punktzahl zuwies. Das Ergebnis war eine numerische Punktezahl, die zeigt, wie groß der Unterschied zwischen der als am besten und der als am schlechtesten empfundenen Übersetzung ist. Darüber hinaus wurden die Evaluatord*innen gebeten anzugeben, welche der Übersetzungen sie für menschliche Übersetzungen und welche für maschinelle Übersetzungen hielten; die Gesamtzahl der maschinellen oder menschlichen Übersetzungen wurde ihnen nicht mitgeteilt.

Das Ergebnis dieser Untersuchung war, dass die Evaluatord*innen beide menschlichen Übersetzer*innen besser bewerteten als die beiden NMÜ-Systeme, wobei die von DeepL erstellten Übersetzungen deutlich besser bewertet wurden als jene, die von Google Translate erstellt wurden. Die von DeepL erstellten Übersetzungen liegen relativ nahe an den zweitbesten, von einem Menschen erstellten Übersetzungen. Die Gesamtbewertung sowohl der beiden menschlichen Übersetzungen als auch der Übersetzung von DeepL liegt im positiven Bereich, während Google Translate negativ abschneidet. Es ist jedoch wichtig, zu beachten, dass es sich hierbei nicht um einen absoluten Wert handelt, sondern um einen relativen Vergleich nur zwischen den in dieser Masterarbeit untersuchten Übersetzungen. Die Ergebnisse der numerischen Bewertung lassen den Schluss zu, dass es für MÜ-Systeme in bestimmten Fällen möglich ist, das Qualitätsniveau einer menschlichen Übersetzung zu erreichen, auch ohne Nachbearbeitung. Die Ergebnisse in Bezug auf die zweite Frage – ob eine bestimmte Übersetzung von einem Menschen oder einer Maschine angefertigt wurde – zeigen jedoch, dass es immer noch einen deutlichen Unterschied zwischen den beiden Arten von Übersetzungen gibt

und dass Expert*innen in diesem bestimmten Fachbereich maschinelle Übersetzungen recht gut als solche erkennen können.

Diese Ergebnisse sind vergleichbar mit zwei jüngsten Untersuchungen im Rahmen von Masterarbeiten, bei denen ebenfalls BWS verwendet wurde und stimmen auch mit früheren Studien überein, bei denen andere Methoden verwendet wurden. Die BWS-Methode hat sich als sehr nützliches Instrument zur Bewertung und zum Vergleich von Übersetzungen erwiesen, die von verschiedenen menschlichen Übersetzern oder MT-Systemen erstellt wurden.

Was die künftige Forschung im Bereich der Qualität der maschinellen Übersetzung betrifft, so gibt es mehrere Aspekte, die noch untersucht werden sollten. So wäre es beispielsweise wichtig, eine solche Studie mit Sprachen durchzuführen, die zu verschiedenen Sprachfamilien gehören, mit Sprachkombinationen, bei denen sich die Anzahl der vorhandenen Textkorpora signifikant unterscheidet, oder, in Hinblick auf musikbezogene Forschung, einen Vergleich der Übersetzungen von Terminologien, die sich auf nichteuropäische Musiktraditionen beziehen. Die Zahl der Studien, die mit BWS durchgeführt werden könnten, ist aufgrund der unzähligen Kombinationen von Elementen, die untersucht werden müssen, enorm: die Ausgangssprache(n), die Zielsprache(n), menschliche Übersetzer*innen, MÜ-Systeme, Übersetzungen mit oder ohne Vor- oder Nachbearbeitung und verschiedene Fachgebiete. Eine relevante Studie wäre es, einzelne MÜ-Systeme mit mehreren menschlichen Übersetzer*innen zu vergleichen, um festzustellen, ob es besser oder schlechter abschneidet, als die menschlichen Übersetzungen, oder ob es irgendwo in der Mitte liegt.

Es kann davon ausgegangen werden, dass die Ergebnisse von Studien zu diesem Thema auch von einer größeren Gruppe von Evaluatord*innen profitieren würden, da die Studie dadurch repräsentativer wäre. Es wäre möglich, eine Studie in größerem Umfang durchzuführen und die Umfrage auf spezialisierte Gemeinschaften auszudehnen oder mehr Teilnehmer*innen zu gewinnen, indem eine finanzielle Vergütung für die Teilnahme angeboten wird.

Die neuronale maschinelle Übersetzung ist eine Technologie, die unbestreitbar erheblich an Bedeutung gewonnen hat und sich in erstaunlicher Weise weiterentwickelt. Und was vielleicht noch wichtiger ist: die Zukunft sieht vielversprechend aus. Angesichts der Tatsache, dass die unbearbeitete Übersetzung von NMT-Systemen in bestimmten Situationen und unter bestimmten Bedingungen ein Qualitätsniveau erreichen kann, das mit dem der menschlichen Übersetzung vergleichbar ist, bleibt abzuwarten, ob die NMT-Systeme so weit verbessert werden können, dass diese Art von Ergebnis durchgängig auftritt. Anhand neuerer Forschungsarbeiten ist es durchaus realistisch anzunehmen, dass das sogar in nur wenigen Jahren der Fall sein könnte.

8. Literaturverzeichnis

- Abe, Hiroya; Kubo, Takahiro; Tsuge, Takahiro & Yamano, Hiroya (2018). Understanding island residents' anxiety about impacts caused by climate change using Best–Worst Scaling: a case study of Amami islands, Japan. *Sustainability science* 14, 131–138.
- Ahrenberg, Lars (2017). Comparing Machine Translation and Human Translation: A Case Study. In: Temnikova, Irina; Orasan, Constantin; Corpas Pastor, Gloria & Vogel, Stephan (Hg.) *Proceedings of the Workshop Human-Informed Translation and Interpreting Technology*. Association for Computational Linguistics, RANLP, Shoumen, Bulgaria, 21–28.
- Akman, Varol; Cicekli, Ilyas & Saygin, Ayse Pinar (2000). Turing Test: 50 Years Later. *Minds and Machines* 10, 463–518.
- Alam, Firoj; Chowdhury, Shammur Absar; Hasan; Md. Arif & Khan, Naira (2019). Neural vs Statistical Machine Translation: Revisiting the Bangla-English Language Pair. *International Conference on Bangla Speech and Language Processing*. Bangladesch: Bangla Language Processing, 27–28.
- Alkhouli, Tamer; Bahar, Parnia; Graça, Miguel; Guta, Andreas; Ney, Hermann; Peter, Jan–Thorsten; Rosendahl, Jan & Rossenbach, Nick (2017). The RWTH Aachen University English–German and German–English Machine Translation System for WMT 2017. In: Bojar, Ondřej; Buck, Christian; Chatterjee, Rajen; Federmann, Christian; Guillou, Liane; Haddow, Barry; Huck, Matthias; Yepes, Antonio Jimeno; Névéal, Aurélie; Neves, Mariana; Pecina, Pavel; Popel, Martin; Koehn, Philipp; Monz, Christof; Negri, Matteo; Post, Matt; Specia, Lucia; Verspoor, Karin; Tiedemann, Jörg & Turchi, Marco (Hg.) *Proceedings of the Conference on Machine Translation (WMT), Volume 2: Shared Task Papers*. Copenhagen: Association for Computational Linguistics, 358–365.
- Almaghout, Hala; Doherty, Stephen & Gaspari, Federico (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives Studies in Translatology* 23 (3), 333–358.
- Almén, Byron & Pearsall, Edward (2006). *Approaches to Meaning in Music*. Bloomington, USA: Indiana University Press.
- ALPAC (1966). *Language and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington D.C.: National Academy of Sciences.

- Arratibel, Lierni Garmendia; Borzeshi, Ehsan Zare; Piccardi, Massimo & Unanue, Inigo Jauregi (2018): English–Basque Statistical and Neural Machine Translation. In: Calzolari, Nicoletta; Choukri, Khalid; Cieri, Christopher; Declerck, Thierry; Goggi, Sara; Hasida, Koiti; Isahara, Hitoshi; Maegaard, Bente; Mariani, Joseph; Mazo, H el ene; Moreno, Asuncion; Odijk, Jan; Piperidis, Stelios; Tokunaga, Takenobu (Hg.) *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Paris: European Language Resources Association (ELRA), 880–885.
- Arnold, D.J. (1994). *Machine translation. An introductory guide*. Manchester: NCC Blackwell
- Arun, Abhishek; Hoang, Hieu & Koehn, Philipp (2008). Towards better Machine Translation Quality for the German–English Language Pairs. In: Callison-Burch, Chris; Koehn, Philipp; Monz, Christof; Schroeder, Josh & Fordyce, Cameron Shaw (Hg.) *Proceedings of the Third Workshop on Statistical Machine Translation*. Columbus, Ohio, USA: Association for Computational Linguistics, 139–142.
- Axin, William; Groves, Robert & Link, Cynthia (2011). Responsive Survey Design, Demographic Data Collection, and Models of Demographic Behavior. *Demography* 48, 1127–1149.
- Banerjee, Satanjeev & Lavie, Alon (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In: Goldstein, Jade; Lavie, Alon; Lin, Chin-Yew & Voss, Clare (Hg.) *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, Ann Arbor: Association for Computational Linguistics, 65–72.
- Bentivogli, Luisa; Bisazza, Arianna; Cettolo, Mauro & Federico, Marcello (2016). Neural versus phrase–based machine translation quality: A case study. In: Carreras, Xavier; Duh, Kevin & Su, Jian (Hg.) *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics: Austin, Texas, USA, 257–267.
- Bergman, Margo; Brownstein, Megumu; Cole, Kate; Kirchhoff, Katrin & Turner, Anne (2014). A Comparison of Human and Machine Translation of Health Promotion Materials for Public Health Practice: Time, Costs, and Quality. *Journal of Public Health Management and Practice* 20 (5), 523–529.
- Blain, Fr ed eric; Ive, Julia & Specia, Lucia (2018). deepQuest: A Framework for Neural-based Quality Estimation. In: Bender, Emily; Derczynski, Leon & Isabelle, Pierre (Hg.) *Proceedings of the 27th International Conference on Computational Linguistics*. Stroudsburg, Pennsylvania: Association for Computational Linguistics, 3146–3157.

- Bouillon, Pierrette; Girletti, Sabrina & Volkart, Lise (2018). Statistical vs. Neural Machine Translation: A Comparison of MTH and DeepL at Swiss Post's Language Service. In: Esteves-Ferreira, João; Macan, Juliet Margaret; Mitkov, Ruslan & Stefanov, Olaf-Michael (Hg.) *Proceedings of the 40th Conference Translating and the Computer*. Geneva: Editions Tradulex, 145–150.
- Bridges, John; Dumanis, Sonya; Fureman, Brandy; Janse, Sarah; Huwig, Tanya & Hyman, Sarah (2019). Patient and caregiver preferences for the potential benefits and risks of a seizure forecasting device: A best–worst scaling. *Epilepsy & Behavior* 96, 183–191.
- Briggs, Neil (2018). Neural Machine Translation Tools in the Language Learning Classroom: Students' Use, Perceptions, and Analyses. *JALT CALL Journal* v14, 2–24.
- Brown, Eileen (2016). Systran launches neural machine translation engine in 30 languages. <https://www.zdnet.com/article/systran-launches-neural-machine-translation-engine-in-30-languages> (Stand: 21.09.2022)
- Burstein, L. Poundie & Straus, Joseph N (2015). *Concise Introduction to Tonal Harmony and Student Workbook*. New York, USA: W. W. Norton & Company
- Cai, Zhiqiang; Graesser, Arthur & Li, Haiying (2014). Comparison of Google Translation with Human Translation. In: Boonthum-Denecke & Eberle, William (Hg.) *Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference*. USA: AAAI, 190–195.
- Coldewey, Devin & Lardinois, Frederic (2017). DeepL: schools other online translators with clever machine learning. <https://techcrunch.com/2017/08/29/deepl-schools-other-online-translators-with-clever-machine-learning> (Stand: 22.06.2021).
- Costa-Jussà, Marta; Farrús, Mireia; Fonollosa, José A. Rodríguez & Mariño Acebal, José B. (2012). Study and comparison of rule-based and statistical Catalan-Spanish machine translation systems. *Computing and Informatics* 31, 245–270.
- Craig, Jonathan; Howard, Kirsten; Howell, Martin; Wong, Germaine; Rose, John & Tong, Allison (2017). Patient Preferences for Outcomes After Kidney Transplantation: A Best–Worst Scaling Survey. *Transplantation* 101 (11), 2765–2773.
- Chanan, Michael (2013). Preface. In: Minors, Helen Julia (Hg.) *Music, Text and Translation*. London: Bloomsbury Publishing, 12–14.
- Chitu, Alex (2007). Google Switches to Its Own Translation System. <https://googlesystem.blogspot.com/2007/10/google-translate-switches-to-googles.html> (Stand: 20.09.2022)

- CMU (2011). Meteor. <https://www.cs.cmu.edu/~alavie/METEOR/> (Stand: 21.09.2022)
- Cugny, Laurent (2019). *Analysis of Jazz: A Comprehensive Approach*. Mississippi, USA: University Press. Ersterscheinung 2009, *Analyser le jazz*, Paris, Éditions Outre Mesure, aus dem Französischen übersetzt von Bérengère Mauduit.
- Curtiss, Marie Joy (1972). Research Project to Provide Materials for Teaching Asian Music in United States Public Schools and Colleges. *Bulletin of the Council for Research in Music Education* 27, 1–6.
- Desblache, Lucile (2018). Translation of music. In: Chan, Sin-wai (Hg.) *An Encyclopedia of Practical Translation and Interpreting*. The Chinese University of Hong Kong Press, 297 – 324.
- Dorr, Bonnie; Madnani, Nitin; Snover, Matthew & Schwartz, Richard (2009). Fluency Adequacy or HTER? Exploring Different Human Judgments with a Tunable MT Metric. In: Callison-Burch, Chris; Koehn, Philipp; Monz, Christof & Schroeder, Josh (Hg.) *Proceedings of the Fourth Workshop on Statistical Machine Translation*. Athen: Association for Computational Linguistics, 259–268.
- Dorr, Bonnie; Makhoul, John; Micciulla, Linnea; Schwartz, Rich & Matthew Snover (2006). A Study of Translation Edit Rate with Targeted Human Annotation. In: Callison-Burch, Chris; Koehn, Philipp; Monz, Christof; Post, Matt; Soricut, Radu & Specia, Lucia (Hg.) *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. Cambridge, Massachusetts, USA: Association for Machine Translation in the Americas, 223–231.
- Dorr, Bonnie; Olive, Joseph; McCary, John & Christianson, Caitlin (2011). Machine Translation Evaluation and Optimization. In: Christianson, Caitlin; McCary, John & Olive, Joseph (Hg.) *Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation*. London: Springer Science, 745-843.
- Dwojak, Tomasz & Jassem, Krzysztof (2019) Statistical versus neural machine translation – a case study for a medium size domain-specific bilingual corpus. *Poznań Studies in Contemporary Linguistics* 55(2) 491–515.
- Engelbach, Susanne (2007). *Beethovens 7. Symphonie - eine Analyse*. München: GRIN Verlag.
- Etzlinger, Maren (2022). *Neural machine translation versus human translation in the field of natural science*. Universität Wien
- Finetti, Fiorenza (2020). *Pre-Editing als Vorschlag zur Verbesserung Neuronaler Maschineller Übersetzung*. Universität Wien.

- Fontelles Rodríguez, Vicent Lluís (2021). *Music didactics - Types, History, Methodologies (models) and other notes for class*. University of València.
- Franzon, Johan; Mateo, Marta; Orero, Pilar und Susam–Sarajeva Şebnem (2008). Translation and Music: A General Bibliography. *The Translator* 14 (2), 453–460.
- Frith, Simon (2004). Introduction. In: Simon Frith (Hg.) *Popular Music. Volume IV. Music and Identity, Critical Concepts in Media and Cultural Studies Series*. London und New York: Routledge, 1–3.
- Gaspari, Federico; Niño, Ana & Somers, Harold (2006). Detecting Inappropriate Use of Free Online Machine Translation by Language Students. A Special Case of Plagiarism Detection. In: Hansen, Viggo & Maegaard, Bente (Hg.) *Proceedings of the 11th Annual conference of the European Association for Machine Translation, June 19–20*. Oslo: European Association for Machine Translation.
- Giannini, Juri (2014). Die italienischen Übersetzungen der musiktheoretischen Terminologie von Arnold Schönberg – zwei Fallbeispiele. In: Werner Hasitschka (Hg.) *Performing Translation. Schnittstellen zwischen Kunst, Pädagogik und Wissenschaft*. Wien: Löcker, 86–105.
- Granado, Paola & Miller, Roger (2018). Some examples from a corpus investigation of potentially confusing polysemy and false friends in the terminology of music, centred on english and french, with equivalents in five other languages. *Rasprave* 44 (2), 395– 410.
- Grimm, Hartmut & Wald-Fuhrmann, Melanie (2017). *Lexikon Schriften über Musik*. Stuttgart: Gemeinschaftsausgabe der Verlage Bärenreiter, Kassel und J. B. Metzler.
- Google Translate (2020). Improve Translate. <https://translate.google.com/about/contribute> (Stand: 21.09.2022)
- González Rey, F. (2016). Advancing the topics of social reality, culture, and subjectivity from a cultural–historical standpoint: Moments, paths, and contradictions. *Journal of Theoretical and Philosophical Psychology* 36, 175–189.
- Guerberof Arenas, Ana (2020). Pre-editing and post-editing. In: Angelone, Erik; Ehrensberger-Dow, Maureen & Massey, Gary (Hg.) *The Bloomsbury Companion to Language Industry Studies*. London: Bloomsbury Academic, 333–360.
- Gudergan, Siegfried; Islam, Towhidu, Louviere, Jordan; Lings, Ian und Flynn, Terry (2013). An introduction to the application of (case 1) best–worst scaling in marketing research. *International journal of research in marketing* 30 (3), 292–303.

- Harnoncourt, Nikolaus (1995). *Baroque Music Today: Music As Speech. Ways to a New Understanding of Music*. Portland, Oregon, USA: Amadeus Press. Ersterscheinung 1982, *Musik als Klangrede*, Salzburg und Wien, Residenz Verlag, aus dem Deutschen übersetzt von Mary O'Neill.
- House, Juliane (1997). *Translation quality assessment: a model revisited*. Tübingen: Narr.
- Hunter RWC (2017). Writing about Music. <http://www.hunter.cuny.edu/rwc/repository/files/WAC/writing-about-music.pdf> (Stand: 22.06.2021).
- Hutchins, W. John (2000). *Early years in machine translation. Memoirs and biographies of pioneers*. Amsterdam: Benjamins.
- Hutchins, W. John & Somers, L. Harold (1992). *An introduction to machine translation*. London: Acad. Press.
- Kadrić, Mira; Kaindl, Klaus & Kaiser–Cooke, Michèle (2009). Translation als kultureller Transfer. In: Kadrić, Mira (Hg.) *Translatorische Methodik*. Wien: Facultas Verlags- und Buchhandels AG, 27–44.
- Kenny, Dorothy (2018). Machine Translation. In: Rawling, Piers und Wilson, Philip (Hg.) *The Routledge Handbook of Translation and Philosophy*. London: Routledge, 428–455.
- Kiš Žuvela, Sanja (2018). Problems of basic Croatian music theory terminology with special emphasis on literature intended for teaching music theory. *Arti Musices* 49, 197–200.
- Koehn, Philipp (2020). *Neural machine translation*. Cambridge: Cambridge University Press.
- Krenz, Michael & Ramlow, Markus. (2008). *Maschinelle Übersetzung und XML im Übersetzungsprozess. Prozesse der Translation und Lokalisierung im Wandel*. Berlin: Frank & Timme.
- Laitz, Steven (2011). *The Complete Musician: An Integrated Approach To Tonal Theory, Analysis, And Listening*. 3. Aufl. New York, USA: Oxford University Press.
- Lauscher, Susanne (2006). Translatqualität – ein Konsens. In: Schippel, Larisa (Hg.) *Übersetzungsqualität: Kritik – Kriterien – Bewertungshandeln*. Berlin: Frank & Timme, 55–75.
- Läubli, Samuel; Sennrich, Rico & Volk, Martin (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. In: Chiang, David; Hockenmaier, Julia; Riloff, Ellen & Tsujii, Jun'ichi (Hg.) *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brüssel: Association for Computational Linguistics, 4791–4796.

- LingoHub (2020). Five great Machine Translation Engines. <https://lingohub.com/blog/2018/11/find-good-machine-translation-engines> (Stand: 22.06.2021).
- Lotz, Susan & Van Rensburg, Alta (2014). Translation technology explored: Has a three-year maturation period done Google Translate any good? *Stellenbosch papers in linguistics plus* 43, 235–259.
- Louviere, Jordan J.; Flynn, Terry N. & Marley, Anthony Alfred John (2015). *Best Worst Scaling: Theory, Methods and Applications*. Cambridge University Press.
- Low, Peter (2003). Translating poetic songs. An attempt at a functional account of strategies. *Target. International Journal of Translation Studies* 15 (1), 91–110.
- Marley, Anthony Alfred John (2009). The Best-Worst Method for the Study of Preferences: Theory and Application. *International Journal of Psychology* 43, 168–186.
- Merkert, Pina (2021). Google: Translate-KI übersetzt dank selbst erlernter Sprache. <https://www.heise.de/newsticker/meldung/Google-Translate-KI-uebersetzt-dank-selbst-erlernter-Sprache-3502351.html> (Stand: 22.06.2021).
- Minors, Helen Julia (2021). Music and multimodal translation. In: Marcelo Winitzer, Gisela & Martin de Leon (Hg.) *En más de un sentido: Multimodalidad y construcción e interpretación*. Universidad de Las Palmas de Gran Canaria, 165–184
- Minors, Helen Julia (2020). Opera and intercultural musicology as modes of translation. In: Chan, Kelly Kar Yue & Serban, Adriana (Hg.) *Opera in Translation. Unity and diversity*. Amsterdam: Benjamins Translation Library, 13–33.
- Montella, Clara & Monti, Johanna (2015). About adequacy, equivalence and translatability in human and Machine Translation. *Conference: New Horizons in Translation and Interpreting Studies, Malaga*. Geneva: Editions Tradulex, 690–701.
- Mounin, Georges (1967). *Die Übersetzung. Geschichte, Theorie, Anwendung*. München: Nymphenburger Verlagshandlung. <https://ai.googleblog.com/2006/04/statistical-machine-translation-live.html> (Stand: 22.06.2021).
- Nida, Eugene Albert (1964). *Toward a science of translating: with special reference to principles and procedures involved in Bible translating*. Leiden: Brill.
- Och, Franz (2006). Statistical machine translation live.
- Open Music Theory (2014). Composing a cantus firmus. <http://openmusictheory.com/cantusFirmus.html> (Stand: 9.9.2022).

- Papineni, Kishore; Roukos, Salim; Ward, Todd & Zhu, Wei-Jing (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In: Charniak, Eugene; Isabelle, Pierre & Lin, Dekang (Hg.) *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) July 2002*. Pennsylvania: Association for Computational Linguistics, 311–318.
- Paetzold, Gustavo Henrique; Scarton, Carolina & Specia Lucia (2018). *Quality Estimation for Machine Translation*. San Rafael: Morgan & Claypool Publishers.
- Poibeau, Thierry (2017). *Machine Translation*. Cambridge, Massachusetts: The MIT Press.
- Popović, Maja (2018). Error Classification and Analysis for Machine Translation Quality Assessment. In: Moorkens, Joss; Sheila Castilho; Federico Gaspari und Stephen Doherty (Hg.) *Translation Quality Assessment: From Principles to Practice. Machine Translation: Technologies and Applications*. Springer, 129–159
- qe-team (2017). MARMOT – Word-Level Quality Estimation and so much more. <https://github.com/qe-team/marmot> (Stand: 22.06.2021).
- Quah, Chiew Kin (2006). *Translation and Technology*. Chippenham: Antony Rowe Ltd.
- Raffel, Burton (1964). Music, Poetry, and Translation. *The Antioch Review* 24 (4), 453–461.
- Reiss, Katharina & Vermeer, Hans J. (1984). *Grundlegung einer allgemeinen Translationstheorie*. Berlin, Boston: De Gruyter.
- Reiss, Katharina (1986). *Möglichkeiten und Grenzen der Übersetzungskritik*. 3. Aufl. München: Max Hueber Verlag.
- Schäfer, Falko (2002). *Die maschinelle Übersetzung von Wirtschaftsfachtexten. Eine Evaluierung anhand des MÜ-Systems der EU-Kommission, SYSTRAN, im Sprachenpaar Französisch-Deutsch*. Frankfurt am Main: Peter Lang GmbH.
- Schäffner, Christina (1998). *Translation and Quality Current Issues. in Language and Society*. North Somerset, England: Multilingual Matters.
- Schmitt, Peter A. (1997). Evaluierung von Fachübersetzungen. In: Schmidt, Heide und Wotjak, Gerd (Hg.) *Modelle der Translation: Festschrift für Albrecht Neubert*. Frankfurt am Main: Vervuert Verlag, 301–332.
- Smartcat (2021). Baidu Translate. <https://www.smartcat.com/integrations/baidu-translate> (Stand: 21.09.2022)
- Spencer-Oatey, H. (2008). *Culturally Speaking. Culture, Communication and Politeness Theory*. London: Continuum.

- Susam–Sarajeva, Şebnem (2008). Translation and Music Changing Perspectives, Frameworks and Significance. *The Translator* 14 (2), 187–200.
- SYSTRAN, 2013. A pioneer and global leader in translation solutions. <https://www.systransoft.com/systran> (Stand: 20.09.2022)
- SYSTRAN, 2018. SYSTRAN: 50 Years of MT Innovation <https://www.systransoft.com/systran/translation-technology/systran-50-years-of-mt-innovation/> (Stand: 20.09.2022)
- Taylor, Christopher (2006). Which Strategy for Which Text? Translation Strategies for Languages for Special Purposes. In: Gotti, Maurizio & Šarcevi Susan (Hg.) *Insights into Specialized Translation*. Bern: Peter Lang Verlag.
- Trujillo, Arturo (1999). *Translation engines. Techniques for machine translation*. Berlin: Springer.
- Vail, Mark (2014). *The Synthesizer. A Comprehensive Guide to Understanding, Programming, Playing, and Recording the Ultimate Electronic Music Instrument*. New York, USA: Oxford University Press.
- Van Slype, Georges (1979). *Critical study of methods for evaluating the quality of machine translation. Final Report*. Brüssel: Bureau Marcel van Dijk.
- Wiener, Oliver (2002). Übersetzung musiktheoretischer Traktate als Phänomen des Kulturtransfers im 18. Jahrhundert. In: Plachta, Bodo & Woesler Winfried (Hg.) *Edition und Übersetzung*. Berlin: Walter de Gruyter Verlag, 405–418.
- Yu, Qiufen (2014). Understanding the Impact of Culture on Interpretation: A Relevance Theoretic Perspective. *Intercultural Communication Studies* 23 (3), 83–102.

9. Anhang 1

Im folgenden Anhang stehen Ausgangstexte, welche als Textproben gedient haben, incl. ihrer Quellen.

Ausgangstext 1:

Exercises in strict voice-leading, or species counterpoint, begin with a single, well formed musical line called the cantus firmus (fixed voice, or fixed melody; pl. cantus firmi). Cantus firmus composition gives us the opportunity to engage the following fundamental musical traits: smoothness, independence and integrity of melodic lines, variety, and motion (towards a goal).

Quelle: <http://openmusictheory.com/cantusFirmus.html> (Stand 9.9.2022)

Ausgangstext 2:

Sequences often are labeled by the pattern of their roots. Of the numerous possible harmonic patterns that may govern a sequence, some are particularly common. The one used most often is the descending fifth sequence—also known as the circle of fifths progression—involving chords whose roots alternately move up by a fourth and down by a fifth.

Quelle: Burstein, L. Poundie & Straus, Joseph N (2015). *Concise Introduction to Tonal Harmony and Student Workbook*. New York, USA: W. W. Norton & Company

Ausgangstext 3:

We label seventh chords by their two most audible features: the type of triad (major, minor, diminished) and the type of seventh above the root (major, minor, diminished). Composers use five important types of seventh chords, though, like the triad types, they are not used with equal frequency. Seventh chords are named according to the quality of the triad and the size of the seventh.

Quelle: Laitz, Steven (2011). *The Complete Musician: An Integrated Approach To Tonal Theory, Analysis, And Listening*. 3. Aufl. New York, USA: Oxford University Press.

Ausgangstext 4:

In attempting to come to grips with the Italian tempo markings of the 17th and 18th Centuries, it must be kept in mind that most of these words (e.g. allegro, largo, presto) were and are part of the Italian vernacular and were used by Italian composers literally, rather than as musical terms. Thus allegro means merry, cheerful, rather than fast; only when the particular nature of gaiety requires a certain tempo does allegro indirectly become a tempo indication. In general these marks should be understood as indications of emotional states, with the absolute tempo deriving from the context.

Quelle: Harnoncourt, Nikolaus (1995). *Baroque Music Today: Music As Speech. Ways to a New Understanding of Music*. Portland, Oregon, USA: Amadeus Press. Ersterscheinung 1982, *Musik als Klangrede*, Salzburg und Wien, Residenz Verlag, aus dem Deutschen übersetzt von Mary O'Neill.

Ausgangstext 5:

Play and sustain a chord on a keyboard that has a built-in arpeggiator or is connected to one and, instead of the chord, you'll hear individual notes playing back one at a time at a tempo determined by a system clock. Depending on the arpeggiator's capabilities, the notes may play in ascending or descending order, both ascending and descending, sequentially in the order that you played the keys, randomly, repeated an octave, two, three, or more higher, and sequencing through different sounds on successive notes. An arpeggiator can be inspiring to work with and often is more interactive than a typical sequencer.

Quelle: Vail, Mark (2014). *The Synthesizer. A Comprehensive Guide to Understanding, Programming, Playing, and Recording the Ultimate Electronic Music Instrument*. New York, USA: Oxford University Press.

Ausgangstext 6:

According to André Hodeir, the characteristic feature of rhythm in jazz lies in an unequal division of time that materializes in playing eighth notes unequally: two eighth notes are actually played as a quarter note and an eighth note in a triplet (long-short triplet pair). Time is thus divided in a ternary manner (triplet subdivision).

Quelle: Cugny, Laurent (2019). *Analysis of Jazz: A Comprehensive Approach*. Mississippi, USA: University Press. Ersterscheinung 2009, *Analyser le jazz*, Paris, Éditions Outre Mesure, aus dem Französischen übersetzt von Bérengère Mauduit.

10. Anhang 2

In diesem Anhang befinden sich die 4 Übersetzungen, in der Reihenfolge, in der Sie in der Umfrage erscheinen. Die zwei menschlichen Übersetzer*innen wurden anonymisiert H1 und H2 benannt. Den Evaluators*innen waren die Übersetzer*innen der 4 Texte unbekannt – dieses Lösungsblatt bekamen sie erst nach dem Ausfüllen des Fragebogens und des Interviews.

Set 1:

<p>DeepL: Übungen in strenger Stimmführung oder Artkontrapunkt beginnen mit einer einzelnen, gut geformten musikalischen Linie, die Cantus firmus (feste Stimme oder feste Melodie; pl. cantus firmi) genannt wird. Die Komposition eines Cantus firmus gibt uns die Möglichkeit, die folgenden grundlegenden musikalischen Eigenschaften zu trainieren: Geschmeidigkeit, Unabhängigkeit und Integrität der melodischen Linien, Vielfalt und Bewegung (auf ein Ziel hin).</p>	<p>H1: Etüden in strenger Stimmführung oder strengem Kontrapunkt beginnen mit einer einzelnen, wohlgeformten Melodielinie namens Cantus firmus (feststehender Gesang oder feststehende Melodie; Pl. Cantus firmi). Die Cantus firmus-Komposition ermöglicht es uns, folgende grundlegende musikalische Merkmale zu aktivieren: Gleichmäßigkeit, Unabhängigkeit und Integrität oder melodische Linien, Varietät und Bewegung (zu einem Ziel hin).</p>
<p>Google Translate: Übungen in strikter Stimmführung oder Artenkontrapunkt beginnen mit einer einzigen, wohlgeformten Musiklinie, die als Cantus Firmus (feste Stimme oder feste Melodie; Pl. Cantus Firmi) bezeichnet wird. Die Komposition von Cantus firmus gibt uns die Möglichkeit, die folgenden grundlegenden musikalischen Merkmale zu berücksichtigen: Geschmeidigkeit, Unabhängigkeit und Integrität oder melodische Linien, Vielfalt und Bewegung (in Richtung eines Ziels).</p>	<p>H2: Übungen zur Stimmführung oder Kontrapunkte verschiedener Gattungen beginnen mit einer einzelnen, ausgestalteten Melodie die „Cantus firmus“ („fester Gesang“ oder „feste Melodie“; pl. Cantus firmi) genannt wird. Die Komposition eines Cantus firmus gibt uns die Möglichkeit, uns mit wesentlichen musikalischen Eigenschaften, wie Wohlklang, Eigenständigkeit und Integrität oder melodischen Linien, Varietät und Bewegung (zu einem Zweck), zu befassen.</p>

Set 2:

<p>Google Translate: Sequenzen sind oft durch das Muster ihrer Wurzeln gekennzeichnet. Von den zahlreichen möglichen harmonischen Mustern, die eine Sequenz steuern können, sind einige besonders häufig. Die am häufigsten verwendete ist die absteigende fünfte Sequenz - auch als Quintenkreis bekannt - mit Akkorden, deren Wurzeln sich abwechselnd um eine vierte und eine fünfte nach unten bewegen.</p>	<p>DeepL: Sequenzen werden oft durch das Muster ihrer Wurzeln gekennzeichnet. Von den zahlreichen möglichen harmonischen Mustern, die eine Sequenz bestimmen können, sind einige besonders häufig. Das am häufigsten verwendete ist die absteigende Quintenfolge - auch als Quintenzirkel bekannt -, bei der Akkorde verwendet werden, deren Grundton sich abwechselnd um eine Quarte nach oben und um eine Quinte nach unten bewegt.</p>
<p>H2: Sequenzen werden nach dem Muster ihrer Grundtöne unterschieden. Von den verschiedenen harmonischen Mustern, die eine Sequenz dominieren können, sind einige besonders gängig. Die meistverwendete Sequenz ist die Quintfallsequenz, auch Quintenzirkel genannt, bei deren Akkordfolge sich die Grundtöne in Quintschritten abwärts- und Quartschritten aufwärtsbewegen.</p>	<p>H1: Sequenzen werden oft nach dem Muster ihrer Grundtöne benannt. Es gibt zahlreiche harmonische Motive, die eine Sequenz lenken können. Einige davon werden besonders oft verwendet. Am häufigsten wird die Quintfallsequenz – oder Quintschrittsequenz – verwendet, bei welcher Akkorde zum Einsatz kommen, deren Grundtöne sich in einem stetigen Wechsel aus aufsteigenden Quartern und fallenden Quinten befinden.</p>

Set 3:

H2: Wir beschreiben Septakkorde durch ihre beiden am deutlichsten heraushörbaren Eigenschaften: die Art des Dreiklangs (Dur, Moll, vermindert) und der Septime (groß, klein, vermindert). Komponistinnen und Komponisten verwenden fünf wichtige Arten von Septakkorden, obwohl, ähnlich wie bei den Dreiklangarten, nicht alle gleich häufig verwendet werden. Septakkorde werden nach der Dreiklangqualität und der Art der Septime benannt.

DeepL: Wir bezeichnen Septakkorde nach ihren beiden am besten hörbaren Merkmalen: dem Typ des Dreiklangs (Dur, Moll, vermindert) und dem Typ der Septime über dem Grundton (Dur, Moll, vermindert). Komponisten verwenden fünf wichtige Typen von Septakkorden, obwohl sie, wie die Dreiklangstypen, nicht mit gleicher Häufigkeit verwendet werden. Septakkorde werden nach der Qualität des Dreiklangs und der Größe der Septime benannt.

H1: Wir benennen Septakkorde nach den beiden Merkmalen, die am besten hörbar sind: die Art des Dreiklangs (Dur, Moll, vermindert) und die Art der Septime über dem Grundton (Dur, Moll, vermindert). Komponisten verwenden fünf wesentliche Arten von Septakkorden, welche allerdings, wie auch die Arten des Dreiklangs, nicht gleich häufig zum Einsatz kommen. Septakkorde werden nach der Qualität des Dreiklangs und der Größe der Septime benannt.

Google Translate: Wir kennzeichnen siebte Akkorde anhand ihrer zwei am besten hörbaren Merkmale: der Art der Triade (Dur, Moll, vermindert) und der Art der siebten über dem Grundton (Dur, Moll, vermindert). Komponisten verwenden fünf wichtige Arten von Septakkorden, die jedoch wie die Triadenarten nicht mit gleicher Häufigkeit verwendet werden. Siebte Akkorde werden nach der Qualität der Triade und der Größe der siebten benannt.

Set 4:

DeepL: Bei dem Versuch, die italienischen Tempobezeichnungen des 17. und 18. Jahrhunderts in den Griff zu bekommen, muss man sich vor Augen halten, dass die meisten dieser Wörter (z. B. allegro, largo, presto) Teil der italienischen Umgangssprache waren und sind und von den italienischen Komponisten wörtlich und nicht als musikalische Begriffe verwendet wurden. So bedeutet allegro eher fröhlich, heiter als schnell; nur wenn die besondere Art der Fröhlichkeit ein bestimmtes Tempo erfordert, wird allegro indirekt zu einer Tempobezeichnung. Im Allgemeinen sollten diese Zeichen als Hinweise auf emotionale Zustände verstanden werden, wobei sich das absolute Tempo aus dem Kontext ergibt.

H2: Beim Versuch, sich mit den italienischen Tempobezeichnungen des 17. und 18. Jahrhunderts auseinanderzusetzen, muss berücksichtigt werden, dass die meisten der dafür verwendeten Wörter (beispielsweise allegro, largo, presto) Teil des italienischen Sprachgebrauchs waren und sind und daher von den italienischen Komponistinnen und Komponisten eher wörtlich anstatt als musikalischer Begriff gebraucht wurden. Dementsprechend meint „Allegro“ eher munter und fröhlich als schnell. Nur wenn eine gewisse Art von Heiterkeit ein spezifisches Tempo benötigt, wird „Allegro“ indirekt zur Tempobezeichnung. Generell sollten diese Bezeichnungen als Ausdrucksbezeichnungen verstanden werden, bei denen sich das tatsächliche Tempo aus dem Kontext ergibt.

Google Translate: Bei dem Versuch, die italienischen Tempomarkierungen des 17. und 18. Jahrhunderts in den Griff zu bekommen, muss berücksichtigt werden, dass die meisten dieser Wörter (z. B. Allegro, Largo, Presto) Teil der italienischen Umgangssprache waren und sind und von Italienern verwendet wurden. Komponisten buchstäblich und nicht als musikalische Begriffe. Allegro bedeutet also eher fröhlich, fröhlich als schnell; Nur wenn die besondere Natur der Fröhlichkeit ein bestimmtes Tempo erfordert, wird Allegro indirekt zu einer Tempoanzeige. Im Allgemeinen sollten diese Zeichen als Hinweise auf emotionale Zustände verstanden werden, wobei sich das absolute Tempo aus dem Kontext ergibt.

H1: Bei dem Versuch, die italienischen Tempobezeichnungen des 17. und 18. Jahrhunderts zu verstehen, muss man sich vor Augen halten, dass die meisten dieser Bezeichnungen (z.B. „allegro“, „largo“, „presto“) Teil der italienischen Umgangssprache waren und noch immer sind und von italienischen Komponisten eher wörtlich und nicht als musikalische Begriffe verwendet wurden. So bedeutet „allegro“ eher „fröhlich“ und „munter“ als „schnell“; nur wenn die besondere Art der Fröhlichkeit ein bestimmtes Tempo erfordert, wird „allegro“ indirekt zu einer Tempoangabe. Im Allgemeinen sollten diese Bezeichnungen als Zeichen emotionaler Zustände verstanden werden, wobei sich das absolute Tempo aus dem Kontext ableitet.

Set 5:

<p>Google Translate: Spielen und halten Sie einen Akkord auf einer Tastatur, die über einen eingebauten Arpeggiator verfügt oder mit einem verbunden ist. Anstelle des Akkords hören Sie einzelne Noten nacheinander in einem von einer Systemuhr festgelegten Tempo. Abhängig von den Fähigkeiten des Arpeggiators können die Noten in aufsteigender oder absteigender Reihenfolge gespielt werden, sowohl aufsteigend als auch absteigend, nacheinander in der Reihenfolge, in der Sie die Tasten gespielt haben, zufällig eine Oktave, zwei, drei oder mehr wiederholt und durch verschiedene Klänge sequenziert auf aufeinanderfolgenden Noten. Ein Arpeggiator kann inspirierend sein und ist oft interaktiver als ein typischer Sequenzer.</p>	<p>DeepL: Wenn Sie einen Akkord auf einem Keyboard spielen und halten, das über einen eingebauten Arpeggiator verfügt oder an einen solchen angeschlossen ist, hören Sie anstelle des Akkords einzelne Noten, die nacheinander in einem von der Systemuhr bestimmten Tempo wiedergegeben werden. Je nach den Fähigkeiten des Arpeggiators können die Noten in aufsteigender oder absteigender Reihenfolge, sowohl auf- als auch absteigend, in der Reihenfolge, in der Sie die Tasten gespielt haben, zufällig, wiederholt eine Oktave, zwei, drei oder mehr höher und in Sequenz durch verschiedene Klänge auf aufeinanderfolgenden Noten abgespielt werden. Die Arbeit mit einem Arpeggiator kann sehr inspirierend sein und ist oft interaktiver als ein typischer Sequenzer.</p>
<p>H2: Wenn man einen Akkord auf einem Keyboard, das einen eingebauten Arpeggiator besitzt oder an einen angeschlossen ist, anschlägt und hält, werden die einzelnen Töne in einem festgelegten Tempo wiederholt. Je nach Arpeggiator gibt es für das Wiedergeben der Töne unterschiedliche Möglichkeiten: in auf- oder absteigender Reihenfolge, abwechselnd, der Reihenfolge folgend, die angeschlagen wurde, zufällig, eine Oktave oder zwei, drei oder mehrere Oktaven höher und durch Wiederholung aufeinanderfolgender Noten mit verschiedenen Soundeffekten. Ein Arpeggiator kann Inspirationsquelle sein und ist oftmals interaktiver aufgebaut als ein üblicher Sequenzer.</p>	<p>H1: Wenn sie auf einem Keyboard, das einen eingebauten Arpeggiator hat oder mit einem verbunden ist, einen Akkord spielen und diesen halten, hören sie anstelle des Akkordes einzelne Noten, die nacheinander in einem von der Systemuhr vorgegebenen Tempo abgespielt werden. Je nach Möglichkeiten des Arpeggiators können die Noten in aufsteigender oder absteigender Reihenfolge, in aufsteigender und absteigender Reihenfolge, in der Reihenfolge der angeschlagenen Tasten, in zufälliger Reihenfolge, wiederholt um eine, zwei oder mehrere Oktaven höher, und mit einer Sequenzierung verschiedener Töne in aufeinanderfolgenden Noten abgespielt werden. Es kann inspirierend sein, mit einem Arpeggiator zu arbeiten. Oft ist dieser auch interaktiver als ein normaler Sequenzer.</p>

Set 6:

<p>H1: André Hodeir zufolge ist das typische Rhythmusmerkmal im Jazz eine ungleichmäßige Zeitaufteilung, die sich im ungleichmäßigen Spielen von Achtelnoten manifestiert: zwei Achtelnoten werden als Viertelnote und Achtelnote in einer Triole (lang-kurzes Triolenpaar) gespielt. Die Zeit ist also ternär aufgeteilt (triolische Unterteilung).</p>	<p>DeepL: Nach André Hodeir liegt das charakteristische Merkmal des Rhythmus im Jazz in einer ungleichen Zeiteinteilung, die sich im ungleichen Spielen von Achtelnoten materialisiert: Zwei Achtelnoten werden eigentlich als Viertelnote und Achtelnote in einer Triole (lang-kurzes Triolenpaar) gespielt. Die Zeit wird also ternär geteilt (triolische Subdivision).</p>
<p>H2: Das charakteristische Merkmal des Jazzrhythmus liegt laut André Hodeir in einer ungleichen Zeitaufteilung, die sich darin zeigt, dass Achtelnoten ungleich gespielt werden: Zwei Achtelnoten werden in einer Triole als eine Viertel- und eine Achtelnote gespielt (lang-kurzes Triolenpaar). Der Takt wird dementsprechend ternär (in Triolen) aufgeteilt.</p>	<p>Google Translate: Laut André Hodeir liegt das charakteristische Merkmal des Rhythmus im Jazz in einer ungleichen Zeiteinteilung, die sich beim ungleichen Spielen von Achtelnoten ergibt: Zwei Achtelnoten werden tatsächlich als Viertelnote und eine Achtelnote in einem Triplet (langes kurzes Triplettpaar) gespielt. Die Zeit wird also ternär aufgeteilt (Triplet-Unterteilung).</p>

11. Anhang 3

Im folgenden Anhang steht die für diese Masterarbeit verwendete Umfrage.

Umfrage zur Qualität von maschinellen und menschlichen Übersetzungen im Fachgebiet Musiktheorie

Daten zur Person (bitte leserlich und in Blockschrift ausfüllen, danke! ☺)

Vor- und Nachname, akademischer Grad:				
Altersgruppe (bitte einkreisen):	18-30	30-45	45-60	60+
Geschlecht:				
Deutsch als Erstsprache?				
Weitere Erstsprache(n)?				
Weitere Sprachkenntnisse; Sprachniveaustufe nach dem Gemeinsamen europäischen Referenzrahmen (A1-C2):				
Angabe zur Fachkompetenz im Bereich Musik (Musikschule, Musikstudium, professionelle(r) Musiker:in ohne formale Ausbildung, etc.):				

Beschreibung der Umfrage:

liebe Teilnehmer:innen,

ich danke Ihnen, dass Sie sich die Zeit nehmen, bei dieser Umfrage teilzunehmen!

Das Ziel dieser Umfrage ist es herauszufinden, ob Musiker:innen einen Unterschied zwischen Humanübersetzungen (erstellt von Menschen) und maschinellen Übersetzungen (erstellt von maschinellen Übersetzungssystemen) musiktheoretischer Texte bemerken. Darüber hinaus soll festgestellt werden, welche Übersetzungen durchschnittlich als die besten bzw. die schlechtesten empfunden werden.

Sie werden Übersetzungen von insgesamt 6 Textproben bewerten. Für jede der Textproben gibt es insgesamt 4 Übersetzungen. Diese Übersetzungen können entweder von Menschen erstellt worden sein, oder auch von Maschinen.

Bei den Fragen handelt es sich um eine Gegenüberstellung der Qualität der vier Übersetzungen, daher gibt es keine richtige oder falsche Antwort. Äußern Sie bitte einfach Ihre Meinung.

Nach dem Ausfüllen des schriftlichen Teils folgt ein kurzes Interview mit Fragen in Bezug auf Sprache, Übersetzen, Musik und diese Umfrage. Für die Umfrage inkl. des Interviews sind insgesamt 75 Minuten vorgesehen.

Ihre Daten und Angaben werden vertraulich behandelt, werden allgemein nicht an Dritte weitergegeben und werden ausschließlich für Zwecke dieser Masterarbeit zur Untersuchung von maschinellen Übersetzungssystemen verwendet.

Set 0 (BEISPIEL):

<p>Übersetzung 1: Musiktheorie ist eines der wichtigsten Themen, die jeder Musiker studieren kann, um sein höchstes Potenzial zu erreichen. Theoretische Kenntnisse helfen Ihnen, Musik auf einer viel tieferen Ebene zu komponieren, aufzuführen, zu verstehen und zu schätzen.</p>	<p>Übersetzung 2: Musiktheorie ist eines von den wichtigsten Themen, die Musiker lernen können, um ihr volles Potenzial auszuschöpfen. Theoretische Kenntnisse ermöglichen es, Musik auf einer viel tieferen Ebene zu schreiben, aufzuführen, zu verstehen und zu genießen.</p>
<p>Übersetzung 3: Musiktheorie ist eines der wichtigsten Themen, die jeder Musiker studieren kann, um sein höchstes Potenzial zu erreichen. Theoriekenntnisse helfen Ihnen, Musik auf einer viel tieferen Ebene zu komponieren, aufzuführen, zu verstehen und zu schätzen.</p>	<p>Übersetzung 4: Musiktheorie gehört zu den wichtigsten Themen, die Musikerinnen und Musiker erforschen können, um ihr volles Potenzial zu erreichen. Theoriekenntnisse ermöglichen es, Musik auf einer viel tieferen Ebene zu komponieren, aufzuführen, zu verstehen und zu würdigen.</p>

1) Bewerten Sie bitte die **beste** und die **schlechteste** dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									X
Übersetzung 2									
Übersetzung 3		X							
Übersetzung 4									

^Bewerten Sie bitte immer nur eine beste Übersetzung (positiv) und nur eine schlechteste Übersetzung (negativ) . Somit werden insgesamt immer nur zwei Übersetzungen bewertet. Hierbei handelt es sich um eine subjektive Einschätzung; es gibt keine richtigen oder falschen Antworten.

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach **von Menschen** erstellt wurden.

Übersetzung 1	X
Übersetzung 2	X
Übersetzung 3	X
Übersetzung 4	

^Wählen Sie bitte immer alle Übersetzungen aus, für die Sie denken, dass sie von Menschen erstellt worden sind. Es ist möglich, dass keine der Übersetzungen von Menschen erstellt wurde, oder auch, dass alle 4 von Menschen erstellt wurden.

Set 1:

<p>Übersetzung 1: Übungen in strenger Stimmführung oder Artkontrapunkt beginnen mit einer einzelnen, gut geformten musikalischen Linie, die Cantus firmus (feste Stimme oder feste Melodie; pl. cantus firmi) genannt wird. Die Komposition eines Cantus firmus gibt uns die Möglichkeit, die folgenden grundlegenden musikalischen Eigenschaften zu trainieren: Geschmeidigkeit, Unabhängigkeit und Integrität der melodischen Linien, Vielfalt und Bewegung (auf ein Ziel hin).</p>	<p>Übersetzung 2: Etüden in strenger Stimmführung oder strengem Kontrapunkt beginnen mit einer einzelnen, wohlgeformten Melodielinie namens Cantus firmus (feststehender Gesang oder feststehende Melodie; Pl. Cantus firmi). Die Cantus firmus-Komposition ermöglicht es uns, folgende grundlegende musikalische Merkmale zu aktivieren: Gleichmäßigkeit, Unabhängigkeit und Integrität oder melodische Linien, Varietät und Bewegung (zu einem Ziel hin).</p>
<p>Übersetzung 3: Übungen in strikter Stimmführung oder Artenkontrapunkt beginnen mit einer einzigen, wohlgeformten Musiklinie, die als Cantus Firmus (feste Stimme oder feste Melodie; Pl. Cantus Firmi) bezeichnet wird. Die Komposition von Cantus firmus gibt uns die Möglichkeit, die folgenden grundlegenden musikalischen Merkmale zu berücksichtigen: Geschmeidigkeit, Unabhängigkeit und Integrität oder melodische Linien, Vielfalt und Bewegung (in Richtung eines Ziels).</p>	<p>Übersetzung 4: Übungen zur Stimmführung oder Kontrapunkte verschiedener Gattungen beginnen mit einer einzelnen, ausgestalteten Melodie die „Cantus firmus“ („fester Gesang“ oder „feste Melodie“; pl. Cantus firmi) genannt wird. Die Komposition eines Cantus firmus gibt uns die Möglichkeit, uns mit wesentlichen musikalischen Eigenschaften, wie Wohlklang, Eigenständigkeit und Integrität oder melodischen Linien, Varietät und Bewegung (zu einem Zweck), zu befassen.</p>

1) Bewerten Sie bitte die beste und die schlechteste dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									
Übersetzung 2									
Übersetzung 3									
Übersetzung 4									

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach von Menschen erstellt wurden.

Übersetzung 1	
Übersetzung 2	
Übersetzung 3	
Übersetzung 4	

Set 2:

<p>Übersetzung 1: Sequenzen sind oft durch das Muster ihrer Wurzeln gekennzeichnet. Von den zahlreichen möglichen harmonischen Mustern, die eine Sequenz steuern können, sind einige besonders häufig. Die am häufigsten verwendete ist die absteigende fünfte Sequenz - auch als Quintenkreis bekannt - mit Akkorden, deren Wurzeln sich abwechselnd um eine vierte und eine fünfte nach unten bewegen.</p>	<p>Übersetzung 2: Sequenzen werden oft durch das Muster ihrer Wurzeln gekennzeichnet. Von den zahlreichen möglichen harmonischen Mustern, die eine Sequenz bestimmen können, sind einige besonders häufig. Das am häufigsten verwendete ist die absteigende Quintenfolge - auch als Quintenzirkel bekannt -, bei der Akkorde verwendet werden, deren Grundton sich abwechselnd um eine Quarte nach oben und um eine Quinte nach unten bewegt.</p>
<p>Übersetzung 3: Sequenzen werden nach dem Muster ihrer Grundtöne unterschieden. Von den verschiedenen harmonischen Mustern, die eine Sequenz dominieren können, sind einige besonders gängig. Die meistverwendete Sequenz ist die Quintfallsequenz, auch Quintenzirkel genannt, bei deren Akkordfolge sich die Grundtöne in Quintschritten abwärts- und Quartschritten aufwärtsbewegen.</p>	<p>Übersetzung 4: Sequenzen werden oft nach dem Muster ihrer Grundtöne benannt. Es gibt zahlreiche harmonische Motive, die eine Sequenz lenken können. Einige davon werden besonders oft verwendet. Am häufigsten wird die Quintfallsequenz – oder Quintschrittsequenz – verwendet, bei welcher Akkorde zum Einsatz kommen, deren Grundtöne sich in einem stetigen Wechsel aus aufsteigenden Quartan und fallenden Quinten befinden.</p>

1) Bewerten Sie bitte die beste und die schlechteste dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									
Übersetzung 2									
Übersetzung 3									
Übersetzung 4									

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach von Menschen erstellt wurden.

Übersetzung 1	
Übersetzung 2	
Übersetzung 3	
Übersetzung 4	

Set 3:

<p>Übersetzung 1: Wir beschreiben Septakkorde durch ihre beiden am deutlichsten heraushörbaren Eigenschaften: die Art des Dreiklangs (Dur, Moll, vermindert) und der Septime (groß, klein, vermindert). Komponistinnen und Komponisten verwenden fünf wichtige Arten von Septakkorden, obwohl, ähnlich wie bei den Dreiklangarten, nicht alle gleich häufig verwendet werden. Septakkorde werden nach der Dreiklangqualität und der Art der Septime benannt.</p>	<p>Übersetzung 2: Wir bezeichnen Septakkorde nach ihren beiden am besten hörbaren Merkmalen: dem Typ des Dreiklangs (Dur, Moll, vermindert) und dem Typ der Septime über dem Grundton (Dur, Moll, vermindert). Komponisten verwenden fünf wichtige Typen von Septakkorden, obwohl sie, wie die Dreiklangstypen, nicht mit gleicher Häufigkeit verwendet werden. Septakkorde werden nach der Qualität des Dreiklangs und der Größe der Septime benannt.</p>
<p>Übersetzung 3: Wir benennen Septakkorde nach den beiden Merkmalen, die am besten hörbar sind: die Art des Dreiklangs (Dur, Moll, vermindert) und die Art der Septime über dem Grundton (Dur, Moll, vermindert). Komponisten verwenden fünf wesentliche Arten von Septakkorden, welche allerdings, wie auch die Arten des Dreiklangs, nicht gleich häufig zum Einsatz kommen. Septakkorde werden nach der Qualität des Dreiklangs und der Größe der Septime benannt.</p>	<p>Übersetzung 4: Wir kennzeichnen siebte Akkorde anhand ihrer zwei am besten hörbaren Merkmale: der Art der Triade (Dur, Moll, vermindert) und der Art der siebten über dem Grundton (Dur, Moll, vermindert). Komponisten verwenden fünf wichtige Arten von Septakkorden, die jedoch wie die Triadenarten nicht mit gleicher Häufigkeit verwendet werden. Siebte Akkorde werden nach der Qualität der Triade und der Größe der siebten benannt.</p>

1) Bewerten Sie bitte die beste und die schlechteste dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									
Übersetzung 2									
Übersetzung 3									
Übersetzung 4									

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach von Menschen erstellt wurden.

Übersetzung 1	
Übersetzung 2	
Übersetzung 3	
Übersetzung 4	

Set 4:

<p>Übersetzung 1: Bei dem Versuch, die italienischen Tempobezeichnungen des 17. und 18. Jahrhunderts in den Griff zu bekommen, muss man sich vor Augen halten, dass die meisten dieser Wörter (z. B. allegro, largo, presto) Teil der italienischen Umgangssprache waren und sind und von den italienischen Komponisten wörtlich und nicht als musikalische Begriffe verwendet wurden. So bedeutet allegro eher fröhlich, heiter als schnell; nur wenn die besondere Art der Fröhlichkeit ein bestimmtes Tempo erfordert, wird allegro indirekt zu einer Tempobezeichnung. Im Allgemeinen sollten diese Zeichen als Hinweise auf emotionale Zustände verstanden werden, wobei sich das absolute Tempo aus dem Kontext ergibt.</p>	<p>Übersetzung 2: Beim Versuch, sich mit den italienischen Tempobezeichnungen des 17. und 18. Jahrhunderts auseinanderzusetzen, muss berücksichtigt werden, dass die meisten der dafür verwendeten Wörter (beispielsweise allegro, largo, presto) Teil des italienischen Sprachgebrauchs waren und sind und daher von den italienischen Komponistinnen und Komponisten eher wörtlich anstatt als musikalischer Begriff gebraucht wurden. Dementsprechend meint "Allegro" eher munter und fröhlich als schnell. Nur wenn eine gewisse Art von Heiterkeit ein spezifisches Tempo benötigt, wird "Allegro" indirekt zur Tempobezeichnung. Generell sollten diese Bezeichnungen als Ausdrucksbezeichnungen verstanden werden, bei denen sich das tatsächliche Tempo aus dem Kontext ergibt.</p>
<p>Übersetzung 3: Bei dem Versuch, die italienischen Tempomarkierungen des 17. und 18. Jahrhunderts in den Griff zu bekommen, muss berücksichtigt werden, dass die meisten dieser Wörter (z. B. Allegro, Largo, Presto) Teil der italienischen Umgangssprache waren und sind und von Italienern verwendet wurden. Komponisten buchstäblich und nicht als musikalische Begriffe. Allegro bedeutet also eher fröhlich, fröhlich als schnell; Nur wenn die besondere Natur der Fröhlichkeit ein bestimmtes Tempo erfordert, wird Allegro indirekt zu einer Tempoanzeige. Im Allgemeinen sollten diese Zeichen als Hinweise auf emotionale Zustände verstanden werden, wobei sich das absolute Tempo aus dem Kontext ergibt.</p>	<p>Übersetzung 4: Bei dem Versuch, die italienischen Tempobezeichnungen des 17. und 18. Jahrhunderts zu verstehen, muss man sich vor Augen halten, dass die meisten dieser Bezeichnungen (z.B. „allegro“, „largo“, „presto“) Teil der italienischen Umgangssprache waren und noch immer sind und von italienischen Komponisten eher wörtlich und nicht als musikalische Begriffe verwendet wurden. So bedeutet „allegro“ eher „fröhlich“ und „munter“ als „schnell“; nur wenn die besondere Art der Fröhlichkeit ein bestimmtes Tempo erfordert, wird „allegro“ indirekt zu einer Tempoangabe. Im Allgemeinen sollten diese Bezeichnungen als Zeichen emotionaler Zustände verstanden werden, wobei sich das absolute Tempo aus dem Kontext ableitet.</p>

1) Bewerten Sie bitte die beste und die schlechteste dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									
Übersetzung 2									
Übersetzung 3									
Übersetzung 4									

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach von Menschen erstellt wurden.

Übersetzung 1	
Übersetzung 2	
Übersetzung 3	
Übersetzung 4	

Set 5:

<p>Übersetzung 1: Spielen und halten Sie einen Akkord auf einer Tastatur, die über einen eingebauten Arpeggiator verfügt oder mit einem verbunden ist. Anstelle des Akkords hören Sie einzelne Noten nacheinander in einem von einer Systemuhr festgelegten Tempo. Abhängig von den Fähigkeiten des Arpeggiators können die Noten in aufsteigender oder absteigender Reihenfolge gespielt werden, sowohl aufsteigend als auch absteigend, nacheinander in der Reihenfolge, in der Sie die Tasten gespielt haben, zufällig eine Oktave, zwei, drei oder mehr wiederholt und durch verschiedene Klänge sequenziert auf aufeinanderfolgenden Noten. Ein Arpeggiator kann inspirierend sein und ist oft interaktiver als ein typischer Sequenzer.</p>	<p>Übersetzung 2: Wenn Sie einen Akkord auf einem Keyboard spielen und halten, das über einen eingebauten Arpeggiator verfügt oder an einen solchen angeschlossen ist, hören Sie anstelle des Akkords einzelne Noten, die nacheinander in einem von der Systemuhr bestimmten Tempo wiedergegeben werden. Je nach den Fähigkeiten des Arpeggiators können die Noten in aufsteigender oder absteigender Reihenfolge, sowohl auf als auch absteigend, in der Reihenfolge, in der Sie die Tasten gespielt haben, zufällig, wiederholt eine Oktave, zwei, drei oder mehr höher und in Sequenz durch verschiedene Klänge auf aufeinanderfolgenden Noten abgespielt werden. Die Arbeit mit einem Arpeggiator kann sehr inspirierend sein und ist oft interaktiver als ein typischer Sequenzer.</p>
<p>Übersetzung 3: Wenn man einen Akkord auf einem Keyboard, das einen eingebauten Arpeggiator besitzt oder an einen angeschlossen ist, anschlägt und hält, werden die einzelnen Töne in einem festgelegten Tempo wiederholt. Je nach Arpeggiator gibt es für das Wiedergeben der Töne unterschiedliche Möglichkeiten: in auf- oder absteigender Reihenfolge, abwechselnd, der Reihenfolge folgend, die angeschlagen wurde, zufällig, eine Oktave oder zwei, drei oder mehrere Oktaven höher und durch Wiederholung aufeinanderfolgender Noten mit verschiedenen Soundeffekten. Ein Arpeggiator kann Inspirationsquelle sein und ist oftmals interaktiver aufgebaut als ein üblicher Sequenzer.</p>	<p>Übersetzung 4: Wenn sie auf einem Keyboard, das einen eingebauten Arpeggiator hat oder mit einem verbunden ist, einen Akkord spielen und diesen halten, hören sie anstelle des Akkordes einzelne Noten, die nacheinander in einem von der Systemuhr vorgegebenen Tempo abgespielt werden. Je nach Möglichkeiten des Arpeggiators können die Noten in aufsteigender oder absteigender Reihenfolge, in aufsteigender und absteigender Reihenfolge, in der Reihenfolge der angeschlagenen Tasten, in zufälliger Reihenfolge, wiederholt um eine, zwei oder mehrere Oktaven höher, und mit einer Sequenzierung verschiedener Töne in aufeinanderfolgenden Noten abgespielt werden. Es kann inspirierend sein, mit einem Arpeggiator zu arbeiten. Oft ist dieser auch interaktiver als ein normaler Sequenzer.</p>

1) Bewerten Sie bitte die beste und die schlechteste dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									
Übersetzung 2									
Übersetzung 3									
Übersetzung 4									

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach von Menschen erstellt wurden.

Übersetzung 1	
Übersetzung 2	
Übersetzung 3	
Übersetzung 4	

Set 6:

<p>Übersetzung 1: André Hodeir zufolge ist das typische Rhythmusmerkmal im Jazz eine ungleichmäßige Zeitaufteilung, die sich im ungleichmäßigen Spielen von Achtelnoten manifestiert: zwei Achtelnoten werden als Viertelnote und Achtelnote in einer Triole (lang-kurzes Triolenpaar) gespielt. Die Zeit ist also ternär aufgeteilt (triolische Unterteilung).</p>	<p>Übersetzung 2: Nach André Hodeir liegt das charakteristische Merkmal des Rhythmus im Jazz in einer ungleichen Zeiteinteilung, die sich im ungleichen Spielen von Achtelnoten materialisiert: Zwei Achtelnoten werden eigentlich als Viertelnote und Achtelnote in einer Triole (lang-kurzes Triolenpaar) gespielt. Die Zeit wird also ternär geteilt (triolische Subdivision).</p>
<p>Übersetzung 3: Das charakteristische Merkmal des Jazzrhythmus liegt laut André Hodeir in einer ungleichen Zeitaufteilung, die sich darin zeigt, dass Achtelnoten ungleich gespielt werden: Zwei Achtelnoten werden in einer Triole als eine Viertel- und eine Achtelnote gespielt (lang-kurzes Triolenpaar). Der Takt wird dementsprechend ternär (in Triolen) aufgeteilt.</p>	<p>Übersetzung 4: Laut André Hodeir liegt das charakteristische Merkmal des Rhythmus im Jazz in einer ungleichen Zeiteinteilung, die sich beim ungleichen Spielen von Achtelnoten ergibt: Zwei Achtelnoten werden tatsächlich als Viertelnote und eine Achtelnote in einem Triplet (langes kurzes Triplettpaar) gespielt. Die Zeit wird also ternär aufgeteilt (Triplet-Unterteilung).</p>

1) Bewerten Sie bitte die beste und die schlechteste dieser 4 Übersetzungen. Bitte bewerten Sie dabei die beste Übersetzung von 0 bis +4 und die schlechteste von -4 bis 0.

	gut				neutral	schlecht			
	+4	+3	+2	+1	0	-1	-2	-3	-4
Übersetzung 1									
Übersetzung 2									
Übersetzung 3									
Übersetzung 4									

2) Geben Sie bitte an, welche der 4 Übersetzungen Ihrer Meinung nach von Menschen erstellt wurden.

Übersetzung 1	
Übersetzung 2	
Übersetzung 3	
Übersetzung 4	

12. Anhang 4

In diesem Anhang steht der Leitfaden für das Interview und die transkribierten Interviews mit den Evaluator*innen.

Der Leitfaden für das Interview:

1) Haben Sie jemals MÜs verwendet?

JA

- a. Für welche Zwecke verwenden Sie MÜ?
- b. Privat oder beruflich?
- c. Verwenden Sie maschinelle Übersetzung im Alltag?
- d. Welches MÜ-System haben Sie verwendet?
- e. Haben Sie jemals MÜs für Themen im Fachgebiet Musik verwendet (Musiktheorie, Songtexte, etc.)?
- f. Wie zufrieden sind Sie insgesamt mit der Qualität von MÜs?
- g. Hat Ihnen MÜ jemals ermöglicht, einen Text zu verstehen, den Sie sonst nicht verstehen hätten können?
- h. Haben Sie jemals von Google Translate gehört? Haben Sie es jemals verwendet? Für welchen Zweck? Wie zufrieden waren Sie? Denken Sie, dass Sie es in Zukunft verwenden werden?
- i. Haben Sie jemals von DeepL gehört? Haben Sie es jemals verwendet? Für welchen Zweck? Wie zufrieden waren Sie? Denken Sie, dass Sie es in Zukunft verwenden werden?

JA und NEIN

2) Sind eher Befürworter:in oder Gegner:in von MÜs?

3) Wo sehen Sie Vor- bzw. Nachteile von MÜs?

4) Hatten Sie bisher Erfahrung mit Übersetzer:innen und/oder Dolmetscher:innen?

5)

- a. Haben Sie jemals eine Übersetzung beauftragt?
- b. Falls ja, waren Sie zufrieden?
- c. Falls ja, woher wussten Sie, ob die Übersetzung qualitativ angemessen war?

6) Haben Sie selbst Texte übersetzt bzw. gedolmetscht? Beruflich oder privat? Mit oder ohne Hilfe von MÜs? Wie ist es Ihnen dabei ergangen?

7) Haben Sie jemals an einer Bewertung von Übersetzungsqualität teilgenommen?

- 8) Kommen Sie als Musiker:in regelmäßig mit musiktheoretischen oder ähnlichen Texten in Kontakt, welche nicht in einer Ihrer Erstsprachen verfasst sind?
- 9) Haben Sie einen Bedarf an Übersetzungen von Texten zum Thema Musiktheorie bzw. Musik im Allgemeinen, oder genügen Ihnen Texte, welche in einer Ihrer Erstsprachen verfasst wurden?
- 10) Empfinden Sie, dass Ihr Bedarf an Übersetzungen von Texten zum Thema Musiktheorie vollkommen mittels MÜ gedeckt werden kann, oder benötigen Sie von Menschen erstellte Übersetzungen?
- 11) Wie ist es Ihnen bei dieser Umfrage ergangen?
 - a. War die Fragestellung klar und deutlich?
 - b. War das Format angemessen?
 - c. Fanden Sie es leicht, sich zu entscheiden, welche der Übersetzungen von Menschen erstellt worden sind?
 - d. Woran erkannten Sie, welche Übersetzungen von Menschen erstellt wurden?
 - e. Wie sicher sind Sie sich, dass Sie alle oder die meisten menschlichen Übersetzungen erkannt haben?
- 12) Sonstiges Feedback zum Thema?
- 13) Sonstiges Feedback zur Umfrage?

12.1. Interview Teilnehmer*in 1:

M: So, das sind die Aufnahmen zur Umfrage Nummer 1. Wie gesagt, deine Daten, incl. dieser Audioaufnahmen, werden streng vertraulich behandelt. Kannst du mir nur noch einmal den Vor- und Nachnamen bestätigen?

T1: (Anonymisiert)

M: Ja, danke dir! Dann gehe ich mal zu den Fragen. Also, meine erste Frage – hast du jemals maschinelle Übersetzungssysteme verwendet, und falls ja, für welche Zwecke; war das eher privat oder beruflich, oder so?

T1: Ja, das habe ich, und eher privat.

M: Eher privat, ja, kannst du dich erinnern wofür?

T1: Ah, ich glaube das war oft für die Schule, genau, dass man da was übersetzen musste, genau, oder dass ich jetzt auch für Tiermedizin da irgendwelche Texte, die sind ja oft diese ganzen Papers auf Englisch, dass man da was eingibt wenn man es nicht versteht.

M: Kannst du dich erinnern, was du verwendet hast, also welches System?

T1: Dieses... Google Translator...

M: Ok, und so insgesamt wie zufrieden bist du mit der Übersetzungsqualität von solchen Systemen.

T1: Also jetzt super, also früher weiß ich noch, da war das nicht so gut, da kam schon manchmal komisches Zeug raus. Aber jetzt ist es super. Auch wenn man manchmal geschachtelte Sätze eingibt ergeben die schon Sinn.

M: Und würdest du sagen, dass dir MÜ hilft, Texte zu verstehen, die du sonst nicht verstehen könntest?

T1: Ja, doch, manchmal schon.

M: Ja, Google Translate hast du schon erwähnt, und damit bist du zufrieden insgesamt und man kann auch annehmen, dass du es in Zukunft verwenden wirst?

T1: Ja.

M: OK, und vom System DeepL hast du vielleicht auch gehört?

T1: Nein.

M: Würdest du sagen du bist eher Befürworterin oder Gegnerin maschineller Übersetzung?

T1: Also, ich finde das gut, ja, Befürworterin, ja, ist eine gute Sache.

M: Wo siehst du Vorteile und Nachteile von MÜ?

T1: Na die Vorteile, dass man das besser verstehen kann usw. Nachteile, dass man dann halt selber faul wird, nö?

M: Kannst du mir sagen, hast du mit menschlichen Übersetzer*innen oder Dolmetscher*innen Erfahrung gehabt, oder dass du eine Übersetzung beauftragt hast oder sowas?

T1: Nein, noch nicht. Eher privat so, aber ja...

M: Hast du Texte selbst übersetzt oder gedolmetscht?

T1: So richtig offiziell oder nur für mich?

M: Halbwegs offiziell oder offiziell, oder auch nur für dich.

T1: Also ja, nur für mich. Offiziell noch nicht.

M: Aber wirklich dann einen ganzen Text erstellt, oder eher nur Textpassagen oder wie?

T1: Ja, eher nur Textpassagen, einen ganzen Text nicht.

M: An einer Bewertung von Übersetzungsqualität, außer dieser hier, hast du nicht teilgenommen, oder?

T1: Nein.

M: Du hast ja einen Musikschulabschluss; kommst du als Musikerin regelmäßig in Kontakt mit musiktheoretischen Texten, die nicht in einer deiner Erstsprachen sind?

T1: Nein, da ich das eher nur privat mache, nicht aktiv beruflich oder so, eher weniger.

M: Dann noch ein paar Fragen zu dieser Umfrage: war die Fragestellung klar und deutlich?

T1: Ja.

M: Und war das Format angemessen?

T1: Ja.

M: Fandest du es leicht zu entscheiden, welche Übersetzungen von Menschen erstellt wurden?

T1: Nein.

M: Wenn du dich entschieden hast, dass eine der Übersetzungen von Menschen ist, woran hast du das erkannt und wie hast du deine Entscheidung getroffen?

T1: Manchmal dachte ich, das vom Menschen wäre doch nicht so kompliziert ausgedrückt oder einfacher, genau, dass man da Wörter benutzt, die bisschen umgangssprachlicher sind, genau, ja, da hatte ich manchmal das Gefühl „und das passt jetzt mehr, das ist vom Menschen, nicht von einer Maschine“.

M: Und wie sicher bist du dir, dass du alle menschlichen Üs richtig erkannt hast?

T1: So 50-50.

M: Sonstiges Feedback zum Thema oder zur Umfrage?

T1: Nein.

M: Ich bedanke mich für das Interview!

12.2. Interview Teilnehmer*in 2:

M: So, das ist das Interview Nummer 2. Vielen Dank noch einmal, dass Sie sich die Zeit für die Umfrage genommen haben, und für das Interview. Wie schon gesagt, Ihre Daten werden streng

vertraulich behandelt, incl. dieser Audioaufnahmen, und werden nicht an Dritte weitergeleitet. Könnten Sie mir bitte nur noch einmal Ihren Vor- und Nachnamen bestätigen.

T2: (Anonymisiert)

M: Vielen Dank! Meine erste Frage wäre: maschinelle Übersetzung insgesamt – haben Sie sie jemals verwendet und für welche Zwecke war das, eher privat oder eher beruflich oder beides?

T2: Eigentlich beides, aber ich tue das dann nur um die Arbeit zu verschnellern, manchmal benützen, aber dann tue ich das dann, jemand anderem, das Resultat, also das Translat, schon manuell verändern...

M: Würden Sie sagen, dass Sie es täglich verwenden, also im Alltag, oder eher selten?

T2: Weder noch. Ich verwende es schon oft, aber nicht jeden Tag, aber oft.

M: Und welches System verwenden Sie?

T2: Ja, ich verwende TransDict.

M: Wie zufrieden sind Sie insgesamt mit der Qualität von MÜs?

T2: Na ja, ich wundere mich, wie das immer besser wird, eigentlich schon zufrieden. Vor allem hilft es mir bei Sprachen, die ich nicht kann, um mehr oder weniger zu verstehen, worum es geht.

M: Also würden Sie auch sagen, dass Ihnen MÜs geholfen haben, Texte zu verstehen, die Sie sonst nicht verstehen hätten können?

T2: Genau, ja, ich kriege z. B. manchmal Texte auf Ungarisch, das könnte ich überhaupt nicht verstehen; so weiß ich circa worum es geht.

M: Haben Sie schon vom System Google Translate gehört und haben Sie es verwendet?

T2: Gehört ja, aber selten verwendet. Eher früher, wo es TransDict nicht gegeben hat.

M: Und was würden Sie sagen – wie zufrieden waren Sie mit Google Translate?

T2: Das geht... nicht so zufrieden.

M: Haben Sie jemals vom System DeepL gehört?

T2: Ja, DeepL, ich benutze DeepL.

M: Und würden Sie sagen, dass Sie da zufrieden sind?

T2: Eigentlich schon, ja.

M: Haben Sie Google Translate und DeepL früher zu selben Zwecken verwendet, wie andere Übersetzungssysteme auch, privat oder beruflich, hauptsächlich um etwas Musik-bezogenes vielleicht?

T2: Nein, Musik-bezogene Texte kann ich eigentlich selber in allen Sprachen lesen, und ich tue das eher, ja, wenn ich manchmal, wie gesagt, so Sachen in einer Sprache, die ich nicht kenne, oder wenn ich schnell was formulieren will, dann schmeiße ich es rein, ich schreibe es auf Deutsch auf DeepL und korrigiere dann die Englische Version.

M: Denken Sie, dass Sie auch in Zukunft Google Translate und DeepL verwenden werden?

T2: Ja, DeepL schon, Google Translate... das benütze ich sehr selten... kaum.

M: Würden Sie sagen, Sie sind eher Befürworter oder Gegner von MÜ, und wo sehen Sie Vor- und Nachteile?

T2: Naja, das kommt darauf an, was Sie damit machen möchten. Also, jetzt für eine Veröffentlichung würde mir nie in den Sinn kommen, dass mit einem Übersetzungsprogramm zu machen, aber irgendwie eine schnelle Kommunikation oder irgendwo etwas, wo es nur darum geht, eine kurze

Information zu wiedergeben, oder was auch immer, warum nicht, ja? Bzw. um was zu lesen, was ich sonst nicht lesen könnte.

M: Also eher Befürworter?

T2: Ja, bin kein Gegner.

M: Haben Sie mit menschlichen Übersetzer*innen und Dolmetscher*innen Erfahrung gemacht, z. B. eine Übersetzung beauftragt oder sowas?

T2: Also, Übersetzungen beauftragt nicht, aber ich habe oft bei Kongressen gesprochen, da wurde in Kabinen gedolmetscht. Ich habe schon oft auch Übersetzungen korrigiert für Publikationen, jetzt wenn es wirklich so Musik-bezogene Texte sind, und ich beschäftige mich auch wissenschaftlich mit dem Thema, also ich würde sagen ja.

M: Und waren Sie mit diesen Übersetzer*innen und Dolmetscher*innen eher zufrieden oder nicht?

T2: Ja, schon sehr zufrieden.

M: Und wie wussten Sie, ob diese Übersetzungen und Dolmetschungen qualitativ hochwertig waren? Haben Sie das irgendwie vielleicht überprüft oder haben Sie einfach vertraut, dass das gut war?

T2: Naja, bei den Kabinen-Dolmetschen habe ich vertraut, also, die waren auch nicht von mir beauftragt worden. Ich rede ja auch, ich kann nicht kontrollieren... Aber ich merke schon oft, dass z. B. wenn jetzt mein Sprechen konsekutiv übersetzt wird, kann es schon sein, dass ich Begriffe benütze, die dann schlecht oder falsch übersetzt werden, und dann sage ich „Nein, Vorsicht, im Musik-Bereich ist das das“, also da merke ich schon und würde dann auch schon einschreiten. Bei schriftlichen Texten merke ich das.

M: Haben Sie auch persönlich übersetzt oder gedolmetscht?

T2: Gedolmetscht habe ich nicht, aber für Publikationen habe ich viel übersetzt.

M: Haben Sie zufällig an einer Bewertung von Übersetzungsqualität teilgenommen, wie bei der hier?

T2: Nein.

M: Sie sind ja auch ein Fachexperte im Fachgebiet Musik; kommen Sie da oft in Kontakt mit Texten, die nicht in einer Ihrer Erstsprachen sind, also weder Deutsch noch Italienisch?

T2: Ja, sehr viel.

M: Und denken Sie, dass Ihr Bedarf an Musik-bezogenen Texten gedeckt ist von den Texten in Ihren Erstsprachen, oder benötigen Sie auch Übersetzungen?

T2: Also, wenn es keine Übersetzungen gab, wäre das schon schlecht. Das ist jetzt auch einer der Hauptprobleme in der Musikwissenschaft, dass wir hauptsächlich nur Texte auf Deutsch und Englisch haben, und natürlich gibt es Texte in anderen Sprachen, aber ziemlich sparsam verbreitet, dadurch dass nicht viele Menschen das lesen können. Es gibt z. B. Italienisch-sprachige oder Französisch-sprachige Literatur, Spanisch-sprachige Literatur, Russisch-sprachig – das wird kaum wahrgenommen von der großen Community. Natürlich ist es auch wichtig, diese Texte zu übersetzen.

M: Und Ihrer Meinung nach, kann dieser Bedarf durch MÜs gedeckt werden, oder benötigen wir immer noch menschliche Übersetzer*innen?

T2: Nein, wir brauchen Menschen.

M: Noch ein paar kurze Fragen zu dieser Umfrage. Wie ist es Ihnen ergangen? Waren die Fragen klar und deutlich? Fanden Sie das Format angemessen?

T2: Ja, und ja.

M: Fanden Sie es leicht zu entscheiden, welche Übersetzungen von Menschen erstellt worden sind?

T2: Also, eher leicht, aber manchmal mit Zweifel.

M: Und woran erkannten Sie, dass eine Übersetzung von einem Menschen ist und nicht von einer Maschine?

T2: Ja, z B. bei den maschinellen Übersetzungen gibt es manchmal syntaktische Fehler, oder es gab bei einer Übersetzung den Terminus „Triplet“ – das gibt es nicht im Deutschen, wahrscheinlich wurde das aus dem Englischen übersetzt. Dann auch die Satzstellung bzw. der Umgang mit Fremdwörtern. „Die Komposition von Cantus Firmus“ ist meiner Meinung nach ziemlich eindeutig ein Hinweis auf maschinelle Übersetzung. Und sagen wir, die menschlichen Übersetzungen gehen auch einen Schritt weiter, indem eigentlich Erklärungen zum Text hinzukommen, Versuche das zu verdeutlichen...

M: Und denken Sie, Sie haben die meisten menschlichen Übersetzungen erkannt, oder eher wenige?

T2: Weiß ich nicht, ich denke schon aber... ich kann mich irren.

M: Sonstiges Feedback?

T2: Nein.

M: Vielen Dank für das Interview!

12.3. Interview Teilnehmer*in 3:

M: Das sind die Aufnahmen zur Umfrage Nummer 3. Kannst du mir bitte noch einmal deinen Namen sagen?

T3: (anonymisiert)

M: Vielen Dank! Ich habe da ein paar Fragen. Als aller erstes zum maschinellen Übersetzen: hast du jemals ein MÜ-System verwendet, entweder privat oder beruflich oder im Alltag?

T3: Ja.

M: Weißt du vielleicht noch welches?

T3: Google Translate.

M: Google Translate verwendest du ab und zu?

T3: Sehr selten, aber ist schon mal vorgekommen.

M: Weißt du, ob das eher privat war oder eher beruflich?

T3: Beides.

M: Kannst du mir sagen, wie du mit Google Translate zufrieden warst, warst du beeindruckt oder findest du es ok oder findest du es schlecht?

T3: Ich finde es ok.

M: Hast du es vielleicht jemals für Themen im Fachgebiet Musik verwendet? Musiktexte oder Musiktheorie oder sowas?

T3: Ja, doch, immer wieder.

M: Hast du vielleicht vom Übersetzungsprogramm DeepL gehört?

T3: Noch nicht.

M: Kannst du mir sagen, würdest du dich eher als Befürworter oder Gegner von MÜ bezeichnen?

T3: Ah, so, das ist mir eigentlich egal, solange es richtig übersetzt wird.

M: Wo siehst du persönlich Vor- und Nachteile von MÜ?

T3: Die Vorteile sind, dass es schnell geht; die Nachteile sind, dass verschiedene Sprachen haben sehr verschiedene Auslegungsmöglichkeiten; etwas, was ich in einer Sprache mit einer Metapher ausdrücken kann, wird nicht zwangsläufig in einer anderen Sprache das gleiche bedeuten. Aber im Grunde genommen, wenn wir 1 zu 1 die Puzzle-Teile in der Sprache und die Wörter so zusammenbasteln, ist es ganz praktisch. Das geht schnell, muss allerdings nicht immer das bedeuten, was man zum Ausdruck bringen möchte.

M: Kannst du mir bitte sagen, ob du Erfahrung gemacht hast, mit menschlichen Übersetzer*innen und Dolmetscher*innen? Z. B. hast du jemals eine Übersetzung beauftragt oder etwas ähnliches?

T3: Ja.

M: Wie zufrieden warst du mit der Qualität?

T3: Spärlich. Nicht zufrieden.

M: Meine nächste Frage wäre gewesen: wie wusstest du, dass die Qualität angemessen war. Aber sagen wir mal so: woher wusstest du, dass die Qualität nicht angemessen war?

T3: Naja, die Sprachen, die gedolmetscht oder übersetzt wurden, ich kenne diese Sprachen und hab gehört, dass es komplett falsch war. Also, das waren einfach falsche Übersetzungen. Komplette falsch. Nicht richtig, nicht schlecht, sondern falsch.

M: Hast du selbst mal Texte übersetzt oder gedolmetscht, privat, beruflich?

T3: Ja.

M: Mit Hilfe von MÜ oder ohne?

T3: Beides. Aber maschinell eher weniger. Durch die Sprachkenntnisse, die ich aufgeschrieben habe, diese Sprachen habe ich ziemlich gut drauf, also dahingehend konnte ich es ziemlich gut, ohne MÜ. Aber ab und zu greift man schon zurück dazu; wie schreibt sich das Wort oder wie übersetze ich dieses und jenes. Manchmal gibt es vier oder fünf Wörter im Englischen für [ein Wort im Deutschen].

M: Und wie ist es dir dabei ergangen? Wie hast du dich bei den Übersetzungen gefühlt? Denkst du, es ist schwer zu übersetzen, oder findest du es eher leicht?

T3: Zweiteres, eher leicht.

M: Ich nehme mal an, du hast nie an so einer Bewertung von Übersetzungsqualität teilgenommen, oder?

T3: Nein.

M: Meine nächste Frage bezieht sich auf Übersetzungen von musikwissenschaftlichen Texten. Empfindest du einen Bedarf an Übersetzungen von Texten zum Thema Musiktheorie, bzw. sind dir Texte zu diesem Thema, die auf deiner Erstsprachen verfasst sind, genug, oder findest du es gut, dass es auch Übersetzungen gibt, aus Sprachen, die du nicht kennst?

T3: Naja, das ist eigentlich eine irrelevante Frage. So wie ich das sehe, wenn ich mich dazu äußern darf, Musiktheorie... die Musik ist eine Sprache, welche keine Übersetzung bedarf... Also, wir haben C-Dur und das ist C-Dur. Bella Bartok hat irgendwas mit Vierteltönen rumexperimentiert, das hat sich nicht durchgesetzt. Ein Kammer-A hat immer 440Hz. Musik ist reinste Physik. Schall ist eine Transversalwelle die sich in alle Richtungen in zwei Dimensionen, zweidimensional ausbreitet.

M: Also eher kein Bedarf an Übersetzungen von Musiktheorie?

T3: Musik eine Sprache für sich selbst. Wie wir sie definieren, wie wir in welcher Sprache die Definition und die Musik als solche... In gewissen Ländern gibt es Kopfbewegungen für nein oder für ja, aber gehüpft wie gesprungen. Aber wir können meiner Meinung nach nicht leugnen, dass eben Musik ein physikalisches Phänomen ist. Es gab ja diesen einen Typen, der hat die Tonleitern kategorisiert, und abhängig von den Frequenzen der Töne schwing unser Körper auch mit, und das ist dann halt die Sprache der Musik. Wie wir sie ausdrücken, jetzt mit Wörtern, das ist egal ob Chinesisch oder eine andere Sprache....

M: Dann ein paar Fragen zu dieser Umfrage. Wie ist es dir ergangen?

T3: Sehr gut!

M: Waren die Fragen klar und deutlich?

T3: Zu 90 Prozent ja, wir hatten vorhin ein paar, wo ich mich nicht ausgekannt hab, aber das haben wir aufgearbeitet.

M: Findest du das das Format angemessen war?

T3: Ja, passt schon.

M: Fandest du es leicht eine Entscheidung zu treffen, welche der Übersetzungen von Menschen erstellt wurden?

T3: Im Grunde genommen ja.

M: Und woran erkanntest du, ob eine Übersetzung von Menschen erstellt worden ist?

T3: Je nachdem, welchen Unsinn jemand geschrieben hat.

M: Eher inhaltlich oder die Wortauswahl?

T3: Beides, also es war schon ein bisschen nach Gefühl eben durch die Sprache, die verwendet wurde, habe ich mir gedacht: das würde doch niemals ein Mensch schreiben, oder bei dem anderen, das würde doch niemals eine Maschine schreiben. Also es ist so: mal so, mal so. Manche waren so hmmm....

M: Sonstiges Feedback zum Thema oder dieser Umfrage?

T3: Nein.

M: Danke!

12.4. Interview Teilnehmer*in 4:

M: Das ist das Interview zur Umfrage Nr. 4. Könntest du mir bitte noch einmal deinen Vor- und Nachnamen bestätigen?

T4: (Anonymisiert)

M: Ja, danke dir, und danke, dass du dir die Zeit genommen hast für die Umfrage und das Interview; diese Audioaufnahmen werden streng vertraulich behandelt. Kannst du mir als allererstes sagen – hast du mal MÜ-Systeme verwendet, also etwas wie Google Translate z. B.

T4: Ich verwende es eigentlich nur auf Social Media, YouTube, Facebook oder so, wenn ein Kommentar in einer Sprache ist, die ich nicht lesen kann. Wenn jemand auf Arabisch antwortet oder so, oder wenn Freunde von mir, die eine andere Muttersprache haben irgendwas in der Sprache schreiben, die ich nicht kann.

M: Ist das eher privat oder beruflich?

T4: Eher privat.

M: Und so im Alltag verwendest du es, also, für Kommentare und so. Wie bist du da zufrieden mit der Qualität solcher Übersetzungen?

T4: Eh, es kommt auf die Sprache an. Also, z. B. wenn Farsi oder sowas übersetzt wird ist es oft sehr schwer verständlich. Bei Europäischen Sprachen wie Französisch oder Spanisch ist die Übersetzung recht gut, aber vor allem bei Arabisch und Farsi ist mir immer aufgefallen, dass Sie irgendwie... es ist sehr schwer zu verstehen, was gemeint ist.

M: Und würdest du sagen, dass dir solche maschinelle Übersetzung mal ermöglicht hat einen Text zu verstehen, denn du sonst nicht verstehen könntest?

T4: Ja.

M: Ich nehme mal an, du verwendest da am meisten Google Translate, oder?

T4: Oder die eingebauten Funktionen, aber ich schätze jetzt YouTube wird wahrscheinlich Google Translate verwenden.

M: Nehme ich auch an, ja.

T4: Und die eingebaute Funktion von Facebook verwende ich auch.

M: Und deine Kommentare bis jetzt beziehen sich hauptsächlich auf Google Translate und auf diese eingebaute Funktion von Facebook.

T4: Ja.

M: Hast du schon vom System DeepL gehört?

T4: Nein.

M: Würdest du sagen du bist eher Befürworter oder Gegner maschineller Übersetzung?

T4: Ich finde es gut, dass es die Möglichkeit gibt, aber wenn ich jetzt z. B. so musiktheoretisches Buch oder sowas lesen muss, oder lernen will, wäre es mir schon lieber... also es kommt drauf an, wenn es die Qualität erreicht einmal, dass es so ist, das es gut ist, aber da war es schon manchmal eindeutig, dass die Maschine nicht versteht, worum es geht, und Ausdrücke verwendet, die nicht passen, und ich würde da gerne was in der Qualität lesen, wenn die Maschine da was zusammen bringt, dann habe ich nichts dagegen.

M: Und wo siehst du Vorteile und Nachteile von MÜ?

T4: Also ich halte es für wirtschaftlich super. Ich will Übersetzern nicht ihre Arbeit absprechen, aber es ist eine irrsinnige Zeitverschwendung da menschliche Leistung zu investieren in etwas, was eine Maschine machen kann, ich werde ja keine Sense verwenden, wenn es einen Mähdrescher gibt.

M: Und Nachteile?

T4: Also den Nachteil, den ich sehen kann, ist eben sich nur auf die Maschine zu verlassen und dass einfach die Qualität nicht stimmt, oder dass der Sinn verloren geht, wenn es niemand, der kompetent ist, es checkt noch einmal, das wäre kompletter Schwachsinn. Aber solange man es als Hilfsmittel verwendet, wenn es nicht Probleme darstellt, dann sehe ich keine großen Nachteile. Wenn der Sinn erhalten bleibt...

M: Mit menschlichen Übersetzer*innen – hattest du da Erfahrung? Oder mit Dolmetscher*innen? Hast du jemals eine Übersetzung beauftragt oder sowas?

T4: Ich übersetze selbst oft Lieder von Englisch auf Dialekt, auf Steirisch, und mein Mitbewohner ist Slowene und immer, wenn etwas auf Serbo-Kroatisch oder Slowenisch ist, dann lass ich mir von ihm helfen, zu erklären, was es ist, aber nicht von professionellen Übersetzern. Im Alltag nur von zweisprachigen Freunden.

M: Ja, meine nächste Frage wäre gewesen, ob du selbst mal etwas übersetzt oder gedolmetscht hast. Und wie ergeht es dir dabei – findest du es leicht, schwer, spannend – wie würdest du es bezeichnen?

T4: Also, ja, Englisch-Deutsch übersetzen ist für mich ganz natürlich dadurch, dass ich Medien sehr viel auf Englisch konsumiere, dadurch ist es für mich wie meine zweite Muttersprache. Und oft, wenn jemand nicht so gut Englisch kann oder so, oder wenn man nur Englisch als die gemeinsame Sprache in einer Gruppe hat, dann würde ich da oft bisschen aushelfen, das finde ich ganz natürlich.

M: Okay, und zurück zur vorherigen Frage in Bezug auf Erfahrung mit menschlichen Übersetzer*innen und Dolmetscher*innen: hast du mal eine Übersetzung beauftragt, oder bei einem Event mit Simultandolmetschen teilgenommen, oder so?

T4: Nie mit einem professionellen Übersetzer. Teilweise durch die Arbeit mit Musik, mit einem Freund von mir, haben wir gemeinsam Lieder gemacht. Ich tue den Text auf Deutsch ein bisschen auslegen, und der übersetzt ihn auf Hebräisch, z. B. und dann machen wir das Lied auf Hebräisch. Aber nie mit professionellen Übersetzern, immer nur mit Leuten mit zwei Sprachen oder so.

M: Und ich nehme mal an, du hast nie an so einer Bewertung von Übersetzungsqualität teilgenommen, oder?

T4: Nein.

M: So, jetzt ein paar Fragen in Bezug auf Übersetzungen von musikwissenschaftlichen Texten. Empfindest du einen Bedarf an Übersetzungen von Texten zum Thema Musiktheorie, oder sind dir Texte zu diesem Thema, die auf deiner Erstsprachen verfasst sind, genug, oder in den Fremdsprachen, die du kennst?

T4: Also, ja, würde auf jeden Fall genügen. Es gibt eh viel auf Deutsch zu Musiktheorie, das kann man nicht in einem Leben lesen. Aber natürlich ist es super, dass es Übersetzungen gibt, dass mehr gemacht wird. Je mehr Sprachen, umso besser.

M: Und würdest du sagen, für Fremdsprachige musiktheoretische Texte, denkst du, dass dieser Bedarf an Übersetzungen durch MÜs gedeckt werden kann, oder benötigen wir immer noch menschliche Übersetzer*innen?

T4: Also, ich glaube eine Kombination eben. Bei repetitiven Sachen macht es Sinn, Maschinen zu verwenden, aber es muss jemand nachkontrollieren, glaube ich, damit der Sinn nicht verloren geht. Eine einzige Sache, wo man sagen kann, dass es mehr Übersetzung geben könnte, obwohl es auf Englisch etwas gibt, ist alles, was die indische Musiktheorie umfasst. Also auch andere größere nicht-westlichen Musikkulturen – ostafrikanisch, westafrikanisch, indisch und südamerikanisch. Und da hat man halt Musiker aus all diesen Gebieten, die verwenden nicht alle Englisch und Spanisch, das heißt: durch die Nachwirkung des Kolonialismus ist es so, dass die Musiker, die sich richtig gut auskennen, immer noch eine Europäische Sprache gut beherrschen.

M: Ein paar Fragen noch zu dieser Umfrage. War die Fragestellung klar und deutlich?

T4: Ja.

M: War das Format angemessen?

T4: Ja, es war sehr gut und übersichtlich.

M: Fandest du es leicht zu entscheiden, welche Übersetzungen von Menschen erstellt wurden?

T4: Nicht immer, bei meisten Fragen schon, bei ein paar war ich mir nicht sicher. Aber da waren halt Sachen wie z. B., auf Englisch heißt es Minor 7th, Major 7th, Diminished 7th, auf Deutsch heißt es große oder kleine 7, oder Septim, und dann hats aber der Bot übersetzt mit Dur und Moll 7, und das würde ein Mensch niemals machen. Oder wo Grundton als Wurzel übersetzt wurde, ich weiß halt, dass das im Englischen, dass das dasselbe Wort ist im Englischen.

M: Und wie sicher bist du dir, dass du alle menschlichen Üs richtig erkannt hast?

T4: Ich bin mir bei 4 sehr sicher, bei 2 bin ich mir nicht 100 Prozent sicher.

M: Sonstiges Feedback zum Thema oder zur Umfrage?

T4: Es hat mir Spaß gemacht. Ich lese gerne musiktheoretische Texte, und es war witzig, das Raten. Und ich finde, es ist eine gute Idee, dass man da was weiterbringt, denn im Endeffekt, wenn es mehr

Übersetzung gibt, aus vielen Sprachen, zu dem Thema vor allem, die Musiker profitieren davon, das finde ich gut.

M: Mich freut es, dass es dir Spaß gemacht hat. Danke!

12.5. Interview Teilnehmer*in 5:

M: So, das sind die Aufnahmen zur Umfrage Nummer 5. Kannst du mir nur noch einmal den Vor- und Nachnamen bestätigen?

T5: (Anonymisiert)

M: Ja, erstmal ein paar Fragen zum maschinellen Übersetzen insgesamt. Verwendest du ab und zu MÜ-Systeme, und falls ja, ist das eher privat oder beruflich, für welche Zwecke?

T5: Ich verwende den DeepL Übersetzer ab und zu, meistens wenn ich zu faul bin, um welche E-Mail-Korrespondenzen [zu erledigen]. Es geht einfach schneller und ich finde es zuverlässig genug.

M: Ist das eher privat oder beruflich?

T5: Eher beruflich.

M: Darfst du sagen, worum es da geht ungefähr?

T5: Ja, musikbezogen, musikwissenschaftsbezogen. Korrespondenzen, wenn ich beispielsweise ein Abstract für eine Konferenz einreiche.

M: Wie zufrieden bist du insgesamt mit der Qualität von MÜs?

T5: Ich finde, die Qualität variiert sehr stark von Maschine zu Maschine. Google Translator finde ich für Englisch mittlerweile auch ganz okay, für Französisch eigentlich auch teilweise ganz okay, aber DeepL finde ich ganz viel besser.

M: Also aus den Sprachen [Englisch und Französisch], oder in die Sprachen?

T5: In die Sprachen.

M: Also du kennst sowohl Google Translate als auch DeepL. Wie würdest du deine Zufriedenheit da bewerten, bei den beiden? Denkst du, dass du sie auch in Zukunft verwenden wirst?

T5: Ja, DeepL werde ich in Zukunft bestimmt verwenden.

M: Und insgesamt wurdest du sagen, dass du zufrieden bist, ja?

T5: Ja, also für die Zwecke, für die ich es verwende, ist es für mich zufriedenstellend.

M: Würdest du sagen du bist eher Befürworterin oder Gegnerin maschineller Übersetzung?

T5: Ehm, das kann ich nicht mit ja oder nein beantworten, weil es stark von der Zweckgebundenheit abhängt. Jetzt rein maschinell irgendwas übersetzen zu lassen, insbesondere wenn es um fachliche Inhalte geht, das würde ich auf garkeinen Fall machen.

M: Und wo siehst du Vorteile und Nachteile von MÜ?

T5: Also die Vorteile liegen mit Sicherheit bei der Geschwindigkeit und das, außer Sachen wie Rechtschreibfehler oder so nicht passieren können, aber wie bei jedem Übersetzungsprozess entstehen Verfälschungen in der Inhaltstransportierung und wenn das automatisiert passiert, sind diese [Spaltungen] die entstehen viel größer. Das ist ein großes Risiko.

M: Dann ein paar Fragen zu menschlichen Übersetzer*innen und Dolmetscher*innen. Kannst du mir sagen, hast du mit Übersetzer*innen oder Dolmetscher*innen Erfahrung gehabt, oder dass du eine Übersetzung beauftragt hast?

T5: Ja.

M: Was genau?

T5: Ich habe einen Vortrag übersetzen lassen und Abstracts für Konferenzen.

M: Warst du zufrieden mit den Übersetzungen?

T5: Mal so, mal so.

M: Wie wusstest du, dass die Übersetzungsqualität angemessen oder nicht angemessen war?

T5: Ich habe erstmal den Inhalt abgeglichen, Satz für Satz, oder eben auch in größeren Abschnitten. Und dann konnte ich eben erkennen, ob das [dem Original] genau entspricht oder nicht. Also meine Sprachkenntnisse, insbesondere Englisch, sind auf sehr hohem Niveau. Ich gib das immer nur in andere Hände, weil ich keine Zeit habe, selbst zu übersetzen, deshalb kann ich schon wirklich prüfen, ob das passt.

M: Und das waren meistens Texte, in Sprachenkombinationen, wo du alle Sprachen mehr oder weniger kennst.

T5: Ja.

M: Vielleicht etwas mit Sprachen, die du überhaupt nicht sprechen kannst?

T5: Nein, da habe ich keine Erfahrungen.

M: Hast du selbst Texte übersetzt oder gedolmetscht, und falls ja, wie ist es dir dabei ergangen bzw. wie hast du dich dabei gefühlt?

T5: So etwas habe ich nur in Alltagssituation erlebt, mit Freunden, die manche Sprachen nicht beherrschen.

M: Hast du dabei auch MÜ verwendet oder nicht?

T5: Nein, bei den Dolmetschungen nicht.

M: Du hast dich dabei also auf deine Sprachkenntnisse verlassen?

T5: Ja.

M: Kommst du als Musikerin regelmäßig in Kontakt mit Texten, die nicht in einer deiner Erstsprachen sind? Also mit Übersetzungen.

T5: Ja.

M: Und als Musikerin bzw. Musikwissenschaftlerin, empfindest du einen Bedarf an übersetzten Musik-bezogenen Texten, oder genügen dir Texte in deiner Erstsprache?

T5: Also ich bevorzuge immer die Lektüre in der Originalsprache, aber es ist schon sehr hilfreich, wenn Aufsätze z. B. auf Italienisch, wobei ich kein Italienisch kann, wenn eine Übersetzung vorliegt, das ist schon hilfreich. Und insbesondere andersrum, das habe ich festgestellt, viele Englischsprachige, also englische Muttersprachler, sind sehr froh, wenn Deutsche Texte eben auch in Englisch vorliegen.

M: Und würdest du sagen, dass dieser Bedarf durch MÜs gedeckt werden, oder benötigen wir immer noch menschliche Übersetzer*innen?

T5: Ich bin ganz fest davon überzeugt, dass eine MÜ, grade wenn es so um etwas spezifisches geht wie Musiktheorie geht, überhaupt nicht ausreicht. Weil das Vokabular so spezifisch ist, da kann eine Maschine das nicht erkennen, wenn sie nicht dafür gebaut worden ist, und ich glaube nicht, dass es eine Maschine gibt, die dafür gebaut worden ist.

M: Kannst du mir noch ein paar Sachen zur Umfrage sagen? War die Fragestellung klar und deutlich? Wie ist es dir da insgesamt ergangen?

T5: Ja, ich finde, du hast es gut erklärt, und mit dem Beispiel vorweg hat das alles für mich sehr deutlich gemacht.

M: Würdest du sagen, dass das Format der Umfrage angemessen war?

T5: Ja, ich würde jetzt überlegen müssen, was der Mehrwert ist, dass wir das jetzt persönlich machen, statt dass es online ist, aber es ist auch eine größere Verbindlichkeit da, wenn man sich vor Ort trifft.

M: Fandest du es leicht, zu entscheiden, welche Übersetzungen von Menschen erstellt wurden?

T5: In manchen Fällen fand ich es sehr eindeutig, ich bin gespannt, was das Ergebnis bringt. Bei manchen anderen hab ich mich schwer getan.

M: Und woran erkanntest du, dass eine Übersetzung von einem Menschen ist und nicht von einer Maschine?

T5: Insbesondere anhand von manchen Fachausdrücken, die einfach inadäquat übersetzt wurden, beispielsweise Wurzel, das ist vom englischen Root nehme ich an, solche Sachen. Oder auch Verben die komisch sind, oder Satzbau...

M: Und wie sicher bist du dir, dass du alle menschlichen Üs richtig erkannt hast?

T5: Boah. Hmmm... Ob ich die Mehrheit geschafft habe, das denke ich schon.

M: Sonstiges Feedback zum Thema oder zur Umfrage?

T5: Ich finde es sehr interessant und gut, dass du das machst. Und ich bin gespannt, was die Ergebnisse bringen werden.

M: Wie gesagt, deine Daten, incl. dieser Audioaufnahmen, werden streng vertraulich behandelt. Ich danke dir für die Umfrage und das Interview!

12.6. Interview Teilnehmer*in 6:

M: So, das sind die Aufnahmen zur Umfrage Nummer 6. Kannst du mir noch einmal deinen Namen bestätigen?

T6: (Anonymisiert)

M: Danke dir! Deine Daten werden streng vertraulich behandelt, incl. dieser Audioaufnahmen, und werden nicht an Dritte weitergeleitet. So, als erstes ein paar Fragen zum MÜ. Hast du jemals MÜ-Systeme verwendet, und falls ja, für welche Zwecke war das? War es eher privat oder beruflich?

T6: In der Schule, beim Latein-Übersetzen als Hilfestellung, den Google Übersetzer. Und wenn ich aus Sprachen übersetzen will, die ich nicht kann, also überhaupt nicht kann.

M: Okay, hast du es vielleicht jemals zum Thema Musik oder Musiktheorie verwendet, oder für Song-Texte oder sowas?

T6: Nein.

M: Wie zufrieden bist du insgesamt mit der Übersetzungsqualität von MÜ?

T6: Also, bis jetzt habe ich immer gelacht darüber, aber jetzt bin ich mir nicht mehr sicher darüber, nach dem Lesen [der Texte in der Umfrage].

M: Hat dir maschinelle Übersetzung mal ermöglicht hat einen Text zu verstehen, denn du sonst überhaupt nicht verstehen könntest?

T6: Ja.

M: Du hast schon Google Translate erwähnt und hast es verwendet. Deine Erfahrung und Zufriedenheit mit der Qualität von MÜ basieren also hauptsächlich auf deiner Erfahrung mit Google Translate, ja?

T6: Ja.

M: Hast du vielleicht vom Übersetzungsprogramm DeepL gehört?

T6: Nein.

M: Würdest du sagen du bist eher Befürworterin oder Gegnerin maschineller Übersetzung?

T6: Also, wenn es gut genug ist und wenn es jetzt nicht auf komplette Genauigkeit darauf ankommt, sondern nur um Informationswiedergabe, warum nicht.

M: Wo siehst du Vorteile und Nachteile von MÜ?

T6: [Ein Nachteil], dass es so schlecht übersetzt wird, dass die Information verloren geht. Vorteil wäre, weniger Arbeit für Menschen. Ich weiß nicht, das könnte auch ein Nachteil sein.

M: Hast du bis jetzt mit menschlichen Übersetzer*innen oder Dolmetscher*innen Erfahrung gehabt, oder dass du eine Übersetzung beauftragt hast oder etwas ähnliches?

T6: Im Freundeskreis habe ich schon Dolmetscher, aber sonst nicht.

M: Hast du selbst Texte übersetzt oder gedolmetscht, und falls ja, wie ist es Ihnen ergangen, wie hast du dich dabei gefühlt?

T6: Ja, es ist manchmal schwieriger und manchmal leichter. Aber manchmal kann man halt Wörter nicht 1 zu 1 übersetzen.

M: War das eher beruflich oder eher privat?

T6: Privat.

M: War das mit Hilfe von MÜ oder ohne?

T6: Meistens mit digitalen Wörterbüchern.

M: An einer Bewertung von Übersetzungsqualität, außer dieser hier, hast du nicht teilgenommen, oder?

T6: Nein.

M: Kommst du als Musikerin regelmäßig in Kontakt mit musiktheoretischen Texten, die nicht in einer deiner Erstsprachen sind?

T6: Nein.

M: Empfindest du einen Bedarf an Übersetzungen von Texten zum Thema Musiktheorie, bzw. sind dir Texte zu diesem Thema, die auf deiner Erstsprachen verfasst sind, genug?

T6: Ich empfinde kein Bedürfnis für Übersetzungen.

M: Dann noch ein paar Fragen zu dieser Umfrage: war die Fragestellung klar und deutlich?

T6: Ja.

M: Und war das Format angemessen?

T6: Ich glaube schon, ja.

M: Fandest du es leicht zu entscheiden, welche Übersetzungen von Menschen erstellt wurden?

T6: Das war eigentlich am leichtesten.

M: Woran erkanntest du, welche Übersetzungen von Menschen erstellt wurden?

T6: Indem sie flüssig zum Lesen waren.

M: Und wie sicher bist du dir, dass du alle menschlichen Übersetzungen richtig erkannt hast?

T6: So 60 Prozent.

M: Sonstiges Feedback zum Thema oder zur Umfrage?

T1: Nein.

M: Ich bedanke mich für das Interview!

12.7. Interview Teilnehmer*in 7:

M: So, das ist das Interview zur Umfrage Nr. 7. Kannst du mir nur noch einmal den Vor- und Nachnamen bestätigen?

T7: (Anonymisiert)

M: Deine Daten, incl. dieser Audioaufnahmen, werden streng vertraulich behandelt. Nun zu den Fragen: hast du jemals MÜ verwendet?

T7: Ja.

M: Kannst du dich erinnern, für welche Zwecke das war, war es eher privat oder eher beruflich?

T7: Beides eigentlich.

M: Vielleicht ein Beispiel, wofür genau?

T7: Beruflich – englische Texte, alle Arten von romanischen Sprachen für einen globalen Helpdesk, um schneller zu sein, um nicht viel per E-Mail machen zu müssen, sondern einfach nur Copy-Paste und dann ausbessern, mit durchwachsenen Ergebnissen. Diverse Online-Lösungen, nix gezahltes, aber war nicht so gut eigentlich. Aber probiert habe ich. Privat eher so... Computer-Spiele, wenn man vielen was geschickt hat, wenn man Communities verwaltet hat oder so, bei Online-Spielen.

M: Welches System hast du dabei verwendet?

T7: Entweder Google-Translate oder BabelFish. Selten was anderes, kann mich nicht erinnern. Die waren so schlecht, dass ich sie nie wieder probiert habe.

M: Und im Alltag verwendest du es eher nicht?

T7: Nein, nicht mehr. Die sind so schlecht, dass ich... sagen wir so, es liefert so schlechte Ergebnisse, dass es eher kontraproduktiv ist. Man ist länger dran mit dem Zeug rumzuspielen, als dass man es gleich selbst schreibt.

M: Also das ist auch deine Antwort zur Frage zur Übersetzungsqualität von MÜ-Systemen insgesamt? Eher unzufrieden?

T7: Zusammenfassend würde ich sagen, dass die Korrektur von maschinellen Übersetzungen, in meiner Erfahrung, die nicht ganz aktuell ist, ein paar Jahre alt, ist, dass man mehr mit dem Ausbessern beschäftigt ist, dass es mehr Zeit in Anspruch nimmt, als es gleich selbst zu schreiben.

M: Und würdest du sagen, dass dir MÜ hilft, Texte zu verstehen, die du sonst nicht verstehen könntest?

T7: Ja, durchaus, also mit meinen Spanisch-, Italienisch-, Französisch-Kursen, die eher nur bis fortschrittlichem Level waren, darum herum waren, konnte ich viel nicht verstehen, dann war das eben, was ich vorhin erwähnt habe, wusste ich mindestens um was es geht. Auch wenn die Details verloren, verstummelt waren - die Satzstellung, die Grammatik sowieso alles falsch – aber, ja, so ein bisschen würde ich sagen war es schon hilfreich.

M: Hast du jemals von DeepL gehört?

T7: Nein. Bin nicht up-to-date.

M: Würdest du sagen du bist eher Befürworter oder Gegner maschineller Übersetzung? Und wo siehst du Vor- und Nachteile?

T7: Also, ich würde jetzt nicht direkt sagen Gegner, aber auf jeden Fall bin ich kritisch, durch die eigenen Erfahrungen. Aber auch wenn sich das wesentlich verbessern würde, so wie in der Umfrage, das war relativ gut eigentlich, erstaunlich gut. Ja, es nimmt Arbeitsplätze weg, also für die, die Sprachen studieren, ist es nicht so lustig, das ist zum Beispiel ein Nachteil. Oder eben die Erfahrung, die ich schon gemacht habe, die hat ja auch Nachteile beinhaltet. Vorteile - sicherlich für Leute mit wenigen Ressourcen, z. B. eine Website betreiben, internationales Publikum bedienen, dann sind diese Dinge, sage ich mal natürlich schon hilfreich, wenn es gratis ist, hat man wenigstens etwas und nicht Garnichts. Oder als Student, mit dem, was in Browsern eingebaut ist auch u. A. sicher im Chromium, dass man wenigstens weiß, worum es geht. Also als Schnellhilfe würde ich es positiv bewerten, aber alles, was ernsthaft ist und halbwegs professionell sein soll, würde ich Menschen machen lassen.

M: Kannst du mir sagen, hast du mit menschlichen Übersetzer*innen oder Dolmetscher*innen Erfahrung gehabt, oder dass du eine Übersetzung beauftragt hast oder sowas?

T7: Sowohl als auch. Ich habe mehr selber übersetzt, als Übersetzungen von anderen gebraucht. Eine gerichtliche Sache auf Spanisch habe ich mir übersetzen lassen, da hat mir der (Anonymisiert) geholfen, das ist kürzlich gewesen, da war die Übersetzung sehr gut. Wobei dann diese Sachen wie Musik oder Gericht dementsprechend schwerer sind. Aber hätte der nicht super Spanisch gesprochen... Ich glaube nicht, dass das bei Online Übersetzern grad Fachthemen [vertreten sind]. Und wenn Geld da ist, auf jeden Fall Menschen beauftragen. Ich finde man kann ja immer neue Technologien einführen, aber schon mit Rücksicht auf die Leute, vielleicht dass dann in dem Tempo einführen, dass alle eine Umschulung kriegen, die dann arbeitslos werden, und nicht einfach zack-bum und dass Schicksale irgendwie... Ehm, was war nochmal die Frage?

M: Hast du mit menschlichen Übersetzer*innen oder Dolmetscher*innen Erfahrung gehabt, oder dass du eine Übersetzung beauftragt hast oder sowas?

T7: Also, wenn ich Geld habe, Menschen. Ich habe auch viel für Kollegen übersetzt, wo es nicht genau sein muss. Es ist für mich auch ein Vertrauensproblem. Wenn ich einen Menschen übersetzen lasse, da weiß ich oder vertraue ich, dass jemand, der vernünftig denkt und Hausverstand hat [den Text übersetzt]. Wenn eine Maschine das macht, und vor allem wie Maschinen weltweit manipuliert werden – das weiß man mittlerweile sehr gut – vertraue ich nicht, dass da etwas korrekt übersetzt ist. Also für alles, was halbwegs ernsthaft, nicht unbedingt professionell, wenn es nur ernsthaft und vertrauensvoll sein soll, keine Maschinen. Und wenn ich eine Übersetzung brauche, dann versuche ich das so zu leben.

M: Als du selbst übersetzt hast, das war teilweise mit Hilfe von MÜs?

T7: Ja, eine Weile lang, bis ich es erprobt habe. Bis ich dieses Urteil gefällt habe, wie vorhin erzählt.

M: An einer Bewertung von Übersetzungsqualität, außer dieser hier, hast du nicht teilgenommen, oder?

T7: Nein.

M: Kommst du als Musikerin regelmäßig in Kontakt mit musiktheoretischen Texten, die nicht in einer deiner Erstsprachen sind?

T7: Ja, weil meine Freunde und Bekannte überall verstreut sind auf der Welt. Wenn dann halt jemand kommt, der Musik macht, z. B. ein rumänischer Freund der in Tschechien lebt, und wenn dann andere mitmusizieren vor allem ist es dann ziemlich schnell Englisch, Tschechisch meistens nicht. Mindestens mit Englisch muss man schon rechnen. Vor allem jetzt auch, in Corona-Zeiten, mit guter Internet-Verbindung online musiziert mit anderen, dann ist natürlich auch die [verwendete] Sprache Englisch. Und das ist ziemlich ätzend, weil Musiktheorie ist schon auf Deutsch ätzend.

M: Das ist dann öfters mündliche Kommunikation. Wie schaut es aus mit schriftlichen Texten?

T7: Da mache ich [mit Fremdsprachen] gar nichts.

M: Noch ein paar kurze Fragen zu dieser Umfrage. Wie ist es Ihnen ergangen? Waren die Fragen klar und deutlich? Fanden Sie das Format angemessen?

T7: Es war alles klar und deutlich. Die mündliche Erklärung war auch gut. Man hätte gar nicht die geschriebene Erklärung durchlesen müssen. Aber ich habe sie trotzdem durchlesen müssen. Jedenfalls, wenn ich ein Problem hatte, dann das, dass ich Musiktheorie hasse und teilweise nicht bewerten konnte, ob das korrekt ist oder nicht, sondern da ich viel nach Gefühl mache, musste ich wirklich nachdenken. Eigentlich mehr über Musiktheorie nachdenken als über das Format oder sonst was. Ich denke es war eh gut gemacht. Ich hätte eher gedacht, dass man jedes Feld eintragen muss in der ersten Tabelle, das war ein bisschen überraschend, dass man nur zwei eintragen muss, aber es ist auch egal eigentlich.

M: Fandest du es leicht zu entscheiden, welche Übersetzungen von Menschen erstellt wurden, und wie hast du deine Entscheidung getroffen?

T7: Es war überhaupt nicht leicht. Wie gesagt waren die erstaunlich gut, vor allem als Fachbereichsübersetzung, die immer noch schlechter war. Es war wirklich schwer. Es gab ein paar, wo es offensichtlich war. Eins, zwei wo einfach keine logischen Zusammenhänge zu erkennen waren. Da war es natürlich leicht, wenn es null Aussagen hat, dann kann ich leicht als MÜ identifizieren. Sonst halt nach Gefühl. Ich habe es mit meinen früheren Erfahrungen verglichen. Ich habe versucht über Satzfehler, Beistrichfehler, Rechtschreibfehler, da hatte ich aber nicht viel Erfolg, weil es schon recht gut war. Sonst nur darüber: wer hat die Logik richtig, und wer hat es schöner formuliert. Menschen tendieren – auch Übersetzer – eine künstlerische Freiheit reinzubringen. Und maschinelle Programme machen so Dinge wie Worte, die ein Mensch in dem Kontext nicht verwenden würde, zu verwenden. Einen Menschen spürt man schon zwischen den Zeilen, auch emotional.

M: Sonstiges Feedback zum Thema oder zur Umfrage?

T1: Nein.

M: Ich bedanke mich für das Interview!

Abstract

Deutsch:

Das Ziel dieser Masterarbeit war es, herauszufinden, wie Musiker*innen und Musikwissenschaftler*innen die unbearbeiteten deutschen Übersetzungen von hochspezialisierten, Fachgebiet-spezifischen englischen Texten aus dem Bereich der Musik erstellt von neuronalen maschinellen Übersetzungssystemen im Vergleich zur menschlichen Übersetzung bewerten und ob sie konsistent zwischen maschineller und menschlicher Übersetzung unterscheiden können.

Dies wurde untersucht, indem insgesamt sechs Textabschnitte jeweils viermal übersetzt wurden (jeweils von zwei menschlichen Übersetzer*innen, DeepL und Google Translate) und Fachexpert*innen im Fachgebiet Musik zur Evaluierung gegeben wurden. Die Evaluierung erfolgte mittels der Best-Worst-Scaling Methode wobei Evaluatord*innen der besten Übersetzung eine positive und der schlechtesten eine negative Punktzahl zuwies. Darüber hinaus wurden die Evaluatord*innen gebeten anzugeben, welche der Übersetzungen sie für menschliche Übersetzungen und welche für maschinelle Übersetzungen hielten.

Das Ergebnis dieser Untersuchung war, dass die Evaluatord*innen beide menschlichen Übersetzer*innen besser bewerteten als die beiden NMÜ-Systeme, wobei die von DeepL erstellten Übersetzungen deutlich besser bewertet wurden als jene, die von Google Translate erstellt wurden. Die Ergebnisse der numerischen Bewertung lassen den Schluss zu, dass es für MÜ-Systeme in bestimmten Fällen möglich ist, das Qualitätsniveau einer menschlichen Übersetzung zu erreichen. Die Ergebnisse in Bezug auf die zweite Frage – ob eine bestimmte Übersetzung von einem Menschen oder einer Maschine angefertigt wurde – zeigen jedoch, dass es immer noch einen deutlichen Unterschied zwischen den beiden Arten von Übersetzungen gibt und dass Expert*innen in diesem bestimmten Fachbereich maschinelle Übersetzungen recht gut als solche erkennen können.

English:

The aim of this master's thesis was to find out how musicians and musicologists evaluate unedited German translations of highly specialized, domain-specific English texts from the field of music produced by neural machine translation systems in comparison to human translation, and whether they can consistently distinguish between machine and human translation.

This was examined by translating a total of six sections of text four times each (by two human translators, DeepL, and Google Translate) and giving them to experts in the field of music for evaluation. The evaluation was done using the best-worst scaling method whereby evaluators assigned a positive score to the best translation and a negative score to the worst. In addition, evaluators were asked to indicate which of the translations they thought were human translations and which were machine translations.

The result of this survey was that the evaluators rated both human translators better than the two NMT systems, with the translations produced by DeepL being rated significantly better than those produced by Google Translate. The results of the numerical evaluation suggest that it is possible for MT systems to reach the quality level of a human translation in certain cases. However, the results related to the second question - whether a particular translation was done by a human or a machine - show that there is still a clear difference between the two types of translations and that experts in this particular field can recognize machine translations as such quite well.