

11 Terminologie in der neuronalen maschinellen Übersetzung

Tom Winter, Daniel Zielinski

11.1 Einleitung

Globale Technologietrends wie maschinelles Lernen und künstliche Intelligenz haben in der jüngsten Vergangenheit die Entwicklung von Produkten und Dienstleistungen stark beeinflusst. Voraussetzungen dafür waren die Verfügbarkeit leistungsstarker Grafikprozessoren (Graphics Processing Unit = GPU), neuer Algorithmen im Bereich des maschinellen Lernens und ausreichend großer, qualitativ hochwertiger Datenmengen.

Viele international tätige Unternehmen haben seit Beginn der Internationalisierung und Globalisierung Ende der 1980er / Anfang der 1990er Jahre eine große Menge ein- oder mehrsprachiger Sprachdaten in Form von Text- und Audiodaten angesammelt – allen voran große Internet-Konzerne wie Amazon, Baidu, Facebook, Google, Microsoft, Yandex etc. Dies hat dazu geführt, dass Sprache schon früh einer der Hauptanwendungsbereiche bei der Entwicklung maschineller Lernsysteme und „intelligenter“ Anwendungen wurde. Schwerpunkte dabei sind Textanalyse und Textklassifikation, Spracherkennung sowie maschinelle Übersetzung.

Heute steht auf dem Markt eine Vielzahl maschineller Übersetzungsdienste für nahezu alle marktrelevanten Sprachkombinationen zur Verfügung, die über Schnittstellen (API) einfach in die eigenen Prozesse und Technologien integriert werden können.

Bei den maschinellen Übersetzungssystemen wird zwischen Systemen mit generischen Übersetzungsmodellen (Englisch: *generic* oder *stock models*) und Systemen mit domänen- oder kundenspezifischen Übersetzungsmodellen (Englisch: *custom models*) unterschieden. Generische Übersetzungsmodelle basieren auf allgemeinen Sprachdaten, während domänen- oder kundenspezifische Modelle mit Sprachdaten eines Fachgebiets oder Kunden trainiert wurden, sodass die entsprechende Sprache, der Stil, die Idiomatik und die Terminologie bei den Übersetzungen bevorzugt werden.

Die maschinelle Übersetzung ist nicht zuletzt auch durch die mediale Berichterstattung bei vielen Unternehmen und Organisationen in den Fokus des Interesses gerückt. Dies betrifft nicht nur Bereiche, die sich seit jeher mit Sprache und Kommunikation beschäftigen, wie Sprachen- und Übersetzungsdienste, sondern auch Bereiche wie IT, Marketing, Produktentwicklung, Kundendienst etc.

Maschinelle Übersetzung wird heute in vielen Kontexten verwendet. Neben der Unterstützung der professionellen Übersetzung kommt sie zur Informationsübersetzung (Gisting) und -gewinnung oder zur Echtzeitübersetzung in Dialogsystemen (Chat, Chatbots, Unified Messaging) zum Einsatz. Dabei variieren die Anforderungen an die Qualität der maschinellen Übersetzungen je nach Einsatzgebiet und -zweck, Textsorte und Sprachpaar.

Neuronale maschinelle Übersetzung (NMT/NMÜ) hat im Vergleich zur statistischen maschinellen Übersetzung (SMT/SMÜ) unabhängig vom Sprachpaar zu einer allgemeinen Verbesserung der Übersetzungsqualität geführt, insbesondere, was die Flüssigkeit (*fluency*) und die Genauigkeit (*accuracy*) betrifft (vgl. Koehn 2010, S. 217–246). Während maschinelle Übersetzungen für Informationszwecke oft schon ausreichend oder „good enough“ sind, müssen sie für Veröffentlichungszwecke meist noch post-editiert werden. Dabei wird zwischen Light Post-Editing und Full Post-Editing unterschieden. Beim Light Post-Editing stehen die inhaltliche Korrektheit und Verständlichkeit im Vordergrund, beim Full Post-Editing werden zusätzlich auch Sprache und Stil berücksichtigt (vgl. ISO 18587:2017).

Während sprachliche und stilistische Fehler relativ einfach von einem Übersetzer oder einer der Zielsprache mächtigen Person zu korrigieren sind, ist die Korrektur von Terminologiefehlern meist schwieriger und aufwändiger, da dazu zusätzliches Wissen über die betroffenen Fachgebiete sowie über kunden- oder organisationsspezifische Kontexte und Präferenzen erforderlich ist. Der Post-Editor wird mehr und mehr ein mehrsprachiger Fachgebiets-experte (Nitzke et al. 2019).

Eine Möglichkeit, den Aufwand für das Post-Editing zu optimieren und den Aufwand für die wiederholte Korrektur derselben Terminologiefehler zu reduzieren, besteht darin, die MÜ-Systeme anzupassen, sodass sie die Terminologie des Fachgebiets oder Kunden lernen und bei der Generierung der Übersetzung anderen möglichen zielsprachlichen Benennungen gegenüber bevorzugen. In der Forschung im Bereich maschinelle Übersetzung wird die Anpassung von MÜ-Modellen an bestimmte Fachgebiete *Domain Adaptation* genannt (vgl. Chu et al. 2018). Domain Adaptation wurde bereits zu Zeiten statistischer maschineller Übersetzung betrieben. Bei der Domain Adaptation

wird methodisch zwischen datenbezogenen (*data centric*) und modellbezogenen (*model centric*) Verfahren unterschieden. Der Fokus der datenbezogenen Verfahren liegt auf der Auswahl, Aufbereitung oder Generierung fachspezifischer Trainingsdaten für das Training eines domänenspezifischen MÜ-Modells (vgl. Moore et al. 2010, Wang et al. 2014, Chen et al. 2016). Modellbezogene Ansätze hingegen befassen sich mit der Anpassung der Trainingsalgorithmen oder dem Tuning bestehender Übersetzungsmodelle (vgl. Sennrich et al. 2013, Durrani et al. 2015, Zhou et al. 2015). Mit Aufkommen der neuronalen maschinellen Übersetzung begannen Forscher, die Prinzipien der Domain Adaptation auch auf neuronale Übersetzungsmodelle zu übertragen.

Die ersten Optionen zur Domain Adaptation in kommerziell verfügbaren neuronalen MÜ-Systemen kamen 2017 auf den Markt (Globalese, Lilt). 2018 kamen weitere Systeme wie Google AutoML (Fei-Fei Li 2018), IBM Language Translator, Microsoft Custom Translator (Microsoft 2018) und SDL ETS 8.0 hinzu (vgl. Savenkov 2018). Dabei kommen in erster Linie datenbezogene Verfahren zum Einsatz, die es dem Nutzer ermöglichen, domänen- oder kundenspezifische Daten hochzuladen, die dann für die Erstellung oder Anpassung eines MÜ-Modells verwendet werden.

Dieser Beitrag widmet sich den Möglichkeiten, die Ausgabe kommerzieller maschineller Übersetzungssysteme in Bezug auf Terminologie zu optimieren. Dazu wird in Kapitel 11.2 zunächst auf die Bedeutung der Terminologie beim Übersetzen eingegangen. Im Anschluss werden in Kapitel 11.3 anhand von Beispielen die verschiedenen typischen Terminologiefehler bei der maschinellen Übersetzung sowie deren Kritikalität beschrieben. Den Hauptteil dieses Artikels stellt Kapitel 11.4 dar, das sich mit den verschiedenen Möglichkeiten befasst, maschinelle Übersetzungsmodelle derart anzupassen, dass diese die korrekte bevorzugte Terminologie verwenden. Kapitel 11.5 fasst die wichtigsten Erkenntnisse zusammen und schließt diesen Beitrag ab.

11.2 Terminologie beim Übersetzen

Bei der Übersetzung von Texten, insbesondere Fachtexten, spielt Terminologie eine zentrale Rolle, da sie einen Großteil der Gesamtbedeutung des Textes trägt. Somit hat Terminologie einen direkten Einfluss auf das Textverständnis, aber auch auf die Rechtssicherheit oder Produkthaftung. Zur Terminologie zählen neben Ein- und Mehrwortbenennungen auch fachsprachliche Wendungen (vgl. Arntz 1999). Die Anzahl fachsprachlicher Benennungen und die Terminologiedichte in Texten variieren dabei in Abhängigkeit vom Fachlichkeitsgrad.

Die Verwendung der korrekten, vom Auftraggeber bevorzugten oder freigegebenen zielsprachlichen Benennungen ist eine der zentralen Qualitätsanforderungen für Übersetzungen.

11.2.1 Textverstehen: Terminologie im Ausgangstext

In der Praxis stehen Übersetzer beim Thema Terminologie häufig vor Problemen, da die Bedeutung von Benennungen im Ausgangstext nicht eindeutig ist und Benennungen im Ausgangstext nicht mit einheitlicher Bedeutung bzw. nicht konsistent verwendet werden. Die Klärung von Begriffen in der Ausgangssprache als Voraussetzung für das Verständnis des Ausgangstextes ist daher oft der erste Arbeitsschritt bei der Fachübersetzung.

Dazu greifen Übersetzer auf verschiedene Arten von Wissen zurück, wie Sprachwissen, Textwissen, Fachwissen oder Wissen über den Auftraggeber. Die bei der Terminologiarbeit im Rahmen von Übersetzungen verwendeten Wissensquellen sind neben den Ausgangstexten die Fachliteratur des entsprechenden Gebiets, Fachwörterbücher, Referenztexte des Auftraggebers, Glossare, Terminologie- und Wissensdatenbanken, Terminologieleitfäden sowie das Internet. In Zweifelsfällen ist es erforderlich, einen Fachexperten auf Auftraggeberseite einzubinden. Alle diese Informationen ergeben den Kontext des Ausgangstextes und den Kontext der Übersetzung. Bei der Klärung der Bedeutung von Begriffen, die im Ausgangstext referenziert sind, bedient sich der Übersetzer der Methode der Merkmalanalyse (vgl. Arntz et al. 2014, S. 56 ff). Dabei werden die diskriminierenden Merkmale erarbeitet, die einen Begriff von anderen Begriffen abgrenzen.

11.2.2 Terminologieübersetzung: Bestimmung zielsprachlicher Äquivalente

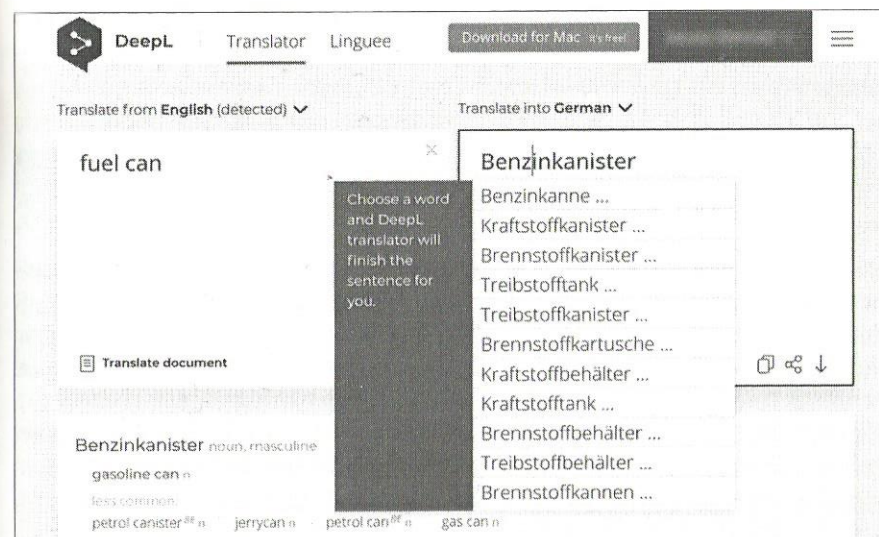
Hat der Übersetzer die Bedeutung der über die Benennungen im Ausgangstext aktivierten Begriffe verstanden, kann er mit der Übertragung der Bedeutung in die Zielsprache beginnen. Bei der Suche nach zielsprachlichen Äquivalenten verwendet der Übersetzer zusätzlich zu den oben genannten Informationsquellen auch Translation Memorys und Paralleltex-te.

Mögliche Übersetzungen für ausgangssprachliche Benennungen werden identifiziert und dann miteinander verglichen, bis eine geeignete Übersetzung gefunden wurde. Auch hier kommt die Methode der Merkmalanalyse zum Einsatz. Zusätzlich empfiehlt sich die Überprüfung der Übersetzungskandidaten anhand einer Liste definierter Benennungsbewertungskriterien. Dazu zählen Eindeutigkeit, Gesetzes- und Normenkonformität, Transparenz, Einheitlichkeit, Gebräuchlichkeit, Zielgruppengerechtigkeit, Kürze, Sprechbarkeit, sprachlich-logische Korrektheit, Konnotationsfreiheit, Ableitbarkeit und Internationalität (vgl. Deutscher Terminologie Tag e.V. 2014, M3-16-24).

Da in der Zielsprache nicht immer eine exakte Entsprechung für eine fachsprachliche oder kundenspezifische Benennung existiert (vgl. Arntz et al. 2014, S. 158), muss der Übersetzer den betreffenden Begriff oder Gegenstand entweder umschreiben, sich der Methode der Entlehnung oder Lehnübersetzung bedienen oder eine neue zielsprachliche Benennung prägen (vgl. Arntz et al. 2014, S. 148–149).

11.2.3 Terminologie in generischen maschinellen Übersetzungsmodellen

Generische maschinelle Übersetzungsmodelle werden mit Millionen Satzpaaren trainiert. Diese beinhalten sowohl allgemeinsprachliche als auch fachsprachliche Inhalte aus verschiedenen Bereichen. Dies führt dazu, dass in einem Übersetzungsmodell für eine ausgangssprachliche fachsprachliche Benennung abhängig vom Kontext mehrere Übersetzungen existieren. Generische Übersetzungsmodelle kennen also keine globalen Vorzugsbenennungen.



DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

Welche der Benennungen am Ende für die Übersetzung gewählt wird, hängt vom Kontext ab. Für die meisten MÜ-Systeme ist dies der zu übersetzende Satz bzw. das zu übersetzende Segment. Gerade bei der professionellen Übersetzung mit Übersetzungstools werden meist nur diejenigen Segmente an das MÜ-System geschickt, für die kein ausreichend hoher Treffer im Translation Memory gefunden wurde. Das bedeutet, dass das MÜ-System keinen Zugriff auf den umgebenden Kontext, den Satz davor oder danach, geschweige denn den gesamten Text hat.

Die Entscheidung für die zielsprachliche Benennung hängt also in diesen Fällen nur vom unmittelbaren textuellen Kontext ab. Dies führt in der Praxis zu inkonsistenten Übersetzungen, wie das folgende Beispiel zeigt.

Englisch	Italienisch
I bought a red <i>fuel can</i> .	Ho comprato una <i>tanica di benzina</i> rossa.
I bought a green <i>fuel can</i> in the online shop.	Ho comprato una <i>tanica di carburante</i> verde nel negozio online.
I bought a golden <i>fuel can</i> .	Ho comprato un <i>bidone</i> d'oro.

Google Translate: <https://translate.google.com>, letzter Zugriff: 15.1.2020

Ein weiteres Beispiel beweist, dass maschinelle Übersetzungssysteme bei der Wahl der zielsprachlichen Terminologie nicht über textexternes Kontextwissen verfügen.

Deutsch	Italienisch
Ich habe bei OBI im Angebot eine Bank für 45,00 EUR gekauft.	Ho comprato una banca da OBI per 45,00 EUR.

Google Translate, <https://translate.google.com/>, letzter Zugriff: 9.1.2020

Die Bedeutung von Bank (als Sitzgelegenheit) wurde im Beispiel nicht korrekt ins Italienische übertragen. Und dies trotz der Hinweise *OBI* (Name eines Baumarktes), *im Angebot* (Finanzinstitute sind eher selten im Angebot erhältlich) und *45,00 EUR* (ein Kaufpreis, der wahrscheinlicher für eine Sitzbank ist). Auch wenn die Terminologiefehler je nach Anwendungsfall mehr oder weniger kritisch sein können, stellt dieses Phänomen bei der professionellen Fachübersetzung ein nicht kalkulierbares Risiko dar (vgl. Ottmann / Canfora in diesem Band). Im Allgemeinen führt es zu einem deutlich höheren Post-Editing-Aufwand bei der Übersetzung von Fachtexten.

Terminologiefehler in generischen maschinellen Übersetzungen reichen von weniger kritischen Fehlern wie Schreibvarianten, die das Verständnis nicht beeinträchtigen, bis hin zu kritischen Bedeutungsverschiebungen. Das folgende Kapitel gibt einen Überblick über die verschiedenen typischen Fehler in maschinellen Übersetzungen.

11.3 Terminologiefehler in maschinellen Übersetzungen

Terminologiefehler in maschinellen Übersetzungen reichen von weniger kritischen Schreib- oder Flexionsfehlern bis hin zu kritischen Bedeutungs- und Konsistenzfehlern. Fehlerwahrscheinlichkeit und -häufigkeit werden durch die folgenden Faktoren begünstigt.

Natürliche Anzahl von Synonymen und/oder Schreibvarianten: Je mehr natürliche Synonyme ein Begriff in einer Sprache hat oder je mehr Schreibweisen zu einer Benennung existieren (z. B. Bindestrich-Varianten), desto unwahrscheinlicher ist es, dass das Datenmaterial, das für das generische Übersetzungsmodell verwendet wurde, konsistent ist. Bei der maschinellen Übersetzung sind folglich mehrere Übersetzungsvarianten zu erwarten.

Entfernung der Fachsprache zur Allgemeinsprache: Die Nähe der Fachsprache zur Allgemeinsprache steigert die Wahrscheinlichkeit, dass entspre-

chende Benennungen im Trainingsmaterial des generischen Übersetzungsmodells enthalten waren. Ihre korrekte maschinelle Übersetzung ist in diesen Fällen somit wahrscheinlicher. So sind beispielsweise Begriffe aus den Bereichen IT oder Mechanik wie „Server“ oder „Achse“ in höherer Anzahl im Trainingsmaterial eines generischen Übersetzungsmodells enthalten als beispielsweise sehr spezifische Terminologie aus dem Bahnbetrieb wie z. B. „Gleisfreimeldeanlage“.

Komplexität der Benennungsstruktur: Mit zunehmender Komplexität einer Benennung (z. B. Komposita oder Mehrwort-Benennungen) steigt logischerweise automatisch auch die Fehleranfälligkeit bei der maschinellen Übersetzung. Mathematisch betrachtet: Bei einer angenommenen durchschnittlichen Anzahl von fünf Varianten pro Benennungsbestandteil, wären bei einer Einwortbenennung fünf Varianten möglich. Bei einem Kompositum aus zwei Bestandteilen stiege die potenzielle Anzahl der Varianten bereits auf 25.

Sprachliche Experimente bestätigen diese Überlegungen: Bei der Übersetzung von drei Sätzen mit der fachspezifischen Benennung „Gleisfreimeldeanlage“ ins Französische durch jeweils vier MÜ-Engines⁹⁶ werden insgesamt zehn französische Benennungsvarianten ausgegeben: *système d'alarme sans voie, système de dégagement de la voie, système de départage de la voie, système de départage des voies, système de détection de la vacance de la voie, système de détection de l'occupation des voies, système de détection des vacances des voies, système de détection d'occupation des voies, système de non-déclaration de la piste, système de validation des voies*. Derselbe Test mit Sätzen, die die Einwortbenennung „Weiche“ enthalten, ergibt hingegen nur vier Varianten: *aiguillage, blagues, commutateur, interrupteur*.

11.3.1 Bedeutungsfehler

Zu Bedeutungsfehlern zählen Bedeutungsverschiebungen oder Bedeutungsänderungen. Diese Art von Fehler tritt häufig in folgenden Fällen auf.

96 Abgefragte Sätze: Die Software für alle Einrichtungen der Gleisfreimeldeanlage musste aktualisiert werden. / Das einwandfreie Funktionieren der Gleisfreimeldeanlage ist von wesentlicher Bedeutung. / Die Reparatur der Gleisfreimeldeanlage hat viel Zeit in Anspruch genommen. Abgefragte Engines: Google Translate: <https://translate.google.com/>; DeepL: [deepl.com](https://www.deepl.com/), IBM Watson: <https://language-translator-demo.ng.bluemix.net/>, Microsoft Translator: <https://www.bing.com/translator/>, letzter Zugriff: 09.01.2020

Homographen: Homographen repräsentieren mindestens zwei Begriffe. Daher ist von Natur aus das Risiko einer semantisch abweichenden Fehlübersetzung höher als bei Benennungen mit nur einer Bedeutung:

Englisch	Deutsch
fuel can	Kraftstoff kann
fuel cans	Kraftstoffdosen
a fuel can	eine Kraftstoffdose

Microsoft Translator, <https://www.bing.com/translator>, letzter Zugriff: 9.1.2020

Das Risiko eines Bedeutungsfehlers steigt, je weniger **Kontext** das zu übersetzende Segment liefert. Teilweise genügt sehr wenig Kontext, z. B. ein unbestimmter Artikel, um, wie im obigen Beispiel, zwischen Verb und Nomen zu unterscheiden und eine korrekte Übersetzung zu generieren.

Wie Kontext den Output der maschinellen Übersetzung beeinflusst, lässt sich an folgendem Beispiel erkennen: Die Übersetzung von „concrete steps“ ins Deutsche ergibt mit DeepL zunächst „konkrete Schritte“. Erweitert man den Satz auf „concrete steps are painted“ wird „steps“ als „Betonstufen“ übersetzt:

Englisch	Deutsch
concrete steps	konkrete Schritte
concrete steps are painted	Betonstufen werden gestrichen

DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

Akronyme werden in der maschinellen Übersetzung in der Regel entweder unverändert in den Zieltext übernommen oder durch ein mögliches zielsprachliches Äquivalent ersetzt. Das kann z. B. die in der Zielsprache gebräuchliche Abkürzung oder die Vollform sein. Das Risiko einer Fehlübersetzung ist in beiden Fällen sehr hoch, da ein Akronym nur selten sprachübergreifend identisch bleibt und das MÜ-System aufgrund der hohen und steigenden Variantenzahl bzw. der Mehrdeutigkeiten von Akronymen nur selten in der Lage ist, diese sicher und richtig aufzulösen.

Französisch	Deutsch
L'ADC demande la fermeture de la voie contiguë pour aller à la visite de son train.	Der ADC fordert die Schließung der angrenzenden Gleise an, um seinen Zug zu besuchen.

DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

Anstelle der Verwendung der deutschen Entsprechung *Triebfahrzeugführer* oder des deutschen Akronyms *Tf* wird im obigen Beispiel das Akronym aus dem Ausgangstext unverändert übernommen. Selbst Kennern der deutschen Eisenbahnsprache erschließt sich die Bedeutung des Satzes nicht, sofern sie nicht über französische Sprachkenntnisse verfügen. Von daher handelt es

sich in diesem Fall um einen kritischen Fehler. Dies ist auch im nächsten Beispiel der Fall:

Französisch	Deutsch
L'ADC demande la fermeture de la voie contiguë pour aller à la visite de son train.	Die CDA beantragt die Schließung des angrenzenden Gleises, um den Besuch ihres Zuges zu ermöglichen.

Falsche Auflösung des Akronyms ADC zu CDA. DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

11.3.2 Auslassungen

In maschinellen Übersetzungen werden manchmal Teile von Benennungen oder ganze Benennungen ausgelassen, was zu einer semantischen Verschiebung oder zu Bedeutungsverlust führen kann.

Deutsch	Französisch
Die Bediener der Stelltische sind nicht an ihrem Platz.	Les opérateurs des tables ne sont pas à leur place.

DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

Die Reduzierung des Wortes „Stelltisch“ auf einen bloßen „Tisch“ im Beispiel erweitert den Interpretationsspielraum stark. Bei dem Bediener kann es sich nun theoretisch ebenso um einen Kellner handeln.

Französisch	Deutsch
Ici, ADC 4712 : le train circulant sur la voie contiguë a un frein bloqué et il y a un jet d'étincelles.	Gestern ADC 4712: der auf dem Nachbargleis fahrende Zug hat eine blockierte Bremse und es gibt einen Funken.

Google Translate, <https://translate.google.com/>, letzter Zugriff: 9.1.2020

Auch bei der Reduktion von „Funkenstrahl“ auf einen „Funken“ im zweiten Beispiel entscheidet der Kontext über die Kritikalität des Fehlers – einer Scheune im Hochsommer genügt ein Funken für eine kritische Einschätzung, wohingegen im stark mechanisch geprägten Eisenbahnbetrieb ein einzelner Funke an der ein oder anderen Stelle keine Gefahr darstellt.

Schwerer wiegt jedoch bei maschinellen Übersetzungen die Auslassung ganzer Wörter. Ohne den Ausgangssatz zu kennen, ist in diesen Fällen das Verständnis der Übersetzung oft unmöglich oder zumindest deutlich erschwert.

Englisch	Deutsch
A 5 liter fuel can can be purchased online.	Ein 5-Liter-Kraftstoff kann online erworben werden.

Auslassung der Information can in fuel can. Microsoft Translator, <https://www.bing.com/translator>, letzter Zugriff: 9.1.2020

Diese Fehler sind sprachpaarübergreifend zu beobachten. Bei der maschinellen Übersetzung von englisch *fuel can* ins Italienische fehlt in der Übersetzung von DeepL die Bedeutung von *fuel*. Es wurde hier die Benennung *lattina* verwendet, die eher im Sinne von *Dose* meist im Kontext von Lebensmitteln verwendet wird. Die bevorzugten Benennungen im Kontext Kunststoffprodukte wären hier *serbatoio* oder *tanica*.

Englisch	Italienisch
fuel cans	barattoli di carburante
5 liter fuel cans	Lattine da 5 litri

DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

11.3.3 Flexionsfehler

Im Allgemeinen sind Flexionsfehler in neuronalen maschinellen Übersetzungen deutlich seltener als bei statistischen maschinellen Übersetzungen. Bei kurzen Segmenten mit wenig Kontext sind sie jedoch öfter zu beobachten. Im folgenden Beispiel wurde das Adjektiv *grün* nicht flektiert. Verändert man den Kontext im Ausgangssatz und ersetzt *green* durch *blue*, wird das Adjektiv jedoch richtig flektiert:

Englisch	Deutsch
green fuel can	grüne Benzinkanister
blue fuel can	blauer Benzinkanister

DeepL Translator: <https://www.deepl.com/translator>, letzter Zugriff: 9.1.2020

11.3.4 Inkonsistenzen

Wie zuvor beschrieben, basieren generische Übersetzungsmodelle auf zahlreichen Textdaten aus unterschiedlichen Quellen. Die Texte sind aus verschiedenen Textsorten, verwenden unterschiedliche Stile und Terminologie und sind meist von variabler Qualität. Entsprechend inkonsistent sind auch die maschinellen Übersetzungen. Die Liste der zur Auswahl stehenden Benennungen, die MÜ-Systeme als alternative Übersetzungen anbieten, verdeutlicht diesen Variantenreichtum (vgl. Beispiel Englisch *fuel can*, Abschnitt Terminologie in generischen maschinellen Übersetzungsmodellen). Bei Google Translate werden die Alternativen zusätzlich mit Hinweisen zur Frequenz sowie Definitionen angereichert. Nichtsdestoweniger wird in bestimmten Kontexten dennoch die weniger gebräuchliche und/oder umgangssprachliche Benennung verwendet, wie folgendes Beispiel veranschaulicht.

Obwohl die Frequenz von *Kraftstoff* und *Brennstoff* als hoch angegeben wird, wird in der Übersetzung *Sprit* verwendet.

Englisch	Deutsch
A red 5 liter fuel can costs 25,99 EUR	Ein roter 5 Liter Sprit kostet 25,99 EUR

Google Translate, <https://translate.google.com/>, letzter Zugriff: 9.1.2020

Zu den Konsistenzfehlern zählen auch inkonsistente Schreibweisen von Benennungen, z. B. mit oder ohne Bindestrich, im Englischen die wechselnde Verwendung von s (en-GB) und z (en-US) oder im deutschen ß (de-DE) und ss (de-CH).

Für die Praxis bedeutet dies, dass mit generischen maschinellen Übersetzungssystemen Benennungen innerhalb eines Auftrages nicht immer zwingend einheitlich übersetzt werden. Bei Fachübersetzungen muss der Post-Editor folglich besonders auf die Terminologie achten.

11.3.5 Kritikalität

Durch eine Gewichtung der Fehlerkategorien (z. B. „minor defects“, „major defects“, „critical defects“) (vgl. DIN 40080) lässt sich eine Kritikalität der Fehler erzeugen, die den Engine-Output auch langfristig vergleichbar macht und eine Basis für Kennzahlen und/oder eine Anforderungsdefinition darstellt. Dabei sollte die Gewichtung stets projektspezifisch festgelegt werden, da die Anforderungen variieren können. Sinnverschiebungen oder -fehler werden in der Regel immer höher gewichtet. Die Verwendung verschiedener Schreibweisen kann aber durchaus in gewissen Kontexten zulässig sein.

Die aufgeführten Beispiele zeigen die hohe Fehleranfälligkeit der generischen maschinellen Übersetzung im Bereich der Terminologie. Um die Fehler und damit auch den eventuellen Nachbereitungsaufwand zu reduzieren, können Übersetzungsmodelle auf einen bestimmten stilistischen, idiomatischen und terminologischen Sprachgebrauch trainiert werden.

11.4 Training und Customization von MÜ-Engines

Zur projekt-, kunden- und/oder fachgebietspezifischen Anpassung von Übersetzungsmodellen bieten sich folgende Möglichkeiten:

Training bedeutet das Erstellen eines neuen Übersetzungsmodells „from scratch“. Diese Option setzt allerdings sehr umfangreiches Trainingsmaterial

voraus (mindestens eine Mio. bilinguale Segmentpaare), das in den meisten Fällen bzw. für die meisten Fachgebiete und Sprachkombinationen nicht in geeigneter Menge und Qualität vorliegt.

Für das bilinguale Training werden Übersetzungsdaten im Austauschformat TMX (Translation Memory Exchange Format) benötigt. Weitere unterstützte Dateiformate sind bilinguale Arbeitsdateien aus TM-Systemen wie MQXLIFF, MQXLZ, MXLIFF, SDLXLIFF, TXLF, XLIFF, XLF oder XLZ. Manche Anbieter wie z. B. Microsoft unterstützen auch die Verwendung von Paralleltexten, die automatisch aligniert werden. Unterstützte Formate hierfür sind TXT, HTML, PDF und DOCX.

Retraining and Customization (auch: Domain Adaption) bedeutet die Anpassung bestehender, generischer Übersetzungsmodelle durch kunden- oder projektspezifische Übersetzungsdaten. Für diese Option genügen bereits kleinere Datenmengen – je nach Anbieter zwischen 5.000 und 15.000 Segmenten. Nach oben sind die Grenzen offen. Allerdings nimmt, nach übereinstimmender Aussage verschiedener Anbieter (Globalese, Microsoft), mit wachsender Datenmenge der proportionale Qualitätszuwachs ab bzw. kann sogar ab einem gewissen Punkt stagnieren.

An dieser Stelle sei darauf hingewiesen, dass *Training* und *Customization* im aktuellen Sprachgebrauch nicht immer unterschieden und oft synonym verwendet werden.

Manche Anbieter ermöglichen beim Training und bei der Customization die **Priorisierung** ausgewählter Übersetzungen (Globalese: *Master vs. Auxiliary Corpora*, Microsoft: *Tuning data*). Wenn beispielsweise 2.000 von 15.000 Segmenten in Stil, Idiomatik und/oder Terminologie exakt dem Projektziel entsprechen, können diese mithilfe der Tuning-Daten beim Training höher gewichtet werden.

Durch **Einbinden von Terminologie** (auch „glossary“ oder „dictionary“) lässt sich die gewünschte Verwendung festgelegter Worte oder Phrasen erzwingen. Dazu erlauben die meisten MÜ-Systeme das Hochladen von Terminologiedaten in den Formaten XLSX, CSV und TSV. Zusätzlich unterstützen einige Systeme wie Globalese auch die Verwendung von TBX-Dateien. Allerdings funktioniert die Einbindung von Terminologie bei den meisten Anbietern durch reines Suchen und Ersetzen starrer Zeichenketten, was die Verwendung ausschließlich auf nicht flektierende Ausdrücke (z. B. Slogans oder Unternehmensnamen wie *Deutsche Bahn AG*) einschränkt. Ansonsten ist eine potenzielle Häufung grammatikalischer Fehler zu erwarten. Microsoft z. B. rät

offen von der Verwendung von Terminologie zur Customization ab und empfiehlt in der Dokumentation des Custom Translators ausschließlich die Terminologiebeeinflussung auf Basis von Korpora.

Um die Qualität eines angepassten maschinellen Übersetzungsmodells zu bewerten, kann ein **Testdatensatz** verwendet werden. Der Testdatensatz ist eine Art „Gold standard“. Er umfasst einen einmalig festgelegten und daraufhin nicht zu verändernden Datensatz projektspezifischer bilingualler Segmentpaare, mit denen der Output der maschinellen Übersetzung kontinuierlich abgeglichen wird. Der Umfang des Testdatensatzes variiert je nach Anbieter. In der Regel genügen ca. 2.000 bilinguale Segmentpaare für eine Evaluierung. Anhand des Vergleichs maschineller Übersetzungen mit den Referenzübersetzungen im Testdatensatz wird von den meisten MÜ-Systemen automatisch der BLEU-Score ermittelt. Dieser gibt an, wie nah die maschinelle Übersetzung den Referenzübersetzungen ist. Er hat zwar wenig konkrete Aussagekraft in Bezug auf die Übersetzungsqualität (vgl. Dougal, 2018, S. 34), ist aber ein guter erster Indikator dafür, ob das Training oder Retraining des Übersetzungsmodells den Output den erwarteten Ergebnissen nähergebracht hat. Einige Anbieter generieren automatisch einen Testdatensatz aus zufällig aus dem Trainingsmaterial gewählten Segmentpaaren, wenn der Benutzer keinen Testdatensatz angibt. In der Regel empfiehlt sich allerdings eine gezielte, projektorientierte Festlegung der Testdaten.

In der Forschung zur maschinellen Übersetzung existieren weitere Ansätze zur Integration von Terminologie- oder Ontologiebeständen. Ein solcher Ansatz ist z. B. das Verfahren der *Term-Injection* (vgl. Dougal, 2018, S. 21–27). Dabei wird durch Eingreifen in das neuronale Netz Terminologie importiert und die Flexion der betreffenden Benennungen sowie aller abhängigen Wörter morphosyntaktisch angepasst. Da diese Ansätze allerdings noch nicht auf dem Markt verfügbar sind, konzentrieren wir uns in diesem Kapitel auf die derzeit gängigste Methode: die Anpassung generischer Übersetzungsmodelle durch *Retraining* bzw. *Customization*.

11.4.1 Training und Customization als Projekt

Die Herausforderung der Planung eines Customization-Projektes besteht in der korrekten Einschätzung der zeitlichen und monetären Aufwände. Zwar ist das reine Retraining relativ gut planbar, jedoch macht es nur einen kleinen Teil des Gesamtprojektes aus. Der weitaus aufwändigere Teil besteht in der Beschaffung und Aufbereitung der Trainingsdaten sowie der Evaluierung des MÜ-Outputs.

Kosten und Zeit für Training und Retraining stehen in der Regel in direktem Verhältnis zur Größe der Trainingsdaten und Anzahl der Sprachpaare. Die meisten Anbieter berechnen entweder einen Preis pro 1.000.000 Zeichen oder stellen die erforderliche Rechenzeit in Rechnung. Je nach Anbieter können noch Kosten für das Hosting der angepassten Übersetzungsmodelle sowie für den Abruf von Übersetzungen auf Basis der Zeichen im Ausgangstext hinzukommen. So wie die Preise je Anbieter variieren (von ca. 5 EUR bis 50 EUR für 1.000.000 übersetzte Zeichen), variieren auch die Trainingszeiten: Während das Retraining mit 20.000 Segmenten bei einigen Anbietern in weniger als einer Stunde abgeschlossen ist, dauert es bei anderen mehrere Stunden (vgl. auch Savenkov 2018).

Der Aufwand für das Sammeln und Aufbereiten der Trainingsdaten sowie die Evaluierung des MÜ-Outputs steht in Abhängigkeit der Projektziele. Oft müssen diese Tätigkeiten mehrmals wiederholt werden. Je höher die Ansprüche an die finale Qualität sind, desto größer ist die Notwendigkeit, mehrere Trainingsdurchläufe zu machen, bis das Qualitätsziel erreicht ist.

Zur Planung kann ein Customization-Projekt in drei Phasen unterteilt werden:

11.4.1.1 Phase 1: Planung und Vorbereitung

In dieser Phase sind unter anderem folgende Aspekte zu berücksichtigen:

■ Definition der Projektziele und Validierungskriterien

- qualitative Anforderungen an das MÜ-Output festlegen
- Fehlerkategorien und Fehlergewichtungen für die Evaluierung definieren

■ Auswahl des Trainingsmaterials

- Sachgebiet, Stil, Idiomatik und Terminologie bestimmen die Auswahl des richtigen Trainingsmaterials
- Materialsammlung und -prüfung: Wie viele Daten stehen zur Verfügung? Hinweis: Durch die spätere automatische Datenaufbereitung innerhalb der MÜ-Systeme vor dem Training kann sich die Datenmenge noch reduzieren. In manchen Fällen führt dies dazu, dass die Mindestdatenmenge dann nicht mehr erreicht wird.
- Datenqualität: Entsprechen die Daten den Qualitätskriterien in Bezug auf Sprachvariante, Textsorte, Inhalt, Orthografie, Konsistenz, Format etc.?

■ Evaluierung der MÜ-Anbieter

- verfügbare Sprachkombination(en) prüfen
- Sprachqualität testen
- direkte vs. indirekte Sprachpaare abfragen
- Preise vergleichen
- Datensicherheitskriterien prüfen
- ggf. weitere projekt- oder kundenspezifische Kriterien

11.4.1.2 Phase 2: Datenaufbereitung und Datengenerierung

Datenaufbereitung

Bei der Datenaufbereitung werden die Trainingsdaten für die Verwendung in einem konkreten maschinellen Übersetzungssystem vorbereitet. Neben der Konvertierung der Daten in ein von den Anbietern unterstütztes Dateiformat werden Daten oft auf verschiedene Qualitätskriterien überprüft und gegebenenfalls bereinigt. Dabei werden oft folgende Aspekte berücksichtigt:

- **Semantische Korrektheit:** Um Sinnverschiebungen im MÜ-Output zu vermeiden, sollten sich Ausgangs- und Zielsegmente unbedingt semantisch entsprechen. Eine häufige Fehlerquelle sind Segmente, deren Ausgangssatz mit mehreren Zielsätzen übersetzt wurde oder Leersegmente.
- **Konsistenz:** Zumindest die zielsprachlichen Segmente sollten auf konsistente Terminologieverwendung überprüft werden. Inkonsistenzen im Ausgangstext hingegen sind für die Engine teilweise sogar förderlich, da so Synonyme erkannt werden, die in der Zielsprache durch eine Benennung konsistent repräsentiert werden.

Hinweis: Sollten Inkonsistenzen im Ausgangstext belassen werden, ist das Trainingsmaterial nicht für beide Sprachrichtungen verwendbar! Für eine Anpassung in beide Sprachrichtungen muss das Trainingskorpus entweder durchgehend konsistent sein oder es muss pro Sprachrichtung ein eigenes Korpus aufgebaut werden.

- **Segmentierung:** Unvollständige grammatikalische Einheiten (z. B. Aufzählungen) beeinflussen die Verarbeitung wohlgeformter Sätze durch die Maschine negativ. Sofern es den Projektanforderungen nicht widerspricht, sollten ausschließlich ganze Sätze im Korpus enthalten sein.

■ **Formatierung:** Die Löschung von Formatierungs-Tags, z. B. mittels Skripten, wird von den meisten Anbietern nicht verlangt; allerdings lassen sich Segmente ohne Tags aufgrund ihrer Übersichtlichkeit leichter auf die oben genannten Aspekte untersuchen.

■ **Anonymisierung und Pseudonymisierung personenbezogener Daten:** Eigennamen, Kundennamen, Geburtstage, Adressen, Telefonnummern, E-Mail-Adressen etc. im Text sollten anonymisiert oder pseudonymisiert werden, damit diese nicht Bestandteil des Übersetzungsmodells werden.

Datengenerierung

Stehen nicht genügend geeignete Trainingsdaten zur Verfügung, bestehen Möglichkeiten, diese zu generieren. Mithilfe von Platzhaltern z. B. lassen sich bestehende oder in geringer Zahl erstellte bilinguale Texte variantenreich vervielfältigen. Dazu sind folgende Arbeitsschritte erforderlich:

- **Erkennen** bestimmter wiederkehrender Elemente im Text ohne zentrale Bedeutung: z. B. Nachnamen, Städtenamen, etc.
- **Ersetzen** dieser Elemente durch dafür definierte Entitäten: z. B. {surname} für alle Nachnamen und {city} für alle Städtenamen.
- **Werte definieren**, mit denen die Platzhalter (automatisch) befüllt werden: Für die Entität {city} wird beispielsweise eine Werteliste mit 50 verschiedenen Städtenamen hinterlegt.
- **Texte generieren**, indem Entitäten mit verschiedenen Werten befüllt werden (je mehr Entitäten in einem Satz erkannt werden, desto häufiger kann dieser variantenreich reproduziert werden).

Originalsatz:

■ Übermittlungscode 16318, Dein Name ist Maier und ich heisse Schmitz.
 ■ ■ Code d'authorisation 16318, tu t'appelles Maier et moi, je m'appelle Schmitz.

Entitätensatz

■ Übermittlungscode {code}, Dein Name ist {surname} und ich heisse {surname}.
 ■ ■ Code d'authorisation {code}, tu t'appelles {surname} et moi, je m'appelle {surname}.

Automatisch generierte Sätze:

■ Übermittlungscode 16318, Dein Name ist Walter und ich heisse Lehmann.
 ■ ■ Code d'authorisation 16318, tu t'appelles Walter et moi, je m'appelle Lehmann.

■ Übermittlungscode BHG-12344, Dein Name ist Maier und ich heisse Bergmann.
 ■ ■ Code d'authorisation BHG-12344, tu t'appelles Maier et moi, je m'appelle Bergmann.

■ Übermittlungscode F-633, Dein Name ist Schmitz und ich heisse Katz.
 ■ ■ Code d'authorisation F-633, tu t'appelles Schmitz et moi, je m'appelle Katz.

Beispiel aus einem Customization-Projekt der Deutschen Bahn AG

Vor der Textgenerierung empfiehlt sich stets eine Überprüfung aller mit Platzhaltern für Entitäten versehenen Sätze auf korrekte Platzierung der Entitäten sowie deren korrekten Verteilung in Ausgangs- und Zieltext. Ansonsten besteht das Risiko, dass fehlerhafte Trainingsdaten generiert werden.

Das Einsetzen von Entitäten sowie die Textgenerierung lassen sich mit Hilfe von Skripten automatisieren. Allerdings bedeutet die Skripterstellung und -anpassung immer einen erhöhten Initial- und Pflegeaufwand, denn auch bestehende Skripte müssen an die projektspezifischen Texte angepasst werden.

11.4.1.3 Phase 3: Training, Evaluierung & Retraining

Wenn die Trainings-, Test- und Tuningdaten vorbereitet sind, werden diese über die Website oder über API in das maschinelle Übersetzungssystem hochgeladen. Bei den meisten Anbietern werden die Trainingsdaten dann zu einem Korpus zusammengefasst. Für das Training oder Retraining eines Übersetzungsmodells wird das Korpus ausgewählt und das Training gestartet. Sobald der Vorgang beendet ist, wird der Benutzer benachrichtigt. Die Benachrichtigung enthält meist zusätzliche Informationen zum Trainingsverlauf und dem Ergebnis der automatischen Evaluierung auf Basis des Testdatensatzes.

Bei der automatischen Evaluierung der angepassten Übersetzungsmodelle kommt bei allen Anbietern aktuell der BLEU-Score zum Einsatz. Diese Metrik

basiert auf der Annahme, dass eine maschinelle Übersetzung umso besser ist, je ähnlicher sie einer menschlichen Referenzübersetzung ist. Die Ähnlichkeit wird auf Basis der Übereinstimmung von n-Grammen berechnet. Der Wert liegt zwischen 0 und 1 und wird häufig in Prozent angegeben. Der BLEU-Score dient als erster Indikator für den Erfolg eines Trainings. Seine Aussagekraft ist jedoch umstritten, da neuronale maschinelle Übersetzungssysteme alternative Übersetzungen generieren können, die zwar den Referenzübersetzungen nicht ähnlich, aber dennoch richtig sein können. Alternativ können weitere automatische Evaluierungen mit anderen Metriken durchgeführt werden (z. B. TER, WER, HTER, hLEPOR etc.). Zur Prüfung spezifischer Qualitätskriterien wie Terminologieeinhalten kommen in der Praxis auch weitere automatische Verfahren wie QA- und Terminologie-Checks oder spezielle Skripte zum Einsatz.

Durch vorgeschaltete automatische Evaluierungsverfahren lässt sich der Aufwand der anschließenden manuellen Fehlerkategorisierung und -gewichtung stark reduzieren. Die manuelle Analyse bleibt allerdings notwendig, um Schwächen der Engine zu konkretisieren und darauf gezielt Daten für das nächste Retraining zu generieren. Auf dieser Basis lassen sich auch Kennzahlen oder Mindestanforderungen kontrollieren, die eine langfristige Vergleichbarkeit der Engine-Entwicklung garantieren und die über die Anzahl weiterer Retrainings entscheiden.

Auf Basis der Evaluierungsergebnisse, d.h. der Anzahl Fehler je Kategorie und der Kritikalität, kann dann gezielt weiteres Trainingsmaterial zur gezielten Optimierung und Korrektur bestimmter Fehler gesammelt oder generiert werden. Der Prozess des Trainings und der anschließenden Evaluierung beginnt dann erneut.

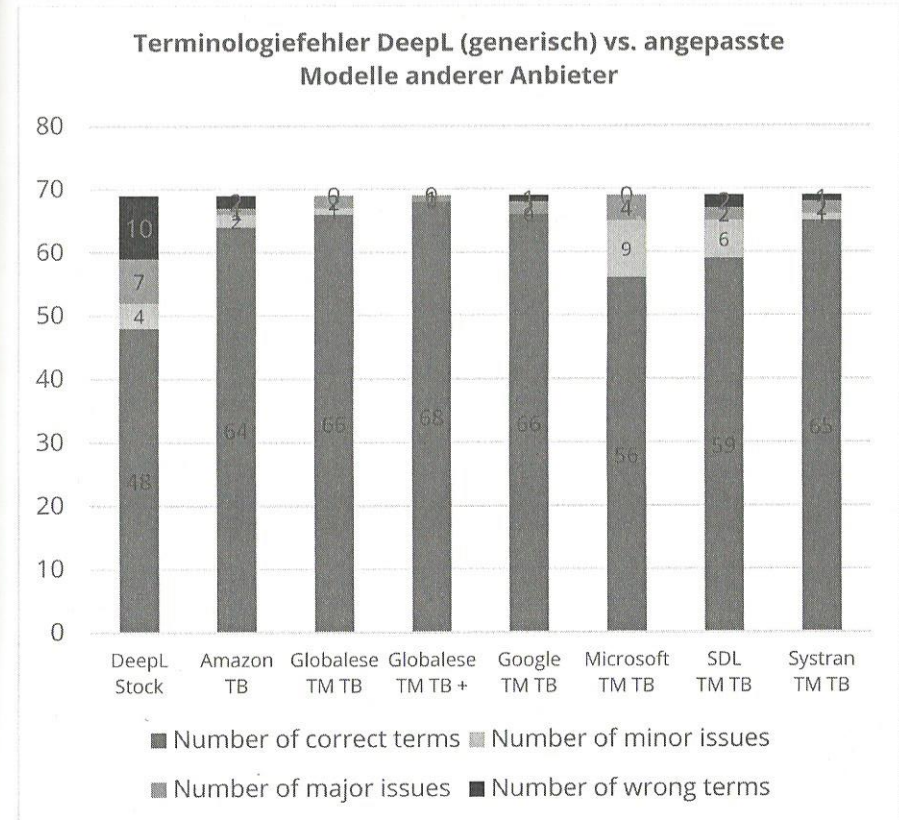
11.4.1.4 Projekt abschließen

Sobald die gewünschte Übersetzungsqualität erreicht ist, kann die angepasste MÜ-Engine eingesetzt werden. Dazu muss diese bei einigen Anbietern zunächst aktiviert bzw. bereitgestellt werden.

Bevor das Trainings- oder Customization-Projekt abgeschlossen wird, empfiehlt es sich, die gewonnenen Erfahrungen in Bezug auf Sprachpaar, Datenmenge, Evaluierungsergebnis, Zeit, Kosten etc. zu dokumentieren, um die gewonnenen Erkenntnisse im Sinn der konstanten Verbesserung für zukünftige Projekte nutzbar zu machen.

11.4.2 Ergebnisse und Optimierungspotenziale

Dass sich das Training oder Retraining von Übersetzungsmodellen lohnt, ist unter MÜ-Experten allgemein bekannt. Wie groß das Optimierungspotenzial ist, hängt von verschiedenen Faktoren ab. Dazu zählen unter anderem das Sprachpaar (direkt vs. indirekt, verwandte vs. nicht verwandte Sprachen), die Datenmenge sowie die Datenqualität. Im Vergleich zu generischen Übersetzungsmodellen, die je nach Fachgebiet nur zwischen 5 % und 60 % der Bezeichnungen richtig übersetzen, lässt sich die Terminologieeinhalten durch Customization auf bis über 90 % optimieren (vgl. Zielinski / Vardaro 2019).



Vergleich Terminologieeinhalten generisches Übersetzungsmodell von DeepL vs. angepasste Übersetzungsmodelle anderer Anbieter Englisch-Deutsch. (vgl. Zielinski / Vardaro 2019)

In einem Projekt der Deutschen Bahn AG z. B. konnte durch ein Retraining mit nur 16.000 Segmenten (davon 453 Segmente mit entsprechendem Terminologiepaar) die korrekte Verwendung von *DE: Gegengleis* und *FR: voie de*

sens contraire von 8 % (4 von 50) auf 86 % (43 von 50) gesteigert werden. Ein weiteres Training mit 71.000 Segmenten (davon 2012 Segmente mit entsprechendem Terminologiepaar) steigerte die korrekte Verwendung auf 90 % (45 von 50).

Das Beispiel macht auch deutlich, dass die Optimierung nicht proportional zur Menge der eingesetzten Trainingsdaten ist. Aufgrund der Komplexität von Sprache und den daraus resultierenden zahlreichen möglichen Ursachen für die weiterhin vorhandene Fehlverwendung (10 %), wird der Aufwand für die weitere Optimierung derart hoch eingeschätzt, dass er der möglicherweise erreichbaren Vervollkommnung nicht Rechnung trägt.

11.5 Zusammenfassung und Ausblick

Das Retraining von MÜ-Modellen hat sich als effektives Mittel erwiesen, um die Übersetzungsleistung generischer Übersetzungsmodelle an fachsprachliche Anforderungen anzupassen. Dabei steigt der Nutzen von Training und Customization in der Regel mit dem Grad der Fachsprachlichkeit der zu übersetzenden Texte.

Eine 100%ige Korrektheit bei der Verwendung der Terminologie in maschinellen Übersetzungen kann allerdings auch trotz Retraining nicht garantiert werden. Selbst bei Verwendung großer, qualitativ hochwertiger Trainingskorpora kommt es bei den Übersetzungen immer noch zu verschiedenen Fehlern in den Bereichen Bedeutung, Wortart, Flexion, Syntax, Schreibweisen etc. Dies liegt daran, dass Sprache ein flexibles System ist und nicht alle Kombinationen bzw. Verwendungszwecke von Benennungen im Trainingsmaterial enthalten sind.

Die Voraussetzung für die Anpassung von MÜ-Engines ist die Verfügbarkeit einer ausreichenden Menge von Trainingsdaten. Hierzu eignen sich in erster Linie Paralleltexthe wie Translation Memorys. Eher weniger eignen sich Terminologiedaten in Form von Glossaren und Terminologiedatenbanken. Dies liegt zum einen daran, dass Terminologiedaten zwar mehrsprachig, aber meist kontextlos sind, zum anderen daran, dass in vielen Systemen die Terminologie mittels Suchen und Ersetzen und ohne sprachlich und kontextuelle Anpassung in die Übersetzung eingefügt wird (vgl. Kapitel 11.4). Terminologiedaten sind jedoch weiterhin für eine automatisierte Terminologieprüfung von Übersetzungen erforderlich.

Die Anpassung von MÜ-Systemen ist mit einem hohen Zeit- und Kostenaufwand verbunden, der nicht unterschätzt werden sollte. Dies gilt insbesondere für die Datenanalyse, -sammlung, -generierung und -aufbereitung sowie für

die anschließende Evaluierung der mit dem angepassten MÜ-Modell angefertigten maschinellen Übersetzungen. Im Vergleich dazu ist das eigentliche Training der MÜ-Modelle deutlich weniger aufwändig, da die Bedienung kommerzieller MÜ-Systeme immer benutzerfreundlicher wird.

Aus den genannten Gründen empfiehlt es sich, vor dem Beginn der Anpassung eines MÜ-Systems den Business Case zu prüfen. Liegt dieser vor, sollte die Anpassung eines MÜ-Modells als Projekt behandelt und entsprechend geplant und budgetiert werden. Die Erfahrungen aus den einzelnen Projekten sollten sauber dokumentiert werden, um den „Lernfortschritt“ der MÜ-Modelle nachvollziehen zu können und Best practices zu etablieren.

Zu guter Letzt sei darauf hingewiesen, dass die Anbieter von MÜ-Systemen von Zeit zu Zeit ihre generischen Übersetzungsmodelle aktualisieren. In diesen Fällen empfiehlt sich ein Retraining der angepassten Modelle, damit die Verbesserungen der neuen generischen Modelle genutzt werden können. Bei einigen Anbietern werden entsprechende Hinweise in der Verwaltungsoberfläche angezeigt. Damit werden Training und Customization maschineller Übersetzungsmodelle zu einer neuen kontinuierlichen Aufgabe im Bereich Übersetzung.

11.6 Bibliografie

- Arntz, R. / Picht, H. / Schmitz, K.-D. (2014): Einführung in die Terminologiearbeit. Hildesheim: Georg Olms Verlag.
- Arntz, R. (1999): Terminologie in der Terminologie. In: Snell-Hornby et. al (Hrsg.): Handbuch Translation, S. 81. Tübingen: Stauffenburg-Verlag.
- Chen, B. / Kuhn, R. / Foster, G. / Cherry, C. / Huang, F. (2016): Bilingual methods for adaptive training data selection for machine translation. In: The Twelfth Conference of The Association for Machine Translation in the Americas, S. 93–106.
- Chenhui Chu / Rui Wang (2018): A Survey of Domain Adaptation for Neural Machine Translation. In: Proceedings of the 27th International Conference on Computational Linguistics, S. 1304–1319. <https://www.aclweb.org/anthology/C18-1111.pdf>, letzter Zugriff: 12.01.2020.
- Deutscher Terminologie Tag e.V. (Hrsg.) (2014): Terminologiearbeit – Best practices 2.0

- Dougal, D. K. (2018): Improving the Quality of Neural Machine Translation Using Terminology Injection, S. 21-27. <https://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=8025&context=etd>, letzter Zugriff: 14.01.2020
- Durrani, N. / Sajjad, H. / Joty, S. / Abdelali, A. / Vogel, S. (2015): Using joint models for domain adaptation in statistical machine translation. In Proceedings of MT Summit XV, S. 117-130.
- Etchegoyhen, T. / Fernández Torné, A. / Azpeitia, A. / Martínez Garcia, E. / Matamala, A. (2018): Evaluating Domain Adaptation for Machine Translation Across Scenarios. LREC. <https://www.aclweb.org/anthology/L18-1002.pdf>, letzter Zugriff: 09.01.2020.
- Fei-Fei Li (2018): Empowering businesses and developers to do more with AI. Google Blog. <https://www.blog.google/products/google-cloud/empowering-businesses-and-developers-to-do-more-with-ai/>, letzter Zugriff: 13.01.2020.
- Freitag, M. / Al-Onaizan, Y. (2016): Fast Domain Adaptation for Neural Machine Translation. <https://arxiv.org/abs/1612.06897>, letzter Zugriff: 09.01.2020.
- ISO 18587 (2017): Translation services – Post-editing of machine translation output – Requirements. <https://www.iso.org/standard/62970.html>, letzter Zugriff: 12.01.2020.
- Koehn, P. (2010): Statistical Machine Translation. Cambridge, Cambridge University Press.
- Microsoft (2018): Customize Microsoft Translator's Neural Machine Translation to translate just the way you want: Custom Translator now in General Availability. Microsoft Translator Blog. <https://www.microsoft.com/en-us/translator/blog/2018/12/04/customize-microsoft-translators-neural-machine-translation-to-translate-just-the-way-you-want-custom-translator-now-in-general-availability/>, letzter Zugriff: 13.01.2020
- Moore, Robert C. / Lewis, W. (2010): Intelligent selection of language model training data. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers, S. 220-224.
- Nitze, J. / Hansen-Schirra, S. / Canfora, C. (2019): Risk management and post-editing competence. In: The Journal of Specialised Translation, S. 239-259.
- Risku, H. (2004): Translationsmanagement. Interkulturelle Fachkommunikation im Informationszeitalter. Tübingen, Gunter Narr Verlag.

- Savenkov, K. (2018): State of the domain-adaptive machine translation. Intento. <https://www.slideshare.net/KonstantinSavenkov/state-of-the-domain-adaptive-machine-translation-by-intento-july-2018>, letzter Zugriff: 13.01.2020.
- Sennrich, R. / Schwenk, H. / Aransa, W. (2013): A multi-domain translation model framework for statistical machine translation. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers, S. 832-840.
- Wang, R. / Finch, A. / Utiyama, M. / Sumita, E. (2014): Neural network based bilingual language model growing for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), S. 189-195.
- Zhou, X. / Cao, H. / Zhao, T. (2015): Domain adaptation for SMT using sentence weight. In: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, S. 153-163.
- Zielinski, D. / Vardaro, J. (2019): Das Terminologieproblem in der maschinellen Übersetzung. Übersetzen und Dolmetschen 4.0 – Neue Wege im digitalen Zeitalter, World Conference Center Bonn, 22.-24. Nov. 2019