

**Evaluation of *Computer Aided Translation* (CAT) Tools in the Translation Process: A
Plea for Complementing CAT-Tools with Corpus Linguistical Tools and Methods**

**Evaluierung von *Computer Aided Translation* (CAT) Tools im Translationsprozess: Ein
Plädoyer für die Ergänzung von CAT-Tools durch korpuslinguistische Werkzeuge und
Methoden**

by

Dora Hinderer

A thesis

presented to the University of Waterloo

and the Universitaet Mannheim

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Intercultural German Studies

Waterloo, Ontario, Canada / Mannheim, Germany, 2024

© Dora Hinderer 2024

Author's declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Ehrenwörtliche Erklärung

Ich versichere, dass ich die Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen in schriftlicher oder elektronischer Form entnommen sind, habe ich als solche unter Angabe der Quelle kenntlich gemacht. Mir ist bekannt, dass im Falle einer falschen Versicherung die Arbeit mit „nicht ausreichend“ bewertet wird. Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form versendet und gespeichert werden kann.

Abstract

The demand for translations of high quality is on a peak while the budget and time resources are low. Consequently, translators need technological support and use Computer-Assisted Translation (CAT) tools. Using the open-source tool OmegaT, this thesis investigates how CAT tools support the translation process. The objective of the study are political texts published by the European Parliament. The analysis further demonstrates that CAT tools have certain constraints. Using parts of the monolingual corpus DeReKo, the thesis exemplifies how these limitations can be overcome by using corpora as an additional resource during the translation process.

This thesis contributes to the improvement of CAT tools as the main technological resource used by translation experts and outlines weaknesses which negatively impact their work. The findings can be used for optimizing all types of technology dealing with language data because they show how natural language must be understood to process it automatically. Furthermore, the approach taken in this work with regard to the critical examination of available resources can be applied in the future to evaluate and optimize new translation tools.

Acknowledgements

I extend my sincere appreciation to my professors Dipl.-Inf. Rainer Perkuhn, Prof. Dr. Andreas Witt and Prof. Dr. Henning Lobin for their advice and insightful feedback, which proved invaluable to the completion of this Master's thesis. Special appreciation is extended to Dipl.-Inf. Rainer Perkuhn whose expertise played a pivotal role in shaping this research.

Recognition is also extended to the university librarians, research assistants, and cohort members whose contributions and ideas impacted and inspired my work.

I must express heartfelt gratitude to my family, notably my partner, parents, and sisters, whose unwavering belief in me has been a constant source of motivation throughout this challenging journey.

Their collective support has been instrumental in the successful completion of this thesis, and I am truly grateful for the encouragement and assistance received from each individual and organization mentioned.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	vi
1. Einleitung.....	1
2. Erläuterung des Untersuchungsrahmens.....	3
3. Teil 1 – <i>Computer Aided Translation (CAT) Tools</i>	6
3.1 <i>Computer Aided Translation versus Machine Translation (MT)</i>	6
3.2 Aufbau und Gebrauch.....	11
3.2.1 Häufige Bestandteile von CAT-Tools.....	11
3.2.2 Gebrauch von CAT-Tools.....	15
3.3 CAT-Tools im Übersetzungsprozess am Beispiel von OmegaT.....	18
3.3.1 Projektvorbereitung.....	20
3.3.2 OmegaT während der Übersetzung.....	23
3.3.3 Übersetzungsabschluss.....	29
4. Teil 2 – Die Nutzung von Korpora im Übersetzungsprozess.....	30
4.1 Begriffsbestimmung.....	30
4.2 Zum Einsatz monolingualer Korpora in der Translationsarbeit.....	33
4.3 Abgrenzung von Korpora zu TMs.....	35
4.4 Das <i>Deutsche Referenzkorpus (DeReKo)</i>	37
4.4.1 Auswahl von Subkorpora in DeReKo.....	39
4.4.2 Suche und Anzeige von Korpusbelegen in COSMAS II.....	42
4.4.3 Kookkurrenzanalyse.....	45
5. Teil 3 – Ergänzung von CAT-Tools durch Korpora.....	48
5.1 Der richtige Textausschnitt für die Übersetzung.....	48
5.2 Einsatzbereich der computergestützten Werkzeuge.....	53
5.3 Automatische Übersetzungsvorschläge.....	60
6. Abschließende Diskussion und Ausblick.....	67
7. Fazit.....	77
Literaturverzeichnis.....	79

Abbildungsverzeichnis

Abbildung 1.....	8
Abbildung 2.....	9
Abbildung 3.....	21
Abbildung 4.....	23
Abbildung 5.....	25
Abbildung 6.....	43
Abbildung 7.....	44
Abbildung 8.....	48
Abbildung 9.....	48
Abbildung 10.....	50
Abbildung 11.....	54
Abbildung 12.....	55
Abbildung 13.....	55
Abbildung 14.....	56
Abbildung 15.....	60
Abbildung 16.....	60
Abbildung 17.....	62
Abbildung 18.....	62
Abbildung 19.....	70
Abbildung 20.....	72

1. Einleitung

Sehr geehrte Damen und Herren, ich interessiere mich sehr für den Übersetzungs-Beruf und wäre sehr dankbar wenn Sie mir einige Fragen beantworten könnten. [...] Ich würde gerne von den Erfahrungen anderer [sic!] hören, die vom Deutschen ins Englischen [sic!] übersetzen. Wie sieht momentan die Arbeitslage für diese Sprach-Kombination aus? [...] Zwecks der Fachprogrammen [sic!] wie Trados, MemoQ u.s.w, ist es sehr schwierig die Benutzung zu erlernen ohne eine Ausbildung als Übersetzer zu absolvieren? Bitte entschuldigen Sie, dass ich so viele Fragen stelle. Ich würde mich wirklich sehr darauf freuen, wenn jemand mir weiterhelfen könnte [...] (Pike 2018).

Hi, Thomas. As your mother tongue is English, I'm replying in English. I've been translating German into English professionally for over 10 years now. [...] If you can get direct clients, they'll usually pay reasonably well, but you're more likely to be getting work from agencies, some of whom will insist that you use a CAT tool and will try to get you to accept really low rates. Good agencies will pay you enough to live on, though. I use Trados and try to make sure that I always have the most up-to-date version (Studio 2017). CAT tools like this are great time-savers once you've used them for a while and built up one or more translation memories (TMs). I'm told MemoQ is good too, but I haven't tried it myself. Translation is a huge industry with people working in all sorts of language pairs. [...] I hope that's useful for starters (Roebuck 2018).

Während sogenannte *Large Language Models* wie *ChatGPT* oder eine Vielzahl von kostenlos nutzbaren maschinellen Übersetzungssystemen in der breiten Gesellschaft für Übersetzungen immer mehr Anwendung finden (vgl. Maulidiyah 2018, S. 1), sind es im Rahmen der professionellen Übersetzungsarbeit überwiegend *Computer Aided Translation* (CAT) Tools, die von Expertinnen und Experten herangezogen werden, um die Translationsarbeit zu unterstützen (vgl. Orozco Agudelo 2022, S. 2). Wie der obige Austausch aus einem Übersetzungsforum zeigt, werden Neulinge der Branche darauf hingewiesen, sich zügig mit CAT-Tools vertraut zu machen, um in der Übersetzungsindustrie mithalten zu können und den Anforderungen der Übersetzungsagenturen gerecht zu werden. Die Übersetzungsarbeit ist, wie der erfahrene Nutzer in seiner Antwort bestätigt, im Zuge der fortlaufenden Globalisierung zu einer Industrie geworden, in der ein stetig wachsender Bedarf an Übersetzungen zusammen mit eng getakteten Deadlines und knappem Budget dazu führt, dass technische Hilfsmittel wie CAT-Tools ein wichtiges Werkzeug der professionellen Übersetzerinnen und Übersetzer darstellen (vgl. Bowker & Barlow 2008, S. 1). Gleichzeitig müssen übersetzte Texte jedoch weiterhin von guter Qualität sein und den Anforderungen der Kundinnen und Kunden genügen. Im Spannungsfeld von Zeitdruck, Kosteneinsparungen und Übersetzungsqualität stellt sich deshalb die Frage, wie CAT-Tools die translatorische Arbeit unterstützen, aber auch welche Grenzen erkennbar werden und

welche anderen technischen Mittel Übersetzerinnen und Übersetzern zur Verfügung stehen.

Diese Arbeit beschäftigt sich mit dem Einsatz von CAT-Tools im Übersetzungsprozess und soll zeigen, welche Art von Unterstützung die Werkzeuge darstellen. Den Untersuchungsrahmen bilden politische Texte, die häufig in mehrere Sprachen übersetzt werden. Am Beispiel des CAT-Tools OmegaT soll neben den Vorteilen außerdem gezeigt werden, welche Einschränkungen die Arbeit mit den computergestützten Werkzeugen mit sich bringt. Auf dieser Grundlage wird anschließend analysiert, wie Korpora herangezogen werden können, um CAT-Tools im Übersetzungsprozess zu ergänzen. Am Beispiel des monolingualen Korpus DeReKo soll gezeigt werden, welche korpuslinguistischen Methoden und Werkzeuge in der Translationsarbeit ergänzend zu CAT-Tools genutzt werden können. Daraus resultierende empirische Erkenntnisse über die Art, wie Sprache analysiert und verstanden werden muss, um qualitativ hochwertige Übersetzungen zu produzieren, können dann zur Weiterentwicklung von CAT-Tools und anderen Automatisierungstools wie maschinellen Übersetzungssystemen beitragen.

Der erste Teil dieser Arbeit widmet sich der Beschreibung und dem Aufbau von CAT-Tools sowie deren Abgrenzung zu maschinellen Übersetzungssystemen. Anschließend wird exemplarisch das CAT-Tool OmegaT vorgestellt und es wird gezeigt, welchen Nutzen OmegaT im Übersetzungsprozess bietet. Der zweite Teil dieser Arbeit beschäftigt sich mit dem Einsatz von Korpora im Übersetzungsprozess. Für die nachfolgende Analyse wird das monolinguale Korpus DeReKo und verschiedene Funktionen, die für die Übersetzungsarbeit genutzt werden können, vorgestellt. Im dritten Teil der Arbeit werden dann die Einschränkungen der Arbeit mit CAT-Tools wie OmegaT aufgezeigt. Die zuvor gewonnenen Erkenntnisse werden zusammengeführt und es soll an konkreten Textbeispielen gezeigt werden, wie OmegaT durch den Einsatz von Korpora und korpuslinguistischen Methoden ergänzt werden kann. In einer abschließenden Diskussion der Ergebnisse sollen die Analyseergebnisse reflektiert werden und es soll erläutert werden, welche Herausforderungen im Übersetzungsprozess auch nach der Ergänzung von CAT-Tools durch Korpora weiter bestehen.

2. Erläuterung des Untersuchungsrahmens

Um zu evaluieren, wie CAT-Tools im Übersetzungsprozess eingesetzt werden, wie sie die Arbeit unterstützen und wo sie an ihre Grenzen stoßen, wurden in der nachfolgenden Analyse politische Texte als Untersuchungsgegenstand herangezogen. Texte dieser Fachdomäne werden häufig und in der Regel in mehrere Sprachen übersetzt und eignen sich deshalb besonders für die Untersuchung. Wie Tatu (2011) schreibt, stellen Texte, die einer Fachdomäne entstammen, in der Translationsarbeit eine besondere Herausforderung dar:

While the translation of a general text between one source language and one target language remains an actual challenge for both human and machine translators, the translation of special domain texts presupposes significant proficiency in handling both the content and the representation of that domain knowledge (S. 167).

Gleichzeitig ist es aber besonders bei politischen Texten, die in der Regel von öffentlichen Institutionen publiziert werden und viele Menschen betreffen und erreichen, von größter Wichtigkeit, dass die Übersetzungen richtig und vollständig sind. Gute Übersetzungsqualität ist bei der Translation politischer Texte demnach essentiell.

Als einer der zu übersetzenden Texte wurde eine legislative EntschlieÙung zur Makrofinanzhilfe für Nordmazedonien gewählt (vgl. Europäisches Parlament 2023b). Der aus der politisch-rechtlichen Domäne entstammende Text wurde am 13. Juni 2023 vom Europäischen Parlament veröffentlicht. Der zweite für die Untersuchung herangezogene Text ist eine Vorrede von Roberta Metsola, in der die Parlamentspräsidentin eine Rede von Olaf Scholz ankündigt, die am neunten Mai 2023 im Europäischen Parlament gehalten wurde (vgl. Metsola 2023b). Beide Texte wurden sowohl im Englischen als auch im Deutschen (vgl. Europäisches Parlament 2023a; Metsola 2023a) auf der Webseite des Europäischen Parlaments veröffentlicht. In beiden Fällen können deshalb die erarbeiteten Untersuchungsergebnisse mit der offiziell publizierten Übersetzung ins Deutsche kontrastiert werden. Auch im Analyseteil dieser Arbeit wurde die Sprachrichtung Englisch-Deutsch fokussiert. In dem Protokoll zur Makrofinanzhilfe für Nordmazedonien wurden die Ergebnisse einer EntschlieÙung schriftlich festgehalten. Beim zweiten Text handelt es sich um eine Verschriftlichung der gesprochenen Rede von Roberta Metsola. Obwohl beide Texte also nicht nur dem politischen, sondern sogar dem europapolitischen Kontext entstammen und

zudem beide Texte ganz aktuell veröffentlicht wurden, weisen sie dennoch einige Unterschiede auf. Wie sich diese Unterschiede unter Umständen auch auf die Übersetzung mit computergestützten Werkzeugen auswirkt, soll im dritten Teil dieser Arbeit diskutiert werden.

Als Datenbasis für das CAT-Tool OmegaT wurde das *DGT-Translation Memory* (DGT-TM) verwendet (vgl. DGT-TM). Dieses von der europäischen Kommission veröffentlichte *Translation Memory* (TM) umfasst eine Sammlung von Texten aus dem *Acquis Communautaire* seit 2007 (Steinberger et al. 2012, S. 454). Darin enthalten sind Texte aus dem politisch-rechtlichen Kontext wie Verträge und Beschlüsse aus dem europäischen Parlament. Die Texte sind in dem speziellen Dateiformat *.tmx* aufbereitet, das sie für den Abruf in CAT-Tools verfügbar macht. Dazu werden ausgangssprachliche Texte zunächst in kleinere Einheiten – sogenannte Segmente – geteilt und dann mit ihrem zielsprachlichen Äquivalent als *Translation Unit* gespeichert (vgl. Steinberger et al. 2012, S. 456). Der Textbestand in DGT-TM umfasst insgesamt 24 Sprachen und über 137 Millionen einzelne *Translation Units* (vgl. Schlüter 2020), aus denen für die nachfolgende Untersuchung ein Teilbestand des Deutschen und Englischen extrahiert wurde.¹ Aus Kapazitätsgründen des CAT-Tools OmegaT konnte nur ein kleiner Teil der verfügbaren Texte des deutsch-englischen Bestands in DGT-TM verwendet werden. Für die nachfolgende Analyse wurden deshalb die zwei neusten Ergänzungen verwendet, die im Jahr 2020 und 2021 veröffentlicht wurden. Die Datengrundlage in OmegaT bildeten dadurch insgesamt 265.248 alignierte *Translation Units* in der Sprachkombination deutsch-englisch.² Wie Orozco Agudelo (2022) feststellt, sind es in der professionellen Übersetzungsarbeit besonders eigene Übersetzungen, die in den TMs gespeichert zur Wiederverwendung in neuen Projekten abgerufen werden (vgl. S. 40). Knapp 30% der Befragten Expertinnen und Experten greifen jedoch auch auf externe Ressourcen zurück (vgl. Orozco Agudelo 2022, S. 40). Diese können ergänzend zu eigenen Beständen genutzt werden, oder als alleinige Ressource dienen. Da in der nachfolgenden Untersuchung kein Altbestand an eigenständig produzierten Texten verwendet werden kann, gelten die nachfolgend gewonnenen Erkenntnisse besonders für Situationen, in denen Überset-

1 Genaue Erläuterungen zur Extraktion eines Teilbestands aus dem *.tmx*-Dateien siehe Steinberger et al. (2012).

2 Diese Information wurde beim Packen der deutsch-englischen *.tmx*-Datei aus dem Gesamtvolumen bereitgestellt.

zende kein eigenes TM besitzen – beispielsweise da es sich um ein Projekt aus einer neuen, bisher nicht übersetzten Fachdomäne handelt.

Um zu untersuchen, wie Korpora die Übersetzungsarbeit unterstützen, wurde ein monolinguales Korpus verwendet. Für die Übersetzungsarbeit werden in der Regel zwar besonders mehrsprachige, insbesondere alignierte Korpora verwendet, die eine Suche nach Übersetzungsäquivalenten und andere kontrastive Analysen erlauben (vgl. Verplaetse & Lambrechts 2019, S. 24), aber besonders für spezielle Fachdomänen sind diese nur begrenzt verfügbar (vgl. Adjali, Morin & Zweigenbaum 2022, S. 3103). Die nachfolgende Analyse soll deshalb zeigen, wie auch monolinguale Sprachdatenbanken für die Translationsarbeit genutzt werden können. Das in dieser Arbeit verwendete *Deutsche Referenzkorpus* (DeReKo) wird im zweiten Teil genauer vorgestellt. Aus dem Gesamtkorpus (Release 2023-I) wurde für die Analyse das Subkorpus *pp – Plenarprotokolle* gewählt, das Plenarprotokolle aller deutschen Parlamente sowie einige Landtagsprotokolle aus Niederösterreich enthält (vgl. Archiv, *Plenarprotokolle*). Die ältesten darin enthaltenen Texte stammen aus Protokollen von 1995 und die neusten Sprachdaten stammen aus dem Jahr 2012 (vgl. Archiv, *Plenarprotokolle*). Insgesamt enthält das Korpus 6642 Texte und über 309 Millionen Wortformen (vgl. Archiv, *Plenarprotokolle*).

3. Teil 1 – *Computer Aided Translation (CAT) Tools*

3.1 *Computer Aided Translation versus Machine Translation (MT)*

Mit der stetig wachsenden Globalisierung in den letzten Jahrzehnten und dem daraus resultierenden steigenden Bedarf an zu übersetzenden Texten hat sich die Arbeit für Übersetzerinnen und Übersetzer verändert (vgl. Federico et al. 2014, S. 6). Die Übersetzungsarbeit ist Teil einer Industrie geworden, in der Automatisierung eine immer wichtigere Rolle spielt, um den Anforderungen des Marktes gerecht zu werden. Enge Deadlines und kleine Budgets kombiniert mit dem Anspruch an qualitativ hochwertige Übersetzungen führen dazu, dass die translatorische Arbeit effizienter gestaltet werden muss. Bereits seit einigen Jahren werden technologische Entwicklungen und Ressourcen deshalb genutzt, um den Übersetzungsprozess oder Teile davon zu automatisieren (Bowker & Barlow 2008, S. 2). Während maschinelle Übersetzungssysteme (*MT-Systeme*) wie *GoogleTranslate* oder *DeepL* die Möglichkeit bieten, Texte schnell, kostenlos und vollständig automatisch übersetzen zu lassen, sind es in der professionellen Übersetzungsarbeit besonders sogenannte *Computer Aided Translation (CAT) Tools*, die von den Expertinnen und Experten genutzt werden, um die Übersetzungsarbeit zu unterstützen (vgl. Orozco Agudelo 2022, S. 2; Esplà-Gomis et al. 2022, S. 7532).

Unter der Bezeichnung *Computer Aided Translation Tool* versteht man Computersoftware, die eingesetzt wird, um den Übersetzungsprozess zu unterstützen. CAT-Tools können beispielsweise genutzt werden, um die Übersetzung von wiederkehrenden Textteilen zu beschleunigen, indem die vorherige Übersetzung einer gleichen oder ähnlichen Textstelle wiederverwendet wird. Besonders bei formelhaften oder schematisch aufgebauten Texten kann dieses Vorgehen den Übersetzungsprozess beschleunigen (vgl. Li & Zhang 2020, S. 1). Die meisten CAT-Tools unterstützen jedoch nicht nur die Arbeit des Übersetzens selbst, sondern werden zusätzlich herangezogen, um eine Automatisierung des Übersetzungsworkflows sowie die Organisation mit anderen Übersetzerinnen und Übersetzern, die im selben Projekt arbeiten, zu unterstützen (vgl. Karpińska 2017, S. 141). Beispielsweise können spezielle Wörterbücher zwischen mehreren Beteiligten geteilt werden, um konsistente und

regelkonforme Übersetzungen über ein größeres Projekt hinweg zu garantieren. CAT-Tools sind deshalb in besonderem Maße an die professionelle Übersetzungsarbeit angepasst (Karpińska 2017, S. 134), was sie von anderen maschinellen Übersetzungssystemen, die häufig auch von Laien genutzt werden, unterscheidet.

Ein weiterer Unterschied zwischen CAT-Tools und MT-Systemen wie *DeepL* ist, dass in CAT-Tools ausgangssprachliche Texte für die Bearbeitung in kleinere, häufig an Satzgrenzen geteilte Bestandteile – sogenannte Segmente³ – gespalten werden. Es wird dann jeweils immer nur ein Segment bearbeitet und übersetzt, während es in den meisten MT-Systemen möglich ist, größere Texte einzufügen und in kurzer Zeit eine Übersetzung für den gesamten Text zu erhalten. Bei *DeepL* kann entweder direkt ein Text bis zu 3000 Zeichen eingegeben werden, oder man lädt eine Datei auf der Plattform hoch und erhält wenige Sekunden später eine Übersetzung des gesamten Dokuments. Abbildung 1 zeigt die Übersetzung eines Textes in *DeepL* während in Abbildung 2 dargestellt ist, wie derselbe Text für die Bearbeitung in dem CAT-Tool OmegaT in einzelne Segmente geteilt wurde. Das aktuell geöffnete Segment ist dabei grün markiert und die Übersetzung würde in den darunterliegenden Zeilen eingefügt werden, die in der Abbildung noch mit ausgangssprachlichem Text befüllt sind.

3 Auf den Segmentbegriff und die zugrundeliegende Definition soll im folgenden Kapitel genauer eingegangen werden.

German ▼ ↔ English (US) ▼ Glossary

Liebe Kolleginnen und Kollegen,
am Europatag, diesem für die Vergangenheit
und die Zukunft so symbolträchtigen Tag,
begrüßen wir den deutschen Bundeskanzler
Olaf Scholz.
Herr Bundeskanzler, herzlich Willkommen im
Europaparlament.
„Der Friede der Welt kann nicht gewahrt werden
ohne schöpferische Anstrengungen, die der
Größe der Bedrohung entsprechen.“ So beginnt
die von Robert Schuman am 9. Mai 1950
vorgelegte Erklärung. Diese Worte haben auch
heute an Bedeutung nicht verloren.
Jedes Jahr feiern wir an diesem Tag Europa.
Ein beispielloses Projekt der Aussöhnung, das
im
Wesentlichen auf Solidarität beruht. Ein Projekt,
das Menschen zusammenbringt, ohne zu
versuchen, uns alle gleich zu machen. Ein
Projekt, dessen Licht selbst einen Eisernen
Vorhang
und Betonmauern durchdringen konnte.

792 / 1500

Dear Colleagues,
On Europe Day, a day so symbolic for the past and
the future,
we welcome the German Chancellor, Olaf Scholz.
Chancellor, welcome to the European Parliament.
"The peace of the world cannot be preserved
without creative efforts commensurate with the
magnitude of the threat." So begins the declaration
presented by Robert Schuman on May 9, 1950
presented. These words have not lost their meaning
today.
Every year on this day we celebrate
unprecedented project of reconciliation
essentially on solidarity.
essentially based on solidarity.
people together without trying to
trying to make us all equal. A project
could penetrate even an iron curtain
and concrete walls.

relevance ...

significance ...

importance ...

poignancy ...

relevancy ...

meanings ...

👍 👎 ✎ 📄 🔗

Abb. 1: Darstellung eines ausgangs- (vgl. Metsola 2023a) und zielsprachlichen Textes im Textfenster des MT-Systems DeepL.

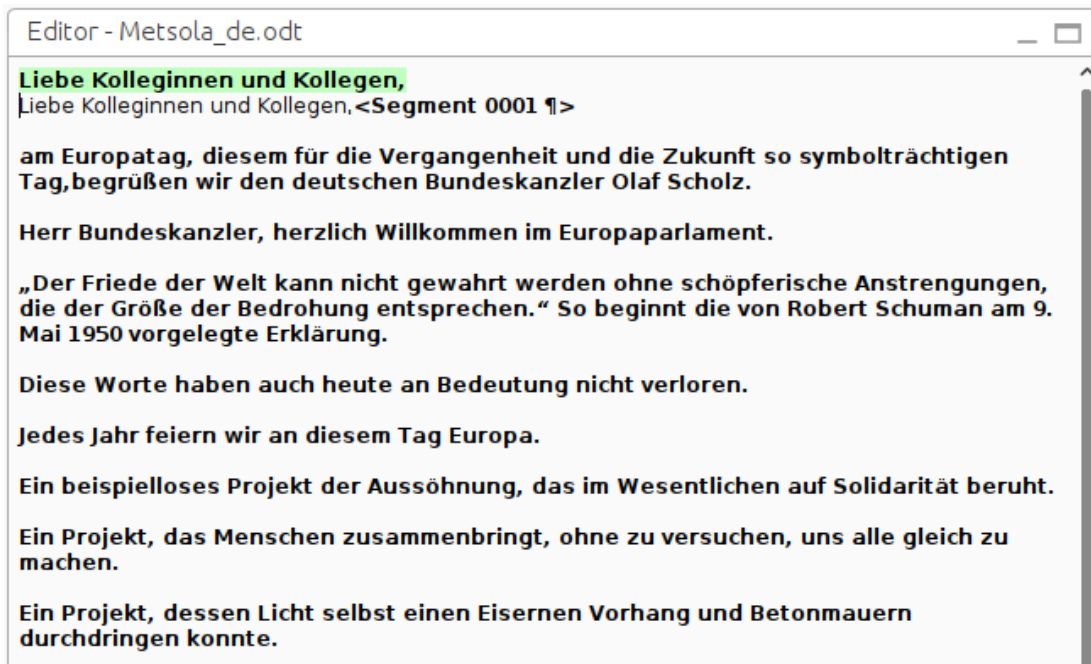


Abb. 2: Darstellung eines ausgangssprachlichen Textes im Editorfenster des CAT-Tools OmegaT (vgl. Metsola 2023a).

Neben der Aufbereitung des ausgangssprachlichen Textes wird zudem ein weiterer Unterschied zwischen den beiden Systemen erkenntlich: In *DeepL* erhält man mit einem Knopfdruck die Übersetzung des gesamten Textes. Für einzelne Wortformen liegen zwar Vorschläge für Übersetzungsalternativen vor, die nachträglich als Drop-Down Liste angezeigt werden können, aber der Text ist bereits fertiggestellt. Im CAT-Tool OmegaT hingegen wird Segment für Segment vorgegangen. Die Bereitstellung eines automatisch produzierten Übersetzungsvorschlags unterscheidet sich dabei von der Vorgehensweise in MT-Systemen. Je nach CAT-Tool werden Übersetzenden verschiedene, aus vergangenen Aufträgen extrahierte Übersetzungsvorschläge oder auch durch maschinelle Übersetzungssysteme produzierte Vorschläge präsentiert. Diese werden jedoch nicht automatisch übernommen und als fertiggestellter Gesamttext bereitgestellt, sondern können vorher akzeptiert, abgelehnt oder modifiziert werden (Vela et al. 2019, S. 1). Eine aktive Übernahme ist notwendig, wodurch Übersetzende bereits in den Entstehungsprozess des Gesamttextes involviert sind (vgl. Yao 2017, S. 106). CAT-Tools nehmen in Bezug auf den Handlungsspielraum der Übersetzenden demnach eine Zwischenposition zwischen reiner Humanübersetzung und rein maschineller Übersetzung ein (vgl. Bowker & Fisher 2010, S. 60).

Sowohl in CAT-Tools als auch in MT-Systemen kann ein Übersetzungsvorschlag nachträglich angepasst werden. Dabei ist jedoch zu beachten, dass eine automatisch produzierte Übersetzung nicht in jedem Fall die beste Option darstellt. Wie Zaretskaya et al. (2016) beschreiben, kann es je nach Übersetzungsszenario, Sprachkombination, Textdomäne, Genre und verwendetem Tool sein, dass eine Nachbearbeitung der gelieferten Übersetzung aufwändiger ist als eine von Grund auf neue Übersetzung und die verfügbaren Systeme deshalb nicht in allen Fällen den Übersetzungsprozess effektiv unterstützen (vgl. S. 91). Um Grenzen wie diese zu erkennen und Verbesserungsmöglichkeiten zu evaluieren, ist es wichtig, sich mit der genauen Funktionsweise von Übersetzungstools vertraut zu machen. Im Folgenden soll deshalb der Aufbau und Gebrauch von CAT-Tools näher beleuchtet werden.

3.2 Aufbau und Gebrauch

3.2.1 Häufige Bestandteile von CAT-Tools

Der Begriff *CAT-Tool* wird für eine Vielzahl verschiedener Systeme verwendet, die für die Übersetzungsarbeit verwendet werden. Die Werkzeuge unterscheiden sich beispielsweise in verfügbaren Lizenzmodellen, den zugrundeliegenden Modulen und Funktionen, den Zugriffs- und Installationsmöglichkeiten und in der Anbindungsmöglichkeit an maschinelle Übersetzungssysteme. Aufgrund der großen Unterschiede zwischen den einzelnen CAT-Werkzeugen werden im folgenden Kapitel nur die wichtigsten Bestandteile gegenwärtig genutzter CAT-Tools vorgestellt. Anschließend wird im nächsten Kapitel OmegaT exemplarisch herangezogen, um ein CAT-Tool und dessen Eigenschaften und Funktionen näher vorzustellen.

Ein zentraler Bestandteil der meisten CAT-Tools sind sogenannte *Translation Memories* (TM) (vgl. Vela et al. 2019, S. 1; Li & Zhang 2020, S. 2). In diesen Datenbanken können übersetzte Texte gespeichert und in zukünftigen Übersetzungsaufträgen recycelt werden. Für die Wiederverwendung bereits übersetzter Texte muss der Ausgangstext mit seiner Übersetzung aligniert werden, das heißt er wird mit seinem zielsprachlichen Äquivalent zusammengeführt (vgl. Bowker & Fisher 2010, S. 61). Dazu werden die Texte in der Ausgangs- und Zielsprache zunächst in kleinere Einheiten – sogenannte Segmente – geteilt und anschließend dem jeweiligen Übersetzungsäquivalent zugeordnet. Zusammen bilden sie dann eine als *Translation Unit* bezeichnete Einheit (vgl. Esplà-Gomis et al. 2022, S. 7532). In der Regel erfolgt die Segmentierung nach Sätzen – die Volltexte werden also in einzelne Sätze geteilt und so gespeichert (vgl. Tatu 2011, S. 4). Wird ein neuer Ausgangstext im CAT-Tool geöffnet, wird auch dieser in einzelne Segmente geteilt. Sobald für eine neue Übersetzung ein noch nicht bearbeitetes Segment geöffnet wird, wird zunächst im TM nach einem in vergangenen Arbeiten bereits übersetzten Segment gesucht, das zu einem vorher festgelegten Ähnlichkeitsgrad – in der Regel als Prozentwert angegeben – mit dem aktuellen Segment formal übereinstimmt. Bei einem Treffer wird das gefundene Segment in der Zielsprache bereitgestellt und kann zumeist per Tastenkombination in den neuen Zieltext übernommen werden. Man unterscheidet dabei zwischen exakter

Übereinstimmung, sogenannten *Exact Matches*, und einer graduellen Übereinstimmung zwischen den beiden Segmenten. Als sogenannte *Fuzzy Matches* werden Übereinstimmungsrelationen bezeichnet, bei denen „some degree of similarity to a segment stored in the TM database“ (Bowker & Fisher 2010, S. 61) vorliegt. Die Mindestübereinstimmung des beim Öffnen des neuen Segments präsentierten Vorschlags kann vorher von der Nutzerin oder dem Nutzer festgelegt werden und liegt – je nach Anforderung und Präferenz – etwa bei 60 bis 70% (vgl. Bowker & Fisher, S. 61). Dabei bezieht sich Ähnlichkeit zwischen den beiden Texteinheiten jedoch ausschließlich auf die formale Übereinstimmung der Zeichenabfolgen. Die Bedeutung eines Satzes oder einzelner Wortformen wird nicht ausgewertet und für die Suche nach wiederverwendbaren Segmenten berücksichtigt (vgl. Vela et al. 2019, S. 2). Neu übersetzte Segmente fließen sofort in die inkrementell wachsende Datenbank mit ein und können so im weiteren Verlauf der Übersetzungsproduktion erneut verwendet werden (vgl. Li & Zhang 2020, S. 2). Je nach Tätigkeit wird den Übersetzenden für ein Projekt entweder ein bestehendes TM mitgeliefert, oder es muss ein eigenes TM erstellt werden. Dies kann entweder über einen längeren Zeitraum durch die iterative Speicherung vergangener Jobs und ein dadurch stetig wachsendes TM erfolgen, oder es wird eine bereits bestehende Ressource wie beispielsweise das in dieser Arbeit verwendete DGT-TM genutzt, das von der europäischen Kommission zur Verfügung gestellt wird. Die darin enthaltenen Texteinheiten wurden für die Nutzung als *Translation Memory* aligniert, aufbereitet – zum Beispiel indem Dopplungen entfernt wurden und als .tmx-Datei zum Download bereitgestellt (vgl. DGT-TM).

Zusätzlich zu TMs bilden auch sogenannte *Terminology Management Systems* (TMS) einen wichtigen Bestandteil der gegenwärtig genutzter CAT-Tools (vgl. Vela et al. 2019, S. 1). Auch in TMS werden Texteinheiten gespeichert. Während TMs jedoch in der Regel ganze Sätze beinhalten, werden in TMS-Datenbanken terminologische Informationen gespeichert (vgl. Bowker & Fisher 2010, S. 61). Mithilfe eines TMS-Moduls lässt sich gespeicherte Terminologie entweder durchsuchen, oder es liefert ähnlich wie ein TM Übersetzungsvorschläge für Termini, die dann übernommen, modifiziert oder abgelehnt werden können. Weiterhin enthalten die Datenbanken Listen nicht zu übersetzender Begriffe, die – beispielsweise da sie Ei-

genannt sind – ohne Veränderung aus dem Ausgangssprachlichen Text übernommen werden sollen (vgl. Trados 2017). Außerdem bieten einige TMS die Möglichkeit, zusätzlich kurze Erklärungen oder Bilder mit weiteren Informationen zu Fachtermini zu speichern, die für die Übersetzung relevant sein können (vgl. Trados 2017). Weiterhin können TMS-Module genutzt werden, um Texte automatisch zu scannen und neue Termini zu extrahieren, die anschließend in die Terminologiedatenbank aufgenommen werden können, oder bestehende Termini durch neu gefundene Begriffe zu ersetzen (vgl. Bowker & Fisher 2010, S. 62). Da Terminologie je nach Textart verschieden sein kann, wird ein TMS häufig von den Kundinnen oder Kunden an die Übersetzungsagentur oder einzelne Übersetzende mitgeliefert (Petrova 2019, S. 90).

Eine weitere Funktion, die häufig Bestandteil von CAT-Systemen ist, ist eine Anbindungsmöglichkeit an maschinelle Übersetzungssysteme (vgl. Vela et al. 2019, S. 1; Petrova 2019, S. 89). CAT-Werkzeuge mit dieser Funktion rücken auf der zuvor erläuterten Skala zwischen humaner und voll automatisierter Übersetzung noch weiter in Richtung rein maschinell produzierter Übersetzung. Dennoch unterscheidet sich diese Art des Übersetzens von klassischen MT-Systemen, denn zumindest in der Theorie ist der Mensch aktiv an der Entstehung des übersetzten Textes beteiligt. MT-Anbindungen als Bestandteil eines CAT-Systems liefern, ähnlich wie zuvor beschriebene TMs, einen Übersetzungsvorschlag, der durch aktives Handeln der Übersetzenden übernommen oder abgelehnt werden kann. Diese Form der automatisierten Übersetzung geht jedoch über die Funktion eines TMs hinaus, da Übersetzungsvorschläge neu generiert werden. Die Funktionsweise kommt also besonders dann zum Einsatz, wenn im TM keine passenden Übersetzungsvorschläge gefunden werden können (vgl. Bowker & Fisher 2010, S. 62). Weiterhin kann in den meisten CAT-Systemen zwischen verschiedenen MT-Anbindungen gewechselt werden und die Möglichkeit, maschinell generierte Übersetzungsvorschläge zu erhalten, kann auch vollständig ausgeschaltet werden (vgl. Trados o. J.). Aus diesem Grund müssen deshalb auch CAT-Werkzeuge mit MT-Anbindung von rein maschinellen Übersetzungssystemen unterschieden werden.

Als letztes Teilelement gegenwärtiger CAT-Systeme soll nun noch die Funktion der *Quality Estimation* (QE) vorgestellt werden. Fast alle heutzutage genutzten

CAT-Tools enthalten automatisierte Prüfroutinen, die einerseits Qualitätsprüfungen im zielsprachlichen Text durchführen und andererseits gefundene Fehler auswerten und in sogenannten *QE Scores* darstellen. Dabei können mithilfe der Datenbanken für Terminologie im letzten Schritt des Übersetzungsworkflows beispielsweise abschließende Terminologieprüfungen durchgeführt werden, es können vergessene, nicht übersetzte Segmente gefunden werden und es können Tippfehler oder Dopplungen sowie URL-Fehler gefunden werden (vgl. Petrova 2019, S. 92f.). Die meisten dieser Prüfungen erfolgen regelbasiert, aber erste Applikationen werben auch mit Qualitätsprüfungen, bei denen neuronale Netze eingesetzt werden. So wird das neueste Update des speziell auf QE spezialisierten Tools *LexiQA* mit folgenden Worten beworben: „Our QA support for these locales comes bundled with brand new neural spellcheckers using our proprietary technology“ (LexiQA o. J.).

Je nach CAT-Tool ist es möglich, mithilfe der Prüfroutine weitere Qualitätsprüfungen durchzuführen, die in den meisten Werkzeugen je nach Bedarf aus- oder abgewählt werden können. Wie das Beispiel von *LexiQA* zeigt, gibt es neben den klassischen CAT-Tools auch computergestützte Applikationen, die auf QE spezialisiert sind und – entweder von Übersetzenden selbst oder nach Lieferung des Produktes von den Auftraggebenden – nochmal separat eingesetzt werden können (vgl. Petrova 2019, S. 92). Je nach Auftrag und Art der Übersetzungen können bestimmte Tools oder Prüffunktionen besser oder weniger geeignet sein. Wie Petrova (2019) jedoch zeigt, ist besonders die korrekte Verwendung von Terminologie von großer Wichtigkeit für qualitativ hochwertige Übersetzungen (vgl. S. 95). Die Prüfung auf korrekt und konsistent verwendete Terminologie ist deshalb eine der wichtigsten Bausteine der *Quality Estimation*.

3.2.2 Gebrauch von CAT-Tools

Im folgenden Kapitel soll ein kurzer Überblick über den Einsatz von CAT-Tools in der Übersetzungsarbeit gegeben werden. Über den Gebrauch von CAT-Tools gibt es viele Studien, die sich mit dem Einsatz von CAT-Tools unterschiedlichster Art und in unterschiedlichsten Personengruppen beschäftigen. Die nachfolgenden Erläuterungen sollen einen Überblick über häufige Einsatzbereiche und Anwendungsmöglichkeiten der computergestützten Werkzeuge geben, stellen aber keine voll umfassende Ausarbeitung zum Gebrauch von CAT-Tools im Übersetzungsprozess dar.

Einen guten Einstiegspunkt für die Skizzierung des Einsatzbereichs von CAT-Tools in der professionellen Übersetzungsarbeit bietet der Artikel von Verplaetse und Lambrechts (2019). Die beiden Autorinnen betrachten darin nicht nur vergangene Studien zum Gebrauch von CAT-Tools, sondern erläutern zudem die Ergebnisse selbst durchgeführter Befragungen, die sie mit anderen Studien in Relation setzen. Die Ergebnisse ihrer eigenen Studie zeigen, dass in den Jahren 2017 und 2018 ein Großteil der befragten Teilnehmerinnen und Teilnehmer (80%) CAT-Tools für die professionelle Übersetzungsarbeit nutzten (vgl. S. 16). Dabei wurden sowohl angestellte als auch selbstständige Übersetzerinnen und Übersetzer befragt. Dies deckt sich in etwa mit den Ergebnissen einer im Jahr 2015 veröffentlichten Studie von Picon et al., die dem Einsatz von CAT-Tools für die Übersetzungsarbeit in der Schweiz nachging.

Eine Studie von Orozco Agudelo veröffentlicht im Jahr 2022 unterstreicht, dass Studien zum Gebrauch von CAT-Tools häufig schwerpunktmäßig die Situation in Europa beleuchten, aber dieser Kontinent nur etwa die Hälfte des globalen Sprachmarktes ausmacht (vgl. S. 10). Im Rahmen seiner Befragung im kolumbianischen Übersetzungskontext wird deutlich, dass nur 47% der dort ansässigen Befragten CAT-Tools oder Teilmodule davon in ihrer Arbeit nutzen (vgl. S. 35). Etwa 40% der Befragten gaben an, dass fehlendes Training im Bereich automatisierter Übersetzungswerkzeuge ein Grund dafür ist, dass sie CAT-Werkzeuge nicht nutzen (vgl. Orozco Agudelo 2022, S. 36). Weiterhin sind 14 der insgesamt 19 in seiner Studie befragten Teilnehmerinnen und Teilnehmer, die angaben, auch in Zukunft kein CAT-Tool für ihre Übersetzungsarbeit nutzen zu wollen, älter als 60 Jahre alt (vgl. S. 68).

Die Tools werden also vorrangig von älteren Personen, die möglicherweise schon langjährige Berufserfahrung ohne die Unterstützung von computergestützten Werkzeugen gesammelt haben, abgelehnt.

Ein Vergleich verschiedener weltweit durchgeführter Studien zum Gebrauch von CAT-Tools in der Translationsarbeit stützt die Annahme, dass die Werkzeuge von der großen Mehrheit aller Befragten genutzt wird. Verschiedene globale Befragungen, die im Zeitraum von 2006 bis 2015 durchgeführt wurden, zeigen, dass insgesamt mindestens 75 Prozent aller Befragten CAT-Tools bei der Arbeit nutzen (vgl. Verplaetse & Lambrechts 2019, S. 10). Während die einzelnen Studien zwar unterschiedliche Berufsprofile innerhalb der Übersetzungsarbeit fokussieren, fassen Verplaetse und Lambrechts (2019) zusammen, dass sowohl selbstständig als *Freelancer* arbeitende Übersetzerinnen und Übersetzer als auch angestellte *in-house* Übersetzende CAT-Tools regelmäßig nutzen (vgl. S. 11). Wichtig ist dabei, was schon im vorigen Kapitel angemerkt wurde: Unter dem Begriff CAT-Tool werden Werkzeuge zusammengefasst, die unterschiedlich zusammengesetzt sein können und je nach Teilmodulen verschiedene Funktionen haben können. Die Definition des Begriffes *CAT-Tool* kann also je nach Studie und je nach Verständnis der oder des Befragten unterschiedlich sein und die Ergebnisse der Befragungen beeinflussen. Zusammenfassend lässt sich aber feststellen, dass computergestützte Werkzeuge verschiedener Art zwar nicht in allen Ländern gleich häufig genutzt werden, aber insgesamt betrachtet dennoch die meisten aller professionellen Übersetzerinnen und Übersetzer Gebrauch von CAT-Tools machen.

Ein wiederkehrendes Argument für den Gebrauch von CAT-Werkzeugen in der Übersetzungsarbeit ist deren Unterstützung bei kollaborativen Arbeiten (vgl. z.B. LeBlanc 2013, S. 7; Gaspari 2003, S. 35). Die Möglichkeit, TMs und Terminologie-datenbanken zu teilen oder Übersetzungsaufträge zentral zu verwalten, eignet CAT-Tools besonders für den Einsatz in größeren Übersetzungsprojekten, in denen mehrere Texte übersetzt werden müssen und mehrere Personen beteiligt sind – gegebenenfalls auch über längere Zeiträume hinweg (vgl. LeBlanc 2013, S. 7).

Weiterhin zeigen verschiedene Studien, dass sich CAT-Tools besonders für die Übersetzung von sachlichen oder technischen Texten wie Betriebsanleitungen oder Verträge eignen, da diese Textsorten besonders häufig repetitive oder standardi-

sierte Strukturen oder Textteile aufweisen (vgl. z.B. Petrova 2019, S. 89). Die von Verplaetse und Lambrechts (2019) durchgeführte Studie zum Einsatz von CAT-Tools aus 2017 und 2018 zeigt zudem, dass Übersetzungsexpertinnen und -experten, die besonders auf Domänen wie Literatur, Geschichte und Kunst spezialisiert sind, seltener auf CAT-Werkzeuge zurückgreifen (vgl. S. 16). Diese Fachbereiche, so die Autorinnen, enthalten weniger wiederholende oder formalisierte Strukturen und sind deshalb weniger für den Gebrauch von CAT-Tools geeignet (vgl. S. 16). Und auch die Befragung von Orozco Agudelo (2022) bestätigt, dass die Befragten, die angaben regelmäßig ein CAT-Tool zu nutzen, eher technische, naturwissenschaftliche oder audiovisuelle Texte übersetzen (vgl. S. 49). Die computergestützten Werkzeuge werden also besonders in bestimmten Textdomänen genutzt, während sich andere Inhalte weniger für den Einsatz von CAT-Tools eignen.

Zusammenfassend lässt sich also festhalten, dass Werkzeuge, die den Übersetzungsprozess mithilfe verschiedener Teilmodule automatisiert unterstützen, von der großen Mehrheit professioneller Übersetzerinnen und Übersetzer weltweit genutzt werden. Während CAT-Tools in bestimmten Ländern und Regionen häufiger genutzt werden als in anderen, ist ein immer wieder genannter Grund für die Nichtnutzung von CAT-Werkzeugen fehlendes Training. Weiterhin zeigen mehrere Studien, dass die Tools für die Übersetzung verschiedener Textarten unterschiedlich gut geeignet sind. Eine Untersuchung der genauen Funktionsweise sowie der Stärken und Schwächen von computergestützten Übersetzungswerkzeugen ist deshalb umso wichtiger, um den Einsatz von CAT-Tools in der Translationsarbeit zu evaluieren und zu optimieren. Im folgenden Kapitel wird ein konkretes Werkzeug herangezogen und in seiner Funktionsweise näher beleuchtet.

3.3 CAT-Tools im Übersetzungsprozess am Beispiel von OmegaT

OmegaT ist ein kostenlos nutzbares und für die professionelle Übersetzungsarbeit konzipiertes CAT-Tool, das über eine Website⁴ für verschiedene Betriebssysteme heruntergeladen und installiert werden kann. Was OmegaT von anderen häufig genutzten Werkzeugen unterscheidet, ist nicht nur die kostenfreie Nutzung, sondern auch die Möglichkeit, das Tool an individuelle Bedürfnisse anpassen zu können, denn durch den öffentlich zugänglichen Quelltext kann die Software modifiziert werden (vgl. OmegaT). Dazu sind allerdings Programmierkenntnisse notwendig. Für diese Arbeit wurde die Version 5.3.0 für Linux genutzt.

Über den Reiter „Documentation“ auf der Website der Software gibt es verschiedene Trainingsmaterialien und Videos zur Nutzung von OmegaT. Hier können sich Übersetzerinnen und Übersetzer über die Funktionalitäten und Einstellungen des Werkzeugs informieren und an konkreten Fallbeispielen den Einsatz von OmegaT für die Übersetzungsarbeit erlernen. Zudem gibt es einen offiziellen User Guide, der neben einer Installationsanleitung detaillierte Ausführungen zu den einzelnen Funktionen von OmegaT enthält.

Während Studien zum Gebrauch von CAT-Tools zeigen, dass Übersetzerinnen und Übersetzer meist kostenpflichtige Werkzeuge wie *SDL Trados* oder *MemoQ* nutzen, ist OmegaT eines der am häufigsten genutzten kostenfreien Werkzeuge (vgl. Attila 2015, S. 198; Orozco Agudelo 2022, S. 34). Auch in Übersetzerforen wird erkennbar, dass OmegaT von Expertinnen und Experten genutzt und teilweise auch kostenpflichtigen CAT-Tools vorgezogen wird:

Als 8-jähriger Trados-Benutzer und 2-jähriger OmegaT-Benutzer musste ich die Erfahrung machen, dass die Argumente eher für OmegaT sprechen [...] Trados wurde hauptsächlich für den Profit entwickelt, von Entwicklern, die wenig Ahnung vom Übersetzen haben. OmegaT hingegen wurde von Linguisten entwickelt, die gleichzeitig auch Programmierer sind, im Grunde von Übersetzer für Übersetzer (Wuelk 2021).

Im Folgenden sollen das Tool und seine Funktionalitäten genauer vorgestellt werden. Die Beschreibung folgt übergeordneten Prozessschritten der Übersetzungsarbeit beginnend mit der Projektvorbereitung und orientiert sich dabei an der Analyse von Keller (2011), die für ihren Vergleich verschiedener TM-Systeme eine ähnliche Beschreibungsstruktur verwendet (vgl. S. 1). Im Folgenden soll zunächst gezeigt wer-

4 Der Link zur Website ist dem Quellenverzeichnis am Ende dieser Arbeit zu entnehmen.

den, welche Unterstützung das Tool im Übersetzungsprozess bietet, während im dritten Teil dieser Arbeit Grenzen diskutiert werden.

3.3.1 Projektvorbereitung

Bevor ein Text mithilfe des CAT-Tools OmegaT übersetzt werden kann, muss zunächst ein neues Projekt angelegt werden (vgl. Keller 2011, S. 53). In einem Dialogfeld lässt sich sowohl der Speicherort des Projektes als auch die Speicherorte und Benennung der jeweiligen Unterordner für TMs, Glossare und Wörterbücher festlegen. Außerdem wird hier die Sprache der Quell- und Zieldateien bestimmt. Zudem können drei Optionen an- oder abgewählt werden: *Satzsegmentierung aktivieren*, *automatische Weitergabe von Übersetzungen* und *Tags entfernen*. Bezüglich der Segmentierungsregeln können auch projektspezifische Regeln definiert werden. So kann, wie in Abbildung 3 durch rot markierte Haken gezeigt, für das geöffnete Projekt festgelegt werden, dass beispielsweise die Satzzeichen *.?!* auch gefolgt von schließenden Anführungszeichen als Segmentgrenze gehandhabt werden sollen, sofern danach ein Leerzeichen folgt (vgl. auch OmegaT HTML Handbuch, Anlage „Reguläre Ausdrücke“).

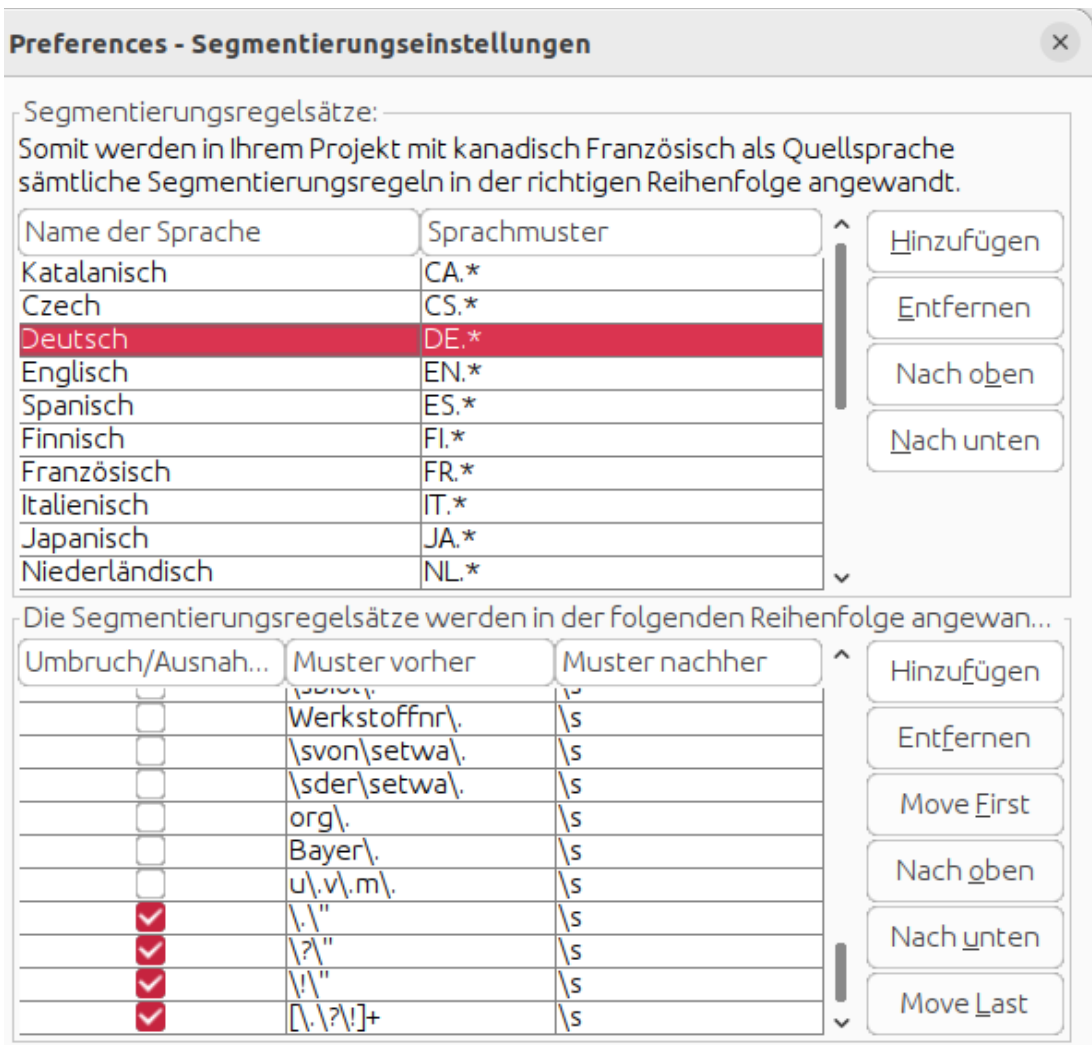


Abb. 3: Ausschnitt aus dem Dialogfenster projektspezifischer Segmentierungsregeln des Deutschen in OmegaT.

Die Option *Automatische Weitergabe von Übersetzungen* legt Regeln für Segmente fest, für die es verschiedene vorgeschlagene Übersetzungen gibt. Übersetzungsvarianten können beispielsweise entstehen, wenn mehrere maschinelle Übersetzungssysteme, die an OmegaT angebunden werden können, unterschiedliche Vorschläge für ein Segment liefern. Wenn diese Funktion aktiviert ist und man eine Variante wählt, wird bei allen nachfolgenden, gleichen Segmenten automatisch diese Übersetzung präsentiert. Bei einer Deaktivierung werden erst am Schluss, wenn sich der oder die Bearbeitende auf eine Übersetzungsvariante festgelegt hat, alle Segmente jeweils übersetzt (vgl. OmegaT Handbuch, S. 18). Projektspezifisch können außerdem verschiedene Einstellungen in Bezug auf zu bearbeitende Dateiformate

te und darin enthaltene Tags vorgenommen werden (OmegaT Handbuch, S. 27). Es können beispielsweise für das jeweilige Projekt geltende Dateifilter aktiviert werden, die bestimmen, welche Dateitypen mit dem Programm bearbeitet werden sollen und welche im Quellordner vorhandenen Dateitypen übersprungen werden sollen. Zudem gibt es unter anderem die Möglichkeit, den Dateinamen für die Suche nach Alternativübersetzungen auszuklammern. So können Übersetzungsvarianten unabhängig vom Dateiformat gefunden und vorgeschlagen werden. Für die nachstehenden Untersuchungen wurden sowohl die Option *Satzsegmentierung aktivieren* als auch die Option *automatische Weitergabe von Übersetzungen* entsprechend der Voreinstellung aktiviert. Auch in Bezug auf projektspezifische weiterführende Einstellungen wurden die Voreinstellungen übernommen.

Bestätigt man sowohl die zuvor beschriebenen Einstellungen als auch die vorgeschlagenen Speicherorte für die einzelnen Inhalte, erhält man im nächsten Fenster eine Übersicht über das neu geladene Projekt. Neben einer Auflistung der einzelnen Dateien im Quellordner des Projekts sieht man hier weiterhin beispielsweise die Gesamtzahl der Segmente sowie die Anzahl einmaliger Segmente. Zusätzlich zu dieser Übersicht wird automatisch außerdem eine detaillierte Statistik des Projekts im Projektordner *omegat* auf dem Laufwerk gespeichert, die beispielsweise auch die genaue Anzahl an Dateien, Wörtern oder Zeichen enthält. Will man nachträglich einzelne Dateien zum Projekt hinzufügen, können diese einfach in den entsprechenden Projektordner namens *source* kopiert werden und anschließend muss das Projekt erneut geladen werden (vgl. Keller 2011, S. 53f.).

Obwohl das Tool in seinen Unterordnern und Einstellungsmöglichkeiten klar strukturiert ist und Schritt für Schritt mithilfe von Dialogfenstern durch den Vorbereitungsprozess geführt wird, ist bereits vor Beginn der eigentlichen Übersetzungsarbeit ersichtlich, dass sowohl technisches Wissen als auch Erfahrung im Umgang mit OmegaT notwendig ist. CAT-Tools zeichnen sich laut verschiedener Studien generell durch eine größere Lernkurve aus (vgl. Bowker & Fisher 2010, S. 63; Yao 2017, S. 112), doch OmegaT sticht dabei besonders hervor. Auch Keller (2011) stellt fest, dass OmegaT im Vergleich zu anderen CAT-Tools in der Handhabung „fundiertes Computerwissen voraus[setzt]“ (S. 60), während dies bei anderen Tools weniger notwendig ist (vgl. S. 22).

3.3.2 OmegaT während der Übersetzung

Während der Übersetzung eines Textes können in OmegaT verschiedene Ansichten angezeigt werden. Oben links befindet sich der Editor, in dem die zu übersetzenden Texte, wie Abbildung 4 zeigt, dargestellt werden (vgl. OmegaT Handbuch, S. 39). Bei der Bearbeitung der einzelnen Dateien werden die Texte entsprechend der eingestellten Segmentierungsregeln getrennt und Segment für Segment präsentiert. Farbliche Markierung kennzeichnet das aktuell zu bearbeitende Quellsegment grün. Das Feld für die Übersetzung öffnet sich beim Auswählen eines Segments und ist zunächst mit der Quellsprache oder einer Vorübersetzung gefüllt, bevor es durch die Übersetzung ersetzt wird. Wurde eine Übersetzung eingefügt, wird der Fettdruck des Quellsegments aufgehoben.

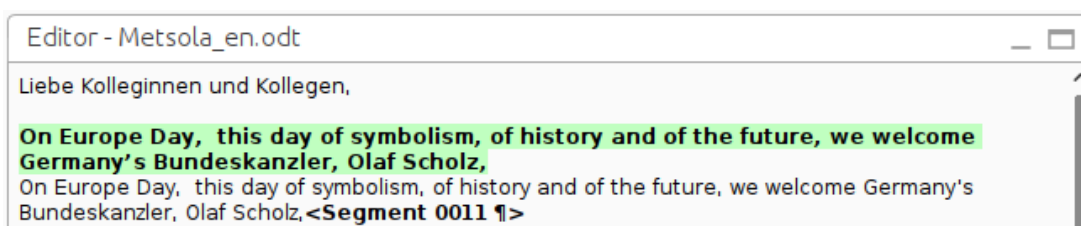


Abb. 4: Bereits ins Deutsche übersetztes und grün markiertes, geöffnetes Segment im Editorfenster in OmegaT (vgl. Metsola 2023b).

Durch die Gliederung des Textes in einzelne Segmente und die Hervorhebung des aktuellen Segments wird der Übersetzungsprozess in OmegaT klar strukturiert. Der Fokus liegt klar auf dem zu bearbeitenden Segment während andere Textteile in den Hintergrund rücken. Einer der in Forschungsarbeiten und Anwendungsstudien am häufigsten genannten Vorteile von CAT-Tools ist, dass sie die Produktivität der Übersetzungsarbeit erhöhen (vgl. Bowker & Fisher 2010, S. 63; LeBlanc 2013, S. 6). Indem sich die Übersetzenden mit OmegaT Schritt für Schritt in Form von aufgrund ihrer Länge leichter zu bewältigenden Zuschnitten linear durch die Texte arbeiten und vom letzten Segment des vorherigen Textes automatisch zum nächsten Segment des nachfolgenden Textes im Projekt übergehen (vgl. Keller 2011, S. 55), trägt die Darstellungsform im Editor zu dieser Produktivitätssteigerung bei. Die Übersetzung eines Textes erfolgt geradlinig und gleichmäßig und kann bis zum Projektende fort-

laufend durchgeführt werden. Übersetzende werden durch die klare Strukturierung, die Auflösung von Textgrenzen sowie die Hervorhebung aktuell bearbeiteter Segmente darin unterstützt, sich auf das aktuelle Segment zu fokussieren und damit effizienter zu arbeiten.

Eine weitere Funktion, die zur Produktivitätssteigerung in der Übersetzungsarbeit beitragen kann, findet sich rechts oben im Fenster *Unschärfe Treffer*. Dort werden mögliche *Matches* aus dem TM angezeigt (vgl. OmegaT Handbuch, S. 35). In OmegaT wird eine komplexe TM-Struktur verwendet, denn es werden verschiedene Ablageordner für verschiedene TMs verwendet (vgl. Keller 2011, S. 54). Übersetzungen aus dem aktuell geöffneten Projekt werden im Projektordner *omegat* in einer Datei im Format *.tmx* gespeichert. Im Ordner *tm* wiederum können beliebige TM-Dateien gespeichert werden, die entweder durch frühere Projekte entstanden sind, oder aus anderen Quellen gezogen wurden und im aktuellen Projekt genutzt werden sollen (vgl. Keller 2011, S. 54). Auch die Auszüge aus dem in diesem Projekt verwendeten DGT-TM wurden in diesem Ordner gespeichert.

Vorschläge aus dem TM werden mit einer Prozentzahl versehen, die einen Hinweis auf den Übereinstimmungsgrad zwischen Quellsegment und im TM gespeicherten Segment liefern. Außerdem lässt sich nachvollziehen, aus welchem Ordner und welcher Datei der Übersetzungsvorschlag entstammt, falls mehrere TM-Ordner angelegt wurden. Ist über ein TM bekannt, dass es weniger gut gepflegt ist und möglicherweise Fehler enthält, kann es in einem separaten Ordner gespeichert werden. Bei der Berechnung der Übereinstimmung von zwei Segmenten wird für dieses TM dann beispielsweise eine festgelegte „penalty“ mitberechnet, die dafür sorgt, dass der Prozentwert für die Übereinstimmung verringert wird und weniger *Matches* aus diesem TM als aus anderen angezeigt werden (vgl. OmegaT Handbuch, S. 43). Voraussetzung hierfür ist jedoch, dass die Qualitätsmängel dem Bearbeitenden bekannt sind.

Bei *Fuzzy Matches* aus einem der verfügbaren TMs werden die Unterschiede zwischen dem Vorschlag und der bearbeiteten Textstelle farblich hervorgehoben (vgl. Keller 2011, S. 56). Gleiche Wortformen sind wie in Abbildung 5 in der Ausgangssprache schwarz markiert während Unterschiede zwischen dem Quellsegment und dem *Fuzzy Match* blau markiert sind. Gibt es mehrere *Matches*, wird dasjenige mit

der höchsten Übereinstimmung ganz oben und fett markiert dargestellt (vgl. Keller 2011, S. 56).

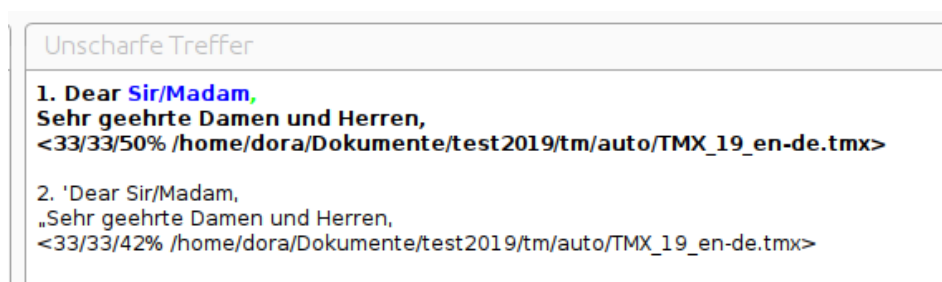


Abb. 5: Fuzzy Match im Fenster „Unschärfe Treffer“ in OmegaT.

Auch die Möglichkeit, bereits übersetzte Textstellen wiederzuverwenden – gegebenenfalls sogar über ganze Projektgruppen hinweg – kann zur Produktivitätssteigerung im Übersetzungsprozess führen: Mithilfe eines TM-Vorschlags werden ähnliche Textstellen konsistenter übersetzt (vgl. Krüger 2016, S. 115), was beispielsweise in Betriebsanleitungen hilfreich sein kann, um Bezugspunkte zwischen einzelnen Textstellen herzustellen und Beschreibungen so deutlich wie möglich zu gestalten. Selbst wenn mehrere Übersetzer und Übersetzerinnen an einem großen Projekt arbeiten, können durch das Teilen und gemeinsame Aufbauen eines TMs konsistente Übersetzungen produziert werden. Andere Beteiligte, so schreibt LeBlanc (2013), profitieren von den „strokes of genius“ (S. 6) ihrer Kolleginnen und Kollegen und arbeiten so schneller und konsistenter.

Außerdem entfällt bei der Übernahme eines Vorschlags durch nur eine einzige Tastenkombination viel Zeit für das Eruiieren und Tippen einer neuen Übersetzung, die einen Mehraufwand an Arbeitszeit und damit verbunden auch Mehrkosten mit sich bringt (vgl. Verplaetse & Lambrechts 2019, S. 10). Übersetzerinnen und Übersetzer können durch das Recycling mehr Text in kürzerer Zeit bearbeiten. Je mehr ähnliche Bestandteile ein Text enthält, desto mehr Material kann wiederverwendet werden. CAT-Tools eignen sich somit im Übersetzungsprozess besonders für repetitive Texte, die besonders viele wiederholende Begriffe und Strukturen enthalten (vgl. Tatu 2011, S. 177; Petrova 2019, S. 89). Im dritten Teil dieser Arbeit soll dies anhand eines konkreten Beispiels weiter verdeutlicht werden.

Werden im TM keine passenden Vorschläge für ein Segment gefunden, ist es in OmegaT zusätzlich möglich, eine Suche nach einzelnen Zeichenfolgen durchzuführen. Beispielsweise über Wildcards oder reguläre Ausdrücke können so ähnliche Zeichenfolgen oder Textstellen sowohl im ganzen Projekt als auch in einzelnen Dateien gefunden werden, die für die aktuelle Übersetzung hilfreich sein können (vgl. OmegaT Handbuch, S. 21ff.). Ein Suchausdruck wird jedoch immer nur im Kontext der *Translation Unit*, in der er sich befindet, angezeigt.

Weitere Übersetzungsvorschläge zu Terminologie und festen Ausdrücken, die für das aktuell geöffnete Segment relevant sein könnten, sind über das Fenster *Glossar* abrufbar. Mithilfe dieser Funktion können terminologische Besonderheiten oder Vorgaben berücksichtigt werden. Wie Petrova (2019) feststellt, ist die Nichteinhaltung von terminologischen Vorgaben der Hauptgrund, warum Übersetzungen zur Korrektur zurückgeschickt werden (S. 90). Ein Glossar kann außerdem nicht nur zu einer besseren Qualität der Übersetzungen beitragen, sondern es erleichtert zudem die Zusammenarbeit mit anderen Übersetzerinnen und Übersetzern. Wird zusammen mit einem Übersetzungsauftrag ein Glossar mitgeliefert oder parallel dazu aufgebaut, nutzen im Idealfall alle Beteiligten dieselben Begriffe. Es resultieren konsistente Übersetzungen, die den Vorgaben des Auftraggebers folgen (vgl. Verplaetse & Lambrechts 2019, S. 21). Voraussetzung dafür ist natürlich wie auch bei TMs, dass die Glossare gut gepflegt sind und nur Einträge beinhalten, die eindeutig formuliert und nach aktuell gültig sind.

Neben den Termini und ihrem Übersetzungsäquivalent kann in OmegaT auch ein Hinweis auf Kundinnen und Kunden oder Fachgebiete als Teil des Eintrags gespeichert werden (vgl. OmegaT Handbuch, S. 56). Allerdings ist der Aufbau von Terminologiedatenbanken im Vergleich zu anderen CAT-Tools in OmegaT eher einfach gehalten, denn es können beispielsweise keine Grafiken zusammen mit terminologischen Einträgen gespeichert werden. Dieser Unterschied ist ein wichtiges Beispiel dafür, dass bei der Vorbereitung eines Übersetzungsauftrages sowie der Pflege eines Glossars berücksichtigt werden muss, welches CAT-Tool im Übersetzungsprozess verwendet wird. Je nach technischen Gegebenheiten müssen verschiedene Anforderungen erfüllt sein, damit ein eindeutiges Verständnis von terminologischen Vorgaben möglich ist. Ist das Glossar nicht entsprechend aufbereitet oder sogar veraltet, kön-

nen die Übersetzungsvorschläge die Produktivität und Übersetzungsqualität sogar verringern (vgl. Verplaetse & Lambrechts 2019, S. 21). Mögliche Einschränkungen der Arbeit mit CAT-Tools sollen jedoch im dritten Teil dieser Arbeit weiter ausgeführt werden.

Das Fenster *Wörterbuch* enthält mögliche Vorschläge aus dem Projektordner *dictionaries*, in dem heruntergeladene Wörterbuchteile gespeichert werden können (vgl. OmegaT Handbuch, S. 54). Ein weiteres Fenster enthält Vorschläge aus maschinellen Übersetzungssystemen. Wenn diese Funktion aktiviert ist, kann durch eine Schnittstelle zu verschiedenen MT-Systemen wie *DeepL* oder *GoogleTranslate* ein Übersetzungsvorschlag automatisch generiert werden. Beide dieser Optionen können auch zur Zeit- und Kostenersparnis und damit zur Produktivitätssteigerung im Übersetzungsprozess beitragen. Während Wörterbücher auch auf bestimmte Fachgebiete spezialisiert sein können, ist bei den über OmegaT abrufbaren maschinellen Übersetzungssystemen wie *GoogleTranslate* jedoch zu beachten, dass diese Tools nicht ausschließlich mit einem fachsprachlichen Datenset trainiert wurden und deshalb – je nach Übersetzungsauftrag – nicht immer geeignete Vorschläge liefern. Dies zeigt auch die Untersuchung von Farahsani, Rini und Jaya (2020) zur Nutzung von *GoogleTranslate* bei der Übersetzung von Texten aus dem Fachbereich Maschinenbau (vgl. S. 433).

Das Fenster *Mehrfachübersetzungen* enthält mögliche Übersetzungsvarianten für Segmente, die bereits bearbeitet wurden. Wenn unterschiedliche Übersetzungen für ein Ausgangssprachliches Segment bereits verwendet wurden, werden die einzelnen Alternativen in diesem Fenster angezeigt (vgl. OmegaT Handbuch, S. 40). In Übersetzungsforen sind sich professionelle Übersetzerinnen und Übersetzer uneinig über die Notwendigkeit dieser Funktion. Ein Beitrag eines Übersetzers auf die Frage, ob er die Funktion für Translationsarbeiten aktiviert hat, enthält ein Beispiel, für das die Funktion seiner Ansicht nach hilfreich ist:

Yes, I do. An example that I give because several translations should be allowed for a unique segment is the following one: In German the word *Bestand* may mean balance, stock or inventory according to the business area to which it belongs and has quite different translation when translating into spanish (*saldo, existencias, inventario, etc.*). However, being eclectic, I believe that they are as many arguments to allow as not to allow multiple translations ;-) (Bouvier 2009).

Die Möglichkeit, verschiedene Übersetzungsalternativen für ein Segment zu hinterlegen, ist eine Hilfestellung, um mit unterschiedlichen, kontextspezifischen Übersetzungsvarianten eines Segment umgehen zu können. Wichtig ist dabei allerdings, dass unterschiedliche Kontexte auch erkannt werden – auch wenn es sich dabei um einzelne Segmente handelt, die durch die Darstellung im Editor vom restlichen Kontext abgehoben werden. Welchen Einfluss die Segmentierung sowie die Darstellung einzelner Segmente auf das Kontextverständnis der Übersetzenden hat, soll im weiteren Verlauf der Arbeit diskutiert werden.

Zuletzt gibt es noch zwei weitere Fenster in OmegaT, die für Notizen und Kommentare eingeblendet werden können. Hier können entweder bei der Bearbeitung kurze Notizen zu einem Segment ergänzen werden, oder es können beispielsweise durch die Übersetzungsagentur vor der eigentlichen Translationsarbeit Kommentare zu einzelnen Textstellen hinterlassen werden (OmegaT Handbuch, S. 41).

3.3.3 Übersetzungsabschluss

Jede abgeschlossene Übersetzung eines einzelnen Segments fließt automatisch in das TM im Projektordner ein und kann im folgenden Segment sofort als Vorschlag wiederverwendet werden (vgl. Keller 2011, S. 60). Hat man eine Datei innerhalb eines Projektes übersetzt, folgt im Editor-Fenster sofort das nächste Segment aus dem nachfolgenden Text. Wie diese Art der Textaufbereitung die Übersetzungsarbeit und die Qualität der Übersetzung beeinflusst, soll im dritten Teil dieser Arbeit näher beleuchtet werden.

Vor der Abgabe eines fertiggestellten Auftrags ist es in den meisten CAT-Tools möglich, Qualitätsprüfungen durchzuführen. Je nach CAT-Tool können unterschiedliche Dinge geprüft werden. In OmegaT kann lediglich eine Rechtschreibprüfung sowie eine automatische Prüfung der Terminologie und Tags durchgeführt werden. Andere CAT-Tools zeichnen sich, wie im Kapitel zum Aufbau von CAT-Tools bereits erläutert, durch deutlich ausgeprägtere Prüffunktionen aus (vgl. Petrova 2019, S. 92). Im Rahmen dieser automatischen Qualitätsprüfung können Fehler gefunden und korrigiert werden, bevor die Übersetzung an die Auftraggeber geht.

Nach Abschluss eines Dokumentes wird dieses automatisch im Projektordner *target* im Originalformat gespeichert und kann dort einzeln eingesehen werden. Je nach Dateiformat sind aufgrund der Konvertierung vereinzelt Änderungen vorzunehmen (vgl. Keller 2011, S. 60). Wurde ein Projekt nicht innerhalb einer Sitzung abgeschlossen, wird in der nächsten Sitzung genau an der zuletzt bearbeiteten Stelle fortgesetzt. Die Suche nach dem zuletzt bearbeiteten Segment entfällt dadurch.

In den vorherigen Ausführungen wurde bereits angedeutet, dass die Arbeit mit CAT-Tools für professionelle Übersetzerinnen und Übersetzer nicht nur Produktivitätssteigerung und Optimierung bedeutet, sondern sich durchaus auch nachteilig auf den Übersetzungsprozess sowie die Qualität der Übersetzung auswirken kann. Im Folgenden sollen Korpora als weitere Hilfsmittel für die computergestützte Übersetzungsarbeit vorgestellt werden, bevor dann im dritten Teil dieser Arbeit erläutert wird, welche Einschränkungen in OmegaT erkennbar sind und welche Vorteile die Ergänzung des Tools durch Korpora und korpuslinguistische Methoden mitbringen würde.

4. Teil 2 – Die Nutzung von Korpora im Übersetzungsprozess

4.1 Begriffsbestimmung

Die im ersten Teil dieser Arbeit vorgestellten CAT-Werkzeuge sind speziell auf die Übersetzungsarbeit ausgerichtet. Darin enthaltene Funktionen und Module dienen dazu, die Übersetzungsarbeit schneller und effizienter zu machen, indem sie Übersetzende gezielt besonders in repetitiven und aufwändigen Arbeiten unterstützen. Korpora sind Bestandteil computergestützter Werkzeuge, die ein breiteres Anwendungsfeld umfassen. Die Datenbanken enthalten digitalisierte Texte oder sprachliche Äußerungen, die maschinenlesbar und typischerweise durch verschiedene Metadaten wie Angaben zur Quelle oder Annotationen beispielsweise zu Wortarten angereichert sind (vgl. Lemnitzer & Zinsmeister 2015, S. 13). In korpuslinguistischen Untersuchungen bilden Korpora eine Datengrundlage für die Analyse von Sprache im Gebrauch (vgl. Perkuhn, Keibel & Kupietz 2012, S. 14) und können Daten in verschiedenen Sprachen, aus verschiedenen Zeiträumen oder zu einer Vielzahl an Themenfeldern umfassen. Doch auch über die wissenschaftliche Forschung hinaus werden Korpora genutzt – beispielsweise bei der Erstellung von Wörterbüchern, im Fremdsprachenunterricht oder als Werkzeuge für die Übersetzungsarbeit (vgl. Lemnitzer & Zinsmeister 2015, S. 34).

Man unterscheidet dabei monolinguale Korpora, die wie das in dieser Arbeit verwendete Korpus DeReKo im Kernbestand nur Texte einer Sprache enthalten, von multilingualen Korpora. Bei den mehrsprachigen Korpora gibt es weitere Differenzierungen: Als Vergleichskorpus werden zumeist Textsammlungen bezeichnet, die Texte mehrerer Sprachen enthalten. Diese Texte stammen zwar inhaltlich aus dem gleichen Diskursfeld, sind aber keine direkten Übersetzungen voneinander (vgl. Lemnitzer & Zinsmeister 2015, S. 138; Mosavi Miangah 2006, S. 3). In der Fachliteratur wird der Begriff *Vergleichskorpus* jedoch unterschiedlich verwendet. Biel (2018) definiert ein Vergleichskorpus als Textsammlung, die nur einsprachige Daten enthält – sowohl Texte, die in die entsprechende Sprache übersetzt wurden, als auch Texte, die schon im Original in derselben Sprache verfasst wurden (vgl. S. 26). Fantiuoli und Zanettin (2015) stellen zusammenfassend fest, dass die Unterscheidung

mehrsprachiger Korpora nicht immer „clear cut“ (S. 3) ist und je nach Standpunkt unterschiedliche Bezeichnungen verwendet werden.

Eine andere Art multilingualer Sprachdatenbanken sind Parallelkorpora. Darunter werden in dieser Arbeit Datenbanken verstanden, in denen Sprachdaten und deren Übersetzung in eine oder mehrere Sprachen gespeichert sind (vgl. Lemnitzer & Zinsmeister 2015, S. 8; Mosavi Miangah 2006, S. 3). Schriftsprachliche Texte liegen in Parallelkorpora beispielsweise auf Satzebene aligniert vor – das heißt, sie sind einander zugeordnet (vgl. Lemnitzer & Zinsmeister 2015, S. 138). In der Übersetzungspraxis kann ein Parallelkorpus unter anderem genutzt werden, um nach der Übersetzung für einen Begriff zu suchen, oder es kann aufzeigen, an welchen Stellen ein direktes Übersetzungsäquivalent nicht vorhanden ist (vgl. Mosavi Miangah 2006, S. 3).

Besonders mehrsprachige Korpora, in denen alignierte Sprachdaten gesucht und so für die Übersetzung genutzt werden können, stellen eine hilfreiche Ressource für die Translationsarbeit dar. Sie bieten den Übersetzenden „a repertoire of translation strategies past translators have resorted to when confronted with similar problems to the ones that have prompted a search in a parallel corpus“ (Zanettin 2002, S. 11). Doch oft sind mehrsprachige Korpora nicht in ausreichendem Maße verfügbar oder umfassen nur wenig Sprachdaten, da sie im Vergleich zu monolingualen Korpora deutlich mehr Aufbereitung erfordern (vgl. Bowker 1998, S. 3; Zanettin 2002, S. 13). Außerdem zeichnen sich Parallelkorpora durch verminderte Sprachqualität der darin enthaltenen Daten aus, da Übersetzungen Besonderheiten aufweisen können, die nicht dem originalsprachlichen Gebrauch entsprechen (vgl. Trawiński & Kupietz 2021, S. 217). So orientieren sich Übersetzungen beispielsweise oft stärker an der Ausgangssprache als an der Zielsprache – ein empirisch belegtes Phänomen, das auch als *shining through* bezeichnet wird (vgl. Trawiński & Kupietz 2021, S. 217). Für die Übersetzungsarbeit ist es wichtig, ausreichend Sprachdaten zu haben, die – wie in den folgenden Ausführungen zum Korpus DeReKo an einem Beispiel gezeigt werden soll – möglichst der gleichen Fachdomäne entstammen wie der zu übersetzende Text. Nicht immer ist es jedoch möglich, auf ein passendes mehrsprachiges Korpus zurückzugreifen. Aus diesem Grund wird in der folgenden Analyse ausschließlich ein einsprachiges Korpus verwendet. Im nächsten Kapitel soll erläutert

werden, welche Möglichkeiten monolinguale Korpora wie DeReKo für den Einsatz in der Translationsarbeit bieten.

4.2 Zum Einsatz monolingualer Korpora in der Translationsarbeit

Da monolinguale Korpora nur Texte in einer Sprache enthalten, kann darin nicht nach Übersetzungen eines Suchausdrucks in andere Sprachen gesucht werden. Dennoch eignen sich auch einsprachige Korpora für den Einsatz im Übersetzungsprozess, denn sie „zeichnen sich durch eine sehr hohe und kontrollierte Sprachqualität aus, da sie typischerweise ausschließlich Originaltexte enthalten und damit den muttersprachlichen Sprachgebrauch reflektieren“ (Kupietz & Trawiński 2022, S. 418). Übersetzte Texte – und damit auch Sprachdaten in Parallelkorpora – weisen Tendenzen für beispielsweise einfachere Sprache, relativ hohe Anteile hochfrequenter Wörter gegenüber niedrigfrequenten Wörtern oder eine übermäßige Orientierung an Konstruktionen der Ausgangssprache auf, die so nicht in originalsprachlichen Texten vorkommen (vgl. Kupietz & Trawiński 2022, S. 422). Monolinguale Korpora hingegen bieten einen Einblick in den tatsächlichen Sprachgebrauch und stellen so sicher, dass Übersetzerinnen und Übersetzer ausschließlich mit Texten konfrontiert sind, die eine gute Basis für die eigene Übersetzung darstellen, sofern das Korpus entsprechend konzipiert ist. Auf die richtige Auswahl der Datengrundlage für einen bestimmten Übersetzungsauftrag soll jedoch in Kapitel 4.4.1 am Beispiel des Korpus DeReKo nochmal eingegangen werden.

In der Übersetzungspraxis können monolinguale Korpora entweder in der Ausgangssprache des zu übersetzenden Segments genutzt werden oder in der Zielsprache, in die ein Text übersetzt werden soll. Steht für das Sprachpaar sowohl ein passendes zielsprachliches als auch ein ausgangssprachliches Korpus zur Verfügung, ist es sinnvoll, beide Korpora für die Translationsarbeit zu nutzen. Die Sprachdatenbanken können dann genutzt werden, um die exakte Bedeutung eines Wortes oder Bedeutungsvarianten in der Ausgangs- und Zielsprache zu ermitteln, um die Entscheidung für ein passendes Übersetzungsäquivalent zu treffen, um über Angaben zur Häufigkeit eines Wortes einen von mehreren zielsprachlichen Begriffen auszuwählen oder um zu bestätigen, dass ein Begriff in der Zielsprache tatsächlich existiert (vgl. Bowker 1998, S. 18). Weitere Arbeiten zeigen zusätzlich, dass einsprachige Textdatenbanken Aufschluss über typische Wortkombinationen bereitstellen können und den Übersetzenden so Hinweise zum Sprachgebrauch liefern (vgl. Mosavi Mian-

gah 2006, S. 5). Außerdem ist es für die Übersetzung eines Textes unbedingt erforderlich, den ausgangssprachlichen Text umfassend zu verstehen, damit die Gesamtbedeutung auch in der Übersetzung richtig übernommen wird (vgl. Bowker 1998, S. 2). Monolinguale Korpora stellen Hintergrundwissen zur Domäne des zu übersetzenden Textes bereit und können so als Recherche- und Informationsquelle dienen.

Dies sind nur einige in der Forschungsliteratur diskutierte Anwendungsfelder von monolingualen Korpora in der Übersetzungspraxis. Nachdem im nachfolgenden Kapitel zunächst Korpora von den Textdatenbanken in CAT-Tools abgegrenzt werden, werden im Anschluss am Beispiel des Korpus DeReKo weitere korpuslinguistische Methoden vorgestellt, die in der Translationsarbeit eingesetzt werden können.

4.3 Abgrenzung von Korpora zu TMs

Sowohl die im Rahmen der Beschreibung von CAT-Tools vorgestellten TMs als auch Parallelkorpora sind durchsuchbare Sprachdatenbanken, die einander zugeordnete Übersetzungen in mehreren Sprachen enthalten. Ein wichtiger Unterschied zwischen den beiden Textsammlungen besteht jedoch in der Zusammenstellung: Während TMs während des Übersetzungsprozesses mit jedem abgeschlossenen Segment inkrementell wachsen und damit zu einem großen Teil eigene Übersetzungen enthalten, werden Korpora in der Regel nicht mit eigenständig produzierten Textdaten zusammengestellt, sondern mit Übersetzungen aus Zeitungen, Büchern oder anderen Quellen (vgl. z.B. DeReKo, Kapitel *Archiv*). Wie Zanettin (2002) schreibt, hat dies zur Folge, dass TMs in der Regel Sprachdaten enthalten, die stark auf vorherigen Projekten basieren und deshalb besonders dann von Nutzen sind, wenn ein bereits bearbeitetes Übersetzungsprojekt weitergeführt werden soll (vgl. S. 11). Korpora hingegen basieren nicht auf vorherigen Arbeiten und können – je nach Bedarf – sehr spezifische oder allgemeinsprachliche Texte enthalten. Nutzen Übersetzende ein TM, das beispielsweise projektübergreifend zusammengestellt wird oder greifen sie auf online verfügbare Ressourcen wie das in dieser Arbeit verwendete DGT-TM zurück, werden für die Übersetzungsvorschläge Ergebnisse der Übersetzung Dritter zusammen mit den eigenen Übersetzungen aus den zuvor übersetzten Segmenten verwendet.

Obwohl in beiden Datenbanken eine Zuordnung der Übersetzung zum Originaltext erfolgt, ist ein weiterer zentraler Unterschied, dass die Segmente in TMs endgültig vom restlichen Text getrennt und als einzelne *Translation Units* gespeichert werden (vgl. Verplaetse & Lambrechts 2019, S. 22). Die Suche nach *Matches* kann demnach auch nur Segmentpaare ausgeben, während in Korpora alignierte Textteile häufig weiterhin im satzübergreifenden Kontext abgerufen werden können. Wie sich dies auf die Übersetzung auswirken kann, soll im dritten Teil dieser Arbeit näher beleuchtet werden. Weiterhin unterscheiden sich Korpora und CAT-Tools in der Bereitstellung von gefundenen Textausschnitten, denn in CAT-Tools werden Übersetzungsvorschläge aus dem TM so geordnet, dass das *Match* mit der größten formalen Übereinstimmung an oberster Stelle steht und per Tastenkombination direkt in das geöffnete Segment übernommen werden kann. Die Anordnung der gefundenen Texteinhei-

ten, die sich laut Verplaetse und Lambrechts (2019) auf den Entscheidungsprozess der Übersetzenden auswirkt (vgl. S. 22), kann in Korpora häufig variiert werden und Texte können beispielsweise zufällig oder chronologisch sortiert dargestellt werden.

Ein weiterer Unterschied besteht darin, dass Sprachdaten in Korpora anders aufbereitet und für die Untersuchung zur Verfügung gestellt werden als Texte in CAT-Tools. Beispielsweise sind die gespeicherten Segmente in TMs nicht durch eine Vielfalt an Metadaten angereichert wie es in Korpora häufig der Fall ist. Dies hat Auswirkungen auf die Suche in den beiden Datenbanken, denn in TMs kann beispielsweise nicht nach Texten aus einem bestimmten Entstehungszeitraum gefiltert werden, während dies in DeReKo über die Auswahl von Subkorpora oder Ergebnisfilter möglich ist. Dies ist nur einer von vielen Unterschieden bezüglich der Textaufbereitung und Suchfunktionen in den beiden Tools. Am Beispiel des Korpus DeReKo sollen im folgenden Kapitel einige der korpuslinguistischen Funktionen und Methoden näher vorgestellt werden.

4.4 Das *Deutsche Referenzkorpus (DeReKo)*

Das *Deutsche Referenzkorpus (DeReKo)* ist eine schriftsprachliche Textdatenbank, die am Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim beheimatet ist. Das Korpus dient verschiedenen wissenschaftlichen Fragestellungen und Anwendungsbereichen (vgl. DeReKo, Kapitel *Einsatzgebiete*), doch im Folgenden soll besonders die Anwendung in der Übersetzungsarbeit fokussiert werden. Dazu werden zunächst einige ausgewählte Funktionen des Korpus vorgestellt, um dann im dritten Teil dieser Arbeit zu diskutieren, welches Potential einige der korpusanalytischen Methoden und Werkzeuge für die Translationsarbeit ergänzend zu CAT-Tools darstellen.

Bereits 1964 am IDS in Mannheim begonnen (vgl. Längen 2017, S. 161), ist die Datensammlung DeReKo mit 55 Milliarden Wörtern (Stand 08.03.2023) mittlerweile „die weltweit größte linguistisch motivierte Sammlung elektronischer Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit“ (DeReKo). DeReKo enthält verschiedene Textarten wie beispielsweise regionale und überregionale Zeitungen, belletristische Werke, oder wissenschaftliche und populärwissenschaftliche Texte (vgl. DeReKo). So wird versucht, „den deutschen Schriftsprachgebrauch zeitbegleitend möglichst repräsentativ abzubilden [und] die Verwendung der deutschen Sprache und ihre Entwicklung zu dokumentieren“ (Längen 2017, S. 161). Wie Perkuhn, Keibel und Kupietz (2012) schreiben, sind „Grundgesamtheiten wie Sprachen oder Sprachdomänen ebenso wie ihre wesentlichen Charakteristika und deren Proportionen erstens nicht allgemeingültig und zweitens auch sonst nicht strikt definierbar“ (S. 47). Repräsentative Stichproben einer Sprache sind deshalb nicht universell erreichbar, weshalb je nach Forschungsfrage oder Einsatzbereich ein unterschiedlicher Ausschnitt an Daten relevant sein kann. Aus diesem Grund ist DeReKo „als eine Art Ur-Stichprobe des Sprachgebrauchs konzipiert, aus der sich jeder Nutzer selbst ein virtuelles Korpus zusammensetzen“ (DeReKo, Kapitel *Einsatzgebiete*) kann oder ein vordefiniertes Korpus nutzen kann. Somit können bei der Komposition eines Korpus Charakteristika wie Textsorten oder Themen, die „für die Sprachdomäne intuitiv als relevant erachtet (bzw. die voraussichtlich Auswirkungen auf Befunde haben werden)“ (Perkuhn, Keibel &

Kupietz 2012, S. 47) vorrangig einfließen. Fast alle Daten in DeReKo sind öffentlich und kostenlos zugänglich (vgl. DeReKo, Kapitel *Verfügbarkeit*).

4.4.1 Auswahl von Subkorpora in DeReKo

Für die nachfolgende Analyse in dieser Arbeit werden nicht alle Textdaten aus DeReKo (Release 2019-II, vgl. DeReKo) verwendet, sondern es wurde das Subkorpus *pp – Plenarprotokolle* aus dem Archiv *W – Archiv der geschriebenen Sprache* in DeReKo ausgewählt. Darin enthalten finden sich „Plenarprotokolle aller deutschen Parlamente, d.h. von Bundestag, Bundesrat sowie aller deutschen Landtage von (mindestens) 2000 bis Mitte 2012“ sowie insgesamt 220 Texte aus dem Landtag von Niederösterreich (vgl. Archiv, Kapitel *Beschreibung*). Die Auswahl des gewünschten Untersuchungskorpus erfolgt über eine Schnittstelle, über die auf die Datensammlung zugegriffen werden kann. Eines dieser Werkzeuge zur Korpusanalyse, das den Zugriff auf die Daten in DeReKo erlaubt und die Sprachdaten durchsuchbar macht, ist das *Corpus Search, Management and Analysis System (COSMAS II)*. Über die Webversion dieser Schnittstelle kann nach der Anmeldung und Korpusfestlegung die ausgewählte Datenmenge durchsucht werden. Neben den vordefinierten Subkorpora können in COSMAS II auch eigene Korpora aus verschiedenen Textdaten zusammengestellt werden. Diese Auswahlmöglichkeiten erlauben es, entweder sehr große, eher breit gefächerte Textbestände zu durchsuchen wie beispielsweise im Korpus *W-öffentlich*, oder es gibt die Möglichkeit, kleinere Ausschnitte auszuwählen und sich so beispielsweise auf eine konkrete Fachdomäne zu beschränken.

In der Translationsarbeit kann dies hilfreich sein, da über domänenspezifische Korpora auf vorrangig für diesen Kontext übliche Textbeispiele zugegriffen werden kann. Anhand einer Suche nach der Wortform *Bund* lässt sich zeigen, wie sich kontextspezifische Bedeutungen in verschiedenen Korpora wiederfinden: Nutzt man die Wortform *Bund* in einer politischen Rede ist mit großer Wahrscheinlichkeit ein Zusammenschluss verschiedener Akteure gemeint – beispielsweise der Bund Deutschland, der die einzelnen Bundesländer vereint (vgl. Duden, Vereinigung). Spricht man jedoch von Möhren oder Spargel ist damit gemeint, dass das Gemüse zu einem Bündel zusammengebunden wurde (vgl. Duden, Bündel). Ein erster Überblick über die Korpusbelege im allgemeinsprachlichen Korpus *W – öffentlich* zeigt, dass in den ersten 20 zufällig sortierten Treffern Belege für beide Bedeutungen vorhanden sind, wie in den Beispielen 1 a und b zu sehen ist. Beispiel 2 wurde aus dem domänenspezifi-

schen Korpus *pp – Plenarprotokolle* extrahiert, in dem bei einer ersten Durchsicht unter den ersten 20 zufällig sortierten Treffern nur die erste Bedeutung von *Bund* im Sinne einer Vereinigung vorkommt.

(1a) Bund, Länder und Kommunen müssen bis zum Jahr 2023 mit 124,3 Milliarden Euro weniger auskommen als angenommen. (aus: Hannover-sche Allgemeine, 10.05.2019; Weniger Steuern als geplant)

(1b) 4 Frühlingszwiebeln 1 Packung Krabbenchips 1/2 Bund Koriander Erdnussbutter, entkernte Chilis, Knoblauch, Limettensaft, Zucker, Gal-gant, Soja-und Fischsauce in einem Mixer glatt pürieren. Erdnüsse zuge-ben und nicht zu fein mixen, sodass ein gewisser Biss spürbar bleibt. (aus: NEWS, 24.09.2016; Vom Unterjubeln des Grünzeugs)

(2) Wenn es wirklich so kommen würde, dass Bremen und andere Bun-desländer und der Bund ihren Anteil bezahlen müssten, dann hätten wir eine Volkswirtschaft, die am Ende ist. (aus: Protokoll der Sitzung des Parlaments Bremische Bürgerschaft am 12.11.2008. 31. Sitzung der 17. Wahlperiode 2007-2011)

Übersetzt man *Bund* in eine andere Sprache, kann das jeweilige zielsprachliche Äquivalent je nach Kontext unterschiedlich sein. So würde das Lexem *Deutscher Bund* beispielsweise im Englischen mit *German Confederation* übersetzt, während man bei der Übersetzung von Beispiel 2 von einem *bunch of coriander* sprechen würde (vgl. dict.cc, Bund). In der Translationsarbeit ist es deshalb wichtig, die passende kontextspezifische Bedeutung zu erkennen, um die richtige Übersetzung aus-wählen zu können. Dazu trägt einerseits die Darstellung des Suchworts in einem grö-ßeren Kontext bei, worauf im folgenden Kapitel näher eingegangen werden soll, aber auch die Eingrenzung eines Subkorpus unterstützt Übersetzende dabei, bei ihren Re-cherchen fachsprachliche Bedeutungsunterschiede zu erkennen und bei der Überset-zung zu beachten. Wie Mosavi Miangah (2006) demonstriert, sind monolinguale Korpora in der Translationsarbeit besonders hilfreich, um Übersetzungsvorschläge aus anderen Systemen und Tools – beispielsweise über TMs abgerufene *Fuzzy Mat-*

ches – zu verifizieren oder abzulehnen (vgl. S. 7). Dazu ist es jedoch, wie das obige Beispiel zeigt, wichtig, auch die richtige Datengrundlage für die Korpusrecherche zu wählen.

Zu beachten ist an dieser Stelle, dass für eine erste Übersicht über die Belegtreffer zu *Bund* nur ein kleiner Teil der Datenmenge in den beiden Korpora betrachtet wurde. Je mehr Sprachdaten ein Korpus enthält und je ungleicher die beiden Lesarten vertreten sind, desto wahrscheinlicher ist es auch, dass die 20 betrachteten Belege trotz zufälliger Sortierung nicht beide Bedeutungsweisen enthalten, obwohl sie eigentlich im Korpus vorhanden wären. Eine Auswertung dieser begrenzten Zufallsstichprobe deutet jedoch bereits an, dass die Wortform *Bund* im Sinne eines Bündels im allgemeinsprachlichen Korpus häufiger vorkommt als im domänenspezifischen Korpus. Diese Annahme soll in Kapitel 4.4.3 zusätzlich durch eine Kookkurrenzanalyse bestätigt werden.

4.4.2 Suche und Anzeige von Korpusbelegen in COSMAS II

Sucht man in einem ausgewählten Korpus nach einem bestimmten schriftsprachlichen Ausdruck, können zunächst verschiedene Operatoren verwendet werden, um die Suche je nach Untersuchungsfrage einzugrenzen oder auszuweiten. Wortformoperatoren wie * ermöglichen es durch den Suchausdruck *Bund** auch Treffer wie *Bundestag* oder *Bundesländer* zu finden, denn er steht für keines oder unendlich viele Zeichen, die dem Suchausdruck folgen können. Abstandsoperatoren wie */+w1* erlauben es, beispielsweise nach Wortverbindungen zu suchen, in denen die zweite Wortform im Suchausdruck direkt auf die erste Wortform folgt. Ist die Suchanfrage eingegeben und bestätigt, wird man zu den Wortformlisten weitergeleitet, in denen sich gefundene Wortformen einzeln an- oder abwählen lassen. So kann die Ergebnismenge weiter eingegrenzt werden. Anschließend gelangt man zum Reiter „Ergebnisse“, der zwar noch keine konkreten Textausschnitte präsentiert, aber eine Übersicht zu gefundenen, nach Metadaten sortierten Textausschnitten bereithält. Hier kann nach Quellen, Textsorten und vielem mehr gefiltert werden.

Die gefundenen Korpusbelege können in COSMAS II pro Rubrik auf verschiedene Weise dargestellt werden: Die *Key-Word-In-Context* (KWIC)-Ansicht zeigt die gesuchte Wortform farblich markiert und eingebettet in begrenzten Kontextumfang. Die Treffer lassen sich beispielsweise chronologisch sortieren, sind nummeriert und mit einer Sigle, welche auf die Quelle und das Entstehungsjahr hinweist, versehen. Einzelne Belege lassen sich durch Anklicken markieren und erscheinen dann oben im Feld *Aktive Treffer*. Abbildung 6 zeigt einen Ausschnitt der zufällig sortierten KWIC-Ansicht für die Suche nach *Bund* im Korpus *pp – Plenarprotokolle*.

Archive Korpus Such. Wortform. Ergebnisse Kook. **KWIC** Volltext Export

Sortierung: KWIC (zufällig sortiert) zu Treffer: 0 springen

Seite 1 von 472

1	PHE/W18.00116	...er neuen Vertragstarife ausgehandelt worden war. Der Bund Deutscher Karneval hat derzeit 5.000 Vereine mi...
2	PHB/W17.00031	...ürde, dass Bremen und andere Bundesländer und der Bund ihren Anteil bezahlen müssten, dann hätten wir ...
3	PBW/W13.00050	...Äußerung des Bundesrechnungshofs war, wonach der Bund dazu verpflichtet war! Sie sollten die ganze Wahr...
4	PNI/W15.00098	...r. 7 gefassten Beschluss ausdrücklich unterstützt. Der Bund hat einen Prüfauftrag bezüglich der Einführung e...
5	PST/W05.00075	...Verfahren, das auf Konsens angelegt ist, bewegen der Bund und die Länder zurzeit mehr als 13 Milliarden Eu...
6	PNO/W18.00002	... am Entwurf der Rechnungswesenverordnung für den Bund . Als Sonderaufgaben gelten die Prüfungen der T...
7	PSL/W14.00009	...der Vergleich gezogen werden. Da zeigt sich, dass der Bund mit über 80 Milliarden dabei ist - Hessen mal vier...
8	PSN/W04.00073	... hat. Gerade im Hinblick auf die Gesetzesaktivitäten im Bund zu einem neuen Beamtenrechtsstatusgesetz sa...
9	PNW/W12.00044	... Wenn man es auf den Punkt bringt, heißt das: Der (d) Bund spielt die Rolle des Anspruchstellers, und diese ...
10	PNO/W16.00027	...ehrs in das Buskonzept gesichert. Denn dafür müssten Bund , Land und Landeshauptstadt ihre Beiträge gerin...
11	PBY/W14.00121	...zurichten, wie das in allen anderen Ländern und beim Bund der Fall ist, sodass letztlich unser bayerisches M...
12	PTH/W04.00095	...langt engagiertes, politisches Handeln international, im Bund , aber auch im Land Thüringen. DIE LINKE wird ...
13	PRP/W15.00004	...Da ist für die Länder nichts mehr, sondern das füllt der Bund aus ", - ich habe daran Zweifel. Unsere Verfassu...
14	PNW/W13.00079	...en, wenn nicht die Regierung? Das heißt: Rot-Grün im Bund und Rot-Grün im Land ist eine Gleichung: Steue...
15	PMV/W06.00003	...en Rehberg'schen Worten von 2001 zu sprechen: Der Bund wird "nicht einen Handschlag rühren ". Auch der ...
16	PBT/W14.00252	...Sie die Flutkatastrophe an, die viel Geld kostet, für den Bund in den nächsten zwei Jahren möglicherweise 5 ...
17	PMV/W03.00039	...om Bund zu Lasten der Länder gehandelt worden. Der Bund hat das Insolvenzverfahren, darunter die Verbra...
18	PBB/W05.00030	...ische Abend, den die Landesverbände von NABU und BUND veranstalten, statt. Dazu sind alle Anwesenden ...
19	PBB/W05.00024	...hen im Sommer dieses Jahres, als die Länder und der Bund versucht haben, gemeinsame Lösungen zu finde...
20	PBT/W15.00135	...n. In einem Bundesstaat mit 16 Einzelstaaten und dem Bund ist das aber doch eine logische Konsequenz, ins...

Abb. 6: Zufällig sortierte KWIC-Ansicht zum Suchausdruck *Bund* in COSMAS II.

Die Volltextansicht im nächsten Reiter zeigt einen größeren Textausschnitt zu den gefundenen Belegen. Aus datenschutzrechtlichen Gründen (vgl. dazu Lünen 2017, S. 165) sind hier keine vollständigen Zeitungsartikel und Texte zu finden, sondern lediglich ein Ausschnitt von bis zu drei Absätzen vor und nach dem Suchbegriff, der aus dem Korpus gezeigt werden kann. Dennoch wird über die erweiterte Kontextdarstellung zusammen mit der Quellenangabe erkenntlich, welchem größeren Kontext ein Belegtreffer entstammt. So sieht man in Abbildung 7, dass der markierte Treffer, der im Korpus *W – öffentlich* bei der Suche nach *Bund* gefunden wurde, aus einem Artikel mit dem Titel „Vom Unterjubeln des Grünzeugs“ entstammt.

	<p>L21/MAR.02590 Berliner Morgenpost, 25.03.2021, S. 3; " Dafür bitte ich um Verzeihung" Welche Regelung nun stattdessen gelten soll, blieb am Mittwoch zunächst unklar. Das Gleiche gilt für Ansammlungen im öffentlichen Raum, die ebenfalls in Punkt vier " grundsätzlich untersagt" wurden. Auch dieser Passus ist nun unwirksam, eine Klärung steht aus. Gestrichen ist auch die " Bitte" von Bund und Ländern, religiöse Veranstaltungen nur virtuell stattfinden zu lassen. Gottesdienste zu Ostern können somit unter Hygieneauflagen stattfinden, sofern es die jeweiligen Länder erlauben. Führende Intensivmediziner äußerten sich enttäuscht über die Rücknahme des geplanten harten Lockdowns zu Ostern. Da aktuell ein exponentielles Wachstum bei den Intensivpatienten zu sehen sei, " hätte die Osterpause sicherlich wieder einige Infektionen verhindern können, die jetzt unvermeidbar stattfinden werden" , sagte der Präsident der Deutschen Interdisziplinären Vereinigung für Intensiv- und Notfallmedizin (DIVI). Gernot Marx.</p>
15	
✓	<p>NEW16/SEP.00344 NEWS, 24.09.2016, S. 84,85; Vom Unterjubeln des Grünzeugs 4 Frühlingszwiebeln 1 Packung Krabbenchips 1/2 Bund Koriander Erdnussbutter, entkernte Chilis, Knoblauch, Limettensaft, Zucker, Galgant, Soja- und Fischeis in einem Mixer glatt pürieren. Erdnüsse zugeben und nicht zu fein mixen, sodass ein gewisser Biss spürbar bleibt. Gemüse waschen. Die Gurke in dicke Scheiben schneiden, die Karotten schälen und in Stifte schneiden, den Broccoli in Röschen zerteilen, kurz blanchieren und abschrecken.</p>
16	
	<p>SOZ07/JUN.01342 Die Südostschweiz, 07.06.2007; In Bundesbern setzt man auf die Organisatoren Dieses Vertrauen ist bei Stahl, aber auch bei Ratskollege Heiner Studer (EVP, Aargau) in hohem Masse vorhanden. Auch Studer sieht nun vornehmlich nicht mehr die Bundespolitik, sondern die Host Citys und Organisationskomitees am Zuge. Und da seien «sehr gute Leute» am Werk, sagt Studer, der sich 2004 noch Heinz Keller (den früheren Direktor des Bundesamts für Sport) als Bundesratsdelegierten für die Euro wünschte, Benedikt Weibel jetzt aber für eine exzellente Wahl hält. Für Nationalrat Jo Lang (Alternative, Zug) trübt ein Parlamentsentscheid vom März die Vorfreude allerdings gewaltig. Lang sieht es mit grossem Unbehagen, dass der Bund an der Euro die Armee einzusetzen gedenkt. Für ihn steht fest: «Der Bundesrat muss garantieren, dass es nicht zu Zusammenstößen zwischen Soldaten und Fans kommen wird.»</p>
17	

Abb. 7: Zufällig sortierte Volltext-Ansicht zum Suchausdruck *Bund* im Korpus *W – öffentlich in COSMAS II*.

Die Darstellung einer Wortform in einem größeren Kontext kann, wie am Beispiel der Wortform *Bund* bereits gezeigt wurde, Übersetzenden helfen, das richtige ziel-sprachliche Übersetzungsäquivalent zu wählen. Gleichzeitig bietet die KWIC-Übersicht eine gute Möglichkeit, einen schnellen Überblick über eine größere Menge an Belegtreffern zu erhalten. Im Folgenden soll die Kookkurrenzanalyse als weitere korpuslinguistische Methode vorgestellt werden, die einen Überblick über die Belegtreffer bieten kann.

4.4.3 Kookkurrenzanalyse

Die Kookkurrenzanalyse ist eine korpuslinguistische Methode, bei der das sprachliche Umfeld eines Lexems oder einer Wortform in den Blick genommen wird. Es wird untersucht, welche Einheiten im Verhältnis zum Gesamtvorkommen im Korpus – also in diesem Fall das gewählte Subkorpus *pp – Plenarprotokolle* – häufiger im Kontext des Untersuchungsobjektes vorkommen und deshalb als auffällig bewertet werden (vgl. Perkuhn, Keibel & Kupietz 2012, S. 113). Die Gesamtheit aller Ergebnisse einer solchen Kookkurrenzanalyse wird als Kookkurrenzprofil (vgl. Belica 2011, S. 162) bezeichnet und gibt Auskunft über die Bedeutung des Untersuchungsobjektes. Eine Analyse dieses Profils ermöglicht eine „detaillierte Auskunft über die syntagmatische und paradigmatische Einbettung des Objekts im Sprachgebrauch“ (Belica 2011, S. 162) und damit einen Einblick in die semantische Ausprägung der untersuchten Einheit. Als primäre Partnerwörter werden dabei Wortformen bezeichnet, die im definierten Trefferkontext für die Kookkurrenzanalyse statistisch auffällig sind, denn ihre „beobachtete Anzahl im Kontext [weicht] von der erwarteten ab[...]“ (Perkuhn, Keibel & Kupietz 2012, S. 117). Dabei wird die Vorkommenshäufigkeit einer Wortform in der Gesamtheit des Korpus und die Vorkommenshäufigkeit im definierten Ausschnitt gegenübergestellt (vgl. Perkuhn, Keibel & Kupietz 2012, S. 117). Über das Korpusanalysesystem COSMAS II gibt es die Möglichkeit, eine solche Kookkurrenzanalyse mit den Sprachdaten aus DeReKo durchzuführen. Führt man die Analyse zum Suchausdruck *Bund* in den beiden zuvor schon gegenübergestellten Korpora *W – gesamt* und *pp – Plenarprotokolle* durch, wird auch über die Analyse des sprachlichen Umfelds bestätigt, dass nur im allgemeinsprachlichen Korpus beide Bedeutungsvarianten für *Bund* im Kookkurrenzprofil vorkommen:

Für das Beispiel wurde in beiden Korpora eine Kookkurrenzanalyse für denselben Suchausdruck durchgeführt. Die Analyseumfang wurde dabei auf einen Kontext von fünf Wörtern rechts und links eingestellt und begrenzt sich auf höchstens einen Satz. Dazu wurde der Autofokus aktiviert, um die besonderen Auffälligkeiten in der engeren Umgebung noch deutlicher anzuzeigen. Eine mittlere Granularität sowie eine normale Zuverlässigkeit wurden gewählt, um relevante Kookkurrenzpartner aber nicht zu detaillierte Profile zu erhalten. Funktionswörter wurden bei der Analyse

ausgeschlossen, da besonders inhaltliche Auffälligkeiten im Kookkurrenzprofil des Suchausdrucks von Interesse waren. Weitere Einstellungen umfassten eine eindeutige Clusterzuordnung, eine maximale Mehrgliedrigkeit von 4, ausgeschaltete Lemmatisierung der Wortformen im Umfeld des Bezugswortes sowie die Anzeige des LLR-Werts und die Nummerierung des Hauptkollokators.⁵

Vergleicht man das Kookkurrenzprofil aus dem Korpus *W – öffentlich* mit dem aus dem domänenspezifischen Korpus *pp – Plenarprotokolle* bestätigt sich die Annahme, die in einer kurzen Durchsicht der ersten 20 zufällig sortierten Belege im vorherigen Kapitel schon deutlich wurde. In der unmittelbaren Umgebung des Suchausdrucks *Bund* finden sich im allgemeinsprachlichen Korpus *W – öffentlich* auf den höchsten Rängen primäre Partnerwörter wie *Land, Umwelt, Gemeinde, Steuerzahler, Kommune, deutsch* oder *Kantone*. Dies deckt sich mit den Partnerwörtern aus der Kookkurrenzanalyse im Korpus *pp – Plenarprotokolle*. Auch dort werden die obersten Ränge durch primäre Partnerwörter wie *Land, Kommune, Gemeinde, Steuerzahler, Verantwortung, Finanzbeziehung* und *Mittel* belegt, die durch ihre syntagmatischen Muster und weitere Partnerwörter eindeutig der Bedeutung von *Bund* im Sinne eines Zusammenschlusses zugeordnet werden können. An keiner Stelle des Kookkurrenzprofils von *Bund* im Korpus *pp – Plenarprotokolle* findet sich ein primäres Partnerwort der anderen Bedeutungsvariante. Im allgemeinsprachlichen Korpus *W* jedoch findet sich auf Rang 42 das Partnerwort *Petersilie* mit sekundären Partnerwörtern wie *Schnittlauch, gehackt* oder *glatte*. Auf Rang 78 ist die Zahl *1* als primäres Partnerwort angeführt – immer in Verbindung mit Gewürzen und Kräutern wie *Schnittlauch, Thymian* oder *Basilikum*. Auf Rang 89 findet sich zudem das primäre Partnerwort *Schnittlauch*.

Auch wenn die Bedeutungsvariante von *Bund* im Sinne eines Zusammenschlusses in beiden Textdatenbanken dominant zu sein scheint⁶, zeigt der Vergleich der Profile in den beiden Korpora dennoch, dass *Bund* im politischen Kontext fast

5 Weitere Erläuterungen zur Definition und Auswirkung der einzelnen Parameter auf das Ergebnis finden sich in Perkuhn, Keibel und Kupietz (2012, S. 118ff.).

6 Eine Erkenntnis, die nicht verwunderlich ist, da auch im Korpus *W – gesamt* beispielsweise viele Zeitungstexte genauso wie die besagten Plenarprotokolle enthalten sind, die diese Lesart begünstigen (vgl. DeReKo, Kapitel *Archiv*). Einen deutlicheren Kontrast könnte man erreichen, wenn man zum Beispiel ein virtuelles Korpus aus Texten des Magazins *chk – Chefkoch* aus dem Archiv *W2* und ähnlichen Quellen erstellt. In diesem Beispiel sollte jedoch gezeigt werden, wie sich die Suchergebnisse in allgemeinsprachlichen Sprachdaten von den Ergebnissen mit Fokus auf eine spezielle Fachdomäne unterscheiden können.

ausschließlich im Sinne eines Zusammenschlusses verwendet wird. Bedeutungsvarianten in unterschiedlichen Korpora können also dargestellt werden, indem Kookkurrenzprofile erstellt werden und über die Analyse der unmittelbaren Umgebung eines Suchausdrucks statistisch signifikante Partnerwörter Aufschluss über kontextspezifische Verwendungsmuster geben. Sowohl wenn ins Deutsche übersetzt wird als auch wenn vom Deutschen in eine andere Zielsprache übersetzt wird, kann das Wissen über kontextspezifische Bedeutungsvarianten für die Translationsarbeit hilfreich sein. Syntagmatische Verwendungsmuster zeigen zudem, in welcher Reihenfolge der Suchausdruck zusammen mit den Partnerwörtern häufig vorkommt. Wie Möhrs (2016) bestätigt, sind diese syntagmatischen Muster bedeutungstiftende Einheiten (vgl. S. 33) und können deshalb insbesondere für zielsprachliche Recherchen in der Übersetzungsarbeit genutzt werden, um typische Vorkommen und Bedeutungskontexte zu ermitteln.

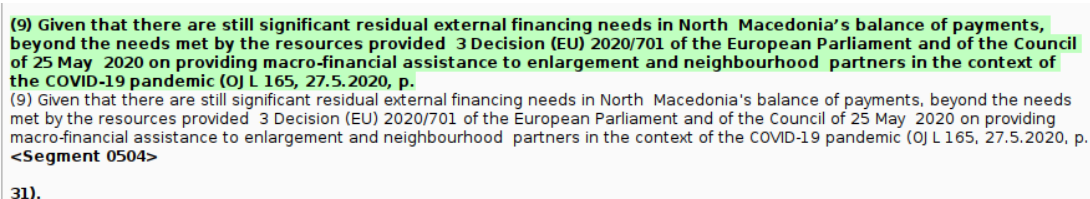
Nachdem nun einige Funktionen und Methoden vorgestellt wurden, die sich über die Korpusanalyseplattform COSMAS II mit den Sprachdaten aus DeReKo durchführen lassen, soll im dritten Teil dieser Arbeit diskutiert werden, wie korpuslinguistische Recherchen und Methoden in der Translationsarbeit ergänzend zu CAT-Tools eingesetzt werden können. Dazu werden im Folgenden zunächst einige Grenzen von CAT-Tools im Übersetzungsprozess herausgearbeitet und die Ergänzungsmöglichkeiten durch Korpora dargestellt.

5. Teil 3 – Ergänzung von CAT-Tools durch Korpora

5.1 Der richtige Textausschnitt für die Übersetzung

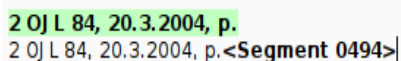
Bereits bei der Beschreibung von OmegaT und dessen Nutzen für die Translationsarbeit in Kapitel 3 wurde deutlich, dass sich die Arbeit mit CAT-Tools nicht nur vorteilhaft auf den Übersetzungsprozess auswirkt. In diesem Kapitel werden einige Grenzen der Arbeit mit CAT-Tools am Beispiel von OmegaT erläutert. Gleichzeitig soll gezeigt werden, wie Korpora genutzt werden können, um die computergestützte Übersetzungsarbeit zu ergänzen.

Einer der am häufigsten genannten Kritikpunkte in Bezug auf die Nutzung von CAT-Tools in der Übersetzungsarbeit ist die Segmentierung. In der Übersetzungsindustrie stetig wachsende Anforderungen an schnell und kostengünstig produzierte Übersetzungen setzen voraus, dass sich Übersetzerinnen und Übersetzer strukturiert und fokussiert durch den Auftrag arbeiten und dabei möglichst viel von bisherigen Arbeiten profitieren, indem sie bereits übersetzte Textstellen wiederverwenden. Die dabei zugrundeliegende Segmentierung der Texte in CAT-Tools wie OmegaT nach zeichenbasierten, festgelegten Regeln kann dazu führen, dass strukturelle oder semantische Einheiten getrennt werden. Dies kann sich, wie die folgenden Beispiele zeigen, auf die Übersetzungsqualität auswirken:



(9) Given that there are still significant residual external financing needs in North Macedonia's balance of payments, beyond the needs met by the resources provided 3 Decision (EU) 2020/701 of the European Parliament and of the Council of 25 May 2020 on providing macro-financial assistance to enlargement and neighbourhood partners in the context of the COVID-19 pandemic (OJ L 165, 27.5.2020, p. 2 OJ L 84, 20.3.2004, p. <Segment 0504>
31).

Abb. 8: Geöffnetes Segment im Editorfenster in OmegaT (vgl. Europäisches Parlament 2023b).



2 OJ L 84, 20.3.2004, p. <Segment 0494>

Abb. 9: Geöffnetes Segment im Editorfenster in OmegaT (vgl. Europäisches Parlament 2023b).

Im ersten Beispiel wurde, analog der Segmentierungsregel „Nach einem Punkt gefolgt von einem Leerzeichen beginnt ein neues Segment“ eine Referenzangabe getrennt. Grün markiert ist das aktuell geöffnete Segment und darunter sieht man das folgende Segment – es enthält einen Teil der Referenzangabe, die das Ende des markierten Segments bildet. Im zweiten Beispiel wurde die gesamte Referenzangabe vom Zitat abgetrennt und bildet ein eigenständiges Segment. Betrachtet man die Referenz völlig ohne Kontext und kennt die darin enthaltenen Kürzel nicht, ist völlig unklar, was die Buchstaben- und Zahlenfolgen bedeuten und wie oder ob sie übersetzt werden sollen. Zwar sehen Übersetzende im Editorfenster – wenn auch deutlich weniger auffällig – in der Regel das unmittelbar vorherige oder nachfolgende Segment wie in Abbildung 8 dargestellt, aber in Übersetzungsaufträgen werden nicht immer alle Textteile übersetzt. Häufig kommt es vor, dass Auftraggebende ihre Texte nur an bestimmten Stellen verändern. Bezahlt wird dann nur das, was ergänzend zum Rest neu übersetzt werden muss, weshalb es in OmegaT die Möglichkeit gibt, beispielsweise von einem nicht bearbeiteten Segment zum nächsten zu springen. Dieses Vorgehen berücksichtigt noch weniger die unmittelbar angrenzenden Segmente für die Bearbeitung des aktuellen Segments. Im Zweifel wird, wie im obigen Beispiel, die Bedeutung des zu übersetzenden Segments nicht voll erfasst und es wird einfach übersprungen oder nicht richtig übersetzt. Einerseits wird dadurch in der aktuellen Übersetzung gegebenenfalls ein Fehler produziert, andererseits wird die fehlerhafte Übersetzung in späteren Arbeiten durch die Übernahme eines *Fuzzy Matches* unter Umständen sogar reproduziert.

In Korpora hingegen ist neben der KWIC-Anzeige, die einen kleinen Ausschnitt des Treffers zeigt, immer auch eine Ergebnispräsentation in Volltextansicht möglich. Während die rein formale Segmentierung durch OmegaT in diesen Beispielen dazu führt, dass strukturell und semantisch zusammengehörige Einheiten vom restlichen Kontext getrennt werden und damit unverständlich sind, ist diese Einschränkung in Korpora durch die Volltextansicht nicht gegeben. Nutzen Übersetzende das Korpus DeReKo über die Korpusanalyseplattform COSMAS II, kann immer auch die gesamte sprachliche Einheit im Kontext angezeigt werden. Dass der Kontext für die richtige Übersetzung von zentraler Bedeutung ist, wurde bereits in Kapitel 4.4.1 am Beispiel des Wortes *Bund* gezeigt.

Wie Bowker und Fisher (2010) argumentieren, geht durch die Segmentierung zudem die sogenannte ‚*notion of text*‘, das Bewusstsein für den Text als Ganzheit, auf Seiten der Übersetzenden verloren (vgl. S. 63) und auch Verplaetse und Lambrechts (2019) unterstreichen diese Erkenntnis:


The segmented approach of TMs also changes the relationship translators have with the text, as it renders translation into a decontextualized activity. Often this approach influences ‚the quality of the final product in terms of syntagmatic cohesion and idiomaticity‘ (S. 22).

Krüger (2016) argumentiert, dass konkrete sprachliche Mittel durch die Segmentierung in der Übersetzung seltener vorkommen oder nicht mehr verwendet werden:

the (usually sentence-based) segmentation enforced by TM systems may affect the translator’s text-pragmatic competence, for example their use of trans-phrastic cohesive devices (such as pronouns or pronominal adverbs used to establish anaphoric or cataphoric reference) (S. 136).

Durch die Segmentierung der Texte in einzelne, eigenständig bearbeitete Teile sowie das kontinuierliche Übersetzen über Textgrenzen hinweg werden Übersetzende dazu angeregt, sich auf das aktuell geöffnete Segment zu fokussieren und der Gesamtkontext wird ausgeblendet. Auch visuell wird diese Wahrnehmung in OmegaT verstärkt, denn obwohl vorherige und nachfolgende Segmente zwar im Editorfenster zu sehen sind, wird dennoch das aktuelle Segment durch seine formale Hervorhebung deutlich in der Vordergrund gerückt, während andere Textteile im Hintergrund stehen. Somit wird jedes einzelne Segment weitgehend isoliert bearbeitet und Referenzen oder Kohäsionsmittel wie Pronomen oder Deiktika werden aufgrund des verminderten Kontextbewusstseins seltener verwendet (vgl. Krüger 2016, S. 136).

Folgendes Beispiel zeigt eine weitere Herausforderung im Umgang mit Referenzen und Deiktika, die durch die Segmentierung entsteht:



It is worth it.
It is worth it.<Segment 0032>

Abb. 10: Geöffnetes Segment im Editorfenster in OmegaT (vgl. Metsola 2023b).

In Abbildung 10 ist ein Segment aus der Vorrede von Roberta Metsola dargestellt, das ins Deutsche übersetzt werden soll. Das Pronomen *it* wird im Deutschen wörtlich mit *es* übersetzt. Aufgrund der Genusunterschiede im Deutschen und Engli-

schen jedoch muss das Pronomen in diesem Fall mit *sie* übersetzt werden, da es sich auf das Substantiv *Europäische Union* bezieht. Der offiziell vom Europäischen Parlament veröffentlichten deutschen Übersetzung ist folgender Kontext zu entnehmen: „Die Europäische Union ist wichtig. Sie ist es wert. Es lebe Europa“ (vgl. Metsola 2023a). Übersetzt man das Pronomen als Bestandteil eines einzelnen Segments, wie es in Abbildung 10 dargestellt ist, und betrachtet es völlig isoliert vom Kontext, führt dies unter Umständen zu einem Fehler in der Übersetzung, denn die Referenz zum Bezugswort kann bei falscher Übersetzung des Pronomens nicht mehr hergestellt werden. Auch an dieser Stelle kann die Segmentierung und visuelle Darstellung einzelner Segmente in OmegaT deshalb dazu führen, dass fehlerhafte Übersetzungen produziert werden.

Durch die visuelle Auflösung der Textgrenzen können außerdem kennzeichnende sprachliche und stilistische Mittel wegfallen. Wie Krieg-Holz und Schütte (2018) schreiben, ist der Textbeginn für viele Textsorten „von höchster pragmatischer Relevanz“ (S. 7), denn darin enthalten sind häufig Gestaltungsmerkmale von Autoren, Epochen, thematische Verknüpfungen oder Hinweise auf textsortenspezifische Prägungen wie beispielsweise Schlagzeilen in Zeitungstexten (vgl. S. 8). Fehlen diese sprachlichen Elemente im ZIELTEXT, führt dies zu verminderter Übersetzungsqualität (vgl. Karpińska 2017, S. 137). Auch hier würde es helfen, das einzelne Segment stärker im Gesamtzusammenhang zu betrachten. In DeReKo wird über die Volltextansicht bereits ein größerer Textausschnitt abrufbar. Darüber hinaus bietet ein Korpus die Möglichkeit, speziell nach bestimmten Textgenres zu filtern. Protokolle aus verschiedenen Parlamenten wie im ausgewählten Korpus sind hier neben Kinder- und Jugendliteratur, verschiedenen Arten von Zeitschriften oder Zeitungen oder Sprachdaten aus Wikipediadiskussionen nur eines von vielen Beispielen (vgl. DeReKo, Kapitel *Archiv*). Übersetzerinnen und Übersetzer können deshalb mithilfe eines Korpus Recherchen in der entsprechenden Textgattung durchführen und werden so für textsortenspezifische Prägungen sensibilisiert, die sie dann für die eigene Übersetzung berücksichtigen können. Das Korpus dient dabei als Recherchewerkzeug vor oder während des Übersetzungsprozesses, in dem gezielt nach bestimmten Textmerkmalen gefiltert werden kann. Bei einem TM ist dies nicht möglich, denn darin kann nicht beispielsweise nach verschiedenen Textarten gefiltert werden und es können

immer nur einzelne *Translation Units* abgerufen werden.

Neben den Auswirkungen der Segmentierung auf die Übersetzungsqualität kommt es zusätzlich zu Auswirkungen auf die Übersetzungsarbeit an sich. Bundgaard, Christensen und Schjoldager (2016) stellen fest, dass der Segment-für-Segment Ansatz den Übersetzungsprozess „more linear [...] more mechanical, less personal, more fragmented and less creative“ (S. 110) macht. In ihren Ausführungen klingt bereits an, dass Übersetzungen außerhalb von CAT-Werkzeugen in einem anderen, weniger linearen Vorgehen produziert werden. CAT-Tools entwerfen durch ihre Art der Textverarbeitung, die den Anforderungen der Übersetzungsindustrie entsprechend auf Produktivitätsmaximierung ausgerichtet ist, ein Translationsvorgehen, das nicht dem intuitiven Vorgehen der Bearbeiter entspricht. Diese Erkenntnis geht auch aus den Befragungen betroffener Übersetzerinnen und Übersetzer in LeBlanc (2013) hervor:

[T]he main drawback of TMs is that they force translators to use a sentence-by-sentence approach [...] This approach is seen as problematic by a majority of translators in that it changes the whole mental process and thus reduces translation to a mere sentence replacement activity (S. 7).

Die Segmentierung eines Textes und die Speicherung einzelner *Translation Units* in den TMs des Systems kann sich negativ auf die Übersetzungsqualität auswirken, da die vorgegebenen Rahmenbedingungen in CAT-Tools nicht den intuitiven Sprachverarbeitungsmechanismen der Übersetzenden entsprechen und gegebenenfalls, wie aus den Ausführungen von Krüger (2016) zu entnehmen ist, textpragmatische Kompetenzen verloren gehen (vgl. S. 136).

5.2 Einsatzbereich der computergestützten Werkzeuge

Eine weitere in der Forschungsliteratur häufig diskutierte Einschränkung der Arbeit mit CAT-Tools ist, dass die Werkzeuge für bestimmte Textarten nicht geeignet sind. In den vorherigen Ausführungen wurde bereits deutlich, dass sich besonders viele wiederverwendbare Strukturen zum Beispiel in technischen Texten finden. Durch den Einsatz von CAT-Tools können diese Texte effizienter und konsistenter übersetzt werden, was sich positiv auf die Qualität der Übersetzung sowie die Bearbeitungs-dauer auswirkt. Konsistenz ist in bestimmten Textarten wie Betriebsanleitungen wichtig, da beispielsweise im Handbuch eines Autos in verschiedenen Kapiteln eine Funktion thematisiert wird. Für die Lesenden wird der Bezug zwischen den Kapiteln besser deutlich, wenn in allen dieser Kapitel einheitlich beispielsweise vom *Einparkassistent* gesprochen wird und nicht plötzlich vom *aktiven Einparksystem*. Um schnell nachschlagen zu können und Bezüge zwischen einzelnen Teilkapiteln herstellen zu können, ist es deshalb wichtig, dass Fachterminologie in technischen Texten wie Betriebsanleitungen einheitlich verwendet wird.

Wie Tatu (2011) jedoch argumentiert, gibt es eine Vielzahl an Texten, die sich nicht für die Bearbeitung mit CAT-Tools eignen, weil die Texte ganz anders aufgebaut und gestaltet sind als Betriebsanleitungen:

[T]he very basic idea of TM[sic!] is fundamented on the concept of repetition of information allowing for recycled material to be used anew. However [...], very few text types lend themselves to such a degree of sentential repeatability that would justify the extensive and exclusive use of TMs within translation agencies (S. 11).

Während die einheitliche Verwendung von Terminologie oder syntaktischen Strukturen in bestimmten Texten wichtig ist, um Referenzen deutlich zu kennzeichnen und Texte möglichst verständlich zu gestalten, kann es beispielsweise in literarischen Texten sinnvoll sein, Varianten für bestimmte Formulierungen beizubehalten. Literarizität und Bedeutungsoffenheit, laut Allkemper und Eke (2014) zwei wiederkehrende Merkmale literarischer Texte, erfordern den kreativen und facettenreichen Umgang mit sprachlichen Mitteln (vgl. S. 26). Geht bei der Übersetzung des Textes aufgrund der Übersetzungsvorschläge aus dem TM jegliche sprachliche Variation verloren, kann sich die Wirkung des Textes und gegebenenfalls sogar seine Bedeutung verändern. Konsistente Übersetzungen sind deshalb nicht für alle Textarten von gleicher Bedeutung und können sogar zu verminderter Übersetzungsqualität füh-

ren.

Doch nicht nur über verschiedene Themenfelder oder Textgenres hinweg eignen sich CAT-Tools unterschiedlich gut für die Übersetzungsarbeit. Für die Untersuchung in dieser Arbeit wurde in OmegaT ein TM verwendet, das Texte aus Beschlüssen und Verträgen der Europäischen Union enthält. Auch für die Übersetzung wurden Texte aus der politischen Domäne herangezogen. Einer dieser vom Europäischen Parlament veröffentlichten Texte behandelt wie schon zu Beginn dieser Arbeit erläutert einen Beschluss für Makrofinanzhilfe in Nordmazedonien (vgl. Europäisches Parlament 2023b). Übersetzt man diesen Text mit OmegaT und dem vorgestellten TM, erhält man einige Übersetzungsvorschläge mit hohem Ähnlichkeitsgrad aus dem DGT-TM wie in Abbildung 11.

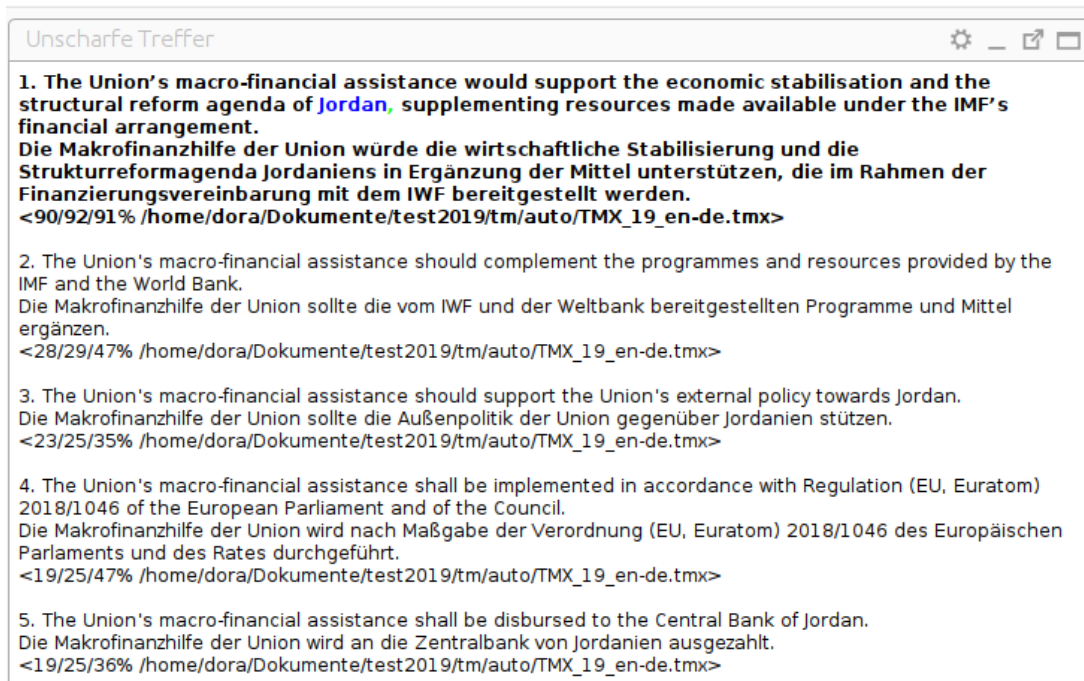


Abb. 11: Übersetzungsvorschläge in OmegaT aus dem DGT-TM.

Das oberste präsentierte *Match* hat einen berechneten Übereinstimmungsgrad von 91% und durch die blaue Markierung wird deutlich, dass der einzige Unterschied zwischen dem Übersetzungsvorschlag und dem aktuell bearbeiteten Segment das Wort *Jordan* ist. An gleicher Stelle enthält das zu übersetzende Segment, wie Abbildung 12 zeigt, die Bezeichnung *North Macedonia*.

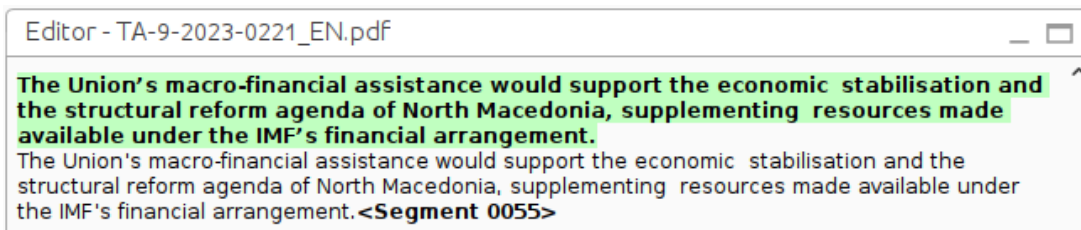


Abb. 12: Zu übersetzendes Segment im Editorfenster von OmegaT (vgl. Europäisches Parlament 2023b).

Aus einer Vielzahl an Übersetzungsvorschlägen wie diesem wird deutlich, dass Beschlüsse zur Makrofinanzhilfe sehr schematisch aufgebaut sind und einige Wiederholungsstrukturen enthalten, die sie für die Bearbeitung mit CAT-Tools eignen. Nutzen Übersetzende das entsprechende TM, können sie große Teile der Übersetzungsvorschläge übernehmen und somit Bearbeitungszeit einsparen.

Doch nicht für alle Arten politischer Texte liefert das TM geeignete *Matches*. Als zweiter Text wurde eine Vorrede von Roberta Metsola herangezogen, in der sie den Redner Olaf Scholz ankündigt (vgl. Metsola 2023b). Auch dieser Text entstammt, wie der zuvor behandelte Text und auch das TM, nicht nur der politischen Domäne, sondern wurde ganz speziell im Europaparlament vorgetragen. Jedoch unterscheidet er sich in der Textart von dem zuvor analysierten Beschluss zur Makrofinanzhilfe, denn übersetzt werden soll eine verschriftlichte Rede. Für diesen Text werden deutlich weniger Übersetzungsvorschläge mit hohem Ähnlichkeitsgrad geliefert. Obwohl der Text also auch der europapolitischen Domäne entstammt wie das TM und der zuvor übersetzte Text, gibt es hier keine größeren Wiederholungsstrukturen, die für eine schnellere Übersetzung genutzt werden könnten. Es werden nur Übersetzungsvorschläge mit geringer Übereinstimmung wie in Abbildung 14 angezeigt, die nicht aus dem TM übernommen werden können, da so viele Änderungen nötig wären, dass die Bearbeitungsdauer nicht verringert würde.

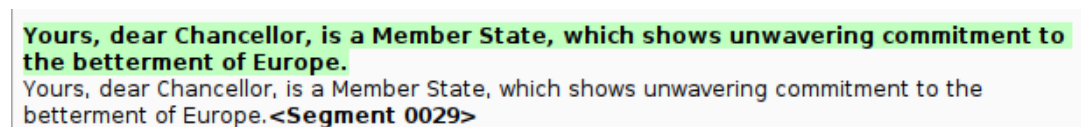


Abb. 13: Zu übersetzendes Segment im Editorfenster von OmegaT (vgl. Metsola 2023b).

1. 'State' means a Member State or the Principality of Liechtenstein.
„Staat“ einen Mitgliedstaat oder das Fürstentum Liechtenstein.
<18/31/42% /home/dora/Dokumente/test2019/tm/auto/TMX_19_en-de.tmx>

Abb. 14: Übersetzungsvorschlag für das in Abb. 13 geöffnete Segment in OmegaT.

Bei der Übersetzung der beiden thematisch aus derselben Domäne entstammenden Texte wird deutlich, wie eingeschränkt der Nutzen von Übersetzungsvorschlägen aus TMs ist, wenn neue Projekte begonnen werden oder Texte ohne stark schematischen Aufbau übersetzt werden sollen. Ähnliche Erfahrungen berichten auch Übersetzende selbst in Forenbeiträgen wie dem Folgenden:

The main point of TMs is to allow you to access repetitive material that you have previously translated. The main way that people do this is by ... working and thereby creating TMs. If you do not do much repetitive work (I don't), they are not of much use, except as a look-up tool for vocabulary that you may not have put into your glossary (Welsh 2012).

Häufig gibt es keine frei verfügbaren und ausreichend spezialisierten Textdatenbanken in Form von TMs, die von Übersetzerinnen und Übersetzern für ihre Arbeit herangezogen werden können. Im Gegensatz zu monolingualen, zielsprachlichen Korpora, so schreiben Esplà-Gomis et al. (2022), stellt deshalb eine der größten Einschränkungen der Nutzung von TMs in der Übersetzungspraxis dar, dass geeignete Ressourcen für spezielle Aufträge häufig nicht verfügbar sind (vgl. S. 7532). Die Textdatenbanken liefern keine geeigneten Übersetzungsvorschläge und tragen somit auch nicht dazu bei, dass Übersetzungen mithilfe von TMs schneller und effizienter produziert werden können.

Korpora wie DeReKo hingegen erlauben die Suche nach einzelnen Bestandteilen eines Segments. So kann beispielsweise nach Nebensätzen, Subjekt-Verb-Verbindungen oder sogar einzelnen Begriffen gesucht werden, die im Zuge der Ähnlichkeitsberechnung bei Übersetzungsvorschlägen gegebenenfalls den Übersetzenden nicht angezeigt werden. Erhält man also keinen oder nur ungeeignete Übersetzungsvorschläge im TM, ist eine weitere Möglichkeit, einzelne Bestandteile in Korpora nachzuschlagen und so wertvolle Hinweise in Bezug auf die Bedeutung oder typische Verwendung einer Wortform oder einer größeren Konstruktion zu erhalten. Das Korpus bietet dabei flexiblere Suchmöglichkeiten als reine Zeichenabgleiche über ganze Segmente hinweg und kann deshalb gezielt hinzugezogen wer-

den, wenn sprachliche Einheiten untersucht und recherchiert werden sollen. Nutzt man ein einsprachiges Korpus ist es natürlich nicht möglich, auch Übersetzungsäquivalente zu recherchieren, aber in mehrsprachigen Korpora könnte man beispielsweise ergänzend zu dem TM-Vorschlag in Abbildung 14 durch eine Korpusrecherche nach der Nomen-Verb-Verbindung *show committment* suchen, um die Übersetzung des Segments zu ergänzen.

Wie der Beitrag aus dem Übersetzerforum unterstreicht, ist es auch in CAT-Tools möglich, gezielte Suchen beispielsweise nach einzelnen Begriffen im TM eines Projektes durchzuführen. Diese Funktion wird insbesondere dann genutzt, wenn keine oder unzureichende *Fuzzy Matches* für ein Segment vorgeschlagen werden können. In ihrem Beitrag stellt die Übersetzerin diese Funktion besonders als „look-up tool for vocabulary“ (vgl. Welsh 2012) vor. Bei der Recherche nach Übersetzungsäquivalenten für einzelne Begriffe gibt es jedoch eine weitere Einschränkung von CAT-Tools im Übersetzungsprozess, denn TMs können in OmegaT zwar in verschiedenen Ordnern hinterlegt werden, aber es kann nicht nur ein bestimmter Teil eines großen TMs verwendet werden. In bestimmten Fällen wie beispielsweise bei der Übersetzung politischer oder rechtlicher Texte kann es unter Umständen hilfreich sein, nur aktuelle Texte aus einem Datenbestand für die Übersetzung heranzuziehen, da sich die Verwendung und Bedeutung eines bestimmten terminologischen Begriffs verändern kann.

Ein Beispiel für einen solchen Bedeutungswandel aus dem politischen Kontext findet sich im Neologismenwörterbuch des Leibniz-Instituts für Deutsche Sprache: Der Begriff *Buschzulage* wurde zunächst „kurz nach der Herstellung der deutschen Einheit am 03.10.1990 eingeführt“, nachdem ein Gesetz verabschiedet wurde, das finanzielle Anreize für Bürger und Bürgerinnen aus alten Ländern festschrieb, die nach der Wiedervereinigung am Aufbau der Verwaltung in neuen Ländern mitarbeiteten (vgl. NeologismenWB, Buschzulage 1). Schon Mitte der 1990er Jahre, nachdem das Gesetz bereits wieder gestrichen wurde, kam es zum Bedeutungswandel und der Begriff bezeichnet seitdem ganz grundsätzlich eine „gehaltliche Vergünstigung als Anreiz, eine Aufgabe in einer weniger attraktiven, oft weit entfernten Region aufzunehmen“ (NeologismenWB, Buschzulage 2). Je nach Entstehungszeitraum eines Textes kann dieser Begriff eine ganz konkrete Zulage bezeichnen, die auf dem 1990

verabschiedeten Gesetz basiert, oder er kann weniger nah an ein bestimmtes Gesetz geknüpft eine generalisierte Bedeutung haben, die nicht mehr direkt mit der Wiedervereinigung Deutschlands zusammenhängt. Im Englischen kann das Wort selbstverständlich nicht 1:1 übersetzt werden, sondern muss über die Definition erläutert werden.

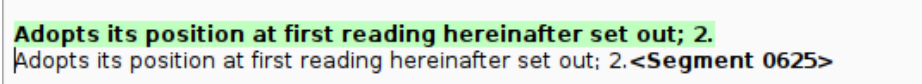
Sucht und findet man den Begriff nahezu isoliert vom restlichen Kontext wie es durch die Segmentierung bei der Suche in TMs üblich ist, wird in Beispielen wie diesem unter Umständen nicht deutlich, dass über die Jahre eine Bedeutungsveränderung erfolgt ist. Um vorrangig Übersetzungsvorschläge in der gewünschten Bedeutungsvariante zu erhalten, müssten in diesem Beispiel nur Sprachdaten aus einem bestimmten Zeitraum aus dem TM ausgewählt werden. In OmegaT ist es jedoch nicht möglich, nur einen Teilbestand des TMs für die Übersetzung zu berücksichtigen. Weder eine zeitliche Gruppierung noch eine Gruppierung nach bestimmten Themenfeldern ist nach der Erstellung des TMs noch möglich. Nutzt man in OmegaT ein großes TM, das beispielsweise aus eigenen Übersetzungen aller Projekte der vergangenen 20 Berufsjahre besteht, kann die darin enthaltene Varianz an Sprachdaten dazu führen, dass eine falsche oder veraltete Übersetzung aus einem Vorschlag übernommen wird und dadurch eine fehlerhafte Übersetzung produziert wird.

Unabhängig davon, ob Unterschiede in der Bedeutung im Umfang von nur einem Segment überhaupt für Übersetzende erkenntlich sind, zeigt das Beispiel, dass Begriffe je nach Verwendungszeitraum eine unterschiedliche Bedeutung haben können und deshalb Übersetzungsvorschläge oder Suchergebnisse aus dem TM unter Umständen zu verminderter Übersetzungsqualität führen können, wenn sie nicht in den zeitlichen Kontext des aktuell bearbeiteten Textes passen. Gerade in politischen Texten kann es beispielsweise durch politische Ereignisse innerhalb weniger Jahre oder sogar Monate zu deutlichen Bedeutungsveränderungen kommen, die für die richtige Übersetzung unbedingt berücksichtigt werden müssen. Das Korpus DeReKo enthält einerseits bereits vordefinierte Subkorpora wie das Korpus *pp – Plenarprotokolle*, die bereits entsprechend bestimmter Kriterien wie Themenfeld oder Entstehungszeitraum zusammengestellt sind. Es kann aber auch aus der Gesamtmenge an Sprachdaten ein eigenes Korpus definiert werden. Zudem kann auch bei der Auswahl eines Subkorpus weiter definiert werden, wenn für die Recherche eines bestimmten

Suchausdrucks beispielsweise nur Belegtreffer aus bestimmten Zeiträumen gezeigt werden sollen. Das Korpus bietet hier entgegen dem TM also flexible Möglichkeiten, die Datengrundlage an die jeweilige Untersuchungsfrage anzupassen. Neben der Anzeige einer Vielzahl von Belegen zum Suchausdruck in der Volltextanzeige können durch die Eingrenzung der Datengrundlage oder Treffermengen Bedeutungsveränderungen im Korpus so stärker berücksichtigt oder verglichen werden werden.

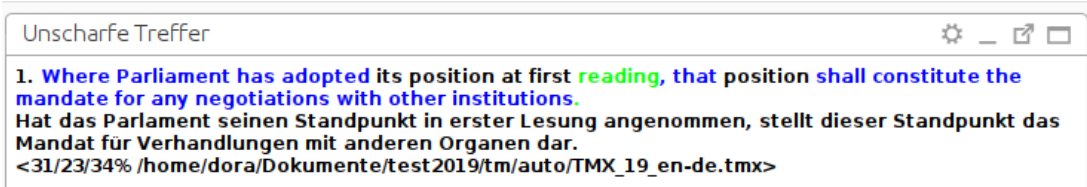
5.3 Automatische Übersetzungsvorschläge

Bereits in den vorherigen Ausführungen zum Bedeutungswandel bestimmter Begriffe wie *Buschzulage* wurde deutlich, dass Übersetzungsvorschläge aus einem TM rein auf einem formalen Abgleich der gespeicherten Daten mit dem aktuell bearbeiteten Segment basieren und Bedeutungsunterschiede nicht differenziert werden. Ein *Exact* oder *Fuzzy Match* wird nur anhand formaler Ähnlichkeit auf Zeichenebene berechnet und die Bedeutung fließt nicht in die Ähnlichkeitsberechnung mit ein. Dies kann, wie Tatu (2011) argumentiert, dazu führen, dass weniger *Matches* angezeigt werden, als es für die Übersetzenden hilfreich wäre: „wherever morphological variance is involved, like inflection or compounding, TMs fail to accomplish their duty“ (S. 176). Abbildung 15 und 16 zeigen ein Beispiel, in dem unterschiedliche Endungen dazu führen, dass das Verb *to adopt* nicht als Übereinstimmung gewertet wurde:



Adopts its position at first reading hereinafter set out; 2.
Adopts its position at first reading hereinafter set out; 2.<Segment 0625>

Abb. 15: Zu übersetzendes Segment im Editorfenster von OmegaT.



Unschärfe Treffer

1. Where Parliament has adopted its position at first reading, that position shall constitute the mandate for any negotiations with other institutions.
Hat das Parlament seinen Standpunkt in erster Lesung angenommen, stellt dieser Standpunkt das Mandat für Verhandlungen mit anderen Organen dar.
<31/23/34% /home/dora/Dokumente/test2019/tm/auto/TMX_19_en-de.tmx>

Abb. 16: Übersetzungsvorschlag für das in Abb. 15 geöffnete Segment in OmegaT.

Obwohl in beiden Segmenten das Verb *to adopt* verwendet wurde und beide Wortformen Teil der Wortverbindung *to adopt a position* sind, wurde aufgrund der morphologischen Variation im formalen Abgleich der beiden Segmente an dieser Stelle keine Ähnlichkeit festgestellt. Morphologische Varianz wirkt sich dabei auf den berechneten Ähnlichkeitsgrad aus, was dazu führen kann, dass ein Übersetzungsvorschlag im Translationsprozess nicht angezeigt wird.

In DeReKo kann über den Lemmatisierungsoperator morphologische Varianz in die Suche nach einem bestimmten Ausdruck eingebunden werden. So findet man beispielsweise im Korpus *pp – Plenarprotokolle* mit dem Suchausdruck *&annehmen*,

mit dem man nach allen Wortformen des Lemmas⁷ *annehmen* sucht, alle in Tabelle 1 dargestellten Wortformen:

Suchausdruck	<i>&annehmen</i>
Gefundene Wortformen	Angenommen, angenommen, annahm, Annahme, annahme, annähme, Annahmen, annahmen, annähmen, annehm, Annehme, annehme, Annehmen, annehmen, annehmend, Annehmende, annehmende, Annehmenden , annehmenden, annehmender, annehmendes, annehmt, annimmst, annimmt, Anzunehmen, anzunehmen, anzunehmend, Anzunehmende, anzunehmende, Anzunehmenden

Tab. 1: Im Korpus *pp – Plenarprotokolle* gefundene Wortformen mit dem Suchausdruck *&annehmen*.

Entgegen der Übersetzungsvorschläge, die Übersetzende mit einem TM erhalten, kann im Korpus entweder nach genau einer Flexionsform eines Wortes gesucht werden, oder es können mehrere Wortformen oder ein ganzes Flexionsparadigma gesucht werden. Je nach Bedarf können Übersetzende dabei selbst steuern, welcher Suchausdruck für die Recherche geeignet ist und es werden nicht automatisch nur Ergebnisse vorgeschlagen, die einen gewissen Grad an formaler Ähnlichkeit aufweisen.

Neben morphologischer Variation können auch semantische Relationen wie Synonymie in den Übersetzungsvorschlägen nicht berücksichtigt werden. Die beiden Lexeme *Orange* und *Apfelsine* sind Synonyme – sie sind auf semantischer Ebene bedeutungsgleich. Auf Zeichenebene jedoch sehen die beiden Wortformen völlig unterschiedlich aus. Bei einer rein auf formalem Abgleich berechneten Ähnlichkeit wie sie in den TM-Modulen eines CAT-Tools vorgenommen wird, würden die beiden Substantive deshalb nicht als *Matches* präsentiert, obwohl die Übersetzung des einen Begriffs durchaus für den anderen wiederverwendet werden könnte.

Einerseits kann der Zeichenabgleich bei der Suche nach wiederverwendbaren Textteilen aus der Datenbank dazu führen, dass eigentlich geeignete Segmente nicht angezeigt werden. Andererseits kann es dabei aber auch dazu kommen, dass *Matches*

7 In den Lemmatisierungsoptionen wurden alle Optionen außer „Flexionsformen“ deaktiviert.

präsentiert werden, die für die richtige Übersetzung nicht zielführend sind. Ein Beispiel hierfür findet sich auch in der Vorrede von Roberta Metsola, wie in Abbildung 17 gezeigt:

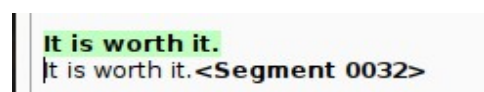


Abb. 17: Zu übersetzendes Segment im Editorfenster in OmegaT (vgl. Metsola 2023b).

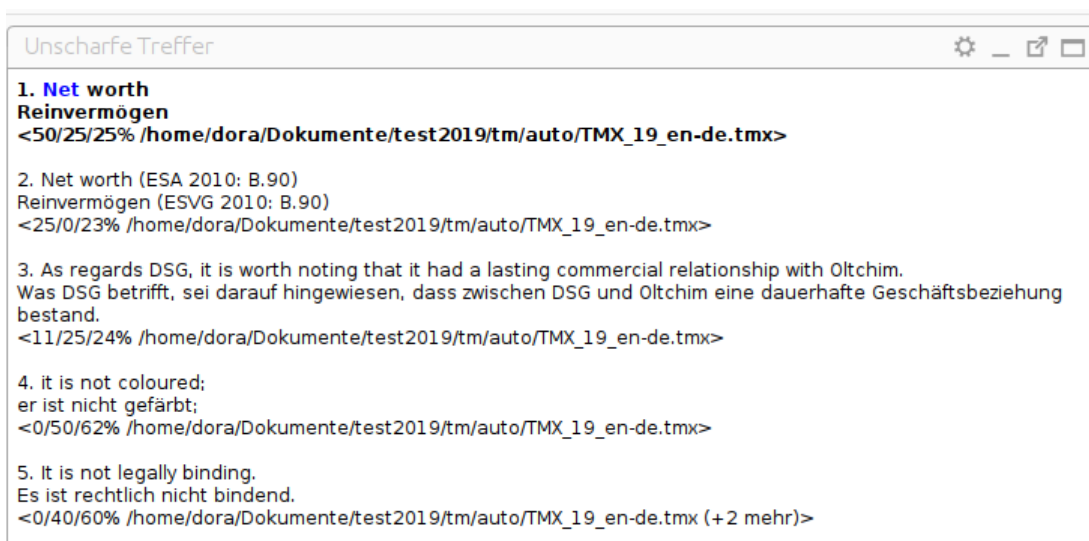


Abb. 18: Übersetzungsvorschlag für das in Abb. 17 geöffnete Segment in OmegaT.

Im ausgangssprachlichen Segment in Abbildung 17 wird das Wort *worth* als Adjektiv verwendet – dies zeigt auch die offizielle Übersetzung des Textes, in der der Satz mit „Sie ist es wert“ übersetzt wurde (vgl. Metsola 2023a). In Abbildung 18 sind die Übersetzungsvorschläge dargestellt, die für dieses Segment aus dem TM präsentiert werden. Im ersten *Fuzzy Match* schwarz markiert und damit als ähnlich klassifiziertes Bestandteil ist die Wortform *worth*, die formal mit dem im ausgangssprachlichen Segment enthaltenen Adjektiv übereinstimmt. Allerdings handelt es sich bei der Wortform im *Fuzzy Match* um das Substantiv *worth*, das im Vorschlag als Teil eines Kompositums mit *Vermögen* übersetzt wird. Wird für die Übersetzung von *worth* im Zieltext *Vermögen* übernommen, entsteht ein Fehler. Durch den reinen Zeichenabgleich bei der Suche nach wiederverwendbaren Textteilen wird ein Vorschlag präsen-

tiert, der für Übersetzende nicht zielführend ist. Obwohl beide Lexeme von *worth* als Adjektiv oder Substantiv demselben Wortfeld angehören und formal identisch sind, können sie doch zumindest in diesem Fall nicht auf gleiche Weise übersetzt werden. Die Anzeige von wiederverwendbaren Textteilen aus dem TM geht über einen ausdrucksseitigen Abgleich nicht hinaus, was zu falschen Übersetzungen führen kann, die auf Basis der Übernahme eines Übersetzungsvorschlags aus dem CAT-Tool produziert werden.

Einige Korpora hingegen enthalten morphosyntaktische Annotationen, die es ermöglichen, eine gezieltere Suche beispielsweise unter Berücksichtigung der Wortart durchzuführen. In DeReKo gibt es einige Teilbestände, die sogenannte *Part-Of-Speech Taggings* enthalten. Jedes einzelne Token ist dabei entsprechend der Wortart mithilfe eines automatischen Taggers annotiert (vgl. COSMAS II, Bemerkungen). Die Suche nach der englischsprachigen Zeichenfolge *worth* ist im deutschsprachigen Korpus DeReKo nicht möglich, weshalb am Beispiel der Wortform *wert* gezeigt werden soll, wie eine Korpusrecherche nach einer bestimmten Wortform durchgeführt werden kann. Ist den Übersetzenden bewusst, dass es sich beim aktuell bearbeiteten Segment um ein Adjektiv handelt, kann mithilfe linguistisch annotierter Korpora die Ergebnismenge eingegrenzt werden. Da das Korpus *pp – Plenarprotokolle* keine Wortartenannotationen enthält, wird stattdessen das vordefinierte Subkorpus *Tagged-T-öffentlich* aus dem *Tagged-T – Archiv morphosyntaktisch annotierter Korpora* verwendet. Über den Suchausdruck *wert /w0 MORPH(N)* kann bei der Suche festgelegt werden, dass die Zeichenfolge *wert* nur mit der Wortformenannotation Nomen gefunden werden soll wie in Beispiel 3. Beispiel 4 wiederum zeigt einen der Belegtreffer mit adjektivischer Verwendung, der mit dem Suchausdruck *wert /w0 MORPH(ADJ)* gefunden wurde.

(3) Das Spektrum der Hintergrundstrahlung, das heute einer Temperatur von 2,7 Kelvin entspricht, wies damals den Messungen an einer bestimmten Galaxie zufolge einen Wert zwischen 6 und 14 Grad über dem absoluten Nullpunkt auf. (spektrumdirekt, 14.05.2008; Ging es früher heiß her?)

(4) 10% aller US-Hausbesitzer schulden nach dem seit Mitte 2006 andauernden Einbruch der Immobilienpreise mit ihren Hypothekenkrediten mehr Geld, als ihre Häuser wert sind. (St. Galler Tagblatt, 03.01.2009; Börsen in Aufruhr)

Obwohl die Suchmöglichkeit unter Berücksichtigung morphosyntaktischer Annotationen nicht in allen Datenbeständen in DeReKo gleichermaßen möglich ist, kann diese Recherchemöglichkeit in der Übersetzungspraxis dennoch hilfreich sein, um die Treffermenge entsprechend bestimmter Kriterien weiter einzugrenzen. Da TMs keine Annotationen enthalten, sind Filtermöglichkeiten wie diese in TMs nicht gegeben. Übersetzungsvorschläge, die automatisch aus einem TM präsentiert werden, sind deshalb nicht immer passend für die aktuelle Übersetzung.

In Korpora hingegen muss aktiv nach bestimmten Wortformen gesucht werden. Dies hat den Vorteil, dass sich Übersetzende unter Umständen stärker mit dem Suchbegriff befassen, es setzt jedoch zumindest bei der Nutzung morphologischer Annotationen voraus, dass die Wortart des Suchausdrucks auch bekannt ist. In Fällen wie diesem, in denen nur ein einsprachiges Korpus in der Zielsprache vorliegt, kann zudem nicht direkt nach möglichen Übersetzungsvarianten der Wortform *worth* in der Verwendung als Adjektiv gesucht werden. Ein mehrsprachiges Korpus würde in diesem Fall deshalb eine weiterführende Hilfestellung für die Wahl des richtigen Übersetzungsäquivalents darstellen.

Die zuvor diskutierten Beispiele zeigen nur besonders deutlich auf, was laut Ingrid und Michael Friedbichler (1997) eine grundsätzliche Herausforderung in der Arbeit mit CAT-Tools ist. In ihren Ausführungen unterstreichen sie, dass Übersetzungsqualität – besonders bei Fachtexten – abhängig davon ist, ob „es dem Übersetzer gelungen ist, die Fachtermini kontext-sensitiv und entsprechend dem jeweiligen Fachidiom, der Textsorte, Kommunikationsabsicht des Autors etc. wiederzugeben“ (S. 50). Es ist also unabhängig von der Relation zwischen Ausdrucks- und Inhaltsseite eines Begriffes immer wichtig, den Gesamtkontext für eine qualitativ hochwertige Übersetzung zu beachten. Die Art, wie Übersetzungsvorschläge in OmegaT präsentiert werden, lässt es nicht zu, den gesamten Kontext des Segments in die Überlegungen miteinzubeziehen. Segmentgrenzen werden klar entsprechend der fest-

gelegten Segmentierungsregeln gesetzt, während der relevante Kontext sich weit über diese Grenzen erstrecken kann. In Korpora kann die Volltextansicht dazu beitragen, Bedeutungsunterschiede zu identifizieren. Zudem wird über die KWIC-Ansicht in COSMAS II eine Übersicht über die gefundenen Belege angezeigt, die einerseits verschieden sortiert werden kann und andererseits keine Belege gegenüber anderen visuell hervorhebt. Entgegen der Darstellungsweise in Vorschlägen aus dem TM werden Übersetzende also nicht angeleitet, sich auf das vermeintlich passendste, fett markierte und auf Rang 1 präsentierte *Match* zu konzentrieren. Gerade in der Auseinandersetzung mit einer Liste von Belegen haben Übersetzende deshalb die Möglichkeit, Bedeutungsvariation zu erkennen. Außerdem gibt es durch die vielfältigen Annotationen die Möglichkeit, Belegtreffer weiter einzugrenzen und die Recherche somit auf möglichst relevante Belege einzugrenzen.

Falsche *Fuzzy Matches* können einerseits dadurch entstehen, dass semantische Relationen oder Mehrdeutigkeiten nicht berücksichtigt werden, aber sie können auch aufgrund schlecht gepflegter Datenbanken entstehen. In einem TM werden inkrementell bisherige Übersetzungen für die Wiederverwendung gespeichert. Häufig werden dabei projektübergreifende TMs angelegt, sodass Übersetzende auch von den vorherigen Arbeiten ihrer Kolleginnen und Kollegen profitieren können. Um gute Übersetzungsvorschläge zu erhalten, ist es aber wichtig, dass nur richtige Übersetzungen in das TM einfließen. Die Realität sieht jedoch anders aus: Besonders große TMs können fehlerhafte oder veraltete Übersetzungseinheiten enthalten, die bei der Wiederverwendung in neuen Arbeiten zu schlechter Übersetzungsqualität führen und sogar reproduziert werden (vgl. LeBlanc 2013, S. 8). Zwar können Übersetzende die Vorschläge vor der Übernahme anpassen, jedoch dauert dies unter Umständen länger als die Übersetzung neu zu produzieren (vgl. Verplaetse & Lambrechts 2019, S. 21). Beispielsweise veraltete Terminologie im Vorschlag muss von den Bearbeiterinnen und Bearbeitern zudem erkannt werden. Korpora hingegen werden in der Regel vor Beginn des Übersetzungsprojektes fertig zusammengestellt und enthalten – zumindest im Falle von DeReKo – keine eigenständig produzierten Übersetzungen, sondern authentische, originalsprachliche Texte. Somit kommt es nicht zur Reproduktion vorheriger Übersetzungsfehler. Durch den Vergleich mehrerer Belegtreffer in Korpora sowie die Nutzung zielsprachlicher Belege als Informationsquelle statt als exaktes

Übersetzungsäquivalent werden sprachliche Fehler in den Belegen außerdem eher erkannt und nicht einfach per Tastenkombination übernommen. Insbesondere monolin-guale Korpora führen durch ihre hohe linguistische Qualität (vgl. Trawiński & Kupi-etz 2021, S. 213) und die Nutzungsweise deshalb in der Translationsarbeit eher nicht zu verminderter Übersetzungsqualität.

Im folgenden Kapitel soll nun abschließend diskutiert werden, welche Ergän-zungsmöglichkeiten von Korpora für die Translationsarbeit mit CAT-Tools in der vorherigen Analyse herausgearbeitet wurden, aber welche Grenzen auch weiterhin im Einsatz von Korpora in der Übersetzungspraxis bestehen.

6. Abschließende Diskussion und Ausblick

Anhand des CAT-Tools OmegaT wurde in den vorherigen Kapiteln beleuchtet, was CAT-Tools sind und welche Unterstützung sie Übersetzenden bei der Translationsarbeit bieten. In einer stetig wachsenden Übersetzungsindustrie werden die computergestützten Werkzeuge insbesondere eingesetzt, um Produktivitätssteigerung zu erzielen und größere Projektgruppen zu koordinieren. Besonders politische Texte werden regelmäßig in viele verschiedene Sprachen für eine breite Leserschaft übersetzt und bilden somit einen festen Bestandteil der Übersetzungsbranche. Während maschinelle Übersetzungssysteme wie *DeepL* und generative künstliche Intelligenz auch für die Übersetzungsarbeit immer stärker eingesetzt werden, zeigt beispielsweise das Stellenprofil für Übersetzerinnen und Übersetzer bei der Europäischen Kommission, dass weiterhin auch Fachkräfte gesucht werden, um Texte aller Tätigkeitsbereiche der Europäischen Union zu übersetzen (vgl. Europäische Kommission, Jobprofil). Dabei sind nicht nur Sprachkompetenzen und fachspezifische Kenntnisse zentrale Anforderungen an Bewerberinnen und Bewerber, sondern auch die „Beherrschung computergestützter Übersetzungshilfen und Terminologiedatenbanken“ (vgl. Europäische Kommission, Jobprofil) ist Grundbestandteil des Jobprofils. CAT-Tools sind demnach weiterhin ein wichtiges Arbeitsmittel in der Übersetzungsbranche.

Wie die vorhergehende Analyse zeigte, können CAT-Tools wie OmegaT in einigen Aufgaben der Translationsarbeit als Hilfsmittel zur Produktivitätssteigerung oder für gute Übersetzungsqualität eingesetzt werden. Am Beispiel zweier Texte aus dem europapolitischen Kontext wurde analysiert, wie OmegaT die Übersetzungsarbeit unterstützen kann. Die Betrachtung des Textes zur Makrofinanzhilfe für Nordmazedonien zeigte, dass insbesondere schematisch aufgebaute Texte für die Bearbeitung mit CAT-Tools geeignet sind. Durch die Wiederverwendung ähnlicher, zuvor übersetzter Segmente mithilfe eines *Translation Memories* kann im Translationsprozess in weniger Zeit mehr Text übersetzt werden. *Terminology Management Systems* und Glossare unterstützen zudem bei der konsistenten Verwendung von Terminologie entsprechend der Anforderungen der Auftraggebenden und automatisierte Prüfverfahren tragen dazu bei, gute Übersetzungsqualität zu sichern.

Neben den Vorteilen, die CAT-Tools für den Einsatz im Übersetzungsprozess eignen, wurde jedoch auch gezeigt, dass sich die computergestützten Tools nicht für alle Textarten gleichermaßen eignen. Insbesondere kreative, weniger schematisch aufgebaute Texte stellen eine Herausforderung für die Übersetzung mit CAT-Tools dar. Die Analyse am Beispiel einer Vorrede von Roberta Metsola zeigte, dass über das TM in OmegaT kaum geeignete Übersetzungsvorschläge präsentiert werden konnten, obwohl sowohl die Datenbasis im TM als auch der zu übersetzende Text dem europapolitischen Kontext entstammten. Obwohl das Korpus *pp – Plenarprotokolle* Texte enthält, die nicht von der europäischen Kommission veröffentlicht wurden wie die zu übersetzenden Texte, bietet das Korpus aufgrund seiner vielfältigeren Recherchemöglichkeiten und der Anzeige von Treffern in einem größeren Kontext dennoch mehr Unterstützung bei der Übersetzung der Rede von Roberta Metsola. So kann mithilfe von Suchoperatoren beispielsweise nach einzelnen Wortformen, Flexionsparadigmen oder Wortverbindungen gesucht werden, oder mithilfe linguistischer Annotationen wie POS-Tagging kann die Suche auf bestimmte Wortformen ausgerichtet werden. Die Aufbereitung und linguistische Annotation der Sprachdaten in Korpora eröffnet für die Translationsarbeit Suchmöglichkeiten, die mit einem TM nicht im gleichen Ausmaß gegeben sind. Zwar kann auch hier das TM – immer beschränkt auf Segmentebene – mithilfe von einigen regulären Ausdrücken durchsucht werden, jedoch ist die automatische Suche nach Übersetzungsvorschlägen nur anhand des exakten Zeichenabgleichs möglich.

Über Korpusanalysen kann außerdem das sprachliche Umfeld eines Suchbegriffs über die Kookkurrenzanalyse systematisch untersucht werden, was besonders für mehrdeutige Lexeme aufschlussreich sein kann. Insgesamt bietet das Korpus also Recherchemöglichkeiten, die über die Suchfunktionen in einem TM hinausgehen. In Fällen, in denen das TM keine oder sogar irreführende Vorschläge liefert, kann es sinnvoll sein, das Korpus als zusätzliches Hilfsmittel für die Übersetzungsarbeit einzusetzen. Im Rahmen eines Ausblicks soll nun abschließend eine weitere Möglichkeit der Korpusrecherche vorgestellt werden, die für die Translationsarbeit genutzt werden könnte.

Die zuvor diskutierte Kookkurrenzanalyse in den Sprachdaten von DeReKo zeigte bereits, dass die Untersuchung des sprachlichen Umfelds eines Suchausdrucks

wertvolle Hinweise für die Bedeutung einer Wortform liefern kann, die unter Umständen auch die Wahl der passenden Übersetzung eines Begriffs beeinflussen kann. Anhand der Wortform *Bund* wurde gezeigt, dass je nach Kontext verschiedene Übersetzungsäquivalente für denselben Begriff in der Ausgangssprache nötig sein können, denn *Bund* hat verschiedene Bedeutungen. Eine weitere Möglichkeit, unterschiedliche semantische Ausprägungen einer Wortform über den unmittelbaren Kontext zu ermitteln, bietet die Kookkurrenzdatenbank CCDB. Die am Leibniz-Institut für deutsche Sprache entwickelte korpuslinguistische Analyseplattform beinhaltet eine Sammlung von 220 309 Kookkurrenzprofilen (Stand 03.08.2023), die auf der Grundlage der Daten aus DeReKo erstellt wurden (vgl. CCDB, Startseite). Die Analysemethode wurde zuletzt 2007 im Rahmen eines nun beendeten Teilprojektes aktualisiert und stellt deshalb kein Produktionssystem dar (vgl. CCDB, Startseite). Ähnliche Analysemethoden zu aktuelleren Datengrundlagen gibt es bereits über andere Schnittstellen (vgl. CCDB, Startseite), aber die Kookkurrenzdatenbank CCDB soll exemplarisch als Ausblick auf weitere für die Translationsarbeit hilfreiche korpuslinguistische Methoden vorgestellt werden.

Mithilfe der Kookkurrenzdatenbank CCDB kann das Bedeutungsspektrum eines Wortes untersucht und visuell dargestellt werden. Am Beispiel des Lexems *Geist*, das mehrere Bedeutungsvarianten aufweist, soll gezeigt werden, wie Analysemethoden der CCDB-Datenbank für die semantische Analyse einer Wortform genutzt werden können. *Geist* kann einerseits im Sinne von Verstand, Gesinnung oder Scharfsinn (vgl. Duden, Verstand) sowie im Sinne von Mensch oder Person mit Fokus auf deren geistige Eigenschaften genutzt werden (vgl. Duden, Person). Weiterhin wird *Geist* auch im Sinne von Gespenst oder Spukgestalt verwendet (vgl. Duden, Person). Abbildung 19 zeigt eine selbstorganisierte Merkmalskarte (SOM-Karte), die Lexeme enthält, deren Kookkurrenzprofile dem des gesuchten Lexems *Geist* besonders ähnlich sind (vgl. Belica 2011, S. 165):

	A	B	C	D	E
1	Grundgedanke Grundprinzip Zusammenhalt Teamgeist Diskurs mündig Journalismus Bürgersinn	Wille Impetus Bewußtsein Bewusstsein Willen Pioniergeist Konstitution	Verstand Intellekt emphatisch Einbildungskraft schöpferisch wahrhaft innewohnen Instinkt	Beschwörung beseelt beschwören Gegenwelt bemächtigen	Dämon Gespenst böse böse Erinnerung Gedanken Gedanke Böse
2	Ideal Ethos Vorkämpfer Grundwert Menschenwürde Gleichheit Präambel Errungenschaft	Spannungsverhältnis impliziert unteilbar konstitutiv zuwiderlaufen implizieren Sittlichkeit	Transzendenz transzendental Negation durchdrungen Manifestation transzendent Widerstreit Naturgesetz	beseelen Seele Lebenskraft idealisiert Einfalt Eingebung Natur lebendig	übernatürlich Genius ach
3	aufklärerisch Universalismus Humanismus humanistisch Liberalismus Spannungsfeld emanzipatorisch egalitär	Gewaltlosigkeit Hüter Hüterin herleiten beschworen universal Humanität universell	allumfassend willen Menschenliebe versöhnen brüderlich Frieden Versöhnung Menschengeschlecht	Sinnbild symbolisieren versinnbildlichen Sündenfall symbolisiert Vollkommenheit zuliebe Liebe	göttlich Schöpfer Herrlichkeit Verklärung Inkarnation Zeugung Gleichnis Mysterium
4	Zeitgeist Rationalismus Materialismus Ungeist materialistisch Nihilismus Menschenbild Marxismus	Dualismus wurzeln Religion atheistisch Religiosität	Symbol Glauben Glaube predigen Nächstenliebe Spiritualität gepredigt	Heiligkeit Barmherzigkeit Erbsünde Menschwerdung bezeugen bezeugt Erlösung Erbarmen	sündig sündigen anbeten Gott Satan Auferstehung unbefleckt Wirken
5	Zeitgenosse Denker neuzeitlich Metaphysik Geschichtswissenschaft Philosophie Anthropologie Hermeneutik	Abendland Mythos Mittelalter Antike Protestantismus vorchristlich Judentum Tradition	Christentum säkularisiert Stätte säkularisieren monotheistisch Glaubenssatz Kirchenvater Hinduismus	Dreifaltigkeit Sakrament fromm Verehrung Dreieinigkeit Gottheit Gebet bekehrt	Täufer Fürbitte Vater Heilige Erlöser Grab verehren verehrt

Abb. 19: Selbstorganisierende lexikalische Merkmalskarte (SOM-Karte) für das Lexem Geist.

Die Karte enthält Lexeme, die in einem ähnlichen festgelegten sprachlichen Kontext verwendet werden wie *Geist*. Mit Blick auf die einzelnen Quadrate der SOM-Karte lässt sich das Ausmaß des Bedeutungsspektrums eines Wortes erkennen, denn „[j]e facettenreicher die [...] semantischen Strukturen sind, die mit einem Lexem [...] typischerweise assoziiert werden, desto heterogener und divergenter ist die Menge der Lexeme, die ihm in ihrem Gebrauch ähnlich sind“ (Belica 2011, S. 164). Über die Anordnung der Quadrate wird zudem deutlich, wie die darin enthaltenen Lemmata zueinander in Beziehung stehen – je näher zwei Begriffe beieinander stehen, desto ähnlicher deren Kookkurrenzprofile (vgl. Belica 2011, S. 165). Die Farbgebung

zeigt, dass zwischen den einzelnen Quadraten fließende Übergänge bestehen. Alle enthaltenen Lemmata sind somit in ihrem Kookkurrenzprofil dem des Untersuchungsobjektes ähnlich, können sich aber untereinander – je nach Platzierung und Farbgebung – in ihren Profilen mehr oder weniger ähneln:

SOMS positionieren lexikalische Ausdrücke und lexikalische Einheiten in einer zweidimensionalen, gitter- bzw. netzartigen Struktur auf die Weise, dass jeweils benachbarte Matrixzellen/Gitterkästchen semantische Nähe zwischen Kollokationsprofilen der darin aufgeführten Ausdrücke implizieren. Je entfernter die Gruppierungen voneinander liegen und je weiter sie farblich auseinandergehen, desto höher ist der Grad der semantischen Differenzen (Storjohann 2013, S. 406).

Bei der Interpretation einer SOM-Karte werden in den Bereichen verschiedene Themenfelder identifiziert, um so einzelne Lesarten des Untersuchungsobjektes zu ermitteln. Mit einem umfassenden aber differenzierten Blick auf das gesamte Bedeutungsspektrum des Lexems sollen semantische Ausprägungen ermittelt werden.

Eine genauere Betrachtung der topografischen Karte zum Suchausdruck *Geist* zeigt, dass oben rechts in Quadrat E1 eher negativ konnotierte Begriffe aufgelistet sind – darunter auch *Gespens*t und *Dämon*. Weiter unten in den Quadraten finden sich eher spirituelle und positiv konnotierte Begriffe wie *Seele* oder *übernatürlich*, bevor dann in den Quadraten im unteren rechten Teil der SOM-Karte religiöse Begriffe wie *Gott*, *Gebet* oder *Fürbitte* zu finden sind. Blickt man von dort aus weiter nach links, finden sich in den Quadraten B5 und C5 eher historische Begriffe, die teilweise auch religiösen Bezug aufweisen wie beispielsweise *Christentum* oder *vorchristlich*. Weiter links in A5 liegt der Schwerpunkt auf philosophischen Disziplinen und Konzepten mit Begriffen wie *Metaphysik*. Weiter oben in den Quadraten A4 und B4 finden sich Begriffe zu (grundlegenden) Werten oder Weltanschauungen wie *Marxismus* und *atheistisch*. Dieses Themenfeld setzt sich über die Mitte in C2 fort bis oben links in A1 bis C3 dann Begriffe wie *Grundgedanke*, *Verstand* oder *Einbildungskraft* gelistet werden, die ganz deutlich auf *Geist* in der Lesart Sinn, Verstand oder Denkweise schließen lassen. Über die SOM-Karte zu *Geist* wird visuell deutlich, dass es sich hierbei um einen mehrdeutigen Begriff handelt.

Zusätzlich zur SOM-Karte gibt es die Möglichkeit, in der CCDB sogenannte *near synonyms* in ihrem Kookkurrenzprofil mit dem des Suchausdrucks zu kontrastieren. Über die Suche nach *related collocation profiles*, die Begriffe anzeigt, welche ein ähnliches sprachliches Umfeld aufweisen wie der Suchausdruck, wurde an zwei-

ter Stelle für *Geist* das Lexem *Gespensst* gelistet. Eine CNS-Karte kann dann genutzt werden, um semantische Überlappungen und Unterschiede der beiden Lexeme weiter zu untersuchen und in einer lexikalischen Merkmalskarte wie in Abbildung 20 visuell darzustellen.

© Cyril Belica: Modelling Semantic Proximity - Contrasting Near-Synonyms (version: 0.21)
Geist : Gespensst
 export SOM as WMF or SVG file

Täufer	sündig	übernatürlich	heraufbeschwören	Bollwerk
Dreifaltigkeit	Verehrung	Schöpfer	Gefahr	Ideologie
Fürbitte	sündigen	Herrlichkeit	heraufbeschwören	Erstarken
Vater	Erbsünde	Verklärung	Deflation	Feindbild
Heilige	Gottheit	Onkel	ableiten	Militarismus
Erlöser	anbeten	zuliebe	Kriegsgefahr	Totalitarismus
Grab	Gott	Zeugung	Weltwirtschaftskrise	Hauptfeind
Sakrament	bezeugen	Mysterium	Hyperinflation	Imperialismus
Glauben	Symbol	Erinnerung	Abstiegsgespensst	heraufziehen
Glaube	göttlich	Seele	Apokalypse	Vergangenheit
Christentum	Heiligkeit	Lebenskraft	Menetekel	Antlitz
predigen	Sinnbild	Einfalt	Schreckgespensst	selbstzerstörerisch
säkularisiert	Barmherzigkeit	ach	herauf	
bekehrt	symbolisieren	Genius	Horrorszenario	
Religion	versinnbildlichen	Eingebung	Unheil	
Stätte	Sündenfall	manch	Horrorvision	
Gewaltlosigkeit	Transzendenz	Verstand	düster	umher
Hüter	schöpferisch	Intellekt	Leinwand	dräuen
Hüterin	idealisiert	beseelen	überlebensgroß	Gewitterwolke
beschwören	durchdrungen	Wille	Allegorie	geisterhaft
Nächstenliebe	Manifestation	beseelt	Schattenspiel	gespensstisch
Menschenliebe	allumfassend	Gedanken	Wand	Fatum
versöhnen	willen	Gedanke	basteln	Morgana
unteilbar	transzendent	Einbildungskraft	Bildchen	gebannt
Ideal	Dualismus	Beschwörung	Geisterstunde	gruselig
Grundgedanke	Spannungsverhältnis	Böse	Phantom	Wiedergänger
Ethos	transzendental	ausstreiben	Horrorfilm	Spuk
Vorkämpfer	Negation	ausgetrieben	untot	leibhaftig
Grundwert	herleiten	beschwören	Zombie	schaurig
Grundprinzip	universal	Magie	Kinderbuch	bannen
Zusammenhalt	Nihilismus		Kindermusical	furchterregend
Menschenwürde	Widerstreit		Kinderstück	dämonisch
aufklärerisch	Rationalismus	Mythos	Dämon	Hexe
Zeitgeist	Zeitgenosse	Vorzeit	Vampir	Elfe
Impetus	Abendland		Ungeheuer	Zauberer
Universalismus	Materialismus		Halloween	Drache
Humanismus	Denker		Werwolf	Drachen
humanistisch	materialistisch		böse	Gnome
Diskurs	wurzeln		bös	Fabelwesen
Journalismus	neuzeitlich		Unwesen	Zwerg

Abb. 20: Kombinierte lexikalische Merkmalskarte für das Wortpaar *Geist/Gespensst*.

Bei der Interpretation der Karte ist die Farbgebung ausschlaggebend, denn sie zeigt auf, zu welchem Anteil *Geist* oder *Gespensst* den in dem farblich markierten Quadrat aufgelisteten Lexemen im Gebrauch ähnlich sind. Ist das Quadrat eher gelb, sind die darin enthaltenen Ausdrücke in ihrem Kookkurrenzprofil dem des Untersuchungsobjekts *Geist* ähnlich. Rot eingefärbte Quadrate deuten auf sprachliche Umfeldern hin, in denen *Gespensst*, nicht aber *Geist* vorkommt. Semantische Überlappungen lassen sich

mit Blick auf die weder klar gelb, noch eindeutig rot eingefärbten Quadrate ermitteln. Die Karte in Abbildung 20 zeigt, dass besonders Begriffe, die vorher mit *Religion*, *Philosophie*, *Verstand* oder *Denkweise* umschrieben wurden, links deutlich gelb markiert sind. Hierbei handelt es sich also um Verwendungskontexte, in denen ausschließlich das Lexem *Geist* vorkommt. Rechts aufgeführte, stärker rot eingefärbte Quadrate und damit semantische Umfelder, in denen eher *Gespens*t vorkommt, beziehen sich auf negativ konnotierte Begriffe wie *Gefahr*, *Apokalypse* und *düster* oder Spukgestalten und Fabelwesen wie *Dämonen*, *Werwölfe*, *Zombies* oder *Unwesen*, die *heraufbeschworen* werden und *Unheil* bringen. Genauso rot markiert sind Begriffe wie *Kinderstück* und *Horrorfilm*, die sich auf Genre und Kategorien für Filme, Aufführungen oder Bücher beziehen. Stärker orange markiert Begriffe wie *Beschwörung* oder *Böse* zeigen, dass beide Lexeme in unheilvollen Kontexten vorkommen können, aber dennoch lässt die Merkmalskarte darauf schließen, dass *Gespens*t deutlich negativer konnotiert ist als *Geist*.

Eine kurze Analyse beider Merkmalskarten zeigte, dass mithilfe der Methodik – ähnlich wie mit der Kookkurrenzanalyse zur Wortform *Bund* in Kapitel 4.4.3 – das semantische Umfeld eines Lexems analysiert werden kann. Mehrdeutigkeit oder Bedeutungsvariation lassen sich durch die Analyse des sprachlichen Umfelds eines Lexems sowie die Kontrastierung mit Wortformen, die als Kookkurrenzpartner gelistet werden, aufdecken und visuell darstellen. Die Bedeutungsvariation eines Begriffes kann sich auch auf die Übersetzung auswirken. Im Falle von *Geist* kommen für die englische Übersetzung mehrere mögliche Termini infrage. Das englische Wort *mind* beispielsweise meint *Geist* im Sinne von Verstand oder Denkweise. *Geist* in seiner Bedeutung als Spukgestalt oder Gespenst wird jedoch nicht mit *mind*, sondern zum Beispiel mit *ghost* übersetzt (vgl. dict.cc, Geist).

An einigen Stellen der vorherigen Analyse wurde jedoch auch deutlich, wo ein einsprachiges Korpus wie DeReKo an seine Grenzen stößt. Mithilfe des Korpus konnten zwar Bedeutungsvariationen aufgedeckt werden, aber nur mehrsprachige Korpora bieten die Möglichkeit, die Recherche dann zusätzlich auf Übersetzungsäquivalente oder -strategien auszuweiten. Im Übersetzungsprozess kann es für solche Untersuchungen deshalb durchaus hilfreich sein, entweder zwei einsprachige Korpora für Quell- und Zielsprache zu nutzen, oder ein multilinguales Korpus heranzuzie-

hen. Ergänzend zur einsprachigen Ressource DeReKo wurde im Rahmen der 2013 gegründeten EuReCo-Initiative ein Großprojekt gestartet, in dem bestehende einsprachige Referenzkorpora zu einem vergleichbaren Korpus zusammengeführt werden sollen (vgl. Trawiński & Kupietz 2021, S. 220). Im Idealfall sollen über die Korpusabfrage- und Analyseplattform KorAP, die als Nachfolge für COSMAS II auch am IDS entwickelt wird, dann Subkorpora anhand bestimmter Kriterien wie Thema, Textsorte oder Erscheinungsjahr erstellt werden können, die mehrsprachige Recherchen in einer Datenbank ermöglichen, die einerseits vergleichbar, aber weiterhin von hoher linguistischer Qualität sind, da es sich um originalsprachliche Texte handelt (vgl. Trawiński & Kupietz 2021, S. 221). Während einzelne Korpora in ersten Pilotprojekten bereits in virtuellen Vergleichskorpora zusammengeführt werden konnten (vgl. Trawiński & Kupietz 2021, S. 223), beschränkt sich EuReCo bisher nur auf einzelne Sprachen. Perspektivisch stellt das Korpus jedoch eine Ressource dar, die für die Translationsarbeit genutzt werden kann.

Neben dem Mangel an geeigneten mehrsprachigen Ressourcen, der zwar mit Projekten wie EuReCo angegangen wird, aber derzeit weiterhin besteht, ist eine zusätzliche Herausforderung für den Einsatz von Korpora in der Übersetzungspraxis, dass für den richtigen Umgang mit Sprachdaten, Analysemethoden und Rechercheergebnissen korpuslinguistisches Fachwissen notwendig ist. Auch der richtige Umgang mit CAT-Tools muss erlernt werden, denn wie auch Bowker und Fisher (2010) bekräftigen, ist deren Einsatz vor allem am Anfang von einer großen Lernkurve begleitet (vgl. S. 63). Doch auch zuvor diskutierte Funktionen im Bereich der Korpusrecherche wie beispielsweise die richtige Einstellung von Konkurrenzparametern oder Lemmatisierungsoptionen können nur dann zum gewünschten Ergebnis führen, wenn ausreichend linguistisches Fachwissen vorhanden ist. Auch die Interpretation von Beleglisten, Konkurrenzprofilen oder Merkmalskarten bedarf fachgerechter Einarbeitung in die Korpusrecherche. Hier stößt die Hinzunahme von Korpora in der praktischen Übersetzungsarbeit an Grenzen, denn eine Einarbeitung ist zeit- und gegebenenfalls kostenintensiv und kann deshalb nicht ohne Weiteres in die tägliche Arbeit integriert werden. Dennoch ist es genau wie bei der Nutzung von CAT-Tools essentiell, dass vor der Hinzunahme von Korpora in die Translationsarbeit zumindest Grundlagen der Korpuslinguistik geschult und verstanden werden. Besonders in der

Ausbildung angehender Übersetzerinnen und Übersetzer sollten Korpora deshalb neben CAT-Tools als weiteres computergestütztes Werkzeug vorgestellt werden, das für die Arbeit genutzt werden kann.

Abschließend ist die wohl größte Herausforderung für den Einsatz von Korpora in der Translationsarbeit, dass Recherchen wie die in dieser Arbeit über COSMAS II durchgeführten sehr zeitaufwändig sind. Bereits zu Beginn der Arbeit wurde betont, dass die stetig wachsende Übersetzungsindustrie besonders von Zeitdruck geprägt ist und Übersetzerinnen und Übersetzer immer mehr Übersetzungen in immer kürzerer Zeit abliefern müssen. Eine umfassende Analyse von Kookkurrenzprofilen in unterschiedlichen Subkorpora oder die Durchsicht einer langen Liste von Belegtreffern ist deshalb in der Praxis so nicht möglich. Umso wichtiger ist es, korpuslinguistische Methoden und Analysen für die Translationsarbeit einerseits zu vereinfachen und andererseits in bestehende Tools, die Übersetzerinnen und Übersetzer bereits kennen und nutzen, zu integrieren. Merkmalskarten sind ein gutes Beispiel für Korpusanalysen, die auf einen Blick und visuell unterstützt darstellen, was für die Übersetzung notwendig ist. Während zwar die Interpretation einer solchen Karte natürlich weder trivial ist noch komplett isoliert von einer klassischen Kookkurrenzanalyse durchgeführt werden sollte, ist die Analyse der semantischen Ausprägung eines Lemmas in dieser Darstellungsform dennoch deutlich weniger komplex als die Analyse langer Listen von Partnerwörtern in COSMAS II. Ausreichendes Training zu den Möglichkeiten und Grenzen der Korpusrecherche können in der Praxis selbst zeitsparend wirken, denn wenn Übersetzende genau wissen, welche Ergebnisse sie mithilfe eines Lemmatisierungsoperators erhalten oder für welche Fragestellungen syntagmatische Muster herangezogen werden können, können sie korpusanalytische Methoden effizient einsetzen.

Um Funktionen wie diese wirklich produktiv nutzen zu können, ist es jedoch auch wichtig, dass Übersetzende nicht in vielen verschiedenen Tools arbeiten. Nur so können Korpusrecherchen wirklich in die Praxis integriert werden und schlussendlich zu verbesserter Übersetzungsqualität beitragen.

Weiterhin ist für jedes Übersetzungsprojekt zu entscheiden, wie die Faktoren Zeit, Kosten und Übersetzungsqualität zueinander im Verhältnis stehen. So kann beispielsweise bei Übersetzungen eines allgemeinen Informationsblatts für die eigenen

Mitarbeiterinnen und Mitarbeiter gegebenenfalls ein höheres Risiko für verminderte Übersetzungsqualität eingegangen werden als beispielsweise bei Übersetzungen einer Betriebsanleitung für Autos, in der falsche Übersetzungen zu sicherheitsrelevanten Fehlern führen können, die in der Konsequenz mit hohen Kosten für ein Unternehmen verbunden wären. Korpusanalysen können dann gezielt herangezogen werden, wenn besonders kritische oder für die Bearbeitung mit CAT-Tools besonders ungeeignete Texte übersetzt werden sollen, oder wenn Übersetzende ganz gezielt weitere Beispiele für einzelne Textteile benötigen.

7. Fazit

Zusammenfassend lässt sich feststellen, dass CAT-Tools als maßgeschneiderte computergestützte Werkzeuge für die Übersetzungsarbeit zwar zu einer Produktivitätsteigerung beitragen und die Translationsarbeit erleichtern können, dass sie aber auch in einigen Fällen an ihre Grenzen stoßen. Die Analyse zeigte, dass durch korpusanalytische Methoden bestimmte Herausforderungen in der Arbeit mit CAT-Tools überwunden werden können und so bessere Übersetzungen erzielt werden können. Dennoch sind Korpora, besonders deshalb, weil die Recherchen darin zeitaufwändig und komplex sind, nur ergänzend zu CAT-Tools einsetzbar. Insbesondere gezieltes Training im Umgang mit Korpora und deren Funktionen sowie die Bereitstellung geeigneter Ressourcen sind zentrale Voraussetzungen für den produktiven Einsatz von Korpora in der Übersetzungsindustrie. Gleichzeitig ist es jedoch genauso wichtig, korpuslinguistische Methoden und Werkzeuge für die Translationsarbeit anzupassen – beispielsweise über die technische Integration einzelner Funktionen in CAT-Tools. Besonders die Aufbereitung von Texten in TMs könnte aber – zumindest in öffentlich zugänglichen Ressourcen wie dem DGT-TM – ähnlich wie in Korpora durch linguistische Annotationen erweitert werden, um Übersetzenden gezielte Recherchemöglichkeiten bereitzustellen.

Die vorherige Analyse zeigte einerseits Möglichkeiten und Grenzen von CAT-Tools und Korpora in der Translationspraxis auf, verdeutlicht aber zudem, mit welchen Herausforderungen die automatisierte Verarbeitung natürlicher Sprache umgehen muss. Auch Tools wie maschinelle Übersetzungssysteme und *Large Language Models* sehen sich mit linguistischen Herausforderungen wie Kontextverständnis oder der Wahl einer geeigneten Datengrundlage für das Training oder Fine-Tuning eines Modells konfrontiert. Zu verstehen, wo die Herausforderungen im automatisierten Umgang mit Sprache liegen, ist die Basis für die Optimierung aller Arten von Systemen, die für automatisierte Translationsprozesse eingesetzt werden. Somit tragen nicht nur die Erkenntnisse der vorherigen Analyse über das Untersuchungsobjekt Sprache zur Weiterentwicklung neuer Systeme bei, sondern auch das Vorgehen dieser Arbeit in Bezug auf die kritische Auseinandersetzung mit verfügbaren Ressourcen ist

perspektivisch essentiell, um neue Tools weiterzuentwickeln und produktiv einzusetzen.

Literaturverzeichnis

Quellenverzeichnis

Bouvier, P. (2009, 21. November). *Do you allow multiple translations for 1 segment?* ProZ.com. Zuletzt abgerufen am 01.08.2023 über https://www2.proz.com/forum/sdl_trados_support/151462-do_you_allow_multiple_translations_for_1_segment.html?print=1&phpv_redirected=1

[CCDB] Belica, C. (2001ff.). *Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform*. Zuletzt abgerufen am 01.08.2023 über <https://corpora.ids-mannheim.de/ccdb/>

[COSMAS II] Leibniz-Institut für Deutsche Sprache Mannheim. *Corpus Search, Management and Analysis System (Version 2.4.2)*. Zuletzt abgerufen am 01.08.2023 über <https://cosmas2.ids-mannheim.de/cosmas2-web/>

[DeReKo] *DeReKo: Das deutsche Referenzkorpus*. Leibniz-Institut für deutsche Sprache Mannheim. Zuletzt abgerufen am 01.08.2023 über <http://www1.ids-mannheim.de/kl/projekte/korpora/>

[DGT-TM] European Commission. (o.J.). *DGT-Translation Memory*. Zuletzt abgerufen am 01.08.2023 über https://joint-research-centre.ec.europa.eu/language-technology-resources/dgt-translation-memory_en

[dict.cc, Bund] Hemetsberger, P. (o.J.). Bund. In *dict.cc. Deutsch-Englisch-Wörterbuch*. Zuletzt abgerufen am 01.08.2023 über <https://www.dict.cc/?s=bund>

[dict.cc, Geist] Hemetsberger, P. (o.J.). Bund. In *dict.cc. Deutsch-Englisch-Wörterbuch*. Zuletzt abgerufen am 01.08.2023 über <https://www.dict.cc/?s=Geist>

[Duden, Bündel] Dudenredaktion. (o.J.). Bund. In *Duden Onlinewörterbuch*. Cornelsen Verlag GmbH. Zuletzt abgerufen am 01.08.2023 über https://www.duden.de/rechtschreibung/Bund_Gebundenes

[Duden, Person] Dudenredaktion. (o.J.). Geist. In *Duden Onlinewörterbuch*. Cornelsen Verlag GmbH. Zuletzt abgerufen am 01.08.2023 über https://www.duden.de/rechtschreibung/Geist_Person

[Duden, Vereinigung] Dudenredaktion. (o.J.). Bund. In *Duden Onlinewörterbuch*. Cornelsen Verlag GmbH. Zuletzt abgerufen am 01.08.2023 über https://www.duden.de/rechtschreibung/Bund_Vereinigung

[Duden, Verstand] Dudenredaktion. (o.J.). Geist. In *Duden Onlinewörterbuch*. Cornelsen Verlag GmbH. Zuletzt abgerufen am 01.08.2023 über https://www.duden.de/rechtschreibung/Geist_Verstand

Europäisches Parlament (2023a, 13. Juni). *Makrofinanzhilfe für Nordmazedonien. Legislative Entschließung des Europäischen Parlaments vom 13. Juni 2023 zu dem Vorschlag für einen Beschluss des Europäischen Parlaments und des Rates über eine Makrofinanzhilfe für die Republik Nordmazedonien (COM(2023)0074 – C9-0015/2023 – 2023/0034(COD))*. Zuletzt abgerufen am 02.08.2023 über https://www.europarl.europa.eu/doceo/document/TA-9-2023-0221_DE.html

Europäisches Parlament (2023b, 13. Juni). *Macro-financial assistance to North Macedonia. European Parliament legislative resolution of 13 June 2023 on the proposal for a decision of the European Parliament and of the Council on providing macro-financial assistance to the Republic of North Macedonia (COM(2023)0074 – C9-0015/2023 – 2023/0034(COD))*. Zuletzt abgerufen am 02.08.2023 über https://www.europarl.europa.eu/doceo/document/TA-9-2023-0221_EN.html

Metsola, R. (2023a, 9. Mai). *We will look to Germany to help reform and prepare our European future [Deutsche Übersetzung]*. Zuletzt abgerufen am 02.08.2023 über <https://the-president.europarl.europa.eu/files/live/sites/president/files/newsroom/2023/05/09-scholz/20230509-scholz-DE.pdf>

Metsola, R. (2023b, 9. Mai). *We will look to Germany to help reform and prepare our European future*. Zuletzt abgerufen am 02.08.2023 über <https://the-president.europarl.europa.eu/home/ep-newsroom/pageContent-area/actualities/we-will-look-to-germany-to-help-reform-and-prepare-our-european-future.html>

[NeologismenWB, Buschzulage 1] Leibniz-Institut für Deutsche Sprache. (o.J.). Buschzulage. In *Neologismenwörterbuch*. Leibniz-Institut für Deutsche Sprache. Zuletzt abgerufen am 01.08.2023 über <https://www.owid.de/artikel/28969/Zulage%20im%20Beitrittsgebiet>

[NeologismenWB, Buschzulage 2] Leibniz-Institut für Deutsche Sprache. (o.J.). Buschzulage. In *Neologismenwörterbuch*. Leibniz-Institut für Deutsche Sprache. Zuletzt abgerufen am 01.08.2023 über <https://www.owid.de/artikel/28969/Zulage%20in%20einer%20weniger%20attraktiven%20Region>

[OmegaT] OmegaT. (o.J.). *OmegaT*. Zuletzt abgerufen am 01.08.2023 über <https://omegat.org/>

Pike, T. (2018, 10. Februar). *Wie ist die aktuelle Arbeitslage für Übersetzer vom Deutschen ins Englischen*. ProZ.com. Zuletzt abgerufen am 11.06.2023 über https://www.proz.com/forum/german/322874-wie_ist_die_aktuelle_arbeitslage_f%C3%BCr_%C3%BCbersetzer_vom_deutschen_ins_englischen.html

Roebuck, M. (2018, 11. Februar). *Work situation for German – English*. ProZ.com. Zuletzt abgerufen am 11.06.2023 über https://www.proz.com/forum/german/322874-wie_ist_die_aktuelle_arbeitslage_f%C3%BCr_%C3%BCbersetzer_vom_deutschen_ins_englischen.html

Welsh, S. (2012, 11. Oktober). *You make them in the course of your work*. ProZ.com. Zuletzt abgerufen am 11.06.2023 über https://www2.proz.com/forum/omegat_support/235016-pre_existing_translation_memories.html?phpv_redirected=1

Wuelk, J. (2021, 15. November). *OmegaT ist SDL Trados überlegen*. ProZ.com. Zuletzt abgerufen am 11.06.2023 über https://www.proz.com/forum/german/354344-omegat_ist_sdl_trados_%C3%BCberlegen.html

Sekundärliteratur

Adjali, O., Morin, E., & Zweigenbaum, P. (2022). Building Comparable Corpora for Assessing Multi-Word Term Alignment. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, & S. Piperidis (Hg.), *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC'2022)* (S. 3103-3112). European Language Resources Association.

Allkemper, A., & Eke, N. O. (2014). *Literaturwissenschaft*. Brill | Fink.

[Archiv, *Plenarprotokolle*] Leibniz-Institut für Deutsche Sprache (o.J.). *Korpora der geschriebenen Sprache. Plenarprotokolle*. Zuletzt abgerufen am 01.08.2023 über <https://www.ids-mannheim.de/digspra/kl/projekte/korpora/archiv/pp/>

Attila, I. (2015). Machine Translation and CAT-tools. In *Nyelvi sokszínűség Európában*. Scientia Könyvkiadó, Kolozsvár (S. 187-203). Scientia Kiadó.

Belica, C. (2011). Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen. In A. Abel, & R. Zanin (Hg.), *Korpora in Lehre und Forschung* (S. 155-178) Bozen University Press.

Biel, L. (2018). Corpora in Institutional Legal Translation: small steps and the big picture. In F. Prieto Ramos (Hg.), *Institutional Translation for International Governance: Enhancing Quality in Multilingual Legal Communication* (S. 25-36). Bloomsbury.

Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Translators' Journal* 43(4), 1-21.

Bowker, L., & Barlow, M. (2008). A comparative evaluation of bilingual concordancers and translation memory systems. In E. Y. Rodrigo (Hg.), *Topics in Language Resources for Translation and Localisation* (S. 1-22). John Benjamins.

Bowker, L., Fisher, D. (2010). Computer-aided translation. In Y. Gambier & L. van Doorslaer (Hg.), *Handbook of translation studies* (S. 60-65). John Benjamins.

Bundgaard, K., Christensen, T. P., & Schjoldager, A. (2016). Translator-computer interaction in action. An observational process study of computer-aided translation. *Journal of Specialised Translation*, 25, 106-130.

[COSMAS II, Bemerkungen] Leibniz-Institut für Deutsche Sprache Mannheim. *Corpus Search, Management and Analysis System. Bemerkungen zum Umgang mit morphosyntaktisch annotierten Korpora*. Zuletzt abgerufen am 01.08.2023 über <https://www2.ids-mannheim.de/cosmas2/projekt/referenz/bemerkungen.html>

Esplà-Gomis, M., Sánchez-Cartagena, V. M., Pérez-Ortiz, J. A., & Sánchez-Martínez, F. (2022). Cross-lingual neural fuzzy matching for exploiting target-language monolingual corpora in computer-aided translation. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Hg.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (S. 7532-7543). Association for Computational Linguistics.

[Europäische Kommission, Jobprofil] Europäische Kommission. (o.J.). *Profil für Übersetzer/-innen*. Zuletzt abgerufen am 01.08.2023 über https://commission.europa.eu/jobs-european-commission/working-eu/translator-profile_de

Fantinuoli, C., & Zanettin, F. (2015). Creating and using multilingual corpora in translation studies. In C. Fantinuoli, & F. Zanettin (Hg.), *New directions in corpus-based translation studies* (S. 1-10). Language Science Press.

Farahsani, Y., Rini, I. P., & Jaya, P. H. (2020). Google Translate Accuracy in Translating Specialized Language From English to Bahasa Indonesia: A Case Study on Mechanical Engineering Terminology. In A. Nurmandi, H. Latief, A. Gomes, Z. Rafique, B. Wadham, B. Barendregt, H. Mustafa, M. R. Bustami, M. van Bruinessen, D. Holmes, & B. Cipto (Hg.), *Proceedings of the 4th International Conference on Sustainable Innovation 2020* (S. 427-435).

Federico, M., Bertoldi, N., Trombetti, M., & Cattelan, A. (2014, 22-26. Oktober). *MateCat: an Open Source CAT Tool for MT Post-Editing* [Präsentation]. AMTA 2014, Vancouver, BC, Kanada.

Friedbichler, I., & Friedbichler, M. (1997). Korpusgestütztes Übersetzen jenseits der Wortgrenzen. Über die Bedeutung von elektronischen Textkorpora und Konkordanzprogrammen für die Translation. *Zeitschrift für fremde Sprachen in Wissenschaft und Praxis*, 42(2), 49-53.

Gaspari, F. (2003). Relevance of Parallel Corpora to the Latest Developments of Machine Translation and Computer-Assisted Translation. *International Journal of Translation*, 15(1), 27-41.

Karpińska, P. (2017). Computer Aided Translation – possibilities, limitations and changes in the field of professional translation. *Journal of Education Culture and Society* 8(2), 133-142.

Keller, N. (2011). *Neun Translation-Memory-Systeme im Praxiseinsatz*. Berufsverband Dolmetscher und Übersetzer.

Krieg-Holz, U., & Schütte, C. (2018). *Textanfänge: Konzepte und Analysen aus linguistischer, literaturwissenschaftlicher und didaktischer Perspektive*. Frank & Timme.

Krüger, R. (2016). Contextualising Computer-Assisted Translation Tools and Modelling Their Usability. *trans-kom*, 9(1), 114-148.

Kupietz, M., & Trawiński, B. (2022). Neue Perspektiven für kontrastive Korpuslinguistik: Das Europäische Referenzkorpus EuReCo. In L. Auteri, N. Barrale, A. Di Bella, & S. Hoffmann (Hg.), *Wege der Germanistik in transkultureller Perspektive. Akten des XIV. Kongresses der Internationalen Vereinigung für Germanistik (IVG) (Bd. 6)* (S. 417-139). Peter Lang.

LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies, *The International Journal for Translation & Interpreting Research*, 5(2), 1-13.

Lemnitzer, L., & Zinsmeister, H. (2015). *Korpuslinguistik. Eine Einführung* (3., überarbeitete und erweiterte Auflage). Narr Francke Attempto.

LexiQA (o. J.) Product Updates. Version 1.16: Released 7 October 2022. What's new. Lexiqa Limited. Zuletzt abgerufen am 11.06.2023 über <https://lexiqa.net/updates/>

Li, C., & Zhang, H. (2020). The Application of Computer Aided Translation in Practice. Taking the Chinese-English Translation of Tourism Texts in Hetao Region as an Example. *Journal of Physics: Conference Series*, 1693(1), 1-5.

Lüngen, H. (2017). DeReKo - Das Deutsche Referenzkorpus. *Schriftkorpora der deutschen Gegenwartssprache am Institut für Deutsche Sprache in Mannheim. Zeitschrift für germanistische Linguistik* 45(1), 161-170.

Maulidiyah, F. (2018). To use or not to use Google Translate. *Journal Linguistik Terapan Politeknik Negeri Malang* 8(2), 1-6.

Möhrs, C. (2016). *Syntagmatische Verwendungsmuster in einsprachigen deutschen Wörterbüchern*. Institut für Deutsche Sprache.

Mosavi Miangah, T. (2006). Applications of corpora in translation. *Iranian Journal of Translation Studies*, 3(12), 1-10.

[OmegaT Handbuch] OmegaT Team. (o. J.). *OmegaT 4.2 - User's Guide*. Zuletzt abgerufen am 11.06.2023 über <https://omegat.org/documentation>

[OmegaT HTML Handbuch] OmegaT Team. (o. J.). *OmegaT 2 Benutzerhandbuch*. Zuletzt abgerufen am 11.06.2023 über https://omegat.sourceforge.io/manual-standard/de/index.html#__sethome

Orozco Agudelo, R. (2022). *Needs and use of CAT tools in Colombia. User survey and quality requirements evaluation* [Masterarbeit]. University of Genève.

Perkuhn, R., Keibel, H., & Kupietz, M. (2012). *Korpuslinguistik*. Wilhelm Fink.

Petrova, V. (2019). Translation Quality Assessment Tools and Processes in Relation to CAT Tools. In I. Temnikova, C. Orasan, G. Corpas Pastor, & R. Mitkov (Hg.), *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology* (S. 89-97). Incoma Ltd.

Picton, A., Fontanet, M., Maradan, M., & Pulitano, D. (2015, Mai). *Corpora in translation: Addressing the gap between the scholar's and the translator's point of view* [Präsentation]. International Conference on Corpus Use and Learning to Translate (CULT), Alicante, Spanien.

Schlüter, P. (2020). *Statistics on the DGT-Translation Memory (DGT-TM)*. Zuletzt abgerufen am 01.08.2023 über https://wt-public.emm4u.eu/Resources/DGT-TM_Statistics.pdf

Steinberger R., Eisele, A., Klocek, S., Pilos, S., & Schlüter, P. (2012). DGT-TM: A freely Available Translation Memory in 22 Languages. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Hg.), *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012)* (S. 454-459). European Language Resources Association.

Storjohann, P. (2013). Korpuslinguistische und lexikografische Ansätze zur Beschreibung deutscher Paronyme. In D. Sava, & H. Scheuringer (Hg.), *Im Dienste des Wortes. Lexikologische und lexikografische Streifzüge. Festschrift für Ioan Lăzărescu* (S. 401-418). Karl Stutz Verlag.

Tatu, O. (2011). Using CAT Systems and/or corpora in the translation profession. *Revue Roumaine de Linguistique*, 56(2), 167-180.

Trados. (2017, 19. September). *What is Terminology Management? – Translation 101 [Video]*. YouTube. Zuletzt abgerufen am 11.06.2023 über <https://www.youtube.com/watch?v=ACW9VFDwmww&t=4s>

Trados. (o. J.). *What is machine translation? SDL Limited*. Zuletzt abgerufen am 11.06.2023 über <https://www.trados.com/solutions/machine-translation/>

Trawiński, B., & Kupietz, M. (2021). Von monolingualen Korpora über Parallel- und Vergleichskorpora zum Europäischen Referenzkorpus EuReCo. In H. Lobin, A. Witt, & A. Wöllstein (Hg.), *Deutsch in Europa. Sprachpolitisch, grammatisch, methodisch* (S. 209-234). De Gruyter.

Vela, M., Pal, S., Zampieri, M., Kumar Naskar, S., & van Genabith, J. (2019). Improving CAT Tools in the Translation Workflow: New Approaches and Evaluation. In M. Forcada, A. Way, B. Haddow, & R. Sennrich (Hg.), *Proceedings of MT Summit XVII* (S. 8-15). European Association for Machine Translation.

Verplaetse, H., & Lambrechts, A. (2019a). Surveying the use of CAT tools, terminology management systems and corpora among professional translators: General state of the art and adoption of corpus support by translator profile. *Parallèles*, 31(2), 3-31.

Yao, S. (2017). Application of Computer-aided Translation in English Teaching. *International Journal of Emerging Technologies in Learning*, 12(8), 105-117.

Zanettin, F. (2002). Corpora in Translation Practice. In E. Yuste, F. Austermühl, G. Budin, M. Olohan, & F. Zanettin (Hg.), *Language Resources for Translation Work and Research. LREC Workshop #8* (S. 10-14). European Language Resources Association.

Zaretskaya, A., Vela, M., Pastor, G. C., & Seghiri, M. (2016). Comparing Post-Editing Difficulty of Different Machine Translation Errors in Spanish and German Translations from English. *International Journal of Language and Linguistics*, 3(3), 91-100.