

Kapitel 4

CAT-Tools: Komponenten und Konzepte

Martin Kappus

Zürcher Hochschule für Angewandte Wissenschaften

In den letzten Jahren wurden die Diskussionen zum Thema Sprach- und Übersetzungstechnologie von neuen Entwicklungen wie neuronaler maschineller Übersetzung und generativen KI-Werkzeugen wie *Large Language Models* (LLMs) dominiert. Aus der professionellen Übersetzung sind aber traditionelle Übersetzungstechnologien wie Computer-Assisted-Translation Tools (CAT-Tools) immer noch nicht wegzudenken. In diesem Artikel werden die wichtigsten Konzepte und Komponenten von CAT-Tools erläutert. Dabei wird auch diskutiert, warum CAT-Tools auch heute noch eine wichtige Rolle spielen. Darüber hinaus wird aufgezeigt, welche Möglichkeiten es gibt, traditionelle Übersetzungstechnologien wie CAT-Tools mit neuronalen maschinellen Übersetzungssystemen und LLMs kombiniert werden können und wie effizientere Verknüpfungen in der Zukunft aussehen könnten

1 Einleitung

In den vergangenen Jahren haben sich die Diskussionen im Bereich der Übersetzungstechnologie in erster Linie auf generative Technologien wie neuronale maschinelle Übersetzungssysteme und generative KI-Systeme, wie die sogenannten *Large Language Models* (LLMs) konzentriert. Dabei sind etablierte Technologien wie CAT-Tools aus dem Fokus geraten, obwohl diese Anwendungen auch gegenwärtig noch ein wichtiger Bestandteil des Werkzeugkastens von professionellen Übersetzer:innen sind. Während maschinelle Überstetzungssysteme automatisch Übersetzungen generieren (oft ohne menschliches Eingreifen), unterstützen CAT-Tools menschliche Übersetzer:innen, indem sie frühere Übersetzungen, Terminologie und Qualitätsprüfungen bereitstellen, wodurch der Übersetzungsprozess effizienter und konsistenter gestaltet wird. Dieser Artikel gibt



einen Überblick über die Geschichte von CAT-Tools, erklärt wichtige Komponenten von CAT-Tools und zeigt Unterschiede zwischen verschiedenen CAT-Tools anhand verschiedener Kategorien auf. Abschließend wird die Relevanz von CAT-Tools auch und gerade in Zeiten von neuronaler maschineller Übersetzung und generativer KI diskutiert¹.

2 Geschichte

Die Geschichte der CAT-Tools lässt sich grob in zwei Phasen unterteilen. Einerseits in die Phase der konzeptuellen Vorüberlegungen, die überhaupt zur Entwicklung und zur kommerziellen Verwendung von CAT-Tools geführt haben. Andererseits in die Phase der Weiterentwicklungen, die diese kommerziellen Produkte im Anschluss bis heute durchlaufen haben. Eine detailliertere Darstellung der geschichtlichen Entwicklung der Übersetzungstechnologie findet man in Chan (2023).

2.1 Vorläufer

Ein Anstoß für die Entwicklung der ersten Übersetzungs-Werkzeuge, wie wir sie heute kennen, war die Enttäuschung Mitte der 1960er Jahre über mangelnde Fortschritte bei der Entwicklung produktiver maschineller Übersetzungssysteme, wie sie im ALPAC-Report deutlich zum Ausdruck kommt: „We have no useful machine translation [and] there is no immediate or predictable prospect of useful machine translation“ (Automatic Language Processing Advisory Committee 1966; vgl. auch Hutchins 1996). In den späten 1970er Jahren verlagerte sich daher der Schwerpunkt von der vollautomatischen maschinellen Übersetzung auf die so genannte „menschlich unterstützte maschinelle Übersetzung“ (d.h. die Vor- und/oder Nachbearbeitung von MT) oder „maschinell unterstützte menschliche Übersetzung“ (d. h. die Verwendung von Sprachtechnologiewerkzeugen). Anstatt neue Übersetzungen automatisch zu erstellen, wurde die Wiederverwendung bereits vorhandener Übersetzungen und Terminologie immer wichtiger. Es wurden Datenbanksysteme entwickelt, um frühere Übersetzungen einfach speichern und abrufen zu können. Die ersten Vorschläge für so genannte Überset-

¹Prinzipiell sind KI-Tools wie neuronale maschinelle Übersetzung und LLMs, wenn sie für translatorische Zwecke genutzt werden, auch CAT-Tools (R. Krüger p. K.). In diesem Artikel wird der Begriff *CAT-Tool* von anderen Übersetzungstechnologien dadurch abgegrenzt, dass bei den traditionellen CAT-Tool-Komponenten wie Translation Memory und Terminologiedatenbank bestehende Ressourcen wiederverwendet werden, während bei der neuronalen maschinellen Übersetzung und bei der Nutzung von LLMs neue Texte automatisiert generiert werden.

zungsarchive, die als Vorläufer der heutigen Translation Memory Tools angesehen werden können, findet man bei Krollmann (1971) und dann 1979 bei Peter Arthern bei der Europäischen Kommission (Hutchins 1998: 294). Anlass für den Vorschlag von Arthern war die Tatsache, dass die Texte der Europäischen Kommission zu dieser Zeit sehr repetitiv waren, was dazu führte, dass die Übersetzer:innen bestimmte Passagen mehr als einmal übersetzen mussten. Die ersten kommerziellen Translation-Memory-Systeme kamen jedoch erst Anfang der 1990er Jahre auf den Markt.

2.2 Geschichte der kommerziellen CAT-Tools

Der IBM Translation Manager (1980er Jahre) ist eines der frühesten Beispiele für ein CAT-Tool. Es war ein umfangreiches System, das auf Großrechnern lief und Übersetzungsspeicher (Translation Memory, TM) sowie Terminologiedatenbanken nutzte. In den 1980er Jahren entwickelte das Unternehmen ALPS eines der ersten CAT-Systeme, das eine Kombination aus Textverarbeitung und Übersetzungsspeicher bot. Es war speziell für große Unternehmen gedacht. In den frühen 90er Jahren kamen weitere kommerzielle Terminologie- und Translation-Memory-Systeme auf den Markt wie z.B MultiTerm (1990) sowie die Translator's Workbench (1994), beide von Trados (D. Brockmann p. K.), STAR Transit im Jahr 1991 (Nimdzi 2022), Deja Vu im Jahr 1993 (Lejosne 2018) und etwas später Wordfast. Die Nachfolger zweier dieser Systeme (STAR Transit und Trados Studio) sind heute noch auf dem Markt, obwohl sie mit den ursprünglichen Versionen nicht mehr viel gemeinsam haben. Ein wichtiger Faktor für den Erfolg dieser Systeme war die Entwicklung der ersten grafischen Benutzeroberflächen. Die ersten Windows-basierten Versionen von Translation-Memory-Tools wurden viel beliebter als die früheren Versionen, die auf DOS basierten.² Die Kerntechnologie von CAT-Tools hat sich in den letzten Jahren nicht verändert. So haben sich beispielsweise die Algorithmen, die für den Vergleich und das Matching von zu übersetzendem Text mit Translation-Memory-Inhalten verwendet werden, seit der Einführung von Translation-Memory-Tools nicht wesentlich verändert. Eine bemerkenswerte Ausnahme ist der Versuch, kleinere Einheiten innerhalb der Translation-Memory-Inhalte zu verwenden (das so genannte Subsegment-Matching), das in verschiedenen CAT-Tools zu finden ist (siehe auch Garcia 2014: 82–83). Neuere Entwicklungen konzentrieren sich auf die Integration anderer Systeme (z. B. NMT) oder die Kombination von Projektmanagementfunktionen und erweiterten Funktionen für die Übersetzungsproduktion (z. B. Online-Vorschau,

²Ausführlichere Übersichten über die frühe Geschichte der Übersetzungstechnologie finden sich bei Hutchins (1998) und Garcia (2014).

Predictive Typing) in einem einzigen Tool. Ein weiterer wichtiger Schritt in der Entwicklung von Übersetzungsmanagementsystemen und CAT-Tools war der Übergang von rein desktop-basierten Anwendungen zum web-basierten Zugriff auf Übersetzungstools (s.u.).

CAT-Tools haben sich also über die Jahre von einfachen Übersetzungsspeichern zu komplexen, kollaborativen Plattformen entwickelt. Sie integrieren maschinelle Übersetzung, Terminologieverwaltung, Workflow-Management und mehr.

3 Komponenten

Oft werden die Begriffe *CAT-Tool* und *Translation-Memory-System* synonym verwendet. Allerdings haben CAT-Tools einen breiteren Funktionsumfang, während ein Translation-Memory-System auf die Verwaltung und Nutzung von gespeicherten Übersetzungen spezialisiert ist. Letztlich ist das Translation Memory eine spezifische Komponente oder Funktionität innerhalb eines CAT-Tools, die sich ausschließlich auf die Speicherung und Wiederverwendung von Übersetzungen konzentriert. In den folgenden Abschnitten werden das Translation Memory und andere wichtige Komponenten von CAT-Tools genauer vorgestellt.

3.1 Translation Memory

Translation Memorys, im Deutschen auch Übersetzungsspeicher, werden vielfach als das Herzstück von CAT-Tools bezeichnet.

Translation Memorys sind Datenbanken, in denen Übersetzungen während der Übersetzung in der Regel als Satzpaare (sogenannte Übersetzungseinheiten), bestehend aus Ausgangssatz und der dazugehörigen Übersetzung, abgespeichert werden. Die in der Datenbank gespeicherten Übersetzungen werden den Übersetzer:innen im weiteren Verlauf der Übersetzungsaufgabe oder bei späteren Übersetzungen immer dann vorgeschlagen, wenn derselbe oder ein ähnlicher Ausgangssatz im zu übersetzenen Text vorkommt. Die Übersetzer:innen können die vorgeschlagene Übersetzung dann übernehmen, anstatt den selben Ausgangssatz noch einmal zu übersetzen.

Vor der Übersetzung werden die zu übersetzenen Texte in kleinere Einheiten zerlegt, die dann einerseits zur Übersetzung angezeigt werden und andererseits als Grundlage für die Suche in der Datenbank (dem Translation Memory) dienen. Dieser Vorgang wird als Segmentierung bezeichnet. Die Segmentierung basiert in der Regel darauf, dass bestimmte (Satz-)Zeichen (wie Punkte, Fragezeichen), aber auch nicht druckbare Zeichen (wie Tabulatoren oder Absatzmarken)

als Indikatoren für das Ende eines Segments verwendet werden (zusammen mit anderen Merkmalen wie Leerzeichen und nachfolgenden Großbuchstaben). Die meisten auf dem Markt erhältlichen CAT-Tools zielen auf eine Segmentierung auf Satzebene ab, d.h. der vorliegende Text wird in einzelne Sätze zerlegt.

In den früheren Versionen von CAT-Tools wurden neben den eigentlichen Übersetzungen (Satzpaaren) nur wenige Zusatzinformationen gespeichert. Typischerweise wurden Informationen zu den Benutzer:innen, die die Übersetzungseinheit erstellt oder zuletzt bearbeitet haben, sowie Datum und Uhrzeit der Erstellung der Übersetzungseinheit und der letzten Bearbeitung mit abgespeichert. Mittlerweile bieten die meisten CAT-Tools die Möglichkeit, eigene Felder anzulegen, in denen Metadaten verschiedener Datentypen verwaltet werden können. Oft werden diese Felder verwendet, um Translation-Memory-Inhalte mit Informationen wie dem Namen des Kunden, dem Fachgebiet oder Informationen zum Übersetzungsprojekt zu speichern. Diese Informationen können verwendet werden, um die Suche im Translation Memory auf bestimmte bereits angefertigte Übersetzungen einzuschränken, die für die anstehende Übersetzung besonders relevant sind. Differenzierte Metadaten können aber auch hilfreich sein, um aus großen TMs zielgerichtet Teilmengen für das Training/Fine-Tuning von MÜ-Systemen oder LLMs zu extrahieren. Seit Ende der 2000er Jahre speichern einige CAT-Tools neben den Übersetzungen und den benutzerspezifischen Metadaten auch Kontextinformationen im Translation Memory. Dabei handelt es sich in der Regel um die Information, welches Segment/welcher Satz dem aktuellen Segment/Satz vorausgeht bzw. folgt. Stimmt neben der Übersetzung aus dem Translation Memory auch der Kontext der Übersetzungseinheit mit dem zu übersetzenden Satz überein, ist die Wahrscheinlichkeit höher, dass die Übersetzung aus dem Translation Memory auch für die neue Übersetzung adäquat ist.

3.2 Terminologiekomponente

Die meisten CAT-Tools verfügen neben dem Translation Memory über eine weitere Datenbank, in der Terminologieeinträge verwaltet werden können und für die Übersetzung bereitgestellt werden. Eine Terminologiedatenbank ist eine spezialisierte Datenbank, die zur Verwaltung, Speicherung und Pflege von Fachbegriffen und deren Definitionen und anderen Metadaten verwendet wird. Bei den Einträgen in der Terminologiedatenbank kann es sich um Einzelwörter, Abkürzungen aber auch um Mehrwortbegriffe handeln. Terminologiedatenbanken unterscheiden sich von Translation Memorys nicht nur in der Größe der gespeicherten Einträge (Satz/Segment im Translation Memory vs. Einzelwort-/Mehrwortbegriffe in der Terminologiedatenbank), in Terminologiedatenbanken

werden auch oft mehr als zwei Sprachen verwaltet und sie zeichnen sich durch eine größere Anzahl von oft strukturierten Metadaten aus, die man in Translation Memory so kaum findet.

Beim Übersetzen werden den Übersetzer:innen Vorschläge aus der Termdatenbank in der Zielsprache angezeigt, wenn der entsprechende Begriff aus der Termdatenbank im Ausgangstext vorkommt. Verschiedene CAT-Tools unterscheiden sich in Bezug auf die Terminologiekomponenten im Hinblick auf:

- die Art ihrer Einbindung
- die Flexibilität der Terminologiekomponente
- deren Funktionsumfang.

Dementsprechend verwenden verschiedene Hersteller auch verschiedene Bezeichnungen für die Terminologiekomponenten ihrer Tools (Termdatenbank, Terminologieverwaltung, u. a.). Im Folgenden wird die Benennung Terminologiekomponente verwendet. Einige CAT-Tools bieten die Terminologiekomponente als eigenständige Anwendung an, die jedoch eng in die Übersetzungsumgebung der Tools integriert sind. Beispiele dafür sind Trados MultiTerm, TermStar und Cross-Term.

Die Idee hinter diesem Ansatz ist, eine eigenständige Oberfläche für die ausführlichere Erstellung und Bearbeitung von Terminologiebeständen zur Verfügung zu stellen. In der Übersetzungsumgebung werden die für die Übersetzung relevanten Ergebnisse aus den Terminologiedatenbanken angezeigt. Darüber hinaus gibt es nur begrenzte Funktionalitäten zur Bearbeitung der Terminologiebestände. Andere CAT-Tools beinhalten alle Funktionalitäten der Terminologiekomponenten im CAT-Tool selbst. Diese Unterscheidung geht in der Regel mit Unterschieden in der Flexibilität der Termbankstrukturen einher. Die Tools mit eigenständigen Anwendungen für die Terminologieverwaltung erlauben meist eine freiere Definition von Terminologiedatenbanken, wohingegen bei den CAT-Tools, bei denen die Terminologiekomponente integriert ist (wie z. B. Phrase oder memoQ), die Struktur der Termdatenbank (also die Anzahl, Art und Benennung der verfügbaren Felder) oft vorgegeben ist. Beide Ansätze haben Vor- und Nachteile. So ermöglicht eine flexiblere Handhabung der Termbankstruktur eine bessere Anpassung auch an komplexere Anforderungen im Bereich der Terminologieverwaltung, wie sie gerade in größeren Organisationen an der Tagesordnung sind. Allerdings bringt diese Flexibilität Nachteile beim Datenaustausch zwischen verschiedenen Termdatenbanken mit sich, da nicht von einer gemeinsamen Struktur ausgegangen werden kann. Eine Reihe von Artikeln

über wichtige Themen aus dem Bereich der Terminologie in der Redaktion und Übersetzung findet man im Sammelband von Hennig & Tjarks-Sobhani (2016).

3.3 Editor

Eine wichtige Komponente in einem CAT-Tool ist der Editor in dem die Übersetzer:innen ihre Übersetzungen erstellen und bearbeiten können. Im Unterschied zu herkömmlichen Texteditoren müssen die Editoren von CAT-Tools standardmäßig die Verarbeitung zweisprachiger Texte ermöglichen. Dazu wird in der Regel die einsprachige zu übersetzende Datei zunächst segmentiert und dann in ein Format umgewandelt, das Platz für die Übersetzung zur Verfügung stellt. Die Übersetzung wird direkt an dieser zweisprachigen Datei vorgenommen. Nach Abschluss der Übersetzung wird aus der zweisprachigen Datei eine einsprachige Datei in der Zielsprache erzeugt. Bis Anfang der 2000er Jahre nutzten viele CAT-Tools das Textverarbeitungssystem MS Word als Editor für die Übersetzer:innen. Die Verbindung zum Translation Memory (und die Terminologie) wurde über Makros und Symbolleisten in Word zur Verfügung gestellt. Mithilfe dieser Makros wurde der Text anhand von Sonderformatierungen und geschützten Zeichen segmentiert und die Platzhalter für die Übersetzung in den Text eingefügt. Ein Vorteil dieses Ansatzes ist die Tatsache, dass die meisten Anwender mit MS Word und seinen Funktionalitäten bereits vertraut sind. Allerdings hatte diese Herangehensweise auch Nachteile, so war die Segmentierung der Texte im Editor nur bedingt geschützt und eine Datei mit beschädigter Segmentierung konnte nicht in eine einsprachig zielsprachige Datei konvertiert werden. Da Ende der 1990er Jahre und Anfang der 2000er Jahre neben Worddateien auch immer mehr andere Dateiformate übersetzt werden mussten, wurden erste eigenständige Editoren für CAT-Tools entwickelt. Für diese Editoren werden die Ausgangsdateien in ein anderes (meist XML-basiertes) Format konvertiert. In diesen modernen Editoren werden Formatierungen und Strukturinformationen meist in Form von sogenannten Tags angezeigt. Beim Übersetzen müssen diese Tags von den Übersetzer:innen entsprechend im Zieltext platziert werden, wobei verschiedene CAT-Tools unterschiedliche Mechanismen bzw. Tastenkombinationen bieten, damit das Platzieren der Tags möglichst effizient ausgeführt werden kann. Einfache Formatierungen (z. B. Kursivschrift, Fettdruck) werden oft direkt in der Übersetzungsansicht so angezeigt, wie sie im fertigen Dokument erscheinen (WYSIWYG), während komplexere Formatierungen nur über die Labels der Tags sichtbar sind.

4 Verschiedene Kategorien von CAT-Tools

Obwohl fast alle CAT-Tools die gleichen Kernfunktionen bieten, gibt es eine Reihe von Unterschieden, anhand derer CAT-Tools bis zu einem gewissen Grad kategorisiert werden können. Die Zuordnung einzelner Produkte zu diesen Kategorien ist jedoch nicht immer eindeutig, und einige Tools können (je nach Konfiguration und Verwendungszweck) mehreren Kategorien zugeordnet werden.

4.1 Art der Plattform: Desktop-basiert vs. cloud-basiert

4.1.1 Desktop-basierte CAT-Tools

Bis ca. 2010 waren alle führenden CAT-Tools desktop-basiert. Desktop-basierte Software wird direkt auf dem lokalen Computer installiert und ausgeführt. Dieser Ansatz hat sowohl Vor- als auch Nachteile gegenüber anderen Plattformen. Die Daten und Programme befinden sich physisch auf der Festplatte der jeweiligen Benutzer:innen (oder auf einem anderen im Netzwerk zugänglichen Laufwerk). Dadurch haben die Benutzer:innen die volle Kontrolle über ihre Daten. Ein Arbeiten ist auch offline (d.h. ohne Internetzugang) möglich. Insbesondere im Vergleich zu den ersten Implementierungen von web-basierten CAT-Tools verfügen desktop-basierte CAT-Tools über einen größeren Funktionsumfang. Dieser Unterschied hat sich jedoch in den letzten Jahren deutlich verringert. Bei desktop-basierten CAT-Tools müssen die für die Abwicklung eines Überstezungsauftrags oder -projekts notwendigen Daten und Dateien auf dem lokalen Rechner verwaltet werden, d.h. es müssen entsprechende Verzeichnisse für die verschiedenen im Prozess anfallenden Daten angelegt werden (manuell oder automatisiert). Dies kann bei einer größeren Organisation oder einem größeren Auftragsvolumen zu recht komplexen Dateistrukturen führen. Außerdem ist der gleichzeitige Zugriff mehrerer Benutzer:innen auf Translation-Memory- und Terminologedaten bei desktop-basierten CAT-Tools nicht ohne Weiteres möglich. Selbst wenn die Daten an einem Speicherort abgelegt werden, der für mehrere Personen zugänglich ist, sind die zugrunde liegenden Datenbanken oft nicht für den gleichzeitigen Zugriff mehrerer Benutzer:innen ausgelegt.

4.1.2 Cloud-basierte CAT-Tools

Mit der Weiterentwicklung des Internets und der zunehmenden Verbreitung von Breitbandanschlüssen entstanden Ende der 2000er und Anfang der 2010er Jahre die ersten cloud-basierten CAT-Tools. Schon im Jahr 2012 wurden die Vor-

und Nachteile von cloud-basierten CAT-Tools von Muegge thematisiert (Muegge 2012).

Cloud-basierte Systeme werden nicht auf dem lokalen Computer, sondern auf Servern eines Anbieters installiert und über das Internet bereitgestellt. Benutzer:innen greifen über einen Webbrower oder spezielle Clients auf die Software zu. Ein großer Vorteil cloud-basierter CAT-Tools besteht darin, dass sie unabhängig vom Betriebssystem auf jedem Computer laufen, der mit dem Internet verbunden ist. Viele der am weitesten verbreiteten kommerziellen desktop-basierten CAT-Tools (z. B. Trados, memoQ, STAR Transit, across) können nur auf Computern mit dem Betriebssystem Windows installiert werden. Benutzer:innen, die ein anderes Betriebssystem wie MacOS oder Linux verwenden, können diese Anwendungen nicht oder nur mit größerem technischen Aufwand, wie z. B. mit der Verwendung von virtuellen Umgebungen, einsetzen. Da ein Großteil der Rechenleistung auf den Servern des jeweiligen Anbieters erfolgt, sind auf Seiten der Benutzer:innen auch keine so leistungsstarken Rechner wie bei desktop-basierten Tools erforderlich. Die Ablage aller benötigten Daten erfolgt bei cloud-basierten Systemen ebenfalls auf den Servern der Anbieter und die erforderlichen Strukturen werden dort meist automatisiert von den Systemen angelegt, so dass sich die Benutzer:innen nicht darum kümmern müssen. Cloud-basierte CAT-Tools sind über das Internet zugänglich und bieten den Vorteil, dass mehrere Benutzer:innen gleichzeitig und ortsunabhängig an einem Projekt arbeiten können. Dies erleichtert die Zusammenarbeit an Projekten in Teams, da auch die für die jeweiligen Schritte benötigten Dateien automatisch den jeweiligen Benutzer:innen zugewiesen werden und nicht manuell im Dateisystem verwaltet und verschoben werden müssen.

4.1.3 Client-Server-Anwendungen

Ein Sonderfall, der nicht in die oben genannten Kategorien fällt, sind die so genannten Client-Server-Anwendungen. Sie stellen eine Erweiterung von Desktop-basierten CAT-Tools dar und können einige der Nachteile von reinen Desktop-Tools ausgleichen. Noch vor dem Aufkommen von Cloud-basierten CAT-Tools boten mehrere CAT-Tool-Hersteller Client-Server-Lösungen an, bei denen die Anwendersoftware (die Client-Software) lokal auf den Rechnern der Anwender:innen installiert wird, die Daten (z. B. Translation-Memory-Daten oder Terminologie-Daten) aber auf einem Server liegen. So können verschiedene Anwender:innen gleichzeitig auf die Daten zugreifen. Diese Daten können auf internen Servern der jeweiligen Organisation gespeichert werden, was eine höhere Datensicherheit für die Organisation mit sich bringt. Die Client-Software

ist meist identisch mit den reinen Desktop-Anwendungen dieser Hersteller. Die Benutzerzugriffe können zentral verwaltet und überwacht werden, was eine bessere Einhaltung von Compliance-Richtlinien ermöglicht. Allerdings ist die Installation solcher Client-Server-Lösungen meist mit großem Aufwand und entsprechendem technischen Know-how verbunden. Außerdem sind die Kosten für Software und Hardware bei Client-Server-Anwendungen oft sehr hoch, so dass diese Lösungen in der Regel nur für größere Organisationen in Frage kommen und für kleinere Gruppen und Netzwerke von Übersetzer:innen nicht sinnvoll sind.

Die Entwicklung von CAT-Tools hat sich von einer fast ausschließlichen Konzentration auf Desktop-basierte Lösungen hin zu einer vielfältigeren Landschaft entwickelt, in der Cloud- und Hybridlösungen eine wichtige Rolle spielen. Diese Entwicklung spiegelt die zunehmende Vernetzung und Globalisierung der Übersetzungsbranche sowie den wachsenden Bedarf an Flexibilität, Mobilität und Zusammenarbeit in Echtzeit wider. Während Desktop-Tools nach wie vor für spezialisierte Aufgaben und größere Projekte beliebt sind, haben sich Cloud-basierte Tools mit interessanten Funktionalitäten und verbesserter Benutzerfreundlichkeit ihren Platz gefunden.

4.2 Unterschiedliche Lizenzmodelle: Kommerziell bis Open-Source

Die CAT-Tools verschiedenen Anbieter unterscheiden sich auch in Bezug auf die Lizenzierung und die Monetarisierung. Bei vielen der verbreiteten CAT-Tools handelt es sich um kommerziell vertriebene CAT-Tools. Die wichtigsten Modelle sind hier rein kommerziell angebotene CAT-Tools auf der einen Seite und Open-Source Software auf der anderen Seite. Kommerzielle Tools werden von Unternehmen entwickelt und vertrieben, und ihre Nutzung ist in der Regel kostenpflichtig. Es gibt dabei verschiedene Lizenzmodelle, darunter (dauerhafte) Einzellizenzen, Abonnements oder Unternehmenslizenzen. In den meisten Fällen bieten die Hersteller darüber hinaus kostenpflichtige Unterstützung bei Problemen mit der Software in Form von Support und Wartungsverträgen an. Open-Source-CAT-Tools sind in der Regel (zumindest in der einen oder anderen Form) kostenlos erhältlich, da sie unter einer Open-Source-Lizenz veröffentlicht werden. Diese Tools können von jedem heruntergeladen, genutzt und verändert werden. Sie werden von Entwicklercommunities und -konsortien entwickelt. Bekannteste Beispiele für Open-Source CAT-Tools sind OmegaT (desktop-basiert) oder Translate 5 (cloud-basiert).

Andere Kategorisierungen von CAT-Tools und verwandten Systemen finden sich z. B. in (Slator 2022) oder (Nimdzi 2022).

5 Die Rolle von CAT-Tools in Zeiten von NMT und generativer KI

CAT-Tools basieren im Wesentlichen auf der Wiederverwendung bereits vorhandener Übersetzungen, während neuere Technologien wie neuronale maschinelle Übersetzungssysteme oder generative KI, die auf LLMs basiert, darauf abzielen, neue Texte zu erstellen. Beide Ansätze weisen Vor- und Nachteile im Bereich der Übersetzung auf. Die Wiederverwendung bereits bestehender, geprüfter Übersetzungen ist für die Konsistenz und Qualität von entscheidender Bedeutung. Generative Systeme hingegen können einen zusätzlichen positiven Effekt auf die Produktivität bei der Erstellung von Übersetzungen haben. Eine der wesentlichen Herausforderungen der kommenden Jahre wird darin bestehen, die jeweiligen Vorteile der Systeme in sinnvoller Weise zu kombinieren. Im Folgenden wird zunächst erörtert, ob CAT-Tools weiterhin ihre Daseinsberechtigung besitzen, und anschließend werden bereits implementierte sowie neuartige Möglichkeiten der Kombination von CAT-Tools mit generativen Technologien dargelegt.

5.1 Warum CAT-Tools immer noch ihre Berechtigung haben

Einer der Hauptgründe für den Einsatz von CAT-Tools für bestimmte Texte, Prozesse und Szenarien ist die terminologische und textliche Konsistenz der Übersetzungen. CAT-Tools nutzen Translation Memorys, in denen frühere Übersetzungen gespeichert und wiederverwendet werden können. Dies spart Zeit und verbessert die Konsistenz, da wiederkehrende Phrasen oder Sätze automatisch vorgeschlagen werden. Generative KI-Modelle hingegen erstellen jede Übersetzung neu, was zu Inkonsistenzen führen kann. Darüber hinaus bieten CAT-Tools spezielle Funktionen zur Verwaltung von Terminologie und Glossaren. Dies ist besonders wichtig in Fachgebieten wie Recht, Medizin oder Technik, in denen terminologische Konsistenz von entscheidender Bedeutung ist. Generative KI kann zwar kreative und natürliche Übersetzungen liefern, hat aber oft Schwierigkeiten, die terminologische Konsistenz über lange Texte hinweg zu gewährleisten. Ein weiterer Punkt, der für CAT-Tools als Übersetzungswerkzeuge spricht, ist die Datensicherheit und Vertraulichkeit. Viele CAT-Tools bieten die Möglichkeit, Daten lokal auf den Rechnern der Übersetzer:innen oder auf einem firmeneigenen Server zu speichern. Generative KI-Modelle, insbesondere solche, die auf externen Cloud-Diensten basieren, werfen oft Fragen des Datenschutzes und der Datensicherheit auf, insbesondere wenn vertrauliche oder geschützte Informationen übersetzt werden sollen. CAT-Tools sind darüber hinaus speziell auf die Bedürfnisse eines Unternehmens oder einer Organisation zugeschnitten und enthalten

in der Regel nur Daten, die für die Organisation relevant sind. D. h. Translation Memorys und Terminologiedatenbanken enthalten Daten, die aus der Organisation stammen und die spezifische Unternehmenssprache widerspiegeln. Generative KI kann zwar bis zu einem gewissen Grad angepasst werden, aber die Feinabstimmung auf spezifische Kundenbedürfnisse ist zumindest derzeit schwieriger und oft weniger präzise.

Weitere Gründe für den Einsatz von CAT-Tools im Zeitalter der generativen KI sind die Integration der Tools in bestehende Arbeitsabläufe und die Nachvollziehbarkeit der Prozesse: CAT-Tools sind häufig in komplexe Übersetzungs- und Lokalisierungsworkflows integriert. Sie ermöglichen die Zusammenarbeit mehrerer Übersetzer:innen, Terminolog:innen und Projektmanager:innen und unterstützen den gesamten Lebenszyklus eines Übersetzungsprojekts von der Auftragsannahme bis zur Lieferung. KI-Tools können dies derzeit nicht in gleichem Maße leisten. Darüber hinaus ermöglichen CAT-Tools die Dokumentation und Rückverfolgbarkeit aller Schritte des Übersetzungsprozesses, was für die Qualitätssicherung und die Einhaltung von Standards wichtig ist. Generative KI liefert zwar schnelle Ergebnisse, die Nachvollziehbarkeit des Prozesses und die Sicherung einer gleichbleibend hohen Qualität sind jedoch oft schwieriger zu gewährleisten.

CAT-Tools haben also nach wie vor ihre Daseinsberechtigung, aber zu den bereits vorhandenen Funktionalitäten werden in den nächsten Jahren sicherlich weitere KI-gestützte Funktionalitäten hinzukommen.

5.2 Integration von KI und CAT-Tools

Moderne CAT-Tools nutzen in den letzten Jahren drei Kerntechnologien: Translation Memory, Terminologie und maschinelle Übersetzung. Mit der generativen künstlichen Intelligenz (KI) kommt nun eine vierte Technologie hinzu, die neue Übersetzungsmöglichkeiten eröffnet und das Potential hat, die Qualität der Übersetzungen zu verbessern. Auch wenn die generativen KI-Anwendungen erst seit ein paar Jahren die breite Masse der Anwender:innen erreicht hat, haben verschiedene CAT-Tool-Hersteller Ansätze entwickelt, wie diese neuen Technologien sinnvoll in die vorhandenen Übersetzungsumgebungen und -prozesse integriert werden. Dabei gibt es Funktionalitäten, die sich auf vorbereitende Schritte wie die Vorübersetzung mit Hilfe von maschineller Übersetzung beziehen. Verschiedene Hersteller haben bereits eine Einschätzung der Qualität von Vorschlägen aus der maschinellen Übersetzung (MT Quality Estimation, MTQE). MTQE schätzt die Qualität der maschinellen Übersetzung auf Basis verschiedener Metriken und Algorithmen, die beispielsweise sprachliche Merkmale, Kohärenz

und Übereinstimmung mit dem Ausgangstext bewerten. Dies erfolgt ohne Referenzübersetzungen, also ohne den direkten Vergleich mit einer menschlichen Übersetzung. Durch die automatische Bewertung sollen Übersetzer:innen und Projektmanager:innen schneller entscheiden können, ob eine maschinell erstellte Übersetzung direkt verwendet werden kann oder ob eine Nachbearbeitung (Post-Editing) erforderlich ist. Dadurch sollen Zeit und Kosten für die manuelle Qualitätsprüfung und das Post-Editing reduziert werden.

Eine gute Übersicht über weitere wichtige KI-gestützte Funktionalitäten, die bereits implementiert worden sind, und die Überlegungen dahinter findet man z. B. bei einem der großen CAT-Tool-Anbieter (RWS, Hersteller unter anderem von Trados Studio). Dort werden folgende Möglichkeiten für ihren KI-gestützten Assistenten aufgelistet (Brockmann 2024):

- Generative KI-Werkzeuge können ohne weitere Bearbeitungsaufforderungen (Prompts) – ähnlich wie NMT Systeme – für die eigentliche Übersetzung verwendet werden. Dies kann als Vorübersetzung des ganzen Textes geschehen oder auf Segmentebene, wenn andere Quellen kein Ergebnis liefern.
- Bei der Vorübersetzung ganzer Texte durch ein generatives KI-Werkzeug können mit zusätzlichen Bearbeitungsaufforderungen (Prompts) bestimmte Vorgaben für die Übersetzung gemacht werden. Dadurch wird die Übersetzung nach individuellen Vorgaben angepasst und auf die ganze Vorübersetzung angewendet. Dies ermöglicht eine Bearbeitung der gesamten Übersetzungsausgabe anstelle einer individuellen Anpassung jeder einzelnen Übersetzung.
- Während der Arbeit im Desktop-Editor kann von den Übersetzer:innen Unterstützung vom generativen KI-Werkzeug für die Bearbeitung der Übersetzung angefordert werden. Während der Arbeit im Editor besteht die Möglichkeit, die Bearbeitungsaufforderungen interaktiv auf Segmentbasis anzuwenden. Der wesentliche Vorteil dieser Vorgehensweise besteht darin, dass die Eingabeaufforderung auf jede Übersetzung anwendbar ist, unabhängig davon, aus welcher Quelle sie stammt (Translation Memory, NMT, generative KI) oder ob das Segment von Grund auf neu übersetzt wurde.
- Eine weitere Möglichkeit für den Einsatz von KI-Werkzeugen betrifft eines der größten Probleme, die sowohl NMT als auch generative KI heute (2024) noch mit sich bringen; die inkonsistente Verwendung von Fachterminologie in den vorgeschlagenen Übersetzungen. Mithilfe von KI-Werk-

zeugen kann die Übernahme von vorgegebener Terminologie forciert werden und das System „gezwungen“ werden, die vorgeschriften Terminologie zu verwenden. Diese automatische Übernahme von Terminologie kann sowohl bei der Erstübersetzung durch die generative KI als auch im Editor interaktiv auf Segmentbasis angewendet werden. In beiden Fällen besteht die Möglichkeit, den Assistenten anzusegnen, bei der Bereitstellung von Übersetzungsvorschlägen die für das Projekt verwendete Terminologie zu berücksichtigen.

Solche und ähnliche KI-gestützte Funktionalitäten findet man auch in anderen CAT-Tools. In den kommenden Jahren werden sicher noch viele neue Anwendungsbereiche für generative KI-Werkzeuge im Zusammenspiel mit CAT-Tools auf den Markt kommen. Angesichts der schnellen Entwicklungen in diesem Bereich in den letzten Jahren lässt sich nur spekulieren, welche dieser Funktionalitäten sich dauerhaft durchsetzen werden und ob generative KI-Werkzeuge die Aufgaben von CAT-Tools mittelfristig nicht doch mit übernehmen werden und so CAT-Tools als die Standardarbeitsumgebung für Übersetzer:innen ablösen können.

Literatur

- Automatic Language Processing Advisory Committee. 1966. *Languages and machines. Computers in translation and linguistics*. Techn. Ber. Washington, D.C.: National Research Council.
- Brockmann, Daniel. 2024. *Introducing Trados Studio 2024 — trados.com*. <https://www.trados.com/blog/introducing-trados-studio-2024/>.
- Chan, Sin-Wai. 2023. The development of translation technology 1967–2023. In Sin-Wai Chan (Hrsg.), *The Routledge encyclopedia of translation technology*, 3–41. London: Routledge.
- Garcia, Ignacio. 2014. Computer-aided translation systems. In Sin-Wai Chan (Hrsg.), *The Routledge encyclopedia of translation technology*, 68–87. Amsterdam: Routledge.
- Hennig, Jörg & Marita Tjarks-Sobhani (Hrsg.). 2016. *Terminologearbeit für Technische Dokumentation* (Schriften zur Technischen Kommunikation). Stuttgart, Germany: tcworld.
- Hutchins, John. 1996. ALPAC: The infamous report. *MT New International* 14. 9–12.
- Hutchins, John W. 1998. The origins of the translator's workstation. *Machine Translation* (13). 287–307. DOI: 10.1023/a:1008123410206.

- Krollmann, Friedrich. 1971. Linguistic data banks and the technical translator. *Meta* 16(1-2). 117–124. DOI: 10.7202/003352ar.
- Lejosne, Arnaud. 2018. *Our story - Atril Solutions*. <https://atril.com/our-story/>.
- Muegge, Uwe. 2012. Cloud-based translation memory tools are changing the way translators work and train. *The Big Wave of Language Technology* 1(3). http://works.bepress.com/uwe_muegge/76/.
- Nimdzi. 2022. *Language technology atlas*. <https://www.nimdzi.com/language-technology-atlas/>.
- Slator. 2022. *Slator machine translation expert-in-the-loop report*. <https://slator.com/machine-translation-expert-in-the-loop-report/> (27 Mai, 2024).

