

## CAT und MÜ – Getrennte Welten?

---

Im vorliegenden Artikel werden die Zusammenhänge zwischen *computerunterstützter Übersetzung* (Computer Assisted Translation, CAT) und *maschineller Übersetzung* (MÜ) untersucht. Im Mittelpunkt stehen die Systeme zur computergestützten Übersetzung sowie ihre Integrierbarkeit mit maschinellen Übersetzungssystemen. Eingangs werden einige terminologische Unterscheidungen getroffen, um die wichtigsten Begrifflichkeiten zu klären. Darüber hinaus werden die Hauptunterschiede zwischen CAT und MÜ erwähnt. Ein Überblick über die wichtigsten Komponenten eines CAT-Systems sowie über die gängigsten Produkte auf dem Markt dient als Grundlage für die Beschreibung der Integrationsmöglichkeiten. Vier Beispielsprozesse veranschaulichen die konkrete Arbeitsweise. Abschließend werden Vorteile und Nachteile einer Integration von CAT und MÜ besprochen

### 1 Terminologie

Einige terminologische Vorbemerkungen dienen zur Unterscheidung der wichtigsten Begrifflichkeiten.

#### 1.1 CAT und MÜ

Computerunterstützte Übersetzung oder maschinengestützte Übersetzung, auch als Computer Assisted Translation, oder Computer Aided Translation (CAT) bekannt, definiert eine Übersetzungsmethode, bei der die Übersetzung in der Regel von einem Menschen mit Unterstützung eines Programms gefertigt wird. Über Ausnahmen wird im Kapitel 5 berichtet.

Maschinelle Übersetzung (MÜ) oder automatische Übersetzung, auch als Machine Translation (MT), Automated oder Automatic Translation bekannt, definiert dagegen eine Übersetzung, die von einer Übersetzungssoftware erstellt wird, gegebenenfalls ohne menschliches Eingreifen.

Insbesondere die englische Bezeichnung MT wird allerdings manchmal in einem sehr allgemeinen Sinn verwendet und schließt ein (nach Reinke (2003)):

**MAHT:** Machine-Aided Human Translation, auch als MAT (Machine Aided Translation) bekannt.

**HAMT:** Human-Aided Machine Translation.

**FAMT:** Fully Automatic Machine Translation, auch als FAHQMT (Fully Automatic High Quality Machine Translation) bekannt.

Angesichts dieser Gliederung, die den Anteil des menschlichen Übersetzers am Übersetzungsablauf berücksichtigt, wäre die CAT in der MAHT anzusiedeln. Jedoch wird MT meistens nur im engeren Sinne von HAMT oder FAMT verwendet, vgl. zum Beispiel Hutchins (2003) und Trujillo (1999). Deswegen ist die CAT eher als eigenständige Übersetzungsmethode zu betrachten und nicht als Unterkategorie der MÜ.

## 1.2 MÜ-Methoden

Aufgrund der unterschiedlichen Methoden, die der MÜ zu Grunde liegen, wird sie wie folgt gegliedert (nach Trujillo (1999), Carl and Way (2003) und Eberle (2006)).<sup>1</sup>

**RBMT:** Rule-Based Machine Translation, regelbasierte maschinelle Übersetzung: Der Ausgangstext wird analysiert, und diese Analyse wird mittels eines Satzes von linguistischen Regeln und eines Wörterbuchs in Strukturen der Zielsprache übersetzt, aus denen der Zieltext generiert wird.

**EBMT:** Example-Based Machine Translation, beispielbasierte maschinelle Übersetzung: Die Übersetzung eines Ausgangstexts wird mittels Regeln aus einem alignierten zweisprachigen Korpus rekonstruiert.

**SMT:** Statistical Machine Translation, statistische maschinelle Übersetzung: Die Übersetzung wird mittels statistischer Wahrscheinlichkeitsmodelle aus einem zweisprachigen Korpus erzeugt, wobei linguistisches Wissen nicht unbedingt einbezogen wird.

## 1.3 CAT, TM und Match

Die Software, um genau zu sein das Software-Paket, welches für die CAT verwendet wird, hat unterschiedliche, zum Teil aus dem Englischen übernommene Bezeichnungen: CAT-Tool, CAT-System, integriertes System sowie Translator's Workstation. Diese Bezeichnungen sind zueinander synonym.

Ein weiterer häufiger Begriff ist Translation Memory System (TM-System). In Kapitel 3 wird das Translation Memory – eine Komponente des CAT-Systems, die Satzpaare

<sup>1</sup>Auf diese Gliederung wird im vorliegenden Artikel nur oberflächlich und vereinfacht eingegangen. Für eine umfassende Einführung und Unterscheidung, siehe TRUJILLO 1999. Für eine kürzere Einführung, siehe EBERLE 2006.

bestehend aus ausgangssprachlichen Sätzen und deren Übersetzung speichert, – näher beschrieben. Die Benennung TM-System nimmt also eine einzelne Komponente für die Bezeichnung des ganzen Software-Pakets her (*pars pro toto*).

Auf der Ebene des Translation Memorys sind drei weitere Begriffe einzuführen. Das erneute Vorkommen desselben Ausgangssegments, für das im Translation Memory eine Übersetzung vorhanden ist, wird 100% Match oder *Perfect Match* genannt.<sup>2</sup> Das Vorkommen eines ähnlichen Ausgangssegments, für das im Translation Memory eine Übersetzung vorhanden ist, wird *Fuzzy-Match* genannt.

Das Vorkommen eines Ausgangssegments, für das im Translation Memory überhaupt keine Übersetzung vorhanden ist, wird *No Match* genannt. Dabei kann der Ausgangstext vollständig neu sein oder nicht ähnlich genug, um als Fuzzy-Match erkannt zu werden.

## 2 Hauptunterschiede

Der wichtigste Punkt bei einer Gegenüberstellung von CAT und MÜ ist, dass ein CAT-System Übersetzungen wiederverwendet und nicht neu erstellt. Aus diesem Grund sind CAT-Systeme eher Information-Retrieval-Systemen ähnlich (siehe Reinke (2003)). Hinzu kommt, dass in einem CAT-System die Übersetzung in der Regel von einem Menschen erstellt wird. Das System beschränkt sich darauf, beim erneuten Vorkommen desselben oder eines ähnlichen Ausgangssegments die bereits erstellte Übersetzung vorzuschlagen.

Zwar sind modernere CAT-Systeme in der Lage, gewisse Textanpassungen vorzunehmen, zum Beispiel bei Zahlen, Tags oder Satzzeichen. Diese Ersetzungen können jedoch nicht als Neutexterzeugung definiert werden. Nur das Teilsatz-Matching, das vorerst lediglich von einigen Systemen angeboten wird, kann als Brückenschlag zur MÜ gesehen werden, wobei diese Aussage angefochten werden kann, weil das Kennzeichen dieser Funktionalitäten die kleinere Match-Einheit ist (eine Phrase statt eines ganzen Segments), und nicht die Synthese einer neuen Übersetzung samt grammatischen Anpassungen.

Die Ähnlichkeit von CAT und EBMT ist zwar unübersehbar. Allerdings werden in der CAT – in Gegensatz zu EBMT – keine automatischen Anpassungen von der Software vorgenommen, welche linguistisches Wissen voraussetzen. Sie bleiben eine Aufgabe des Humanübersetzers (siehe Trujillo (1999)).

Der zweite wichtige Punkt bei der Unterscheidung von MÜ-Systemen und CAT-Systemen ist der Entscheidungsträger. Bei den letzteren kann der Übersetzer immer eingreifen und bestimmen, ob eine angebotene Übersetzung bzw. vorgenommene Anpassung übernommen werden soll oder nicht (siehe Somers (2003a)). Es sind zwar Funktionen vorhanden, welche die ungeprüfte Übernahme von 100% Matches ermöglichen,

<sup>2</sup>Der Begriff *Perfect Match* kann auch andere Bedeutungen haben, zum Beispiel in SDL Trados. Im Allgemeinen sind weitere produktspezifische Begrifflichkeiten möglich.

aber selbst in diesem Fall liegt die Entscheidung beim Übersetzer, diese Funktionalität zu verwenden.

### 3 Komponenten eines CAT-Systems

Die Benennung CAT-System weist auf die Tatsache hin, dass dieser Systemtyp mehrere Komponenten verbindet, welche unterschiedliche und zum Teil unabhängige Funktionen ausführen.<sup>3</sup>

Die einzelnen Komponenten sind vom Produkt abhängig, es gibt jedoch einige, die bei allen Produkten zu finden sind und das Kernstück solcher Systeme bilden.

Der Rahmen dieses Artikels ermöglicht lediglich eine teilweise Aufzählung. Für eine vollständige und detaillierte Beschreibung, siehe Massion (2005).<sup>4</sup>

**Übersetzungsspeicher:** Er wird auch als Translation Memory (TM) bezeichnet und ist die wichtigste Komponente solcher Systeme. Bei den meisten Systemen handelt es sich um eine Datenbank, in einigen Fällen kommt aber auch eine Sammlung von Referenzdateien zum Einsatz.<sup>5</sup> Der Inhalt der Datenbank besteht aus Segmentpaaren, wobei ein ausgangssprachliches und mindestens ein zielsprachliches Segment vorhanden sind. Dazu kommen teilweise konfigurierbare Informationseinheiten, zum Beispiel zum Autor, Erzeugungsdatum, Fachgebiet und andere.

**Editor:** Anwendung, in der die Übersetzung angefertigt wird. Je nach Produkt kann der Editor integriert sein oder auf eine externe Anwendung zur Textverarbeitung zurückgreifen (in der Regel MS Word).

**Formatfilter:** Diese Filter ermöglichen die Bearbeitung unterschiedlicher Formate im Editor. Ihre Aktualisierung seitens der Hersteller, die saubere Trennung von Inhalt und Format sowie deren korrekte Zusammenführung nach der Übersetzung und schließlich ihre Konfigurierbarkeit seitens der Anwender können entscheidende Kaufkriterien sein.

**Alignment-Komponente:** Sie dient zum Aufbau eines Übersetzungsspeichers auf der Basis von Paralleltextrn, das heißt Paare von ausgangssprachigen und zielsprachigen Segmenten, wenn ihre Übersetzung ohne CAT-System angefertigt wurde.

<sup>3</sup>Ein Beispiel liefert das Produkt SDL Trados: Das Paket beinhaltet unter anderem die Anwendung zur Terminologieverwaltung SDL Trados MultiTerm. Diese Anwendung kann jedoch auch eigenständig verwendet werden, ohne jegliche Anbindung an das Translation Memory.

<sup>4</sup>Für diesen Artikel, wenn nicht anders angegeben, wurden Across Personal Edition (4.00), Déjà Vu X Professional (7.5), Heartsome Translation Studio (7.0), MemoQ (3.2), MultiTrans (4.3), SDL Trados Freelance (8.3.0), STAR Transit NXT (Informationen entnommen aus dem Benutzerhandbuch) sowie Wordfast (5.5) berücksichtigt.

<sup>5</sup>Zum Beispiel bei STAR Transit. Auch die so genannten TextBases von MultiTrans, eines Produktes des kanadischen Herstellers MultiCorpora, weisen gewisse Ähnlichkeiten mit diesem Ansatz auf.

KOMPONENTE	FUNKTION
Translator's Workbench	Übersetzungsspeicher
TagEditor	Integrierter Editor
Filter Settings u.a.	Formatfilter
WinAlign	Alignment Komponente
Synergy	Projektmanagement-Anwendung
MultiTerm	Terminologie-Datenbank

**Tabelle 1:** Komponenten eines CAT-Systems

SYSTEM	MÜ-INTEGRATION
Across	Ja
Heartsome	Ja (über API)
Déjà Vu	Ja (über API)
SDL Trados	Ja
Transit	Ja
Wordfast	Ja

**Tabelle 2:** Technische Integration von CAT und MÜ

**Projektmanagement-Anwendung:** Viele Produkte bieten auch eine Software, welche die Verwaltung (im weiteren Sinne) der Übersetzungsprojekte vereinfacht. Bei Unternehmenslösungen sind die Funktionalitäten dieser Komponenten erwartungsgemäß besonders ausgebaut.

**Terminologie-Datenbank:** Analog zu dem Übersetzungsspeicher beinhaltet sie die ein- oder mehrsprachigen Terminologieinträge. Je nach Produkt kann sie eine einfache Liste sein oder mit zusätzlichen Feldern und Informationen (einschließlich Bilder) versehen werden.

Ein Beispiel anhand eines marktüblichen Produkts (SDL Trados) liefert Tabelle 1.

Abschließend folgt eine alphabetische Auflistung der gängigsten – keineswegs aber aller erhältlichen – Systeme (vgl. Lagoudaki (2006)): Across, Déjà Vu, Heartsome, MemoQ, MultiTrans, SDL Trados, Transit, Wordfast.

#### 4 Integration

Die Integration zwischen CAT-Systemen und MÜ-Systemen hat in den letzten Jahren an Bedeutung gewonnen und wird mittlerweile von mehreren Systemen angeboten, wie Tabelle 2 zeigt (aus Massion (2008)):

CAT-System	MÜ-System
Across	Language Weaver
SDL Trados	SDL, Systran, Logos
Transit	Logos, Reverso, Systran
Wordfast	Systran, Power Translator ...

**Tabelle 3:** Zusammenarbeit zwischen Herstellern von CAT und MÜ-Systemen

Allerdings ist nur bei einigen Systemen eine direkte Integration vorhanden. In den übrigen Fällen muss die Interaktion über eine API-Implementierung (Application Programming Interface) erfolgen, also über eine Programmierschnittstelle, die von einer Software zur Verfügung gestellt wird, und mit deren Hilfe andere Programme an die Software angebunden werden können.

Die Integration über API birgt die Gefahr, dass bei Software-Updates die Anbindung nicht mehr funktioniert. Damit solche Kompatibilitätsprobleme gelöst werden können, ist eine kontinuierliche Pflege der API notwendig.

Die Integration der Systeme hat auch die Zusammenarbeit zwischen den Herstellern von CAT-Systemen und MÜ-Systemen gefördert. Tabelle 3 im Kapitel 4.1 bietet einen Überblick. Es muss jedoch vorausgeschickt werden, dass bei großen Unternehmenslösungen weitere, im vorliegenden Artikel nicht aufgelistete Integrationen möglich sind.

#### 4.1 CAT- und MÜ-Systeme

Im vorliegenden Artikel wird die Integration ausschließlich aus dem Blickwinkel der CAT-Systeme betrachtet. Es wird nicht darauf eingegangen, welche MÜ-Systeme über Schnittstellen zu CAT-Systemen verfügen.

Seit 2007 besteht eine Partnerschaft zwischen Across Systems GmbH und Language Weaver Inc. Language Weaver ist ein statistisches maschinelles Übersetzungssystem. Die Integrationsmöglichkeit besteht vorerst nur für den Across Language Server, also nicht für die Across Einzelplatzversion. Die Anbindung erfolgt über eine Dynamic Link Library (DLL).

SDL Trados verfügt über eine Exportfunktion, die ein mit Systran bzw. Logos kompatibles Format bietet. Für nähere Informationen siehe 5.1.2. Die Firma SDL verfügt außerdem über eine eigene Lösung, den SDL Enterprise Translation Server. Seit der Version 8.3 bietet SDL Trados auch für Einzelplatzversionen den Zugang über das Internet zu diesem maschinellen Übersetzungssystem.



Abbildung 1: Transit Ressourcenleiste

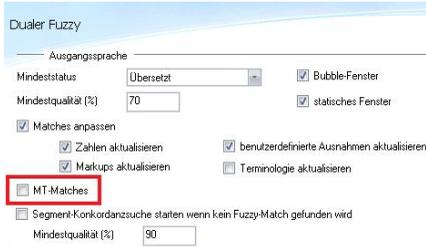


Abbildung 2: Abzug für MÜ-Matches in Transit NXT

Mit der neuen Version NXT<sup>6</sup> bietet Transit eine erweiterte Integration mit MÜ-Systemen (vgl. StarAG (2008)), die kundenspezifisch erfolgen wird. Durch die Schaltfläche Masch. Übers. auf der Ressourcenleiste kann eine Verbindung zum gewünschten MÜ-System hergestellt werden (siehe Abbildung 1).

Standardmäßig behandelt Transit NXT die Übersetzungen aus einem MÜ-System gesondert, indem sie nicht automatisch übernommen werden. Dies lässt sich in den Grundeinstellungen der Funktion Dualer Fuzzy durch die Option MT Matches ändern, siehe Abbildung 2.

Wird die Option aktiviert, werden 100% Matches, die auf maschinellen Übersetzungen gründen, automatisch übernommen.

Unter den CAT-Systemen stellt Wordfast in gewisser Hinsicht einen Spezialfall dar. Dieses Produkt ist keine eigenständige Anwendung, sondern ein Add In für MS Word.<sup>7</sup> Aus diesem Grund kann Wordfast prinzipiell mit allen MÜ-Systemen kommunizieren, die ebenfalls zumindest auch als Add-In in MS Word arbeiten können.

Die Marktentwicklung belegt das wachsende Interesse zum Thema Integration. Weitere Hersteller von CAT-Systemen werden solche Funktionalitäten künftig auch anbieten, zum Beispiel MultiCorpora. In der Version 4.4 von MultiTrans wird eine Integration mit @prompt und Systran möglich sein.

Die Integration zwischen Heartsome und Asia Online war zum Redaktionsschluss noch nicht fertig. Sie wird über eine API erfolgen.

<sup>6</sup>Transit NXT ist Ende 2008 zu Redaktionsschluss auf den Markt gekommen. Aus diesem Grund waren für den vorliegenden Artikel nur allgemeine Informationen verfügbar und das Produkt wird nicht näher beschrieben.

<sup>7</sup>Mit der Version 6.0 bietet Wordfast eine eigenständige Anwendung.

## 4.2 EBMT-Methoden in CAT-Systemen

Etliche CAT-Systeme integrieren EBMT-Methoden, um ihre Retrieval-Leistung zu verbessern.

Déjà Vu bietet eine eigene Funktion namens EBMT<sup>8</sup>, welche unter bestimmten Bedingungen eingreift:

Das Ausgangssegment im Translation Memory und das im Text unterscheiden sich lediglich durch einen Terminus. Beispiel: „Hauptschalter ausschalten“ und „Näherungsschalter ausschalten“.

Beide Termini sind in der Terminologiedatenbank und verfügen über eine Übersetzung in die Zielsprache: „main switch“ und „proximity switch“.

Im Translation Memory ist schon ein Fuzzy-Match für das Ausgangssegment vorhanden: „Switch off the main switch“.

Unter diesen Bedingungen ersetzt Déjà Vu automatisch die Übersetzung des alten Terminus im Fuzzy-Match durch die Übersetzung des neuen Terminus: „Switch off the proximity switch“. Dadurch wird das Fuzzy-Match zu einem Perfect Match. Dabei wird jedoch keine grammatische Anpassung vorgenommen, welche unter Umständen notwendig sein könnte.

Genau über die gleiche Funktion wie EBMT von Déjà Vu verfügen auch Transit – sie wird aber *Terminologie aktualisieren* benannt – sowie Heartsome – mit dem Namen *Quick Translation*.

In Déjà Vu ist EBMT ein Teil der Funktion *Assemble*, die Subsegmente zusammenführen kann, um die Übersetzung eines Segmentes anzubieten.

Über eine mit *Assemble* vergleichbare Funktion verfügt MemoQ: Die Übersetzung kann ebenfalls aus Fragmenten zusammengeführt werden, wenn für das ganze Segment keine Entsprechung gefunden werden konnte.

Dies zeigt, dass die Zusammenführung unterschiedlicher Technologien in eine Anwendung ebenfalls möglich ist.

## 5 Prozesse

Die Prozessintegration von CAT-Systemen und MÜ-Systemen kann sich auf verschiedene Weise realisieren. Zunächst muss definiert werden, welche Software im Mittelpunkt des Prozesses steht: Im vorliegenden Artikel werden die CAT-Systeme betrachtet, möglich sind aber auch die MÜ-Systeme. Für eine nähere Beschreibung siehe Geldbach and Seewald-Heeg (2006).

<sup>8</sup>Der Begriff EBMT wird im Benutzerhandbuch von Déjà Vu verwendet. Auf die Frage, ob diese Benennung für diese Funktion optimal ist, wird nicht eingegangen.

Selbst wenn das CAT-System im Mittelpunkt steht, sind unterschiedliche Prozesse denkbar. In diesem Artikel werden zwei Varianten beschrieben. Die Qualität der gelieferten maschinellen Übersetzung wird nicht thematisiert.

### 5.1 Sukzessive Bearbeitung

In diesem Prozess bearbeiten das CAT- und das MÜ-System die Dateien sukzessive. Eine gleichzeitige Bearbeitung durch die Interaktion von beiden Systemen ist nicht möglich.

Im Allgemeinen lässt sich der Prozess folgendermaßen skizzieren: Die Ausgangsdatei wird durch das CAT-System vorbereitet. Nach der Übersetzung durch das MÜ-System erfolgen die Korrektur des maschinell übersetzten Textes sowie die etwaige Vervollständigung seitens des Humanübersetzers wieder im CAT-System, aus dem die Zieldatei erzeugt wird.

#### 5.1.1 Across

Dank der Partnerschaft mit Language Weaver Inc. bietet Across die Integration mit einem MÜ-System. Es ist zwar kein Export aus der Anwendung notwendig (vgl. 5.1.2 SDL Trados), trotzdem erfolgt die Bearbeitung durch das MÜ-System und das CAT-System in zwei getrennten Schritten.<sup>9</sup>

Die Bearbeitung von Übersetzungen erfolgt in Across projektbasiert. Ein Schritt der Projektvorbereitung ist die Vorübersetzung. Der Text wird mit dem Übersetzungsspeicher abgeglichen und die Übersetzungen oberhalb einer definierten Ähnlichkeitsgrenze – unter Tools > Profileinstellungen > crossTank > Erweiterte Einstellungen > Vorübersetzung (ab) – werden ins Dokument eingefügt. Wenn der Übersetzer mit seiner Tätigkeit beginnt, liegt das Dokument zum Teil schon in der Zielsprache vor, es sei denn, aus dem Übersetzungsspeicher kam kein einziges Match. Die Vorübersetzungsfunktion ist auch in den anderen CAT-Systemen vorhanden.

Bei der Vorübersetzung in Across kann Language Weaver für diejenigen Matches verwendet werden, die keine 100% Matches sind. Das heißt, nur die perfekten Treffer werden aus dem Übersetzungsspeicher genommen. Der Rest wird hingegen mittels Language Weaver übersetzt. Diese Übersetzungen von Language Weaver werden nicht unmittelbar in den Übersetzungsspeicher aufgenommen. Sie sollen vom Übersetzer geprüft und bestätigt werden. Ein spezielles Symbol dient zur Unterscheidung dieser Übersetzungen von jenen aus dem Übersetzungsspeicher. In der Analyse (Report) für das Dokument (oder die Dokumente) werden diese automatisch übersetzten Segmente gesondert ausgewiesen.

---

<sup>9</sup>Damit die maschinelle Übersetzung mit Language Weaver eingesetzt werden kann, müssen gewisse Vorarbeiten geleistet werden, insbesondere rund um das Trainingskorpus für das System. Eine Beschreibung dieser Vorarbeiten ist außerhalb des Rahmens dieses Artikels. Alle Informationen über Across sind von Keller entnommen.

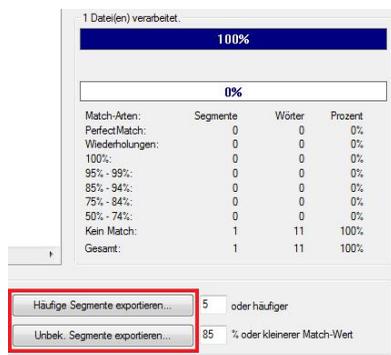


Abbildung 3: Analyse mit Export-Funktionen in SDL Trados

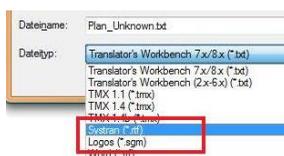


Abbildung 4: Verfügbare Exportformate in SDL Trados

### 5.1.2 SDL Trados

Das folgende Beispiel basiert auf SDL Trados Freelance, das heißt die Einzelplatzversion von SDL Trados für selbstständige Übersetzer. Bevor die Übersetzung einer Ausgangsdatei begonnen wird, ist die Analyse der Ausgangsdatei ein zwar nicht notwendiger aber üblicher Schritt. Die Analyse ist ein Abgleich zwischen der Ausgangsdatei und dem Übersetzungsspeicher und beziffert, wie viele Segmente des zu übersetzenden Textes im Übersetzungsspeicher schon enthalten sind.<sup>10</sup>

Nach Abschluss der Analyse wird in SDL Trados ein Fenster mit den Analyseergebnissen (Abbildung 3) angezeigt:

Mit den Funktionen **Häufige Segmente exportieren** und **Unbek. Segmente exportieren** lassen sich die gewünschten Segmente aus der Datei extrahieren und exportieren. Dieser Schritt löst den Text als Einheit auf. Als Dateitypen für den Export stehen unter anderem zwei MÜ relevante Formate zur Verfügung (Abbildung 4):

Die gewünschten Segmente werden in ein Format exportiert, das durch Systran bzw. Logos bearbeitet werden kann. Die Bearbeitung im jeweiligen MÜ-System kann in

<sup>10</sup>Auf eine genauere Beschreibung einer Analyse wird verzichtet. Sie kann, je nach Konfiguration und CAT-System, auch weitere Informationen liefern wie z.B. interne Wiederholungen.

```

<TrU>
<CrD>15122005, 09:34:52
<CrU>MT!
<Att L=Kunde>
<Txt L=Auftragsnummer>
<Seg L=DE-CH>Stabiles und sicheres
<Seg L=IT-CH>Strumento di lavoro st
</TrU>
    
```

Abbildung 5: Attribut-Wert maschinell übersetzter Segmente in Translator's Workbench

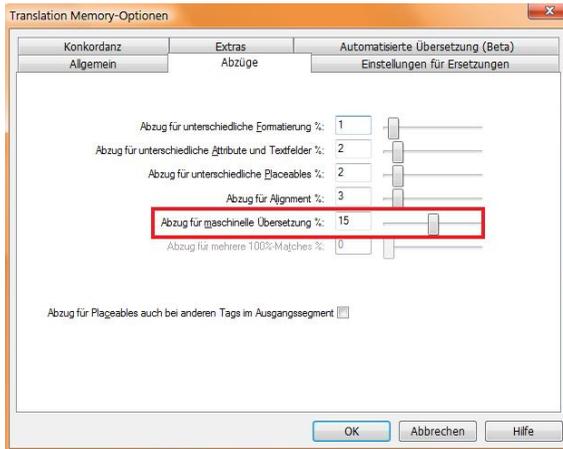


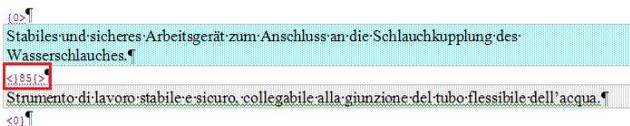
Abbildung 6: Abzüge für maschinell übersetzte Segmente in Translator's Workbench

diesem Artikel nicht behandelt werden und kann im Detail in der Dokumentation des jeweiligen Programms nachgeschlagen werden. Für einen Überblick über Systran siehe Geldbach and Seewald-Heeg (2006).

Nach der Übersetzung im MÜ-System werden die Segmente in den Übersetzungsspeicher importiert und dabei automatisch mit einem speziellen Attributwert versehen. Der Import erfolgt über Datei > Import im Hauptfenster von Translator's Workbench. Das Attributfeld heißt CrU (Creation User) und bekommt den Wert MT! (Machine Translation), siehe Abbildung 5.

Die Übersetzungseinheiten aus dem MÜ-System können mit einem speziellen Abzug versehen werden, um zum Beispiel die ungeprüfte Übernahme zu vermeiden. Dafür ist es erforderlich, Optionen > Translation Memory-Optionen > Abzüge in Translator's Workbench auszuwählen (Abbildung 6).

Alle Abzüge, einschließlich Abzug für maschinelle Übersetzung, sind durch Schieber frei konfigurierbar. Wenn Translator's Workbench im Attributfeld CrU den Wert



**Abbildung 7:** Match aus MÜ-System

MT! findet, wird der Abzug angewendet. Dies hat wichtige Folgen für die Übersetzung. Verschiedene Szenarien sind vorstellbar:

1. Der zu übersetzende Text ist gleich dem im Übersetzungsspeicher vorhandenen Text.
2. Der zu übersetzende Text ist dem im Übersetzungsspeicher vorhandenen Text ähnlich.
3. Der zu übersetzende Text ist im Übersetzungsspeicher nicht vorhanden.

Das letzte Szenario wird nicht näher beschrieben und ist uninteressant, da der Abzug irrelevant ist. Kein Ähnlichkeitswert kann dabei vermindert werden.

Das erste Szenario würde im Normalfall ein 100% Match zurückgeben, weil kein Unterschied im Text den Ähnlichkeitswert vermindert. Durch den Attributwert MT! und die Abzugseinstellung von 15%, wird hingegen nur ein Fuzzy-Match (85%) angeboten (Abbildung 7):

Der graue Hintergrund ist eine Besonderheit derjenigen Fuzzy-Matches, für die ein Abzug wegen ihres Ursprungs aus einem MÜ-System angewendet wurde. Sonst wäre der Hintergrund hellgelb.<sup>11</sup>

Das zweite Szenario unterscheidet sich nur geringfügig vom ersten. Der Ähnlichkeitswert des Fuzzy-Match, zum Beispiel 93%, wird durch den Abzug zusätzlich vermindert und wird 78%. Sollte der Ähnlichkeitswert durch den MÜ-Abzug unter die Mindestähnlichkeitsgrenze fallen, die vom Übersetzer unter Optionen > Translation Memory-Optionen > Allgemein > Minimaler Match-Wert in Translator's Workbench eingestellt worden ist, wird das Segment als No Match angeboten.

Nachdem das Segment vom Übersetzer überprüft und gegebenenfalls überarbeitet worden ist, wird es bestätigt und im Übersetzungsspeicher gesichert. Ein zusätzliches Attributfeld wird hinzugefügt, ChU (Change User) und bekommt als Wert die Identifikationskennung (User ID) des Übersetzers (Abbildung 8):

Wenn der Ausgangstext wieder in einem zu übersetzenden Text vorkommt, wird er als 100% Match angeboten. Der Abzug greift nicht mehr. Auf diese Weise werden

<sup>11</sup>Standardmäßige Farbeinstellungen, welche angepasst werden können.

```

<TrU>
<CrD>15122005, 09:34:52
<CrU>MTI
<ChD>25062008, 21:24:07
<ChU>DA
<Att L=Runde>
<Txt L=Auftragsnummer>
<Seg L=DE-CH>Stabiles und siche
<Seg L=IT-CH>Strumento di lavor
</TrU>
    
```

Abbildung 8: Attributfeld Change User

Übersetzungen aus dem MÜ-System nach Überprüfung wie alle anderen Übersetzungen behandelt.

## 5.2 Gleichzeitige Bearbeitung

Die Dateien werden vom MÜ-System und vom CAT-System gleichzeitig bearbeitet. Eine Interaktion von beiden Systemen ist möglich.

### 5.2.1 SDL Trados

Ab der Version 8.3 bietet SDL Trados Freelance eine Beta-MÜ-Funktion, die über das Internet arbeitet.<sup>12</sup> Diese Funktion kann über Optionen > Translation Memory-Optionen > Automatisierte Übersetzung (Beta) aktiviert werden, siehe Abbildung 9.

Translator's Workbench kontaktiert den SDL Automated Translation Server, der eine maschinelle Übersetzung zurückliefert. Sie wird ins Zielsegment zur weiteren Bearbeitung hinzugefügt. Im Gegensatz zu Wordfast (5.2.2) erfolgt die maschinelle Übersetzung in jedem Fall über das Internet. Auf Vor- und Nachteile dieser Möglichkeit, auch im Hinblick auf Datensicherheit, wird hier nicht eingegangen.

Die maschinelle Übersetzung kann schon bei der Vorübersetzung zum Einsatz kommen, vorausgesetzt, dass der Match-Wert für zu übersetzende Segmente (unter Extras > Übersetzung in Translator's Workbench) unterhalb 100% eingestellt ist. In diesem Fall wird für diejenigen Segmente, die kein 100% Match oder Fuzzy-Match haben, eine automatische Übersetzung hinzugefügt. Diese Segmente werden im Vorübersetzungsbericht nicht gesondert ausgezeichnet und gelten als unübersetzt. In der zweisprachigen Datei erscheinen sie aber als 1% Match, siehe Abbildung 10.

Die maschinelle Übersetzung kann außerdem während der interaktiven Übersetzung zum Einsatz kommen. Sie greift nur dann ein, wenn für ein Ausgangssegment weder ein 100% Match noch ein Fuzzy-Match gefunden werden konnte.

<sup>12</sup>Diese Funktion steht nicht für alle Sprachpaare zur Verfügung.

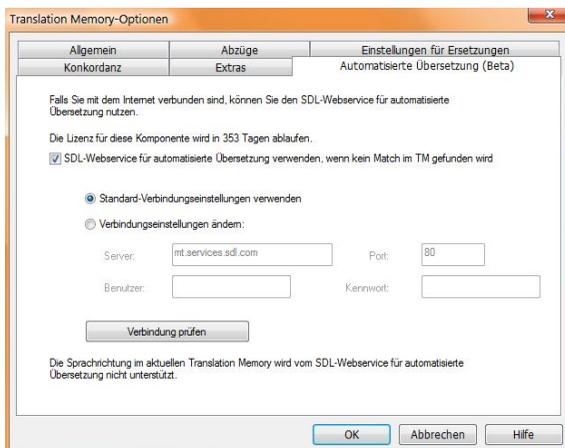


Abbildung 9: SDL Trados Automated Translation

TU **cd** **bold="on">** **Näherungsschalter** </b> </td>, auch Näherungsinitiator oder Annäherungsschalter genannt, verwenden Sensoren die auf Annäherung, das heißt ohne direkten Kontakt berührungsfrei reagieren. ① approximation switch, also approximation initiator or approach switch named, use -freely- react sensors that to approach, that is without direct contact contacts. < /td> TU TU Näherungsschalter werden bei technischen Prozessen zur Positionserkennung von

Abbildung 10: Vorübersetzter Text



Abbildung 11: SDL Trados MÜ-Match



Abbildung 12: Wordfast Symbolleiste in MS Word

Die maschinelle Übersetzung im Zielsegment wird im Editor durch einen grauen Hintergrund gekennzeichnet. Darüber hinaus werden in Translator's Workbench folgende Informationen angezeigt, siehe Abbildung 11: Der Status *Automatisch übersetzt* in der Statusleiste; Der Fuzzy-Wert von 1%.

### 5.2.2 Wordfast

Das CAT-System Wordfast ist in MS Word integriert.<sup>13</sup> Wordfast besteht aus einer Word-Vorlage (*wordfast.dot*), welche eine Sammlung von Makros beinhaltet und als Add-In verwendet wird. Ein Add-In ist ein „Programm zum Hinzufügen von [...] Befehlen oder Features“ (WORD 2003) in ein Hauptprogramm.

Über solch ein Add-In für MS Word verfügen auch diverse MÜ-Systeme, welche sonst auch als unabhängige Anwendung arbeiten, zum Beispiel *translate pro*, *Systran*, *Personal Translator*, *@prompt*, *T1*. Dies ist die Voraussetzung für die gleichzeitige Bearbeitung mit Wordfast.

Wordfast verwendet MS Word als Editor, wobei Formate aus anderen Programmen ebenfalls bearbeitet werden können.<sup>14</sup> Wenn Wordfast korrekt installiert worden ist, erscheint in MS Word eine zusätzliche Symbolleiste (Abbildung 12).<sup>15</sup>

Bevor die Übersetzung begonnen werden kann, müssen die Aufrufparameter für das jeweilige MÜ-System in *Wordfast > Setup > MT* definiert werden (Abbildung 13):

<sup>13</sup>In der Version 6.0 von Wordfast ist ein proprietärer Editor verfügbar. Da zum Redaktionsschluss nur eine Pre-release Version verfügbar war, und da das Add-In-Konzept parallel beibehalten wird, wird im vorliegenden Artikel die Version 5.5 beschrieben.

<sup>14</sup>Weitere Formate, die Wordfast 5.5 direkt bearbeiten kann, sind MS Excel, MS PowerPoint, MS Access und getaggte Formate (zum Beispiel aus SDL Trados S-Tagger). HTML, SGML, XML können nur mit +Tools bearbeitet werden. Mit der Version 6 wird die Formatsunterstützung ausgeweitet.

<sup>15</sup>Diese Beschreibung bezieht sich auf Wordfast 5.5 in Verbindung mit MS Word 2003.

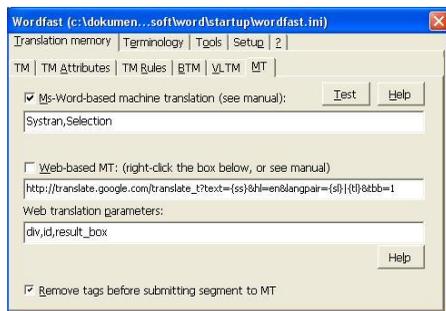


Abbildung 13: MÜ-System in Wordfast

SYSTEM	ABZUG
Across	Nein, aber spezielles Symbol
SDL Trados	Ja
Transit	Ja
Wordfast	Ja

Tabelle 4: CAT-Systeme und Abzug

In Abbildung 13 sind beispielsweise die Aufrufparameter von Systran angegeben. außerdem können auch im Internet verfügbare automatische Übersetzungsdienste verwendet werden (Option *Web-based MT*). Die Suche nach der Übersetzung erfolgt in folgender Reihenfolge: Zunächst wird der Wordfast-Übersetzungsspeicher abgefragt. Wird kein Treffer gefunden, wird eine Übersetzung vom MÜ-System angefordert und in das Dokument eingefügt.

### 5.3 Auszeichnung im Fokus

Wie in den Kapiteln 4.1, 5.1.1 sowie 5.1.2 bereits erwähnt, verwenden diverse CAT-Systeme Abzüge für die Treffer aus MÜ-Systemen bzw. heben sie gesondert hervor. Damit wird der Übersetzer auf sie hingewiesen und kann sie entsprechend prüfen. Eine Übersicht der Kennzeichnungsmöglichkeiten liefert Tabelle 4.

Across, das sonst ebenfalls einen Ähnlichkeitswert für Fuzzy-Matches angibt, verzichtet auf einen Abzug und zeichnet die Treffer mit einem speziellen Symbol aus.

## 6 Vorteile und Nachteile

Die Integration von CAT- und MÜ-Systemen bietet wichtige Vorteile der maschinellen Übersetzung. In erster Linie können damit größere Übersetzungsvolumina im Vergleich zur Humanübersetzung bearbeitet werden.

Beachtliche Kosteneinsparungen können ebenfalls erzielt werden. Selbst wenn maschinell übersetzte Texte von Humanübersetzern geprüft werden sollen, werden sie in der Regel als Fuzzy-Match (SDL Trados) oder als Sondertreffer (Across) angeboten. Dafür sind Preisstaffelungen möglich, die günstiger sind als eine Neuübersetzung.

Diesen Vorteilen stehen jedoch auch Nachteile gegenüber. Wenn der Korrekturaufwand für die Übersetzer hoch ist, könnte die beabsichtigte Prozessbeschleunigung nicht erzielt werden. Dabei zeigen Vergleichstests, dass die Korrektur einer schlechten und insbesondere einer mittelmäßigen (maschinellen) Übersetzung mehr Zeit in Anspruch nimmt als eine Neuübersetzung. Für eine detaillierte Behandlung dieses Themas siehe Krings (1998). Der erhöhte Zeitaufwand kann in höheren Kosten resultieren.

Das entscheidende Kriterium für den erfolgreichen Einsatz der kombinierten Lösungen (MÜ-System und CAT-System) ist also die Qualität der maschinellen Übersetzung. Dafür ist ein entsprechender nicht unerheblicher Aufwand einzubringen, wie zum Beispiel in Trojanus (2002) und Geldbach and Seewald-Heeg (2006) beschrieben. Eine unbedachte Integration kann hingegen die Erwartungen enttäuschen.

### Literatur

Atril (2003). Madrid.

Carl, M. and Way, A. (2003). Introduction. In *Recent Advances in Example-Based Machine Translation*, pages XVII–XXXI. Kluwer Academic Publisher.

Doug, A. (2003). Why translation is difficult for computers. pages 119–142. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Eberle, K. (2006). Maschinelle Übersetzung - Hopp oder top? In der Dolmetscher und Übersetzer e.V., B., editor, *MDÜ - Fachzeitschrift für Übersetzer*, volume 4 of 9-15. BDÜ.

Geldbach, S. and Seewald-Heeg, U. (2006). MÜ ist so gut wie ihre Funktionalität, Prä- und Postedition. In *MDÜ - Fachzeitschrift für Übersetzer*, pages 9–15. BDÜ.

Hutchins, J. (2003). Commercial Systems: The state of the Art. In Somers, H., editor, *Computers and Translation: a translator's guide*, pages 161–174. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Keller, N. Die Anbindung des MÜ-Systems Language Weaver an Across. Unveröffentlichter Vortrag beim Workshop "Maschinelle Übersetzung – Von der Theorie zur Anwendung" der GLDV, AK Maschinelle Übersetzung. Köthen, Hochschule Anhalt, 4.07.2008.

Krings, H. (1998). *Texte reparieren*. Gunter Narr Verlag.

- Lagoudaki, E. (2006). Translation Memories Survey 2006: Enlightning users' perspective.
- Lange, C. A. and Bennett, W. S. (2000). Combining Machine Translation with Translation Memory at Baan. In Sprung, R. C., editor, *Translating into Success*, pages 203–218. John Benjamins Publishing Company, Amsterdam, Philadelphia.
- Massion, F. (2005). *Translation Memory Systeme im Vergleich*. Doculine Verlag, Reutlingen.
- Massion, F. (2008). Integration durch Standards. In *Produkt Global*, pages 22–25. Hüthig, Heidelberg.
- Microsoft (2003). Online-Hilfe von Microsoft Word.
- Reinke, U. (2003). *Translation Memories: Systeme – Konzepte – Linguistische Optimierung*. Peter Lang Verlag.
- Seewald-Heeg, U. (2007). Vielfalt auf dem Markt. 4:12–25.
- Somers, H. (2003a). An overview of EBMT. In *Recent Advances in Example-Based Machine Translation*, pages 3–57. Kluwer Academic Publisher.
- Somers, H. (2003b). Translation memory systems. pages 31–47. John Benjamins Publishing Company, Amsterdam, Philadelphia.
- StarAG (2008). *Transit NXT - Benutzerhandbuch (Vorabversion)*. Ramsen.
- Trojanus, K.-H. (2002). Anspruch und Wirklichkeit. In *MDÜ - Fachzeitschrift für Übersetzer*, pages 19–24. BDÜ.
- Trujillo, A. (1999). *Translation Engines: Techniques for Machine Translation*. Springer Verlag, London.