

2002

Lokalisierung und Maschinelle Übersetzungssysteme

Uwe Muegge, *Monterey Institute of International Studies*

Uwe Mügge

Lokalisierung und Maschinelle Übersetzungssysteme

1 Einleitung

Maschinelle Übersetzungssysteme werden seit über 50 Jahren entwickelt und die Geschichte dieser Technologie ist ausgesprochen bewegt. Befürworter und Kritiker der Maschinellen Übersetzung stehen sich in erbitterter Gegnerschaft gegenüber und die Diskussion um die Brauchbarkeit von maschinell erstellten Übersetzungen wird mitunter hitzig und emotional geführt. Die Thesen beider Seiten standen und stehen sich diametral gegenüber: So behaupten die Einen, dass Maschinelle Übersetzungssysteme in der Lage sind, vollautomatisch hochwertige Übersetzungen (*Fully-Automatic High-Quality Translation, FAHQT*) zu erstellen, während die Anderen die prinzipielle Unmöglichkeit eines solchen Unterfangens annehmen.

Ich möchte diese Diskussion hier nicht nachzeichnen, das ist an anderer Stelle bereits ausführlich geschehen (z.B. in Hutchins 1999, Rico 2001, Weber 1998). Mein Anliegen ist vielmehr zu zeigen, dass der Einsatz von Maschinellen Übersetzungssystemen bei der Lokalisierung von Technischen Dokumenten sinnvoll und vorteilhaft sein kann, und zwar auch für kleine und mittelständische Unternehmen. Außerdem will ich die Bedingungen darstellen, unter denen ein solcher Einsatz Erfolg versprechend ist.

Um es gleich vorweg zu nehmen: Wenn eine Technische Redaktion bereits Dokumente nach Verständlichkeitskriterien erstellt und die dahin gehenden Empfehlungen der tecom (vgl. Gabriel u. a. 1995, Gabriel u. a. 2001) sowie die Forderungen der einschlägigen DIN-Normen (z.B. DIN EN 62079:2001-11 2001, DIN 2345:1998-04 1998) erfüllt, dann ist es zum effizienten Einsatz von Maschinellen Übersetzungssystemen nur noch ein relativ kleiner Schritt, allerdings mit der wichtigen Einschränkung, dass die Maschinellen Übersetzungssysteme den Menschen nicht ersetzen. Das ist selbst in den komplexen Installationen in der Großindustrie bislang nicht gelungen (vgl. Gereke-Bornemann u. a. 2000, Lockwood 2000). Im

hier skizzierten Szenario stelle ich vielmehr dar, wie Maschinelle Übersetzungssysteme hoch qualifizierte Übersetzende in die Lage versetzen, erheblich produktiver und konsistenter zu arbeiten. Mit anderen Worten: Maschinelle Übersetzungssysteme erstellen eine Rohübersetzung, die dann von menschlichen Übersetzenden den jeweiligen Qualitätsvorgaben für den Zieltext entsprechend nachbearbeitet wird.

2 Was sind Maschinelle Übersetzungssysteme?

2.1 Definition

Ein Maschinelles Übersetzungssystem ist ein Computerprogramm, das weit gehend selbstständig maschinenlesbare Texte von einer Ausgangssprache in eine Zielsprache übersetzt.

2.2 Aufbau

Allgemeines

Prinzipiell ist es schwierig, allgemein gültige Aussagen über Maschinelle Übersetzungssysteme zu machen, denn dazu ist dieses Gebiet zu groß und zu heterogen. Schließlich sind heute weit über 400 Maschinelle Übersetzungssysteme als kommerzielle Produkte erhältlich (vgl. z.B. Hutchins 2000, Schmitt 2000b); die Zahl der rein im Forschungsbetrieb eingesetzten Systeme dürfte ebenfalls erheblich sein. Außerdem gibt es nicht eine Basistechnologie für Maschinelle Übersetzungssysteme, sondern gleich eine ganze Palette:

- Regelbasierte (Transfer-) Systeme
- Wissensbasierte Systeme
- Beispielbasierte Systeme
- Statistische Systeme
- Interlinguasysteme
- Direktsysteme.

Zu den einzelnen Übersetzungstechnologien/-philosophien möchte ich hier nur sagen, dass es für die verschiedensten Einsatzbereiche ganz unterschiedliche Lösungen gibt. Eine detaillierte Diskussion der technologischen Grundlagen der verschiedenen Maschinellen Übersetzungssysteme ist der Fachliteratur zu entnehmen (z.B. Trujillo 1999). Im Folgenden möchte ich mich auf die Dar-

stellung der Regelbasierten Systeme beschränken, die im übrigen (zumindest im europäischen Sprachraum) die populärsten Vertreter Maschineller Übersetzungssysteme stellen (z.B. *Langenscheidts T1*, *linguatec Personal Translator*, *Systran*).

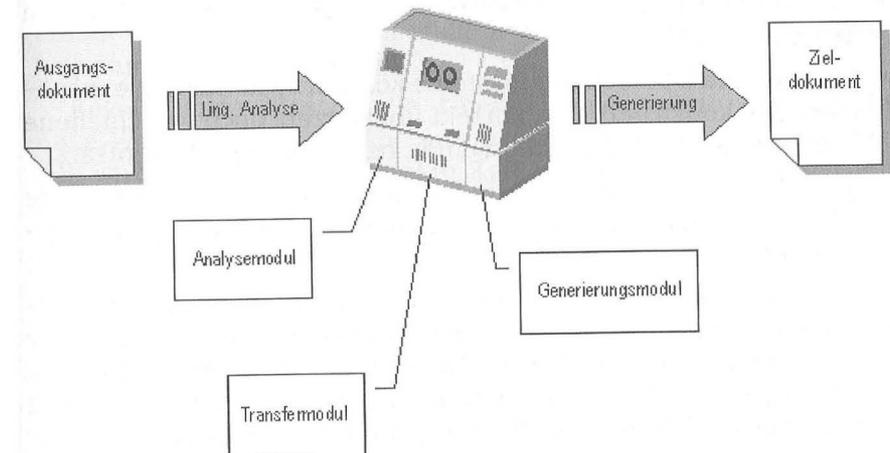


Bild 1: Aufbau Regelbasierter Maschineller Übersetzungssysteme

Regelbasierte Systeme bestehen aus drei Modulen:

- Das Analysemodul *parst* (analysiert) den Ausgangstext satzweise in Zusammenarbeit mit dem Wörterbuch. Im Lauf der Analyse erkennt dieses Modul Wortarten und -stämme sowie die Beugung der einzelnen Satzbestandteile und erstellt eine abstrakte Strukturbeschreibung des Ausgangssatzes.
- Das Transfermodul überträgt die Strukturbeschreibung nach produktspezifischen Regeln in eine entsprechende Strukturbeschreibung für die Zielsprache und fügt in Zusammenarbeit mit dem Wörterbuch die benötigten zielsprachlichen Wörter ein.
- Das Generierungsmodul hat die Aufgabe, aus der Strukturbeschreibung einen wohlgeformten Satz in der Zielsprache zu erzeugen, indem es den Regeln entsprechend Wörter beugt und deren Stellung im Satz anpasst.

Regelwerk

In den einzelnen Modulen dieser Maschinellen Übersetzungssysteme sind sprachrichtungsspezifische Regeln implementiert. Diese

Regelwerke können im Allgemeinen nur mit speziellen Kenntnissen modifiziert bzw. erweitert werden. Ich möchte deshalb an dieser Stelle nicht in die technischen Details gehen, sondern vielmehr auf einige allgemein gültige Prinzipien hinweisen.

- Maschinelle Übersetzungssysteme analysieren und übersetzen Texte auf Satzbasis.
- Maschinelle Übersetzungssysteme können zwar nicht alle möglichen Satzkonstruktionen korrekt übersetzen, doch verschiedene Produkte beherrschen schon einen relativ großen Prozentsatz.
- Je einfacher und gängiger die Satzkonstruktion, desto höher die Wahrscheinlichkeit, dass ein Satz korrekt analysiert und übersetzt wird.
- Besonders hohe Erfolgsquoten erzielen Maschinelle Übersetzungssysteme beim Übersetzen von Ausgangstexten, die in einer so genannten ‚Kontrollierten Sprache‘ erstellt wurden. Kontrollierte Sprachen sind im Vergleich zur Allgemeinsprache in Bezug auf zulässige Wörter und Satzkonstruktionen mehr oder weniger stark eingeschränkt (vgl. z.B. Lehrndorfer 1996, Mügge 2002b).
- Bei Satzkonstruktionen, für die im Maschinellen Übersetzungssystem keine Regeln hinterlegt sind, ist die Wahrscheinlichkeit besonders hoch, dass entsprechende Sätze nicht korrekt übersetzt werden.

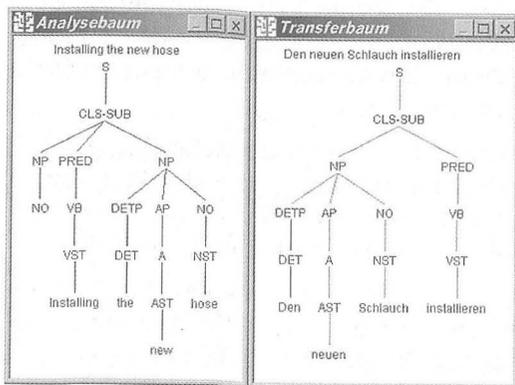


Bild 2: Beispiel für Analyse und Transfer nach der Phrasenstrukturgrammatik

Wörterbuch

Das Wörterbuch spielt beim Arbeiten mit Maschinellen Übersetzungssystemen eine entscheidende Rolle. Es liegt auf der Hand,

dass ein Maschinelles Übersetzungssystem nur dann ein Wort richtig erkennen und übersetzen kann, wenn die entsprechenden Wortpaare im Wörterbuch hinterlegt sind. Weniger bekannt dürfte sein, dass ohne vollständiges Wörterbuch auch die Analyse der Satzkonstruktion oft fehl schlägt, denn wenn ein Substantiv oder Verb in einem Satz nicht als solches erkannt werden kann, ist die Satzanalyse zum Scheitern verurteilt.

Regelbasierte Maschinelle Übersetzungssysteme sind in der Lage, Wortformen in der Übersetzung korrekt zu beugen, d. h., dass z.B. Verbformen in die korrekte Zeit und Person gesetzt werden. Damit die Beugung funktionieren kann, benötigen Maschinelle Übersetzungssysteme umfassende linguistische Informationen zu jedem Wort. Viele der heute auf dem Markt angebotenen Maschinellen Übersetzungssysteme verfügen über umfangreiche Wörterbücher der Allgemeinsprache und je nach Produkt auch über mehr oder weniger umfassende Wörterbücher zu verschiedenen Sachgebieten. Wenn bei einem Übersetzungsprojekt eine andere als die im Systemwörterbuch vorhandene Entsprechung gewünscht wird oder noch gar kein Eintrag vorhanden ist, muss das Systemwörterbuch bearbeitet oder ein Benutzerwörterbuch erstellt werden, wenn man im Zieltext nicht immer wieder dieselben Fehler korrigieren will. In diesem Zusammenhang sind zwei Herausforderungen zu bewältigen: das Erkennen unbekannter Terminologie und der Import bestehender Terminologiebestände.

Wie bereits oben erwähnt, benötigen Maschinelle Übersetzungssysteme ein vollständiges Wörterbuch, um Texte erfolgreich übersetzen zu können. Damit unvollständige Wörterbücher ergänzt werden können, bieten die gängigen Maschinellen Übersetzungssysteme Funktionen zur Erkennung unbekannter Terminologie. Diese Funktion ist in der Regel wortbasiert, weshalb die Erkennung unbekannter Terminologie in manchen Sprachen sehr gute, in anderen dagegen unbefriedigende Ergebnisse liefert.

Damit auch in den Fällen, in denen die integrierte Terminologieerkennung eines Maschinellen Übersetzungssystems keine befriedigenden Resultate liefert, umfassende Wörterbücher schnell und kostengünstig aufgebaut werden können, ist der Einsatz externer Werkzeuge für die Terminologieextraktion sinnvoll. Inzwischen ist eine ganze Reihe von Produkten und Dienstleistungen in diesem Bereich verfügbar. Bild 3 zeigt das Ergebnis eines von mir entwickelten Extraktionsverfahrens, das neben dem Terminologiekandi-

Term Candidate	Part of Speech	Example Sentence
auto	adj	Auto synchronization of billing plans makes it easier to handle the billing changes required by a new sales document.
automate	verb	The process of changing the original license and creating a new one is automated using a function called a Serial Number Split.
autosynchronization indicator	noun	'AutoSynch' (autosynchronization indicator)
availability period	noun	Define contract availability periods in Customizing for 'IS-HT-SW' by choosing Assign availability period to sales document type.
background job	noun	A list of order documents can be created from which from the documents can be processed individually, or selected for synchronization via a background job.
background process	noun	'VCP' is a background process that automatically creates customer orders for new product licenses as soon as they are available.
basic	adj	Basic business information can be maintained directly in the 'BASE'.
basic function	noun	You can maintain these by choosing Basic Functions 'Listing/Exclusion' Customizing.
bill of material usage type	noun	Enter B as bill of material usage type to represent the bundle.
billing activity	noun	subsequent billing activity
billing block	noun	In addition, you can fill in the Billing block field if you want to activate a billing block against the credit lines created by the system for periods that have already been billed.

Bild 3: Beispiel einer automatischen Terminologieextraktion

daten auch die Wortart und einen Beispielsatz automatisch extrahiert (vgl. Mügge 2002a). Die mit einem derartigen Verfahren erstellten Extraktionslisten sind eine ausgezeichnete Grundlage für das Erstellen umfassender Benutzerwörterbücher für Maschinelle Übersetzungssysteme.

Bei vielen Maschinellen Übersetzungssystemen ist der Import bestehender Terminologiebestände relativ einfach zu bewerkstelligen. Eine wichtige Voraussetzung für den Terminologieimport ist allerdings, dass zu jedem Eintrag in der bestehenden Terminologie die Wortart angegeben ist. Aus diesem Grund sollte diese Eintragskategorie in keiner Terminologiedatenbank fehlen. Außerdem ist beim Terminologieimport in Maschinelle Übersetzungssysteme zu beachten, dass es bei verschiedenen Produkten z. T. gravierende Einschränkungen zu beachten gilt. Hier einige Beispiele für Einträge, die u. U. nicht ohne weiteres in ein Benutzerwörterbuch importiert werden können:

- Wörter, die bereits im Systemwörterbuch hinterlegt sind
- Wörter, die eine Ziffer als Bestandteil haben
- Einträge, die aus mehreren Wörtern bestehen, insbesondere Redewendungen
- Bestimmte Wortarten und Beugungsformen.

Wie oben bereits erwähnt, benötigen Maschinelle Übersetzungssysteme detaillierte linguistische Informationen zu jedem Wörterbucheintrag, damit Wörter in der Übersetzung korrekt gebeugt werden können. Selbst wenn in einer bestehenden Terminologiedatenbank entsprechende Informationen hinterlegt sind, können derartige Daten heute in der Regel nicht an die gängigen Maschinellen Übersetzungssysteme weiter gegeben werden, weil es dafür noch keine geeignete Plattform gibt – das vom TQPro-Konsortium entwickelte Austauschformat OLIF wird noch nicht auf breiter Basis unterstützt (vgl. Thurmair 2001). Die nahe liegende Konsequenz wäre, alle importierten Einträge im Wörterbucheditor des Maschinellen Übersetzungssystems manuell zu vervollständigen. Die Komplexität der Einträge und das dazu notwendige linguistische Wissen ist von Produkt zu Produkt unterschiedlich.

Ein weiteres charakteristisches Merkmal der Wörterbücher von Maschinellen Übersetzungssystemen ist die Möglichkeit, jedem Eintrag ein Sachgebiet sowie Kunden- und Produktangaben hinzuzufügen. Diese Angaben ermöglichen es, identische Ausgangssprachliche Wörter in unterschiedlichen Kontexten jeweils anders zu übersetzen. So könnte z.B. festgelegt werden, dass die Entsprechung für den deutschen Term *Festplatte* im Englischen für Kunde A *disk drive* ist, während für Kunde B stets *hard disk* erzeugt wird.

Die Wörterbucharbeit in Maschinellen Übersetzungssystemen ist in jedem Fall mit einem erheblichen Zeitaufwand verbunden. Wenn dieser Zeitfaktor möglichst minimiert und gleichzeitig eine möglichst hohe Übersetzungsqualität erzielt werden soll, dann hat sich in der Praxis folgende Strategie bewährt:

- Nur bei Wörtern, die im Zieltext mehrfach (z.B. ab drei Mal) falsch gebeugt werden, nehmen die Übersetzenden korrigierende Einstellungen im Wörterbucheditor des Maschinellen Übersetzungssystems vor.
- Bei Wörtern, die im Zieltext nur vereinzelt falsch gebeugt werden, erfolgt die Korrektur beim Nachbearbeiten durch menschliche

che Übersetzende (dem so genannten Postediting), idealerweise in einer Translation-Memory-Umgebung.

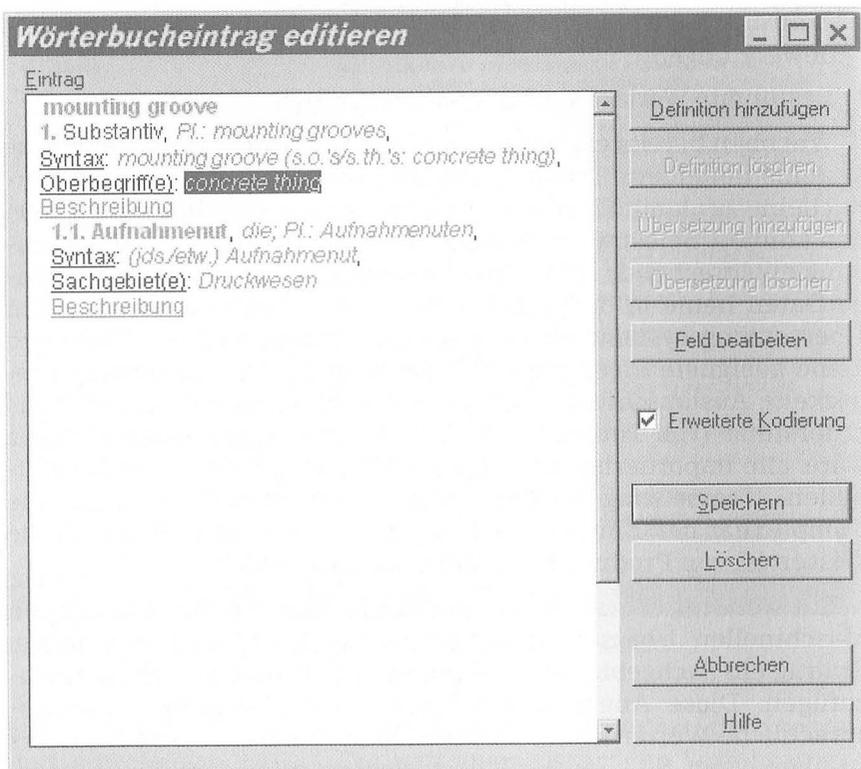


Bild 4: Wörterbucheintrag im Benutzerwörterbuch (linguaterc Personal Translator)

3 Stärken Maschinelles Übersetzungssysteme

Wenn Aussagen zur Leistungsfähigkeit Maschinelles Übersetzungssysteme gemacht werden sollen, muss zunächst deutlich zwischen zwei Varianten der Regelbasierten Maschinelles Übersetzungssysteme unterschieden werden: den auf einem PC oder einem Server im Intranet installierten Produkten und den (meist kostenlosen) Übersetzungsdiensten im Internet. Es ist richtig, dass bei beiden Varianten oft dieselbe Technologie zum Einsatz kommt – der älteste und bekannteste Internetübersetzungsdienst *Babelfish* (www.babelfish.com) basiert z.B. auf *Systran* – und doch gibt es

schwer wiegende, systembedingte Unterschiede zwischen diesen Anwendungen.

Die auf Maschinelles Übersetzungssystemen basierenden Übersetzungsdienste im Internet sind auf hohen Datendurchsatz und schnelle Rückmeldung ausgerichtet. Deshalb ist es bei diesen Diensten in der Regel nicht möglich, ein Sachgebiet zu definieren oder gar ein Benutzerwörterbuch für eine Übersetzung bereit zu stellen. Aufgrund der Tatsache, dass derartigen Maschinelles Übersetzungssystemen weder Kontextinformation noch ein projektspezifisches Wörterbuch zur Verfügung steht, sind solche Systeme oftmals nur in der Lage „blanken Dadaismus“ zu produzieren, um es mit den Worten des Übersetzungswissenschaftlers Peter Schmitt zu sagen (vgl. Schmitt 2000, 231).

Meiner Meinung nach sind die auf Maschinelles Übersetzungssystemen basierenden Übersetzungsdienste im Internet dafür verantwortlich, dass die Maschinelle Übersetzung insgesamt einen, vorsichtig ausgedrückt, zweifelhaften Ruf genießt. Ich betone deshalb nochmals, dass derartige Dienste lediglich über eine eingeschränkte Übersetzungsfunktionalität verfügen und deshalb nur sehr eingeschränkt dafür geeignet sind, Auskunft über die Übersetzungsqualität eines intern installierten Maschinelles Übersetzungssystem zu geben. Seriöse Aussagen über die Übersetzungsqualität Maschinelles Übersetzungssysteme können nur gemacht werden, wenn auch das volle Potenzial dieser Systeme ausgeschöpft wird, d. h., wenn es möglich ist, bei der Übersetzung das korrekte Sachgebiet zu wählen und ein projektspezifisches Wörterbuch zu nutzen.

3.1 Übersetzungsgeschwindigkeit

Es ist allgemein bekannt, dass Maschinelle Übersetzungssysteme in kurzer Zeit große Textmengen bearbeiten können; Übersetzungsleistungen von 100.000 bis über 1.000.000 Wörtern *pro Stunde* sind heute Realität (vgl. Gruber 2002, 5). Wenn man diese Zahl mit dem Richtwert von 3.000 Wörtern *pro Tag* für menschliche Übersetzende vergleicht, wird deutlich, welches enorme Potenzial in dieser Technologie steckt.

Aber wie bereits mehrfach angeklungen sind die marktgängigen Maschinelles Übersetzungssysteme nicht ohne weiteres in der Lage, typische Dokumente aus der Technischen Redaktion (z.B. Spezifikationen, Anleitungen und Hilfedokumente) so zu bearbei-

ten, dass die Zieltexte den Anforderungen an eine Informationsübersetzung (vollständig, verständlich, korrekte Terminologie), geschweige denn eine publikationsreife Übersetzung genügen. Um zu einem brauchbaren Zieldokument zu gelangen, muss in aller Regel entweder das Wörterbuch des Maschinellen Übersetzungssystems bearbeitet und/oder die Übersetzung durch menschliche Übersetzer posteditiert werden – immer vorausgesetzt, dass sich der Ausgangstext in Inhalt und Satzbau überhaupt für die Maschinelle Übersetzung eignet.

Angesichts der Vielfalt von Maschinellen Übersetzungssystemen und unterstützten Sprachkombinationen einerseits und der Heterogenität von Übersetzungsprojekten andererseits sind verlässliche und allgemeingültige Zahlenangaben für die mit Maschinellen Übersetzungssystemen mögliche Beschleunigung des Übersetzungsvorgangs kaum zu erhalten. Ich kann aus eigener praktischer Erfahrung jedoch sagen, dass bei einer Maschinellen Übersetzung eines nach Verständlichkeitskriterien erstellten Ausgangstextes (nach vorangegangener Wörterbucharbeit) ca. 3-4 Standardseiten pro Stunde zu publikationsreifer Qualität posteditiert werden können, während in der selben Zeit von einem vergleichbaren Ausgangstext ca. 1-1,5 Seiten manuell übersetzt werden. Diese Werte decken sich weitgehend mit den Angaben in der Literatur (vgl. Lehrndorfer u. a. 1997, 12).

3.2 Übersetzungskosten

Ganz gleich, ob eine Einzelperson eine PC-Lösung wie den *linguatic Personal Translator Office Plus* für ca. € 300 anschafft oder ein Unternehmen ein Client-Server-System wie den *IBM WebSphere Translation Server* für ein Vielfaches dieser Summe implementiert: Nach Installation und Schulung sind die Kosten für das Erstellen der einzelnen Übersetzung vernachlässigbar. Ins Gewicht fallen im laufenden Betrieb lediglich die Kosten für das Postediting und/oder die Wörterbucharbeit. Aufgrund dieser Tatsache können mit Maschinellen Übersetzungssystemen auch solche Dokumente übersetzt werden, für die eine menschliche Übersetzung wegen der damit verbundenen relativ hohen Kosten ausscheidet.

Ein Bereich, in dem Maschinelle Übersetzungssysteme die Kostenkarte voll ausspielen, ist die interne Kommunikation via E-Mail und *Chatting* in großen multinationalen Organisationen. In diesen beiden Medien sind vor allem hohe Geschwindigkeit, niedrige Kos-

ten pro Übersetzung und große Anzahl der unterstützten Sprachrichtungen die entscheidenden Faktoren. Der *IBM WebSphere Translation Server*, der bei der Deutschen Bank eingesetzt wird, ist ein Beispiel einer derartigen Anwendung eines Maschinellen Übersetzungssystems (vgl. Gruber 2002).

Ein weiterer aus der Kostenperspektive interessanter Einsatzbereich für Maschinelle Übersetzungssysteme ist die so genannte Informationsübersetzung. Informationsübersetzungen haben den Zweck, den Inhalt eines Dokuments vollständig und verständlich wieder zu geben. Dabei kommt es mehr auf die korrekte Terminologie als auf die korrekte Grammatik an. Die Stilistik ist von vollkommen untergeordneter Bedeutung und deshalb werden derartige Informationsübersetzungen von menschlichen Übersetzern vor allem im Bereich der Terminologie nachbearbeitet. Mit Maschinellen Übersetzungssystemen erstellte Informationsübersetzungen von im Internet erhältlichen Produktspezifikationen und Marketingmaterial von Konkurrenzunternehmen können z.B. dazu verwendet werden, um Fachabteilungen im Unternehmen außerordentlich kostengünstig mit tagesaktuellen Informationen zur Wettbewerbssituation zu versorgen.

3.3 Übersetzungsqualität

Maschinelle Übersetzungssysteme sind wegen ihrer angeblich mangelhaften Übersetzungsqualität seit Jahren Gegenstand von Häme und Spott. Bei genauer Betrachtung vieler fehlgeschlagener Übersetzungsprojekte zeigt sich jedoch oft, dass für die Maschinelle Übersetzung ungeeignetes Ausgangsmaterial (z.B. literarische Texte) verwendet sowie weder Wörterbucharbeit geleistet noch das Sachgebiet definiert wurde. Eine besonders beliebte Methode zur Demonstration der Unbrauchbarkeit Maschineller Übersetzungssysteme ist die Rückübersetzung von maschinell übersetzten Ausgangstexten – ein Test, der Fehlerquoten potenziert, da die Maschinelle Übersetzung naturgemäß nur mit fehlerfreien Ausgangstexten akzeptable Ergebnisse liefern kann. Textsammlungen, die die o. g. Methoden zur Illustration der Untauglichkeit Maschineller Übersetzungssysteme für den professionellen Übersetzungsbereich einsetzen (vgl. z.B. Domzalski u. a. 2001), haben meiner Meinung nach bestenfalls hohen Unterhaltungswert.

Richtig ist, dass mit den heute auf dem Markt verfügbaren Übersetzungssystemen in der Regel keine publikationsreifen Überset-

zungen erzeugt werden können. Es ist aber genauso richtig, dass bei geeigneten Ausgangstexten eine Maschinelle Rohübersetzung mit relativ geringem Aufwand zu einer publikationsreifen Übersetzung editiert werden kann. Bild 5 zeigt jeweils einen Ausschnitt des Ausgangs- und Zieltextes einer Wartungsanleitung. In diesem typischen Beispiel ist im nummerierten Fließtext tatsächlich kein einziger Satz fehlerfrei übersetzt worden. Damit geht der Punkt zunächst an die Kritiker. Viel wichtiger als die Frage, wie viele Sätze richtig oder falsch übersetzt wurden, ist die Frage, wie lange es dauert, um aus dieser Rohübersetzung einen publikationsreifen Text zu erstellen. Weil die meisten der hier illustrierten Fehler Wortstellungs- und Beugungsfehler sind und nur einer von sechs Sätzen komplett neu übersetzt werden muss, benötigen menschliche Übersetzende für das Postediting einer solchen Rohübersetzung erfahrungsgemäß ca. ein Viertel der Zeit, die für die rein manuelle Übersetzung auf gleicher Qualitätsstufe benötigt würde.

Ich möchte an dieser Stelle betonen, dass die Maschinelle Übersetzung nach dem Postediting nicht mehr von einer rein manuell erstellten Übersetzung zu unterscheiden ist. Und nicht nur das: Maschinelle Übersetzungen sind in mehreren Qualitätskategorien rein manuellen Übersetzungen sogar deutlich überlegen.

- Terminologische Konsistenz: Maschinelle Übersetzungssysteme kennen keine persönlichen Vorlieben und verfügen durch die

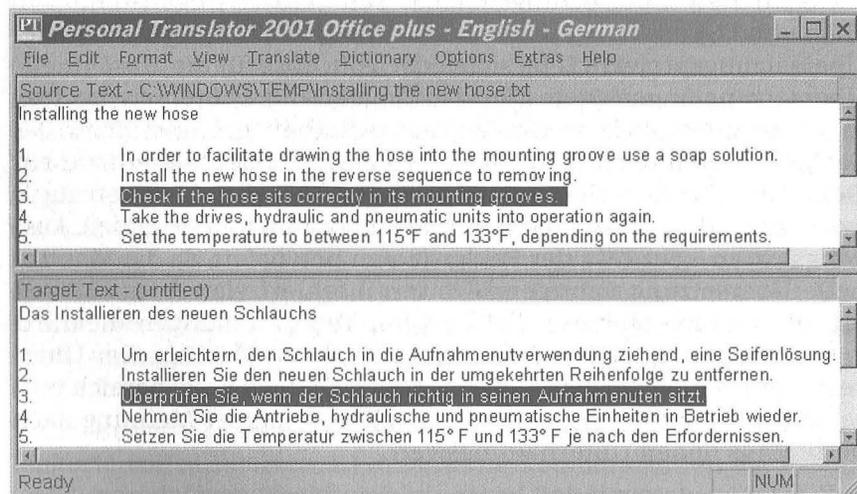


Bild 5: Beispielübersetzung einer Wartungsanleitung

Wörterbuchkomponente über ein ausgezeichnetes Gedächtnis. Deshalb übersetzen Maschinelle Übersetzungssysteme Terminologie innerhalb von Dokumenten und über Dokumente sowie lange Zeiträume hinweg gleich.

- Stilistische Konsistenz: Bei der Maschinellen Übersetzung bewältigt ein einziges System enorme Textmengen. Aufgrund der im System implementierten Regeln werden gleiche Ausgangssprachliche Satzstrukturen stets mit gleichen Zielsprachlichen Satzstrukturen übersetzt. Diese Eigenschaft garantiert die stilistische Konsistenz im Sinne einer *Corporate Language* über Dokumente und lange Zeiträume hinweg.
- Vollständigkeit: Maschinelle Übersetzungssysteme ermüden nicht und lassen sich auch von nichts und niemandem bei der Arbeit ablenken. Bei korrekter Formatierung des Ausgangstextes bietet der Einsatz eines Maschinellen Übersetzungssystems die Gewähr dafür, dass die Übersetzung zumindest auf Satzebene vollständig ist.

4 Schwächen Maschineller Übersetzungssysteme

Die bekannteste Schwäche Maschineller Übersetzungssysteme ist natürlich die, dass diese Systeme nicht in der Lage sind, alle möglichen Satzkonstruktionen und Satzbestandteile korrekt zu erkennen und zu übersetzen. Diese Schwäche kann jedoch zum großen Teil behoben werden, indem die Ausgangstexte kurz und präzise formuliert sowie projektspezifische Wörterbücher erstellt werden. Neben dieser allgemein bekannten Herausforderung gibt es verschiedene weitere Hindernisse, die den Einsatz Maschineller Übersetzungssysteme erschweren. Aber auch für diese Bereiche sind Lösungen verfügbar, wie nachfolgend gezeigt wird.

4.1 Lokalisierung

Lokalisierung bedeutet, einen Ausgangstext an die sprachlichen und *kulturellen* Anforderungen eines Marktes anzupassen. In den Bereich der kulturellen Anpassung fallen so wichtige Dinge wie z.B. das

- Konvertieren von Maßangaben und Einheiten
z.B. „20 °Celsius“ in Deutschland entsprechen „68 °Fahrenheit“ in den USA

- Konvertieren von Dezimalgruppen
z.B. „1.234,56 GB“ in Deutschland entspricht „1,234.56 GB“ in den USA
- Konvertieren von Zeit- und Datumsangaben
z.B. „05.12.2002“ in Deutschland entspricht „12/05/2002“ in den USA.

Leider unterstützen die gängigen Maschinellen Übersetzungssysteme derartige Konvertierungen nicht. Wie Beispielsatz 5 in Bild 5 illustriert, belassen gängige Maschinelle Übersetzungssysteme derartige Angaben beim Übersetzen im Original. Bei Ausgangsdokumenten, die einen hohen Anteil an Zahlenangaben aufweisen und die von einem SI-konformen (*Système International d'Unités*) Herstellungsland wie z.B. Deutschland in einen nicht-SI-konformen Markt wie z.B. die USA exportiert werden, hat dies einen erheblichen Nachbearbeitungsaufwand zur Folge. Und nicht nur das: Es besteht darüber hinaus die Möglichkeit, dass beim manuellen Postediting entweder eine notwendige Konvertierung übersehen oder fehlerhaft ausgeführt wird, was als schwerer Übersetzungsfehler zu werten wäre.

Diese gravierende Schwäche vieler Maschineller Übersetzungssysteme kann jedoch dadurch behoben werden, dass das Postediting in einem Translation-Memory-System durchgeführt wird, das die Lokalisierung von Zahlenangaben unterstützt. Bild 6 illustriert die automatische Erkennung von Maßangaben in der *Trados Workbench*. In dieser Umgebung kann in den Optionen die Konvertierung individuell eingestellt werden. Beim Übersetzen zeigt das Translation Memory automatisch solche Zahlenangaben an, die auf Tastendruck vom System konvertiert werden können.

Das Postediting von Maschinellen Übersetzungen in einem Standard Translation-Memory-System wie der *Trados Workbench* war bislang aufgrund fehlender Schnittstellen nur für wenige Maschinelle Übersetzungssysteme möglich. Ich habe ein Verfahren entwickelt, das über die Nutzung des TMX-Austauschformats die Integration jedes beliebigen Maschinellen Übersetzungssystems in die gängigen Translation-Memory-Systeme ermöglicht (vgl. Mügge 2001). Das auf diese Weise automatisierte Lokalisieren von Zahlenangaben in Translation-Memory-Systemen beschleunigt das Postediting nicht nur erheblich, dieses Verfahren ist auch entschieden zuverlässiger als das manuelle Umrechnen und Umformatieren.

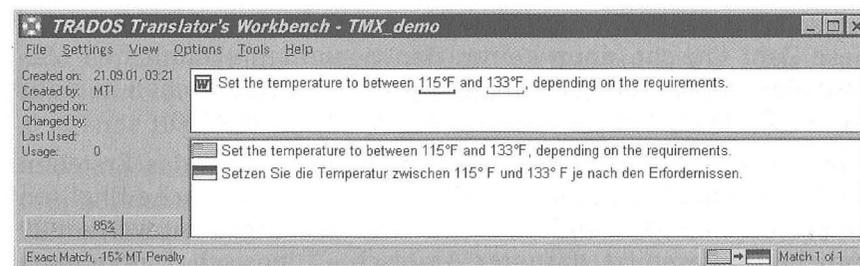


Bild 6: Automatische Erkennung von Maßeinheiten im Translation Memory

4.2 Dateiformate

Selbst die Maschinellen Übersetzungssysteme der Preisklasse unter € 500, wie z.B. *Langenscheidts T1*, *linguatec Personal Translator*, *Systran Professional*, unterstützen die heute am weitesten verbreiteten Dateiformate in der Technischen Dokumentation: DOC bzw. RTF, HTML und reine Textformate (ASCII bzw. ANSI). Die Beschränkung der gängigen Maschinellen Übersetzungssysteme auf die populärsten Dateiformate dürfte für viele Unternehmen deshalb gar nicht negativ ins Gewicht fallen. Unternehmen, die gegenwärtig ihre Dokumentation mit einem Publishing-System wie *Interleaf* oder *PageMaker* erstellen, können Texte nur auf Umwegen maschinell übersetzen und müssen ihre Dokumente je nach Komplexität der Formatierungen aufwändig vor- bzw. nachbearbeiten.

Mit dem Einstieg in die strukturierte und formatneutrale Dokumentationserstellung mit SGML/XML kann diese Problematik umgangen werden. In einer derartigen Redaktionsumgebung kommen in erster Linie für den Einsatz in Unternehmen konzipierte Maschinelle Übersetzungssysteme wie *LOGOS* oder *Systran Enterprise* in Frage, die allerdings auch preislich in einer anderen Dimension als die Konsumentenprodukte angesiedelt sind.

4.3 Sprachkombinationen

Auf globalen Märkten aktive Unternehmen benötigen Übersetzungen ihrer Technischen Dokumentation oftmals in einem Dutzend oder mehr Sprachen. Viele Maschinelle Übersetzungssysteme unterstützen jedoch nur eine oder bestenfalls einige wenige Sprachrichtungen. Erschwerend kommt hinzu, dass für Deutsch als Ausgangssprache im gesamten Markt nur sieben Zielsprachen zur Verfügung stehen: neben Englisch werden Französisch, Italienisch, Ja-

panisch, Russisch, Spanisch und Ukrainisch unterstützt. Und auch diese Zahl täuscht, denn keines der momentan erhältlichen Maschinellen Übersetzungssysteme kann aus dem Deutschen in mehr als drei Fremdsprachen übersetzen (vgl. Hutchins 2000).

Ein möglicher Ausweg aus diesem Dilemma wäre das Erstellen der Dokumentation in Englisch. Für die Ausgangssprache Englisch bietet sich eine ganz andere Marktsituation: Insgesamt werden 29 Sprachen unterstützt, darunter Arabisch, Chinesisch, Indonesisch und Koreanisch, und einzelne Produkte bieten die Übersetzung in sieben Zielsprachen an (vgl. Hutchins 2000). Englisch ist als Ausgangssprache für die Maschinelle Übersetzung auch deshalb besonders geeignet, weil für Englisch im Unterschied zu Deutsch zahlreiche Regelwerke und Werkzeuge für Kontrollierte Sprachen zur Verfügung stehen (vgl. Huijsen 1998), die das Erstellen übersetzungsorientierter Texte erleichtern.

5 Welche Projekte sind für die Maschinelle Übersetzung geeignet?

Im Vorangegangenen habe ich dargestellt, dass mit Maschinellen Übersetzungssystemen im professionellen Übersetzungsbereich wertvolle Zeit-, Kosten- und Qualitätsvorteile erzielt werden können. Dieser potentielle Nutzen ist aber nicht bei allen Übersetzungsprojekten gleich. Und wie bereits angeklungen, gibt es ganz unterschiedliche Einsatzbereiche für Maschinelle Übersetzungssysteme, wie z.B. die Echtzeitübersetzung von E-Mails und die Informationsübersetzung für firmeninterne Zwecke. Ich möchte im Folgenden die Parameter ansprechen, die für Texte gelten, die mit einem Maschinellen Übersetzungssystem vorübersetzt und mit möglichst geringem Aufwand von menschlichen Übersetzenden zu publikationsreifen Dokumenten posteditiert werden sollen.

5.1 Textsorte

Für die Maschinelle Übersetzung sind vor allem Texte geeignet, in denen der kulturspezifische Kontext (Geschichte, Religion, Humor usw.) eine untergeordnete Rolle spielt, die mit eher kurzen und einfach strukturierten Sätzen auskommen und bei denen das Sachgebiet genau definiert ist. Diese Bedingungen werden tendenziell von der gesamten Bandbreite der klassischen Technischen Dokumentation erfüllt, die von der Produktspezifikation über die Handbuchli-

teratur bis zum Schulungsmaterial reicht. Bei Instruktionstexten wie z.B. Installations- und Wartungsanleitungen sowie Hilfesystemen erzielen Maschinelle Übersetzungssysteme besonders überzeugende Ergebnisse.

Weniger für die Maschinelle Übersetzung geeignet sind Textsorten, bei deren Erstellung vor allem journalistische Prinzipien wie z.B. eine variationsreiche und bildhafte Sprache im Vordergrund stehen und bei deren Übersetzung feine Bedeutungsnuancen zu erfassen sind. In diese Kategorie fallen vor allem Marketingtexte wie z.B. Imagebroschüren, Kundenzeitschriften und ganz besonders Texte für die Anzeigenwerbung.

5.2 Konsistenz

Konsistente Ausgangstexte fördern das Qualitätsniveau und die Effizienz jedes Übersetzungsprojekts. Für maschinell zu übersetzende Texte ist die Forderung nach inhaltlicher, terminologischer und stilistischer Einheitlichkeit innerhalb von Dokumenten und Projekten von nicht zu überschätzender Bedeutung:

Inhaltliche Konsistenz

In einem Übersetzungsprojekt sollten möglichst wenige Sachgebiete behandelt werden, weil je Übersetzungsdurchlauf für jedes Wort nur eine Übersetzung verwendet werden kann, in unterschiedlichen Sachgebieten dieselben Wörter jedoch anders übersetzt werden müssen. So bezeichnet das englische Wort *shunting* in einem Finanztext die Arbitrage, in einem Text aus dem Verkehrswesen das Weichenstellen. Die Technische Dokumentation ist wegen der für diese Gattung typischen Beschränkung auf einige wenige Sachgebiete pro Übersetzungsprojekt wesentlich besser für die Maschinelle Übersetzung geeignet als z.B. eine Ausgabe einer Tageszeitung, die ein ungleich breiteres Wissensspektrum abdeckt.

Terminologische Konsistenz

Jedes neue Wort in einem Übersetzungsprojekt erfordert einen Eintrag im Wörterbuch des Maschinellen Übersetzungssystems, andernfalls wird nicht nur das einzelne unbekannte Wort entweder gar nicht oder falsch übersetzt, sondern u. U. der gesamte Satz, in dem das unbekannte Wort vorkommt. Diese Regel gilt auch für Varianten, d. h., dass der Eintrag ‚Motoröl‘ die Variante ‚Motorenöl‘ u. U. nicht abdeckt. Deshalb sind besonders solche Texte für die Maschi-

nelle Übersetzung geeignet, die bereits bei der Erstellung mit einem bestehenden Wörterbuch abgeglichen werden.

Stilistische Konsistenz

Maschinelle Übersetzungssysteme können nicht alle Satzkonstruktionen erfolgreich übersetzen, die nach den Grammatikregeln der jeweiligen Sprache zulässig sind. Kurze, einfache und vollständige Ausgangssätze versprechen das beste Übersetzungsergebnis – nicht nur bei der Maschinellen Übersetzung. Je geringer die Anzahl der unterschiedlichen Satzkonstruktionen in einem Ausgangstext, desto einfacher ist das Postediting, insbesondere in einer Translation-Memory-Umgebung. Aus diesem Grund bietet sich die Maschinelle Übersetzung für Texte an, die nach einem umfassenden Redaktionsleitfaden, besser noch: einer Kontrollierten Sprache, erstellt werden. An dieser Stelle möchte ich darauf hinweisen, dass effektive Kontrollierte Sprachen bei Bedarf bereits mit relativ geringem Aufwand realisiert werden können und deshalb auch für kleine und mittelständische Unternehmen interessant sind (vgl. Mügge 2002b).

5.3 Textvolumen

Da bei der Maschinellen Übersetzung von Technischen Dokumenten i. R. aufwändige Arbeiten in den Bereichen Wörterbuchanpassung und Postediting anfallen, eignet sich diese Technologie eher für das Übersetzen größerer Textmengen. Die Schwelle, ab der sich die Maschinelle Übersetzung unter ökonomischen Vorzeichen lohnt, hängt von der Qualität des Ausgangstextes, den Erwartungen an den Zieltext und dem aus diesen beiden Faktoren resultierenden Aufwand für die Wörterbucharbeit und das Postediting ab. Bei Ausgangstexten, die in einer Kontrollierten Sprache und ohne neue Terminologie erstellt wurden, können sogar einzelne Seiten effizient übersetzt werden; bei komplexeren Ausgangstexten liegt der Schwellenwert nach meinen Erfahrungen bei 50 bis 100 Seiten.

5.4 Zeitdruck

Je schneller ein Projekt übersetzt werden muss, desto interessanter wird die Maschinelle Übersetzung. Das gilt besonders für solche Projekte, bei denen Abstriche bei der stilistischen Qualität des Zieltextes akzeptabel sind. So würde unter hohem Zeitdruck der 2. Satz in der Beispielübersetzung in Bild 5 nur minimal posteditiert, z.B.:

- Rohübersetzung: „Installieren Sie den neuen Schlauch in der umgekehrten Reihenfolge zu entfernen“
- Minimales Postediting: „Installieren Sie den neuen Schlauch in der umgekehrten Reihenfolge *zum Entfernen*“
- Stilistisch korrektes Postediting: „Installieren Sie den neuen Schlauch in der umgekehrten Reihenfolge *des Ausbaus*“.

6 Fazit

Die Maschinelle Übersetzung ist den Kinderschuhen inzwischen entwachsen und hat sich in einzelnen Nischen und vor allem in großen Organisationen einen festen Platz erobert. Angesichts immer kürzerer Produktzyklen und der Forderung nach schneller Markteinführung bei möglichst geringen Kosten für die Dokumentation und die Übersetzung wird die Maschinelle Übersetzung auch für kleine und mittelständische Unternehmen interessant. Der Schlüssel zum erfolgreichen Einsatz dieser viel versprechenden Übertetzungstechnologie liegt in der Technischen Redaktion: Wenn Ausgangstexte

- nach einem umfassenden Redaktionsleitfaden, der im Idealfall auf die Bedürfnisse der Maschinellen Übersetzung ausgerichtet ist,
- unter Berücksichtigung eines unternehmens- oder projektspezifischen Wörterbuchs unter Vermeidung von Homonymen, Synonymen und Varianten

erstellt werden, dann können Maschinelle Übersetzungssysteme innerhalb kurzer Zeit zum effizienten Einsatz gebracht werden. Zum Abschluss möchte ich noch darauf hinweisen, dass das Optimieren von Ausgangstexten für die Maschinelle Übersetzung bedeutet, Texte von allem zu befreien, was mehrdeutig und missverständlich sein kann. Das Resultat sind Dokumente, die nicht nur die Maschine verarbeiten kann, sondern die auch für den menschlichen Leser maximal verständlich und nutzbar sind.

7 Literatur

DIN 2345:1998-04 (1998): Übersetzungsaufträge. Berlin: Beuth.

DIN EN 62079:2001-11 (2001): Erstellen von Anleitungen. Gliederung, Inhalt und Darstellung. Berlin: Beuth.

- Domzalski, Oliver Thomas u. a. (2001): Vor Gebrauch kräftig schütteln – vorn nutzen aus mächtig wackeln. Berühmte Text – vom Computer neu übersetzt. Frankfurt am Main: Eichborn.
- Gabriel, Carl-Heinz u. a. (1995): tekomp-Richtlinie. Technische Dokumentation beurteilen. Stuttgart: Gesellschaft für technische Kommunikation (tekomp).
- Gabriel, Carl-Heinz u. a. (2001): Leitfaden Betriebsanleitungen. Stuttgart: Gesellschaft für technische Kommunikation (tekomp).
- Gereke-Bornemann u. a. (2000): Fallstudie. Projekt MULTIDOC/MULTILINT. <http://www.uni-hildesheim.de/~kger0009/>
- Gruber, Stefanie (2002): German version will be available soon! oder Wie der Translation Server das Leben in einer Firma verändert. <http://www-5.ibm.com/de/software/enews/essay/2002-01-01-ess-1.html>
- Huijzen, Willem-Olaf (1998): Controlled language – an introduction. In: Mitamura, Teruko u. a. (Hrg): Proceedings of the second international workshop on controlled language applications. Pittsburgh, PA: CLAW. S. 1-15.
- Hutchins, John (1999): The development and use of machine translation systems and computer-based translation tools. <http://ourworld.compuserve.com/homepages/WJHutchins/Beijing.htm>
- Hutchins, John (2000): Compendium of translation software. Commercial machine translation systems and computer-aided translation support tools. Carouge: European Association for Machine Translation (EAMT).
- Lehrndorfer, Anne (1996): Kontrollierte Sprache für die Technische Dokumentation. Hintergründe, Fakten und Beispiele. Weinstadt: Adolph.
- Lehrndorfer, Anne u. a. (1997): How to save money in translation cost. In: Technical Communicators' Forum Nr. 2. S. 12-14.
- Lockwood, Rose (2000): Machine translation and controlled authoring at Caterpillar. In: Sprung, Robert C. u. a. (Hrg): Translating into success. Cutting-edge strategies for going multilingual in a global age. Amsterdam: Benjamins. (American Translators Association scholarly monograph series), S. 187-202.
- Mügge, Uwe (2001): The best of two worlds. Integrating machine translation into standard translation memories. A universal approach based on the TMX standard. In: Language International Nr. 6. S. 26-29.
- Mügge, Uwe (2002a): Automatische Terminologieextraktion als Kernelement terminologischer Qualitätssicherung. In: Schmitt, Peter A. (Hrg): Translationsqualität. Ausgewählte Beiträge der Internationalen Fachtagung Translationsqualität in Leipzig 28./29.11.1999. Tübingen: Stauffenburg. (Studien zur Translation), [im Druck].
- Mügge, Uwe (2002b): Kontrolliertes Deutsch – eine Entmystifizierung. In: Lebende Sprachen Nr. 3, S. 110-114.
- Rico, Celia (2001): From novelty to ubiquity: computers and translation at the close of the industrial age. In: Translation Journal Nr. 1. <http://accurapid.com/journal/15mt2.htm>
- Schmitt, Peter A. (2000a): Automatische Übersetzung funktioniert nicht à la Plug and Play. In: c't Nr. 22. S. 231.
- Schmitt, Peter A. (2000b): XLATIO. Übersetzen und Dolmetschen. Automatic Translation. <http://www.uni-leipzig.de/~xlatio/soft-axlo.htm>
- Thurmair, Gregor (2001): Terminologie-Management in TQPro. In: Gräfe, Elisabeth (Hrg): tekomp Jahrestagung 2001 in Wiesbaden. Zusammenfassung der Referate. Stuttgart: Gesellschaft für technische Kommunikation (tekomp). S. 244-247.
- Trujillo, Arturo (1999): Translation engines: techniques for machine translation. London: Springer. (Applied Computing).