



Michael Krenz/Markus Ramlow

Maschinelle Übersetzung und XML im Übersetzungsprozess

Prozesse der Translation und Lokalisierung
im Wandel

Zwei Beiträge, hg. von Uta Seewald-Heeg

F Frank & Timme

Michael Krenz/Markus Ramlow
Maschinelle Übersetzung und XML im Übersetzungsprozess
Zwei Beiträge, hg. von Uta Seewald-Heeg

Hartwig Kalverkämper/Larisa Schippel (Hg.)

TRANSÜD.

Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens

Band 19

Michael Krenz/Markus Ramlow

Maschinelle Übersetzung und XML im Übersetzungsprozess

Prozesse der Translation und Lokalisierung im Wandel

Zwei Beiträge, hg. von Uta Seewald-Heeg

F Frank & Timme

Verlag für wissenschaftliche Literatur

Umschlagabbildung: Aquädukt von Roquefavour © Claus Ferber / PIXELIO

ISBN 978-3-86596-184-6

ISSN 1438-2636

© Frank & Timme GmbH Verlag für wissenschaftliche Literatur
Berlin 2008. Alle Rechte vorbehalten.

Das Werk einschließlich aller Teile ist urheberrechtlich geschützt.
Jede Verwertung außerhalb der engen Grenzen des Urheberrechts-
gesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar.
Das gilt insbesondere für Vervielfältigungen, Übersetzungen,
Mikroverfilmungen und die Einspeicherung und Verarbeitung in
elektronischen Systemen.

Herstellung durch das atelier eilenberger, Leipzig.

Printed in Germany.

Gedruckt auf säurefreiem, alterungsbeständigem Papier.

www.frank-timme.de

Inhaltsverzeichnis

UTA SEEWALD-HEEG	
Tätigkeitsfelder im Wandel. Vorwort zum Band	7
MARKUS RAMLOW	
Maschinelle Übersetzungssysteme im Vergleich	15
MICHAEL KRENZ	
XML im Übersetzungsprozess	151
Index	361

Tätigkeitsfelder im Wandel. Vorwort zum Band

Die internationale Präsenz vieler Unternehmen, die ihre Produkte nahezu zeitgleich auf unterschiedlichen Märkten sprachlich und kulturell angepasst absetzen, hat zu einer umfangreichen Nutzung computergestützter Verfahren bei der Textproduktion und der Übersetzung geführt. Mit dem Einsatz computergestützter Übersetzungstechnologie verfolgen die Unternehmen das Ziel, den Lokalisierungsprozess zu automatisieren, um die anfallenden Übersetzungs- und Lokalisierungsaufwände zeitlich und kostenmäßig zu reduzieren.

Durch den Einzug technologischer Werkzeuge haben sich Arbeitsabläufe, zum Teil auch Tätigkeitsschwerpunkte von Redakteuren und Übersetzern verändert. Die Verkürzung der Zeitspanne von der Entwicklung eines Produktes bis zu seiner Einführung auf dem Markt, verkürzte Produktlebenszyklen sowie die gleichzeitige Auslieferung verschiedener Länder- und Sprachvarianten (das sogenannte *simultaneous shipment* oder kurz *simship*) führte dazu, dass Übersetzungs- und Lokalisierungsdienstleister, und zwar Freiberufler ebenso wie Agenturen, zum Teil unter enormem zeitlichen Druck qualitativ hochwertige Ergebnisse liefern müssen. Ohne den Einsatz von Technologien, die die Arbeit von Redakteuren, Übersetzern und Lokalisierern effizient unterstützen, wären Produktions- und Lokalisierungszyklen mit derart engen zeitlichen Margen und hohem Kostendruck gar nicht möglich.

Technische Dokumentationen von Geräten, die weiterentwickelt werden, müssen langfristig gepflegt werden. Sie verursachen beträchtliche Kosten, insbesondere wenn sie in mehrere Sprachen übersetzt werden. Dies hat dazu geführt, dass man im Bereich der Technischen Dokumentation Standardisierungen anstrebt, um eine Kostensenkung zu erzielen. Die Einführung von XML, Akronym für *Extensible Markup Language*, einer Auszeichnungssprache zur Erstellung und zum Austausch strukturierter Daten, bietet die Grundlage

dazu. Mit XML können Dokumente modular aufgebaut und die einzelnen Module bzw. Textbausteine separat verwaltet und wiederverwendet werden. Metadaten erlauben eine Zuordnung einzelner Textbausteine zu bestimmten Geräte- bzw. Maschinenteilen oder auch zu bestimmten Zielgruppen. Auf diese Weise ermöglicht die Verwendung von XML das sogenannte *Single-Source*- oder Einquellenprinzip: Eine gemeinsame Textquelle wird in unterschiedliche Ausgabeformate gewandelt, anstatt hierzu verschiedene Quelldokumente zu erstellen. Bausteine in Form von XML-Modulen bestehen ausschließlich aus Inhalten wie Text und Bildern. Die konkrete Darstellung oder Formatierung wird mithilfe zusätzlicher Darstellungsdefinitionen festgelegt.

Redakteure erstellen heute vielfach Textbausteine wie z. B. Warnhinweise in sprachlichen Formen, die in Handbüchern unterschiedlicher Produkte verwendet werden können bzw. in unterschiedliche Dokumenttypen, wie die Druckausgabe eines Handbuchs, die Online-Hilfe oder einen Reparaturleitfaden, einfließen können. Das in dieser Form vorliegende Ausgangsmaterial wird schließlich für die Anpassung an verschiedene Märkte von Übersetzern und Lokalisierern bearbeitet. Dabei kann die Übersetzung von Dokumenten, die in einem XML-Format vorliegen, entsprechend der modularen Struktur solcher Dokumente komponentenweise erfolgen.

Da XML-Dokumente unterschiedlich aufgebaut sein können, d. h. der zu übersetzende Inhalt von Anwendung zu Anwendung unterschiedlich gekennzeichnet sein kann, ist die Konfigurierbarkeit von Extraktionswerkzeugen und die korrekte Analyse von XML-Dokumenten eine Anforderung an moderne Übersetzungsumgebungen mit dem Ziel, eine fehlerfreie Extraktion der zu übersetzenden Segmente aus dem Dokument zu gewährleisten.

Durch die Verbreitung von XML werden Übersetzer immer häufiger mit der Verarbeitung von XML-Dokumenten konfrontiert. Aus diesem Grund müssen Übersetzer und Lokalisierer zunehmend auch über Kenntnisse der Eigenschaften von XML und der unterschiedlichen Voraussetzungen zur Verarbeitung von XML-Dokumenten in satzarchivbasierten Übersetzungsumgebungen (Translation-Memory-Systemen) verfügen, um die jeweils im Einsatz befindlichen Systeme optimal für solche Aufgaben konfigurieren zu können.

Damit auf Änderungen von Produkten reagiert werden kann, deren Lokalisierung häufig bereits einsetzt, bevor eine finale Version freigegeben wird, sind computerbasierte Übersetzungsumgebungen wie Translation-Memory-Programme und Lokalisierungswerkzeuge, von Jost Zetzsche auch als *Translation Environment Tools* (TEnT) bezeichnet,¹ unumgänglich. Erst die durch diese Werkzeuggruppe gegebene Wiederverwendung einmal übersetzter Textbausteine ermöglicht eine schnelle Anpassung an Veränderungen, ohne dass bereits getätigte Übersetzungen verworfen werden müssen. Sie sind darüber hinaus zur Sicherstellung terminologischer Konsistenz erforderlich. Diese trägt ebenso wie die Wiederverwendung im Rahmen früherer Versionen übersetzter Texte zur Senkung von Lokalisierungskosten bei.

Um anfallende Übersetzungs- und Lokalisierungsaufwände über die mit diesen Werkzeugen zu erzielenden Einsparungen hinaus weiter zeitlich und kostenmäßig zu reduzieren, wird seit einigen Jahren wieder vermehrt über den Einsatz vollautomatischer maschineller Übersetzung (MÜ) im Lokalisierungsprozess diskutiert. In zahlreichen Unternehmen ist MÜ bereits fester Bestandteil des Übersetzungsprozesses, häufig eng verzahnt mit Translation-Memory-Technologie.²

Die Wissensdatenbank der Firma Microsoft³ ist eines der bekanntesten Beispiele für den Einsatz automatischer Übersetzung. Zahlreiche Beschreibungen von Problemsituationen bei *Microsoft*-Produkten, die der Nutzer auf Anfragen an die Datenbank erhält, wurden nur in englischer Sprache verfasst, können aber in anderen Sprachen ebenfalls abgerufen werden, da sie maschinell übersetzt in der Datenbank hinterlegt wurden. Nutzer der Datenbank werden darauf hingewiesen, dass entsprechend gekennzeichnete Texte maschinell erstellt wurden und unter Umständen qualitative Mängel aufweisen. Der Anbieter nimmt die sprachlichen Mängel in Kauf und macht sich die Tatsache zunutze, dass thematisch sachkundige Nutzer in der Regel in der Lage sind, solche von einem Übersetzungsprogramm erzeugten Rohübersetzungen

.....
1 Siehe die von Jost Zetzsche betriebene Internetpräsenz „translatorstraining.com“, <http://www.translatorstraining.com> [09.05.08]

2 Vgl. z. B. LISA (2004) und Symantec (2007)

3 Microsoft Knowledge Base, <http://support.microsoft.com> [09.05.08]

inhaltlich zu erschließen und als Informativübersetzungen⁴ zur Problemlösung zu nutzen.

Damit aus dem Einsatz von Programmen zur maschinellen Übersetzung ein Nutzen erwächst, müssen zahlreiche Rahmenbedingungen erfüllt sein. Schon die Entwicklung des kanadischen Übersetzungssystems METEO, das seit 1977 Wetterberichte vom Englischen ins Französische übersetzt und seit 1981 von der kanadischen Regierung eingesetzt wird,⁵ hat gezeigt, dass Maschinelle Übersetzung nur in bestimmten Bereichen und unter bestimmten Bedingungen sinnvoll eingesetzt werden kann. Mit METEO war der Nachweis geführt worden, dass Texte eines eingeschränkten Sachgebiets, deren syntaktische Strukturen keine große Varianz aufweisen und deren Wortschatz fest umgrenzt ist, mit gutem Ergebnis maschinell übersetzt werden können.

Beschränkungen, die im Fall von METEO durch die Art der Textsorte Wetterbericht vorgegeben sind, versucht man im Bereich der Technischen Dokumentation heute vielfach durch den Einsatz sogenannter kontrollierter Sprachen zu erzielen. In der Luft- und Raumfahrtindustrie ebenso wie bei IT-Unternehmen und in der Automobilindustrie werden kontrollierte Sprachen des Englischen, Deutschen, Französischen oder Schwedischen für die Redaktion technischer Berichte, Handbücher, Hilfesysteme usw. eingesetzt.⁶ Um sicherzustellen, dass nur die autorisierte Terminologie und nur die in der kontrollierten Sprache festgelegten Satzstrukturen und Stilregeln beim Verfassen verwendet werden, setzen Redakteure hierzu mittlerweile auch spezialisierte Werkzeuge ein. Dies ist zum einen die Werkzeuggruppe des Maschinellen Lektorats, Prüfprogramme, die im Anschluss an die Texterstellung eingesetzt werden, und zum anderen Satzspeicherprogramme, sogenannte *Authoring-Memory-Systeme*, die zum Teil auch während der Texterstellung bereits Vorschläge aus dem Satzarchiv und Rückmeldungen aus der Terminologiekomponente liefern.⁷ Die mittels kontrollierter Sprachen vorgenommenen Standardisierungen von Ausgangstexten ermöglichen eine leichtere Weiterverarbeitung der betreffen-

.....
4 Zimmermann (o.J.)

5 Vgl. Isabelle (1987)

6 Vgl. Muegge (2008a)

7 Vgl. Schmidt (2007) und Berns (2008)

den Texte. Sie bilden die Voraussetzung dafür, dass maschinell übersetzte Texte mit einem Nachbearbeitungsaufwand durch den Humanübersetzer zu einem qualitativ hochwertigen Produkt überarbeitet werden können, der deutlich unter dem zeitlichen Aufwand einer Übersetzung durch einen Humanübersetzer liegt.

Neben Texten, die nach Stilregeln kontrollierter Sprachen erstellt werden, erfüllen einzelne Textsorten, wie Stücklisten oder Katalogeinträge, an sich bereits die Voraussetzungen für einen effizienten Einsatz geeigneter maschineller Übersetzungsverfahren. Dies zeigen Einsatzszenarien in Unternehmen.⁸

Außer Systemen, die auf der Grundlage morphologischer und syntaktischer Regeln übersetzen, sogenannten regelbasierten Systemen, sind für fachlich genau umrissene Einsatzzwecke wie z. B. den Bereich des Maschinenbaus, für den eine Vielzahl von Paralleltexten existieren, inzwischen auch korpusbasierte Systeme im Einsatz, die an zweisprachigen Textkorpora trainiert werden und Übersetzungen statistisch berechnen.⁹ Während dieser Typ von Übersetzungssystemen bislang im wesentlichen im industriellen Umfeld im Einsatz ist, werden regelbasierte Systeme von Herstellern und Verlagen auch als Bürolösungen für Nichtexperten angeboten. Bei der Vermarktung der Produkte wird vielfach vollständig darauf verzichtet, auf das im professionellen Umfeld längst berücksichtigte Faktum aufmerksam zu machen, dass Maschinenübersetzung nur für bestimmte Textsorten in fachlich dafür geeigneten Gebieten sinnvoll einsetzbar ist und darüber hinaus im Vorfeld ein umfangreiches Terminologiemanagement erforderlich macht. Ohne terminologische Vorarbeiten und Eingrenzungen im Hinblick auf ein Fachgebiet und bestimmte Textsorten fallen die Übersetzungsergebnisse häufig syntaktisch mangelhaft und inhaltlich unverständlich aus. Auch wenn die Qualität der Übersetzung bei regelbasierten Texten von Sprachpaar zu Sprachpaar und innerhalb eines Sprachpaars von Sprachrichtung zu Sprachrichtung große Unterschiede aufweisen kann und terminologische Vorarbeiten beim Einsatz dieser Systeme als essentiell betrachtet werden müssen, unterstreichen die Ergebnisse der Evaluierung von

.....
8 Siehe z. B. Rychtycky (2007) oder Muegge (2008b)

9 So beispielsweise das MÜ-Produkt Language Weaver. Vgl. Gerber (2005)

MÜ-Systemen, dass der Einsatz von MÜ nur sinnvoll ist und Einsparungen ermöglicht, wenn bereits beim Verfassen der Ausgangstexte die spätere Verarbeitung durch die Maschine berücksichtigt wird.

Übersetzern fällt beim Einsatz automatischer Übersetzung die Aufgabe der Qualitätssicherung zu. Im Rahmen der Posteditingsphase lekturieren sie die maschinell übersetzten Texte und korrigieren die sprachlichen Mängel des Übersetzungsprogramms. Wo dagegen die Aufgabe des Übersetzens in der Hand des Übersetzers bleibt, der als Hilfsmittel computerbasierte Übersetzungsumgebungen einsetzt, ist ein hohes Maß an technologischer Kenntnis und zum Teil auch informationstechnisches Wissen erforderlich. In beiden Fällen muss sich der Übersetzer technologischen Herausforderungen stellen und sich an veränderte Arbeitsbedingungen anpassen. Dies dokumentieren auch die beiden im vorliegenden Band vereinten Arbeiten von Markus Ramlow und Michael Krenz. Beide rücken Technologien in den Mittelpunkt, mit denen sich Übersetzer und Lokalisierer heute vielfach konfrontiert sehen. Die Arbeiten führen in jeweils unterschiedliche Gebiete ein und vermitteln ein Bild verschiedener Aufgabenstellungen, die im Lokalisierungsprozess anfallen können:

Am Beispiel der Evaluierung von Übersetzungssystemen mit der Sprachrichtung Französisch-Deutsch zeigt Markus Ramlow in seiner im Juli 2006 vorgelegten Arbeit, wie die Übersetzungsqualität regelbasierter automatischer Übersetzungssysteme, auch vor dem Hintergrund der von den Herstellern angegebenen Leistungsfähigkeit, einzuschätzen ist. Der Analyse der Übersetzungsergebnisse geht ein historischer Abriss der Entwicklung von MÜ sowie eine Gegenüberstellung verschiedener in der Literatur dokumentierter Evaluierungsverfahren von MÜ-Systemen voran. Am Beispiel von Texten aus dem Bereich der Wirtschaft und der Jurisprudenz wird dargestellt, welche sprachlichen Erscheinungen den Systemen bei der Verarbeitung Probleme bereiten. Die Problemanalyse macht gleichzeitig deutlich, in welchem Umfang eine Nachbearbeitung durch den Humanübersetzer gegebenenfalls erforderlich ist.

Michael Krenz, dessen Arbeit bereits 2005 entstand, beschreibt den Aufbau von XML-Dokumenten, die aufgrund der Entwicklungen im Bereich der

Technischen Dokumentation eine immer größere Rolle im Lokalisierungsprozess spielen. Anhand zahlreicher Beispiele erhält der Leser einen Eindruck über Methoden zur Bearbeitung von XML-Dateien und mögliche Problemfälle bei deren Verarbeitung mit Translation-Memory-Systemen. Auch die auf XML aufbauenden standardisierten Formate wie TMX (*Translation Memory Exchange*, Austauschformat für Satzarchive, d. h. Translation-Memory-Daten), SRX (*Segmentation Rule Exchange*, Austauschformat zur Übermittlung von Segmentierungsregeln verschiedener Übersetzungssysteme) oder XLIFF (*XML Localization Interchange File Format*, Format zum Austausch zu lokalisierender Daten), die im Arbeitsablauf der Lokalisierung von Bedeutung sind, werden vorgestellt und strukturell beschrieben, bevor deren Unterstützung durch Übersetzungssysteme exemplarisch veranschaulicht wird.

Die beiden Arbeiten von Markus Ramlow und Michael Krenz skizzieren die technologischen Herausforderungen, denen sich die am Übersetzungs- und Lokalisierungsprozess Beteiligten stellen müssen, von moderner Übersetzungstechnologie, wie automatischen Übersetzungssystemen und satzarchivbasierten Werkzeugen, bis hin zu verschiedenen Dateiformaten wie XML und deren Manipulation. Sie geben damit einen Einblick in den Wandel, dem die Tätigkeitsfelder von Übersetzern und Lokalisierern und die an sie gestellten Anforderungen unterworfen sind.

Literaturverzeichnis

- Berns, Kerstin (2008): Wiederverwendung von Textmodulen. Vom Translation Memory zum Authoring Memory. In: Produkt Global 2/2008, 26-28
- Gerber, Laurie (2005): Large Engineering Company Benefits from Language Weaver's Translation System: Case Study. Online im Internet
http://www.imakenews.com/lweaver/e_article000475299.cfm?x=b5TlkgN,0,w
[19.05.08]
- Isabelle, Pierre (1987): Machine translation at the TAUM group. In: King, Margaret (Hrg.): Machine translation today: the state of the art. Edinburgh: Edinburgh University Press, 247-277
- LISA (2004): LISA Best Practice Guide „Implementing Machine Translation“. Online im Internet: <http://www.lisa.org/Business-Decision-Da.512.0.html> [19.05.08]
- Muegge, Uwe (2008a): Machine Translation – Controlled Language – Translation Standards. Online im Internet: www.muegge.cc/controlled-language.htm [19.05.08]
- Muegge, Uwe (2008b): Wie mit proaktivem Terminologiemanagement Übersetzungen auf Knopfdruck möglich werden. In: Mayer, Felix/ Schmitz, Klaus-Dirk (Hrsg.): Terminologie & Fachkommunikation. Akten des DTT-Symposiums, Mannheim 2008. München, Köln: DTT, 115-121
- Rychtycky, Nestor (2007): Intelligent Systems for Manufacturing at Ford Motor Company. In: IEEE Intelligent Systems, Jan./Feb. 2007, 16-19. Online im Internet: http://www.computer.org/portal/cms_docs_intelligent/intelligent/content/promo3.pdf [19.05.08]
- Schmidt, Ulrich (2007): Kontrollierte Sprache. Einsparpotenziale ausschöpfen. In: Produkt Global, 1/2007, 30-33
- Symantec (2007): Systran-Fallstudie. Online im Internet:
<http://www.systran.de/ubersetzung/systran/unternehmensprofil/fallstudien/symantec> [19.05.08]
- Zimmermann, Harald (o.J.): Informationeller Mehrwert durch automatische Übersetzung? Online im Internet: http://scidok.sulb.uni-saarland.de/volltexte/2007/742/pdf/OJ_a.pdf [19.05.08]

Maschinelle Übersetzungssysteme im Vergleich

1	Theoretischer Teil.....	19
1.1	Einleitung.....	20
1.1.1	Der Verstehensprozess bei Mensch und Maschine.....	21
1.1.2	Gründe für den Einsatz maschineller Übersetzungssysteme.....	24
1.1.3	Anwendungsbereiche der Maschinellen Übersetzung	26
1.2	Geschichte der Maschinellen Übersetzung.....	28
1.2.1	Die Anfänge der Maschinellen Übersetzung	28
1.2.2	Der ALPAC-Bericht und seine Folgen.....	31
1.2.3	Die 70er und 80er Jahre.....	33
1.2.4	Neuere Entwicklungen und gegenwärtiger Stand der Forschung.....	35
1.2.5	Stand der wissenschaftlichen Darstellung.....	36
1.3	Klassifizierungskriterien für Übersetzungssysteme	38
1.3.1	Transferstrategie.....	38
1.3.1.1	Regelbasierte MÜ.....	38
1.3.1.2	Neuere Ansätze	42
1.3.2	Anwendungsverfahren	44
1.3.3	Zeitpunkt des menschlichen Eingreifens.....	47
1.3.4	Anteil des menschlichen Eingreifens	50
1.3.5	Weitere Kriterien	53
1.3.6	Methodologische Entscheidungen	54
1.4	Linguistische Probleme im maschinellen Übersetzungs- prozess	56
1.4.1	Morphologische Mehrdeutigkeit	56
1.4.2	Lexikalische Mehrdeutigkeit.....	59
1.4.3	Syntaktische Mehrdeutigkeit und Komplexität.....	64
1.4.4	Lexikalische und strukturelle Inkongruenz.....	66

1.4.5	Mehrdeutigkeit durch Pronominalisierung	68
1.4.6	Zusammenfassung	69
2	Empirischer Teil.....	70
2.1	Einleitung	70
2.1.1	Herstellerangaben zu den Übersetzungssystemen	70
2.1.2	Ansätze zur Evaluierung von Übersetzungssystemen.....	73
2.1.2.1	Der Ansatz von van Slype (1979).....	73
2.1.2.2	Der Ansatz von Lehrberger/Bourbeau (1988)	76
2.1.2.3	Der Ansatz von Arnold (1994).....	77
2.1.2.4	Der Ansatz von Hutchins (1997)	79
2.1.2.5	Kritische Bewertung der Ansätze.....	80
2.1.3	Vorstellung der Methodik der in dieser Arbeit durchgeführten Evaluierung.....	83
2.2	Linguistische Evaluierung.....	86
2.2.1	Ein Beispiel: Qu'est-ce que la compétitivité?	88
2.2.2	Systematische Evaluierung der nicht präeditierten Texte	101
2.2.2.1	Morphologie	101
2.2.2.2	Lexik.....	104
2.2.2.3	Syntax.....	109
2.2.2.4	Pronomenreferenz	115
2.2.3	Systematische Evaluierung der präeditierten Texte	117
2.2.3.1	Lexik.....	118
2.2.3.2	Syntax.....	119
2.2.4	Orthographie und Interpunktion	120
2.3	Schlussfolgerung.....	122
2.3.1	Bewertung der Fehler nach Fehlerklassen	122
2.3.1.1	Lexik.....	122
2.3.1.2	Syntax.....	125
2.3.1.3	Weitere Fehlerklassen.....	126
2.3.2	Optimierung der Übersetzung durch Präedition	129

2.3.3	Bewertung der Qualität der Übersetzungssysteme	130
2.3.3.1	@prompt.....	130
2.3.3.2	Personal Translator und translate	132
2.3.3.3	Power Translator	134
2.3.3.4	Systran	137
2.3.3.5	T1	138
2.3.4	Bewertung des PreisLeistungsverhältnisses.....	140
2.3.5	Ausblick.....	142

Kurzfassung

Elektronische Hilfsmittel wie ein- und mehrsprachige elektronische Wörterbücher oder Terminologiedatenbanken erleichtern zweifelsohne die Arbeit des professionellen Übersetzers dahingehend, dass ihr Einsatz den Zeit- und Arbeitsaufwand beim Übersetzen reduziert. Demzufolge müsste der Einsatz maschineller Übersetzungssysteme eine noch bedeutend umfangreichere Arbeitserleichterung für den professionellen Übersetzer nach sich ziehen. Dies setzt allerdings voraus, dass die Qualität maschinell übersetzter Texte zumindest so hoch ist, dass deren Korrektur durch den Humanübersetzer weniger zeit- und arbeitsaufwändig ist als eine vollständige Humanübersetzung. Angesichts der Komplexität menschlicher Sprache setzt dies wiederum voraus, dass maschinelle Übersetzungssysteme im Übersetzungsprozess auf das Wissen – zumindest sprachliches, möglichst aber auch außersprachliches Wissen – zurückgreifen können, auf das der Humanübersetzer zurückgreift.

Die vorliegende Arbeit wird anhand einer vergleichenden linguistischen Evaluierung von sechs Übersetzungssystemen – @prompt Professional 7.0, Personal Translator 2006 Pro, Power Translator 10 Professional, Systran Professional Premium 5.0, T1 Professional 6.0, translate pro Version 8 – zweierlei aufzeigen: zum einen die Möglichkeiten und Grenzen der maschinellen Simulierbarkeit des Übersetzungsprozesses im Allgemeinen, zum anderen die jeweiligen Stärken und Schwächen der evaluierten Übersetzungssysteme im Besonderen.

Der Zielsetzung der Arbeit entsprechend, die darin besteht zu erörtern, ob und inwieweit der Einsatz maschineller Übersetzungssysteme den Arbeitsaufwand des professionellen Übersetzers reduziert, handelt es sich bei den ausgewählten Versionen der Übersetzungssysteme um diejenigen Versionen, die auf die Bedürfnisse des professionellen Übersetzers abgestimmt sind. Die Evaluierung wird für die Übersetzungsrichtung Französisch-Deutsch durchgeführt.

1 Theoretischer Teil

1.1 Einleitung

Vor nunmehr etwa sechzig Jahren sind die ersten Versuche unternommen worden, das Übersetzen – eine Tätigkeit, die einst dem Menschen vorbehalten war – zu automatisieren, man könnte auch sagen: maschinell zu simulieren.

Die Disziplin, die dieses Ziel verfolgt, ist in den Vereinigten Staaten von Amerika geboren und auf den Namen *Machine Translation* getauft worden. In Analogie dazu ist auch im Deutschen von *Maschinellem Übersetzen* die Rede. Im Französischen hingegen heißt diese Disziplin *traduction automatique*, im Italienischen *traduzione automatica*, im Spanischen *traducción automática*. Im Grunde bezeichnen die jeweiligen Benennungen dasselbe Phänomen, nämlich die „Übersetzung eines Textes aus einer natürlichen Sprache in eine andere natürliche Sprache mittels eines Computers“¹. Einen Unterschied gibt es dennoch: Die Übersetzung wird in den germanischen Sprachen als maschinell, in den romanischen Sprachen aber als automatisch beschrieben. Die Bezeichnung *maschinell* geht auf gr. *mēchanē* ‚Werkzeug‘ zurück². Die Bezeichnung *automatisch* hingegen geht auf gr. *autómatos* ‚aus eigenem Antrieb‘ zurück³. Der Gegensatz, der in dieser diachronen Betrachtung der beiden Bezeichnungen zum Ausdruck kommt, tritt ebenso bei einer synchronen Betrachtung des Bedeutungsgehalts der Bezeichnungen *maschinell* und *automatisch* zu Tage. Der mit *maschinell* bezeichnete Begriff ist produzentenbezogen, er antwortet also auf die Frage, *wer* etwas tut. Der mit *automatisch* bezeichnete Begriff jedoch ist prozessbezogen, er antwortet also auf die Frage, *wie* etwas getan wird. Die Bezeichnungen *maschinelle Übersetzung* und *automatische Übersetzung* repräsentieren also verschiedene Begriffe. Somit besteht ein Kontrast zwischen Germania und Romania. In den germanischen Sprachen wird die maschinelle Übersetzung als eine von einem Übersetzungsproduzenten – der Maschine – angefertigte Übersetzung beschrieben, wohingegen der Begriff in

.....
1 Grein (o.J.): Einleitung
2 Vgl. Kluge (2002:602)
3 Vgl. Kluge (2002:78)

den romanischen Sprachen den Übersetzungsprozess beschreibt und nichts über den Produzenten der Übersetzung aussagt.

In diesem begrifflichen Kontrast zwischen *Germania* und *Romania* kommt ein Phänomen zum Ausdruck, das für Sprache und Sprachen überhaupt charakteristisch ist. Der semantische Gehalt eines Begriffs (und das dadurch bezeichnete Konzept) in einer Sprache und der semantische Gehalt eines Begriffs (und das dadurch bezeichnete Konzept) in einer anderen sind nicht zwangsläufig identisch. Selbst Begriffe, die oberflächlich betrachtet das gleiche Phänomen beschreiben, können bei genauer Betrachtung in ihrer Bedeutung voneinander abweichen, auch wenn es sich wie im vorliegenden Beispiel nur um eine Bedeutungsnuance handelt.

Obwohl die semantische Analyse der scheinbar gleichen Begriffe der *maschinellen Übersetzung* und der *automatischen Übersetzung* gezeigt hat, dass in den beiden Begriffen eine unterschiedliche Konzeptualisierung des durch sie bezeichneten Gegenstands zum Ausdruck kommt, haben sie doch etwas gemeinsam. Sie bilden nämlich insofern eine Begriffsgemeinschaft als beide die Übersetzung als etwas nicht Menschliches beschreiben. Man könnte also sagen, dass sie sich beide auf die extrahumane Übersetzung beziehen und damit gemeinschaftlich den Gegensatz zur Humanübersetzung darstellen.

1.1.1 Der Verstehensprozess bei Mensch und Maschine

*The translation machine (...) is now on our doorstep. In order to set it to work, it remains to complete the exploration of linguistic data by means of comparative lexical and structural analyses...*⁴

Dass diese Einschätzung aus dem Jahre 1960 allzu optimistisch war, liegt aus heutiger Sicht auf der Hand. Inzwischen sind mehr als 45 Jahre vergangen und noch immer existiert kein vollautomatisches Übersetzungssystem, das ohne menschliches Eingreifen qualitativ hochwertige Übersetzungen generiert;

.....
4 Delavenay (1960:114)

noch immer sind die Probleme, die eine vollautomatische maschinelle Übersetzung bereitet, nicht gelöst worden.

Warum ist dies so? Übersetzen ist ein bilingualer und bikultureller Kommunikationsprozess. Gemäß dem Kommunikationsmodell von Kalverkämper ist Kommunikation die Mitteilung eines Textes von einem Sender an einen Empfänger. Voraussetzung für den reibungslosen Ablauf der Kommunikation ist, dass Sender und Empfänger dasselbe Sprachsystem bzw. denselben Code als Grundlage verwenden. Die sich so konstituierende Kommunikationssituation ist eingebettet in eine Kultur.⁵ Überträgt man dieses Modell der monolingualen und -kulturellen Kommunikation auf den Übersetzungsprozess, so muss das Modell in Anlehnung an Kade (1968) folgendermaßen erweitert werden:⁶

Sender → AT → Übersetzer → ZT → Empfänger

Der Übersetzer ist dabei Empfänger des AT (Ausgangstextes) und Sender des ZT (Zieltextes), wobei er den Übergang vom AT zum ZT als Umkodierung vollzieht. Die Umschreibung der Übersetzungskommunikation als Umkodierung wird zwar der Komplexität des translatorischen Handelns des Humanübersetzers nicht gerecht⁷, doch im Hinblick auf den maschinellen Übersetzungsprozess ist sie durchaus zutreffend, da die Maschine lediglich eine Umkodierung von sprachlichem Material vornimmt und gezwungenermaßen weitgehend ohne die Einbeziehung außersprachlicher Informationen arbeitet. Daraus folgt auch, dass im Kontext der Maschinellen Übersetzung (MÜ) die Dimension der interkulturellen Kommunikation im Sinne einer Vermittlung zwischen Kulturen nicht vorhanden ist. Somit ist die MÜ allenfalls als bilingualer, nicht aber als bikultureller Kommunikationsprozess zu betrachten.

Voraussetzung für das Funktionieren von Kommunikation ist, dass die am Kommunikationsprozess beteiligten Partner – Sender und Empfänger – den

.....
5 Vgl. Kalverkämper (2004:28)
6 Vgl. Koller (2001:106)
7 Vgl. Koller (2001:107)

Text verstehen. Dies gilt zwangsläufig nicht nur für die einsprachige, sondern auch für die mehrsprachige Kommunikation: „Wer Texte übersetzen möchte, muss sie zunächst verstehen.“⁸ Aber genau da liegt das Problem:

Das Haupthindernis für effektive MÜ ist, dass der Computer, oder genauer: das Computerprogramm, den zu übersetzenden Text nicht „versteht“, jedenfalls nicht im menschlichen Sinn dieses Wortes.⁹

Dies ist darauf zurückzuführen, dass dem Verstehensvorgang bei Mensch und Maschine unterschiedliche Prinzipien zugrunde liegen. Die Maschine ist nur beschränkt zum Verstehen fähig. Hierfür gibt es mehrere Gründe.¹⁰ Die Maschine

- kann keine vagen oder unspezifischen Aufgaben ausführen. Ein Übersetzungssystem benötigt klare Regeln. Die menschliche Sprache ist zwar wie jedes System durch Regeln gekennzeichnet, doch das der MÜ zugrunde liegende System, eine Programmiersprache, ist ein geschlossenes System, die menschliche Sprache hingegen ist ein offenes und dynamisches System, das auch Regeländerungen oder -abweichungen zulässt. Die Verarbeitung natürlicher Sprache durch maschinelle Systeme ist kompliziert, da für die Gesamtheit der sprachlichen Regeln präzise in Algorithmen darstellbare Regeln gefunden und definiert werden müssen.
- kann nicht lernen. Sie ist auf die Informationen angewiesen, über die sie verfügt. Computer können – sieht man von speziellen Algorithmen des sogenannten „maschinellen Lernens“ ab – in der Regel nicht lernen, weil dies Klassifizierung und Kreativität voraussetzt. Klassifizierung setzt die Erkennung von Ähnlichkeiten voraus und Ähnlichkeit wiederum ist eine vage Beziehung zwischen Sachverhalten. Wie bereits erwähnt, sind Computersysteme aber nicht in der

.....
8 Paepcke (1986:104)

9 Wilss (1994:170)

10 Vgl. Arnold (2003:121-122), Wilss (1988:236-237)

Lage, vage Aufgaben zu bewältigen. Kreativität macht es erforderlich, neue Regeln zu erfinden und sich nicht auf die Einhaltung bestehender Regeln zu beschränken. Es existieren zwar Algorithmen, die das Erlernen gewisser Aufgaben ermöglichen. Algorithmen, die den Erwerb des für die MÜ erforderlichen Wissens ermöglichen, gibt es aber nicht. Daher muss alles, was ein MÜ-System wissen soll, in Form von genauen Regeln programmiert werden.

- kann nicht logisch denken. Logisches Denken erfordert nämlich Weltwissen, über das der Computer nicht verfügt. Das gesamte Weltwissen, das beim Übersetzen einbezogen werden muss, in ein MÜ-System einzuspeisen ist praktisch unmöglich, so dass einem MÜ-System der Rückgriff auf das logische Denken verwehrt bleibt.
- kann keine Probleme bewältigen, bei denen potentiell mehr als eine Lösung möglich ist. Ein Übersetzungssystem kann z. B. keine lexikalischen oder syntaktisch Mehrdeutigkeiten disambiguieren bzw. die Disambiguierung gelingt nur bis zu einem gewissen Grad.

Trotz der bestehenden Mängel kann durch den Einsatz maschineller bzw. maschinengestützter Übersetzungsverfahren eine Effizienzsteigerung beim Übersetzen erzielt werden: „Machine aids for translators currently improve productivity between three and six-fold.“¹¹ In der Annahme, dass Goshawkes Einschätzung aus dem Jahre 1987 zutreffend ist, müsste der Einsatz maschineller bzw. maschinengestützter Übersetzungssysteme mittlerweile wohl eine noch größere Produktivitätssteigerung ermöglichen, da inzwischen fast 20 Jahre vergangen sind und davon auszugehen ist, dass in diesem Zeitraum in der MÜ-Forschung Fortschritte erzielt worden sind.

1.1.2 Gründe für den Einsatz maschineller Übersetzungssysteme

Die MÜ ist aus gesellschaftspolitischer, wirtschaftlicher, wissenschaftlicher und philosophischer Sicht von Bedeutung.

.....
11 Goshawke (1987: 40)

Sie ist gesellschaftspolitisch bedeutend, weil Übersetzungen in allen Gemeinschaften unerlässlich sind, in denen mehr als eine Sprache gesprochen wird. Die Alternative besteht in der Verwendung einer Lingua franca. Dieser Weg erscheint aber wenig attraktiv, weil damit eine Vormachtstellung der betreffenden Sprache und Kultur einhergeht. Übersetzungen sind eine unabdingbare Voraussetzung für die Kommunikation in fast allen Lebensbereichen, wie z. B. Wissenschaft, Wirtschaft, Industrie und Technik. Informationen in der eigenen Muttersprache erhalten und verbreiten zu können ist ein wichtiges Recht. Allerdings ist die Zahl der menschlichen Übersetzer gegenwärtig zu klein, um den ständig steigenden Bedarf an Übersetzungen für eine wachsende Zahl von Sprachenpaaren zu decken, zumal die Produktivität eines menschlichen Übersetzers begrenzt ist. So kann ein professioneller Übersetzer kaum mehr als 1 500 Wörter pro Tag übersetzen. Angesichts der Zeit, die für Terminologierecherche, Überarbeitung u. ä. benötigt wird, dürfte der Durchschnitt eher bei nur etwa halb so vielen Wörtern liegen.¹²

Sie ist wirtschaftlich in zweierlei Hinsicht bedeutend. Zum einen, weil es für den Absatz von Produkten, wie z. B. technischen Geräten, im Ausland unerlässlich ist, dass die entsprechenden Unterlagen (Bedienungsanleitungen u. ä.) in der Muttersprache der potentiellen Käufer abgefasst sind. Ist dies nicht der Fall, wird der Kunde das Produkt nicht bedienen können und auch nicht kaufen. Zum anderen, weil Übersetzungen hohe Kosten verursachen. Neben der Bezahlung des Übersetzers verursachen auch Verzögerungen bei der Fertigstellung von Übersetzungen enorme Kosten. So kann sich die Markteinführung eines Produktes verzögern, wenn die dafür notwendigen Übersetzungen nicht rechtzeitig fertiggestellt werden. Dies hat Gewinneinbußen zur Folge.¹³

Sie ist wissenschaftlich bedeutend, weil die Forschung auf dem Gebiet der MÜ geeignete Anwendungs- und Versuchsfelder für die Forschung in Bereichen wie Informatik, Künstliche Intelligenz, Linguistik und Computerlinguistik bietet. Die MÜ gibt z. B. Aufschluss darüber, inwieweit es möglich ist, spezifi-

.....
12 Vgl. Arnold (1994:4), Goshawke (1987:39-40), Hutchins (1986:15)

13 Vgl. Arnold (1994:4-5)

sche geistige Fähigkeiten des Menschen (Denken, Lernfähigkeit) maschinell nachzuahmen.¹⁴

Sie ist philosophisch bedeutend, weil sie den Versuch unternimmt, einen Prozess nachzuahmen, der die Fähigkeit menschlichen Denkens voraussetzt und nur gelingt, da der Übersetzer sprachliches Wissen, Weltwissen und Situationswissen anwendet. Somit liefert die MÜ Erkenntnisse darüber, ob und inwieweit das Denken automatisiert werden kann.¹⁵

1.1.3 Anwendungsbereiche der Maschinellen Übersetzung

Was die Anwendungsbereiche der MÜ anbelangt, lassen sich verschiedene Ebenen unterscheiden.

- a. Private Anwendung: Die Möglichkeit der Nutzung eines MÜ-Systems durch einen individuellen Nutzer ist aus verschiedenen Gründen erst seit den 90er Jahren möglich. Die Leistungsfähigkeit der PCs hat sich seitdem stark erhöht, praktisch anwendbare MÜ-Systeme setzten noch bis 1990 große und teure Hardware voraus, und die Kosten für die Anschaffung von MÜ-Systemen waren aufgrund der geringen Zahl der Anwender sehr hoch. Bei der privaten Anwendung der MÜ existieren keine allgemeingültigen Qualitätsstandards, die für die Übersetzungen gelten, weil die Anwendungsbereiche sehr heterogen sind. Bei den Nutzern kann es sich z. B. um Personen handeln, die nur die wesentlichen Informationen eines fremdsprachigen Textes benötigen, ein MÜ-System kann aber auch von einem professionellen Übersetzer als Hilfsmittel eingesetzt werden.¹⁶ Ist Letzteres der Fall, sollte die Qualität der Übersetzung möglichst hoch sein, damit der zur Erstellung einer qualitativ hochwertigen Übersetzung erforderliche Posteditationsaufwand möglichst gering ist.

.....
14 Vgl. Arnold (1994:5), Hutchins (1986:15)

15 Vgl. Arnold (1994:5)

16 Vgl. Kunze (o.J.:124)

- b. Institutionalisierte Anwendung: Dies war in der Vergangenheit der einzige Bereich. Er umfasst die Anwendung in Übersetzungsagenturen, Unternehmen und internationalen Institutionen. Diese Einrichtungen haben die Arbeitsabläufe zumindest teilweise auf die MÜ umgestellt bzw. dadurch ergänzt. Der Einsatz von MÜ-Systemen wird allerdings nicht ausschließlich positiv bewertet, aber bis zu einem gewissen Grad können dadurch die für die Übersetzung benötigte Zeit und die entstehenden Kosten reduziert werden.¹⁷
- c. Öffentliche Anwendung: Dieser Bereich betrifft v. a. die Nutzung von mehrsprachigen Informations- und Auskunftssystemen an Kundenterminals oder im Internet (z. B. für Fahrplanauskünfte). Derartige Systeme entwickeln sich erst, aber hier besteht ein großer Bedarf, weil dadurch Sprachbarrieren schnell überwunden werden können. Die Qualität der Übersetzung ist in diesem Bereich anders zu bewerten als in den Bereichen a. und b., da die Korrektur der generierten Übersetzung praktisch ausgeschlossen ist. Somit ist es erforderlich, bei der Systementwicklung einen Kompromiss zwischen der möglichen Qualität und der Akzeptabilität von Übersetzungsfehlern zu finden. Noch zeigt die MÜ-Forschung auf diesem Gebiet keine überzeugenden Lösungen auf, doch ist dieser Anwendungsbereich von großem praktischem Wert. Daher ist zukünftig von verstärkten Bemühungen in Forschung und Entwicklung zur Optimierung derartiger Systeme auszugehen.¹⁸
- d. Online-Übersetzung: Die Bereiche a. bis c. betreffen die Übersetzung von Texten, die bereits schriftlich vorliegen. Die Online-Übersetzung hingegen ist dadurch gekennzeichnet, dass Texte in ein Übersetzungssystem eingespeist und dann umgehend in eine Fremdsprache oder auch simultan in mehrere Sprachen übersetzt werden. Dieses Verfahren kann in Ländern, in denen mehrere Sprachen gesprochen werden, bei der Erstellung internationaler Doku-

.....
17 Vgl. Kunze (o.J.:124-125)

18 Vgl. Kunze (o.J.:125)

mente, bei der Übersetzung von Internetseiten usw. gewinnbringend eingesetzt werden. Wie in anderen Bereichen der MÜ ist auch hier eine Nachbearbeitung der Übersetzungen unerlässlich. Besteht das Ziel der Übersetzung aber darin, lediglich den Inhalt fremdsprachiger Texte verständlich zu machen, erweist sich dieses Verfahren als nützlich.¹⁹

- e. Maschinelles Dolmetschen: Sowohl beim AT als auch beim ZT handelt es sich um einen mündlich dargebotenen Text. Die Übersetzung gesprochener Sprache ist insofern problematisch, als die Spracherkennung Voraussetzung für die Übertragung ist. Im Vergleich zur geschriebenen Sprache ist die Fehlerquote bei der Verwendung gesprochener Sprache deutlich höher. So treten u. a. syntaktisch unvollständige Äußerungen, Anordnungsfehler oder gar gänzlich falsche Formen auf, die in der menschlichen Kommunikation durch den Gesprächspartner automatisch korrigiert werden. Eine Maschine vermag dies jedoch nicht.²⁰

1.2 Geschichte der Maschinellen Übersetzung

1.2.1 Die Anfänge der Maschinellen Übersetzung

Die ersten Versuche, den Übersetzungsprozess zu automatisieren, wurden 1933 unternommen. Der Franzose Georges Artsrouni und der Russe Petr Petrovich Smirnov-Trojanskij entwickelten unabhängig voneinander maschinengestützte Übersetzungssysteme. Beide Systeme beruhten auf dem maschinellen Vergleich von Lexikoneinträgen, wobei keine flektierten Wörter, sondern nur Grundformen berücksichtigt wurden. Artsrounis Übersetzungsmaschine ermöglichte die Übersetzung einzelner Wörter mit Hilfe eines elektrischen Motors und Lexikoneinträgen, die nach dem Lochkartenprinzip gespeichert waren. Bei dem von Smirnov-Trojanskij entwickelten Übersetzungssystem war der Übersetzungsprozess in drei Phasen gegliedert. In der

.....

19 Vgl. Kunze (o.J.:126)

20 Vgl. Kunze (o.J.:126, 142-143)

ersten Phase wurde der zu übersetzende Text durch einen Menschen prädi-
tiert, in der zweiten automatisierten Phase wurde der ausgangssprachliche Text
analysiert und in eine zielsprachliche Repräsentation übertragen, in der dritten
Phase wurde der ZT wiederum von einem Menschen posteditiert. Smirnov-
Trojanskij stellte 1939 eine verbesserte Version seines Systems vor, jedoch
fanden seine Arbeiten in der Sowjetunion kaum Beachtung.²¹

Die Geschichte der MÜ im engeren Sinne beginnt allerdings erst kurz nach
dem Zweiten Weltkrieg. Die ersten elektronischen Rechner waren bereits
während des Zweiten Weltkriegs für militärische Zwecke entwickelt worden,
aber erst 1946 gaben Andrew Donald Booth und Warren Weaver den Anstoß
zur Entwicklung maschineller Übersetzungssysteme. 1942 war der erste
Computer an der Harvard University entwickelt worden. Um eine finanzielle
Förderung für die Entwicklung eines Großrechners für die Universität zu
erhalten, mussten potentielle Investoren durch ein geeignetes Projekt ange-
lockt werden. Daher entwickelten Booth und Weaver die Idee der MÜ, die
zunächst ausschließlich auf einem Wörterbuchvergleich basierte. So wurde
1948 von R. H. Richens ein erstes in Zusammenarbeit mit Booth entwickeltes
Übersetzungssystem im Versuchsstadium vorgestellt. Dieses beschränkte sich
erstmalig nicht nur auf eine Wort-für-Wort-Übersetzung, sondern konnte
durch den Vergleich von Wortstämmen und Flexionsendungen in beschränk-
tem Umfang auch eine syntaktische Analyse vornehmen.²²

Die Entwicklungen auf dem Gebiet der MÜ waren zu jener Zeit außerhalb von
Fachkreisen kaum bekannt. Dies änderte sich 1949 mit der Veröffentlichung
des Weaver-Memorandums. Weaver vertrat darin die optimistische Auffas-
sung, dass es v. a. in technisch-fachsprachlichen Texten möglich sei, die beste-
henden Probleme – insbesondere die Probleme der Mehrdeutigkeit – zu lösen.
Er begründete seine Einschätzung, indem er auf die sprachlichen Universalien
verwies. Weavers Ansicht ist aus heutiger Sicht zu optimistisch, da die sprach-
lichen Universalien noch immer ein Problem der linguistischen Forschung
und damit auch der MÜ sind. Sein Bericht war aber insofern zukunftsweisend,

.....
21 Vgl. Schäfer (2002:19), Schwanke (1991:69)

22 Vgl. Schäfer (2002:20), Schwanke (1999:69-70)