

TRANSLATIONSTECHNOLOGIE – ÜBERBLICK UND AUSSICHT

Peter Sandrini, Innsbruck

Die altgriechische Mythologie erklärt durch die Prometheussage, wie der Mensch zu seinem wichtigsten Werkzeug jener Epoche, dem Feuer gekommen ist. Der Titan Prometheus hat das Feuer aus dem Olymp gestohlen und es den Menschen gebracht; dafür wurde er von Zeus bestraft und an einen Felsen gekettet, wo ihm jeden Tag ein Adler die Leber aus dem Bauch frisst, während sie in der Nacht wieder nachwächst. Prometheus war er ein Menschenfreund, ein Kulturbringender, einer, der die Götter zum Wohle der Menschen überlistete. Prometheus wurde zum Inbegriff des Kampfes um Fortschritt und Überleben.

Die Benutzung von Werkzeugen ist seither zu einem bestimmenden Merkmal des Menschen geworden. Sehr oft wurde gerade diese Eigenschaft als einzige Unterscheidung vom werkzeuglosen Tier herangezogen und der Mensch als ein werkzeugbenutzendes bzw. werkzeugherstellendes Tier, der *homo faber* bezeichnet. Die intellektuelle Leistung des Menschen wurde erst viel später in den Vordergrund gestellt, als der Mensch als *homo sapiens* bezeichnet wurde, als jemand der weiß, der Wissende: Ein Begriff, der sich insbesondere mit den aktuellen Schlagwörtern der Wissensgesellschaft und der Wissensvermittlung deckt. Dass beide Aspekte zur Charakterisierung des Menschen zu kurz greifen, zeigt die neuzeitliche Betonung des *homo oeconomicus*, die den Menschen in seinen Aktivitäten des Handelns und des Wirtschaftens zu erklären versucht. Dass auch dies noch nicht für ein holistisches Bild des Menschen reichen kann, hat der holländische Historiker Johan Huizinga in den 30er Jahren des 20sten Jhds. aufgezeigt; für ihn ist der Mensch vor allem durch seine Spielfähigkeit gekennzeichnet: Der *homo ludens*. Diese These war so überraschend, daß der schockierte Übersetzer des Buches die ausdrückliche Behauptung Huizingas, alle Kultur sei eine Art Spiel, absichtlich in die konventionellere Aussage umwandelte, das Spiel sei ein Element der Kultur.

Homo faber der Mensch als Verfertiger Arbeit	Homo sapiens der einsichtige Mensch Wissen	Homo oeconomicus der wirtschaftende Mensch Geld	Homo ludens der spielende Mensch Spiel
--	--	---	--

Den Übersetzer bringt Chesterman in einem Beitrag zu den 8 Phasen der Entwicklung der Translationswissenschaft (1995) als den *homo transferens* ins Spiel. Es ist der homo sapiens, der mit Texten und (Fach-)Wissen umgeht; in

seiner Tätigkeit als Übersetzer ist er der homo faber, der Texte in anderen Sprachen produziert; und schließlich handelt es sich natürlich auch um einen wirtschaftlich denkenden Menschen, da er in der Regel durch seine Tätigkeit seinen Lebensunterhalt bestreiten muss. Zu seiner eigenen Zufriedenheit und seinem inneren Ausgleich sollte er aber auch eine gewisse Genugtuung in seiner Arbeit finden und das, was heute mit dem englischen Modewort *fun* beschreiben wird: er ist in diesem Sinn zum Teil auch ein homo ludens.

Im Folgenden wenden wir uns dem Translator als homo faber, dem produzierenden und daher unter wirtschaftlichem Druck stehenden Übersetzer zu. Der Leistungsdruck wird gerade in der globalen Gesellschaft, in der Kommunikation über Sprach- und Kulturgrenzen hinweg zur Alltäglichkeit geworden ist, besonders groß. Rationalisierung, Produktivitätssteigerung sind die Schlagworte, die dazu führen, dass Werkzeuge eingesetzt werden müssen, um dem Menschen die Sisyphusarbeit abzunehmen. Dadurch kann wertvolle Zeit gewonnen werden, um darüber nachzudenken, wie die Übersetzung verbessert werden kann bzw. was im Produktionsablauf verbessert werden kann, um letzten Endes den wirtschaftlichen Nutzen zu maximieren.

Werkzeug ist für den Übersetzer nicht mehr das promethische Feuer, sondern heute vor allem die Technik, wobei der Technikbegriff folgendermaßen definiert wird: „Unter Technik sind alle künstlich hervorgebrachten Verfahren und Gebilde zu verstehen, die in soziale Handlungszusammenhänge zur Steigerung ausgewählter Wirkungen eingebaut werden“ (Wörterbuch der Soziologie, Band 3 s.v. Techniksoziologie 1989, 725). Dabei müssen einerseits das Inventar an Instrumenten und Installationen (sachliche Artefakte = Gebilde) bzw. die Sachtechnik und andererseits das Repertoire an Kunstfertigkeit und Kenntnissen (= Verfahren) voneinander unterschieden werden. Unter letzterem werden alle Verfahrensweisen verstanden, die einen Handlungsablauf zu kombinierbaren Operationsschritten umformen, dabei methodischen Operationsregeln folgen und strategisch einen bestimmten Zweck anstreben, z.B. formelhaftes Sprechen und ritualisierte Handlungssequenzen von der Gebetstechnik bis zur Rhetorik, trainierte Bewegungsabläufe der Schwimmtechnik und habitualisierte Verhaltensschemata der schauspielerischen Ausdruckstechnik.

Für das Übersetzen beinhaltet das einerseits die Technik des Übersetzens (methodische Operationsregeln, Handlungssequenzen), die in der Übersetzungswissenschaft der frühen Jahre gerne in den Mittelpunkt gestellt wurde. Ob man das Übersetzen in diesem Sinn ausschließlich als eine Technik sehen kann, bei der der Handlungsablauf in einzelne Operationsschritte mit präzisen Anwendungsregeln zerfällt, ist eine Frage, die von der modernen Übersetzungswissenschaft verneint wird: Zu groß ist die Heterogenität von Texten und Translationssituationen.

Die sogenannte Sachtechnik (Inventar an Instrumenten und Installationen) ist für

das Übersetzen wohl nur auf digitale Werkzeuge, d.h. auf *Software* zu beziehen, sieht man von trivialen Werkzeugen wie Telefon, Kugelschreiber, Papier etc. einmal ab. Dabei stellt sich durchaus die Frage, ob die modernen computer-gestützten Hilfsmittel des Übersetzers nicht durch die Hintertür wieder einen solchen regelhaften operationalen Handlungsablauf einführen. Dies würde aber den Rahmen dieses Beitrages sprengen und müsste Gegenstand weiterführender Überlegungen sein.

Läßt man die Entwicklung automatischer Übersetzungssysteme außer acht, die bereits ab 1945 eingesetzt hat, beginnt die Entwicklung technischer Hilfsmittel zur Unterstützung des Übersetzers mit der Einführung von Computern auf breiter Basis. Analog den Entwicklungsphasen der Translationswissenschaft nach Chesterman (1995) beginnen die computergestützten Werkzeuge auf der Wortebene. Diese Hilfe leisten elektronische Wörterbücher, die das Nachschlagen eines Wortes auf unproblematische Weise im Computerspeicher erlauben. Ab ca 1969 wurden Terminologiedatenbanken wie Eurodicautom entwickelt, die auf großen Mainframe-Computern liefen und über Terminals abfragbar waren. Erst ab Mitte der achtziger Jahre verbreitet sich der Personal-Computer und es entstanden Programme zum Terminologiemanagement, mit deren Hilfe jeder Übersetzer seine eigenen Terminologiebestände verwalten konnte.

In der zweiten Phase, in der Chesterman (1995) die wortwörtliche Übersetzung ansiedelt, werden diese einfachen Wortdatenbanken in Textproduktionsumgebungen integriert und ermöglichen das automatische Suchen und Ersetzen von Wörtern oder Wortfolgen auf Tastendruck. Mit der allgemeinen Verbreitung der PCs entstand auch die Notwendigkeit der Softwarelokalisierung, für die ab ca 1990 eigene Softwarewerkzeuge entwickelt wurden. Das Erkennen und Ersetzen von Wortfolgen aus Datenbanken führte in konsequenter Weiterentwicklung zur Erfindung der Translation-Memory-Systeme ab ca. 1989, kurz bevor das World-Wide-Web seinen Siegeszug antrat. Das WWW wurde bald zur globalen Kommunikationsplattform, das einerseits einen völlig neuen Markt für Übersetzungsdienstleistungen hervorbrachte, andererseits aber auch die technischen Voraussetzungen zur gemeinsamen Nutzung von sprachlichen Ressourcen auf globaler Ebene schuf.

Anforderungen an Translationstechnologie

An moderne Software für Übersetzer, die den Namen Workbench oder Desktop verdient, werden relativ hohe Anforderungen gestellt. Nicht nur soll der eigentliche Übersetzungsprozeß unterstützt werden durch das sprachliche Erkennen von Texten und das Suchen von Termini bzw. Textstellen in Datenbanken oder das automatische Übersetzen, sondern es müssen auch Funktionen integriert sein,

die den Ablauf eines Übersetzungsprojektes unterstützen wie Textstatistiken (Zeilen, Zeichen, Seiten, Wörter), Projektmanagement, Qualitätsprüfroutinen und Content-Management. „Translation tools of the future will need to handle terminology management, translation memory, word processing, desktop publishing, filtering of content in and out of various platforms, natural language processing, project management, team translation functions, database management and global content management all in one software package.“ (Hunt 2003) Eine Vielfalt an Funktionen, die für eine solche Software gefordert wird. Eine logische Gliederung der verschiedenen Funktionen im Rahmen der Translationstechnologie hat Melby (1998) in folgendem Schema vorgelegt:

Infrastruktur		
Translation	Terminusebene	Segmentebene
vor	·Termextraktion ·Terminologierecherche	·Textsegmentierung ·AT- und ZT-Alignment ·Indexierung
während	·Automatische Terminologiesuche	·TranslationMemory-Suche ·Maschinenübersetzung
nach	·Terminologiekonsistenzprüfung ·Prüfen auf nicht erlaubte Terminologie	·Prüfen auf fehlende Segmente ·Format- und Grammtikprüfung

Workflow und Abrechnung

nach Alan Melby (1998): Eight Types of Translation Technology

Die im engeren Sinne translationsspezifischen Funktionen können auf die zwei Bereiche Terminologie und Textabschnitte bzw. Textsegmente reduziert werden, während allgemeine Texteditoren (als Teil der Infrastruktur bei Melby) und Projektmanagement sowie Rechnungslegung zwar für eine rationelle Dienstleistung ebenso wichtig sind, hier aber nur am Rande erörtert werden. Zudem werden die zwei Bereiche auf der zeitlichen Ebene mit dem Translationsprozesses in Verbindung gesetzt, je nachdem, ob sie vor, während oder nach dem Übersetzen durch den Menschen zur Anwendung kommen. Im folgenden soll auf die wichtigsten der in dieser Übersicht genannten Tools kurz eingegangen werden.

Auf Terminusebene gibt es bereits seit längerem entsprechende computergestützte Werkzeuge zur Terminologieverwaltung. Interaktion mit Texteditoren, automatische Suche und Übernahme von Termini sind dabei zum Standard

geworden. Neuere Entwicklungen stellen Tools zur automatischen Identifikation und Extraktion von Terminologie aus zweisprachigen Textkorpora sowie die Kooperation bei der Bearbeitung und Nutzung von Terminologiebeständen über das WWW dar.

Auf die Ebene einzelner Textabschnitte bzw. Textsegmente haben sich Translation-Memory-Programme spezialisiert. Der Name bezieht sich auf einen Übersetzungsspeicher, in dem das Ausgangstextsegment mit seiner Übersetzung in einer Datenbank abgespeichert wird. Bei jeder neuen Übersetzung wird das Ausgangstextsegment mit dem Speicher verglichen, und eine im Speicher gefundene totale oder partielle Übereinstimmung dem Übersetzer angeboten. Auf diese Weise wird keine Textstelle zweimal übersetzt. Die meisten Softwarepakete mit dieser Funktionalität bieten darüber hinaus auch noch andere Leistungen, wie z.B. das Alignment, das die Integration bestehender Übersetzungen in einen Übersetzungsspeicher erlaubt: Ausgangstext und Zieldtext werden in einzelne Segmente zerlegt und die korrespondierenden Segmentpaare in einem Translation-Memory (TM) gespeichert, das dann für neue Übersetzungen zur Verfügung steht.

TM-Systeme rationalisieren und beschleunigen den Übersetzungsprozess und sind daher für professionelle Übersetzer bei Texttypen mit wiederkehrenden Textstellen sehr interessant. Die Verwendung solcher Systeme wurde auf der Grundlage einer Studie (Höcker 2003) im Rahmen eines EU-Projektes (eColore: <http://ecolore.leeds.ac.uk>) erhoben. Dazu wurde eine Umfrage unter Übersetzern durchgeführt (BDÜ und ITI Mitglieder), die folgende Ergebnisse ergab: 29 % der Befragten benutzen TM-Systeme täglich, 15 % wöchentlich, 8 % monatlich und 12 % weniger oft, 36 % überhaupt nicht. Keine Unterschiede bestehen bezüglich der verschiedenen Altersgruppen, sehr wohl aber bezüglich der Texttypen, die mit Hilfe von TM-Systemen übersetzt werden: Bei den 29 %, die TM täglich verwenden, zählen 76 % der übersetzten Texte zur technischen Dokumentation, 58 % zu Software-User-Interfaces, 43 % zu Hilfedateien und Webseiten. Bei der Gruppe der Non-User ergibt sich bei den übersetzten Texttypen ein diametral entgegengesetztes Bild: 63 % Rechtstexte, 62 % Wirtschaftskorrespondenz, 61 % Wirtschaftstexte, 21 % literarische Texte, 53 % technische Dokumentation und 16 % Software UI.

Ein weiterer Teil der Umfrage beschäftigte sich mit den Gründen, die zum Einsatz dieser Programme geführt haben. Klar voran liegen hier Produktivitätssteigerung, Kundenwünsche Konsistenz von Text und Terminologie, Möglichkeiten der Qualitätskontrolle und Terminologiemanagement. Ebenso erhoben wurden die Gründe, die einen Einsatz von TM-Systemen verhindern bzw. erschweren: Anggeführt werden hier vor allem die hohen Anschaffungskosten, mangelnder Return-on-Investment (ROI), komplexere Abrechnungsmodalitäten, Lernschwierigkeiten und Schulungsbedarf. Während Open-Source-Software die

Anschaffungskosten wenigstens zum Teil relativiert, besteht doch ein großer Bedarf an entsprechenden Ausbildungs- und Einschulungsangeboten.

Während der Einsatz von TM-Systemen auf breiter Basis zunimmt, gibt es in der Entwicklung solcher Programme unterschiedliche Tendenzen. Der technologische Knackpunkt bei TM-Systemen ist ohne Zweifel die Segmentierung, d.h. das Aufspalten des Textes in kleinere Abschnitte. Die Segmentierung entscheidet über die Effizienz des Einsatzes von TM (Produktivität und Konsistenz), daher soll auf dieses Problem näher eingegangen werden. Es lassen sich 3 Arten von Segmentierungsstrategien unterscheiden, und zwar auf der Basis von

- reinen Zeichenfolgen (z.B. in der Software MultiTrans),
- ganzen Sätzen (der traditionelle Ansatz z.B. bei Trados), oder
- Absätzen (z.B. bei OmegaT).

Je kleiner die Segmentierungseinheit ist, desto eher und desto mehr Ergebnisse werden bei gleichen Voraussetzungen gefunden. Bei kleineren Segmentierungseinheiten unterhalb der Satzebene verschwimmt zusehends die Grenze zu terminologischen Einheiten. Auswahl und Einfügen der angezeigten Treffer aus einer langen Liste erfordern vom Übersetzer jedoch wesentlich mehr Aufwand. Je größer die Segmentierungseinheit, desto mehr Zeitersparnis für den Übersetzer. Eine Studie zur Evaluierung von TM-Systemen (Gow 2003) führt die Effizienz der unterschiedlichen Segmentierungsstrategien auf unterschiedliche Texttypen zurück: Tritt eine hohe Wiederholung auf Absatzebene auf, z.B. bei standardisierten Texten (Bedienungsanleitungen, Handbücher, etc.) haben entsprechende Segmentierungsstrategien einen Vorteil und bringen hohen Nutzen. Bei Wiederholungen unterhalb der Satzebene, d.h. bei kaum standardisierten Texten, bringen Systeme wie MultiTrans einen höheren Nutzen. Erfahrene Übersetzer, die bereits eine hohe terminologische Vorbereitung im Fachgebiet besitzen, erwarten sich von einem TM-Tool eine Geschwindigkeitssteigerung bzw. Zeitersparnis, die sich am besten mit größeren Segmenten erreichen läßt. Übersetzer am Beginn ihrer Karriere werden eine hohe Trefferquote bevorzugen, um möglichst viel Information aus der Datenbank gewinnen zu können, wobei der Zeitfaktor eher im Hintergrund steht.

Die unterschiedlichen Segmentierungsarten führen auch beim Austausch von TM zu Problemen: Das TM eines Systems, das auf Absatzebene segmentiert, wird beim Einsatz in einem System, das auf Satz- oder Wortebene segmentiert, kaum zu Übereinstimmungen führen. Daher ist der Austausch zwischen Systemen mit unterschiedlichen Segmentierungsverfahren nicht sinnvoll, außer es läßt sich das Segmentierungsverfahren anpassen. In diesem Fall sollte für die Wiederverwendung genau das Verfahren gewählt werden, das auch für den Aufbau des TM zur Anwendung gekommen ist. Auf die Frage der Austauschformate TMX und SRX werden wir weiter unten eingehen.

Ein weiterer wichtiger Punkt bei der Anwendung solcher Systeme ist die Such-

und Vergleichsstrategie. Hier hat sich das sogenannte Fuzzy-Matching bewährt, das eine ungenaue Suche nach ähnlichen Segmenten erlaubt und gegenüber einer Suche auf der Basis eines reinen Zeichenvergleichs Vorteile bringt. Besteht die Möglichkeit, den Prozentsatz der Übereinstimmung einzustellen, können auch wenig übereinstimmende Segmente gefunden werden, oder am anderen Ende der Skala auch nur Segmente, die beinahe ganz übereinsimmen.

Zur Anwendung kommen in der Praxis eine Reihe von Softwarepaketen. Die neuesten Daten (Lommel 2004) geben Trados als Marktführer mit 71 % an, gefolgt von SDLX mit 28 %, Déjà-Vu mit 24 % und Alchemy Catalyst mit 20 % Marktanteil. Größere Unternehmen benutzen durchaus auch mehrere unterschiedliche Tools. Die Vielfalt an TM-Systemen ist mittlerweile beeindruckend: Neben den genannten bekannteren Programmen, die alle in die erste Gruppe fallen, sind folgende Arten an TM-Systemen zu unterscheiden:

- Eigenständige Applikationen: Star Transit, Cypresoft Trans Suite, MultiCorpora Multitrans, Xerox Multilingual Memory Manager, Heartsome Translation Editor, Termseek Translator's Intuition, Bridgeterm Promemoria;
- Softwarepakete, die auf MS-Word-Makroprogrammen beruhen: Wordfast, Wordfisher, Meta-Texis, Sprint, Memcat;
- kleinere Anwendungen wie King Memo, Ecco, Frankenstein, Araya Suite;
- Open-Source-Programme wie Foreigndesk, OmegaT, Kbabel, Mimer's Brunn.

Zusätzlich gibt es eine Reihe von Programmen, die der Verwaltung von TM dienen, wie z.B. der Bearbeitung und dem Zusammenfügen bzw. Aufteilen von TM. Hier sind zu nennen: Heartsome TMX Editor, Waikoloa.

Im oben dargestellten Diagramm (Melby 1998) werden auch die nicht translationsspezifischen Funktionen mit einbezogen, die meistens durch allgemeine Softwarepakete bereits mehr oder weniger gut abgedeckt werden. Der wichtigste Bereich ist dabei die Textverarbeitung mit allen damit verbundenen Funktionen, wie z.B. Textkonvertierung, Rechtschreibprüfung, Wörterbücher, DTP, Textsuche, Textstatistiken, etc.

Translationsspezifisch sind hingegen alle Funktionen, die der Verwaltung von Übersetzungsprojekten im weitesten Sinn dienen: Planen des Workflows, Projektmanagement, Teamübersetzen, Erstellen von Kostenvoranschlägen und Abrechnungen, etc. Auch zu diesem Zweck wurden selbständige Programme entwickelt, die auf mehr oder weniger Akzeptanz bei den professionellen Übersetzern gestoßen sind: z.B. PractiCount, LTC-Organiser, Translation Office 3000.

Vielfach ist diese Funktionalität, zumindest teilweise, bereits in umfassenderen TM-Programmen integriert: Bei Trados beispielweise übernimmt der Workspace Projektmanagementfunktionen und die Analyse-Funktion kann für Kostenvoranschläge herangezogen werden.

Standards und Normen

Die Arbeit mit unterschiedlichen Softwareprogrammen und unterschiedlichen Dateiformaten kann zu großen Problemen führen, wenn sprachliche Ressourcen wie Terminologien oder TM-Dateien ausgetauscht bzw. von unterschiedlichen Systemen eingesetzt werden müssen. Um die allgemeine Verwendbarkeit sprachlicher Ressourcen unabhängig von einzelnen Softwareproduzenten und Betriebssystemen zu unterstützen, wurden entsprechende Normen geschaffen: Das Translation Memory Exchange Format TMX mit dem Segmentation Rules Exchange Standard SRX und das Termbase Exchange Format TBX, die alle als Anwendung der allgemeinen Auszeichnungssprache Extended Markup Language XML entwickelt wurden. Bevor die einzelnen Austauschformate beschrieben werden können, soll eine kurze Einführung in XML die Grundlagen erklären.

XML wurde als ein plattformunabhängiges, software- und hardwareneutrales Instrument entworfen, um Informationen bzw. Daten zu strukturieren, zu speichern und zu übermitteln.. XML ist eine Textauszeichnungssprache, die sich aus dem alten Standard SGML und dem durch das WWW weit verbreiteten HTML entwickelt hat. Ähnlich wie HTML beschreibt XML Daten anhand einzelner Auszeichnungseinheiten oder *Tags*, die anders als bei HTML aber nicht vordefiniert sind, sondern für jede Anwendung neu definiert werden können. So kann der Autor einer XML-Datei Tags für seinen Zweck „erfinden“ und in eine geeignete hierarchische Struktur bringen. Der jeweils zulässige anwendungsspezifische Satz an Tags kann in einer Document Type Definition DTD oder einem XML-Schema spezifiziert werden. Daraus resultiert die breite Anwendungsmöglichkeit von XML.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<vortrag>
<titel>Translationstechnologie</titel>
<datum>07.01.2004</datum>
<Inhalt>
<intro>Einführung</intro>
<überschrift1>historischer Überblick</überschrift1>
<überschrift2>Anforderungen</überschrift2>
<überschrift3>Tools</überschrift3>
<überschrift4>...</überschrift4>
</Inhalt>
</vortrag>
```

Dieses XML-Beispiel beginnt mit einer Deklaration, dass es sich hier um ein XML-Dokument der Version 1 handelt und dass darin der Zeichensatz ISO Latin 1 verwendet wird. Das Dokument beginnt mit einem sogenannten Root-Tag, durch den das Dokument als ein Vortrag gekennzeichnet wird, danach kommen

Titel, Datum und Inhalt mit Unterelementen. XML ist immer streng hierarchisch geordnet, die einzelnen Elemente beginnen und schliessen mit einem Tag. XML enthält keine Programmieranweisungen, es gliedert lediglich die vorhandenen Informationen und macht sie durch entsprechende Auszeichnungen explizit. Die Darstellung und Formatierung überlässt XML den einzelnen Softwareprogrammen. So wird in unserem Beispiel lediglich ausgesagt, dass es sich bei der Zeichenfolge *Translationstechnologie* um einen Titel handelt, es wird aber nicht festgelegt, ob der Titel nun fett, zentriert, etc. dargestellt werden soll. Die Tags Vortrag, Titel usw. sind nicht in irgendeinem Standard festgelegt, sondern ad hoc für dieses Beispiel erfunden worden. Auf diese Weise ist XML ungebunden und kann für jeden erdenklichen Zweck oder Anwendung eingesetzt werden.

Eine dieser Anwendungen von XML ist TMX, das *Translation Memory Exchange Format*, das den Zweck hat, den Import und den Export von TM-Datenbanken zwischen den verschiedenen Softwareprodukten zu erlauben. Dadurch können die Benutzer unabhängig von dem eingesetzten Tool ihre Daten weiterverwenden, wenn sie auf ein anderes Tool umsteigen, oder auch mit anderen Benutzern Daten austauschen.

```
<?xml version="1.0" ?>
<tmx version="1.4">
  <header creationtool="XYZTool" creationtoolversion="1.01-023"
    datatype="PlainText" segtype="sentence"
    adminlang="en-us" srclang="EN"
    o-tmf="ABCTransMem">
  </header>
  <body>
    <tuv>
      <tuv xml:lang="EN">
        <seg>Text in <bpt i="1"&lt;B></bpt>bold<ept i="1"&lt;/B></ept>.</seg>
      </tuv>
      <tuv xml:lang="FR">
        <seg>Texte en <bpt i="1"&lt;B></bpt>gras<ept i="1"&lt;/B></ept>.</seg>
      </tuv>
    </tuv>
  </body>
</tmx>
```

Die zentralen in TMX verwendeten XML-Tags sind die *Translation Units* <TU>, die jeweils ein Segment kennzeichnen; sie enthalten die einzelnen Sprachversionen oder *Translation Unit Variants* <TUV>, die durch das Attribut *lang* gekennzeichnet werden und das Segment <SEG> bzw. den Textabschnitt in dieser Sprache enthalten. Theoretisch könnten durchaus mehrere Sprachen bzw. mehrere TUVs in TMX abgespeichert werden, zur Zeit werden von dem meisten TM-Systemen aber lediglich zweisprachige TM unterstützt.

Die Version TMX 1.1 speichert reinen Text ab, ohne Berücksichtigung von Formatierungen oder anderen Textauszeichnungen. Ab Version 1.4 werden Forma-

tierungen und Inline-Tags in TMX mitgespeichert. Ein Abspeichern von Textsegmenten als Translation Units im reinen Textformat bringt bei gleichen Textsegmenten, die jedoch anders formatiert sind bzw. durch Formatierungstags unterbrochen sind, keine vollkommenen Übereinstimmungen mehr, obwohl der Text genau derselbe ist.

Auf das Problem der unterschiedlichen Segmentierungsverfahren wurde bereits oben verwiesen. Um den Austausch von TMX-Dateien dennoch möglichst transparent zu machen und eine Anpassung der Segmentierungsverfahren in den Softwaresystemen zu erlauben, wurde ein Verfahren entwickelt, das die Segmentierungsregeln in die TMX-Datei integriert, das *Segmentation Rules Exchange Format SRX*. Dabei werden in die TMX-Datei die genauen Segmentierungsregeln, die beim Erstellen dieses TM zur Anwendung gekommen sind, integriert. Bei der Verwendung dieses TM in einem anderen System, kann dieses die ursprünglichen Segmentierungsregeln auslesen und das System auf diese Regeln umstellen, um eine möglichst hohe Trefferquote zu erreichen.

Für den Austausch von terminologischen Daten wurde ebenfalls ein auf XML basierendes Format entwickelt, das TBX.

```
<?xml version='1.0'?>
<!DOCTYPE martif PUBLIC "ISO 12200:1999A//DTD MARTIF core (MSCcdV04)//EN">
<martif type='DXLT' lang='en' >
<martifHeader>
  <fileDesc><sourceDesc><p>from an Oracle termBase</p></sourceDesc></fileDesc>
  <encodingDesc><p type='DCSName'>MSCdmV04</p></encodingDesc>
</martifHeader>
<text> <body>
  <termEntry id='ID67'>
    <descrip type='subjectField'>manufacturing</descrip>
    <descrip type='definition'>A value between 0 and 1 used in </descrip>
    <langSet lang='en'>
      <tig>
        <term>alpha smoothing factor</term>
        <termNote type='termType' >fullForm</termNote>
      </tig>
    </langSet>
    <langSet lang='hu'>
      <tig><term>Alfa simitási tényező </term></tig>
    </langSet>
  </termEntry>
</body> </text>
</martif>
```

TBX beruht auf dem älteren Terminologieaustauschformat Martif (Machine-Readable Terminology Interchange Format) und besitzt als zentrales Element den terminologischen Eintrag `<termEntry>`, das in verschiedene Spracheinheiten `<langSet>` zerfällt, die wiederum verschiedene *Term Information Groups* `<tig>` enthalten. Jeder Terminus wird durch die im TIG enthaltenen Informationen aus-

fürhlich beschrieben.

Für die Durchführung von Lokalisierungsprojekten wurde ein einheitliches Format geschaffen, das die Vielfalt an lokalisierbaren Inhalten auf ein einheitliches Datenformat reduziert und damit den Austausch von Dateien zum Zwecke der Lokalisierung erleichtert. Das *Localisation Interchange File Format* XLIFF beruht ebenfalls auf XML und unterstützt den verlustfreien Austausch lokalisierbarer Daten und der gesamten dazu gehörigen Information.

```
<xliff version='1.1'  
  xmlns='urn:oasis:names:tc:xliff:document:1.1'  
  <file original='hello.txt' source-language='en' target-language='fr'  
    datatype='plaintext'  
    <body>  
      <trans-unit id='hi'>  
        <source>Hello world</source>  
        <target>Bonjour le monde</target>  
      </trans-unit>  
    </body>  
  </file>  
</xliff>
```

XLIFF wird u.a. unterstützt von folgenden Tools: der XLIFF-Konverter Rainbow von Enlaso (HTML, XML und RTF nach XLIFF), XLIFF Translation Editor von Heartsome, ein TM-System, dessen Dateiformat auf XLIFF beruht.

Die Lokalisierung ist ein typisches Beispiel für einen Bereich, in dem sich die Anwendung von computergestützten Übersetzungstools konsequent durchgesetzt hat. Unter Lokalisierung versteht man, einfach ausgedrückt, die Anpassung eines Produktes an lokale Anforderungen. Dies beinhaltet einerseits die sprachliche Übersetzung, andererseits die kulturelle Adaptation für spezifische lokale Märkte. Lokalisierung umfasst zwei große Bereiche: Die Lokalisierung von Software und die Lokalisierung von Webinhalten. Softwarelokalisierung ist ohne entsprechende Softwarewerkzeuge nicht durchführbar, da der Programmcode von den nötigen Anweisungen zur Interaktion mit dem Anwender (Menüführung, Dialogfelder, Programm- und Fehlermeldungen) getrennt werden muss. Auch in der Lokalisierung von Webinhalten sind entsprechende Softwarewerkzeuge nicht mehr wegzudenken, weil auch hier der HTML-Code vom sprachlichen Inhalt getrennt werden muss; vor allem aber müssen aufgrund des hohen Aktualisierungsgrades Änderungen am Text automatisch erkannt werden können.

Komponenten solcher Lokalisierungstools sind ein Editor, der den Programmcode vom reinen Text trennt, ein Terminologiemodul sowie ein TM-Modul. Bei der Weblokalisierung werden für größere und komplexere Webauftritte Content-Management-Systeme verwendet, große Datenbanken, in denen Textsegmente und Grafiken zentral verwaltet werden. Neuere Systeme dieser Art, die so ge-

nannten Global-Content-Management-Systeme GMS vereinen viele translations-spezifische Funktionen unter einer Oberfläche: TM-Komponente, Terminologie-management, Textstatistiken, Kostenvoranschläge und Dienstleistermanagement, Projektmanagement, Workflow, Abrechnung.

An spezifischen Lokalisierungstools sind derzeit u.a auf dem Markt:

- zur Softwarelokalisierung: Alchemy Catalyst, Passolo, Multilizer, AIT Visual Localize, Software Builders Applocalize;
- zur Weblokalisierung: Webbudget XT, Cat's Cradle, WholeTree SiteTrak, Logoport;
- Globalization Management Systems (GMS): Globalsight, SDL, Uniscape, Idiom.

Maschinenübersetzung

Für die Maschinenübersetzung ist es bisher trotz hoher Förderungen nicht zu einem Durchbruch in der Ausgabequalität gekommen. Sehr wohl aber hat das WWW zu einer weiten Verbreitung automatischer Übersetzungssysteme geführt, wobei auf Knopfdruck eine Webseitenübersetzung durchgeführt werden kann. Eine Änderung im Anspruch der Benutzer hat dieser Art von Dienstleistung zum Erfolg geführt: Niemand erwartet sich eine perfekte Übersetzung; es genügt in vielen Fällen eine reine Inhaltsangabe über den Ausgangstext. Die Maschinenübersetzung liefert suboptimale Texte, die zur reinen Informationswiedergabe durchaus genügen können, denkbar z.B. im Rahmen von unternehmens- bzw. organisationsinterner Kommunikation (EU) oder bei fachinterner Informationswiedergabe. Jedenfalls ist für den professionellen Übersetzer durch diese Anwendungen keine Konkurrenz zu befürchten. Auch können diese Systeme nicht zur Unterstützung des Übersetzungsprozesses herangezogen werden, da die Korrektur bzw. Nachbearbeitung der Resultate häufig mehr Zeit in Anspruch nimmt als eine Humanübersetzung.

Interessant wird die Maschinenübersetzung für Unternehmen oder Organisationen aber dann, wenn solche Systeme in eine optimierte Anwendungsumgebung integriert werden. Ein wichtiger Baustein dazu ist die Standardisierung der Ausgangstexte durch die Einführung sogenannter kontrollierter Sprachen, die eine Vereinfachung und Disambiguierung der Syntax sowie eine absolute Terminologiekonsistenz durchsetzen. Der Satz *clean surfaces* kann mehrfach interpretiert werden, je nachdem ob *clean* als Substantiv oder Verb verstanden wird; im *controlled English* wird dieser Satz entweder zu *Clean the surfaces* oder zu *the clean surfaces*. Der mißverständliche *noun cluster* „main gear inboard door retraction winch handle“ wird im *Simplified English*, eine weitere Variante einer kontrollierten Sprache, zu „the retraction winch handle to the

inboard door of the main gear“. Solche kontrollierten Sprachen sind von großen internationalen Unternehmen bereits mit Erfolg eingesetzt worden (z.B. Boeing, Caterpillar, Rank Xerox, IBM). Sie eignen sich gemeinsam mit der automatischen Übersetzung besonders für Texte wie Handbücher, Anleitungen, Product-support, u.ä., weniger aber für kreative Sprache, wie sie im Bereich von Werbe- und Marketingtexten zum Einsatz kommt.

Daneben versuchen neuere Entwicklungen an die erfolgreiche Anwendung des Translation-Memory anzuknüpfen und den alten regelbasierten Ansatz durch einen statistischen, auf bereits durchgeführten Übersetzungen sowie vorliegenden zweisprachigen Paralleltextrn beruhenden Ansatz zu ersetzen. Lexeme, Phrasen und syntaktische Strukturen werden aufgrund von Wahrscheinlichkeitsanalysen übersetzt: Je größer der zur Verfügung stehende Korpus, desto besser wird auch die automatische Übersetzung ausfallen (siehe *Language Weaver*).

Open Source Software

Softwarewerkzeuge sind für den professionellen Übersetzer also kaum mehr wegzudenken. Der Einsatz dieser Programme trifft aber häufig immer noch auf Widerstand bei den betroffenen Personen, sei es aus Mangel an Schulung oder aufgrund der teilweise sehr hohen Kosten. Zumindest die finanziellen Hürden für den Einstieg in das computergestützte Übersetzen können heute durch alternative Programmentwicklung überwunden werden.

Im Bereich der Softwareproduktion ist im Rahmen der Globalisierung ein neuer Trend zu beobachten. IT-Unternehmen verstehen sich nicht mehr unbedingt als Produktionsbetriebe, sondern immer mehr als Dienstleistungsunternehmen. Die Produktion von Software muss daher nicht unbedingt von angestellten Programmierern als Auftragsarbeit durchgeführt werden, sondern z.B. auch durch kooperative Anstrengungen in einem Netzwerk verschiedener Nutzer-Programmierer, wie das heute von der aufstrebenden Open-Source-Community vorgemacht wird. Hierbei vernetzen sich teils IT-versierte Anwender teils Programmierer, um gemeinsam eine neue Software zu erstellen bzw. ein bestehendes Softwarepaket zu verbessern. Durch die internationale Vernetzung und die andauernde Arbeit am Programm kann dieses immer weiter verbessert und angepasst werden. Die Open-Source-Bewegung kann proportional viel mehr Energie und Aufwand an Personal bzw. an Fachwissen für ein einziges Programmierproblem aufwenden, als es jemals Softwarebetriebe tun könnten, die stets unter wirtschaftlichem Druck stehen. Unter der freien Lizenz *General Public License GNU* werden solche Programme über das WWW kostenlos zur Verfügung gestellt. Voraussetzung ist lediglich, dass sie absolut kostenlos sein müssen und dass auch der Quellcode veröffentlicht wird. Die entscheidenden Vorteile dabei sind folgende:

- Verwendung nach freiem Belieben und zu jedem Zweck, auf beliebigen PCs und auf jeder Betriebssystemplattform;
- Absolute Verfügungsmacht über die Software für das Anpassen an eigene Anforderungen: Verbessern und Erhöhen der Funktionalität, Fehlerkorrektur, Verstehen der internen Vorgänge;
- Weiterverteilen an andere Benutzer.

Da der Programmiercode offen ist, steigt auch die Sicherheit der Anwendung, weil jeder Einblick in die internen Vorgänge und Abläufe der Software hat. Das beste Beispiel für eine solche Software ist Open Office, eine Büroanwendung, die als Open-Source zur Verfügung steht und durchaus mit dem Marktführer vergleichbar ist, und heute bereits von sehr vielen Übersetzern eingesetzt wird. Zusätzlich wird das offene Dateiformat von Open-Office zu einer breiten Unterstützung dieser Textverarbeitung führen: So werden Open-Office-Textdateien direkt von den TM-Systemen OmegaT und DéjàVu gelesen und verarbeitet.

An übersetzungsspezifischer Open-Source-Software sind vor allem ForeignDesk und OmegaT zu nennen. ForeignDesk wurde vom Lokalisierungsunternehmen Lionbridge zunächst als hausinternes Tool entwickelt und danach unter der GPL-Lizenz frei zur Verfügung gestellt. OmegaT hingegen ist als Kooperationsprojekt entstanden und erfreut sich einer regen Entwicklungsarbeit. Kennzeichnend für OmegaT ist vor allem seine Plattformunabhängigkeit, da es in JAVA programmiert ist und daher sowohl auf Windows- als auch auf Unix- und Applerechnern lauffähig ist.

Durch diese Entwicklungen ist es für einen Übersetzer derzeit ohne weiteres möglich, eine komplette Softwareausstattung auf der Basis von Open-Source für seine tägliche Arbeit einzurichten: Die Basis dafür bildet das freie Betriebssystem Linux, das seit ca 10 Jahren als Open-Source im WWW frei heruntergeladen werden kann. Die meisten Linux-Distributionen enthalten bereits OpenOffice (Textverarbeitung, Tabellenkalkulation und Präsentationsmodul). Mit OmegaT kann ein freies TM-System eingesetzt werden. Woran noch ein gewisser Mangel an überzeugenden Open-Source-Projekten besteht, ist vor allem im Bereich Terminologieverwaltung und translationsspezifischer Projektverwaltung.

Ausblick

Zu erwarten ist in Zukunft ein Ausbau des Open-Source-Bereiches auch im Bereich Translationstechnologie. Hier wird sich der Einsatz allgemeiner genormter Dokumenten- und Austauschformate verstärkt durchsetzen (vgl. OASIS Organization for the Advancement of Structured Information Standards und OSCAR Open Standards for Container/Content Allowing Re-use).

Im allgemeinen aber wird eine weitere Integration von CAT-Tools in die traditionellen Arbeitsabläufe und Standardsoftware zu erwarten sein. Beste Voraussetzungen dafür hat Trados, das teilweise von Microsoft aufgekauft wurde und dadurch eine hohe Kompatibilität zu Microsofts Officeprodukten erreicht hat. Eine sehr enge Zusammenarbeit wird vor allem in Hinblick auf Microsofts .net-Strategien (Webservices) zu beobachten sein. Dieser Schritt wird eine tiefgehende Kooperation von Übersetzern über das Internet ermöglichen: z.B. den Austausch und die gemeinsame Nutzung und Wartung von Translation-Memories über das Netz, von Terminologiedatenbanken und allgemein jeder Art von Sprachressourcen.

Bei allen Vorteilen, die der Einsatz computergestützter Werkzeuge bereits heute und verstärkt in der Zukunft bringen wird, dürfen die damit verbundenen Nachteile aber nicht ungenannt bleiben. Zu nennen sind in diesem Zusammenhang vor allem folgende Punkte:

- Notwendigkeit der Aneignung von Zusatzkompetenzen (Umgang mit dem PC: Betriebssystem, Textformate, Austauschformate, etc.);
- erhöhter Lernaufwand (Software-Updates, spezifische Software) und Zeitaufwand (in der Verwendung, Konvertierungen, Formate, etc.);
- Ändern des Workflows: Umstellen des gewohnten Ablaufs bei Übersetzungen auf die u.U. von der Software vorgegebenen Abläufe. Die Flexibilität des Übersetzers wird häufig durch stringent von der Software vorgegebene Abläufe eingeengt.

Große Herausforderungen stellt diese Entwicklung an die Übersetzer Ausbildung, die den angehenden professionellen Übersetzer nicht nur an bestehende Tools und ihre Funktionsweise heranführen muss, sondern vor allem auch die damit verbundenen Änderungen im Ablauf des Translationsprozesses und in der Abwicklung von Translationsprojekten mit berücksichtigen sollte.

Hürden für die Ausbildung bilden sowohl die Knappheit an Experten, die zur Ausbildung herangezogen werden können, als auch die Kosten für die Software. Für das zweite Problem könnte der Bereich Open-Source einen Ausweg bilden. Das Konzept der Optimum-Cost-Ware (OCW) beinhaltet für die Übersetzer Ausbildung einerseits das Verwenden bestehender Softwareressourcen zu einem geringen Preis oder als Open-Source, andererseits aber auch ein Abgehen vom vielfach angewandten Prinzip, wonach durch das Verwenden *einer* bestimmten Software die Methodik computergestützter Übersetzungs- und Lokalisierungsprojekte vermittelt werden soll. Dieses Prinzip wird durch OCW umgekehrt zum Vermitteln von Vorgehensweisen und Methodologien anhand *mehrerer* unterschiedlicher Softwareprodukte (vgl. Mata 2004).

Mit welchen Inhalten die Translationstechnologie im Curriculum der Ausbildungsgänge vertreten sein sollte, wurde unter anderem auch im Rahmen von EU-Projekten erarbeitet. Das **Letrac**-Projekt, im Rahmen des 4. Rahmenprogramms

der EU gefördert, versuchte, die notwendigen Basiselemente im Curriculum festzulegen und ein Gerüst an Kompetenzen der Translationstechnologie zu beschreiben, das in die einzelnen Ausbildungsprogramme integriert werden kann. Die aktuellere Grundlage bietet das bereits erwähnte Projekt **eColoRe**, ein EU-finanziertes Projekt zur Entwicklung allgemein zugänglicher und erneuerbarer eContent-Lokalisierungs-Ressourcen zur Unterstützung der Übersetzerausbildung im Bereich Informations- und Kommunikationstechnologie, mit dessen Hilfe "aufgabenorientierte Mustertexte und pädagogische Hilfsmittel" zur Verfügung gestellt werden sollen. Auch wenn hier spezifisch auf den Bereich der Lokalisierung abgestellt wird, können doch sinnvolle Rückschlüsse auf die Erfordernisse einer allgemeinen Ausbildung in Translationstechnologie gezogen werden.

Die vielfältigen Möglichkeiten, die von der Translationstechnologie heute geboten werden, lassen den Vergleich mit dem Feuer des Prometheus durchaus zu. Für den professionellen Übersetzer, der von seiner Arbeit lebt, ist die Translationstechnologie kaum mehr wegzudenken. Damit kommt zu den verschiedenen Epitheta, mit denen der moderne Mensch bezeichnet wird, durch die technologische EDV-Kompetenz das *digitalis* dazu. Der informierte Anwender wird zum *homo digitalis sapiens*. Im Bereich der Translation ist dies der Übersetzer, der über die gegenwärtig möglichen Anwendungen Bescheid weiß. Der Übersetzer, der mit diesen Instrumenten arbeitet und produziert, wird zum *homo digitalis faber*, während der *homo digitalis oeconomicus* das Kosten-Nutzen-Verhältnis abwägt, d.h. mögliche lohnende Einsatzgebiete (Texttypen, Fachbereiche, Alignment, etc.) kennt.

<i>homo transferens digitalis</i>			
<i>sapiens</i> Kenntnis der Möglichkeiten	<i>faber</i> Einsatz der Translations- technologie	<i>oeconomicus</i> Kenntnis über die Einsatzoptimierung	<i>ludens</i> Ausschöpfen der Funktionalität, Kritik und Weiterentwicklung

Die Faszination durch die stetige Weiterentwicklung und die potentielle Arbeitserleichterung lässt einen Teil der Übersetzer zum *homo digitalis ludens* (Nerd, Freak) werden, der seine digitalen Werkzeuge in den Mittelpunkt stellt und damit spielt, d.h. die volle Funktionalität ausschöpft und Neues ausprobiert sowie der Übersetzer, der an der Softwareentwicklung mitarbeitet und sein ganzes Know-How einbringt. Wünschenswert ist daher, dass dieser spielerische Effekt beim Umgang mit Translationstechnologie möglichst groß bleibt. Schließlich sollte die Arbeit ja auch Spass machen.

Links

TMX Translation-Memory-Exchange Standard

<http://www.lisa.org/tmx/>

SRX Segmentation Rules Exchange Format

<http://www.lisa.org/oscar/seg/drafts/srx/srx05-20031027.htm>

TBX TermBaseExchange-format

<http://lisa.org/tbx/>

Human Language Technology:

<http://www.hltcentral.org/>

Tools Garage:

<http://www.lai.com/tg.html>

Translatum Forum:

<http://www.translatum.gr/dics/translation-memory.htm>

Google Directory zu CAT:

http://directory.google.com/Top/Computers/Software/Globalization/Computer_Aided_Translation/Software/

DMOZ Open Directory zu CAT:

http://dmoz.org/Computers/Software/Globalization/Computer_Aided_Translation/

GNU General Public License:

<http://www.gnu.org/copyleft/gpl.html>

Linux-Distributionen im WWW:

<http://www.linuxiso.org/>

Office-Paket mit Textverarbeitung, Tabellenkalkulation und Präsentationsmodul:

<http://www.openoffice.org>

Translation-Memory OmegaT:

<http://www.omegat.org/omegat/omegat.html>

Literaturverzeichnis

Chesterman, A. (1995): „The successful translator: The evolution of homo transferens“.

In: *Perspectives: Studies in Translatology* 1995:2, 253-270

Gow, F. (2003): “Extracting Useful Information from TM Databases” In: *Localisation Focus*, Vol. 2 Issue 4, 18-20

Höcker, M. (2003): “*eCoLoRe Translation Memory Survey 2003*”. Berlin, Bundesverband der Dolmetscher und Übersetzer e.V. (BDÜ)

http://ecolore.leeds.ac.uk/downloads/2003.05_bdue_survey_analysis.doc

Hunt, T. (2003): “Translation Technology Failures and Future”. In: *Lisa Newsletter*, vol XII nr. 1.4, 2003

Lommel, A. (2004): *Lisa 2004 Translation Memory Survey*. Translation Memory and Translation Memory Standards. *LISA*, www.lisa.org

- Mata, M. (2004): „Optimum Cost-Ware' in Translator Trainig for the Localisation Market" in: *Localisation Focus*, Vol. 3 Issue 4. 11
- Melby, A. (1998): „Eight Types of Translation Technology" online paper
<http://www.ttt.org/technology/8types.pdf> 01/2005