

2007

Automatische Terminologieextraktion

Uwe Muegge, *Monterey Institute of International Studies*

Uwe Mügge

Automatische Terminologieextraktion

1 Einleitung

1.1 Terminologie als ein zentrales Übersetzungsproblem

¹Ein Wesensmerkmal von Fachtexten ist deren sprachliche Spezialisierung. Diese Spezialisierung kommt vor allem in der Verwendung von fachspezifischen Benennungen zum Ausdruck, d. h., in der Fachsprache werden Benennungen verwendet, die entweder in der Gemeinsprache überhaupt nicht verwendet werden (z. B. Benutzeroberfläche in der Informatik) oder in der Fachsprache für einen anderen Begriff stehen (z. B. Mutter in der Mechanik) als in der Gemeinsprache. Deshalb können Fachtexte nur dann sachlich richtig von einer Sprache in eine andere übertragen werden, wenn bei der Übersetzung die entsprechende multilinguale Terminologie zur Verfügung steht.

1.1.1 Normierung

1.1.1.1 DIN 2345, LISA QA Model und SAE J2450

In den letzten Jahren hat sich die Einsicht, dass systematische Terminologiearbeit einen unverzichtbaren Bestandteil jedes Übersetzungsprojekts darstellt, in verschiedenen Normen niedergeschlagen. So fordert DIN 2345 Übersetzungsaufträge, die im deutschsprachigen Raum wohl bekannteste Norm im Übersetzungsbereich, im Abschnitt „Analyse des Ausgangstextes“:

Vor und während des Übersetzungsprozesses muss der Übersetzer feststellen, inwieweit die benötigte Terminologie bereits vorliegt oder noch erarbeitet werden muss und inwieweit eine weitere Einarbeitung in das Fachgebiet erforderlich ist (DIN 2345, 1998:11).

Das für Übersetzung/Lokalisierung von Software, Dokumentation, Hilfesystemen usw. entwickelte Qualitätssicherungsmodell der Localisation Industry Standards Association (LISA) sieht ebenfalls die Glossarerstellung/-aktualisierung vor der eigentlichen Übersetzungsphase vor (LISA 1998:5). Interessant ist in diesem Zusammenhang auch die im Entstehen befindliche Industrienorm J2450 für die Bewertung von Übersetzungsqualität, die von der amerikanischen Society of Automotive Engineers (SAE) entwickelt wird. Diese Norm steht beispielhaft für die sich in Teilen der Industrie durchsetzende Einsicht, dass terminologische Fehler zu den für die Handhabung eines Produktes folgenschwersten gehören: Falsche Terminologie (Wrong Term) wird nach

¹ Dieser Beitrag ist eine punktuell aktualisierte Fassung eines Vortrags auf der TQ2000, Internationale Fachtagung Translationsqualität, in Leipzig 28./29.10.1999.

J2450 als einzige Kategorie mit der maximalen Fehlerpunktzahl belegt (Godden 1999:14).

1.1.1.2 Normen unterstützen nur ansatzweise bei der praktischen Umsetzung

Die in den oben erwähnten Normen zum Ausdruck kommende strikte Forderung nach effektiver Terminologearbeit steht jedoch ein evidenten Mangel an konkreten Hilfen zur Umsetzung in der Praxis gegenüber. So finden sich einerseits in keiner der angesprochenen Normen Hinweise zu Prozessen und Werkzeugen, mit denen die Terminologearbeit vor Übersetzungsbeginn bewältigt werden kann. Andererseits ist die Aufgabe, auch nur für die heute wichtigsten Fachgebiete und Sprachkombinationen genormte multilinguale Terminologien bereitzustellen, von den bestehenden Normierungsorganisationen nicht zu leisten. So kann es nicht verwundern, dass z. B. die auf CD-ROM erhältliche Terminologiedatenbank des DIN (DIN-TERM) lediglich drei Sprachen abdeckt, nämlich Deutsch, Englisch und Französisch (vgl. Beuth 2000).

1.1.2 Effizientes Terminologiemanagement

Terminologische Qualität bedeutet konsistente Verwendung einer korrekten Benennung in der Zielsprache für die entsprechende Benennung in der Ausgangssprache. Die Gewährleistung dieser Qualität auf effiziente Art und Weise ist Aufgabe des Terminologiemanagements. Neben der gleichlautenden Forderung der oben dargestellten Normen ist es vor allem der Effizienzgedanke, der die Terminologearbeit vor dem Übersetzungsbeginn nahe legt. Denn nur wenn bereits beim Übersetzungsbeginn ein komplettes Wörterbuch zur Verfügung steht, ist es möglich, dass

- Übersetzer mit maximaler Produktivität arbeiten, weil sie sich ganz aufs Übersetzen konzentrieren können. Das heißt, Übersetzer müssen ihre Arbeit nicht wie heute üblich immer wieder für die Terminologearbeit unterbrechen, sondern bekommen für sämtliche ausgangssprachlichen Benennungen in einem zu übersetzenden Text automatisch die zielsprachlichen Entsprechungen angezeigt;
- bei umfangreichen Projekten, bei denen Teams von räumlich verteilt arbeitenden Übersetzern eingesetzt werden, der Managementaufwand für die Terminologieangleichung entfällt und
- Kundenkorrekturen im Terminologiebereich weitgehend vermieden werden können, da das Glossar dem Auftraggeber bereits vor Übersetzungsbeginn zur Autorisierung vorgelegt werden kann.

1.2 Bisher gängige Terminologieextraktionsverfahren

Terminologearbeit vor dem Übersetzungsbeginn bedeutet, dass in einem ersten Schritt die in einem ausgangssprachlichen Text verwendeten Benennungen identifiziert, d. h. extrahiert werden, damit in einem zweiten Schritt die ziel-

sprachlichen Entsprechungen recherchiert werden können. Es folgt eine kurze Übersicht über die drei bislang allgemein verfügbaren Verfahren zur Terminologieextraktion.

1.2.1 Manuelle Terminologieextraktion

Der unbekannte Text wird von einem Terminologen, Übersetzer oder einer Hilfskraft von Anfang bis Ende gelesen/überflogen, und dabei werden Terminologiekandidaten ggf. mithilfe geeigneter Makros in eine Extraktionsliste eingetragen. Dieses Verfahren ist die einzige allgemein bekannte Methode zur Terminologieextraktion, die sich jedoch ungeachtet ihres geringen Komplexitätsgrades keiner großen Beliebtheit erfreut, da die manuelle Extraktion insbesondere bei großen Übersetzungsprojekten ausgesprochen personal, zeit und kostenintensiv ist.

1.2.2 Indexwerkzeuge

Derartige Anwendungen erstellen Listen sämtlicher in einem Text verwendeten Wörter, die ggf. mit einem bestehenden Wörterbuch oder so genannten Stopplisten abgeglichen werden können. Die im IBM TranslationManager bereitgestellte Funktion „New Terms List“ ist ein Beispiel eines in eine Translation-Memory-Umgebung integrierten Indexwerkzeugs (vgl. IBM 1998). Ein schwerwiegender Nachteil derartiger Werkzeuge ist die Tatsache, dass sie aufgrund ihrer mangelnden linguistischen Intelligenz nicht in der Lage sind, zwischen Gemeinsprache und Fachsprache zu unterscheiden. Da Indexwerkzeuge vielmehr nach dem simplen Prinzip des Mustervergleichs auf Wortebene arbeiten, ist es diesen Anwendungen auch nicht möglich, zu einem Wort gehörende Wortformen (z. B. „spleißen“ und „spleißt“ im Deutschen) oder aus mehreren Wörtern bestehende Benennungen (z. B. „Flat Panel Display“ im Englischen) zu erkennen. Deshalb sind mit Indexwerkzeugen erstellte Extraktionslisten einerseits unvollständig (Silence), und zum anderen enthalten sie einen hohen Anteil von terminologisch nicht relevanten Wörtern (Noise). Je nach Sprache des zu bearbeitenden Textes und Umfang der zur Verfügung stehenden Wörterbücher/Stopplisten müssen mit Indexwerkzeugen erstellte Extraktionslisten entweder aufwendig manuell nachbearbeitet werden oder sind für Übersetzungszwecke völlig unbrauchbar.

1.2.3 Komplexe Konkordanzwerkzeuge

Stehen Paralleltexte, d. h. Originaltext und Übersetzung, in maschinenlesbarer Form zur Verfügung, werden in einem ersten Schritt die einzelnen Sätze/Segmente im Quelltext ihrer jeweiligen Entsprechung in der Übersetzung zugeordnet. In einem zweiten Schritt wird dann mithilfe linguistischer und/oder statistischer Methoden, ggf. unter Einsatz von Wörterbüchern, ein zweisprachiges Glossar der in diesen Texten verwendeten fachsprachlichen Benennungen erzeugt. Ein Beispiel für ein kommerzielles Produkt, das Terminologie auf diesem Wege extrahiert, ist der Xerox TermFinder (vgl. Xerox 1999). Der Umstand, dass dieses Werkzeug für den stolzen Preis von USD 25.000 angeboten

wird, macht die Terminologieextraktion nach diesem Verfahren allerdings nur für einen sehr beschränkten Anwenderkreis interessant.

2 Automatische Terminologieextraktion

2.1 Terminologieextraktion mit gängigen Maschinellen Übersetzungssystemen

Es ist eine wenig bekannte Tatsache, dass viele der heute populären Maschinellen Übersetzungssysteme für die automatische Terminologieextraktion brauchbar sind. Die Eignung für diesen Zweck scheint selbst den Herstellern dieser Systeme nicht bewusst zu sein, denn bislang ist die Terminologieextraktion mit den gängigen Maschinellen Übersetzungsprodukten erstens nur auf Umwegen möglich und zweitens in der Produktliteratur nicht dokumentiert.

Dabei ist das zu Grunde liegende Prinzip relativ simpel. Die gängigen kommerziellen maschinellen Übersetzungssysteme sind nicht nur mit leistungsfähigen linguistischen Funktionen ausgestattet, sondern verfügen auch über Wörterbücher, die zwischen Vokabular der Gemeinsprache und verschiedener Fachsprachen in den unterstützten Landessprachen unterscheiden. Bei der maschinellen Übersetzung legen diese Systeme eine Liste neuer, d. h. dem Systemlexikon nicht bekannter Wörter an, bei denen es sich mit hoher Wahrscheinlichkeit um Terminologie handelt. Einige Maschinelle Übersetzungssysteme gestatten darüber hinaus den Zugriff auf eine Liste der bei der jeweiligen Übersetzung als Terminologie erkannten Wörter. Die Kombination beider Dokumente ergibt, wie im Folgenden gezeigt wird, eine qualitativ hochwertige Terminologieextraktionsliste.

2.2. Das Extraktionsverfahren Schritt für Schritt

2.2.1 Maschinelles Übersetzen

Aufgrund der Tatsache, dass die gängigen Maschinellen Übersetzungssysteme die Terminologieextraktion nicht als eigenständige Funktion anbieten, muss die Extraktionsliste auf dem Umweg der maschinellen Übersetzung erzeugt werden. Im Rahmen der Maschinellen Übersetzung erstellen die gängigen Maschinellen Übersetzungssysteme zumindest eine Liste neuer Wörter und ggf. ein Glossar der bei der Übersetzung verwendeten Terminologie.

2.2.2 Bearbeiten des Glossars

Das Glossar ist je nach Spezialisierungsgrad des Quelldokuments und Sachkenntnis der Übersetzerin von mehr oder minder großem Nutzen. Da die für die einzelnen Sachgebiete in Maschinellen Übersetzungssystemen angelegten Terminologiebestände oftmals nur den Grundwortschatz des jeweiligen Sachgebiets abdecken und aktuelle Wortschöpfungen nicht berücksichtigen, ist die Ausbeute an nützlicher Terminologie zumindest im Fall von hoch spezialisierten und/oder innovativen Texten für Übersetzerinnen mit langjähriger Erfahrung auf dem je-

weiligen Sachgebiet eher gering. Umgekehrt gilt, dass bei Texten mit Einführungscharakter, die von Übersetzern bearbeitet werden, die über wenig oder gar keine relevanten Fachkenntnisse verfügen, die Glossare durchaus brauchbar sind.

2.2.2.1 Filtern des Glossars

Das Glossar enthält im Rohformat z. B. die im Quelldokument verwendeten quellsprachlichen Benennungen, die zielsprachlichen Entsprechungen sowie eventuell Kontextbeispiele, Angaben zum Sachgebiet, Steuerzeichen usw.

Mit einem geeigneten Word-Makro kann das Glossar auf Tastendruck gefiltert, sortiert und auf die weitere Verarbeitung (z. B. die Kombination mit der Liste neuer Wörter) vorbereitet werden (s. Abb. 1).

area	Bereich (Datenverarbeitung)	Most companies start with one of these areas and then plan to add more in the future.
availability	Eingabebereitschaft (Datenverarbeitung)	In addition, the ability to offer online information on order history, product availability, price calculations and delivery times enables customers to make immediate buying decisions.
basic	basisch (Chemie)	Because it is the basic technology of the system which will enable companies to take advantage of what the very near future will bring.
benefit	Vorteil (Allgemeines Technisches Vokabular, Datenverarbeitung, Wirtschaft und Handel)	Here and now benefits
box	Box (Datenverarbeitung)	in box:
business need	Anforderung in Ihrem Unternehmen (Allgemeines Technisches Vokabular)	Traditional ERP systems meet traditional business needs:

Abb. 1

2.2.3 Bearbeiten der Liste neuer Wörter

2.2.3.1 Filtern der Liste neuer Wörter

Die Liste neuer Wörter ist im Rohformat recht unübersichtlich. Neben dem wesentlichen Gehalt an Terminologiekandidaten und Kontext enthält die Liste neuer Wörter für die weitere Verarbeitung irrelevante Informationen (z. B. Steuerzeichen).

Mit einem geeigneten Word-Makro kann die Liste neuer Wörter auf Tastendruck gefiltert, sortiert und auf die weitere Verarbeitung (z. B. Kombination mit dem Glossar) vorbereitet werden (s. Abb. 2).

3-tier architecture		The secret behind the system is its 3-tier architecture, which means that Web functionality is fully integrated into the solution.
applications mean		Will adding new applications mean starting again from scratch and re-coding the existing ERP business logic?
business community		As a supplier of ERP systems for mid-sized companies, the needs and challenges facing this segment of the business community are naturally our main focus.
business community integration		In brief, Business Community Integration (BCI) makes the Internet a part of everyday business operations by moving business processes online, connecting suppliers, partners and customers.
business dimension		Customers are the business dimension that companies neglect at their own peril.
business environment		the competitive landscape the company operates within, including the company's position and potential

Abb. 2

2.2.3.2 Filtern von terminologisch irrelevanten Einträgen

Da die Liste neuer Wörter als Basis für den Aufbau bzw. die Erweiterung einer Terminologiedatenbank dienen soll, müssen terminologisch irrelevante Einträge (Noise) aus der Liste entfernt werden. Die Entscheidung, welche Terminologiekandidaten in der Liste verbleiben und welche entfernt werden sollen, liegt im Ermessen des mit der Terminologieextraktion beauftragten Übersetzers/Terminologen und muss deshalb manuell erfolgen. In der Beispieldatei könnte z. B. der zweite Eintrag von oben „applications mean“ als terminologisch irrelevant verworfen werden.

2.2.3.3 Filtern überflüssiger Kontextinformation

In der Liste unbekannter Wörter werden u. U. mehrere Kontexte angegeben. Diese Funktion ist durchaus sinnvoll, wenn möglichst aussagekräftige Beispielsätze für den Aufbau einer Terminologiedatenbank gewünscht werden, denn

wird nur ein Kontext extrahiert, könnte dies z. B. eine Überschrift sein, die nur aus dem Terminologiekandidaten besteht, was den Informationswert des Kontexts auf null reduziert. Die Entscheidung, welche Kontexte in der Liste verbleiben und welche entfernt werden sollen, liegt im Ermessen des mit der Terminologieextraktion beauftragten Übersetzers/Terminologen und muss deshalb manuell erfolgen.

Steht bei der Terminologieextraktion in erster Linie der Effizienzgedanke, d. h. die schnelle Erstellung eines Wörterbuchs, im Vordergrund, kann die Liste unbekannter Wörter mit einem geeigneten Word-Makro so gefiltert werden, dass wie in der Beispieldatei stets nur ein Kontext zu den einzelnen Terminologiekandidaten angezeigt wird.

2.2.3 Terminologieextraktionsliste als Kombination von Glossar und Liste neuer Wörter

Weder das bei der Übersetzung eines Dokuments vom Maschinellen Übersetzungssystem erzeugte Glossar noch die Liste neuer Wörter enthält für sich genommen in der Regel nicht den gesamten Terminologiebestand des Quelldokuments. Dieser Mangel ist systembedingt: Im Glossar fehlen Benennungen, die nicht im Systemlexikon des Maschinellen Übersetzungssystems enthalten sind, und die Liste neuer Wörter enthält keinen Grundwortschatz. Wird eine möglichst vollständige Terminologieextraktionsliste angestrebt, liegt die Kombination von Glossar und Liste neuer Wörter nahe.

In den vorangegangenen Schritten wurde die Liste neuer Wörter dahin gehend manuell optimiert, dass irrelevante Einträge (Noise) entfernt wurden. Das Glossar wurde dagegen nicht weiter manuell nachbearbeitet, da zumindest bei hoch spezialisierten Quelldokumenten die Mehrzahl der relevanten Benennungen nicht erfasst wird (Silence) und deshalb das Glossar bestenfalls als Auswahlliste infrage kommt.

Mit einem geeigneten Word-Makro kann ein ausgewählter Eintrag des Glossars auf Tastendruck zunächst gefiltert (z. B. Entfernen der Angabe des Sachgebiets) und dann in die Liste neuer Wörter eingefügt werden. Die daraus resultierende Terminologieextraktionsliste ist aufgrund des niedrigen Noise/Silence-Anteils qualitativ sehr hochwertig (s. Abb. 3).

2.2.5 Nutzen der Terminologieextraktionsliste

2.2.5.1 Verbesserte Kalkulation

Die Terminologieextraktionsliste ist bereits in der vorläufigen, hauptsächlich einsprachigen Form ein wertvolles Hilfsmittel: Anhand dieser Liste kann der Übersetzungsdienstleister/Terminologe den für die Terminologearbeit im Rahmen eines Übersetzungsprojekts notwendigen Ressourcenbedarf detailliert ermitteln. Dadurch können in der Angebotsphase für komplexe Übersetzungspro-

jekte fundierte Vorhersagen bezüglich Liefertermin und Kosten gemacht werden.

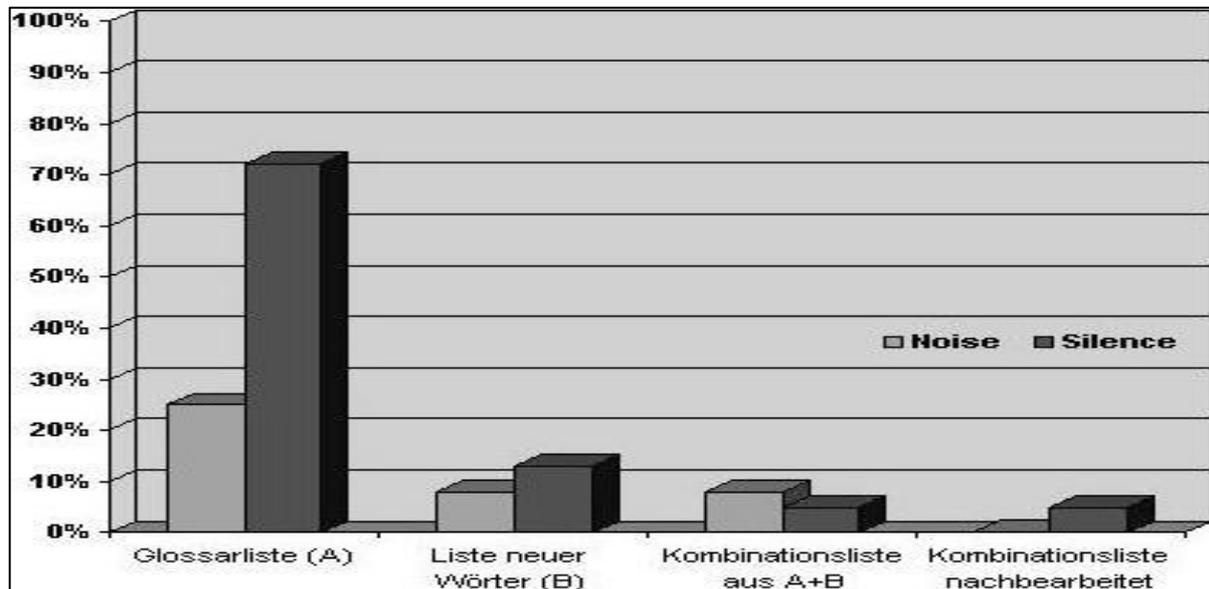


Abb. 3

2.2.5.2 Einsprachiges Glossar

Normierter Sprachgebrauch im Sinne einer Corporate Language macht es in Umgebungen, in denen keine statische kontrollierte Sprache eingesetzt wird, notwendig, dass die der Corporate Language zugrunde liegenden Glossare/Wörterbücher ständig aktualisiert werden. Mit dem oben beschriebenen Verfahren können aus großen Datenbeständen weitgehend automatisch innerhalb kürzester Zeit und mit relativ geringem Ressourceneinsatz für diesen Zweck geeignete Glossare erstellt werden.

2.3 Terminologierecherche auf der Basis der Terminologieextraktionsliste

Die nach dem oben beschriebenen Verfahren erstellte Terminologieextraktionsliste stellt eine hervorragende Ausgangsbasis für die Terminologierecherche vor dem Übersetzungsbeginn dar. Im Folgenden wird die Erweiterung der Terminologieextraktionsliste zu einem vollständigen zweisprachigen Wörterbuch dargestellt.

2.3.1 Erstellen des Importformats für das Terminologieverwaltungssystem

Bei der Terminologierecherche sollten nach Möglichkeit die in der Datenbankdefinition des jeweiligen Terminologieverwaltungssystems spezifizierten Eintragskategorien berücksichtigt werden. Deshalb ist es notwendig, die Terminologieextraktionsliste zunächst so zu erweitern, dass für jede Benennung die entsprechenden Felder zur Verfügung stehen.

Mit einem geeigneten Word-Makro kann die Terminologieextraktionsliste auf Tastendruck automatisch mit den benötigten Feldern erweitert werden. In Abb. 4 ist eine Beispieldatei für den späteren Import in Trados MultiTerm dargestellt.

```

**
<English>business community
<Part of Speech>noun
<Stem>business community
<Context>As a supplier of ERP systems for mid-sized companies, the needs
and challenges facing this segment of the business community are naturally
our main focus.
<Source>Booklet e-business
<German>
<Part of Speech>noun
<Stem>
<Context>
<Source>
**
<English>business community integration
<Part of Speech>noun
<Stem>business community integration
<Context>Business Community Integration (BCI) is the name given by AMR to
the next wave of the e-business revolution - and it's already happening.
<Source>Booklet e-business
<German>
<Part of Speech>noun
<Stem>
<Context>
<Source>
**

```

Abb. 4

2.3.2 Terminologierecherche in Trados-Translation-Memories mit dem Konkordanzwerkzeug eines Maschinellen Übersetzungssystems

Translation Memories gehören zu den zuverlässigsten Quellen terminologischer Information, insbesondere wenn es sich um projektspezifische, vom Auftraggeber autorisierte Datenbestände handelt. In den gängigen Translation-Memory-Produkten sind Konkordanzwerkzeuge integriert, die den Zugriff auf einzelne Worte, Satzfragmente oder Sätze in der Ausgangssprache und deren Entsprechung in der Zielsprache des jeweiligen Translation Memory ermöglichen sollen.

Das Konkordanzwerkzeug der Trados Workbench weist allerdings für die Terminologierecherche u. U. hinderliche Eigenschaften auf. So kann z. B. neben dem aktuell geöffneten Translation Memory nur ein weiteres Referenz-Translation-Memory durchsucht werden, und für die Suche in der Zielsprache muss auf umständliche Weise die Sprachrichtung umgekehrt werden.

Die in manchen Maschinellen Übersetzungssystemen integrierten Konkordanzwerkzeuge hingegen unterliegen keinen derartigen Einschränkungen, und

deshalb bietet sich die Nutzung von Trados-Translation-Memories außerhalb der Workbench an.

2.3.2.1 Recherche in Trados-Translation-Memories in der Word-Umgebung

Ein externes Konkordanzverfahren bietet sich vor allem dann anstelle der Trados-Word-Konkordanzfunktion an, wenn für ein Übersetzungsprojekt mehrere Translation Memories von Belang sind. Das ist z. B. bei Softwarelokalisierungsprojekten regelmäßig der Fall – hier sind bei der Terminologierecherche für die neue Version eines Anwendungsprogramms zumindest die Translation Memories der Ressource-Dateien, des Hilfesystems, der Dokumentation und des Marketingmaterials der früheren Version sowie ggf. das in ein Translation Memory konvertierte Microsoft-Glossar relevant.

Besonders interessant ist an einer derartigen Konkordanzlösung, dass

- eine unbegrenzte Anzahl von Translation Memories
- in beliebiger Kombination und
- in beiden Sprachrichtungen

durchsucht werden kann.

2.3.3 Terminologierecherche in anderen Quellen und mit anderen Werkzeugen

Für die Terminologierecherche kann ein breites Spektrum von Quellen relevant sein: von bestehenden autorisierten Wörterbüchern, Glossaren und Translation Memories (im Idealfall) über Parallelliteratur (elektronische Dokumente einschließlich Webseiten sowie gedruckte Materialien) bis zu einsprachigen Druckdokumenten, deren Inhalt nur unter erschwerten Bedingungen in ein elektronisches Terminologieverwaltungssystem übernommen werden kann (z. B. Dokumente in den Präsenzbeständen von Bibliotheken, Patentauslegestellen und Archiven). Für jedes dieser Szenarien stehen Lösungen zur Verfügung, die Übersetzerinnen/Terminologinnen die Terminologierecherche auf der Basis einer nach dem oben beschriebenen Verfahren erstellten Terminologieextraktionsliste erheblich erleichtern (s. dazu Mügge 2000).

2.3.4 Import der Terminologiedatei in MultiTerm

Nach Abschluss der Terminologierecherche kann das nunmehr vollständige Wörterbuch aus Word in ein Terminologieverwaltungssystem exportiert werden, in diesem Fall in Trados MultiTerm. Da bereits die Terminologieextraktionsliste im MultiTerm-Importformat angelegt wurde, muss das daraus resultierende Wörterbuch nun lediglich mit Word im Textformat gespeichert werden und kann dann nach dem Standardverfahren in MultiTerm importiert werden. Abb. 5 zeigt den in den vorangegangenen Schritten bearbeiteten Beispieleintrag, wie er nach dem Import in MultiTerm angezeigt wird.



Abb. 5

2.4 Genehmigung des Wörterbuchs durch den Auftraggeber

An dieser Stelle soll nochmals darauf hingewiesen werden, dass mit dem hier beschriebenen Verfahren umfassende, projektspezifische Wörterbücher bereits vor dem Übersetzungsbeginn erstellt werden können. Deshalb ist es möglich und sinnvoll, dem Auftraggeber das für die Verwendung in einem konkreten Übersetzungsprojekt vorgesehene Wörterbuch zur Korrektur und Genehmigung vorzulegen. So wird sichergestellt, dass im Übersetzungsprodukt nur die vom Auftraggeber gewünschten Benennungen erscheinen. Die Verwendung eines vom Auftraggeber autorisierten Wörterbuchs gibt den Übersetzerinnen Sicherheit und macht bei konsequenter Nutzung des Wörterbuchs eine Validierung der Terminologie nach der Übersetzung weitgehend überflüssig.

2.5 Nutzung des Wörterbuchs in der Übersetzungsumgebung

Sobald das Wörterbuch in MultiTerm importiert ist, steht die erstellte Terminologie in der Übersetzungsumgebung zur Verfügung. Die Übersetzungsumgebung besteht aus folgenden Komponenten: Translation Memory (Trados Workbench), Terminologieverwaltungssystem (MultiTerm) und Textverarbeitungsprogramm (Word), das über Schnittstellen Abfragen in den beiden vorgenannten Produkten ermöglicht.

2.5.1 Manuelle Abfrage mit einem Terminologieverwaltungssystem

Im Prinzip wäre die Nutzung des Wörterbuchs auch ohne Translation Memory möglich. In diesem Fall muss die Terminologie allerdings aktiv vom Übersetzer abgefragt werden, d. h., der Übersetzer steht vor der Aufgabe, jede Benennung im Quelltext auszuwählen, über eine Tastenkombination eine Abfrage im Terminologieverwaltungssystem durchzuführen, das Abfrageergebnis über eine weitere Tastenkombination in die Übersetzung einzufügen sowie das Abfrageergebnis ggf. manuell zu beugen. Dieser Vorgang ist so umständlich wie zeitaufwendig und stellt deshalb ein Produktivitätshindernis dar. Das verleitet Übersetzer dazu, im Falle von Benennungen, für die sie eine Entsprechung präsent

haben oder für die sich eine wörtliche Übersetzung anbietet, diese eigene Benennung zu verwenden, ohne über die Abfrage des Terminologieverwaltungssystems deren Zulässigkeit zu prüfen. Die Verwendung von vom autorisierten Wörterbuch abweichenden Benennungen stellt jedoch, wie eingangs dargestellt, einen schwerwiegenden Übersetzungsfehler dar. Zur Beseitigung derartiger systematischer Mängel sind bei Übersetzungsprojekten aufwendige Vorkehrungen im Bereich der Qualitätssicherung notwendig (z. B. intensives terminologisches Lektorat), was unter Effizienz Gesichtspunkten nicht wünschenswert ist.

2.5.2 Automatische Terminologieerkennung mit einem Translation-Memory-System

In einer Translation-Memory-Umgebung wie der Trados Workbench erfolgt die Terminologieerkennung dagegen automatisch, sofern diese Funktion aktiviert ist. Mit automatischer Terminologieerkennung werden im Quelltextfenster der Workbench sämtliche Benennungen mit einem farbigen Balken markiert, für die ein Wörterbucheintrag vorhanden ist (s. Abb. 6). Die zielsprachlichen Benennungen können durch Betätigen einer Schaltfläche, über einen Menübefehl oder eine Tastenkombination direkt in die Übersetzung eingefügt werden. Die Tatsache, dass zielsprachliche Benennungen der Übersetzerin automatisch angeboten werden und nicht umständlich aktiv abgefragt werden müssen, fördert die Akzeptanz des bereitgestellten Wörterbuchs bei Übersetzern ganz erheblich.

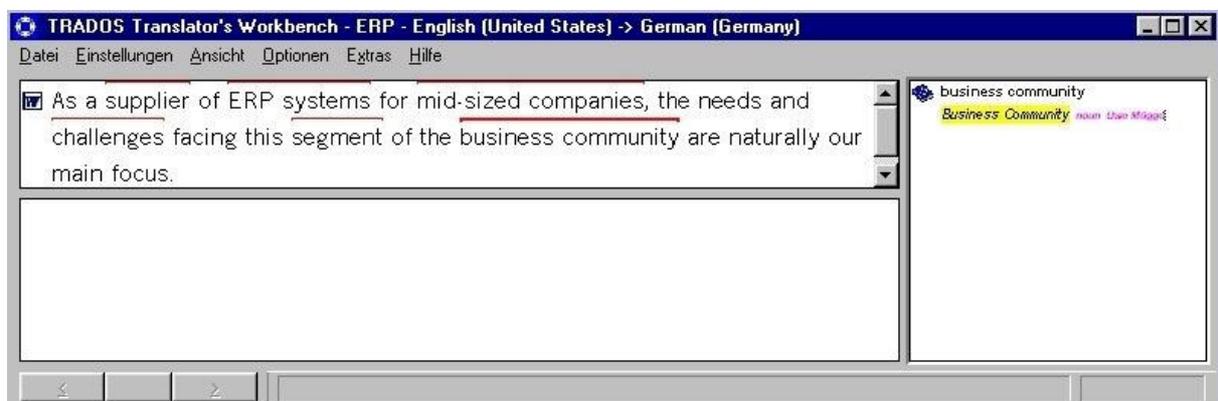


Abb. 6

Deshalb ist im Hinblick auf terminologische Konsistenz nicht nur der Einsatz eines Terminologieverwaltungssystems, sondern eines Translation-Memory-Systems zu fordern. Diese Forderung gilt unabhängig davon, ob für ein Übersetzungsprojekt wiederverwendbare Segmente vorliegen bzw. eine erneute Nutzung der aktuellen Übersetzung zu erwarten ist (das klassische Szenarium für den Einsatz von Translation-Memory-Technologie).

Mit dem hier beschriebenen Verfahren steht den Übersetzern zu Beginn des Übersetzungsprojekts ein umfassendes, projektspezifisches Wörterbuch zur Verfügung. Bei konsequentem Einsatz der automatischen Terminolo-

gieerkennungsfunktion ist deshalb die terminologische Konsistenz und damit ein wesentlicher Qualitätsfaktor gesichert.

2.6 Terminologievalidierung

Wie im vorangegangenen Absatz beschrieben, bietet die automatische Terminologieerkennung bei vollständigem Wörterbuch die Gewähr für terminologische Konsistenz. Die in den bislang mit dem hier beschriebenen Verfahren im praktischen Einsatz gemachten Erfahrungen zeigen jedoch, dass Übersetzerinnen das Angebot des automatischen Übersetzungsvorschlags im Rahmen der automatischen Terminologieerkennung nicht immer annehmen.

2.6.1 Mögliche Produktivitätseinbuße bei automatischer Terminologieerkennung

Ein nahe liegender Grund für das oben beschriebene qualitätsschädliche Verhalten von Übersetzern ist die gegenüber der Nutzung eigener Benennungen / wörtlicher Übersetzungen langsamere Verarbeitungsgeschwindigkeit der Workbench bei aktivierter automatischer Terminologieerkennung: In dieser Übersetzungsumgebung muss zumindest eine Tastenkombination betätigt und ggf. der Übersetzungsvorschlag noch gebeugt werden. Deshalb wird selbst von Expertenseite die Deaktivierung der automatischen Terminologieerkennung in der Workbench zur Steigerung der Verarbeitungsgeschwindigkeit empfohlen (s. Pleinc 1999). Denn eine höhere Verarbeitungsgeschwindigkeit schlägt sich zumindest für freiberufliche Übersetzer, die nach Zeilenhonorar arbeiten, in einem höheren Einkommen nieder. Aus diesem Grunde ist im Hinblick auf hohe terminologische Qualität des Übersetzungsprodukts ein Werkzeug zur möglichst automatischen Überprüfung/Validierung der Terminologie nach Übersetzungsende notwendig.

2.6.2 Automatische Terminologievalidierung

Da in der Workbench keine entsprechende Validierungsfunktionalität bereitgestellt wird, muss für diesen Zweck auf das Produkt eines Drittherstellers zurückgegriffen werden.

In dem hier eingesetzten Validierungswerkzeug kommt ein ebenso simples wie effektives Verfahren für die automatische Terminologievalidierung zum Einsatz: Aus einem bestehenden Wörterbuch wird eine zweispaltige Tabelle erzeugt, die in der linken Spalte die ausgangssprachlichen und in der rechten die zielsprachlichen Benennungen enthält. Damit gebeugte Wortformen erkannt werden können, muss der Eintrag aus dem unveränderlichen Wortstamm und dem nachfolgenden Platzhalter „*“ bestehen. Anhand dieser Tabelle wird ein simpler Mustervergleich in den mit der Workbench bearbeiteten Segmenten durchgeführt. Wird in einem zielsprachlichen Segment die in der Tabelle verzeichnete Entsprechung für eine im ausgangssprachlichen Segment gefundene Benennung nicht gefunden, erhält der Anwender eine Meldung (s. Abb. 7).

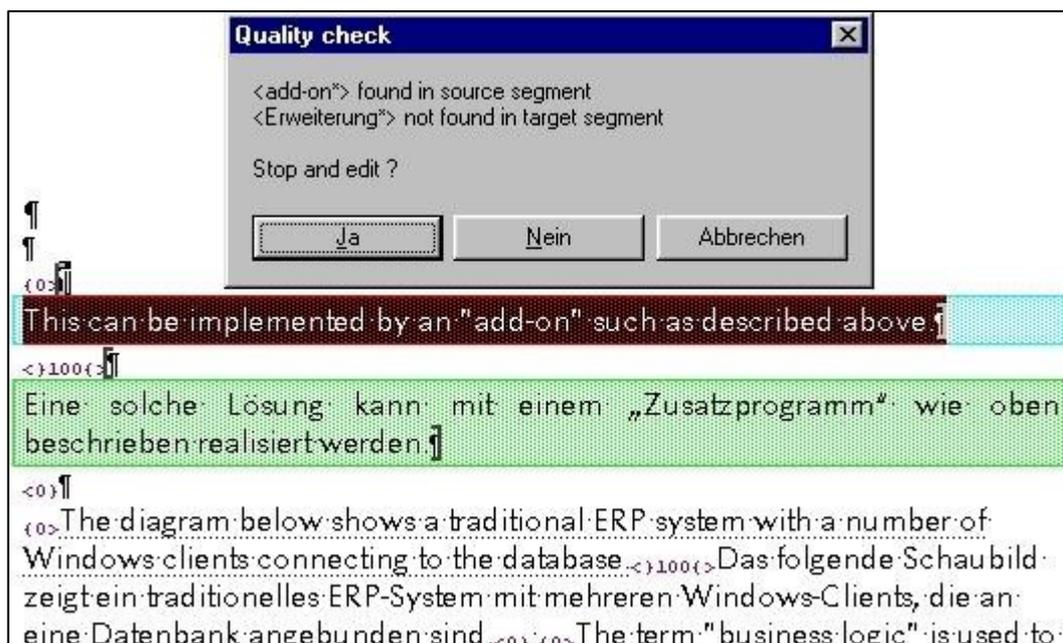


Abb. 7

Die Anforderungen des Validierungswerkzeugs können bereits bei der Erstellung der Terminologieextraktionsliste berücksichtigt werden, indem der Terminologiekandidat vom Word-Makro automatisch in eine zusätzliche Kategorie, z. B. Stem (s. Abb. 5), kopiert wird. Der Export aus MultiTerm in das Validierungswerkzeug wird durch entsprechende Filtereinstellungen in MultiTerm und eine makrobasierte Bearbeitung in Word realisiert.

Mit einem derartigen Validierungswerkzeug können selbst große Dokumente vollständig auf terminologische Konsistenz überprüft werden, weil dieses Werkzeug automatisch arbeitet und deshalb die Validierung vom Ressourcenbedarf kaum ins Gewicht fällt.

3 Fazit

Das hier vorgestellte Verfahren zur automatischen Terminologieextraktion sowie die darauf aufbauenden Prozesse zeichnen sich dadurch aus, dass sie einerseits sehr leistungsfähig sind und daher eine wesentliche Qualitäts- sowie Produktivitätssteigerung gegenüber herkömmlichen Übersetzungsumgebungen versprechen und andererseits mit heute existierenden Standardanwendungen realisiert werden können.

Die Vorteile im Einzelnen:

- Die automatische Terminologieextraktion erfasst Terminologie unabhängig vom Fachgebiet und erkennt auch Mehrwortbenennungen zuverlässig.
- Die Terminologieextraktionsliste enthält einen geringen Anteil von irrelevanten Einträgen und bietet Kontextinformationen zu jeder Benennung.
- Der Terminologieextraktionsprozess kann in die gängige Textverarbeitungsumgebung Word integriert und von da gesteuert werden.

- Als Resultat der Terminologieextraktion steht vor Beginn der eigentlichen Übersetzung ein umfassendes, projektspezifisches Wörterbuch zur Verfügung. Folglich fällt während der Übersetzung keine Terminologiarbeit in wesentlichem Umfang mehr an, und bei großen Übersetzungsprojekten muss terminologische Konsistenz nicht mehr aufwendig nachträglich in Dokumente hineinkorrigiert werden.
- In der Summe ergeben die einzelnen hier vorgestellten Verfahren ein vollständiges und durchgängiges Terminologieverwaltungssystem, das von der Terminologieextraktion über die Terminologierecherche bis zur Speicherung/Wiedergabe und der Validierung alle wesentlichen Bereiche der computergestützten Terminologiarbeit abdeckt.

Diese Werkzeuge und Verfahren sind vom Investitionsbedarf auch für selbstständige Übersetzer und Terminologen interessant und können bei entsprechenden technischen Kenntnissen umgehend in die berufliche Praxis umgesetzt werden.

Bibliographie

Beuth (2000): *Eintragsstruktur der Terminologiedatenbank DIN-TERM*.

<http://www.beuth.de/beuthprodukte/emedien/datenbanken/dinterm/struktur.html>.

DIN 2345 (1998): *Übersetzungsaufträge*. Berlin: Beuth.

Godden, Kurt (1999): *J2450. Society of Automotive Engineers Translation Quality Metric. Microsoft-PowerPoint-Präsentation*. Detroit: GM Service Operations.

IBM (1998): *IBM TranslationManager. Translator's reference part 3. Working with TranslationManager components and functions*.

<http://www-4.ibm.com/software/ad/translat/tm/tm2docu/tmref3.htm>.

LISA (1998): *LISA Quality Assurance Model*. Version 1. Adobe-Acrobat-Dokument. Féchy: Localisation Industry Standards Association.

Mügge, Uwe (2000): „Terminology work: Tools and processes that make a difference.“ *ATA Chronicle* 4: 15–20.

Pleinc, Denis (1999): *Translation Tools*. <http://www.open.hr/~dpleic/tools.html>.

Xerox (1999): *The Xerox Terminology Suite*.

<http://www.xerox-emea.com/xms/softsols/termin.htm>