

Professionelles Übersetzen mit STAR Transit

Judith Klein

STAR Deutschland GmbH

1 Translate it !

Transit (**Translate it**) ist ein professionelles Translation-Memory (TM) System entwickelt auf der Basis der langjährigen praktischen Erfahrung von STAR, eines der größten und erfolgreichsten Übersetzungsunternehmens der Welt. Die Kunden von STAR kommen aus der Automobilindustrie, der Informationstechnologie und dem Bereich Maschinenbau. Transit, das seit 1994 kommerziell als Produkt vertrieben wird, kommt Ende 1999 als 32-Bit-Version mit Unicode- und XML-Unterstützung als Version 3.0 auf den Markt.¹

Vorrangiges Ziel von Transit ist die Steigerung der Übersetzungsproduktivität durch automatisches Vorübersetzen auf der Grundlage geprüfter manuell angefertigter Übersetzungen. In Kombination mit integrierter Terminologieprüfung, ausgereifter übersetzungsspezifischer Funktionalität beim computergestützten manuellen Übersetzen und vollkompatiblen Schnittstellen zu den gängigen Textverarbeitungs- und Desktop-Publishing (DTP) Systemen bietet Transit daher eine gute Lösung für die im Übersetzungsgeschäft aktuellen Aufgaben:

- Die Übersetzung von Dokumenten, die mit unterschiedlichen Textverarbeitungs- und DTP-Systemen erstellt wurden, kann mit einer einzigen Software bearbeitet werden, wobei aufwendige Layoutaufgaben, die mit dem Übersetzungsvorgang primär nichts zu tun haben, nicht erforderlich sind.
- In der technischen Dokumentation muß die Übersetzung sehr umfangreicher Texte bei immer kürzern Lieferzeiten zu kostengünstigen Konditionen angefertigt werden. Da häufig aktualisierte Versionen bereits existierender Dokumente zu übersetzen sind und die technischen Texte selbst überdurchschnittlich viel sprachliche Wiederholungen aufweisen, können viele Passagen früherer Übersetzungen wiederverwendet werden. Während das Wiederverwenden bereits angefertigter Übersetzungen ohne TM-Technologie (über „Copy & Paste“) zwar möglich, aber fehleranfällig und sehr zeitaufwendig ist,

greift Transit automatisch auf bereits vorhandenes Übersetzungsmaterial zu.

- Die Qualität der Übersetzung, die auf sprachlicher Klarheit sowie inhaltlicher und terminologischer Korrektheit basiert, ist ein Kernfaktor in der technischen Dokumentation, da Fehler in Bedienungsanleitungen und Service-Handbüchern zu teuren Produkt- und Folgeschäden führen können. Die automatische Wiederverwendung manuell angefertigter und geprüfter Übersetzungen sowie die automatische Terminologieprüfung mittels der integrierten Terminologiekomponente TermStar dienen der Qualitätssicherung und unterstützen die terminologische Konsistenz der Übersetzung.

2 Arbeitsweise von Transit

Die komplette Version des Transit-Softwarepakets umfasst den übersetzungsspezifischen Transit-Editor mit integrierten Import- und Exportmechanismen zur Bearbeitung von Dokumenten unterschiedlicher Dateiformate², das vollintegrierte Terminologieverwaltungssystem TermStar, das Translation-Memory als Kernstück des Systems, sowie ein interaktives Alignment-Tool zum Konvertieren existierender Übersetzungen in Transit-Format. Transit operiert im Gegensatz zu anderen TM-Systemen nicht auf einem Datenbanksystem, sondern arbeitet mit einem effizienten Dateiverwaltungssystem. Die ausgangssprachliche Datei und die zielsprachliche Datei bilden ein *Sprachpaar*, wobei das Dateikürzel die jeweilige Sprache anzeigt. Liegt ein fertig übersetztes Sprachpaar in Transit-Format vor, dient es dem Translation-Memory als so genanntes *Referenzmaterial* für künftige Übersetzungsprojekte.

In Abbildung 1 ist der Übersetzungsprozess in Transit am Beispiel der englisch-deutschen Übersetzung eines Worddokuments dargestellt: Beim Import wird die Formatierungsinformation des englischen Worddokuments (Text.doc) herausgefiltert und in einer separaten Datei (Text.cod) abgelegt. Nach dem Filtern entsteht ein Transit-Sprachpaar (Text.eng, Text.ger), in dem die Formatierungsinformation in XML-Format kodiert ist³ und im Text durch Platzhalter (*Tags*) ersetzt ist. Im zweiten Schritt wird der Text in so genannte Segmente (Überschriften, Sätze, usw.) unterteilt, für die anschließend eine Vorübersetzung mit Zugriff auf das projektspezifische Referenzmaterial durchgeführt wird: die Übersetzung identischer Texteinheiten wird aus dem Translation-Memory automatisch in die ziel-

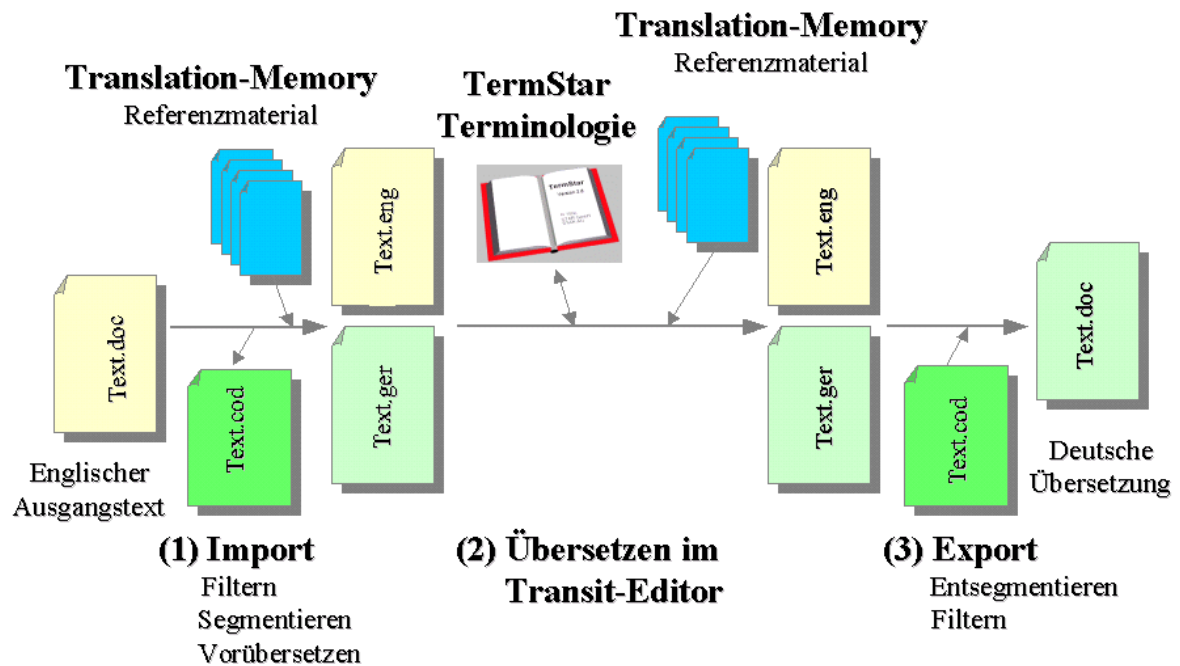


Abb. 1: (1) Import eines englischen Worddokuments, (2) Übersetzen im Transit-Editor, (3) Export der deutschen Transit-Datei ins Word-Format.

sprachliche Datei (Text.ger) eingefügt, so dass sie nach dem Import bereits vollständig übersetzte Segmente enthält. Im Transit-Editor wird der deutsche Text mit Zugriff auf die projektspezifischen Terminologiewörterbücher von TermStar und die Übersetzungsvorschläge aus dem Translation-Memory fertig bearbeitet. Anschließend wird die zielsprachliche Datei aus Transit exportiert, wobei die Segmentierung aufgehoben und die Formatierungsinformation wieder hinzugefügt wird. Als Ergebnis liegt die deutsche Übersetzung im Wordformat vor.

3 Übersetzen mit Transit

Der übersetzungsspezifische Transit-Editor (vgl. Abb.2) arbeitet mit fünf Fenstern, deren Anordnung benutzerspezifisch definierbar ist: (1) Ausgangstext-Fenster, (2) Zieltext-Fenster, (3) Fuzzy-Index Fenster mit Übersetzungsvorschlägen, (3) Wörterbuchfenster und (4) Notizfenster mit Informationen zum Status der Segmente (z.B. übersetzt, nicht übersetzt etc.). Der Transit-Editor basiert auf einer Emulation des weitverbreiteten Textverarbeitungssystems WinWord⁴ und bietet eine Vielzahl von Tastenkombinationen zum effizienten manuellen Übersetzen an. Globale Editieroperationen, wie z.B. gezieltes Suchen/Ersetzen, können für alle

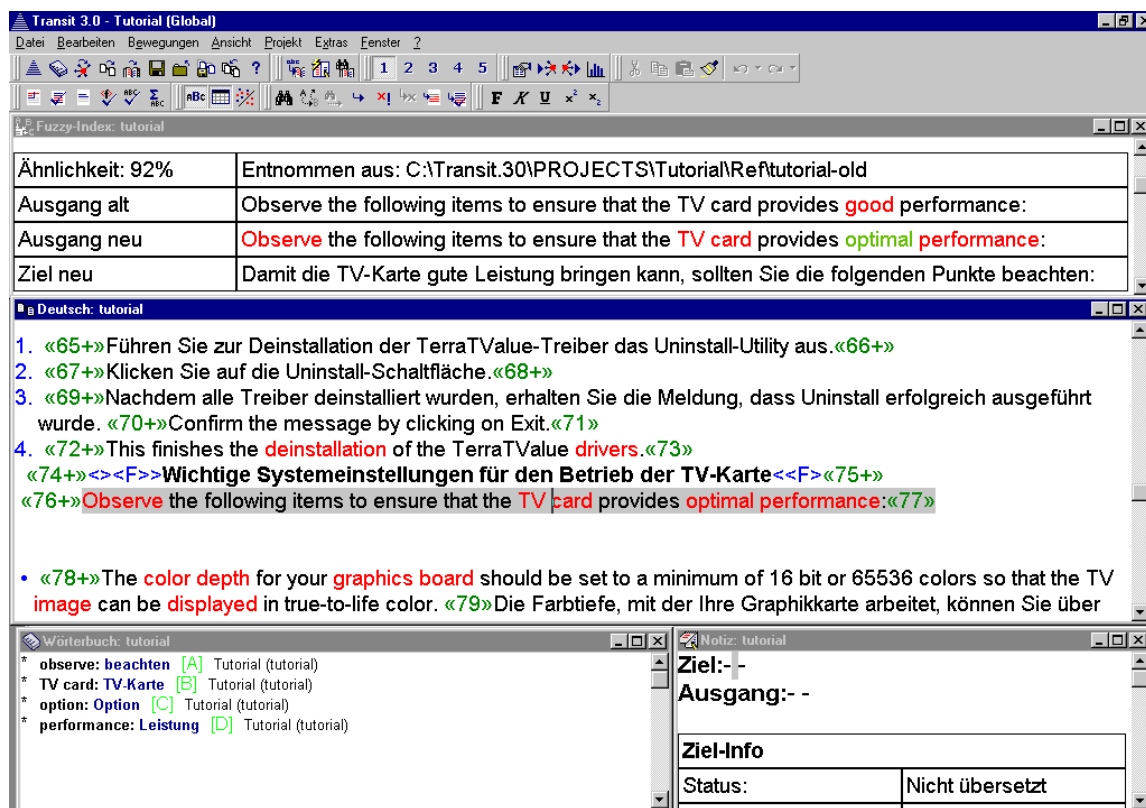


Abb. 2: Transit-Editor 3.0 mit Zieltext-Fenster, Fuzzy-Index Fenster, Wörterbuchfenster und Notizfenster. Das Ausgangstext-Fenster liegt im Hintergrund.

Dateien eines Übersetzungsprojekts gleichzeitig durchgeführt werden und vereinfachen so die beim Übersetzen anfallenden Routinarbeiten. Ab Version 3.0 wird der Text als Quasi-WYSIWYG angezeigt, so dass die für die Übersetzung notwendigen Formatierungsinformationen (z.B. die Spaltenbreite einer Tabelle, die für zielsprachliche Formulierungen beachtet werden sollte) direkt sichtbar sind.

Eine detaillierte Übersicht über die Übersetzung (wie viele Segmente automatisch vorübersetzt wurden, ob ein Segment auf der Basis eines Übersetzungsvorschlages übersetzt wurde, etc.) kann jederzeit abgerufen werden. Zur Feinabstimmung von Übersetzungseinheiten (Konkordanzen) können mit Hilfe eines ausgefeilten Filtermechanismus genau diejenigen Segmente angezeigt (bzw. ausgeblendet) werden, die z.B. ein bestimmtes Wort oder eine bestimmte Phrase enthalten. Integrierte Prüffunktionen erlauben es, den Text auf korrekte Orthographie⁵ und projektspezifische Fachterminologie zu überprüfen.

3.1 Interaktion zwischen Transit und Translation-Memory

Das Translation-Memory von Transit besteht aus Referenzdateien, den Transit-Sprachpaaren, die projektspezifisch zusammengestellt werden können. Neben Referenztexten, d.h. Übersetzungen, die zuvor mit Transit erstellt wurden oder über das Alignment-Tool ins Transit-Format umgewandelt wurden, werden auch Segmente aus dem aktuellen Übersetzungsprojekt als Referenzbasis verwendet: Sobald eine Texteinheit in Transit übersetzt wurde, dient diese Übersetzung sofort im Transit-Editor als Übersetzungsvorschlag für nachfolgende Segmente. Aus den Referenztexten kann ein projektspezifischer Referenzextrakt generiert werden, aus dem alle redundanten Segmente entfernt wurden, wobei Übersetzungsvarianten jedoch erhalten bleiben. Außerdem kann ein projektspezifisches Fuzzy-Match-Sprachpaar erzeugt werden, das alle diejenigen ausgangssprachlichen Segmente (plus deren Übersetzung) enthält, die zu den Segmenten des aktuellen Ausgangstextes einen zuvor festgesetzten Ähnlichkeitsgrad aufweisen.

Auf die Translation-Memory-Komponente wird zweifach zugegriffen: beim automatischen Vorübersetzen während des Imports und beim Übersetzen im Transit-Editor. Während des Imports werden alle Segmente, für die im Referenzmaterial identische Segmente (*Exact Match*) gefunden werden, automatisch vorübersetzt und brauchen nicht mehr bearbeitet zu werden. Außerdem kann optional eine teilweise Vorübersetzung durchgeführt werden, bei der Transit geringe Abweichungen (symmetrische Zahlen- und Formatierungsdifferenzen oder benutzerdefinierte Ausnahmen) in der Übersetzung automatisch anpaßt, die anschließend nur überprüft werden müssen (vgl. Abb.3).

| | |
|-------------|----------------------------------|
| Ausgang alt | Windows 95 or Windows 98 |
| Ausgang neu | Windows 98 or Windows 2000 |
| Ziel neu | Windows >98< oder Windows >2000< |

Abb. 3: Die Zahlen 95 und 98 wurden automatisch an die Zahlen des aktuellen Segments angepasst.

Beim manuellen Übersetzen im Transit-Editor liefert das Translation-Memory automatisch Übersetzungsvorschläge und bietet sie im Fuzzy-Index-Fenster an. Die Unterschiede zwischen ausgangssprachlichem Referenzsegment und dem aktuellen Segment werden farblich hervorgehoben und erleichtern somit das Bearbei-

ten des Übersetzungsvorschlages (siehe Abb.4): Die Wörter *drivers and* sind im aktuellen Segment hinzugekommen und werden in „Ausgang neu“ farblich markiert. Die Zahlenunterschiede zwischen „Ausgang alt“ und „Ausgang neu“ sind ebenfalls farblich hervorgehoben und wurden automatisch im Übersetzungsvorschlag angepaßt.

| | |
|------------------|---|
| Ähnlichkeit: 84% | Entnommen aus: D:\transit-30\PROJECTS\tutorial\Reftutorial-old |
| Ausgang alt | CD-ROM drive (to install the software under Windows 95 or Windows 98) |
| Ausgang neu | CD-ROM d rive (to install the drivers and software under Windows 98 or Windows 2000) |
| Ziel neu | CD-ROM Laufwerk (zur Installation der Software unter Windows >98< oder Windows >2000<) |

Abb. 4: Übersetzungsvorschlag im Fuzzy-Index-Fenster.

Der beste Übersetzungsvorschlag, d.h. derjenige der den höchsten Ähnlichkeitsgrad aufweist, wird zuerst angezeigt. Weitere Vorschläge können per Tastendruck angefordert werden. Das Referenzsprachpaar kann jederzeit (z.B. zum Korrigieren der Referenzübersetzung oder zum Einsehen des Kontextes des Referenzsegments) jederzeit im Transit-Editor geöffnet werden. Sobald der Übersetzungsvorschlag angepaßt ist, wird er per Tastendruck in den Zieltext übernommen.

3.2 Interaktion zwischen Transit und TermStar

Das integrierte Terminologieverwaltungssystem TermStar dient der effizienten Nutzung von Fachterminologie beim Übersetzen mit Transit. Für das aktuelle Sprachpaar führt Transit automatisch in allen projektspezifischen Wörterbüchern eine Hintergrundsuche durch und hebt die ausgangssprachlichen Wörter, für die ein Eintrag gefunden wurde, im Text farblich hervor. Die gefundenen Fachbegriffe werden im Wörterbuchfenster des Transit-Editors zu einem einzigen, großen, virtuellen Wörterbuch zusammengefasst. Zu jedem Segment werden im Wörterbuchfenster die Fachbegriffe und die zugehörige(n) Übersetzung(en) angezeigt, wobei zu jedem Eintrag angegeben ist, aus welchem Wörterbuch er stammt (vgl. Abb.2). Die Übersetzung eines Fachbegriffs kann per Tastendruck direkt in den Zieltext übernommen werden. Durch das interaktive Zusammenspiel zwischen TermStar und Transit kann automatisch überprüft werden, ob ein Fachbegriff tatsächlich mit einer in den Projektwörterbüchern vorhandenen Übersetzung übersetzt wurde. Im Transit-Editor können neue Fachbegriffe über die Schnelleingabe-Maske oder per Mausklick Terminologie direkt erfaßt werden, so dass sie für die weitere Übersetzung sofort zur Verfügung stehen.

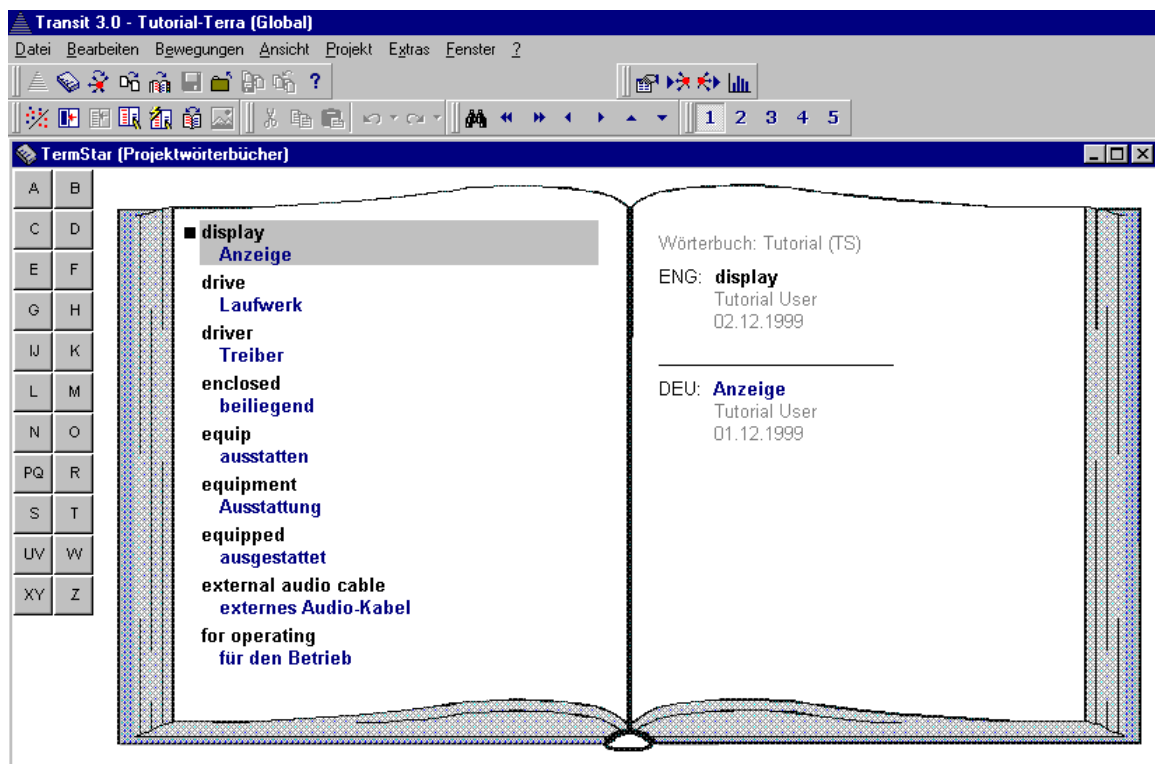


Abb. 5: TermStar Arbeitsplatz. Das Layout der Wörterbücher ist individuell konfigurierbar.

TermStar selbst ist ein professionelles Datenbanksystem für die multilinguale Terminologearbeit, das in einem benutzerfreundlichen Buchdesign (vgl. Abb. 5) eine Vielzahl leistungsfähiger Funktionen zur Terminologiepflege und zum Terminologieabruf bietet:⁶

- **Komplexes Annotationsschema:** Jeder Fachbegriff kann mit einer Vielzahl unterschiedlicher Informationen (Querverweise, Definition, Wortart, Eintragsdatum, Benutzerkennung, graphische Illustrationen) versehen (annotiert) sein, wobei Auswahllisten und Vorgabewerte die Konsistenz terminologischer Information sowie die sachgebiets- oder kundenorientierte Klassifikation der Einträge ermöglichen.
- **Effiziente Abrufmechanismen:** Neben Navigiermechanismen (z.B. über die Buchstabenleiste) und benutzerfreundlicher Hyperlink-Technologie, über die Querverweise zu verwandten Fachbegriffen direkt angesteuert werden können, bietet TermStar effiziente Suchfunktionen (u.a. über reguläre Ausdrücke) zum gezielten Abrufen spezieller Datensät-

ze. Außerdem ist es möglich, über benutzerdefinierbare Filter nur bestimmte Datensätze aller zu einem Projekt geladenen Wörterbücher anzeigen zu lassen.

- **Import-/Export-Schnittstellen:** Existierende Terminologielisten können in (neue oder bereits bestehende) TermStar-Wörterbücher integriert werden, wobei eine detaillierte Kontrollfunktion zur Vermeidung von Doppelinträgen bereitgestellt wird. Über benutzerspezifisch konfigurierbare Exportmechanismen können (vollständige) Wörterbücher aus TermStar exportiert werden oder Auszüge eines Wörterbuchs, die zuvor über die oben genannten Filter definiert wurden, in ein neues TermStar-Wörterbuch umgewandelt werden.

4 Kennzeichen von Transit

Transit zeichnet sich durch folgende Kriterien besonders aus:

Unterstützte Sprachen: Neben west- und osteuropäischen Sprachen⁷ werden auch asiatische Sprachen, wie Japanisch, Chinesisch, Thailändisch oder Koreanisch, unterstützt. Die Version 3.0 wird zusätzlich Arabisch (und später auch Hebräisch) unterstützen. Transit 3.0 basiert auf Unicode und erleichtert damit die Handhabung unterschiedlicher Zeichensätze.

Unterstützte Dateiformate: Vollintegrierte Import- und Exportmechanismen erlauben die reibungslose Bearbeitung einer Vielzahl gängiger Dateiformate, ohne dass zusätzliche Konvertierungstools benötigt werden. Eingebettete Objekte eines unterstützten Dateiformats (z.B. MS-Powerpoint-Graphik in einem Word-Dokument) können in Transit bearbeitet werden.

Performanz der TM-Komponente: Im Gegensatz zu Datenbank-basierten TM-Systemen treten Performanzprobleme beim Zugriff auf das Datei-basierte Translation-Memory von Transit wesentlich weniger auf. Die Retrieval-Geschwindigkeit, die u.a. von Größe und Verwaltung des TM abhängt, kann in Transit über die benutzerspezifische Zusammenstellung des Referenzmaterials direkt positiv beeinflusst werden.

Transparenz der TM-Komponente: Das Prinzip des Dateiverwaltungssystems erlaubt es, nur genau diejenigen Übersetzungen als Referenzbasis auszuwählen, die für das aktuelle Übersetzungsprojekt relevant sind und den ausgewählten Referenzdateien zudem eine bestimmte Priorität zuzuweisen. Da das Referenzmaterial im Translation-Memory in Dateiformat vorliegt, kann ein Referenz-Sprachpaar im Transit-Editor jederzeit bearbeitet werden, und der Kontext einer Übersetzungseinheit ist direkt einsehbar.

Effektivität der TM-Komponente: Aufgrund der XML-basierten Kodierung der Formatierungsinformation in Transit-Sprachpaaren werden die Unterschiede zwischen Segmenten, die aus unterschiedlichen Dateiformaten in Transit importiert werden, reduziert. Dadurch kann das Referenzmaterial im Translation Memory noch effizienter zwischen Übersetzungsprojekten genutzt werden, deren Ausgangssprachliche Texte in verschiedenen Dateiformaten erstellt wurden. Ab Version 3.0 erlaubt Transit den Austausch von Referenzmaterial über TMX (Translation Memory Exchange Format).

Interaktives Alignment-Tool: Existierende Übersetzungen in Dateiformaten, die Transit unterstützt, können über das Alignment-Tool zu Transit-Sprachpaaren aufbereitet und in das Translation-Memory integriert werden. Der Alignment-Prozess läuft weitgehend automatisch ab und verlangt nur dann ein Eingreifen, wenn Strukturunterschiede zwischen Ausgangstext und Übersetzung vorliegen.

5 Schlussbemerkung

Grundlage einer qualitativ hochwertigen TM-Komponente sind die Qualität der darin verfügbaren, manuell erstellten Übersetzungen und die Qualität des Retrieval-Algorithmus zum Auffinden der besten Referenzübersetzungen. Da ein Translation-Memory-System (im Gegensatz zu einem automatischen Übersetzungssystem) nicht selbständig übersetzt, sondern nur auf vorliegende Übersetzungen zurückgreift, ist die Korrektheit der Übersetzungseinheiten, die entweder im Rahmen der Vorübersetzung automatisch eingefügt wurden oder im Transit-Editor angeboten werden, gewährleistet, sofern die zuvor manuell angefertigte Übersetzung fehlerfrei ist.

Transit soll und kann qualifizierte Übersetzerinnen und Übersetzer nicht ersetzen, sondern dient ihnen als professionelle Arbeitshilfe. Die Gesamtheit der Transit-Funktionalitäten bildet eine vollständige, integrierte Übersetzungsumgebung, die bei gleichzeitiger Verbesserung der Übersetzungsqualität umfassende Möglichkeiten zur Produktivitätssteigerung bietet.

Danksagung

Mein Dank gilt meinen Kolleginnen und Kollegen von der Transit Abteilung der STAR Deutschland GmbH, insbesondere Christiane Gläser, Bettina Steudel und Oliver Rau, die mit ihrem umfassenden Fachwissen viele wesentliche Punkte zu diesem Artikel beigetragen haben.

Literatur

- [Andrés Lange98] Andrés Lange, C. (1998): 'Tying the Knot. How Baan wed machine translation to translation memory – and survived the honeymoon', in: *Language International*, 34-36.
- [Benis99] Bensi, M. (1999): 'Translation Memory. From Q to R', in: *Bulletin of the Institute of Translation and Interpretation*, 4/99, 4-19.
- [Steudel95] Steudel, B. (1995): *Management von Änderungsübersetzungen. Untersuchung zur Einführung eines Übersetzungsspeichers in einem industriellen Sprachendienst*. Diplomarbeit.

ANMERKUNGEN

- ¹ Die STAR Firmengruppe (<http://www.star-group.net>) mit Hauptsitz in der Schweiz wurde vor 15 Jahren gegründet und hat sich vor allem auf den Bereich technische Dokumentation spezialisiert. Das Transit System wird seit Ende der 80er Jahre von der STAR Deutschland GmbH entwickelt und ist seit 1990 bei STAR intern im Einsatz.
- ² Neben den Dateiformaten Winword, Wordperfect, Amipro, Windows Resource-Dateien, Windows Hilfe-Dateien, Interleaf, FrameMaker, PageMaker, SGML, HTML, Ventura Publisher, QuarkXPress, C-Quellcode oder andere Quellcodes, ASCII, und Ansi-basierten Dateiformaten wird Transit 3.0 zudem Powerpoint und Excel unterstützen. Nicht alle Filter sind Bestandteil des Transit-Standardpakets.
- ³ Die XML-konforme Kodierung wird ab Version 3.0 von Transit angeboten.

-
- ⁴ Emulationen sind Nachbildungen der Texteditoren, d.h. bestimmte Tastenkombinationen der Editoren (wie z.B. STRG + S zum Speichern) haben im Transit-Editor die selbe Funktion.
 - ⁵ Das Transit-Standardpaket umfasst zwei (von insgesamt 18) Rechtschreibprüfprogrammen. Für Deutsch kann nach der alten und nach der neuen deutschen Rechtschreibung geprüft werden.
 - ⁶ TermStar kann als integraler Bestandteil von Transit genutzt werden, ist aber auch als eigenes Softwarepaket erhältlich.
 - ⁷ Deutsch, Englisch, Französisch, Italienisch, Spanisch, Portugiesisch, Niederländisch, Schwedisch, Dänisch, Norwegisch, Finnisch, Griechisch, Türkisch, Indonesisch, Russisch, Tschechisch, Polnisch, Ungarisch, Rumänisch, Bulgarisch, Slowenisch, Ukrainisch.