

Der Einsatz von Translation-Memory-Systemen am Übersetzerarbeitsplatz

Aufbau, Funktionsweise und allgemeine Kaufkriterien

Uta Seewald-Heeg

Für viele Übersetzer sind Translation-Memory (TM)-Systeme selbstverständliche Arbeitswerkzeuge, andere probieren lediglich ein Programm aus oder haben keinerlei Erfahrungen mit dieser Technologie. In diesem Artikel gibt Prof. Dr. Uta Seewald-Heeg einen Überblick darüber, wie solche Systeme aufgebaut sind, welche Funktionen sie aufweisen und wie sie den Übersetzungsprozess verändern. Im Anschluss daran beschreibt die Autorin das Testszenarium und listet die untersuchten Systeme auf, die schließlich in einem Systemvergleich nebeneinander gestellt werden.

Steigerung der Effizienz

Der moderne Übersetzerarbeitsplatz hat sich in den vergangenen Jahren in weiten Bereichen in ein elektronisches Büro gewandelt. Computer und Internet gehören inzwischen zur Standardausrüstung eines Übersetzers. Durch die Nutzung des Internets und elektronischer Wörterbücher lässt sich die terminologische Recherche heute viel effizienter gestalten, als dies noch vor 10 Jahren möglich gewesen wäre.

Der Einsatz von Übersetzungstechnologie bietet weiteres Potential, Projekt- und Übersetzungsabläufe noch effizienter zu gestalten, und ermöglicht gerade in größeren Übersetzungsdienstleistungsunternehmen zum Teil deutlich kürze Bearbeitungszeiträume und dadurch auch erhebliche Kostenreduktion.

Mit Werkzeugen zur computerunterstützten Übersetzung (eng. *Computer-Aided Translation Tools*, kurz: CAT-Tools) stehen dem Übersetzer heute Hilfsmittel zur Verfügung, die ihn zum Teil bereits bei der Angebotserstellung unterstützen können. Die Suche nach Wörtern und Phrasen in bereits früher erstellten Übersetzungen bietet die Möglichkeit, den konsistenten Gebrauch von Terminologie sicherzustellen und einmal erstellte Übersetzungen mehrfach zu nutzen und damit unter Umständen den für eine Übersetzung erforderlichen Zeitaufwand erheblich zu reduzieren.

Kern der meisten heute auf dem Markt befindlichen CAT-Systeme ist die Translation-Memory-Technologie. *Translation Memories*, kurz TMs, können gleichsam als die elektronische Erweiterung des Gedächtnisses des Übersetzers betrachtet werden, der nicht mehr selbst mühevoll in abgelegten bzw. gespeicherten Dokumenten nach Textpassagen oder Formulierungen suchen muss, sondern diese vom System automatisch angezeigt bekommt. Einige Systeme umfassen mittlerweile eine Vielzahl von Funktionen, die neben der Unterstützung des eigentlichen Übersetzungsprozesses auch die Verwaltung ganzer Übersetzungsprojekte erlauben, weshalb hier verschiedentlich auch von *Translation-Management*-Systemen gesprochen wird.

Vorraussetzungen für einen sinnvollen Einsatz

Nicht alle systemseitig vorgesehenen Möglichkeiten von TM-Systemen – die zwar von Hersteller zu Hersteller variieren – lassen sich in jeder Übersetzungssituation mit dem gleichen Nutzen einsetzen. Um die Vorzüge eines TM maximal nutzen zu können, müssen einige Voraussetzungen gegeben sein. Hierzu gehört z.B., dass der zu übersetzende Text in elektronischer Form vorliegt, was bei guter Druckqualität und entsprechendem Umfang von in Papierform vorliegenden Texten auch durch Einscannen der Dokumente erreicht werden

kann. Doch unter Umständen ist das betreffende Dokument ohne den Einsatz eines CAT-Tools schneller übersetzt, als das Dokument in geeignete maschinenlesbare Form gebracht werden kann. Einzelne Textsorten, wie z.B. Werbetexte, sind für einen nutzbringenden Einsatz von TM-Technologie oft weniger geeignet, da zur Erzielung bestimmter Effekte in der Zielsprache gegebenenfalls ganz andere sprachliche Mittel oder Satzstrukturen erforderlich sind als in der Ausgangssprache.

Erleichterung der Übersetzungsarbeit

Im Bereich der Technischen Dokumentation und der Softwarelokalisierung gibt es heute aufgrund der Textvolumina und des in der Regel hohen Zeitdrucks keine Alternative zum Einsatz von TM-Systemen. Um als Unterauftragnehmer großer Übersetzungsdienste oder Lokalisierungsunternehmen arbeiten zu können, muss ein Übersetzer heute über mindestens *ein* TM-System verfügen. In verschiedenen Fällen ist – in Abhängigkeit von den von den Auftraggebern geforderten Formaten – sogar die Anschaffung mehrerer Systeme erforderlich. Auch in zahlreichen anderen Übersetzungssituationen können TM-Systeme die Tätigkeit des Übersetzers sehr erleichtern. Dabei sei daran erinnert, dass TMs dem Übersetzer das Übersetzen nicht abnehmen. TM-Systeme übersetzen nicht selbst, sie verfügen lediglich über intelligente Such- und Speichermechanismen. Dies unterscheidet TM-Systeme wesentlich von maschinellen Übersetzungssystemen.

Aufbau des Artikels

Der vorliegende Beitrag beschreibt zunächst Eigenschaften und Funktionen von TM-Systemen im Allgemeinen, um im Anschluss daran einzelne derzeit auf dem Markt angebotene Systeme einer genaueren Betrachtung zu unterziehen, so dass der Leser erste Anhaltspunkte darüber erhält, welches der hier genannten Systeme seinen individuellen Anforderungen gerecht wird. Einsteigern in die TM-Technologie mag das Glossar als Verständnishilfe bei der Lektüre dieses Beitrags sowie bei der Durchsicht von Produktbeschreibungen dienen. Die tabellarische Aufstellung von Systemeigenschaften am Ende dieses Artikels ist in erster Linie als Unterstützung für den eigenen Entscheidungsprozess gedacht. Die Beschreibung der nachfolgend aufgeführten Produkte konzentriert sich auf Einzelplatzsysteme, die bei freiberuflich tätigen Übersetzern oder in kleineren Agenturen zum Einsatz kommen. Die Verfügbarkeit von Netzwerkversionen und damit unter Umständen verbundenen zusätzlichen Funktionen wird jedoch auch berücksichtigt und ist im tabellarischen Überblick ebenfalls aufgeführt.

Veränderungen seit Herbst 2002

In Heft 4-5/2002 des MDÜ gab es bereits einen sehr breit angelegten Überblick über damals auf dem Markt verfügbare TM-Systeme, der von Karl-Heinz Freigang und Uwe Reinke zusammengestellt wurde. Der Markt der CAT-Tools ist ebenso wie jener der Standard-Büroapplikationen und der Computer-Hardware in ständiger Bewegung und hat sich in den vergangenen drei Jahren deutlich verändert. Während von den Produkten der Hersteller Trados, Star, SDL, Atril und Champollion (Wordfast), die bereits 2002 in der Übersicht enthalten waren, inzwischen neue Versionen erhältlich sind, die zusätzliche Funktionen bieten und weitere Dateiformate unterstützen, werden andere Produkte (TRANS Suite von Cypressoft oder das an der Universität Athen entwickelte System Tr·Aid) nicht mehr angeboten. Schließlich haben sich auch neue Systeme auf dem Markt etabliert, wie beispielsweise across von Nero oder MetaTaxis, das in der Zwischenzeit auf dem Markt an Sichtbarkeit gewonnen hat. Waren zum damaligen Zeitpunkt für fast alle TM-Produkte Hardware-Dongles erforderlich, was vor allem beim Einsatz mehrerer Systeme recht unpraktisch war, so werden die heutigen Systeme in der Regel über einen Software-Schlüssel freigeschaltet.

Schon jetzt ist sicher, dass sich der TM-Markt weiter entwickeln und verändern wird. Erst im Juni wurde die Übernahme von Trados durch SDL bekannt. Beide Hersteller haben im Sommer 2005 neue Versionen ihrer Produkte auf den Markt gebracht, für die es auch in den kommenden Jahren Supportleistung geben soll („[...] we are offering a guarantee of support for the latest products for 5 years“, so Terry Lawlor von SDL in den transline tecNews vom 24.6.2005). Doch mittel- oder längerfristig ist damit zu rechnen, dass die TM-Produkte von Trados und SDL zu einer Produktlinie verschmelzen werden. „TRADOS und SDLX werden erst mal als eigenständige Tools weitergeführt. Es wird ein SDLX 2006 und ein TRADOS 8 Release geben. Mittelfristig soll die Code-Basis zusammengeführt werden, um so das Beste beider Produkte in einer Anwendung anbieten zu können“, so Jochen Hummel, vormals Trados, jetzt SDL, in transline tecNews vom 24.6.2005.

Aufbau von Translation-Memory-Systemen

TM-Systeme haben sich in den vergangenen Jahren zum Teil zu äußerst komplexen Systemen mit einer Vielfalt an Funktionen entwickelt. Um ihre grundlegenden Funktionsweisen darzustellen, seien zunächst die wesentlichen Programmmodule eines TM-Systems aufgeführt. Im Kern setzt sich ein solches System aus den folgenden Komponenten zusammen (vgl. auch Abb. 1):

- (1) Einem Übersetzungsspeicher, dem Translation Memory selbst, das dem beschriebenen Systemtyp ursprünglich seinen Namen gab. Der Übersetzungsspeicher bzw. das TM ist ein multilinguales Text- oder Satzarchiv, in dem quellsprachige und übersetzte Textsegmente einander zugeordnet abgelegt werden. Der Inhalt eines TM dient beim Übersetzen als Referenzmaterial, das dem Übersetzer bei ähnlichen Texten zur Wiederverwendung angeboten wird.
- (2) Einer Terminologiekomponente, in der Termini zwei- oder mehrsprachig aufgenommen werden können, wobei in den meisten Fällen auch die Angabe von Definitionen, Kontextbeispielen, grammatischen, projektbezogenen und administrativen Informationen möglich ist.
- (3) Einem Editor, in dem der zu übersetzende Text bearbeitet wird. Einige Systeme arbeiten mit eigenen Editoren, in denen die Texte unabhängig vom Dateiformat der zu übersetzenden Dokumente übersetzt werden. Andere Hersteller nutzen für bestimmte Dateiformate vorhandene Textverarbeitungsprogramme, in der Regel Microsoft Word.
- (4) Einer Filterkomponente, die es erlaubt, Dokumente, TMs und Terminologie unterschiedlicher Formate in das System zu importieren oder in für andere Arbeitsumgebungen geeignete Formate zu exportieren.

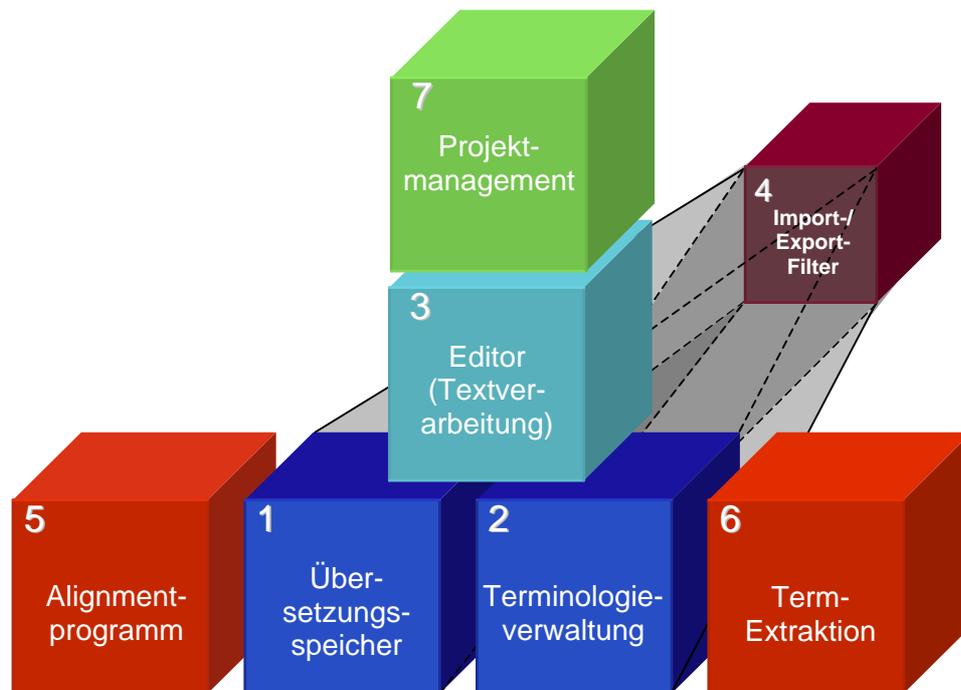


Abb. 1: Programmmodule eines Translation-Memory-Systems.

Über diese Kernkomponenten hinaus können TM-Systeme über weitere Module verfügen:

- (5) Ein Alignment-Programm, mit dem vorhandene quell- und zielsprachige Textbestände in ein TM überführt werden. Dabei werden die betreffenden Texte in einzelne Sätze segmentiert, die dann automatisch zu Paaren ausgangs- und zielsprachiger Segmente zusammengeführt werden. Auf diese Weise können auch Texte, die ohne den Einsatz von TM-Technologie übersetzt wurden, schon zu Beginn des Einsatzes eines TM-Systems als Referenzmaterial genutzt werden.
- (6) Ein Term-Extraktions-Programm, mit dem im Vorfeld einer Übersetzung automatisch Termkandidaten aus einem Dokument extrahiert werden. Für die als Termini bestätigten Kandidaten müssen schließlich manuell fremdsprachige Äquivalente recherchiert werden, auf die beim Übersetzen dann zugegriffen wird. Aus vorhandenen bilingualen Textbeständen lassen sich mit Hilfe solcher Programme auch im Nachgang einer Übersetzung ausgangssprachige Termini und ihre zielsprachigen Entsprechungen extrahieren, die dann, zum Teil zusammen mit automatisch extrahierten Kontextbeispielen, in eine vorhandene Terminologiekomponente übernommen werden können.
- (7) Ein Projektmanagement-Werkzeug, mit dem – in Abhängigkeit vom Hersteller – Benutzergruppen und -profile sowie Zugriffsrechte verwaltet und Funktionen zur Angebotskalkulation und zur Workflow-Überwachung bereitgestellt werden.

Funktionsweise

Full- und Fuzzy-Matches

Beim Übersetzen eines Textes mit einem TM wird der Text zunächst in einzelne Übersetzungseinheiten zerlegt bzw. segmentiert. Diese Segmentierung erfolgt auf der Basis eines Analyse-Algorithmus, der Satzendezeichen wie Punkt, Frage- und Ausrufezeichen oder Absatzmarken als Segmentgrenzen interpretiert. Jedes vom TM-System identifizierte Segment (Überschrift, Satz, Abbildungsunterschrift oder Listenelement etc.) bildet eine Übersetzungseinheit und wird zunächst mit dem Inhalt des Übersetzungsspeichers verglichen. Dieser Vergleich ist bei den derzeit auf dem Markt befindlichen Systemen ein reiner Zeichenkettenvergleich, der einen mathematisch errechneten Ähnlichkeitswert (Match-Wert),

in den meisten Fällen in Form eines Prozentwerts, als Ergebnis liefert. Stimmt das zu übersetzende Ausgangstextsegment vollständig mit einem im TM gespeicherten Segment überein, spricht man von einem so genannten 100%- oder *Full-Match* (Abb. 2). Übersetzungen von 100%-Matches können vom TM-System automatisch in das Zieldokument übernommen werden. Bestehen zwischen dem zu übersetzenden Ausgangstextsegment und einem im Referenzmaterial enthaltenen Segment Abweichungen, so dass die beiden Zeichenketten nicht vollständig zur Deckung kommen, liegt ein so genannter *Fuzzy-Match* vor (Abb. 3). Je größer die Abweichungen zwischen Ausgangstextsegment und einem Segment des Referenzmaterials sind, desto niedriger wird der errechnete Match-Wert. Übersetzungen von Segmenten, die nur geringfügige Abweichungen aufweisen, lassen sich vom Übersetzer in der Regel schnell anpassen, so dass die Anpassung der Übersetzung eines Fuzzy-Match-Segments in vielen Fällen mit einem deutlich geringeren Zeitaufwand möglich ist als die Neuübersetzung eines Segments, für das kein Übersetzungsvorschlag vorliegt. Da der Aufwand, eine Übersetzung an das aktuell zu übersetzende Segment anzupassen, bei einem geringen Grad an Übereinstimmung zwischen zwei Segmenten höher sein kann als die Neuübersetzung eines Segments, spezifiziert der Übersetzer einen Schwellenwert bzw. minimalen Match-Wert, unterhalb dessen als ähnlich berechnete Segmente nicht mehr als Übersetzungsvorschlag angezeigt werden, um nicht unnötig Zeit zur Durchsicht oder zur Änderung einer Übersetzung zu verbrauchen.

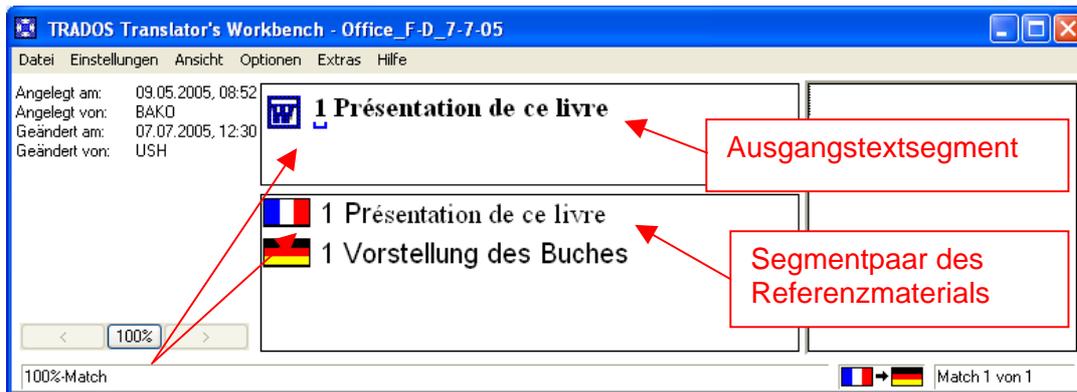


Abb. 2: Darstellung eines 100%-Match in der Translator's Workbench von **TRADOS**.

Die in den einzelnen Systemen implementierten Fuzzy-Match-Algorithmen unterscheiden sich voneinander, so dass die zwischen zwei Segmenten berechnete Ähnlichkeit sich je nach System in einem anderen Wert ausdrücken kann. Auch bei der Festlegung der minimalen Match-Wert-Grenze durch den Benutzer unterscheiden sich die Systeme. Während sich mit der Translator's Workbench von **TRADOS** Matches bis zu einem Wert von 30% anzeigen lassen, sind bei **across** nur Match-Werte ab 50% einstellbar.

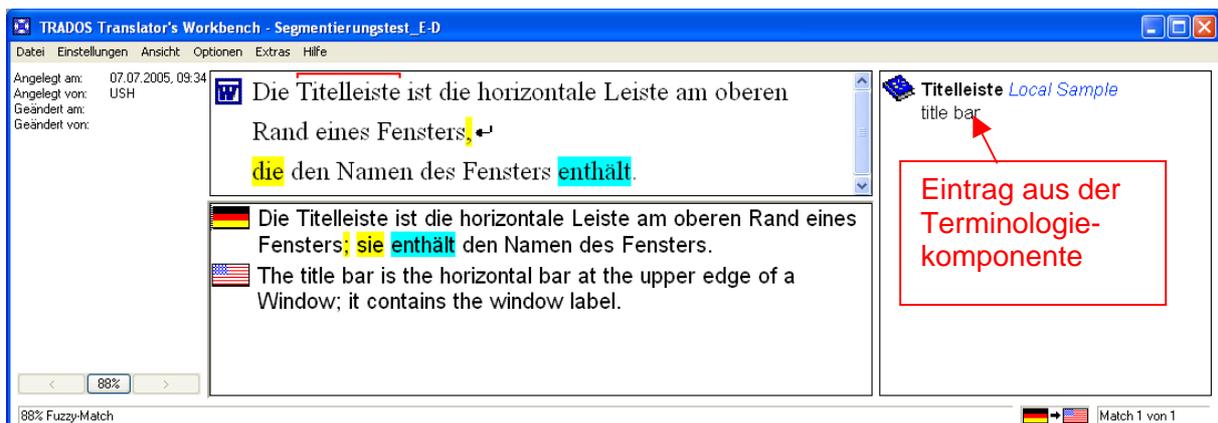


Abb. 3: Darstellung eines Fuzzy-Match in der Translator's Workbench von **TRADOS**.

Terminologearbeit und Wiederverwendung von Übersetzungseinheiten

Über den Abgleich zwischen Segmenten des Ausgangstextes mit dem Referenzmaterial hinaus werden die Wörter eines Segments in den meisten Fällen mit der in der Terminologiekomponente hinterlegten Terminologie verglichen. Im Falle von terminologischen Übereinstimmungen werden dem Übersetzer hinterlegte Termini in der Regel angezeigt (Abb. 3), so dass er diese ebenfalls per Mausklick oder Tastenkombination in das Zieltextsegment übernehmen kann, eine Funktionalität, die insbesondere bei der Vorgabe von Terminologie durch den Auftraggeber äußerst hilfreich ist und die terminologische Konsistenz von Texten unterstützt.

Beim Einsatz eines TM bearbeitet der Übersetzer ein Dokument segmentweise. Nach der Übersetzung eines Segments in die gewünschte Zielsprache werden Ausgangstext- und Zieltextsegment zusammen abgespeichert und im Übersetzungsspeicher als Referenzmaterial abgelegt, wo sie bereits beim Übersetzen des nachfolgenden Segments als Referenzmaterial zur Verfügung stehen. Insbesondere bei Texten mit erheblichen internen Redundanzen wirkt sich die Möglichkeit der unmittelbaren Wiederverwendung von Übersetzungseinheiten auf die Übersetzungsgeschwindigkeit beschleunigend aus. In den meisten TM-Systemen besteht heute die Möglichkeit, für ein Ausgangstextsegment mehrere zielsprachige Übersetzungen (hier wird zum Teil auch von mehreren 100%-Matches gesprochen) abzulegen. Dies kann erforderlich sein, wenn für verschiedene Kunden derselben Branche dasselbe TM benutzt wird, firmenspezifische Vorgaben jedoch im Einzelnen andere Übersetzungen erfordern oder wenn aufgrund textsortenspezifischer Konventionen andere Übersetzungen nahe gelegt werden.

Zugriff auf Kontextinformationen

TM-Systeme unterscheiden sich in der Art der Speicherung von einander zugeordneten Ausgangstext- und Zieltextsegmenten. Die meisten Systeme arbeiten Datenbank-basiert und speichern die Satzpaare im TM wie in einem gemeinsamen „Container“. Bei den meisten Systemen geht dabei die Information verloren, aus welchen Dokumenten die jeweiligen Segmente ursprünglich entstammen. Dies kann zu Fehlern im Zieltext führen, wenn mehrere 100%-Matches vorliegen oder das Zielsegment eines vorhandenen 100%-Match aufgrund eines anderen Textzusammenhangs oder eines anderen Sachgebiets im zu bearbeitenden Dokument keine zutreffende Übersetzung darstellt. In einigen Datenbank-basierten Systemen wie **across** werden daher auch Angaben des betreffenden Kunden und des jeweiligen Projekts als Informationen zusammen mit jedem Segmentpaar abgelegt (Abb. 4a) und im Übersetzungseditor angezeigt.

Die Alternative zu Datenbank-basierten Systemen sind Referenztext-basierte Systeme. Sie erzeugen TMs, indem sie jeweils auf den Ausgangs- und Zieltext der einzelnen Segmente zugreifen und die betreffende Quelle bei den einzelnen Satzpaaren als Information mit abspeichern (Abb. 4b). Auf diese Weise hat der Übersetzer die Möglichkeit, in Zweifelsfällen die Kontextinformation, d.h. den Text, in dem das betreffende Segment eingebettet war, als Information mit heranzuziehen.

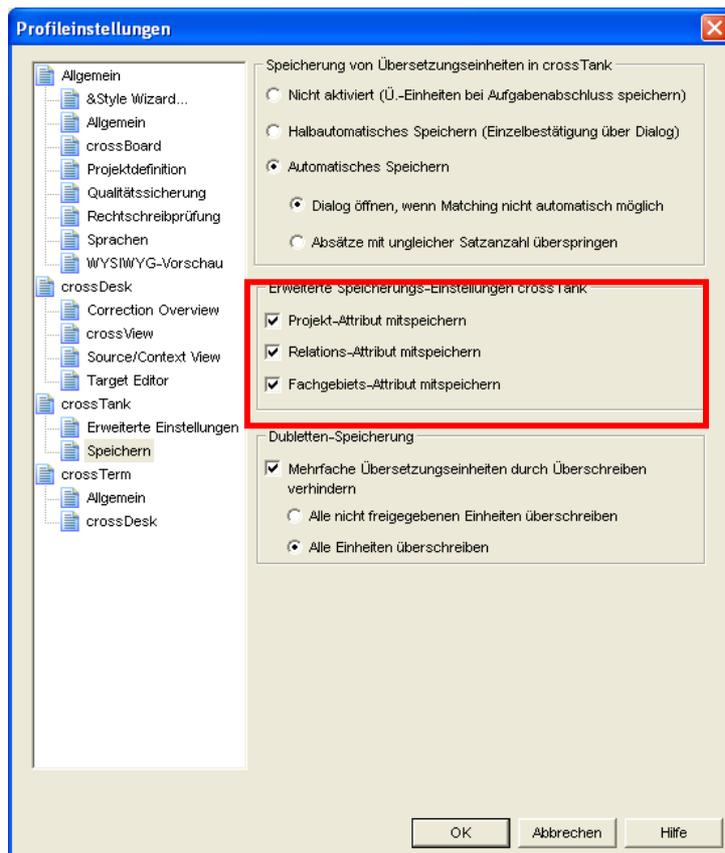


Abb. 4a: Kontextinformation zu Segmenpaaren in **across**.

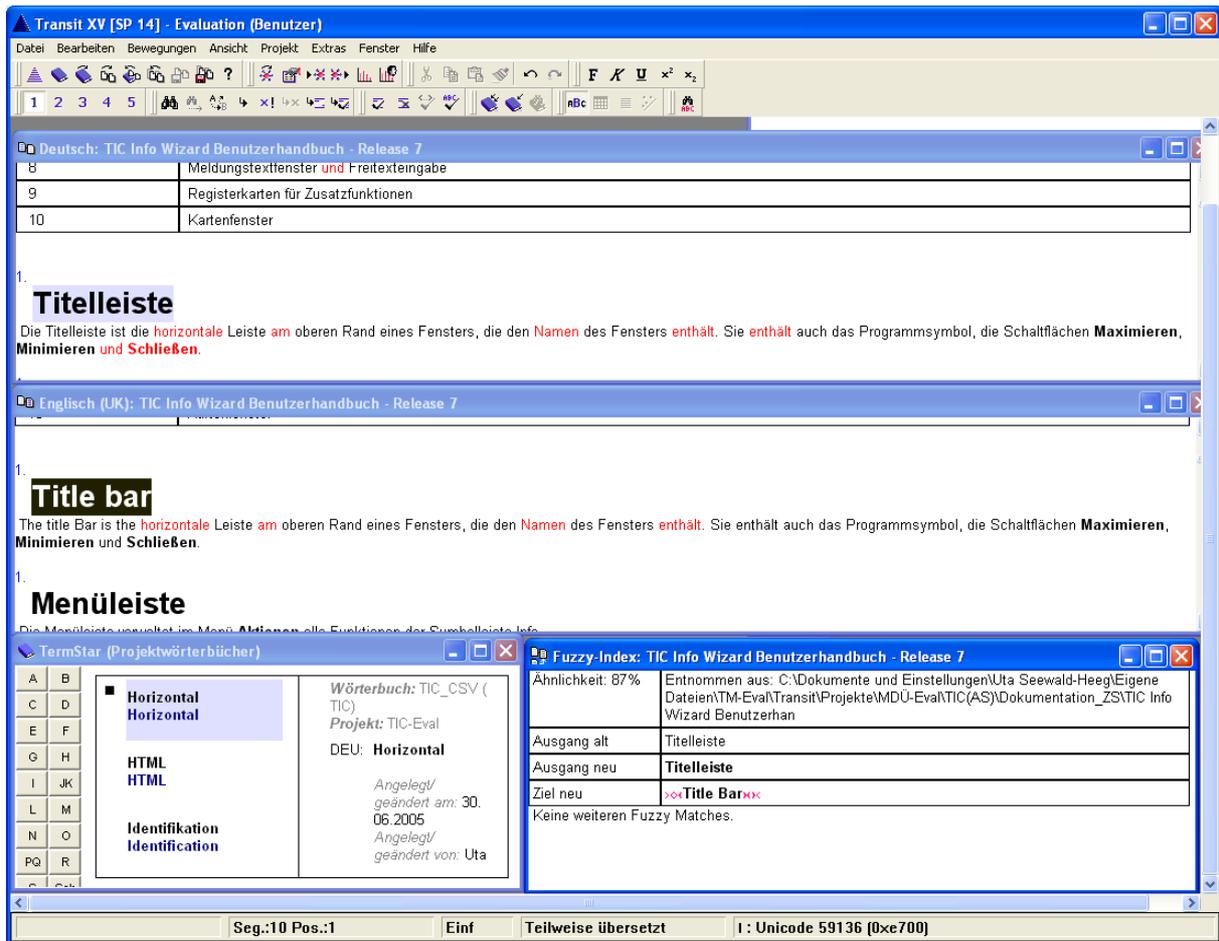


Abb. 4b: Kontextinformation des bei der Fuzzy-Match-Suche ermittelten Segments durch Angabe der Quelle, dem das betreffende Segment (hier: „Ausgang alt“) entstammt, im rechten unteren Teilfenster des TM **Transit XV von Star**.

Dass die aus dem Kontext herausgelöste Betrachtung einer Übersetzungseinheit und ihrer zielsprachigen Entsprechung linguistisch problematisch sein kann, verdeutlichen die Beispiele in Tab. 1, bei denen der Vergleich eines zu übersetzenden ausgangssprachigen Segments (AS-Segment „neu“) mit einem im Referenzmaterial enthaltenen Segment (AS-Segment „alt“) einen 100%-Match liefert. Die im Referenzmaterial abgelegten Übersetzungen der betreffenden Segmente (ZS-Segment „alt“) führen im vorliegenden Kontext jedoch zu Übersetzungsfehlern.

Ausgangstext		Referenzmaterial	
AS-Text (EN) „neu“	ZS-Text (DE) „neu“	AS-Text (EN) „alt“	ZS-Text (DE) „alt“
Visually inspect <u>the motor</u> without removing it from the housing. Remove and clean it only if excessive dust or other contaminants are visible.	[...] Nehmen Sie ihn nur heraus und säubern Sie ihn nur, wenn zu viel Staub oder andere Verunreinigungen sichtbar sind.	Visually inspect <u>the air filter</u> without removing it from the housing. Remove and clean it only if excessive dust or other contaminants are visible	[...] Nehmen Sie es nur heraus und säubern Sie es nur, wenn zu viel Staub oder andere Verunreinigungen sichtbar sind.
AS-Text (DE) „neu“	ZS-Text (FR) „neu“	AS-Text (DE) „alt“	ZS-Text (FR) „alt“
Er hat den Schlüssel im Schloss gelassen. (Schloss = Gebäude)	Il a laissé la clé au château .	Er hat den Schlüssel im Schloss gelassen. (Schloss = Schließvorrichtung)	Il a laissé la clé dans la serrure .

Tab. 1: 100%-Matches zwischen zu übersetzendem Segment (AS-Text „neu“) und im Referenzmaterial gespeichertem Segment (AS-Text „alt“), die im vorliegenden Kontext zu Fehlern bei der Übersetzung führen.

Segmentierungsregeln

Wie eingangs bereits erläutert, wird ein zu übersetzendes Dokument anhand von Satzendezeichen und bei getaggten Dateiformaten (HTML etc.) zusätzlich auf der Grundlage bestimmter Tags in Übersetzungseinheiten segmentiert. Die Segmentierungsalgorithmen der angebotenen Systeme unterscheiden sich hierbei allerdings, so dass ein gegebenes Dokument nicht in jeder TM-Umgebung in dieselben Segmente aufgespalten wird (vgl. Tab. 2). Wird das Semikolon beispielsweise von einem System nicht als Segmentende behandelt, von einem anderen aber doch, differieren Umfang und Zahl der für ein Dokument hinterlegten Segmente von TM zu TM. Die Interpretation von Satzendezeichen als Segmenttrennzeichen wirkt sich auch auf die Wiederverwendbarkeit einzelner im TM enthaltener Segmente aus. So kann ein TM-System unter Umständen für ein im Dokument enthaltenes Teilsegment eines größeren bereits im TM abgelegten Segments keinen Übersetzungsvorschlag unterbreiten, da aufgrund der unterschiedlichen Länge der Segmente kein Ähnlichkeitswert oberhalb der spezifizierten Fuzzy-Match-Grenze errechnet wird (vgl. Abb. 5a und Abb. 5b).

Zeichen	across	Déjà Vu	MetaTaxis	SDLX	Trados	Transit	Wordfast
Tabulator	kein Segmentende	Segmentende nach Nummerierungsfeld, sonst nicht	Kein Segmentende	kein Segmentende	Segmentende bei Standardeinstellung	kein Segmentende bei Standardeinstellung	Segmentende, sofern nicht nach Nummerierung
Semikolon	kein Segmentende	Segmentende	Segmentende bei Standardeinstellung	kein Segmentende	kein Segmentende bei Standardeinstellung (vgl. Tab. 3)	Segmentende bei Standardeinstellung	kein Segmentende
Doppelpunkt	kein Segmentende	Segmentende	Segmentende bei Standardeinstellung	kein Segmentende	Segmentende bei Standardeinstellung	Segmentende bei Standardeinstellung	Segmentende
Weiche Zeilenschaltung	kein Segmentende	Segmentende in Word; kein Segmentende in PPT	Kein Segmentende	Segmentende in Word; kein Segmentende in PPT	kein Segmentende bei Standardeinstellung	kein Segmentende	kein Segmentende

Tab. 2: Unterschiedliche Behandlung von Interpunktionszeichen und von Tags bei der Segmentierung von Dokumenten in Übersetzungseinheiten (Segmente).

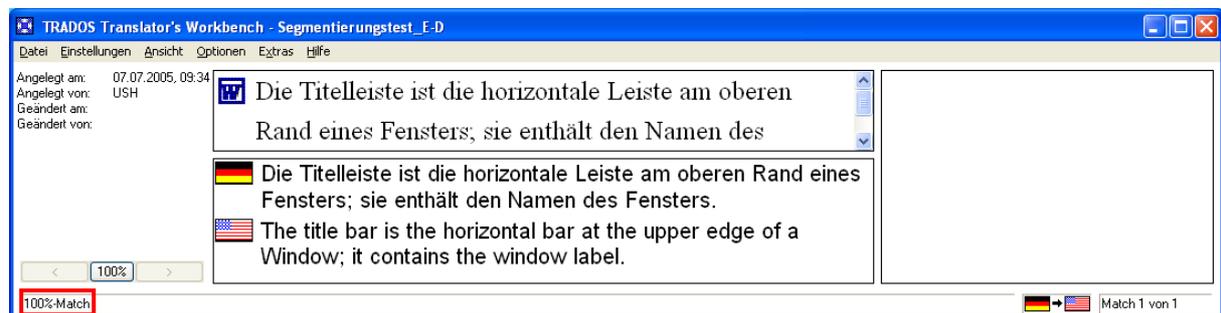


Abb. 5a: 100%-Match eines Segments mit Semikolon.

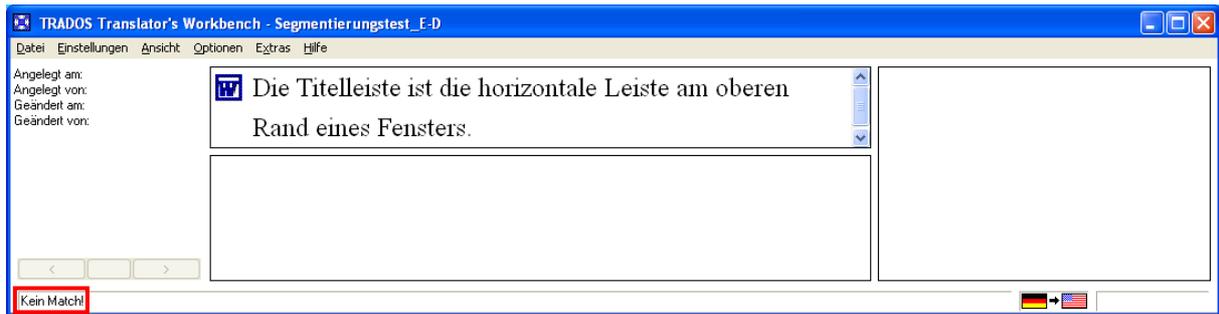


Abb. 5b: Bei einem minimalen Match-Wert von 60 % kein Match des Teilsegments, das Bestandteil des in Abb. 5a dargestellten und im TM enthaltenen Segments ist. (Der von TRADOS hier errechnete Match-Wert beträgt 54%. Wird nach dem zweiten Teil des in Abb. 5a dargestellten Segments „Sie enthält den Namen des Fensters.“ gesucht, wird das Segment auch bei der in TRADOS minimalen Fuzzy-Match-Grenze von 30% nicht mehr ermittelt.)

Die Segmentierungsregeln lassen sich bei einigen Systemen verändern und an bestimmte Eigenschaften eines gegebenen Textes anpassen (vgl. Abb. 6). Bei der Translator's Workbench von **TRADOS** z.B. können bei der Anpassung der Segmentierungsregeln verschiedene Zustände unterschieden werden, die in Tab. 3 am Beispiel der Segmentierungsregeln für das Semikolon dargestellt werden.

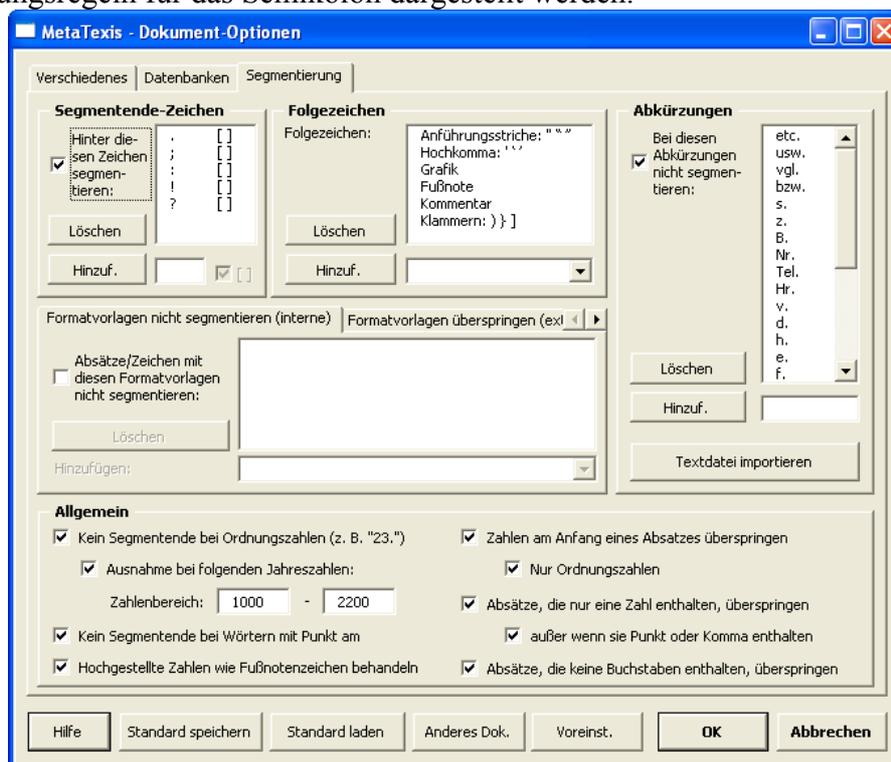


Abb. 6: Dialog zur Einstellung der Segmentierung bei **MetaTaxis**.

Semikolon	Standardeinstellung (kein Segmentende nach Semikolon)	Semikolon zur Liste der Segmentende-Zeichen hinzugefügt.		
		Regel ohne Einschränkung aktiviert (Kontrollkästchen für Kleinschreibung nach dem Semikolon <u>grün</u>)	Regel wird nur bei Großschreibung nach dem Semikolon aktiviert (Kontrollkästchen für Kleinschreibung nach dem Semikolon <u>weiß</u>)	Regel wird nur bei Kleinschreibung nach dem Semikolon aktiviert (Kontrollkästchen für Kleinschreibung nach dem Semikolon durch ein <u>Häkchen</u> aktiviert)
[...] eines Fensters; sie [...]	kein Segmentende	Segmentende	kein Segmentende	Segmentende
[...] eines Fensters; Sie [...]	kein Segmentende	Segmentende	Segmentende	kein Segmentende

Tab. 3: Parametrisierung der Segmentierungsregeln in der Translator's Workbench von **TRADOS**.

Suche nach Termini im Referenzmaterial

Erinnert sich der Übersetzer während der Übersetzung daran, einen in dem aktuell zu übersetzenden Segment enthaltenen Terminus bereits an anderer Stelle übersetzt zu haben, ohne dass das System einen Match aus dem TM vorschlägt oder die Terminologiekomponente für den in der Übersetzungseinheit enthaltenen Terminus einen Treffer ausweist, kann der Übersetzer die Funktion der Konkordanzsuche nutzen, die es erlaubt, nach dem betreffenden Wort im Referenzmaterial zu suchen (Abb. 7). Alle im TM abgelegten Segmente, die die gesuchte Zeichenkette enthalten, werden daraufhin vom System als Liste zusammengestellt. Der Übersetzer kann auf die einzelnen Segmente zugreifen und ein Segment oder Teile des Segments in die Zwischenablage kopieren, um sie sodann in das zu übersetzende Zieldokument einzusetzen.

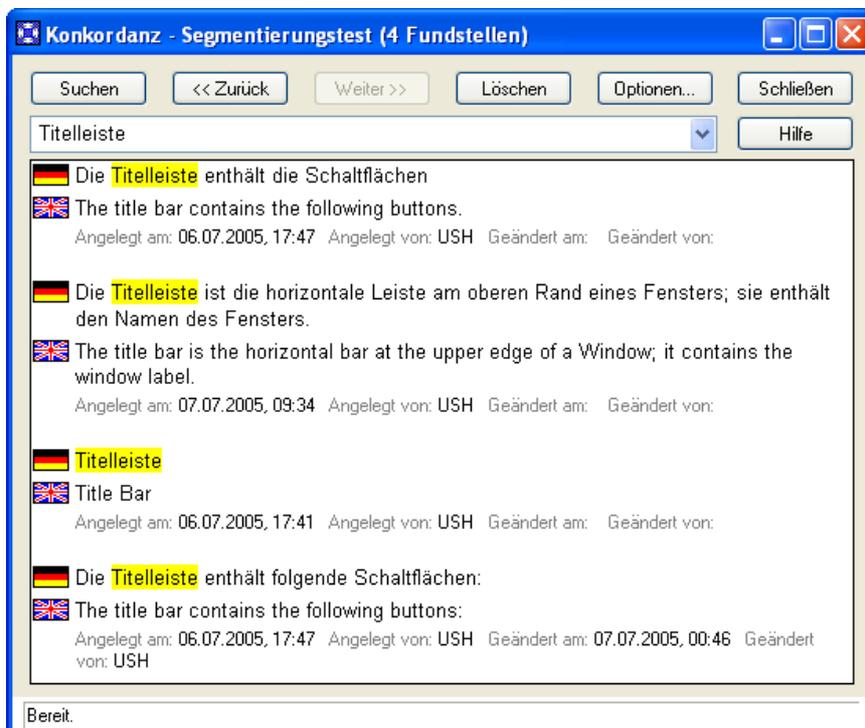


Abb. 7: Ergebnis der Konkordanzsuche in der Translator's Workbench von **TRADOS**. Das Suchwort wird in den Segmenten, in denen es auftritt, jeweils farbig markiert.

Aufbau von Satzarchiven aus Paralleltextbeständen

Verfügt ein Übersetzungsdienst oder ein Übersetzer über große Mengen Ausgangssprachiger Texte und deren Übersetzungen in elektronischer Form, können diese Texte dazu genutzt werden, ein Translation Memory aufzubauen. Dies kann vor allem dann von großem Nutzen sein, wenn sie Fachgebieten und Themen entstammen, die auch weiterhin Gegenstand von Übersetzungsaufträgen sind. Hierzu werden die entsprechenden Dokumentpaare in ein so genanntes Alignment-Programm geladen, das die vorliegenden Texte in Übersetzungseinheiten segmentiert und die ermittelten Segmente anschließend auf der Grundlage von Interpunktionszeichen und Formatangaben einander zuordnet. In Fällen, in denen Ausgangs- und ZIELTEXT in der Form voneinander abweichen oder ein Ausgangssprachiger Satz mit zwei Sätzen der Zielsprache wiedergegeben wird, kann es gelegentlich zu fehlerhaften Zuordnungen zwischen Segmenten des Ausgangs- und des ZIELTEXTES kommen, die vom Benutzer allerdings manuell bereinigt werden können. Steht für eine manuelle Kontrolle und Korrektur nicht ausreichend Zeit zur Verfügung, so dass die automatisch erstellte Zuordnung, das *Alignment*, unbearbeitet in das TM importiert werden muss, empfiehlt es sich, den berechneten Ähnlichkeitswert zwischen einem Ausgangssegment und einem durch Alignment gewonnenen Segment des Referenzmaterials mit einem Strafpunkt (*Penalty*) zu versehen. Der Benutzer kann in Abhängigkeit von seiner Einschätzung der Güte eines Alignments festlegen, wie hoch der Abzug vom errechneten Ähnlichkeitswert sein soll. Wird für ein automatisch aligniertes Segment ein Abzug von 5% bestimmt, wird ein 100%-Match beispielsweise als Fuzzy-Match von 95% angezeigt, über dessen unveränderte Übernahme in das Zieldokument der Übersetzer dann zu entscheiden hat, so dass diese Segmente nicht ungeprüft automatisch in den ZIELTEXT eingesetzt werden.

Abzüge können vom Benutzer eines TM je nach Hersteller auch für andere Segmenteigenschaften vorgenommen werden, so z.B. für unterschiedliche Formatierung oder für eine andere Regionalvariante einer Sprache (z.B. amerikanisches stattritisches Englisch).

Terminologieerfassung und -recherche

Im Idealfall erfolgt vor Beginn der Bearbeitung eines Übersetzungsauftrags eine Terminologierecherche. Unter Umständen liefert der Auftraggeber die Terminologie bereits. Häufig stößt man aber auch während des Übersetzens noch auf Termini, die hinsichtlich ihrer Bedeutung und des gebräuchlichen Zielsprachigen Äquivalents recherchiert werden müssen. Um solche Termini während des Übersetzens erfassen zu können, ohne den Übersetzungseditor zu verlassen, verfügen Systeme mit umfangreichen Terminologieverwaltungsfunktionen zusätzlich über so genannte Schnelleingabemodi, mit denen Ausgangs- und Zielsprachige Termini, gegebenenfalls mit einigen zusätzlichen terminusbezogenen Angaben, in die Terminologiekomponente aufgenommen werden können (Abb. 8a und 8b).

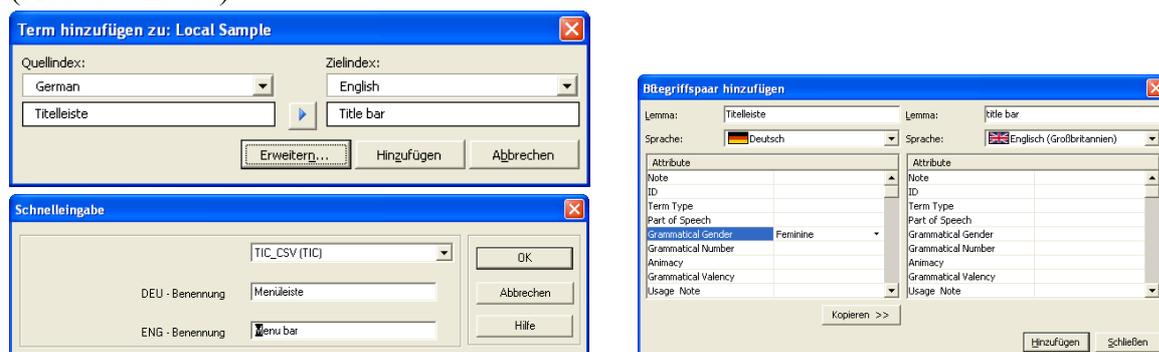


Abb. 8a: Schnelleingabemodus bei **TRADOS** und **Transit** von Star.

Abb. 8b: (Schnell-)Eingabemodus bei **Déjà Vu**

Zur Unterstützung des Aufbaus neuer Terminologie bieten einige Hersteller auch Programmmodule zur Terminologieextraktion an. Dabei können Termkandidaten auf der Grundlage eines einsprachigen Textes ermittelt werden. Für die so ermittelten Termini müssen anschließend manuell zielsprachige Äquivalente recherchiert werden. Darüber hinaus besteht in der Regel auch die Möglichkeit, aus alignierten Texten bilinguale Termpaare, gegebenenfalls mit Kontextbeispielen, automatisch ermitteln zu lassen. Dieses Verfahren kann als Unterstützung für den Aufbau neuer Terminologie herangezogen werden, verlangt aber immer noch erheblichen manuellen Aufwand bei der Bereinigung der Termkandidaten und der Erfassung der Terminologie.

Unterstützung für Projektmanagementaufgaben

Viele Übersetzungswerkzeuge bieten Unterstützung für Projektkalkulation und Managementaufgaben. Die von den meisten Systemen zur Verfügung gestellten Funktionen erlauben es, die zu übersetzenden Dokumente zu analysieren und mit dem vorhandenen Referenzmaterial und gegebenenfalls auch mit der Terminologie abzugleichen. Eine so durchgeführte Analyse listet neben der in einem Dokument enthaltenen Zahl der Wörter auch die Zahl der auftretenden Segmente nach verschiedenen Match-Typen auf (vgl. z.B. **across** in Abb. 9), so dass der Projektverantwortliche auf dieser Grundlage den Zeitaufwand sowie die Kosten für die Angebotserstellung ermitteln kann. Einige Systeme bieten hier sogar die Möglichkeit, konkrete Preise für die verschiedenen Match-Typen einzusetzen, so dass die Kalkulation bereits vom System übernommen wird und nicht in einer separaten Anwendung durchgeführt werden muss.

Ebenso wie bei der Errechnung von Match-Werten, wo je nach System für dasselbe Segment beim Vergleich mit dem Referenzmaterial verschiedene Match-Werte ermittelt werden können, variieren auch bei der Berechnung der Wortzahl eines Dokuments die Angaben zwischen den Systemen, wie aus Tab. 4 ersichtlich ist, so dass Kalkulationen auf Wortbasis hier durchaus zu unterschiedlichen Ergebnissen führen können.

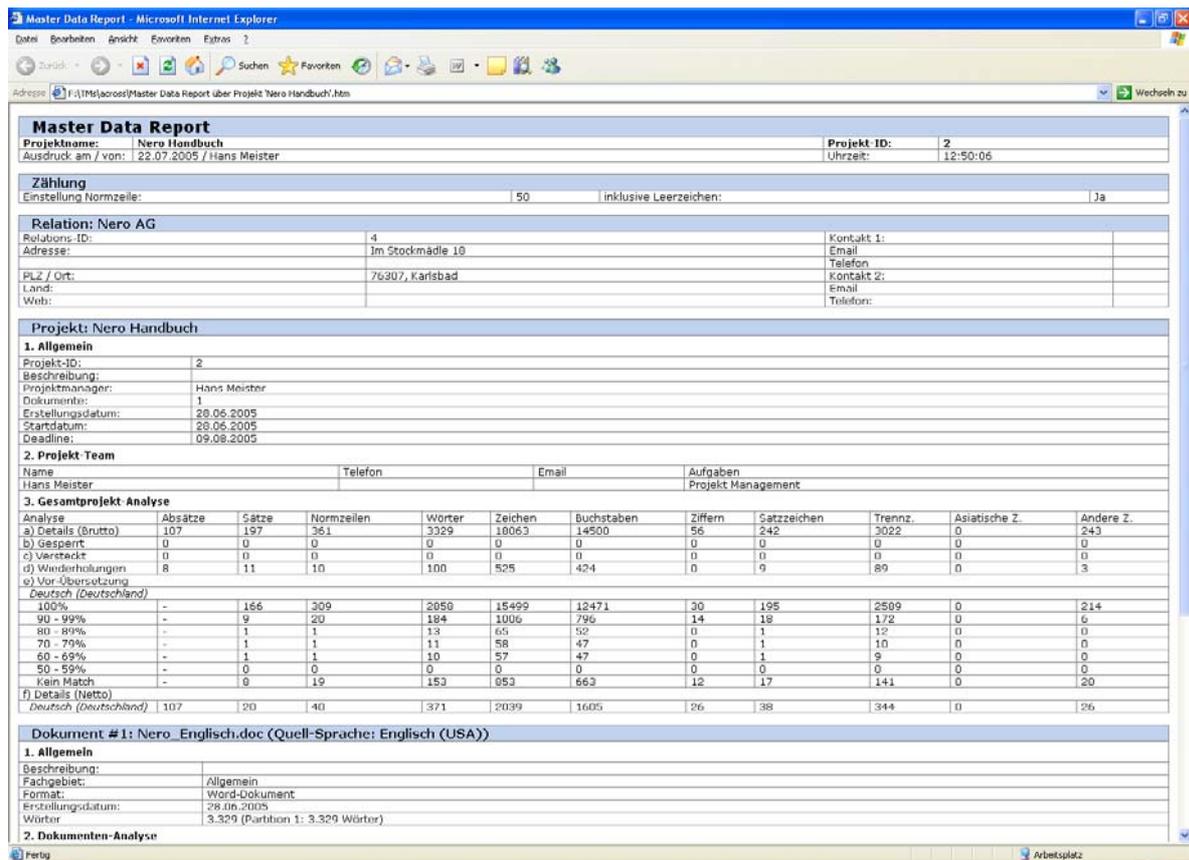


Abb. 9: Bericht einer Dokumentenanalyse mit der Team Edition von **across**.

	Word	across	Déjà Vu	MetaTaxis	SDLX	TRADOS	Transit	Wordfast
Anzahl Wörter	32	29	28	28	30	28	31	28

Tab. 4: Von verschiedenen Systemen ermittelte Wortzahl für ein Dokument mit folgenden Eigenschaften: 2 Ordnungszahlen als Kapitelzählung, davon 1 mit Feldfunktion, 2 Nummerierungen in einer Aufzählungsliste, 1 Kompositum mit Trennzeichenschreibung, 1 Akronym, 1 Abkürzung. Vom Textverarbeitungsprogramm MS Word werden alle hier aufgeführten Zeichen als Wörter gezählt.

Neben Funktionen zur Dokumentenanalyse bieten vor allem Netzwerk-Versionen von TM-Systemen darüber hinaus heute zum Teil sehr umfangreiche Workflow-Unterstützung, ermöglichen die Verwaltung von Benutzergruppen und -profilen und gestatten, wie etwa im Fall der Team Edition von **across**, von der selben Arbeitsumgebung aus Aufgaben an Team-Mitarbeiter zu übergeben, Projektberichte zu erstellen sowie den Fortschritt eines Übersetzungsprojekts zu überwachen.

Qualitätssicherung

Auch zur Qualitätskontrolle während und nach Abschluss der eigentlichen Übersetzung bieten die TM-Systeme eine Reihe von Programmmodulen an, die vor der Übergabe einer Übersetzung an den Auftraggeber in jedem Fall genutzt werden sollten. Zur Überprüfung der sprachlichen, d.h. in erster Linie orthografischen Korrektheit eines Textes werden Rechtschreibprogramme eingesetzt. Die Rechtschreibprüfung wird entweder durch TM-eigene Rechtschreibprogramme oder die Nutzung der Rechtschreibprüfung von Microsoft Word durchgeführt.

Zur Kontrolle der terminologischen Konsistenz dient bereits während des Übersetzens die Terminologieerkennung, durch die Termini und deren zielsprachige Äquivalente dem Übersetzer visuell dargeboten werden. Bei einigen Systemen kann auch im Anschluss an die Übersetzung überprüft werden, ob die vom Auftraggeber vorgeschriebene Terminologie – sofern sie als Terminologiedatenbank geliefert oder in die Terminologiekomponente importiert wurde – auch tatsächlich bei der Übersetzung berücksichtigt wurde. Da diese Prüfroutinen aber in der Regel mit den in der Terminologieliste als Lemmata enthaltenen Zeichenketten einen exakten Zeichenkettenvergleich vornehmen, steht der Nutzen bei Sprachen mit ausgeprägter Flexion häufig nicht im Verhältnis zu den Benutzerinteraktionen, die erforderlich sind, um Meldungen zu verwerfen, in denen verwendete Termini aufgrund hinzugefügter Flexionssuffixe vom System nicht erkannt werden.

Auch die Funktion zur Dokumentenanalyse kann zur Qualitätskontrolle eingesetzt werden. Mit ihr lässt sich die Vollständigkeit eines übersetzten Dokuments prüfen, die vorliegt, wenn nach Abschluss einer Übersetzung beim Vergleich des Referenzmaterials mit dem quellsprachigen Dokument nur noch 100%-Matches identifiziert werden.

Vor allem bei getaggtten Dateiformaten wie HTML oder XML, die explizite Auszeichnungselemente, so genannte Tags mit Informationen zum Layout oder zur Struktur eines Dokuments enthalten, besteht bei einigen Systemen die Möglichkeit, die Formatierung bzw. die in der Übersetzung vorhandenen Tags auf ihre Übereinstimmung mit den im Original enthaltenen Auszeichnungselementen hin zu überprüfen bzw. festzustellen, ob keines der Tags beim Übersetzen beschädigt wurde.

Erzeugen einer auslieferbaren Zieltextdatei

Bevor ein übersetzter Text an den Auftraggeber ausgeliefert werden kann, sind in der Regel einige abschließende Schritte notwendig, die aus der Übersetzung im TM-Editor einen dem Ausgangstext in Layout und Dateiformat analogen Zieltext erzeugen. Hierbei lassen sich im Wesentlichen zwei verschiedene Fälle unterscheiden.

Wird ein Editor zur Übersetzung verwendet, in dem der Zieltext in dasselbe Dokument, das den Ausgangstext enthält, eingefügt wird (z. B. Microsoft Word), wobei ein bilinguales Dokument entsteht, so muss dieses bilinguale Dokument nach Abschluss der Übersetzung vom Ausgangstext „gesäubert“ werden (*Clean-up*).

TM-Systeme, die für alle Dateiformate denselben Editor verwenden, trennen beim Import der zu übersetzenden Dokumente die Formatinformationen von dem zu übersetzenden Text. Ausgangs- und Zieltext werden bei Systemen dieser Art in der Regel in zwei verschiedenen Teilfenstern dargestellt. Nach Abschluss einer Übersetzung muss der übersetzte Text exportiert werden, wobei Layout und Formatinformationen wieder mit dem Text zusammengefügt werden.

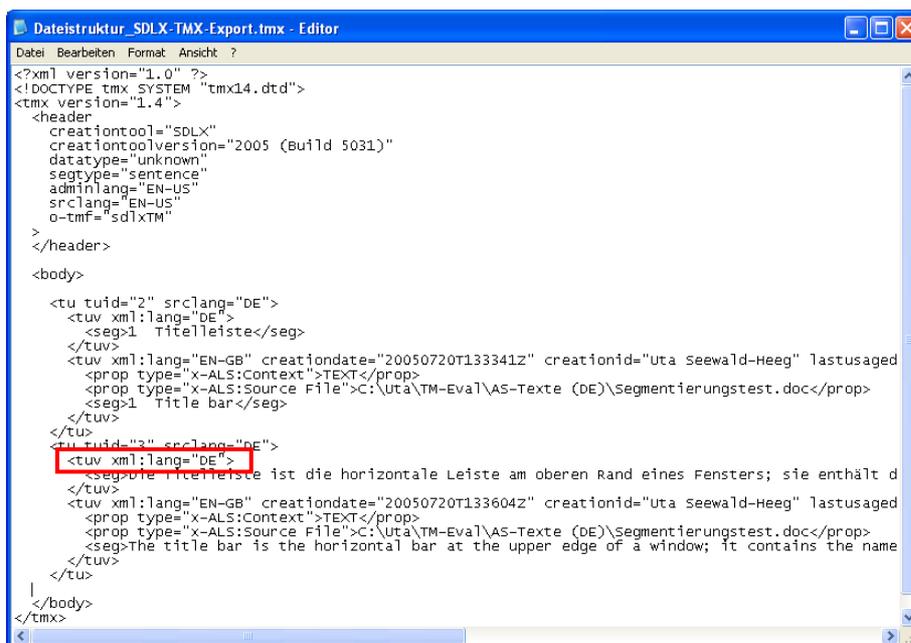
Datenaustausch

Der Austausch von TM-Daten zwischen verschiedenen Systemen und Anwendungen spielt heute eine immer entscheidendere Rolle, insbesondere in der Lokalisierungsbranche, wo der Datentransfer zwischen Lokalisierungswerkzeugen, TM- und Terminologieverwaltungssystemen inzwischen in den Lokalisierungs-Workflow integriert ist. Aber auch die Nutzung eines TM oder die Verwendung von Terminologie, die mit dem System eines anderen Herstellers angelegt wurden, kann für den Freiberufler von Bedeutung sein, wenn ein Kunde die Bearbeitung von Dokumenten mit einem bestimmten System vorschreibt. Umgekehrt stellen TM- und Terminologieressourcen auch für ein Unternehmen einen wertvollen

Wissensbestand dar, der seinen Wert allerdings erst zur eigentlichen Entfaltung bringt, wenn er sowohl unabhängig von einem einzelnen System als auch unabhängig von einem bestimmten Dienstleister einsetzbar ist, eine Bedingung, die nur durch standardisierte Austauschformate erreichbar ist.

In den vergangenen Jahren sind Standards entwickelt worden, die den Austausch von TMs und Terminologie ermöglichen. Federführend hierbei war und ist die Arbeitsgruppe OSCAR der *Localization Industry Standards Organization* (LISA). Sie entwickelte einen inzwischen von fast allen TM-Systemen unterstützten Standard zum Austausch von TM-Daten, das TMX-Format (*Translation Memory Exchange Format*). TMX basiert auf der Auszeichnungssprache XML und liegt seit Oktober 2004 in der Version 1.4b vor. Wenngleich alle Hersteller inzwischen TMX als Exportformat anbieten, so steht dieser in der Version 1.4b nur bei den TM-Produkten **SDLX 2005** und **TRADOS 7** und **Transit XV** zur Verfügung. **SDLX 2005** und **TRADOS 7** sind darüber hinaus beide LISA-zertifiziert, womit sie nachweisen, dass das von ihnen unterstützte TMX-Format vollständig der LISA-Spezifikation entspricht. Beide Systeme unterstützen darüber hinaus auch die TMX-Erweiterung (*Level 2*), die es erlaubt, Angaben zu Segmentierungsregeln (*Segmentation Rule Exchange*, SRX) des exportierenden Systems zu kodieren. **Déjà Vu** und **MetaTaxis** unterstützen ebenfalls die Version 1.4 von TMX. **Wordfast** und **across** unterstützen zurzeit TMX in der Version 1.1, das zusätzlich zu Version 1.4 von **TRADOS** und **MetaTaxis** und als Importformat auch von **Déjà Vu**, **SDLX** und **Transit** unterstützt wird.

Aus einem TM im TMX-Format exportierte Satzpaare werden wie folgt dargestellt: Der Beginn eines Segments mit seinen zielsprachigen Äquivalenten wird in TMX mit dem Tag `<tu>` (*translation unit*) gekennzeichnet. Dem XML-Standard entsprechend wird die betreffende Einheit mit `</tu>` abgeschlossen. Innerhalb dieser Klammer befinden sich Ausgangs- und zielsprachige Segmente, jeweils mit Angabe der Sprache des jeweiligen Segments, die seit Version 1.3 als Wert des Attributs `xml:lang` (`<tuv xml:lang>`), das in Version 1.1 lediglich die Bezeichnung `lang` trägt, angegeben wird. Ein konkretes Segment wird innerhalb der Tags `<seg>` und `</seg>` aufgeführt (vgl. Abb. 10).



```
<?xml version="1.0" ?>
<!DOCTYPE tmx SYSTEM "tmx14.dtd">
<tmx version="1.4">
  <header>
    creationtool="SDLX"
    creationtoolversion="2005 (Build 5031)"
    datatype="unknown"
    segtype="sentence"
    adminlang="EN-US"
    srclang="EN-US"
    o-tmf="sdlxTM"
  >
</header>
  <body>
    <tu tuid="2" srclang="DE">
      <tuv xml:lang="DE">
        <seg>1 Titelleiste</seg>
      </tuv>
      <tuv xml:lang="EN-GB" creationdate="20050720T133341Z" creationid="Uta Seewald-Heeg" lastused=
        <prop type="x-ALS:Context">TEXT</prop>
        <prop type="x-ALS:Source File">C:\Uta\TM-Eval\AS-Texte (DE)\Segmentierungstest.doc</prop>
        <seg>1 Title bar</seg>
      </tuv>
    </tu>
    <tu tuid="3" srclang="DE">
      <tuv xml:lang="DE">
        <seg>Die Titelleiste ist die horizontale Leiste am oberen Rand eines Fensters; sie enthält d
      </tuv>
      <tuv xml:lang="EN-GB" creationdate="20050720T133604Z" creationid="Uta Seewald-Heeg" lastused=
        <prop type="x-ALS:Context">TEXT</prop>
        <prop type="x-ALS:Source File">C:\Uta\TM-Eval\AS-Texte (DE)\Segmentierungstest.doc</prop>
        <seg>The title bar is the horizontal bar at the upper edge of a window; it contains the name
      </tuv>
    </tu>
  </body>
</tmx>
```

Abb. 10: Gerüst einer aus SDLX 2005 exportierten TMX-Datei in der Version 1.4.

Für den Austausch terminologischer Daten wurden weitere Standards entwickelt. Die Arbeitsgruppe OSCAR der LISA veröffentlichte im April 2002 das *TermBase eXchange*

Format (TBX), ein ebenfalls auf XML basierendes Austauschformat. Bislang wird es allerdings nur von wenigen Produkten, wie **across**, unterstützt. Verbreitet ist hingegen die Unterstützung des Im- und Exports von Termini aus mit Trennzeichen (Tabulator, Komma, Semikolon) voneinander getrennten Einträgen, die mit Excel oder Word erzeugt werden können, d.h. mit Programmen, die von zahlreichen Übersetzern für die Erfassung der eigenen Terminologie verwendet werden, jedoch weder effiziente Suchmöglichkeiten, noch flexible Beschreibungsmöglichkeiten für Termini bieten.

Testszenarium

Die Beschreibung der nachfolgend aufgeführten Systeme erfolgt anhand eines Beispielszenariums, das die wesentlichen Vorbereitungs- und Bearbeitungsschritte mit einem TM-System enthält und damit Aufschluss über verschiedene Nutzungsmöglichkeiten der einzelnen Systeme geben soll:

- Einrichten eines Übersetzungsprojekts
 - a. Unterstützte Dokumentenformate
 - b. Unterstützte Sprachen
- Bereitstellen von Terminologie
- Bereitstellen vorhandenen Referenzmaterials
- Aufwands- und Kostenkalkulation
- Übersetzen
 - a. Editor
 - b. Terminologieerkennung
 - c. Konkordanzsuche
- Zieldokument erzeugen
- Pflege der TM-Daten
- Einarbeitungsaufwand
- Kriterien für den Kauf eines TM-Systems

Die Systeme im Vergleich

Die hier untersuchten Systeme sind:

- **across** Personal Edition, Version 3.00.43, von der Firma Nero, das jüngste der auf dem TM-Markt angebotenen Systeme
- **Déjà Vu X Professional**, Version 7.0.273, von Atril
- **MetaTaxis**, Version 2.692, von Hermann Bruns, MetaTaxis Software and Services
- **SDLX 2005**, Build 5031, von SDL
- **Transit** und **TermStar XV Professional**, Version XV SP 14 Build 518, von Star
- **TRADOS 7 Freelance** (Translator's Workbench 7.0.0.615 und MultiTerm 7.0.0.315 von Trados)
- **Wordfast**, Version 5.0, von Champollion

Einrichten eines Übersetzungsprojekts

Beim Anlegen eines Übersetzungsprojekts bieten alle Systeme einen Benutzerdialog, anhand dessen die verschiedenen Schritte der Projekterstellung unterstützt werden.

Beim Start von **across** gelangt man ebenso wie beim Start von **SDLX** zunächst in eine Auswahloberfläche, die den Zugriff auf alle Komponenten des Systems ermöglicht und einen Project Wizard anbietet (Abb. 11a und 11b).



Abb. 11a: Startfenster von **across**.

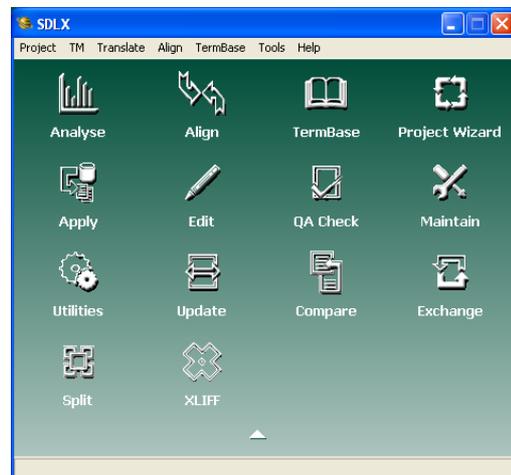


Abb. 11b: Startfenster von **SDLX**.

Transit startet mit dem Programmfenster, in dem der Projektauswahldialog angezeigt wird (Abb. 12a). **Déjà Vu** startet ebenfalls mit dem Programmfenster, von wo aus über das Dateimenü die Optionen zum Anlegen eines neuen Projekts aufgerufen werden können (Abb. 12b).

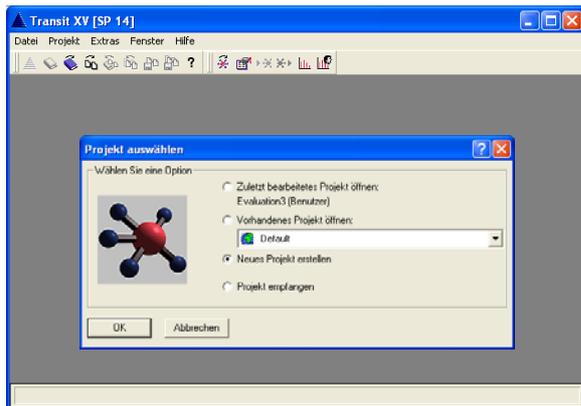


Abb. 12a: Startfenster von **Transit**.

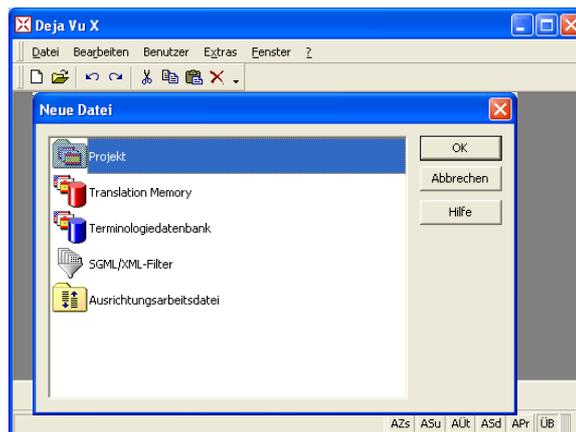


Abb. 12b: Startfenster von **Déjà Vu**.

Um ein Übersetzungsprojekt mit **TRADOS 7 Freelance** einzurichten, muss zunächst die TM-Komponente, Translator's Workbench, gestartet werden. Von hier aus kann über das Menü „Optionen“ sowohl ein neues Translation Memory als auch alle auf das TM-Projekt bezogenen Einstellungen, wie Ersetzungen, Match-Wert-Grenzen für die Suche im TM, für die Konkordanzsuche sowie die Suche nach Termini in der Terminologiedatenbank, Abzüge (*Penalties*) etc., vorgenommen werden (Abb. 13).

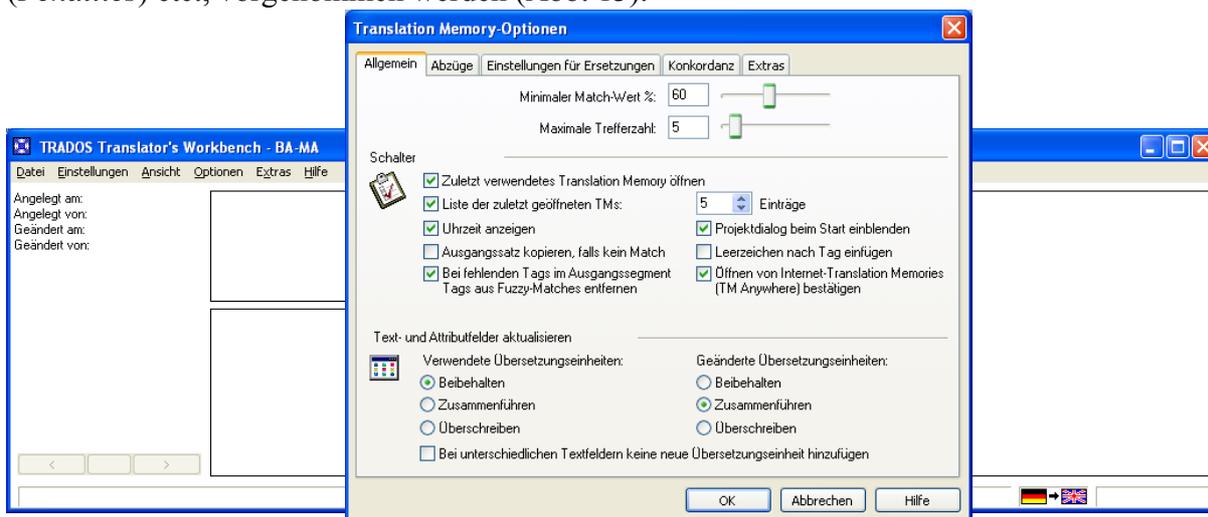


Abb. 13: **TRADOS** Translator's Workbench mit Dialog für Translation-Memory-Optionen.

Werden Übersetzungsprojekte mit **MetaTaxis** oder **Wordfast** angelegt, wird zunächst Word gestartet und das entsprechende Add-In geladen (Abb. 14a und 14b). Bei **Wordfast** werden die Projekteinstellungen, d.h. Erstellen oder Laden eines TM sowie gegebenenfalls vorhandener Terminologie (Abb. 15), über das **Wordfast**-Symbol (), vorgenommen. **MetaTaxis** verfügt über einen Start-Assistenten, der über die Unteroption „Start-Assistent starten...“ der Option „Datei“ des **MetaTaxis**-Menüs aufgerufen werden kann, aber auch nach Öffnen einer zu übersetzenden Datei in Word und Ausführen einer Navigationsfunktion (z.B. Öffnen eines Segments) automatisch gestartet wird. Mit Hilfe des Start-Assistenten wird man in sechs Schritten angeleitet, die erforderlichen Einstellungen eines Projekts vorzunehmen (vgl. Abb. 16).



Abb. 14a: **Wordfast**-Symbolleiste in Word und Symbol für die Zusatzkomponente +Tools.



Abb. 14b: **MetaTaxis**-Symbolleiste in Word.

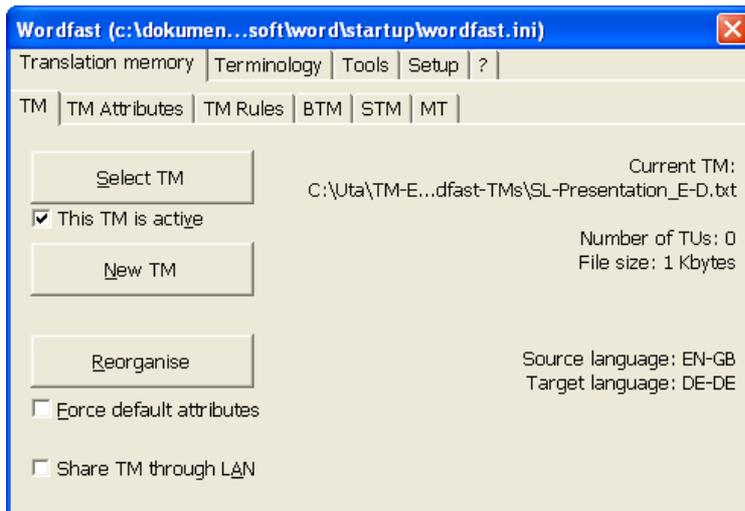


Abb. 15: Einrichten eines Übersetzungsprojekts in **Wordfast**.

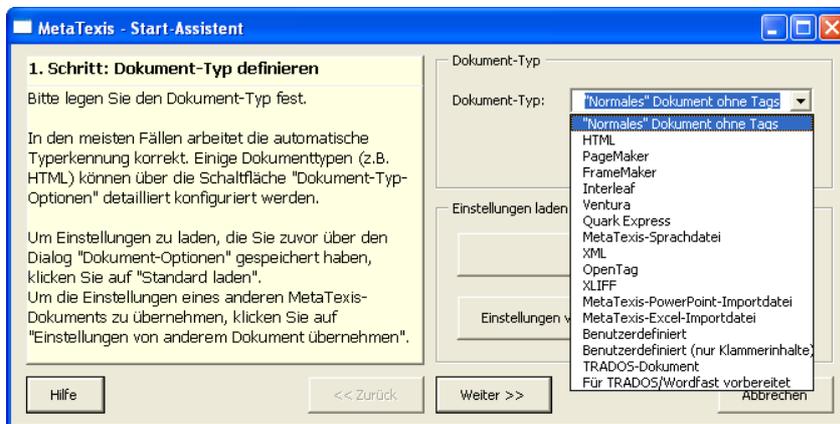


Abb. 16: Angabe des Dokumenttyps im Start-Assistenten von **MetaTaxis**.

Arbeit mit verschiedenen Dateiformaten

Bei den als Add-In von Word konzipierten TM-Systemen sind die unterstützten Dokumentenformate im Wesentlichen auf jene Dateiformate beschränkt, die in Word geöffnet werden können. Mit **MetaTaxis** lassen sich darüber hinaus über die Option „Office“ im **MetaTaxis**-Menü auch PowerPoint- und Excel-Dateien direkt importieren. Daneben sind alle getaggten Formate mit **MetaTaxis** bearbeitbar. Ist **Wordfast** um die so genannten +Tools (Abb. 14) ergänzt, ein zusätzliches kostenfrei erhältliches Add-In, das in Word installiert werden muss, lassen sich hiermit ebenfalls Excel-, PowerPoint-, Access-, HTML- und andere getaggte Dokumentenformate übersetzen. Zur Übersetzung von Excel, PowerPoint und Access muss allerdings jeweils die Anwendung der Ausgangsdatei mit dem betreffenden Dokument und gleichzeitig ein leeres Word-Dokument geöffnet werden.

Die übrigen Programme unterstützen eine Vielzahl von Formaten. Während **across**, **Déjà Vu**, **SDLX** und **Transit** unabhängig vom Dokumenttyp stets den in der Programmumgebung integrierten produkt-eigenen Übersetzungseditor verwenden, besteht bei der Nutzung von **TRADOS** die Möglichkeit, Word-Dokumente mittels eines Add-In in Word selbst zu bearbeiten. Während dies in früheren Versionen von **TRADOS** sogar die einzige Möglichkeit der Übersetzung von Word-Dokumenten war, besteht inzwischen die Möglichkeit, Word-Dokumente, ebenso wie HTML-, PowerPoint- und andere Dateiformate im TagEditor, dem Trados-eigenen Editor, zu übersetzen (Abb. 17), der nun auch mit AutoText-Funktionen sowie Such- und Schnelleingabefunktion für Termini in der Terminologiedatenbank

ausgestattet ist. Für Dateiformate, die von den Systemen nicht direkt unterstützt werden, bieten sowohl **across** mit der crossBox als auch **TRADOS** mit T-Window for Clipboard und **SDLX** mit SDL Clipboard die Möglichkeit, Text aus der Zwischenablage zu übersetzen.

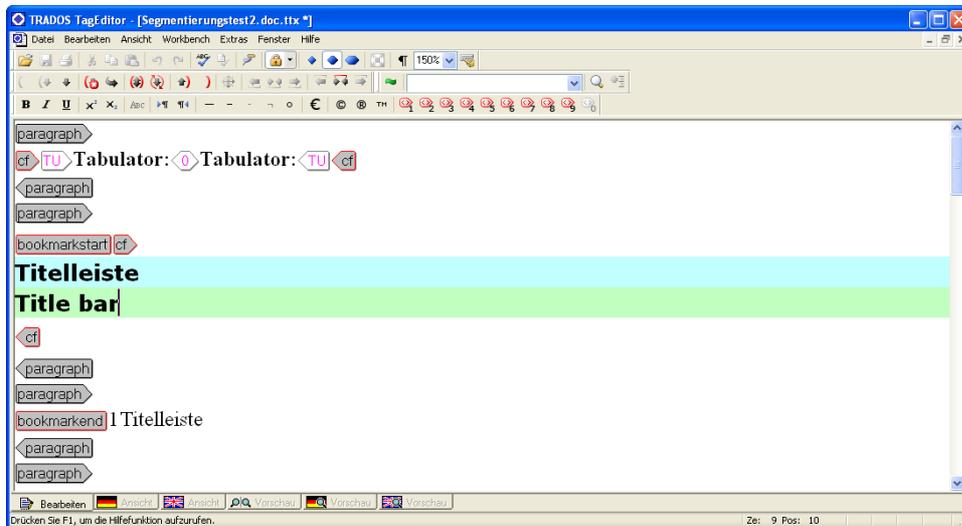


Abb. 17: Übersetzen im TagEditor von TRADOS.

Festlegung der Projekteinstellungen

Zur Konfigurierung eines Übersetzungsprojekts gehören neben der Angabe der Dokumente und der Sprachen (Abb. 18a), in die ein Dokument übersetzt werden soll, auch die Einstellungen der Match-Wert-Grenzen für den Abgleich der zu übersetzenden Segmente mit dem Referenzmaterial (Abb. 18b) und die Suche nach einzelnen Wörtern oder Wortgruppen im TM (in den meisten Fällen als Konkordanzsuche bezeichnet). Auch die Vergabe von Abzügen für bestimmte Attribute (Format, Kontext etc.) des Referenzmaterials, Einstellungen für automatische Ersetzungen, wie z.B. der automatischen Anpassung von sprach- und kulturspezifischen Formaten für Datum und Zeit, nicht zu übersetzenden Produktbezeichnungen etc. müssen an dieser Stelle spezifiziert werden. Insbesondere die Möglichkeit der automatischen Anpassung von Zahlenformaten kann beim Übersetzen äußerst hilfreich sein, wenn neben dem Format auch noch Umrechnungen erforderlich sind, wie dies im Fall der Lokalisierung von Inch-Angaben in amerikanischen Texten notwendig ist. Können, wie im Fall von **TRADOS** oder **SDLX**, entsprechende Einstellungen vorgenommen werden, übernimmt das TM auch die Umrechnung bei der Anpassung des amerikanischen an das metrische System.

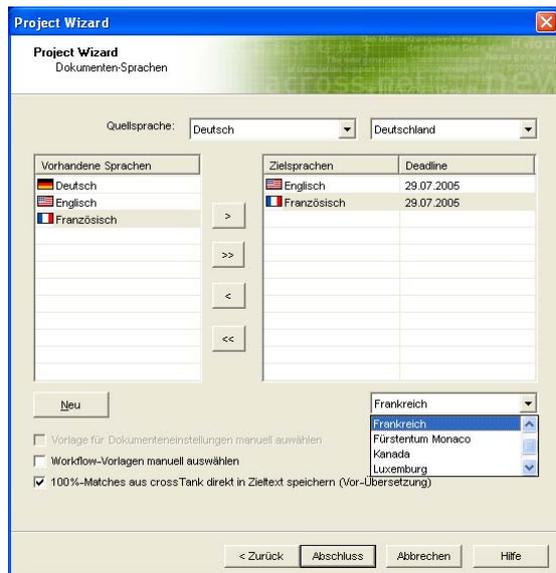


Abb. 18a: Projekteinstellungen in **across**.

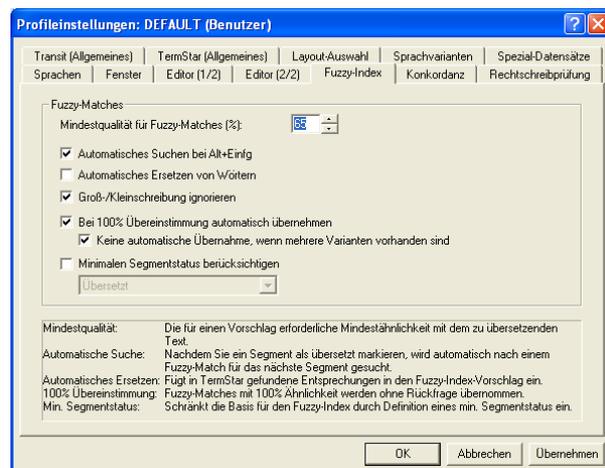


Abb. 18b: Einstellungen der TM-Optionen in **Transit**.

Sollen abweichend von den Standardeinstellungen Anpassungen der Segmentierungsregeln oder die Verwendung bestimmter Farben zur Darstellung von Vorschlägen aus dem Referenzmaterial vorgenommen werden, werden diese ebenfalls im Rahmen der Projekteinstellungen und zum Teil in zusätzlichen Profileinstellungen, die nicht projektabhängig geändert werden, festgelegt.

Ein TM oder verschiedene TMs

Für bestimmte Dokumenttypen oder Kunden werden häufig neue Translation Memories angelegt, so dass für ein zu bearbeitendes Projekt gegebenenfalls Referenzmaterial in verschiedenen TMs vorliegen kann. Bei einigen Produkten können daher auch mehrere TMs im Lese-Modus geöffnet werden. Lediglich **across** verfolgt hier eine andere Philosophie. Es werden grundsätzlich alle Segmentpaare in demselben TM, crossTank, abgespeichert. Ein Zugriff auf eine bestimmte Kunden- oder Sachgebiets-spezifische Teilmenge des TM ist durch Definition von Filterkriterien möglich. Laut Angaben des Herstellers soll es auch bei sehr großen TMs mit mehreren Zehn- oder Hunderttausend Segmentpaaren keine Einbußen in der Performanz des Systems geben.

Definition von Verwaltungsaufgaben

Nicht nur hinsichtlich der Datenhaltung von TM-Daten, auch bei der Verwaltung projektbezogener Daten gibt es wesentliche Unterschiede zwischen den Systemen. An dieser Stelle unterscheidet sich **across** von den übrigen Systemen dadurch, dass hier bereits im Rahmen des Projektdialogs administrative Angaben, die das Projektmanagement betreffen, festgehalten werden. So lassen sich mit **across** unter anderem kundenspezifische Daten verwalten und der Abgabetermin für einen Auftrag erfassen (Abb. 19). Darüber hinaus kann auch zwischen bestimmten Workflow-Szenarien ausgewählt werden. Hier stehen neben einem vordefinierten Übersetzungs-Workflow auch weitere mit „Terminologiearbeit und Übersetzen“ sowie „Extern Editieren“ bezeichnete Workflows zur Verfügung. In Abhängigkeit vom gewählten Workflow werden auf dem so genannten crossBoard unterschiedliche Aufgaben zur Bearbeitung aufgelistet (Abb. 20), und die Arbeitsumgebung wird bei der Auswahl einer entsprechenden Aufgabe automatisch entsprechend der Aufgabenstellung eingerichtet.

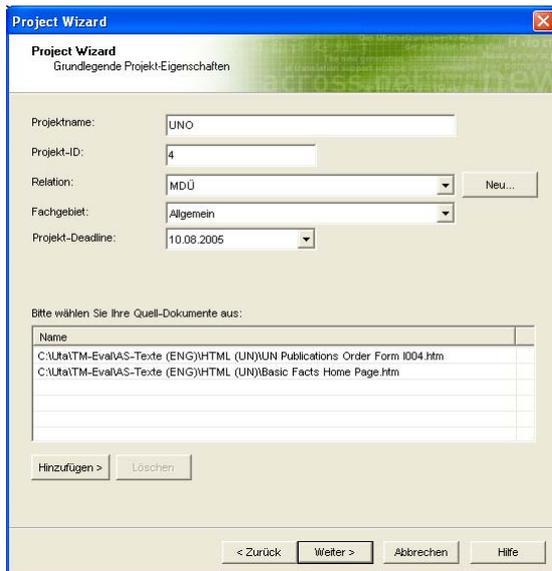


Abb. 19: Erfassen administrativer Angaben beim Einrichten eines Projekts in **across**.

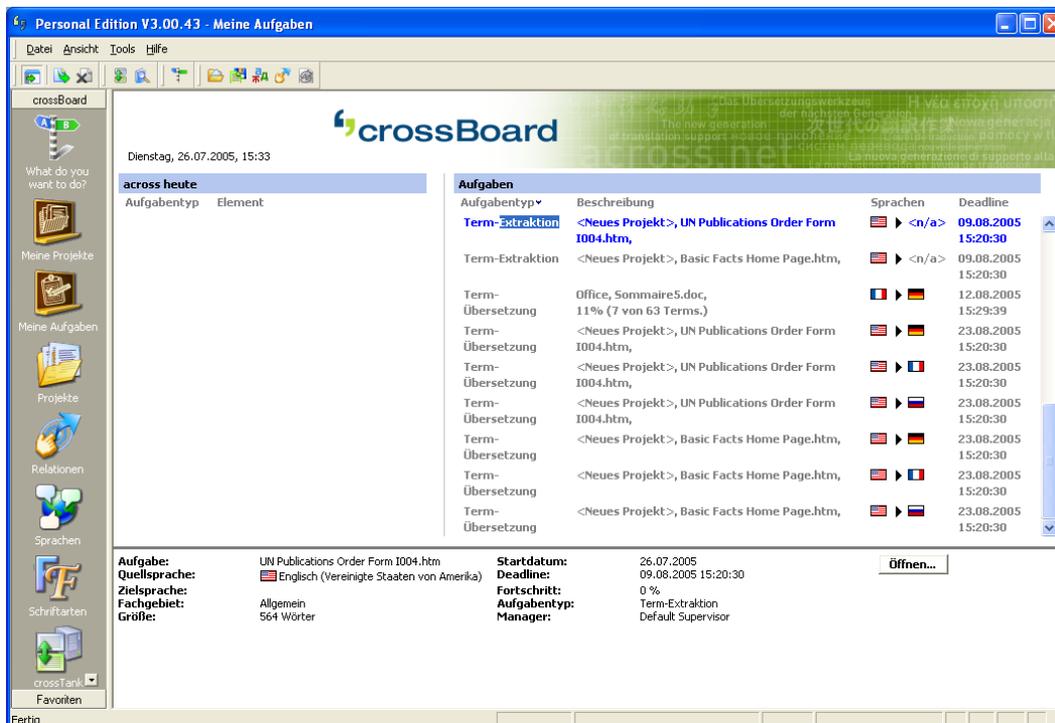


Abb. 20: Projektübersicht im crossBoard von **across**.

Bereitstellen von Terminologie

Im Zuge der Vorbereitung der Terminologie für ein Projekt kann das von einigen Systemen bereitgestellte Modul zur Termextraktion genutzt werden. In **across** ist dieses Modul Bestandteil der Einzelplatzversion, von **TRADOS** steht hierzu das Programm „MultiTerm Extract“ zur Verfügung und von **SDL** wird neuerdings mit SDL PhraseFinder ein zusätzliches Tool angeboten, das diese Aufgabe ebenfalls unterstützt. Während man für die Termextraktion bei **TRADOS** und **SDL** separate Anwendungen aufruft, wählt man bei **across** in den Projekteinstellungen lediglich den Workflow „Terminologiarbeit und Übersetzen“ und gelangt so automatisch in die Termextraktionskomponente (Abb. 21). In **Wordfast** kann man über das Wordfast-Menü ebenfalls Termkandidaten extrahieren, die

entsprechend des in **Wordfast** unterstützten Formats als Liste im Text-Format gespeichert werden.

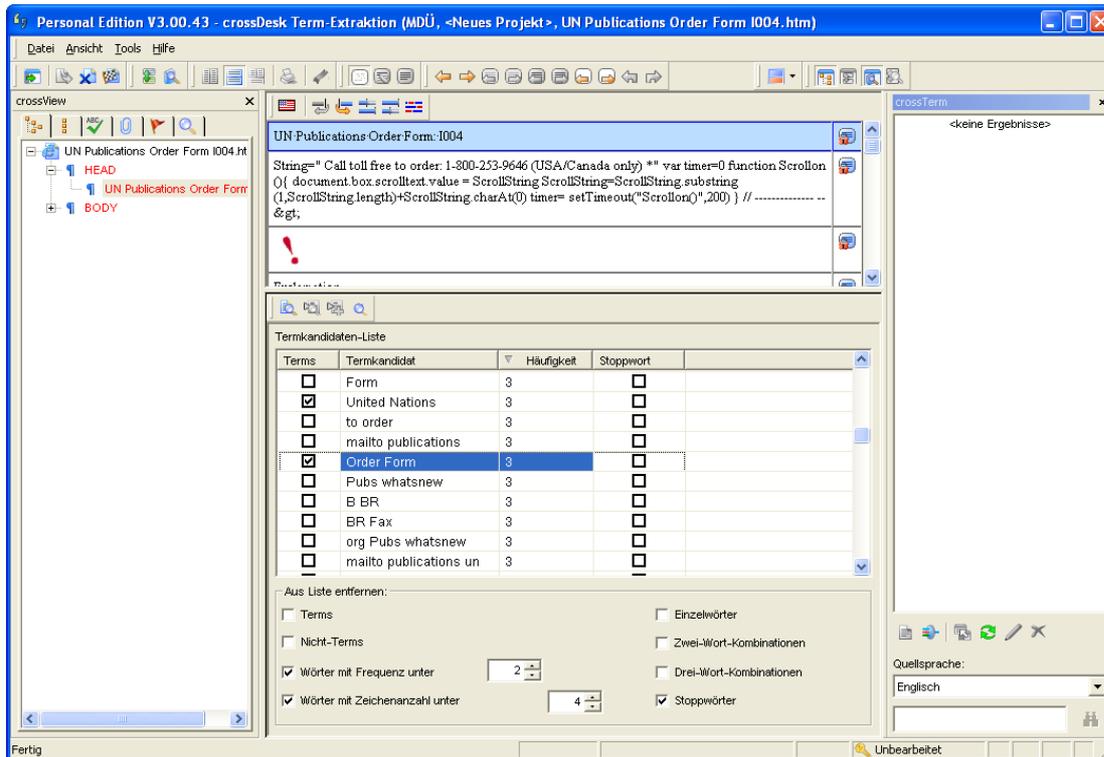


Abb. 21: Arbeitsumgebung zur Termextraktion in **across**.

Bei den Optionen zur Bereitstellung von Terminologie gibt es zwischen den hier betrachteten Systemen erhebliche Unterschiede. Während die auf Word aufbauenden TM-Systeme **MetaTaxis** und **Wordfast** in erster Linie bilinguale Textdateien mit Äquivalentpaaren, gegebenenfalls auch mit einigen zusätzlichen Attributen, als Terminologielisten unterstützen und **Wordfast** hier auch nur einen Datenimport aus Textdateien per Kopieren und Einfügen ermöglicht, sind die übrigen Systeme mit zum Teil sehr umfangreichen Funktionen zum Import von externen Terminologiedaten im Excel- oder Textformat oder dem Terminologieformat eines anderen Systems ausgestattet (Abb. 22 und 23a). Bei **TRADOS** übernimmt diese Funktion das Programm-Modul „MultiTerm Convert“ (Abb. 23b), das allerdings vorab gesondert aufgerufen werden muss.

Die Terminologiekomponenten von **across**, **Déjà Vu**, **SDLX**, **STAR** und **TRADOS** erlauben sowohl eine multilinguale als auch eine begriffsorientierte Datenhaltung, bei der synonyme Benennungen eines Begriffs in einem Eintrag abgelegt werden können, was die Nutzung der terminologischen Daten auch zu Zwecken der fachlichen Informationsrecherche sowie für Übersetzungsprojekte mit ganz unterschiedlichen Sprachpaaren möglich macht. Die Terminologiekomponenten **MultiTerm** von **TRADOS** und **TermStar** von **STAR** sind darüber hinaus auch als separate Produkte erhältlich und stehen somit auch für terminologische Arbeiten außerhalb des Übersetzungskontextes zur Verfügung.

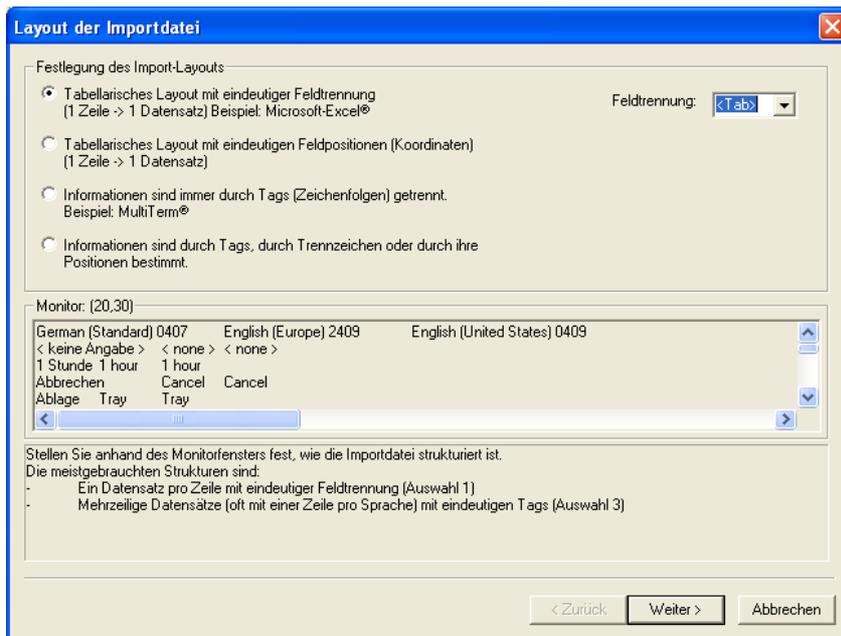


Abb. 22: Unterstützung beim Import von so genannten benutzerdefinierten Formaten in **TermStar**, der Terminologiekomponente von **Transit**; hier: Import einer Terminologieliste im Tabulator-getrennten Textformat, das von Excel aus erzeugt werden kann.

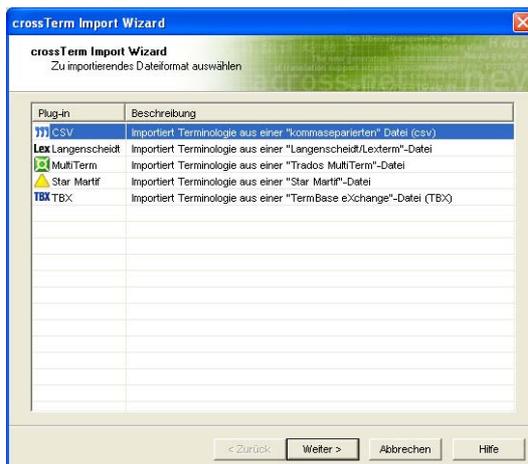


Abb. 23a: Unterstützung beim Import von Terminologie in **crossTerm**, der Terminologiekomponente von **across**.

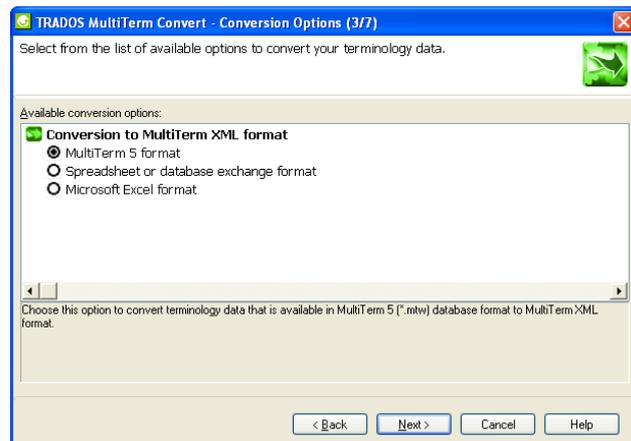


Abb. 23b: Unterstützung beim Import von Terminologie in **MultiTerm**, der Terminologiekomponente von **TRADOS** mit MultiTerm Convert.

Bereitstellen vorhandenen Referenzmaterials

Liegen elektronische Fassungen von Ausgangs- und Zieldateien früherer Übersetzungsprojekte zur Verfügung, die nicht mit TM-Technologie bearbeitet wurden, kann durch Alignment ein TM erstellt werden. Die Produkte **across** (Abb. 24), **Déjà Vu** (Ausrichtungsbearbeitung, Abb. 25), **SDLX** (SDL Align, Abb. 26), **TRADOS** (WinAlign, Abb. 27) und **Transit** (Alignment-Projekt, Abb. 28a und Abb. 28b) stellen hierzu entsprechende Programmmodule zur Verfügung, die synchronisierte Segmentpaare in das entsprechende TM-Format konvertieren oder in ein angegebenes TM importieren können. **Wordfast**, das ein Alignment über die +Tools erstellen kann, und **MetaTaxis** verwenden für das Alignment zwei sich parallel öffnende Wordfenster bzw. eine Tabelle, in der die synchronisierten Segmente nummeriert untereinander angezeigt werden.

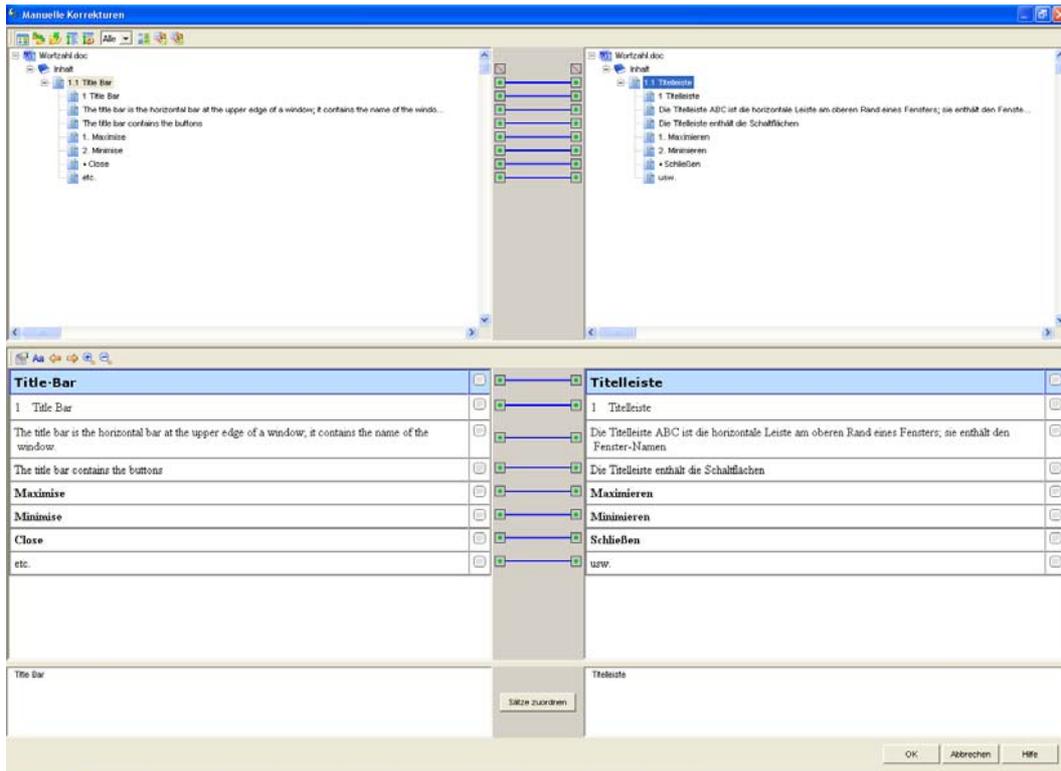


Abb. 24: Alignment-Oberfläche bei **across**.

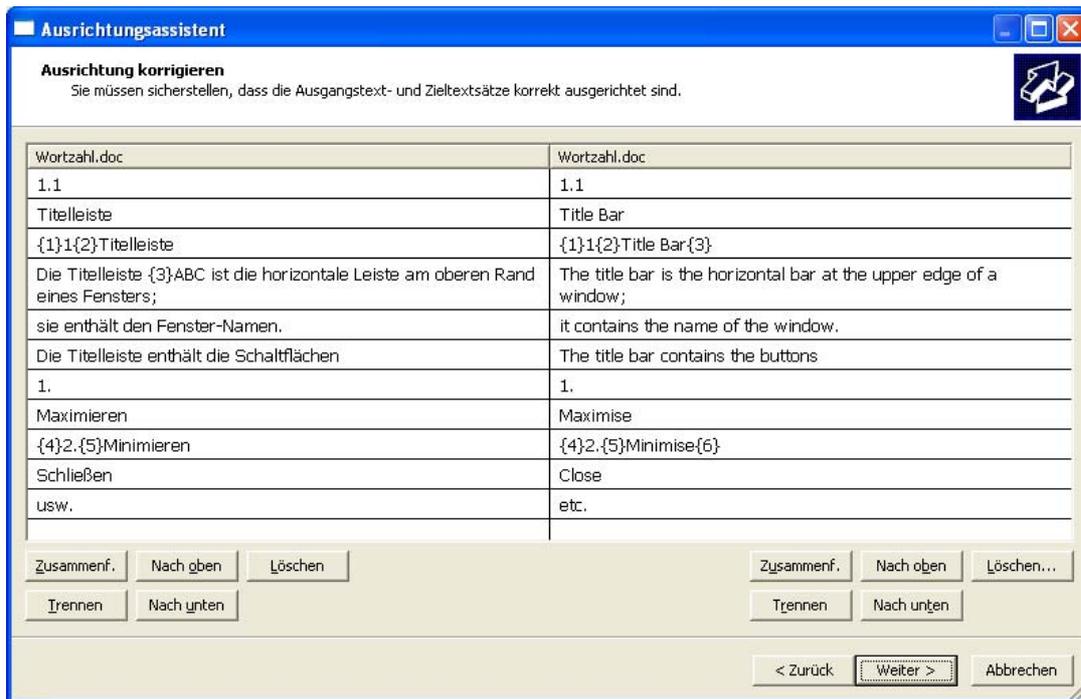


Abb. 25: Alignment-Oberfläche bei **Déjà Vu**.

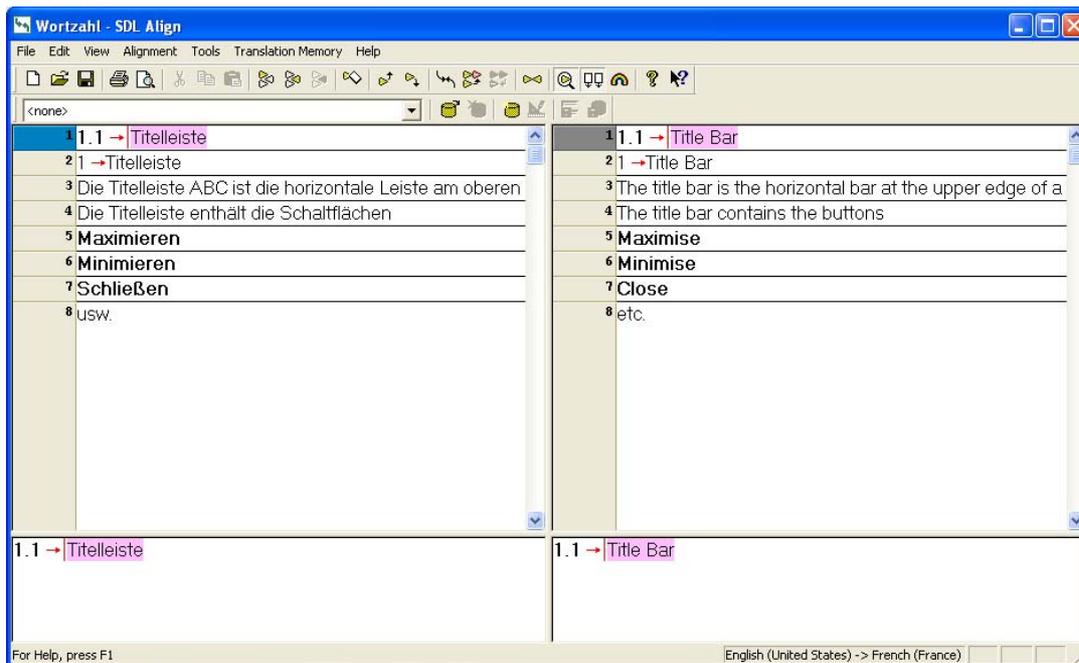


Abb. 26: Alignment-Oberfläche bei **SDL Align**.

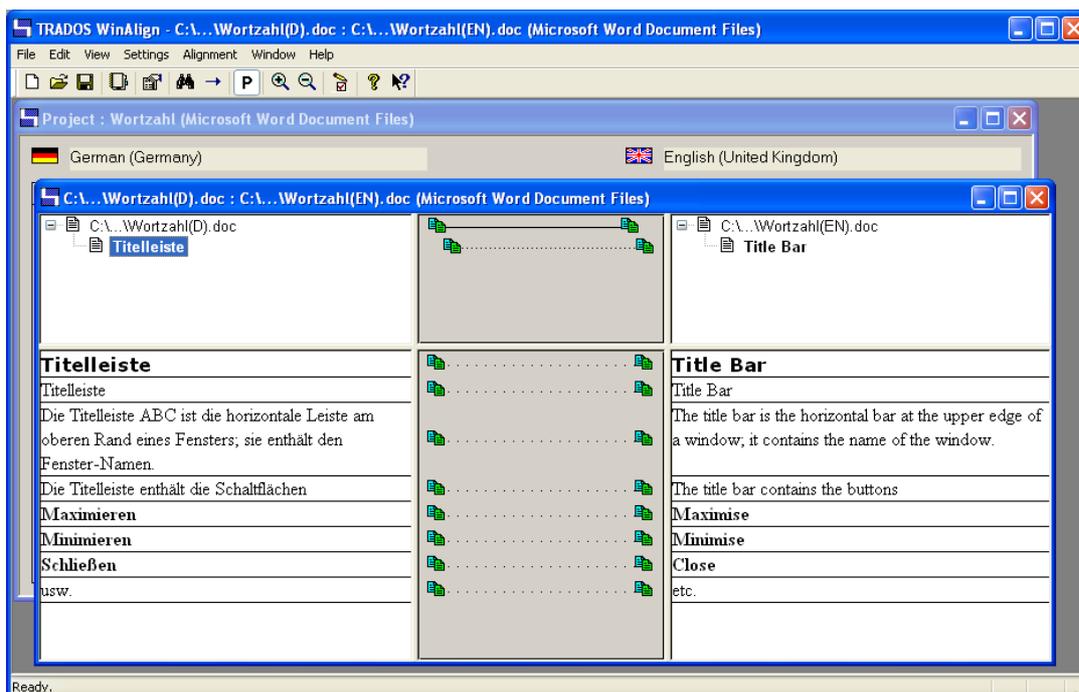


Abb. 27: Alignment-Oberfläche bei **TRADOS WinAlign**.

Da sich Übersetzungen in der Struktur oder Zahl der Segmente vom betreffenden Ausgangstext unterscheiden können, lassen sich bestimmte Faktoren, die bei der Analyse der Dateien berücksichtigt werden, in Abhängigkeit von den zu synchronisierenden Dokumenten vom Benutzer festlegen. Besonders detailreiche Gewichtungsoptionen bietet hier z.B. **Transit** von Star (Abb. 28a).

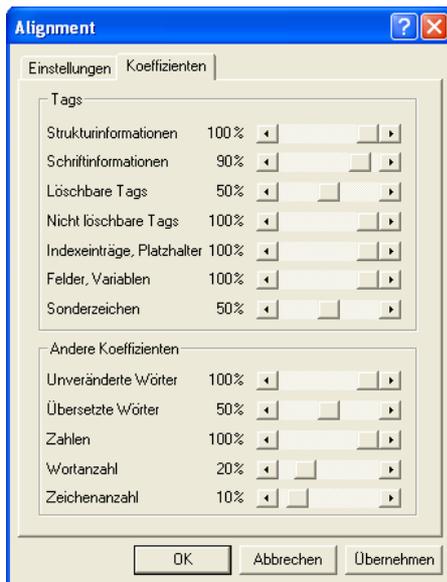


Abb. 28a: Gewichtungsfaktoren für die automatische Zuordnung von Segmenten beim Alignment mit **Transit**.

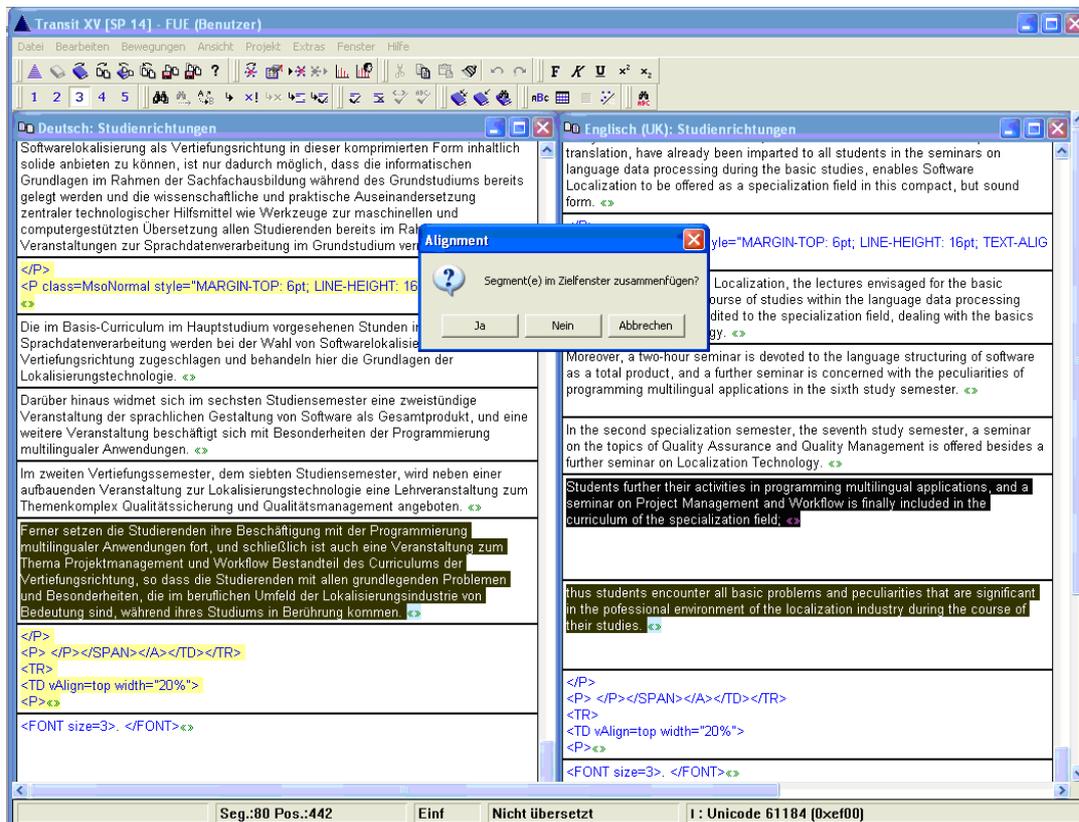


Abb. 28b: Alignment-Oberfläche bei **Transit** mit Dialog zum Zusammenfügen von Segmenten.

Aufwands- und Kostenkalkulation

Ist ein Übersetzungsprojekt eingerichtet worden, bei dem auf bereits vorhandenes Referenzmaterial zugegriffen werden kann, lassen sich durch die Analyse und den Abgleich der zu übersetzenden Dokumente mit dem Referenzmaterial Aufwand und Kosten des Auftrags kalkulieren. Hierzu stellen alle Systeme Analysefunktionen bereit, mit denen neben der Wort-, Segment- oder Zeilenzahl der Dokumente zum Teil auch bereits Preise berechnet werden können (vgl. Abb. 9, Abb. 29a und 29b). Die Ergebnisse lassen sich in der Regel als

Excel-, RTF- oder HTML-Datei exportieren, um die Daten für die Angebotserstellung weiterverwenden zu können (Abb. 30a, 30b).

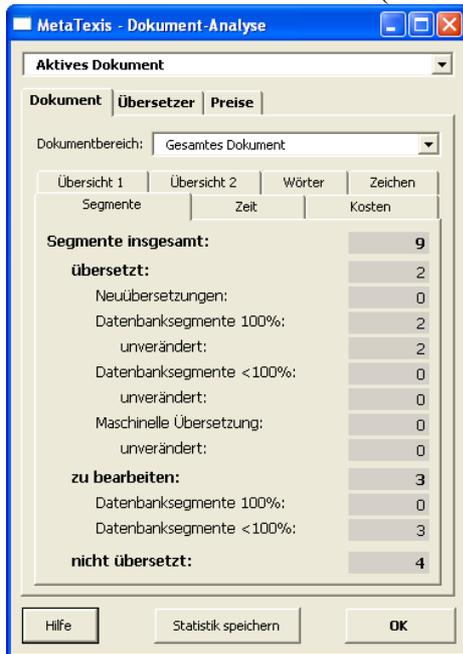


Abb. 29a: Dokumentenanalyse mit Angaben über die Zahl und den Match-Wert-Status der zu bearbeitenden Segmente in **MetaTaxis**.



Abb. 29b: Einstellungsmöglichkeiten für die Preiskalkulation in **MetaTaxis**.

Document	Segment type	Segment count	Source word count	Source alternative word count	Source character count	Source character count (no spaces)	Target character count	Target character count (no spaces)
1	Confirmed (transl.)	0	0	0	0	0	0	0
2	100% matched	2	5	4	28	26	24	20
3	95% to 99%	0	0	0	0	0	0	0
4	85% to 94%	0	0	0	0	0	0	0
5	75% to 84%	1	16	16	100	93	102	82
6	50% to 74%	0	0	0	0	0	0	0
7	Less than 50%	0	0	0	0	0	0	0
8	Machine translate	0	0	0	0	0	0	0
9	Untranslated	5	9	9	74	70	74	70
10	Total	8	30	29	210	189	200	172

Abb. 30a: Excel-Export einer Dokumentenanalyse mit **SDLX**.

Datei(en)	Vorübersetzt	Teilweise übersetzt	100-95%	94-85%	84-75%	74-50%	Verbleibende nicht übersetzte Einheiten	Summe
Basic Facts Home Page	0.02 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	14.03 €	14.05 €
UN Publications Order Form I004	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	27.48 €	27.48 €
Summe (nicht vermindert durch Int. Repetitions)	0.02 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	41.51 €	41.53 €
Gefundene Int. Repetitions (reduziert durch den Grenzwert)	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €
Summe (vermindert durch Int. Repetitions)	0.02 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	41.51 €	41.53 €
Summe mit Gewichtungsfaktor	0.02 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	41.51 €	41.53 €
Summe mit Dehnungsfaktor	0.02 €	0.00 €	0.00 €	0.00 €	0.00 €	0.00 €	49.81 €	49.83 €

Abb. 30b: RTF-Export einer Dokumentenanalyse mit **Transit**.

Übersetzen

Nach den erfolgten Vorarbeiten und Projekteinstellungen kann mit der Übersetzungsarbeit begonnen und in den Übersetzungseditor gewechselt werden, der bei **TRADOS** als Programm (z.B. TagEditor) gesondert gestartet werden muss, bei **SDLX** über das Programmfenster und bei **across** über das crossBoard-Modul „Meine Aufgaben“ aufgerufen wird. Bei **Transit** wird ein im Rahmen der Projektdefinition angelegtes Sprachpaar über das Dateimenü geöffnet, und bei **Déjà Vu** werden die zu übersetzenden Dokumente nach Abschluss der Projekteinstellungen in den Übersetzungseditor importiert.

Übersetzungseditor

Alle Programme zeigen Matches mit dem Referenzmaterial an und verfügen über eine Terminologieerkennung beim Übersetzen. Termini der Terminologiekomponente werden angezeigt und können per Mausklick oder Tastenkombination in den Zieltext übernommen werden, sofern sie nicht im Rahmen einer Vorübersetzung bereits automatisch in den Zieltext eingefügt wurden. Dies gilt für alle 100%-Matches, unter Umständen auch Matches mit niedrigeren Match-Werten. Letzteres ist jedoch bei mehreren möglichen Äquivalenten eines AS-Terminus und der vielfach veränderten Wortstellung in der betreffenden Zielsprache nur bei bestimmten Dokumenten und Sprachen nützlich. Während Termini aus der Terminologiedatenbank in **MetaTaxis** im AS-Segment in grünen eckigen Klammern zusammen mit ihrem ZS-Äquivalent angezeigt und im ZS-Segment an der Stelle des AS-Segments substituiert werden, werden Termini in **Wordfast** in der Statusleiste des Word-Fensters unten links angezeigt und können über Tastenkombinationen sowohl mit dem jeweiligen zielsprachigen Äquivalent in der Terminologieleiste angezeigt oder in den Zieltext übernommen werden. Full- und Fuzzy-Matches werden durch die Farbe des Textfensters, das die Übersetzung aufnimmt, gekennzeichnet. Dabei werden 100%-Matches standardmäßig grün und Fuzzy-Matches gelb dargestellt. Den genauen Match-Wert ersieht man als Übersetzer aus den zwischen geschweiften Klammern stehenden Zahlenangaben (Abb. 31).

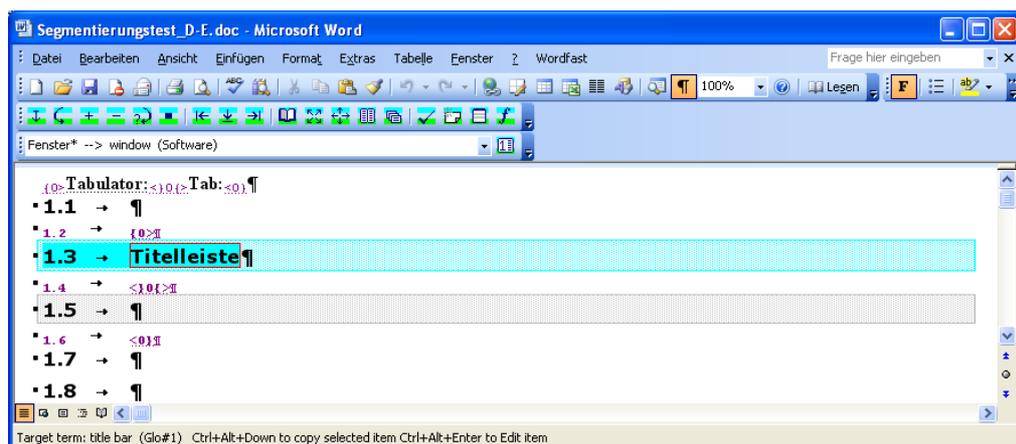


Abb. 31: Übersetzen mit **Wordfast**. Im Glossar hinterlegter Terminus wird im AS-Textfeld blau hinterlegt und durch Tastenkombination selektiert, was im Programm durch einen roten Rahmen signalisiert wird.

In **MetaTaxis** können die Match-Werte wahlweise angegeben werden. Dies wird in den Dokument-Optionen, also der Projekteinrichtung, festgelegt. Daneben werden die verschiedenen Matchtypen auch farblich unterschieden. Abweichungen zwischen zu übersetzendem Segment und Segment im Referenzmaterial wird mittels gestrichelter roter Rahmen um die betreffenden Textstellen angezeigt (Abb. 32a). Außer der Bearbeitung von Dokumenten in Word bietet **MetaTaxis** einen so genannten Dialogbox-Modus an (Abb. 32b),

der insbesondere bei der Übersetzung von Dokumenten mit Textfeldern, die zu klein sind, um Ausgangs- und Zielsegment darin darzustellen, vorteilhaft ist.

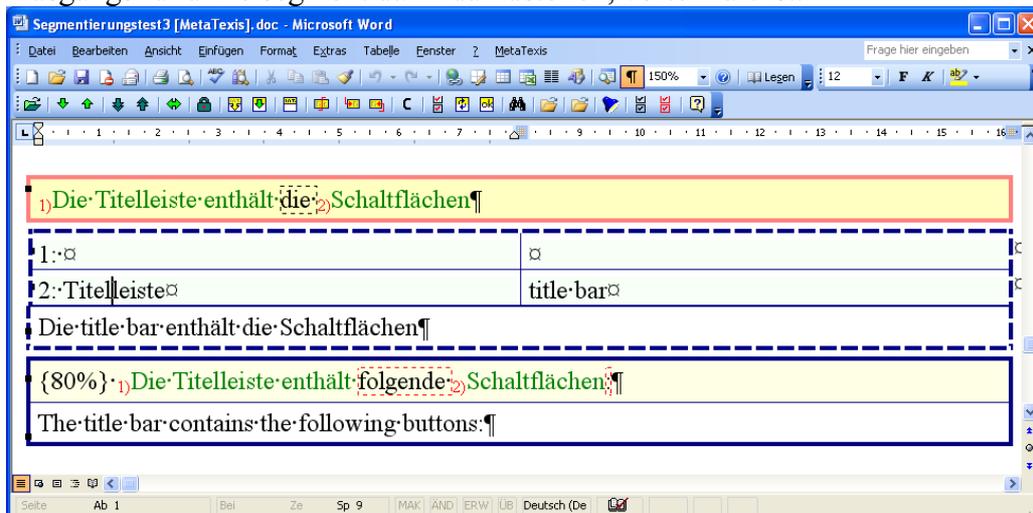


Abb. 32a: Übersetzen mit **MetaTaxis**. Fuzzy-Match wird in einem zusätzlich eingeblendeten Textfeld mit farbiger Hervorhebung der Abweichungen angezeigt.

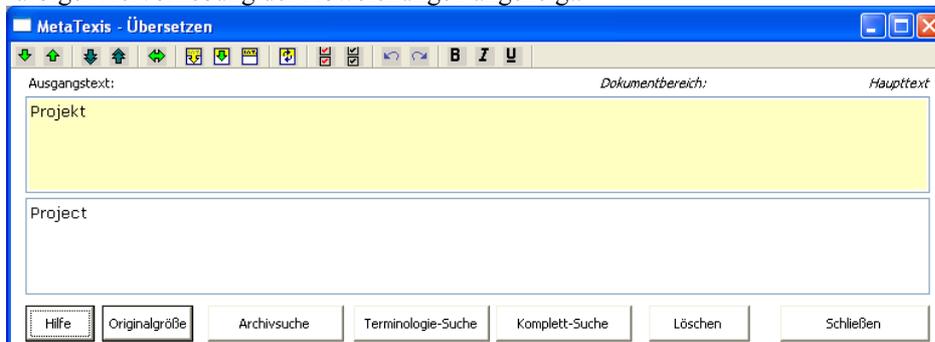


Abb. 32b: Übersetzen mit **MetaTaxis** im Dialogbox-Modus.

Verwendet man **across** (Abb. 33), **Déjà Vu** (Abb. 34), **SDLX** (Abb. 35), **TRADOS** (Abb. 36) oder **Transit** (Abb. 37), so werden Übersetzungsvorschläge aus dem Referenzmaterial sowie die jeweils in einem Segment enthaltene Terminologie in einem dafür vorgesehenen Teilfenster der Übersetzungsumgebung eingeblendet, sofern die entsprechenden Voreinstellungen und Ansichtsoptionen ausgewählt wurden.

Unterschiede zwischen dem zu übersetzenden Segment und dem Vorschlag aus dem Referenzmaterial werden in den Übersetzungsumgebungen ebenfalls farblich hervorgehoben, wodurch die Unterschiede insbesondere in der **Translator's Workbench** (Abb. 36), auf dem crossDesk von **across** (Abb. 33) und im TM Difference View von **SDLX** (Abb. 35) schnell erfasst werden können. Nicht immer ist der Vorschlag aus dem Referenzmaterial mit dem höchsten Match-Wert derjenige, der im zu bearbeitenden Dokument wiederverwendet werden kann. Daher ist eine Auflistung der verschiedenen Matches, wie sie im Teilfenster „Fuzzy“ von **across** (Abb. 33) erfolgt, besonders effizient, weil andere Matches hier nicht erst per Mausklick sichtbar gemacht werden müssen.

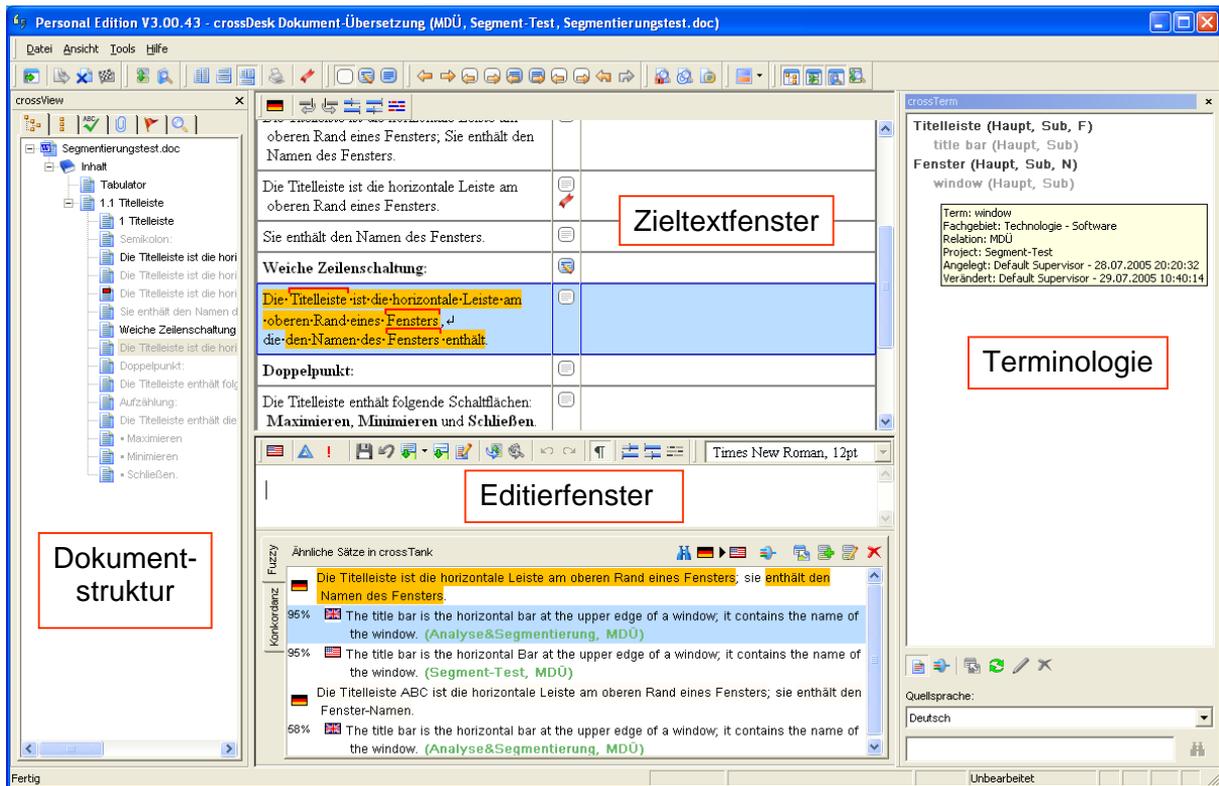


Abb. 33: Übersetzen mit **across**.

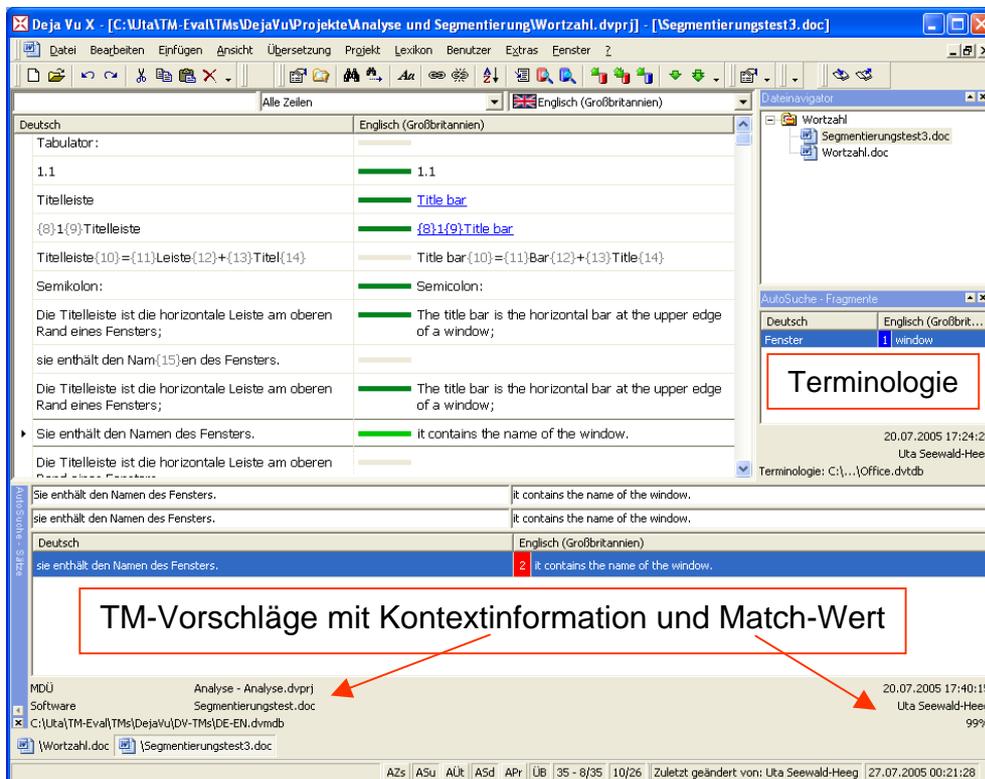


Abb. 34: Übersetzen mit **Déjà Vu**.

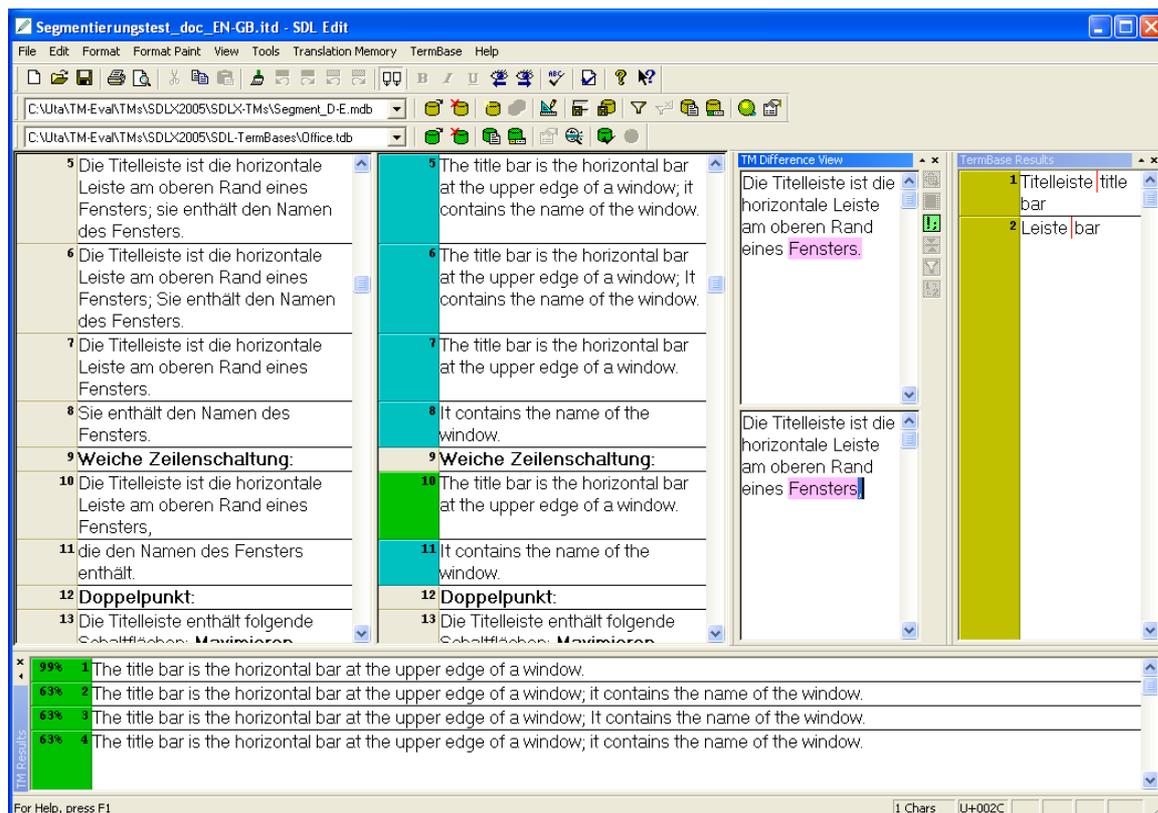


Abb. 35: Übersetzen mit **SDLX**.

across, **SDLX** und **TRADOS** TagEditor zeigen das zu übersetzende Dokument ebenso wie die übrigen Systeme in einem Semi-WYSIWYG-Format an, verfügen allerdings zusätzlich über eine Vorschaufunktion, mit der das jeweils in Bearbeitung befindliche Dokument in seiner AS- oder seiner ZS-Fassung betrachtet werden kann. Bei **across** lässt sich die Vorschau über die Symbole „Vorschau Quelltext“ bzw. "Vorschau Zieltext" in der Symbolleiste aufrufen, bei **SDLX** über die Augensymbole in der Symbolleiste und im TagEditor über die Registerkarten am unteren Rand des Editorfensters.

Nicht selten müssen getroffene Entscheidungen beim Übersetzen noch einmal verifiziert, unter Umständen auch Rücksprache mit dem Kunden genommen werden, um inhaltliche Fragen klarzustellen. Während man hierzu früher Notizen extern aufschreiben und zusätzlich markieren musste, auf welche Stelle des AS-Textes die Notiz Bezug nahm, erlauben die TM-Systeme, einzelnen Segmenten Kommentare bzw. Notizen zuzuordnen. So hat **Transit** beispielsweise auch für diese klassischen Hilfsmittel in der Übersetzungsumgebung ein separates Fenster vorgesehen, das Notizen zu einzelnen Segmenten aufnimmt (Abb. 37), so dass Wörterbuch, Notizblock und Editor auf einer Oberfläche integriert sind.

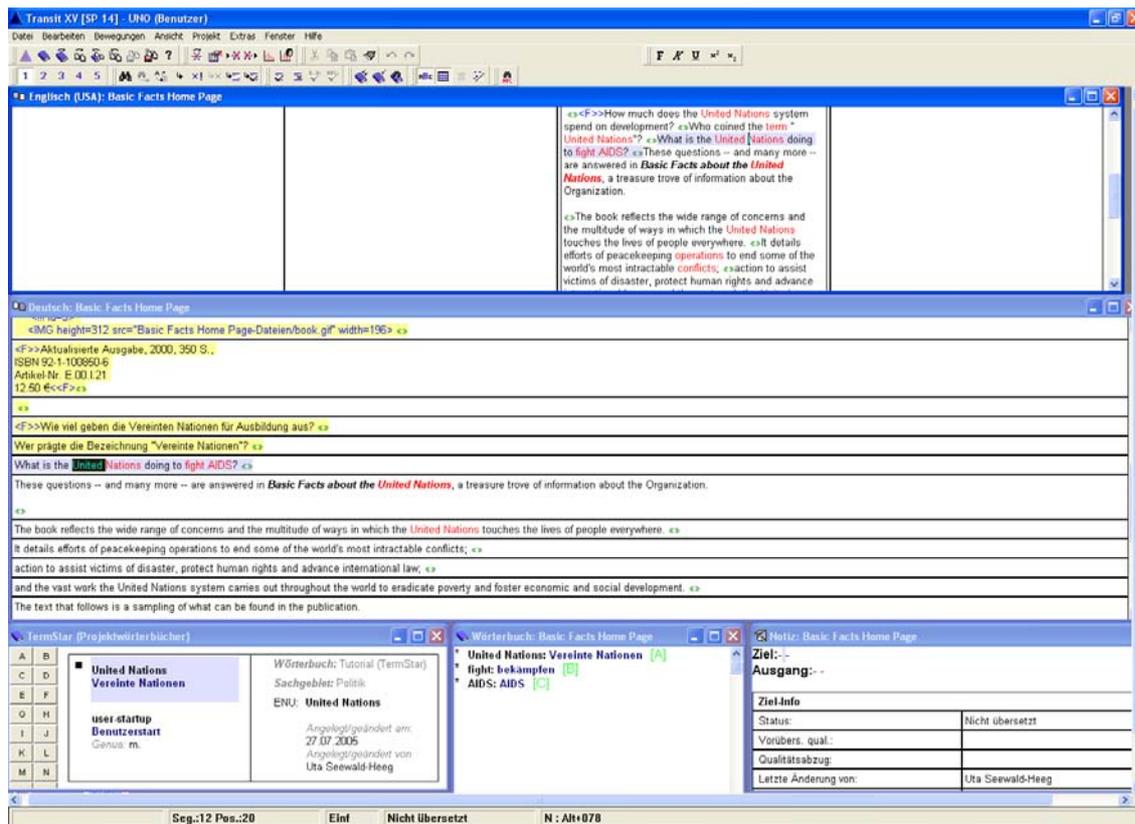


Abb. 37: Übersetzen mit **Transit** von Star. Links unten ein Eintrag in TermStar, der Terminologieverwaltung von Transit, und in der Mitte die im geöffneten Segment hinterlegten Termini mit den jeweiligen Äquivalenten.

Terminologieerkennung

Ein Wort in einem gerade geöffneten Segment, das in der Terminologiedatenbank hinterlegt ist, wird bei den verschiedenen Systemen entweder durch eine farbliche Markierung des Terminus im AS-Segment hervorgehoben oder durch die Anzeige des Terminus und seiner möglichen Äquivalente in einem Teilfenster der Übersetzungsumgebung oder aber durch beide Varianten (**across**, **TRADOS**, **Transit**, **SDLX**) markiert. Im Bedarfsfall lässt sich auch der vollständige terminologische Eintrag aus der Terminologiedatenbank einblenden, um so gegebenenfalls besser entscheiden zu können, ob ein aufgeführtes Äquivalent in einem gegebenen Kontext auch tatsächlich eine zutreffende Übersetzung ist. Bei **Transit** wird der Eintrag im Termstar-Wörterbuch (Abb. 37), das als Fenster während der Übersetzungssitzung stets eingeblendet sein kann, über die Symbolleiste aufgerufen. Per Doppelklick auf das Wörterbuchsymbol des Terminus im Terminologiefenster der Translator's Workbench wird auch hier der vollständige Eintrag aus dem Wörterbuch eingeblendet und etwaige Synonyme in einer ebenfalls recherchierbaren Liste aufgeführt (Abb. 38). In **SDLX** kann der Eintrag eines selektierten Terminus in der SDL Termbase in einem so genannten MiniBrowser über die Symbolleiste des SDLX-Editors eingesehen werden (Abb. 39). In **crossTerm** für einen im gerade geöffneten Segment auftretenden Terminus hinterlegte Informationen werden als Tooltipp (Kontextinformation, die beim Gleiten des Cursors über den AS-Terminus oder einen seiner Äquivalente eingeblendet werden) direkt auf der Übersetzungsoberfläche crossDesk angezeigt. Bei Bedarf kann der Eintrag per Doppelklick auf den AS-Terminus auch vollständig in der crossTerm-Umgebung angezeigt werden (Abb. 40).

Gerade bei der Übersetzung von Sprachen mit ausgeprägter Flexion kann die Terminologieerkennung allerdings nur dann auch zur Sicherstellung der terminologischen

Konsistenz eingesetzt werden, wenn ein System auch flektierte Wortformen erkennt. Dies ist bei **across**, **TRADOS**, **Transit** und **Déjà Vu** gegeben, während **Wordfast** flektierte Formen nur erkennt, wenn sie trunkiert (mit Sternchen gekennzeichnet) im Wörterbuch hinterlegt sind, eine Darstellungsmöglichkeit, die auch für **MetaTaxis** angekündigt ist.



Abb. 38: Von der **Translator's Workbench** aus aufgerufener Eintrag im Terminologieverwaltungsprogramm MultiTerm. Die Liste am linken Fensterrand führt zusätzlich alle zu diesem Eintrag in MultiTerm erfassten alternativen Benennungen auf.



Abb. 39: MiniBrowser der SDL Termbase, der über die Symbolleiste von **SDLX** aufgerufen wird.

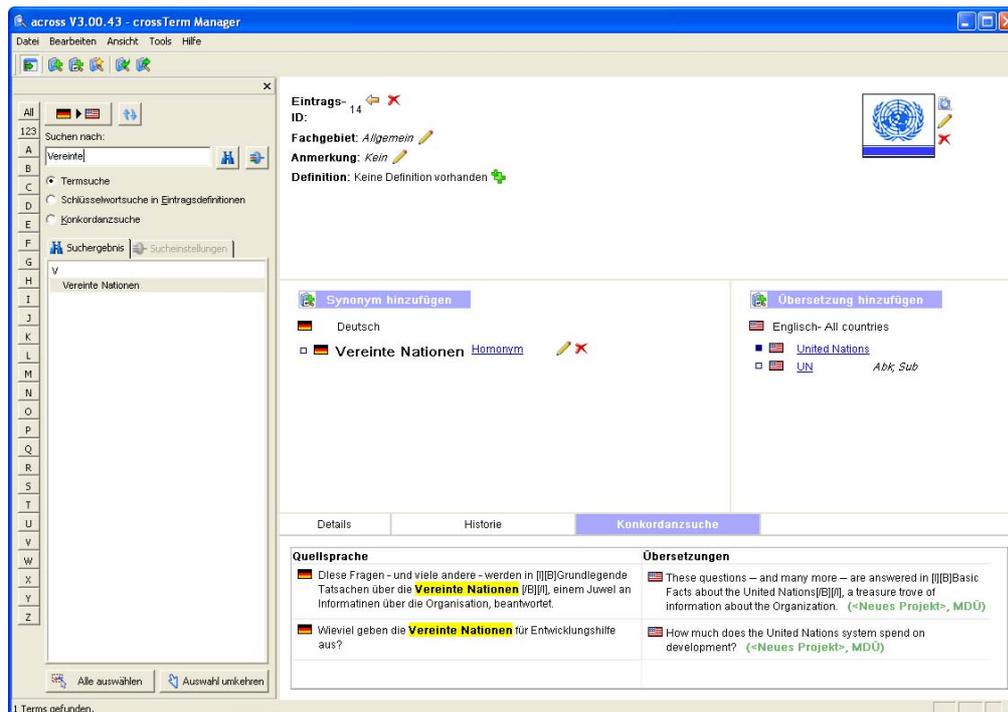


Abb. 40: **crossTerm**-Oberfläche mit eingblendetem Suchergebnis für den Terminus und Konkordanz für das Suchwort.

Konkordanzsuche

Erhält man für ein Segment, das einen bereits recherchierten Terminus enthält, der jedoch nicht in die Terminologiedatenbank aufgenommen wurde, keinen Vorschlag aus dem Referenzmaterial, so bieten die Systeme eine Zeichenkettensuche im Referenzmaterial, die häufig als Konkordanzsuche bezeichnet wird.

In **MetaTaxis** kann eine Konkordanzsuche mittels der Funktion „TM anzeigen“ durchgeführt werden, die in der Symbolleiste über das Fernglas aufgerufen werden kann. Ähnlich wird bei **Déjà Vu** verfahren, wo über eine erweiterte Suche im TM alle Segmente, die eine zuvor markierte Zeichenkette enthalten, in einem Fenster aufgelistet werden. Auch **Wordfast** bietet eine solche Funktion, die über die Wordfast-Symbolleiste in Word aufgerufen werden kann (Abb. 41). Das entsprechende Symbol bei **TRADOS** mit zwei sich diagonal überkreuzenden Pfeilen stand hier wohl Pate. Das Ergebnis der Konkordanzsuche wird bei **TRADOS** in einem Fenster aufgelistet, in dem das Suchwort im AS-Segment der angezeigten Segmentpaare farbig hinterlegt ist und vom System vergebene Angaben zum Speicherdatum und zum Bearbeiter der betreffenden Übersetzung angezeigt werden (Abb. 7). Ähnlich ist die Anzeigefunktion in **across**, wo die Ergebnisse der Konkordanzsuche in demselben Teilfensterbereich wie die Fuzzy-Matches angezeigt werden (Abb. 42). Als Zusatzinformationen zu den einzelnen Segmenten werden hier Kunde und Projekt angezeigt. In **across** kann ebenso wie in **SDLX** sowohl nach Zeichenketten in der Ausgangssprache als auch in der Zielsprache gesucht werden. Während man in **SDLX** (Abb. 43) hier lediglich zwei verschiedene Menüoptionen oder Tastenkombinationen für die Suche in der Ausgangs- bzw. in der Zielsprache ausführt, muss man in **across** zuvor in der Suchleiste, in die das Suchwort manuell eingetragen werden muss, die Sprachrichtung entsprechend ändern. Allerdings lässt **across**, das ja multilinguale TMs anlegt, auch Suchen in anderen Sprachpaaren als derjenigen zu, die im Übersetzungseditor bearbeitet wird. Auch in **Transit** ist sowohl eine Suche in der

Quell- als auch in der Zielsprache möglich. Die Suchergebnisse werden in einem zusätzlichen sich öffnenden Fenster angezeigt (Abb. 44).

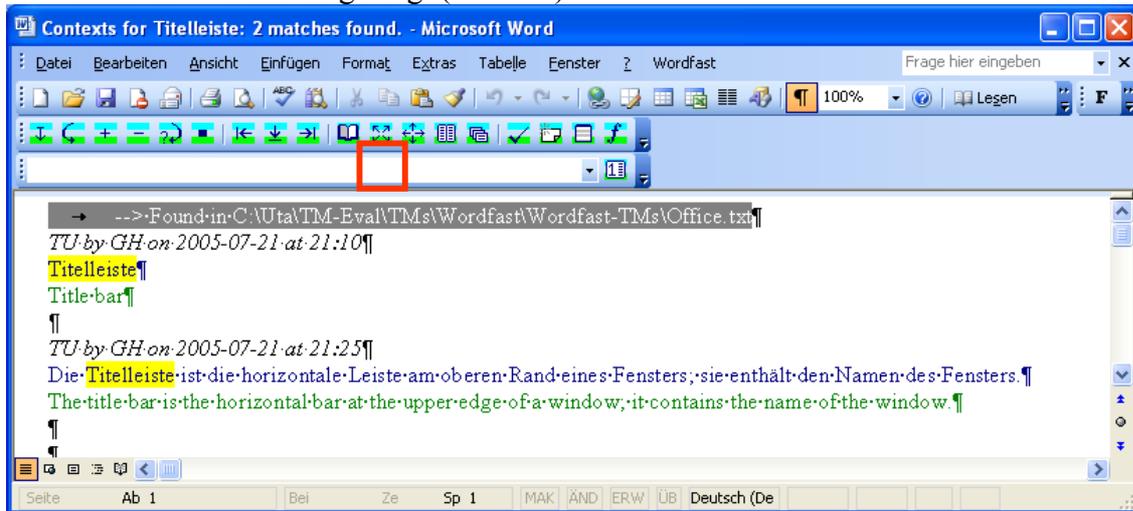


Abb. 41: Ergebnisfenster der Konkordanz- bzw. Kontextsuche in **Wordfast**.

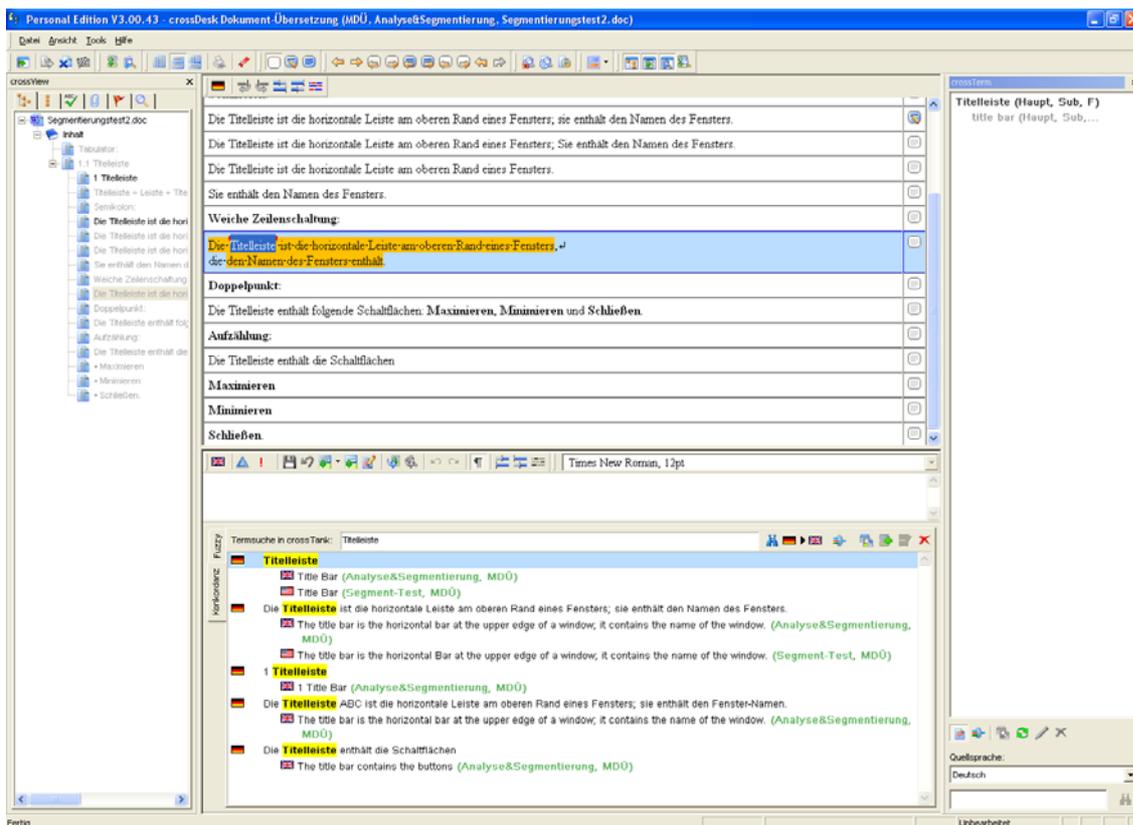


Abb. 42: Konkordanzsuche in **across**.

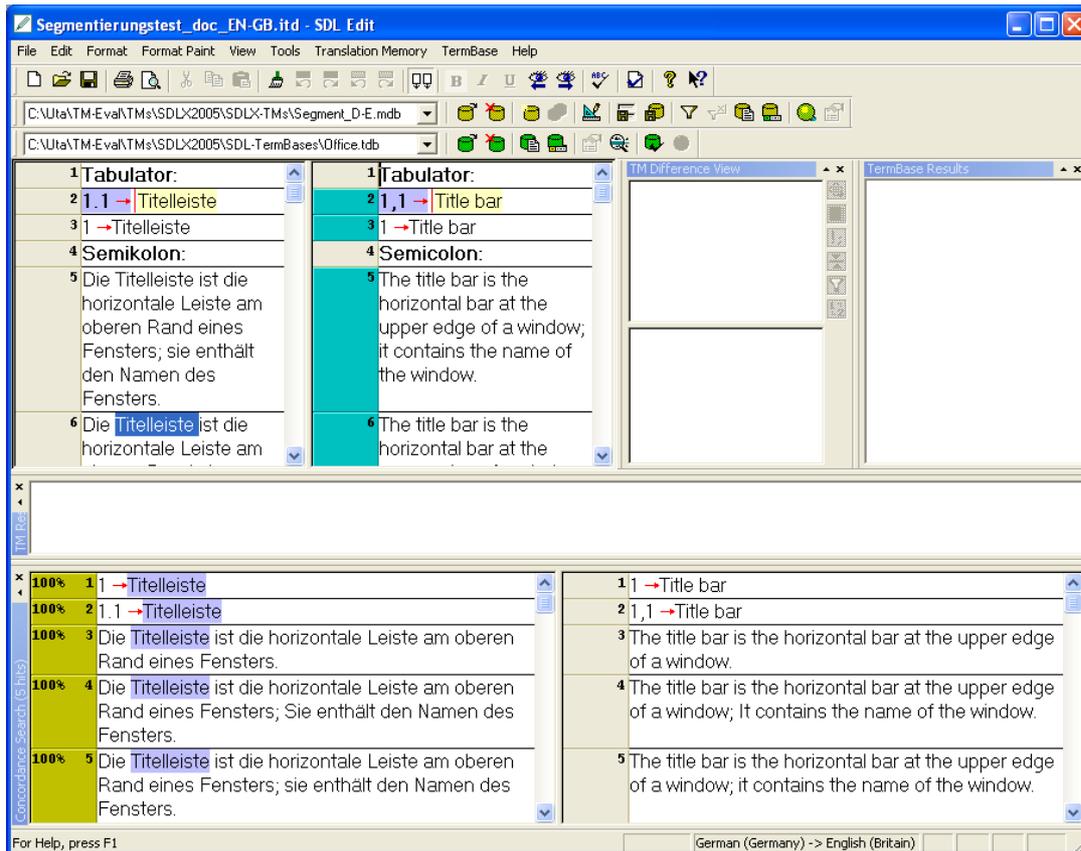


Abb. 43: Konkordanzsuche in **SDLX**.

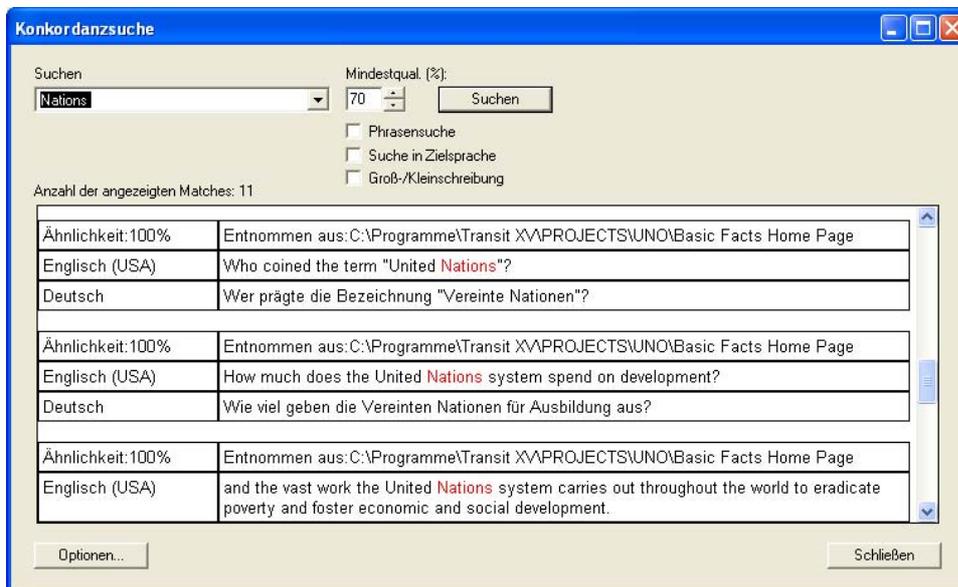


Abb. 44: Konkordanzsuche in **Transit**.

Erzeugen des Zieldokuments

Nach Abschluss einer Übersetzung und erfolgreicher Überprüfung der übersetzten Dokumente auf orthografische Korrektheit, terminologische Konsistenz, Formatierung etc. müssen die Dokumente wieder in ihr ursprüngliches Format überführt werden. Bei der Erzeugung von Zieldokumenten im Word-Format, die in Word übersetzt wurden (z.B. bei **MetaTaxis**, **Wordfast** und **Translator's Workbench** mit Word als Editor), werden bei diesem Prozess

die als verborgener Text enthaltenen AS-Segmente aus dem Dokument gelöscht, weshalb man hier auch von Clean-up spricht.

Von den übrigen Editoren aus müssen die betreffenden Zieldtexte durch eine Exportfunktion erzeugt werden. Bei **across** wird hierzu der so genannte Check-Out Wizard aufgerufen (Abb. 45), bei **SDLX** ist dies eine Funktion des Project Wizard. Vom **TRADOS** TagEditor aus wird das zu exportierende Dokument mit der Speicheroption „Zieldtext speichern unter...“ abgespeichert. Bei den übrigen Programmen werden die Zieldtexte mit einer Exportfunktion erzeugt (vgl. **Transit** in Abb. 46).

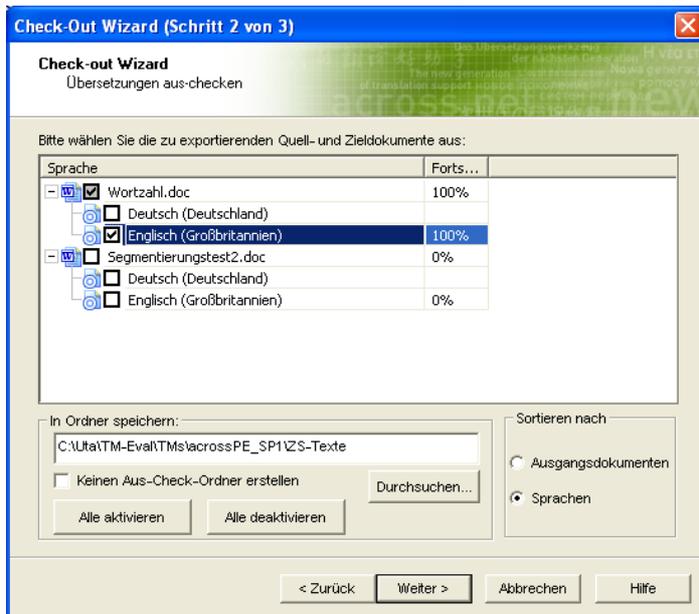


Abb. 45: Erzeugung eines Zieldokuments mit dem Check-Out Wizard in **across**.

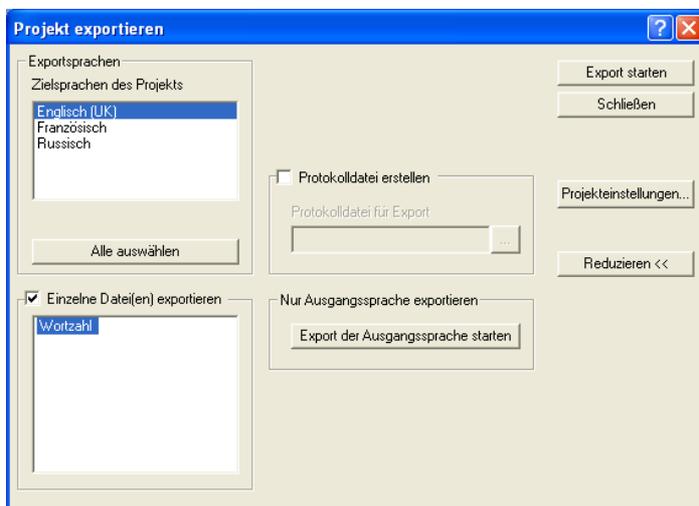


Abb. 46: Exportdialog bei **Transit**.

Pflege der TM-Daten

Außer bei **Transit**, wo das jeweilige Referenzmaterial in Form von Referenzdateien vorliegt, legen die TM-Systeme die Segmentpaare als Datenbank ab. Um in einer solchen Datenbank nachträglich Korrekturen oder Änderungen vornehmen zu können, bieten die Systeme eine Funktion zur Datenbankpflege an, mit Hilfe derer globale Ersetzungen, etwa von Produktbezeichnungen, oder Formatänderungen vorgenommen werden können oder aber unter Verwendung eines Filters Auszüge aus einem TM erstellt werden können. Auch

Segmente lassen sich in dieser Umgebung nachträglich zusammenführen oder teilen. Bei **SDLX** übernimmt diese Funktionen das Modul **SDL Maintain** (Abb. 47), bei **Déjà Vu** wird hierzu im Datei-Menü das zu bearbeitende TM geöffnet (Abb. 48). Globale Änderungen können hier nur mit SQL-Befehlen durchgeführt werden. Da **crossTank**, das TM von **across**, die Segmentpaare aller Übersetzungsprojekte enthält, lassen sich nach Auswahl der Sprachrichtung Segmente, die ein bestimmtes Suchkriterium erfüllen, jeweils in einer Ansicht auflisten (Abb. 49). Von **crossTank** aus können auch nach unterschiedlichen Filterkriterien verschiedene TMX-Exportdateien erzeugt werden. In **MetaTaxis** lassen sich Änderungen an den gespeicherten Segmenten über die Option „Datenbankverwaltung“ vornehmen. **Wordfast** bietet Funktionen zur Datenbankpflege über die Option „EditMemory“ des Wordfast-Menüs (Abb. 50). Per Doppelklick lassen sich die aufgelisteten Segmentpaare in einem Editierfenster samt der zusammen mit einem Attribut abgespeicherten Attribute bearbeiten. Globale Änderungen sind über diese Editierfunktion nicht möglich.

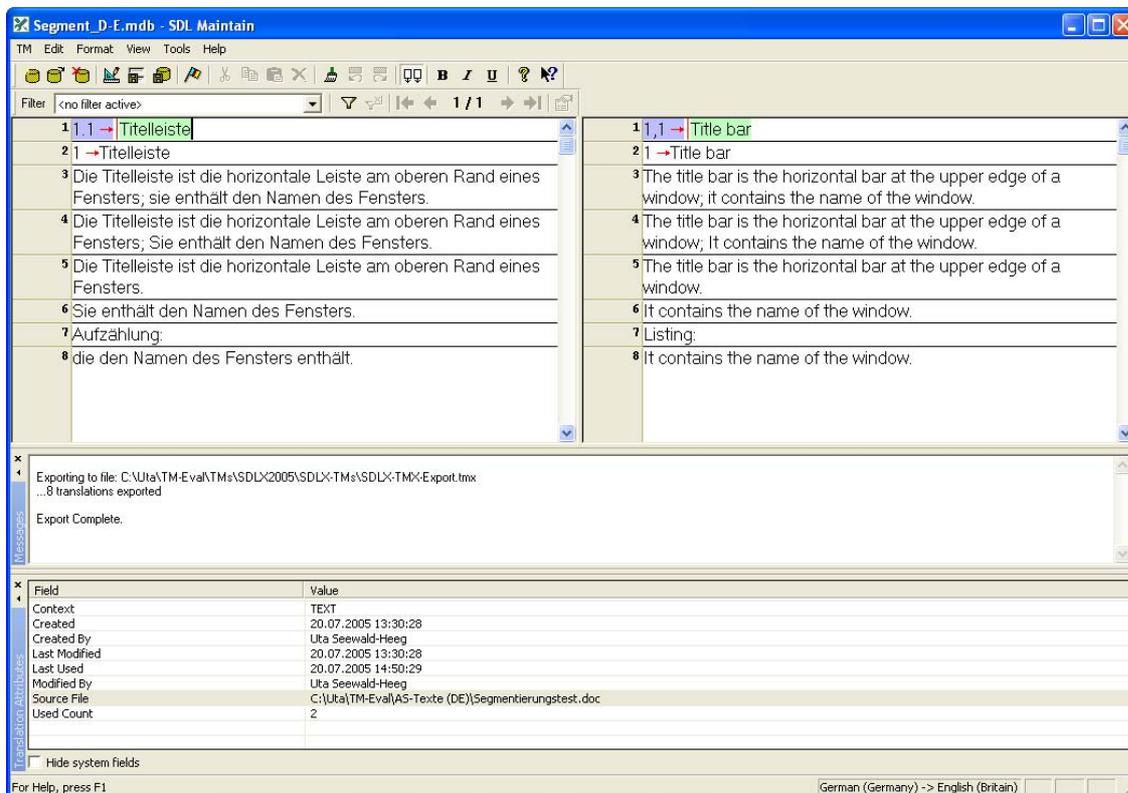


Abb. 47: TM-Datenbankpflege in **SDLX** mit dem Modul **SDL Maintain**.

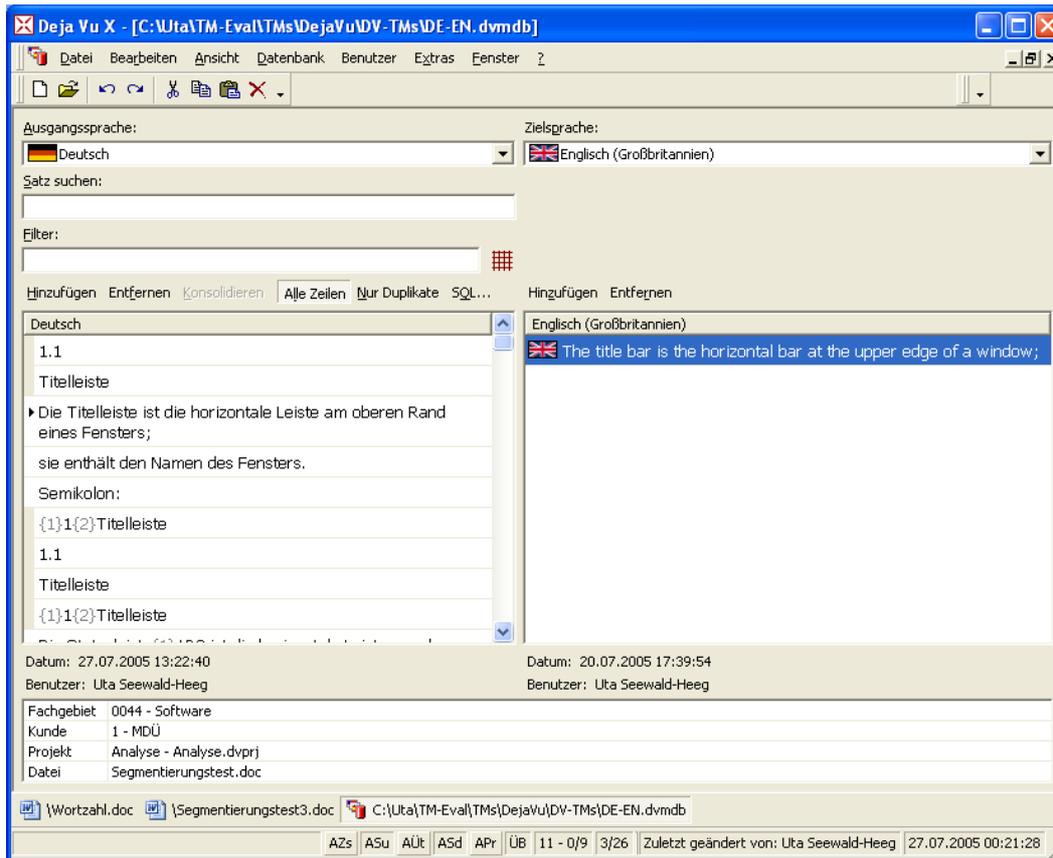


Abb. 48: TM-Datenbankpflege in **Déjà Vu**.

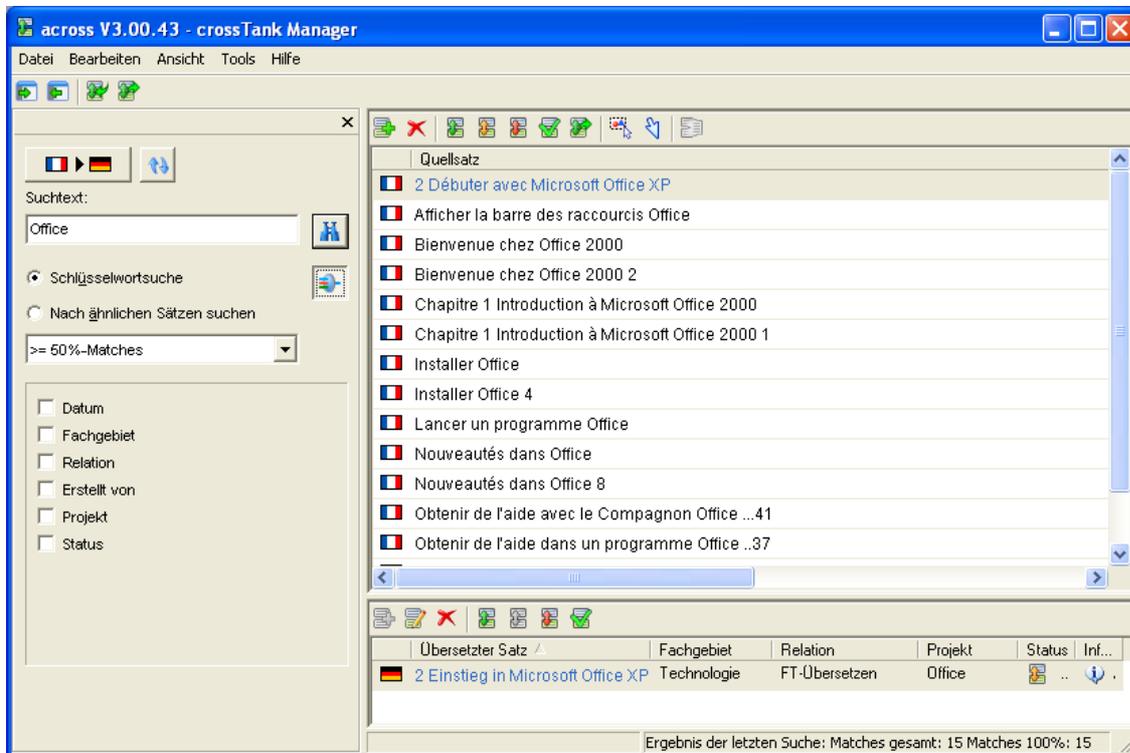


Abb. 49: TM-Datenbankpflege in **across** mit dem crossTank Manager.

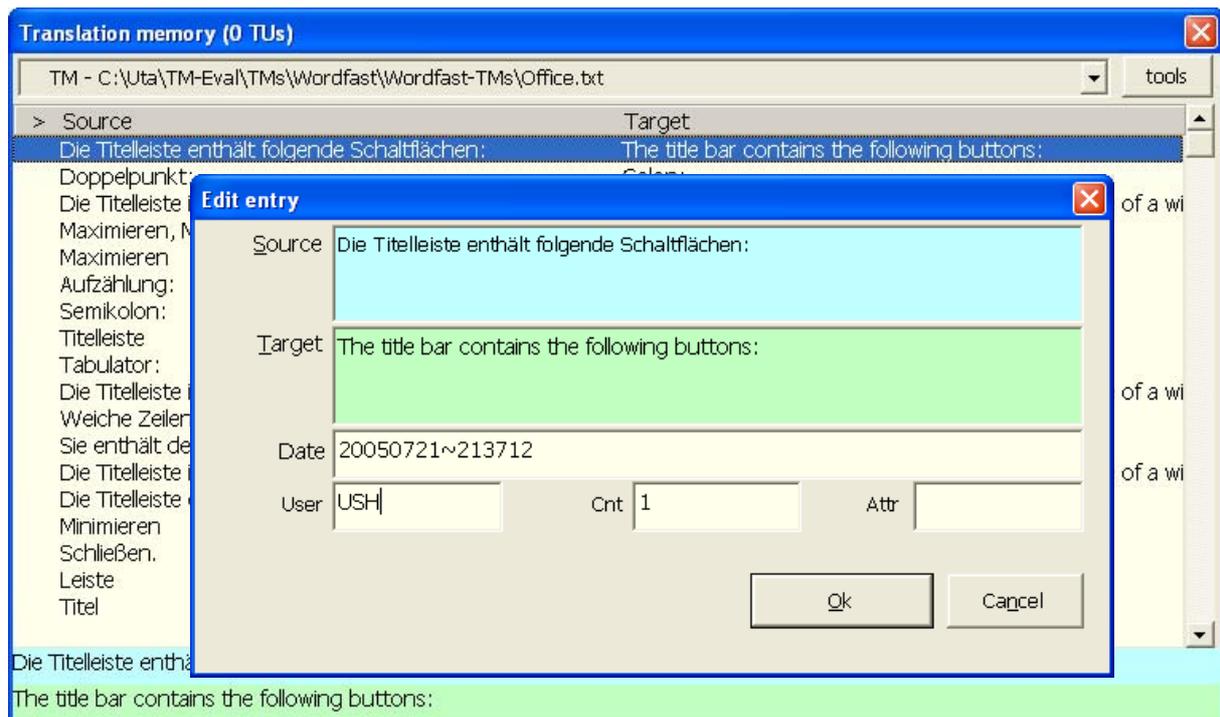


Abb. 50: TM-Datenbankpflege in **Wordfast**.

Einarbeitungsaufwand

Um die gesamte Palette der Funktionen eines TM-Systems nutzen und sinnvoll einsetzen zu können, bedarf es – unabhängig vom Hersteller – stets einer Einarbeitungszeit. Diese umfasst neben der Einarbeitung in die verschiedenen Module und Arbeitsschritte mit einem TM-System zum Teil auch den Erwerb gewisser technischer Grundlagenkenntnisse. So verlangt die Bearbeitung von getaggten Dateiformaten wie HTML in zahlreichen Fällen zumindest grundlegende Kenntnisse in der Interpretation von Tags, so dass unter Umständen erforderliche Anpassungen von Hyperlinks in HTML-Dateien vom Übersetzungseditor aus auch korrekt vorgenommen werden können.

Die als Add-Ins von Word vorliegenden Systeme **MetaTaxis** und **Wordfast**, die bei den Anschaffungskosten bei einem Bruchteil dessen liegen, was man für die Anschaffung eines der anderen Systeme investieren muss – dies gilt insbesondere für **MetaTaxis** –, sind insbesondere für diejenigen interessant, die vor allem Word-Dateien und andere Office-Dokumente übersetzen und keine besonderen Ansprüche an die Terminologieverwaltung stellen, also in erster Linie mit Wortlisten arbeiten. Es ist überraschend, wie viele Funktionalitäten **MetaTaxis** und **Wordfast** innerhalb von Word bieten. Da aber über die Funktionen hinaus, die nicht über die Symbolleiste erreichbar sind, zahlreiche Funktionen über Unteroptionen der Optionen des **MetaTaxis**- bzw. **Wordfast**-Menüs selektiert werden müssen, ist für ein effizientes Arbeiten eine gewisse Virtuosität auf der Tastatur erforderlich, um die betreffenden Funktionen über Tastenkombinationen aufrufen zu können. Darüber hinaus ist für ein reibungsloses Funktionieren und eine volle Ausnutzung dieser beiden Systeme eine intensive Vertrautheit mit dem Textverarbeitungsprogramm Word angebracht.

Auch die anderen hier betrachteten Systeme erfordern eine intensive Einarbeitung. Auch für sie gilt letztendlich, dass etliche Systemfunktionen schneller aufgerufen werden können, wenn man die der Funktion zugeordnete Tastenkombination beherrscht und nicht jedes Mal mit

dem Cursor zwischen Editierfeld und Symbolleiste wechseln muss. Dies gilt insbesondere auch für das Übernehmen von Termini aus der Terminologiekomponente.

Nicht nur für bislang ungeübte Nutzer von TM-Technologie bieten Systeme, welche die in den verschiedenen Schritten eines Übersetzungsprojektes erforderlichen Programmmodule in einer Auswahl übersichtlich darstellen, wie etwa **across** und **SDLX**, einen einfacheren Zugang zu bestimmten Funktionen des Systems.

Eine Erleichterung im Umgang mit den Systemen stellt für deutsche Muttersprachler auch eine deutsche Benutzeroberfläche dar, wie sie außer von **Wordfast** von allen hier vorgestellten Produkten zur Verfügung gestellt wird, im Fall der neuen **SDLX**-Version aber zum Testzeitpunkt noch nicht vorlag.

Tutorials und Dokumentationen

In jedem Fall sind Tutorials zur Demonstration und Einübung bestimmter Arbeitsabläufe und Funktionsaufrufe äußerst instruktiv, wenn es darum geht, das Arbeiten mit einem neuen System zu erlernen. **STAR** bietet hier ein Computer-Based-Training-Programm für Transit Satellite und **across**, **SDL** und **TRADOS** stellen Flash-Dateien mit animierten Inhalten als Tutorial zur Verfügung. **across** bietet darüber hinaus individuelle Systemvorführungen per Internet und Telefon (Live-Demos) an.

Die in der Regel als PDF-Datei vorliegenden Dokumentationen sind inzwischen zu einem Umfang angewachsen, dass es vielfach empfehlenswert ist, zur Einarbeitung mit einem System zunächst die häufig ebenfalls verfügbare Kurzeinführung („Getting Started“) zu lesen und die verschiedentlich nicht weniger als 600 bis 700 Seiten umfassenden vollständigen Dokumentationen online zu konsultieren. Hier erleichtern Hersteller die Einarbeitung zusätzlich, wenn sie zu Systemen, die mit einer deutschen Benutzeroberfläche ausgestattet sind (**across**, **Déjà Vu**, **MetaTaxis**, **TRADOS**, **Transit**, in Kürze wohl auch **SDLX**), auch eine Dokumentation in deutscher Sprache anbieten (**across**, **Transit**, demnächst auch **MetaTaxis**).

Kriterien für den Kauf eines TM-Systems

Wie bereits erwähnt und aus der Übersichtstabelle ersichtlich ist, bieten die Hersteller unterschiedliche Versionen ihrer TM-Systeme an. Soweit nichts Anderslautendes vermerkt ist, beziehen sich die hier aufgeführten Funktionen der Systeme auf Einzelplatzversionen.

Vor dem Erwerb eines TM sollte man unbedingt einen Kriterienkatalog erstellen, in dem alle Anforderungen aufgeführt werden, die ein TM für den eigenen Einsatz erfüllen muss. Hier spielt die Unterstützung von Dokumentformaten sowie der benötigten Sprachen und Zeichensätze eine entscheidende Rolle, die der Berücksichtigung der zur Verfügung stehenden finanziellen Mittel vorangehen muss. Sofern vorhandene Terminologiebestände genutzt werden sollen, ist zu klären, ob das betreffende TM das vorhandene Format unterstützt und die Informationen verlustfrei importiert werden können. Auch die Möglichkeit des Datenaustauschs mit anderen Systemen kann von Bedeutung sein, wenn man mit größeren Agenturen oder mit anderen freiberuflich tätigen Kollegen zusammenarbeitet.

Ein so in die engere Wahl gezogenes System sollte anschließend einem eigenen praktischen Test unterzogen werden. Hierzu eignen sich die Evaluierungsversionen, die auf den Websites der Hersteller in der Regel als Download zur Verfügung gestellt werden. Insbesondere bei der Arbeit mit Arabisch oder asiatischen Sprachen empfiehlt sich ein eingehender Test der Sprachenunterstützung durch die Systeme. Hilfreich ist an dieser Stelle, wenn – wie im

vordefinierten Tutorial-Projekt von **Transit** – Sprachen wie Arabisch oder Chinesisch schon vorgesehen werden. Wie bereits erwähnt, bieten einige Hersteller auch online frei verfügbare Schulungsunterlagen (Tutorials) an, die den Einstieg in die Arbeit mit einem TM-System erleichtern, insbesondere wenn man sich zum ersten Mal mit dieser Technologie auseinandersetzt.

Bei Fragen im Vorfeld einer Anschaffung oder beim späteren Umgang und bei Problemen mit einem System kann es unter Umständen auch hilfreich sein, die Diskussionsforen zur computerunterstützten Übersetzung (siehe Quellen) oder die Mailinglisten, die zu verschiedenen TM-Produkten existieren (siehe Websites der Hersteller), zu konsultieren. Kolleginnen und Kollegen haben sich häufig bereits mit denselben Problemen auseinandergesetzt und hier Antwort auf Fragen bekommen, mit denen man sich gerade selbst beschäftigt.

[Autorenkasten: Uta Seewald-Heeg]

[Kasten]

Quellen

transline tecNews. Online im Internet: <http://www.transline.de/transline-tecNews/sdl-trados-kaufen-uebernahme> [13.07.2005]

Mailinglisten für Technische Redakteure und Übersetzer: <http://www.techwriter.de/thema/u-cat.htm> [22.07.2005]

CATMT, Diskussionsliste zur computerunterstützten und maschinellen Übersetzung: <http://groups.yahoo.com/group/catmt/> [29.07.2005]

Glossar

Algorithmus: Verarbeitungsvorschrift, die programmiert und maschinell ausgeführt werden kann.

Alignment: Prozess der Zuordnung quell- und zielsprachiger Übersetzungseinheiten zum Zweck des Aufbaus eines Translation-Memory-Systems.

AS: Ausgangssprache.

AS-Segment: ausgangssprachiges Segment oder auch Ausgangstextsegment.

Ausgangstextsegment: zu übersetzende segmentierte Übersetzungseinheit des ausgangssprachigen Dokuments.

Austauschformat: Datenformat, mit dem der Export und Import von Daten zwischen verschiedenen Anwendungen ohne nennenswerte Verluste ermöglicht wird.

Computer Based Training: Computerlernprogramm, meist in Form einer multimedialen Präsentation von Lernstoffen.

Dongle: Kopierschutzstecker, der dazu dient, Software vor unautorisierter Vervielfältigung zu schützen.

Exact-Match: Vollständige Entsprechung eines zu übersetzenden Segments des Quelltexts mit einem quellsprachigen Segment im Referenzmaterial; verschiedentlich auch als Full-Match bezeichnet.

Full-Match: siehe Exact-Match.

Fuzzy-Match: Übersetzungsvorschlag, der durch Vergleich eines Ausgangstextsegments mit dem Referenzmaterial angeboten wird und Unterschiede zum Referenzmaterial aufweist, die vom Übersetzer gegebenenfalls angepasst werden müssen.

Match: Übereinstimmung zweier Übersetzungseinheiten.

Match-Wert: Mathematisch errechneter Wert, der die Ähnlichkeit zweier Segmente, und zwar des zu übersetzenden Segments und eines im TM abgelegten Segments, angibt.

Referenzmaterial: Im Translation Memory gespeicherte Segmentpaare, mit denen die zu übersetzenden Texteinheiten des Ausgangstextes verglichen werden.

Segment: Übersetzungseinheit.

Segmentierung: Zerlegung eines Dokuments in Übersetzungseinheiten.

SRX (Segmentation Rule Exchange): Von der Arbeitsgruppe OSCAR der Localization Industry Standards Association (LISA) erarbeitetes Format zur standardisierten Beschreibung von Segmentierungsregeln.

Tag: Auszeichnungselement, das Informationen zum Layout oder zur inhaltlichen Struktur eines Dokuments angibt. Tags werden in der Regel in spitzen Klammern dargestellt.

TBX (TermBase Exchange Format): Von der Arbeitsgruppe OSCAR der Localization Industry Standards Association (LISA) erarbeiteter Standard auf der Grundlage von XML zum Austausch terminologischer Daten.

Term: Aus dem Englischen entlehnte Bezeichnung für *Terminus*.

TMX (Translation Memory eXchange Format): Von der Arbeitsgruppe OSCAR der Localization Industry Standards Association (LISA) entwickelter Standard auf der Grundlage von XML zum Austausch von Translation Memories.

Übersetzungsspeicher: Translation Memory; zum Teil auch als Satzarchiv bezeichnet.

Vorübersetzung: Abgleich eines zu übersetzenden Dokuments mit dem Referenzmaterial und automatische Ersetzung aller 100%-Matches mit den jeweiligen zielsprachigen Segmenten. Durch eine Vorübersetzung entsteht ein Dokument, das ausgangs- und zielsprachige Textpassagen enthält.

XML (Extensible Markup Language): Metasprache zur Beschreibung strukturierter Dokumente. XML ist die Basis einer Vielzahl von Darstellungs- und Austauschformaten wie TMX, TBX oder XHTML, einer Varietät der Standard-Auszeichnungssprache HTML für Dokumente des World Wide Web.

Zieltextsegment: Übersetzung des Ausgangstextsegments in eine vorgegebene Zielsprache.

ZS: Zielsprache.

ZS-Segment: Zielsprachiges Segment oder auch Zieltextsegment.