

Silvia Hansen-Schirra

Nutzbarkeit von Sprachtechnologien für die Translation

The Usability of Language Technologies for Translation

This paper describes the integration of language technologies for translation and interpreting. Starting from the perspective of professional translators and interpreters, applications developed in computational linguistics are adapted and their usability is tested. For instance, corpus linguistics is used to extract typical translation shifts and to quantify contrastive differences in the source and target language. In addition, corpus tools can be used to build up web-based ad-hoc corpora for the extraction of terminology and corporate language. Furthermore, the use of text summarization as means of knowledge acquisition is discussed. For the optimization of the interpreting process, voice recognition is used to prompt figures and named entities. Finally, eye-tracking in connection with corpus linguistics is exploited to identify translation problems and to prompt translation candidates. The exploitation of language technologies for our discipline paves the way for interactive computer-assisted translation and interpreting workflows.

1 Einleitung

Die Integration von Sprachtechnologien in die Translation ist noch nicht besonders weit fortgeschritten. Es gibt zwar denjenigen Bereich der Maschinellen Übersetzung (MÜ), der für Laien brauchbare Ergebnisse liefert – als Beispiele sind hier frei verfügbare Systeme wie der Google-Übersetzer (Google Übersetzer o.J.) zu nennen; als Vorreiter für allerneuste, zumeist hybride Entwicklungen soll hier das Euromatrix Plus-Projekt (EuroMatrixPlus o.J.) aufgeführt werden. In der professionellen Übersetzung jedoch findet die MÜ noch kaum Anwendung – selbst große Sprachdienstleister beziehungsweise Sprachabteilungen in internationalen Firmen führen erst erste Machbarkeitsstudien durch (vgl. Gassmann 2009). Dies liegt unter anderem daran, dass die MÜ für die professionelle Translation qualitativ immer noch nicht als ausreichend erachtet wird und der MÜ-Qualitätsbegriff ein anderer ist als in der Humanübersetzung (vgl. Mertin 2006).

Über basale Technologien wie Rechtschreib- und Grammatikprüfung sowie Internetsuchen und Online-Ressourcen hinausgehend (Austermühl 2001), haben Sprachtechnologien in Form von Autorensystemen und Sprachkontrollsystemen zur Umsetzung von kontrollierten Sprachen und *Corporate Languages* Einzug in die technische Redaktion und Übersetzung gehalten (Göpferich 1998). Weiterhin sind sogenannte Translation Workbenches zu nennen, die Translation Memories (d.h. Übersetzungsspeicher), Terminologieverwaltung, Projektmanagement, Alignierungsprogramme, Vor-

übersetzung etc. vereinen.¹ Abschließend ist die Terminologieextraktion zu nennen, die ebenfalls bereits im Rahmen von Translation Memories, zum Beispiel *MultiTerm Extract* (SDL MultiTerm Extract o.J.), in der professionellen Translation eingesetzt wird.

Zu kritisieren ist allerdings, dass die im Übersetzer- und Dolmetscheralltag eingesetzten Translationstechnologien im Verhältnis zum Stand der Forschung in der Computerlinguistik weit hinter den aktuellen Möglichkeiten bleiben. Sie machen beispielsweise keinen Gebrauch von linguistischem und/oder statistischem Wissen, so wie es in der Computerlinguistik längst zum Standard gehört. Als Beispiel sollen die oben erwähnten Translation Memories aufgeführt werden: Entgegen der Vorschläge, die Reinke schon im Jahr 2004 gemacht hat, nutzen Translation Memories nach wie vor keine morphologische Analyse, um die *Fuzzy Match*-Rate zu verbessern. Weiterhin operieren sie nur auf der Satzebene, eine Übersetzungseinheit, die in der Translationswissenschaft als überholt gilt (siehe Alves et al. 2010). Eine Segmentierung auf Teilsatz- oder Phrasenebene, ähnlich wie es in der beispielbasierten MÜ schon lange umgesetzt wird (z.B. Langlais et al. 2005), würde *Precision* und *Recall* bei der Arbeit mit Translation Memories verbessern. Eine Suche nach grammatischen Konstruktionen, zum Beispiel nach typologischen Sprachunterschieden oder Fachsprachmerkmalen, ist wegen der wort- und zeichenkettenbasierten Algorithmen der Translationstools ebenfalls nicht möglich. Der Termextraktion und -verwaltung ist anzulasten, dass der für die Translation dringend notwendige Kontext oftmals gänzlich außer Acht gelassen wird. Die wenigen Technologien, die für den dolmetschspezifischen Einsatz ausgerichtet sind, stellen oftmals nichts anderes als die Verwaltung von Termlisten dar – zum Beispiel Interplex (Interplex Glossary Software o.J.), Terminus (Terminus o.J.), LookUp (LookUp o.J.). Die Bereiche der Wissensakquisition und -verwaltung stellen eine Forschungslücke in der Translationswissenschaft dar.

Der vorliegende Aufsatz möchte daher den Einsatz von Sprachtechnologien für die Translation testen. Im zweiten Kapitel werden daher Usability Tests für korpuslinguistische Anwendungen (Abschnitt 2.1), Textzusammenfassungssysteme (Abschnitt 2.2), Spracherkennung (Abschnitt 2.3) sowie für die Termextraktion (Abschnitt 2.4) präsentiert. Hierbei wird immer die Perspektive des professionellen Übersetzens und Dolmetschens beleuchtet. Beim Übersetzen steht die Fachübersetzung im Vordergrund. Beim Dolmetschen kann vom Konferenzdolmetschen ausgegangen werden, *Community Interpreting* (z.B. im medizinischen Bereich) oder Gerichtsdolmetschen sind ebenso denkbar. Das Schlusskapitel diskutiert die sich aus den Fallstudien ergebenden Perspektiven sowie die Forschungsdesiderata, die bei deren Umsetzung noch überwunden werden müssen. Weiterhin wird eine visionäre Fallstudie diskutiert, die die Translation nicht nur mit Sprachtechnologien, sondern auch mit Blickbewegungssteuerung zusammenbringt.

¹ Eine Übersicht über aktuell verfügbare Translation Workbenches und Translation Memories ist in Seewald-Heeg (2005) zu finden.

2 Sprachtechnologien für die professionelle Translation

Da die Integration von Sprachtechnologien in die Translationspraxis noch einige Desiderata an die Forschung stellt, werden im Folgenden einige Pilotstudien und Usability Tests vorgestellt, die den Einsatz von computerlinguistischen Methoden im Übersetzeralltag operationalisieren. Die Pilotstudien beziehungsweise Implementierungen fanden alle im Rahmen von Abschluss- und Doktorarbeiten sowie Projekten am Lehrstuhl für Englische Sprach- und Übersetzungswissenschaft am Fachbereich Translations-, Sprach- und Kulturwissenschaft der Universität Mainz in Germersheim statt. Die Daten und Methoden werden hier zusammengetragen beziehungsweise reinterpretiert und im Sinne der Integration von Sprachtechnologien in die Translation bewertet.

2.1 Korpuslinguistik

Auf der Basis von Übersetzungskorpora lassen sich typische Übersetzungsentscheidungen in Abhängigkeit von der Übersetzungsrichtung und der Textsorte modellieren. Diese können einerseits als interaktive Grammatiken für Humanübersetzer genutzt werden (Hansen-Schirra 2008) oder in Form von linguistischen Regeln die Maschinelle Übersetzung optimieren (Čulo 2010). So lässt sich am Beispiel des englisch-deutschen CroCo-Korpus (Hansen-Schirra et al. 2012) empirisch nachweisen, welche Kategorienwechsel – von Vinay und Darbelnet (1958/1995) als Transposition bezeichnet – für die jeweiligen Übersetzungsrichtungen am häufigsten vorkommen. Abbildung 1 zeigt, dass in 42,95 % aller Shifts in der Übersetzungsrichtung Englisch-Deutsch englische Verben zu Adverbien, Adjektiven oder Nomen im Deutschen werden. Die letzten beiden Arten von Transpositionen sind typologisch bedingt, da das Deutsche inhaltsorientierter und somit oft auch nominaler ist als das Englische (vgl. House 1997). Das folgende Beispiel belegt einen Verb-Nomen-Shift in Instruktionstexten [INSTR] vom Englischen [EO] ins Deutsche [GTrans].

- (a) When replacing the lamp be sure to turn off power and unplug the power cord. (EO_INSTR)
- (b) Schalten Sie vor dem Austausch der Lampe das Gerät aus und entfernen Sie das Netzkabel. (GTrans_INSTR)

Aus den gleichen typologischen Gründen findet man die oben beschriebenen Arten von Transpositionen seltener in deutsch-englischen Übersetzungen, wohingegen die Nomen-Verb- und Nomen-Adjektiv-Transpositionen erheblich zunehmen (sie belaufen sich insgesamt auf 39,69 %, siehe Abb. 1). Die richtige Frequenz von Übersetzungsshifts trägt dazu bei, den Konventionen der Zielsprache zu entsprechen und die Übersetzungen der Zielsprache anzupassen.

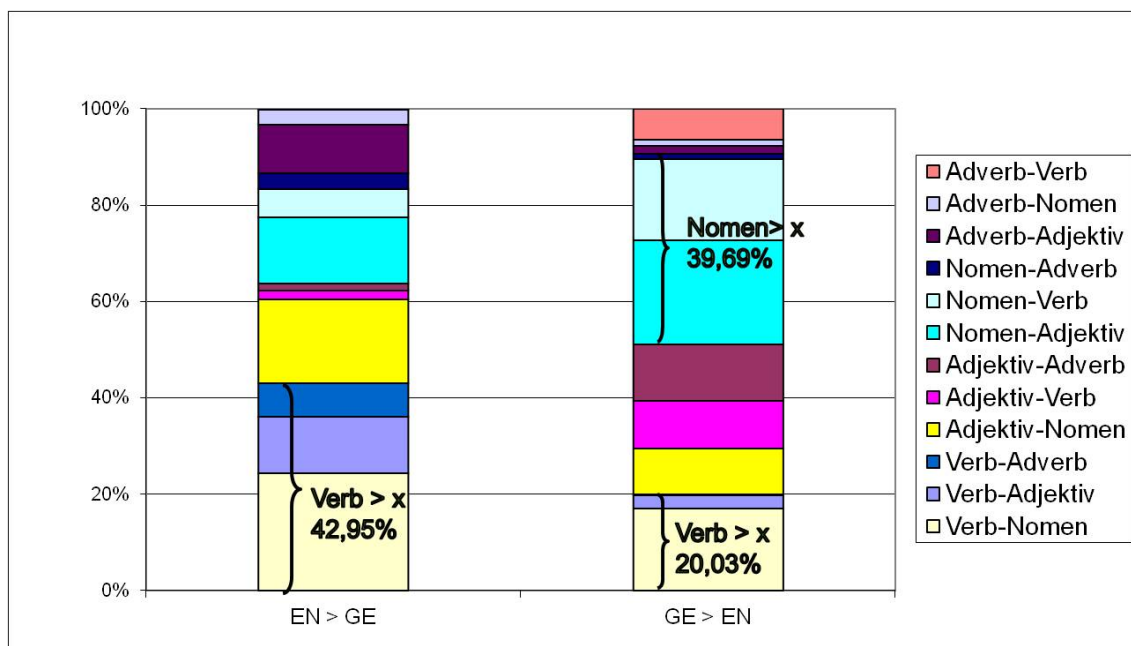


Abb. 1: Kategorienwechsel im CroCo-Korpus (Hansen-Schirra et al. 2012)

Weiterhin lässt sich auf der Basis von Übersetzungskorpora zeigen, welche Textsorten am häufigsten auf semantische Perspektivenwechsel, die in der Übersetzungswissenschaft als Modulation bezeichnet werden (Vinay/Darbelnet 1958/1995), zurückgreifen. Abbildung 2 zeigt auf der Basis dreier unterschiedlicher Textsorten des CroCo-Korpus (politische Essays [ESSAY], fiktionale Texte [FICTION] und Instruktionstexte [INSTR]), wie häufig Wörter des Ausgangs- und Zieltexes miteinander aligniert sind, obwohl sie unterschiedlichen grammatischen Funktionen zugeordnet sind.² Auf diese Art und Weise können beispielsweise Subjekt-Objekt-Wechsel, aber auch alle anderen Funktionswechsel automatisch extrahiert werden. Aus Abbildung 2 wird ersichtlich, dass sowohl die Textsorte als auch die Übersetzungsrichtung einen großen Einfluss auf die Anzahl und Art der Shifts haben.

² Die Frequenzen sind hier in prozentualen Häufigkeiten zur Gesamtanzahl der Shifts pro Textsorte und Übersetzungsrichtung angegeben.

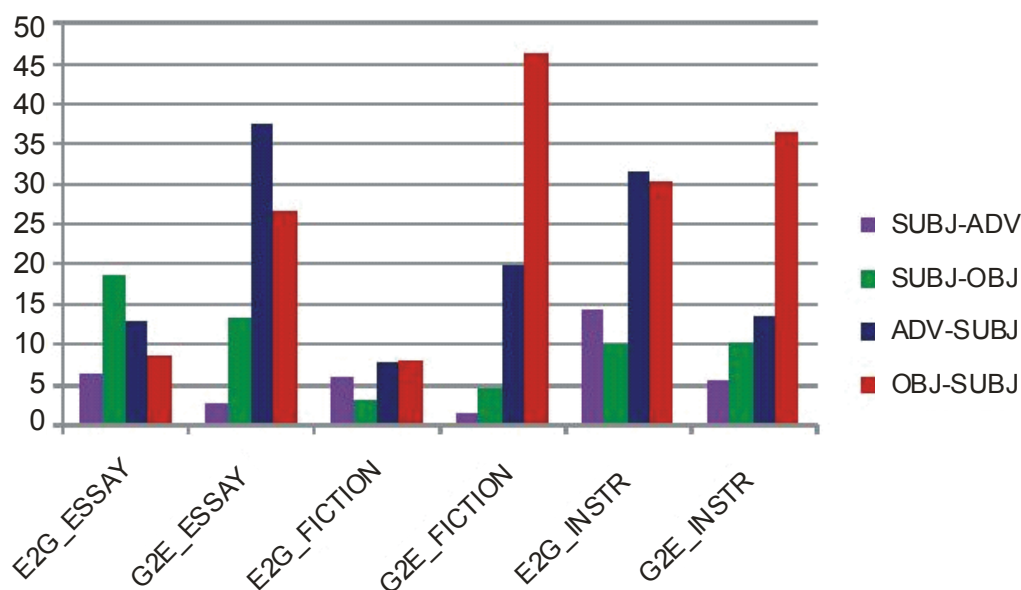


Abb. 2: Funktionswechsel im CroCo-Korpus

Diese Divergenzen lassen sich wiederum mit kontrastiven Unterschieden erklären. So sind Objekt-Subjekt-Shifts für die Übersetzungsrichtung Deutsch-Englisch häufiger als in die umgekehrte Übersetzungsrichtung. Dies liegt an der relativ freien Satzstellung im Deutschen, die beispielsweise auch Objekte in thematischer Position erlaubt und die in der englischen Übersetzung in die fixe englische SVO-Satzstruktur transferiert werden muss (vgl. Hawkins 1986). Das folgende Beispiel belegt diese Erklärung:

- (a) Die Frauen hat das nicht gerade zimperlich gemacht. (GO_FICTION)
- (b) The women weren't exactly prudes. (ETrans_FICTION)

Die hier diskutierten Zahlen verdeutlichen, dass kontrastive Unterschiede je nach Textsorte durchaus stärker oder schwächer ausgeprägt sind. Die hohe Anzahl an Objekt-Subjekt-Shifts in deutsch-englischen Übersetzungen fiktionaler Texte spricht für eine kreative, teilweise markierte Belegung der thematischen Position im Deutschen in dieser Textsorte.

Das Wissen um diese kontrastiven Textsortenkonventionen ermöglicht es Übersetzern, nicht nur korrekte, sondern auch idiomatisch passende Übersetzungen zu produzieren und somit einen "kulturellen Filter" anzuwenden (House 1997). Dies betrifft sowohl die Auswahl an linguistischen Merkmalen als auch deren Häufigkeitsverteilung. Übersetzungskorpora können daher eine wichtige Ressource sein, um zu belegen, welche Shifts typisch sind, um Übersetzungen der Zielsprache und -kultur anzupassen.

2.2 Textzusammenfassungssysteme

Textzusammenfassungssysteme haben in vielerlei Hinsicht Anwendungspotenzial. So werden sie beispielsweise zur Zusammenfassung von Nachrichten genutzt oder um Suchmaschinentreffer zu filtern (vgl. Marcu 2000). Sie haben bisher jedoch noch keinen Einzug in die Translationspraxis gehalten. Dabei wären sie ideale Hilfen zur Wissensakquisition für Dolmetscher und Fachübersetzer, die sich in fremde Themengebiete einarbeiten müssen. Oftmals muss diese Einarbeitung innerhalb kürzester Zeit erfolgen, so dass kaum die Zeit bleibt, um sich das notwendige Wissen durch das Lesen von Fachartikeln anzueignen. Hierbei ist die für die Translation notwendige Wissensakquisition nicht mit reiner Terminologiarbeit gleichzusetzen. Die Erarbeitung von Wortfeldern, Kollokationen, einer adäquaten Phraseologie etc. ist ebenfalls wichtig für die Übersetzung, kann jedoch das fehlende Fach- bzw. Hintergrundwissen nicht ersetzen. Das Textverständnis spielt für die Translation eine zentrale Rolle (vgl. Göpferich 2008). Nur wenn der Übersetzer und Dolmetscher den Ausgangstext und den damit verbundenen situativen Kontext versteht, kann er die Übersetzung adäquat in die Zielsprache transferieren und den Erwartungen der Rezipienten anpassen (vgl. House 2009).

Nitzke (2011) hat daher in einer translationsbezogenen Studie die Qualität verschiedener Textzusammenfassungssysteme verglichen. Als Grundlage hierfür hat sie einen technischen Fachtext in englischer Sprache, der 454 Wörter verteilt auf 22 Sätze beinhaltet, von 40 Übersetzungsstudenten, die einen einschlägigen Masterstudiengang belegten und die als potenzielle Nutzer dieser Systeme angesehen werden konnten, zusammenfassen lassen. Bei dieser Zusammenfassung sollte der Text auf 20 % seines Umfangs reduziert werden. Hierfür sollten die Studierenden die 5 Sätze auswählen, die ihrer Meinung nach für Informationsgehalt und -übermittlung des Textes am wichtigsten waren. Das heißt, es waren die Sätze betroffen, die in Vorbereitung auf eine Übersetzung am besten den Inhalt des Textes vermittelt hätten. Bei der Bewertung konnten die Studierenden zwischen den Werten "sehr wichtig", "wichtig" und "unwichtig" wählen. Auf der Basis der Einschätzungen wurde ein Gold-Standard ermittelt, der die Sätze mit den fünf besten Durchschnittswerten enthielt.

Getestet wurden die Systeme SweSum (SweSum o.J.), GreatSummary (GreatSummary o.J.), Tools4Noobs (Tools4Noobs o.J.), AutoSummarize (AutoSummarize o.J.) und MEAD (MEAD o.J.). Die Systeme sollten den gleichen technischen Fachtext auf eine Kompressionsrate von 20 % Prozent reduzieren. Im Zuge einer Reinterpretation von Nitzkes (2011) Ergebnissen, wurden die folgenden Werte ermittelt: Es wurde ausgezählt, wie viele Sätze aus dem Gold-Standard das jeweilige System wählte (*Recall*). Weiterhin wurde quantifiziert, wie viele Sätze das jeweilige System gewählt hat, die die Studierenden durchschnittlich mit "sehr wichtig" und "wichtig" bewertet haben (*Precision*). Gemäß der studentischen Bewertung konnte abschließend berechnet werden, wie gut die Systeme insgesamt waren (*Qualitatives Ranking*). Hierfür wurde ein Durchschnittswert für jedes System errechnet, der die Anzahl der

ausgewählten wichtigen, sehr wichtigen und unwichtigen Sätze berücksichtigte. Tabelle 1 zeigt die Ergebnisse der Evaluierung.

	Precision	Recall	Qualitatives Ranking
SweSum	50 %	60 %	3. Platz
GreatSummary	60 %	0 %	5. Platz
Tools4Noobs	80 %	40 %	2. Platz
AutoSummarize	67 %	60 %	1. Platz
MEAD	50 %	20 %	4. Platz

Tab. 1: Evaluierung von Textzusammenfassungssystemen

Es wird ersichtlich, dass die Systeme teilweise recht gut abschneiden, was die Präzision angeht (bis zu 80 %). Es gelingt ihnen also, Sätze mit hohem Informationsgehalt auszuwählen. Im Vergleich zum Gold-Standard haben die Systeme höchstens 60 % der besten Sätze ausgewählt. Der Recall ist also verbesserungswürdig, ihm sollte aber auch nicht zu viel Wert beigemessen werden, da er sich tatsächlich nur auf die 5 besten Sätze bezieht und andere Sätze, die vom Informationsgehalt her zwar etwas schlechter, aber auch gut eingeschätzt wurden, vernachlässigt. Das qualitative Ranking spiegelt eine Mischung zwischen Precision und Recall wieder – das System mit dem zweitbesten Precision- und dem besten Recall-Wert schnitt als bestes ab (Auto-Summarize), das System mit der besten Precision und dem zweitbesten Recall als zweitbestes (Tools4Noobs).

Diese Evaluierung zeigt, dass die guten Textzusammenfassungssysteme dem Gold-Standard der Nutzer schon relativ nahe kommen. Daraus lässt sich schließen, dass sie als Vorbereitungshilfe für einen Übersetzungs- und Dolmetschauftrag im Sinne der Wissensakquisition durchaus sinnvoll eingesetzt werden können.

2.3 Spracherkennung

Automatische Spracherkennung ist zwar für das maschinelle Dolmetschen bereits zum Einsatz gekommen (z.B. im Verbmobil-Projekt; Verbmobil o.J.), jedoch für das humane Dolmetschen noch nie in Betracht gezogen worden. Nun ist beim Simultandolmetschen aber die Verarbeitungs- und Aufmerksamkeitskapazität extrem beansprucht, da der Dolmetscher gleichzeitig mehrere Sprachverarbeitungsprozesse aktivieren muss. Gile (2009) unterscheidet in seinem *Effort-Modell* die folgenden Prozesse: *Listening & Analysis*, *Production*, *Memory* sowie *Coordination*. Sobald ein *Effort* mehr beansprucht wird, leiden laut Gile die anderen darunter. So kann beispielsweise durch eine Ausgangsrede mit einem hohen Anteil an Zahlen beziehungsweise Eigennamen das Gedächtnis mehr beansprucht werden, worunter möglicherweise die zielsprachliche Produktion leiden könnte. Bei der Analyse könnte eine hohe Termdichte ein Hemmnis darstellen, was wiederum auf das Gedächtnis und auf die Produktionsfähigkeit Auswirkungen haben könnte. Als eine mögliche Lösungsstrategie führt Kalina (1998) die

Verlängerung der *Décalage*³ an, wobei der Dolmetscher einige Sekunden länger wartet, um mehr sprachliches Input zu erhalten und z.B. das Satzende abzuwarten. Als weiteren verstehensunterstützenden Faktor gibt Gile (2009) den Kabinenpartner an, der typischerweise Dolmetschinput, das Wahrnehmungsprobleme auslösen könnte (z.B. Zahlen, Eigennamen, Terme), notiert. Es gibt aber durchaus Situationen, in denen der Kabinenpartner diese entlastende Funktion nicht wahrnehmen kann (weil er sich vom letzten Einsatz erholen bzw. auf den nächsten vorbereiten muss oder nicht in der Kabine anwesend sein kann), so dass sich der Dolmetscher nicht per se auf diese Hilfe verlassen kann; vgl. Gile 2009).

In einer Pilotstudie testete Richter (2010), inwiefern die Funktion des Kabinenpartners von Spracherkennungssoftware übernommen werden kann. Sie ließ eine deutsche fünfminütige Rede aus einer Bilanzpressekonferenz eines Verbandes von einer deutschen Muttersprachlerin vortragen. Die Rede enthielt insgesamt 49 Zahlen und 10 Eigennamen. Als Spracherkennungssysteme wurden die folgenden verwendet: *DRAGON NaturallySpeaking Premium 11.0 Mobile* (Dragon NaturallySpeaking o.J.) und die Spracherkennung von *Windows 7* (Windows Spracherkennung o.J.). Beide Systeme wurden auf die Rednerin trainiert, so dass ein individualisiertes Sprachmodell vorlag. Beide Systeme gaben die komplette Rede als *Speech-to-Text-Output* aus, so dass bei dieser Pilotstudie noch keine Informationsextraktion betrieben wurde. Die Texte wurden aber im Speziellen auf Zahlen und Eigennamen ausgewertet (siehe Tabelle 2).

	Dragon	Windows 7
Zahlen	93,62 %	85,11 %
Eigennamen	60 %	40 %
Verzögerung	2 Sek.	3 Sek.

Tab. 2: Evaluierung von Spracherkennungssystemen (siehe Richter 2010: 78ff.)

Wie aus der Tabelle ersichtlich wird, schneidet die Dragon-Software sowohl was die Zahlen angeht als auch in Bezug auf die Eigennamen besser ab als die Windows-Spracherkennung. Die Zahlen werden mit einer Trefferquote von 93,62 % erkannt, bei den Eigennamen liegt die Trefferquote nur bei 60 %. Dies deutet daraufhin, dass in Bezug auf die Zahlen die Spracherkennung durchaus als verstehens- und gedächtnisfördernde Methode beim Simultandolmetschen eingesetzt werden kann (vgl. Richter 2010). Bei den Eigennamen wäre sie zum derzeitigen Entwicklungsstand nicht derart ergiebig.

Der nächste logische Evaluierungsschritt wäre, die Terminologie mit in die Erkennung einzubeziehen. Terme können oftmals aus mehrgliedrigen Komposita bestehen, so dass die automatische Erkennung inklusive der automatischen Suche und Anzeige von Übersetzungsäquivalenten eine erhebliche Entlastung für den Dolmetscher

³ Unter *Décalage* versteht man die zeitliche Verschiebung zwischen Dolmetschinput und -output.

darstellen könnte. Weiterhin müsste getestet werden, ob es effizienter ist, das Spracherkennungsoutput einer Informationsextraktion (in Bezug auf Zahlen, Eigennamen und Terme) zu unterziehen oder ob sich ein *Speech Mining* in Kombination mit der Spracherkennung als ergiebiger erweist. Abschließend müsste die Dolmetschspracherkennung für einen realen Einsatz natürlich auch nutzerunabhängig funktionieren, was aber viele mobile (Handy-)Anwendungen ohnehin schon zulassen.

2.4 Terminologieextraktion

Für die Translation erfolgt die Terminologieextraktion bisher lediglich auf der Basis von vorhandenen Korpora, Datenbanken oder Dokumenten, die dem Übersetzer von vorherigen Aufträgen bereits vorliegen oder die ihm der Auftraggeber zur Verfügung stellt. Liegen diese Daten nicht vor, ist keine Termextraktion möglich. Der Ressourcenvielfalt, die das Internet zur Verfügung stellt, wird mit traditionellen Extraktionsalgorithmen kaum Rechnung getragen. Fantinuoli (2011) zeigt, wie bei der Vorbereitung auf einen Dolmetscheinsatz – prinzipiell könnte es aber auch für die Einarbeitung im Rahmen einer Fachübersetzung sein – die Wissensakquisition auf der Basis strukturierter und semistrukturierter Internetdaten erfolgen kann. Auf der Basis von 4 bis 6 *Seeds*, das heißt spezifischen und für das Thema relevanten Termen, erstellt Fantinouli's Workbench *InterpretBank* (InterpretBank o.J.) ein Korpus, dessen Größe der Nutzer wiederum selbst spezifizieren kann. Dass dieser Arbeitsschritt – und somit die Zusammenstellung des Hintergrundmaterials für einen Dolmetsch- oder Übersetzungseinsatz – sehr schnell erfolgen kann, wird in der folgenden Tabelle veranschaulicht.

Korpusgröße (Anzahl der Tokens)	Zeit (Minuten)
114.202	1,22
162.139	2,08
369.238	5,85
792.614	15,94

Tab. 3: Performanz von InterpretBank (siehe Fantinuoli 2011: 138ff.)

Das resultierende Korpus wird bereinigt (d.h. HTML-Tags etc. werden entfernt) und mit Part-of-Speech-Tagging sowie Lemmatisierung versehen. Auf dieser Basis werden Ein-Wort- und Multi-Wort-Terme extrahiert.⁴ Das Korpus und die Termliste werden suchbar gemacht; die Treffer werden in der Workbench angezeigt (siehe Abb. 3).

Über diese Funktionalitäten hinausgehend werden auf der Basis von statistischen Häufigkeiten die Kookkurrenzen und Kollokationen des jeweiligen Suchwortes berechnet und angezeigt. Weiterhin werden die entsprechenden Konkordanzen aus dem generierten Korpus zur Verfügung gestellt. Das Suchwort wird weiterhin mit semi-

⁴ Die genaue Prozedur wird in Fantinuoli (2011: 114ff.) beschrieben.

strukturierten Internetdaten verlinkt: Es werden Wikipedia-Informationen sowie Übersetzungen aus leo.org oder dict.cc angezeigt. Mit Hilfe dieser Informationen kann der Übersetzer beziehungsweise Dolmetscher sich innerhalb kürzester Zeit in die linguistischen Eigenschaften eines Themengebiets einarbeiten. Die Kookkurrenzen und Kollokationen ermöglichen die Erarbeitung von Wortfeldern – durch die Verknüpfung mit Online-Wörterbüchern kann dies auch multilingual erfolgen. Durch die Kombination der Korpusdaten mit den semistrukturierten Internetdaten können die extrahierten Terme mit Wissen verknüpft werden. Auf der Basis der so gewonnenen Definitionen können sich Übersetzer und Dolmetscher schnell und effizient in das Sachgebiet einarbeiten. Workbenches wie die InterpretBank helfen demnach nicht nur bei der Wissensverwaltung und -repräsentation sondern auch bei der Wissensakquisition.

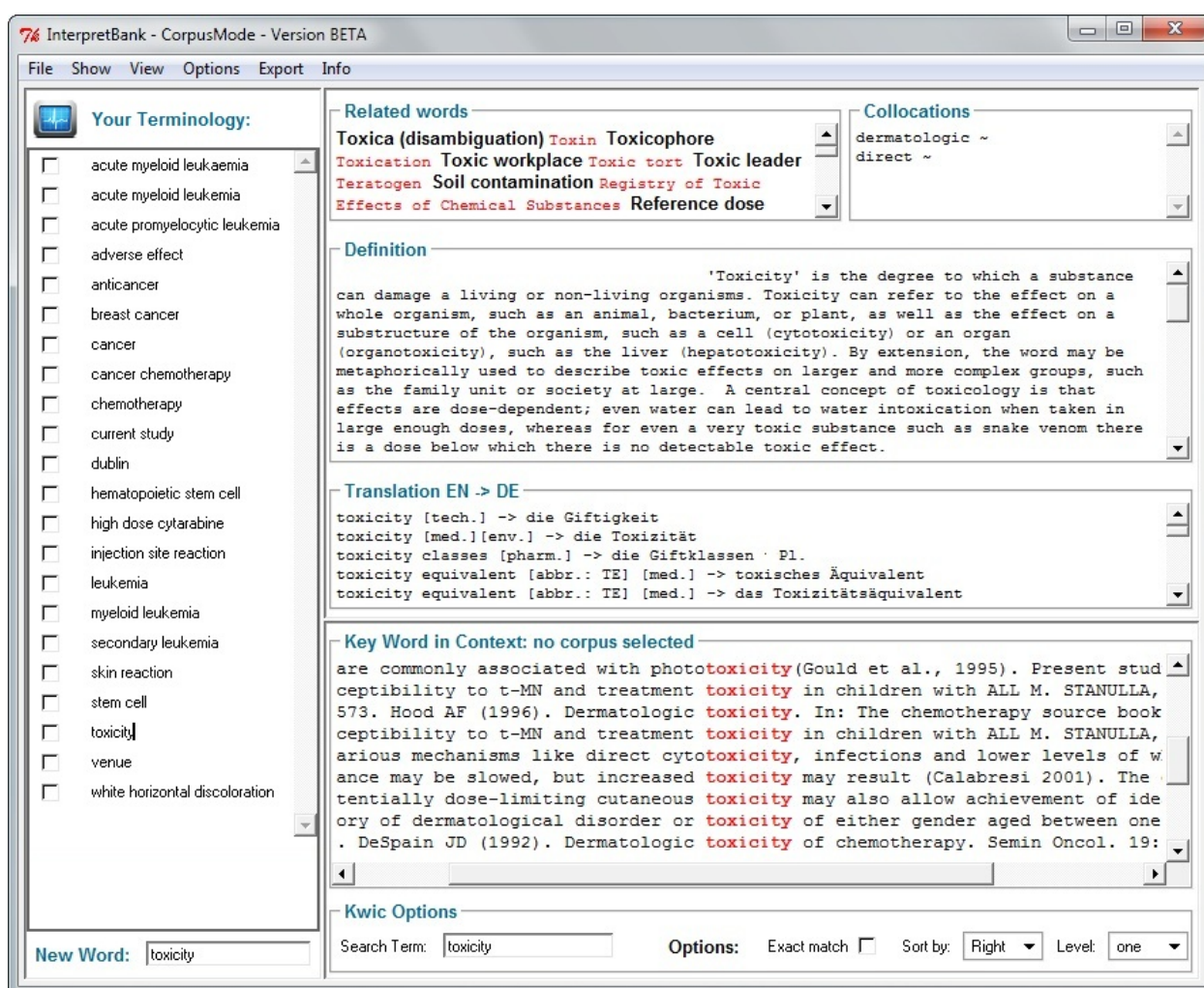


Abb. 3: Korpusbasierte Termextraktion mit InterpretBank

3 Perspektiven und Forschungsdesiderata

Die Fallstudien in Abschnitt 2 haben verdeutlicht, dass Sprachtechnologien die translatorische Tätigkeit effizienter gestalten können. Dies trifft einerseits auf die Geschwindigkeit des Translationsprozesses selbst, ebenso wie auf die der Vorbereitung und der Rechercharbeit, zu, andererseits aber auch auf die Qualität des Translats, wenn es um Konsistenzstiftung sowie Wissensverwaltung und -repräsentation geht.

In der künftigen Forschung geht es mehr noch als bisher darum, die computerlinguistischen Entwicklungen in die bisher auf dem Markt zur Verfügung stehenden Translationstechnologien (v.a. Terminologieverwaltung und Translation Memories) zu integrieren. So arbeiten wir gerade an der Integration der Spracherkennung für Dolmetscher in die Workbench *InterpretBank*. Auf diese Weise kann der Dolmetscher bei der Vorbereitung auf korpuslinguistische Methoden, Termextraktion und die Verknüpfung zu semistrukturierten Daten zurückgreifen. Zudem kann er während des Einsatzes nach Termen, Kollokationen und Übersetzungskandidaten suchen sowie die Spracherkennung von Eigennamen, Zahlen und Termen zur Memorisierung des Ausgangstextes nutzen.

Weiterhin erscheint es vielversprechend, die Sprachtechnologien aus dem Bereich der Korpuslinguistik in Translation Memories einzuspeisen. So können die Funktionalitäten (Vorübersetzung, Fuzzy Matches, Alignment etc.) und die Effizienz bereits existierender Systeme genutzt und um linguistisches Wissen erweitert werden. Ähnlich wie Reinke (2004) es mit der Einbindung von morphologischen Analysen vorschlägt, gehen wir einen Schritt weiter und planen die Integration von Baumbankstrukturen in Translation Memories (siehe Vela/Hansen-Schirra 2006). Dadurch können Translation Memories Fuzzy-Suchen auch auf grammatischer Ebene vollziehen, was bei der Bewältigung von kontrastiven Lücken helfen könnte. Die Übersetzer könnten sich beispielsweise für Spaltsätze und Anhebungen in der Übersetzungsrichtung Englisch-Deutsch (siehe Hansen-Schirra 2008, 2012) sowie für Pronominaladverbien in der Übersetzungsrichtung Deutsch-Englisch (siehe Schreiber demn.) typische Übersetzungsstrategien anzeigen lassen, die Anregungen für das eigentliche Übersetzungsproblem liefern können. Annotierte Übersetzungskorpora können durch die Such- und Speicherfunktionalitäten von Translation Memories demnach als interaktive Transfergrammatik genutzt werden.

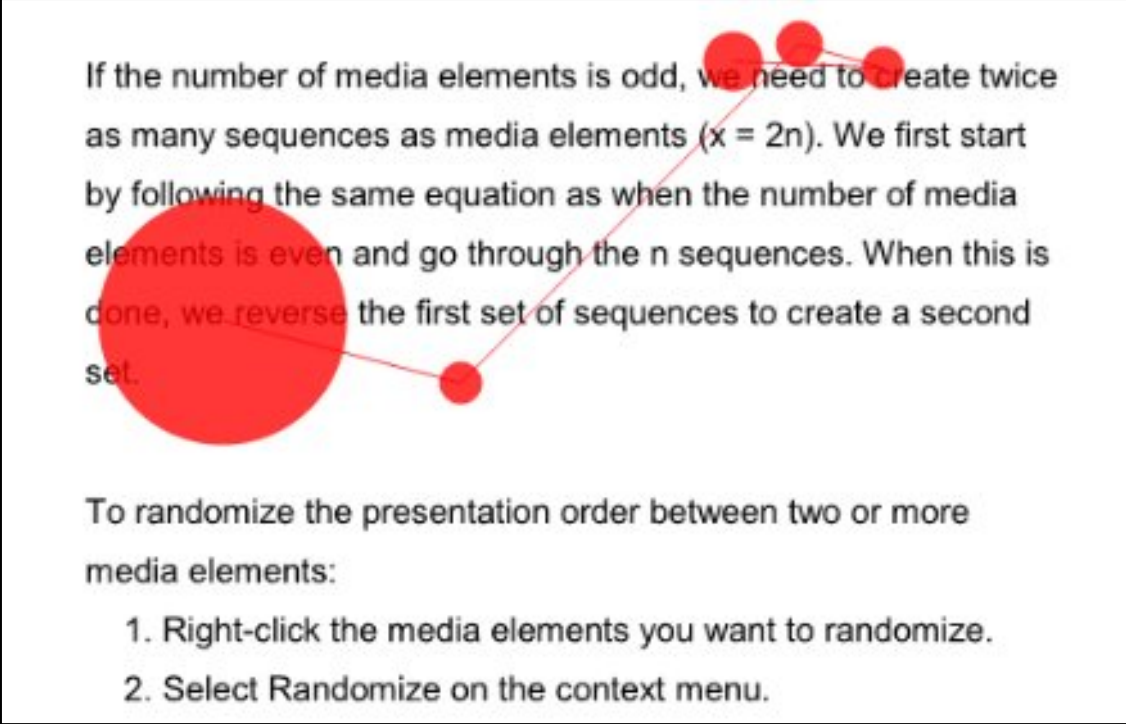
Ein anderes Problem stellt die Komplexität von Translationstechnologien und ihre Anwendung dar. Die Benutzerfreundlichkeit leidet unter der immer größer werdenden Anzahl von Programmen und Fenstern innerhalb der Programme, die Translatoren beim Übersetzen und Dolmetschen benötigen. Da aber die Vielzahl der verwendeten Ressourcen und Programme zur Texterstellung und Qualitätssicherung dienen und somit keine Reduktion diesbezüglich zulassen, versucht man die Usability dieser komplexen Workbenches zu verbessern, indem man augeninduzierte Steuerung zulässt. So bieten beispielsweise die meisten Eyetracker Softwarepakete an, mit denen die Augensteuerung von verschiedenen Programmen möglich ist. Auf diese Weise können

zum Beispiel Fenster durch Blickbewegungen aktiv werden, was dem Übersetzer bei häufigen Funktionswechseln (Internetrecherche, Termsuche, Konkordanzsuche, Textverarbeitung, etc.) das Klicken spart. Weiterhin gibt es Ansätze, bei denen Blickbewegungen zur Erkennung von Übersetzungsproblemen genutzt werden. Das System iDict (Hyrskykari et al. 2000) blendet beispielsweise wörterbuchbasiert Übersetzungskandidaten für einzelne Wörter ein, wenn die Fixation des Übersetzers einen gewissen Schwellenwert überschreitet (typischerweise ist dieser Schwellenwert einstellbar) oder wenn durch den Regressionspfad das betreffende Wort als Übersetzungsproblem eingestuft wird. Der Vorteil dieser Technologie liegt darin, dass dem Übersetzer die Recherchearbeit abgenommen wird und potenzielle Übersetzungslösungen direkt angezeigt werden, sobald die Verarbeitung ins Stocken gerät.

Solchen Systemen lässt sich jedoch die gleiche Kritik entgegenbringen wie den Translationstechnologien in Abschnitt 1; sie operieren nämlich nur auf Wortebene und lassen nur String-Matches zu. Außerdem sind sie nicht an andere Sprachtechnologien gekoppelt, die die Recherche effizienter und die Suche nach Translationskandidaten zielgerichteter gestalten könnten. Daher arbeiten wir daran, nicht nur Terme, sondern auch Kookkurrenzen und Kollokationen einzublenden, wenn durch längere Fixationen oder Regressionen entsprechende Multi-Wort-Problemstellen identifiziert werden. Als Visualisierungen sind, vor allem beim Dolmetschen, auch Word Clouds denkbar, die dem Nutzer Wortfelder in der Zielsprache als Translationsoptionen zur Verfügung stellen. Beim Übersetzen haben sich in einer ersten Pilotstudie Konkordanzen als Visualisierungshilfe als brauchbar erwiesen, da sie auch den für den Übersetzer wichtigen Kontext mitberücksichtigen (siehe Abb. 4). Auch hier erscheint die Einbindung von linguistischem Wissen als vielversprechend. Sobald kontrastive Lücken (z.B. Spaltsätze in englischen Originalen) durch das Blickverhalten als problematisch eingestuft werden, könnte man korpusbasierte Übersetzungslösungen einblenden. Ähnliches trifft auf kontrastive Fachsprachenmerkmale zu. Im Folgenden soll kurz eine Pilotstudie beschrieben werden, die zeigt, welches Potenzial die Integration von Technologien zur Sprachverarbeitung in Übersetzungshilfen birgt.

In unserer Fallstudie baten wir zwei Übersetzer einen englischsprachigen Fachtext jeweils ins Deutsche und Französische zu übersetzen. Es handelte sich dabei um einen Auszug aus dem Handbuch von Tobii Studio, der Software des Tobii-Eyetrackers aus dem Germersheimer Eyetracking-Labor. Die Blicherfassung erfolgte mit dem Tobii TX300. Wir nutzten das Eyetracking dazu, ad hoc ein Lesemodell für diese Art der Fachübersetzung zu erstellen. Problematische Stellen im Ausgangstext definierten wir durch die Fixationslänge (in unserem Beispiel waren das 3 Sekunden; dieser Wert ist aber prinzipiell veränderbar) sowie durch Regressionen. Abbildung 4 zeigt, dass beide Kriterien auf das Personalpronomen *we* (lange Fixation in der 5. Zeile, Regression in der 1. Zeile) zutreffen. Tatsächlich stellt dieses Personalpronomen für eine Übersetzung ins Deutsche oder Französische dahingehend ein Problem dar, als dass englische Fachtexte aus der technischen Redaktion personenorientiert und weniger inhaltsorientiert sind als das Deutsche oder das Französische (siehe House 1997). Dahin-

gehend stellt die Verwendung von Personalpronomen für diese Fachtextsorte eine kontrastive Lücke und somit ein Übersetzungsproblem dar. Eine wörtliche Übersetzung würde sowohl im Deutschen als auch im Französischen zwar nicht falsch, aber doch untypisch auf den Rezipienten wirken. Als Übersetzungshilfe wird daher an dieser Stelle eine Konkordanzsuche in den Subkorpora des OPUS-Korpus (OPUS Corpus query o.J.) mit der entsprechenden Textsorte (KDE und OpenOffice) gestartet und die drei ersten Treffer werden angezeigt (siehe Abb. 4).



If the number of media elements is odd, we need to create twice as many sequences as media elements ($x = 2n$). We first start by following the same equation as when the number of media elements is even and go through the n sequences. When this is done, we reverse the first set of sequences to create a second set.

To randomize the presentation order between two or more media elements:

1. Right-click the media elements you want to randomize.
2. Select Randomize on the context menu.

47088	defined the following selection : Here	we	will automatically create a for shelve 1 a chart
de	Hier wird automatisch ein Diagramm für Januar , März , April und Mai für das Regal 1 erstellt .		
fr	Un diagramme correspondant aux mois de janvier , mars , avril et mai pour le rayon 1 est créé .		
1390772	with another item in the schedule ,	we	cannot change it !
de	kollidiert mit einem anderen Eintrag im Planer , aktueller Eintrag kann nicht geändert werden .		
fr	Cette plage entre en conflit avec une autre , impossible de la modifier!		
1388561	Automatically do a full data check when	we	find to many corrupted chunks .
de	Automatisch vollständige Datenprüfung durchführen , wenn zu viele defekte Blöcke gefunden werden .		
fr	Réalise automatiquement une vérification complète quand de nombreux morceaux sont corrompus .		

Abb. 4: Augengesteuerte Übersetzungshilfen

Natürlich stellen diese Konkordanzen eine Zufallsauswahl dar, aber der Übersetzer erhält auf dieser Basis schon erste Anregungen für mögliche Übersetzungskandidaten.

Tatsächlich scheint für beide Zielsprachen die Passivierung eine konventionalisierte Übersetzungsstrategie zu sein. Wenn man nun das blickinduzierte Lesemodell des Übersetzers mit den quantitativen Korpusauswertungen verknüpft, die in Abschnitt 2.1 diskutiert wurden, lassen sich tatsächlich auch hochfrequente Strategien ermitteln. Da es sich bei diesem System jedoch lediglich um ein Pilotprojekt handelt, gilt es das vollständige Potenzial, das in der Integration von Sprach(verarbeitungs)technologien in Translationshilfen liegt, noch in der künftigen Forschung auszuschöpfen.

Literatur

- Alves, Fabio; Adriana Pagano, Stella Neumann, Erich Steiner, Silvia Hansen-Schirra (2010): "Translation Units and Grammatical Shifts: Towards an Integration of Product- and Process-based Translation Research." Gregory M. Shreve, Erik Anglone (Hg.): *Translation and Cognition*. Amsterdam: Benjamins, 109-142
- Austermühl, Frank (2001): *Übersetzen im Informationszeitalter. Überlegungen zur Zukunft fachkommunikativen und interkulturellen Handelns im Global Village*. Trier: Wissenschaftlicher Verlag
- AutoSummarize (o.J.) – <http://www.microsoft.com/education/autosummarize.aspx> (06.12.2012)
- Čulo, Oliver (2010): Automatische Extraktion von bilingualen Valenzwörterbüchern aus deutsch-englischen Parallelkorpora: eine Pilotstudie. Diss. Saarbrücken: Universität des Saarlandes
- Dragon NaturallySpeaking (o.J.) – <http://www.nuance.de/for-individuals/by-product/dragon-for-pc/premium-version/index.htm> (06.12.2012)
- EuroMatrixPlus (o.J.) – <http://www.euromatrixplus.net/> (06.12.2012)
- Fantinuoli, Claudio (2011): InterpretBank: Design and Implementation of a Terminology and Knowledge Management Software for Conference Interpreters. Diss. Germersheim: Johannes Gutenberg-Universität Mainz
- Gassmann, Tatjana (2009): Kombination maschineller und computergestützter Übersetzungsprozesse bei Corporate Language Management der Daimler AG. MA-Arbeit. Germersheim: Johannes Gutenberg-Universität Mainz
- Gile, Daniel (2009): *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: Benjamins
- Google Übersetzer (o.J.) – <http://translate.google.de> (06.12.2012)
- Göpferich, Susanne (1998): *Interkulturelles Technical Writing: Fachliches adressatengerecht vermitteln*. (Forum für Fachsprachenforschung 40.) Tübingen: Narr
- Göpferich, Susanne (2008): *Translationsprozessforschung: Stand – Methoden – Perspektiven*. (Translationswissenschaft 4.) Tübingen: Narr
- GreatSummary (o.J.) – <https://addons.mozilla.org/de/firefox/addon/greatsummary/> (06.12.2012)
- Hansen-Schirra, Silvia (2008): "Interactive Reference Grammars: Exploiting Parallel and Comparable Treebanks for Translation." Elia Yuste (Hg.): *Topics in Language Resources for Translation and Localisation*. (Benjamins Translation Library 79.) Amsterdam/Philadelphia: Benjamins, 23-38
- Hansen-Schirra, Silvia (2012): "Vom Suchen und Finden der passenden Übersetzung – Annotation und Abfrage paralleler Korpora für Englisch, Deutsch und Französisch." Barbara Ahrens, Silvia Hansen-Schirra, Monika Krein-Kühle, Michael Schreiber, Ursula Wiene (Hg.): *Translationswissenschaftliches Kolloquium II. Beiträge zur Übersetzungs- und Dolmetschwissenschaft*. Frankfurt am Main u.a.: Lang

- Hansen-Schirra, Silvia; Stella Neumann, Erich Steiner (2012): *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. Berlin: de Gruyter
- Hawkins, John (1986): *A Comparative Typology of English and German. Unifying the Contrasts*. London: Croom Helm
- House, Juliane (1997): *Translation Quality Assessment. A Model Revisited*. Tübingen: Narr
- House, Juliane (2009): *Translation*. Oxford: Oxford University Press
- Hyrskykari, Aulikki; Päivi Majaranta, Antti Aaltonen, Kari-Jouko Räihä (2000): "Design Issues of iDICT: A Gaze-assisted Translation Aid." *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications*. New York: ACM, 9-14
- Interplex Glossary Software (o.J.) – <http://www.fourwillows.com/interplex> (06.12.2012)
- InterpretBank (o.J.) – <http://www.interpretbank.de/> (06.12.2012)
- Kalina, Sylvia (1998): "Kognitive Verarbeitungsprozesse." Mary Snell-Hornby, Hans G. Höning, Paul Kußmaul, Peter A. Schmitt (Hg.): *Handbuch Translation*. (Stauffenburg-Handbücher.) Tübingen: Stauffenburg, 330-335
- Langlais, Philippe; Fabrizio Gotti, Didier Bourigault, Claude Coulombe (2005): "EBMT by Tree-Phrasing: a Pilot Study." Michael Carl, Andy Way (Hg.): *Proceedings of the Second Workshop on Example-Based Machine Translation*. (Workshop at MT-SUMMT X). Phuket – <http://www.mt-archive.info/MTS-2005-Langlais.pdf> (18.12.2012)
- LookUp (o.J.) – <http://www.lookup-web.de> (06.12.2012)
- Marcu, Daniel (2000): *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge/London: MIT Press
- MEAD (o.J.) – <http://old-site.cslp.jhu.edu/ws01/groups/asmd/news.html> (06.12.2012)
- Mertin, Elvira (2006): *Prozessorientiertes Qualitätsmanagement im Dienstleistungsbereich Übersetzen*. Frankfurt am Main u.a.: Lang
- Nitzke, Jean (2011): *Automatic Summarization and Its Relevance for Technical Translations and Scientific Writing*. MA-Arbeit. Gernersheim: Johannes Gutenberg-Universität Mainz
- OPUS Corpus query (o.J.) – <http://opus.lingfil.uu.se/bin/opuscqp.pl> (06.12.2012)

trans-kom

ISSN 1867-4844

trans-kom ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

trans-kom veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

trans-kom wird ausschließlich im Internet publiziert: <http://www.trans-kom.eu>

Redaktion

Leona Van Vaerenbergh
Artesis Hogeschool Antwerpen
Vertalers en Tolken
Schilderstraat 41
B-2000 Antwerpen
Belgien
leona.vanvaerenbergh@scarlet.be

Klaus Schubert
Universität Hildesheim
Institut für Übersetzungswissenschaft
und Fachkommunikation
Marienburger Platz 22
D-31141 Hildesheim
Deutschland
klaus.schubert@uni-hildesheim.de

- Reinke, Uwe (2004): *Translation Memories: Systeme – Konzepte – Linguistische Optimierung*. Frankfurt am Main u.a.: Lang
- Richter, Yvonne (2010): Der Einsatz von Spracherkennungssoftware beim Simultandolmetschen – Das Programmkonzept "BoothMate". Diplomarbeit. Germersheim: Johannes Gutenberg-Universität Mainz
- Schreiber, Michael (demn.): "Anaphorik als Übersetzungsproblem. Am Beispiel der deutschen Pronominaladverbien." Melanie Arnold, Silvia Hansen-Schirra, Michael Poerner (Hg.): *Streifzüge durch die Welt der Sprachen und Kulturen*. Frankfurt am Main u.a.: Lang
- SDL MultiTerm Extract (o.J.) – <http://www.translationzone.com/de/translation-agency-products/desktop/sdl-multiterm-extract/> (06.12.2012)
- Seewald-Heeg, Uta (2005): "Der Einsatz von Translation-Memory-Systemen am Übersetzerarbeitsplatz." *MDÜ – Mitteilungen für Dolmetscher und Übersetzer* 52 [4-5]: 8-43
- SweSum (o.J.) – <http://swesum.nada.kth.se/index-eng.html> (06.12.2012)
- Terminus (o.J.) – <http://www.terminus.wintringham.ch> (06.12.2012)
- Tools4Noobs (o.J.) – <http://www.tools4noobs.com/summarize/> (06.12.2012)
- Vela, Mihaela; Silvia Hansen-Schirra (2006): "The Use of Multi-level Annotation and Alignment for the Translator." *Proceedings of ASLIB Conference: Translating and the Computer 28*, London
- Verbmobil (o.J.) – <http://verbmobil.dfki.de/> (06.12.2012)
- Vinay, Jean-Paul; Jean Darbelnet (1958): *Stylistique comparée du français et de l'anglais*. (Bibliothèque de stylistique comparée 1.) Paris: Didier / Montréal: Beauchemin – Übersetzung (1995): *Comparative Stylistics of French and English: A Methodology for Translation*. Amsterdam: Benjamins
- Windows Spracherkennung (o.J.) – <http://windows.microsoft.com/de-de/windows7/Set-up-Speech-Recognition> (06.12.2012)

Autorin

Silvia Hansen-Schirra studierte Angewandte Sprach- und Übersetzungswissenschaft an der Universität des Saarlandes. 2007 folgte sie dem Ruf auf die Juniorprofessur für Computerlinguistik und Translation an die Universität Mainz am Fachbereich Translations-, Sprach- und Kulturwissenschaft in Germersheim. Seit 2010 ist sie dort als Universitätsprofessorin für Englische Sprach- und Übersetzungswissenschaft tätig. Sie beschäftigt sich mit den folgenden Forschungsschwerpunkten: Sprachtypologie, Fachkommunikation, Übersetzungsuniversalien, kognitive Aspekte der Sprachverarbeitung, computergestützte Translation, Korpuslinguistik.
Email: hansenss@uni-mainz.de

Buchempfehlungen von Frank & Timme

FFF – Forum für Fachsprachen-Forschung

Herausgegeben von
Prof. Dr. Dr. h.c. Hartwig Kalverkämper

Klaus-Dieter Baumann (Hg.): **Fach – Translat – Kultur.** Interdisziplinäre Aspekte der vernetzten Vielfalt. ISBN 978-3-86596-209-6

Radegundis Stolze: **Fachübersetzen – Ein Lehrbuch für Theorie und Praxis.**
2. Auflage. ISBN 978-3-86596-257-7

Laurent Gautier (éd.): **Les discours de la bourse et de la finance.**
ISBN 978-3-86596-302-4

TRANSÜD. Arbeiten zur Theorie und Praxis des Übersetzens und Dolmetschens

Herausgegeben von
Prof. Dr. Dörte Andres, Dr. Martina Behr
und Prof. Dr. Dr. h. c. Hartwig Kalverkämper

Lavinia Heller: **Translationswissenschaftliche Begriffsbildung und das Problem der performativen Unauffälligkeit von Translation.**
ISBN 978-3-86596-470-0.

Regina Bouchehri: **Translation von Medien-Titeln.** Der interkulturelle Transfer von Titeln in Literatur, Theater, Film und Bildender Kunst. ISBN 978-3-86596-400-7.

Martina Behr: **Evaluation und Stimmung.** Ein neuer Blick auf Qualität im (Simultan)Dolmetschen. ISBN 978-3-86596-485-4. (lieferbar ab Januar 2013)

Vorankündigung:

Claudia Dathe/Renata Makarska/Schamma Schahadat (Hg.): **Zwischentexte.** Literarisches Übersetzen in Theorie und Praxis.
ISBN 978-3-86596-442-7.



F Frank & Timme

Verlag für wissenschaftliche Literatur

Wittelsbacherstraße 27a, D-10707 Berlin

Telefon (0 30) 88 66 79 11, Fax (0 30) 86 39 87 31

info@frank-timme.de, www.frank-timme.de

Darüber hinaus:

Michaela Heinz (éd.): **Dictionnaires et Traduction.** ISBN 978-3-86596-372-7

F Frank & Timme
Verlag für wissenschaftliche Literatur

Frank & Timme GmbH

Wittelsbacherstraße 27a, 10707 Berlin

Telefon: (0 30) 88 66 79 11

Fax: (0 30) 86 39 87 31

info@frank-timme.de

www.frank-timme.de