

TEXTTECHNOLOGIE UND TRANSLATION

Peter Sandrini, Innsbruck (Österreich)

1. Von der Gutenberg-Galaxis zum digitalen Zeitalter

Die Veränderungen der letzten 20 – 30 Jahre haben unter dem Schlagwort der Globalisierung sowohl das ökonomische und politische Zusammenwachsen als auch vielfältige Veränderungen im Bereich der Kultur, gekennzeichnet durch die gegensätzlichen Tendenzen der Homogenisierung und Heterogenisierung (Globalisierung) gebracht. Eine entscheidende Rolle spielt dabei die technologische Revolution der weltweiten Kommunikationsmöglichkeiten mit dem Internet und dem WWW. Zahlreiche Forscher haben sich mit den sich daraus ergebenden Konsequenzen in allen Bereichen auseinandergesetzt: Marshal McLuhan (1995) läutete mit dem Begriff des „Global Village“ ein neues Zeitalter ein und beschrieb das Ende der sogenannten Gutenberg-Galaxis. Arjun Appadurai (1998) fasste die durch Globalisierung hervorgerufenen Veränderungen mit der Beschreibung von 5 sogenannten „scapes“ zusammen, mit denen der Anthropologe und Kulturwissenschaftler die Komplexität der modernen Welt zu erklären versucht: „ethnoscape“, „technoscape“, „financescape“, „ideoscape“ sowie vor allem „mediascape“. Während Appadurai in seiner „mediascape“-Analyse noch relativ vage bleibt, und kaum auf die Besonderheiten der Neuen Medien eingeht, versucht Marc Poster (2006) die Merkmale und Einflüsse des Web und dessen Rückwirkungen auf Kultur und Politik in das Zentrum seiner Ausführungen zu stellen.

Mit dem elektronischen Zeitalter und den neuen globalen technischen Kommunikationsmöglichkeiten sind jedenfalls auch die Prozesse der Texterstellung und der Textrezeption anderen Rahmenbedingungen unterworfen: Für jede aktive und passive Teilnahme an digitalen Kommunikationsprozessen werden bestimmte Kenntnisse und Fertigkeiten notwendig, die als digitale Medienkompetenz eine der wichtigsten Kulturtechniken unseres Zeitalters zu sehen sind.

2. Digitale Textforschung

Seit Entstehen der Textlinguistik und schon früher hat sich die Sprachwissenschaft und auch die Translationswissenschaft mit dem Text als sprachlichem Konstrukt und Gegenstand der translatorischen Tätigkeit auseinandergesetzt. Verschiedene Ansätze wurden dabei verfolgt, und jeweils unterschiedliche Aspekte des Textes beleuchtet: Zu Beginn standen strukturalistische Ansätze im Mittelpunkt, die dann kommunikationsbezogenen und handlungstheoretisch

ausgerichteten Ansätzen Platz machten, um in der Folge von der kognitions-wissenschaftlichen Textproduktions- und Rezeptionsforschung abgelöst zu werden. Gegenstand dieses Beitrages ist nun die texttechnologische Betrachtungs-weise, wobei betont werden muss, dass die hier angeführten einzelnen Tenden-zen der Textforschung sich gegenseitig keinesfalls ausschließen, sondern ergän-zend zu sehen sind und den Blickwinkel auf den Text jeweils vervollständigen. Für ein vollständiges Erfassen des Textes im digitalen Zeitalter muss natürlich das Medium entsprechend mitberücksichtigt werden: Dies bedeutet, die text-technologische Sicht auf den Text und auf textuelle Daten mit einzubeziehen.

Was unter einem Text zu verstehen sei, wurde in der Sprachwissenschaft und in der Translationswissenschaft bereits umfassend diskutiert und es soll an dieser Stelle nicht näher darauf eingegangen werden. Wesentliche Merkmale des Text-begriffes lassen sich anhand der folgenden Textdefinition aus der Translations-wissenschaft festmachen:

Ein Text ist ein thematisch und/oder funktional orientierter, kohärenter Komplex aus verbalen und/oder nonverbalen Zeichen, der eine für den Adressaten erkenn-bare kommunikative Funktion erfüllt und eine inhaltlich und funktional abge-schlossene Einheit bildet (Schmitt 1997, 25).

Texte werden im elektronischen Zeitalter im Wesentlichen als Dokumente ge-handhabt, der Begriff Dokument kommt aus der Informationswissenschaft und wird verstanden als „eine inhaltlich begrenzbare Einheit von Wissen, welches auf einem materiellen Träger dauerhaft fixiert ist“ (Manecke / Seeger 1997, 21). Im Gegensatz zur Textdefinition oben wird hier nun auch der Träger, das elekt-ronische Medium in all seinen Varianten thematisiert. In diesem Sinne wird der digitale Text auch unter dem informationswissenschaftlichen Aspekt des Daten enthaltenden Dokumentes gesehen und mit der Schlagwortkette „Datei – Doku-ment – Text – Informationseinheit – Daten“ gekennzeichnet.

Bevor wir zur eigentlichen Texttechnologie kommen, sollen hier die wesentli-chen Neuerungen, die eine Sichtweise auf den Text als digitales Dokument mit sich bringt, dargestellt werden. Die Unterschiede, die eine Adaptierung des her-kömmlichen Textbegriffes notwendig erscheinen lassen, können nach Storrer (2004, 13) in drei Merkmalen identifiziert werden:

- Die Struktur und die Speicherung von digitalen Dokumenten erfolgt unab-hängig von ihrer Visualisierung. Realisiert wird dies durch das Trennen von Inhalt und Form, wobei die Struktur und die inhaltliche Auszeichnung von Dokumenten die Voraussetzung dafür bilden, dass mithilfe von Layout-vorlagen und Scripts relativ leicht eine veränderte Visualisierung des Doku-mentes erzeugt werden kann bzw. das Dokument für unterschiedliche Ver-

wendungsszenarien und Medien flexibel gestaltet, angepasst und verändert werden kann: „Strukturierte Texte bilden deshalb ein Art informationellen Rohstoff im Publikationsprozess, der für verschiedene Zwecke eingesetzt und ‚raffiniert‘ werden kann“ (Lobin und Lemnitzer 2004, 5). Man spricht in diesem Zusammenhang von „single source publishing“ bzw. „cross media publishing“.

- Die Verarbeitung digitaler Texte ist ein andauernder zyklischer Prozess, bei dem textuelle Daten produziert, gepflegt und genutzt werden: Dies beginnt mit der Strukturierung der textuellen Daten in formalen Dokumentengrammatiken, geht über die Datenerfassung bzw. Neueingabe über die Bearbeitung und Konvertierung zur Revisionierung und wieder zurück zur Überprüfung der Strukturierung. Es fehlt dabei ein Endprodukt wie in der Gutenberg-Galaxis, als ein Buch geschrieben, publiziert und gedruckt wurde und danach in Buchform vorlag. Digitale Texte unterliegen einem andauernden Verarbeitungsprozess, in dem es lediglich temporäre Versionen und zeitlich abgestimmte Online-Verfügbarkeit gibt. Storrer (2004, 13) spricht in diesem Zusammenhang von einem zyklischen Prozess der Verarbeitung von digitalen Dokumenten, „bei dem ein Ausgangsdokument immer wieder inhaltlich modifiziert, strukturell annotiert und an neue Anwendungsszenarien angepasst werden kann“.
- Texte und Textteile werden als Daten verlinkt. Die Strukturierung des digitalen Textes erlaubt es, Textteile als Daten zu sehen, die einerseits kategorisiert und dokumentiert werden können, andererseits Verbindungen zu anderen Daten eingehen können.

Strukturierung, Bearbeitungszyklus und Dokumentation beruhen im digitalen Zeitalter stets auf der Kooperation vieler Personen und fördern die Differenzierung der Textkompetenzen aller am Dokumentenzyklus Beteiligten; die kooperative Texterstellung wird dadurch in vielen Bereichen zur Regel, wobei das Erstellen multilingualer Dokumente und das Übersetzen in den Lebenszyklus digitaler Dokumente nahtlos integriert wird. Beispiele für digitale Texte, die heute vor allem auch Gegenstand von Translation sind: Dokumente, Websites, ComputerGames, Software, Multimediaprodukte (CD-ROM) und Datenbanken. Der Umgang mit diesen digitalen Texten erfordert spezifische Kenntnisse über die damit verbundenen Formate und Dateitypen. So wäre es undenkbar für einen Übersetzer, der im Bereich der Website-Übersetzung tätig ist, nicht die entsprechenden XHTML-Kenntnisse sowie die dazugehörigen Stylesheet-Formate CSS oder XSLT zu beherrschen. Das digitale Texthandling verlangt vom Translator, sich in verschiedenste Dokumentenformate einzuarbeiten.

3. *Texttechnologie*

Zusammenfassen lassen sich diese praktischen und theoretischen Anforderungen unter dem Begriff der Texttechnologie, die einen der zwei wesentlichen Teilbereiche der maschinellen Sprachverarbeitung im Rahmen der Informationstechnologie bildet, und zwar

Sprachtechnologie im engeren Sinne (Spracherkennung, Sprachsynthese, Sprachdialogsysteme) und Texttechnologie (Textklassifikation, Textzusammenfassung, Textmining, Informationsextraktion, natürlichsprachige Suche, Frage-/Antwort-Systeme) (Carstensen et al. 2010, 554).

Die Texttechnologie befindet sich an der Schnittstelle zwischen der Computerlinguistik und der Computerphilologie, hier aber näher an der Linguistik als an den Philologien und beschäftigt sich mit der maschinellen Verarbeitung geschriebener Sprache. „Für eine effiziente und auch auf Texttypen abgestimmte adäquate Verarbeitung von Texten ist es derzeit gängige Praxis, diese durch die Hinzufügung von Zusatzinformationen aufzubereiten. Hierbei spielt die Verwendung von Auszeichnungssprachen eine zentrale Rolle“ (Witt / Metzging 2006). Die Online-Enzyklopädie Wikipedia nennt als wichtigste Anwendungsgebiete der Texttechnologie Auszeichnungssprachen, Textkorpora, Hypertext-Theorie und -Technologien, Datamining sowie Ontologien.

Ziel der Texttechnologie ist die „Entwicklung von Verfahren, die präzise und deshalb algorithmisierbar sind und die auf Texte als semistrukturierte Daten angewendet werden“ (Lobin / Lemnitzer 2004, 1). Der zentrale Begriff der semistrukturierten Daten fokussiert das Anwendungsgebiet auf textuelle Daten und grenzt sich auf der einen Seite von geisteswissenschaftlichen Disziplinen ab, die sich mit dem Text und seiner Exegese beschäftigen, wie die theologischen, philologischen und juristischen Verfahren, auf der anderen Seite hingegen von den Datenbanken, deren Untersuchungsgegenstand strukturierte Daten sind. Auf der gleichen Ebene stehen der Texttechnologie andere sprachtechnologische Disziplinen gegenüber, deren Gegenstand zumindest nicht in erster Linie Texte sind.

Nach Lobin / Lemnitzer (2004, 7) wendet die Texttechnologie vier grundlegende Verfahren an:

1. Verfahren, bei denen Texte mit zusätzlichen Informationen angereichert werden;
2. Verfahren, mit denen Informationen aus Texten extrahiert werden;
3. Verfahren, mittels derer Texte zugänglich gemacht werden, um ein bestimmtes Informationsbedürfnis zu befriedigen;
4. Verfahren, mittels derer aus Texten abgeleitete Objekte erstellt werden (Lobin / Lemnitzer 2004, 7).

Nah verwandt sind das Dokumentenmanagement und das Elektronische Publizieren, die sich jedoch stärker mit dem Dokument als solchem und seiner äußeren Erscheinungsform beschäftigen.

Anders als bei dedizierten Studieninhalten wie beispielsweise der Nebenfachstudiengang *Texttechnologie* im Bachelor und im Master der geisteswissenschaftlichen Kernfächer an der Universität Bielefeld, wird die Texttechnologie in der Translationsausbildung funktional zum Fach Translationswissenschaft gesehen und daher die Inhalte auch entsprechend angepasst. Bevor wir jedoch den neuen Begriff der translationsspezifischen Texttechnologie beschreiben und dessen Inhalte festlegen können, sollen zunächst die Bezüge der Translation zur Textforschung dargestellt werden.

4. *Übersetzen als multilinguale Textarbeit*

Ohne Zweifel nimmt der Text eine zentrale Rolle im Translationsprozess ein. Der Text ist stets Ausgangspunkt und Zielpunkt des translatorischen Handelns. Holz-Mäntari (1984) geht sogar soweit, das Übersetzen als „funktionsspezifische Textarbeit“ zu definieren.

Neuere Ausprägungen translatorischer Dienstleistungen beziehen explizit digitale Texte in den umfassenden globalen Textlebens- bzw. Informationszyklus mit ein. So definiert sich z.B. die Lokalisierung als Hyperonym für Software-, Website- und Multimedialokalisierung als „services and technologies for the management of multilinguality across the global information flow“ (Schäler 1999). Ebenso spricht man im Rahmen der internationalen Produktdokumentationserstellung nicht mehr von Übersetzung als einem für sich allein stehenden Bereich, sondern integriert sie in die mehrsprachige technische Redaktion, die dann als spezifische Ausbildung angeboten wird, u.a. an der Fachhochschule für angewandte Sprache SDI in München.

Somit steht der digitale Text im Mittelpunkt des modernen (professionellen) Übersetzers, und zwar beinahe in jedem Fachbereich, sogar literarische Texte werden digital erfasst, bearbeitet, gesetzt und gedruckt, oder auch im Netz als Hypertexte kooperativ erstellt und weiterentwickelt, wobei zugegebenermaßen in manchen philologischen Nischen die texttechnologische Dimension eine eher geringere Rolle spielt.

Der Übersetzer empfängt digitale Texte in einem bestimmten Format mit einer speziellen Auszeichnung, die er bearbeiten, in eine andere Sprache übersetzen, für ein anderes Zielpublikum und eine unterschiedliche Zielkultur aufbereiten und für den vom Auftraggeber festgelegten Zweck funktional anpassen muss. Der Ausgangstext bildet den informationellen Rohstoff, auf dessen Basis der Zieltext erstellt wird, d.h. der informationelle Rohstoff wird bearbeitet, verän-

dert, angepasst und als Zieltext publiziert, in einem weiteren Zyklus wieder aktualisiert, angepasst, neu übersetzt, etc. Auszeichnungselemente und Formatfragen werden im eigentlichen Übersetzungsprozess möglichst ausgeklammert bzw. im Zieltext funktional wiedergegeben. Konzeptualisiert man mit Lobin / Lemnitzer (2004, 3) „die Verwendung textueller Daten als das Durchlaufen eines Lebenszyklus“, so eröffnet das Übersetzen quasi einen neuen Lebenszyklus der textuellen Daten, eingebettet in ein neues sprachliches und kulturelles Umfeld. Umgekehrt integriert sich das Übersetzen als ein weiterer Schritt im mehrsprachigen Textlebenszyklus mit eigenen Verfahren und Prozessen.

Ansatzweise könnte man damit sogar so weit gehen, das Übersetzen als eine besondere Art der Transformation digitaler Texte zu sehen und damit die texttechnologische Sicht auf das Übersetzen absolut setzen. Unbestreitbar ist hingegen die Tatsache, dass „medientechnische Kompetenz“ (Austermühl 2001, 147) und ein umfassendes texttechnologisches Wissen für den Übersetzer von zentraler Bedeutung sind.

5. *Translationsspezifische Texttechnologie*

Aufgrund dieser Prämissen soll nun versucht werden, den Begriff der *translationsspezifischen Texttechnologie* einzuführen. Für die Zwecke der Translation und damit eingeschlossen der Translationsdidaktik können wir die translationsspezifische Texttechnologie als eine Anwendung der Texttechnologie auf Translation und Lokalisierung definieren, wobei unter der selbstverständlichen Voraussetzung, dass der (Ausgangs- und Zielt-)Text jedenfalls digital vorliegt und entsprechende Rechnerkonfigurationen sowie Kodierungssysteme vorhanden sind (Lobin / Lemnitzer 2004, 2), vor allem die Aspekte der multilingualen Informations- und Dokumentationsarbeit, der Konversions- und Transformationsarbeit im Translationsprozess sowie der multilingualen Austauschformate im Vordergrund stehen.

Insbesondere die beiden letzten Themen möchte ich im Folgenden näher beleuchten. Die Konversions- und Transformationsarbeit ist im modernen Translationsprozess omnipräsent: „Freelance translators also need to adapt to an environment using electronic document interchange where a key factor is the ability to handle a variety of IT systems, software, tools, and resources“ (Letrac 1998). Dies ist bei weitem nicht eine formale Nebensache, sondern vielmehr ein integraler Bestandteil der Translation digitaler Texte, da der Translationsprozess sonst außerhalb des beschriebenen Lebenszyklus digitaler Dokumente stattfinden würde. Dies würde einen Bruch im Verarbeitungsprozess darstellen, der nicht nur einen hohen zusätzlichen Aufwand bedeutet, sondern vor allem auch den Übersetzer von der digitalen Produktion abkoppelt, mit allen Konsequenzen für

seinen professionellen Status und seine Verdienstmöglichkeiten. In diesem Fall würde der Übersetzer wieder einmal als ein notwendiges und kostenintensives – also möglichst zu vermeidendes – Übel angesehen. Bei einer selbstverständlichen Integration der Übersetzung in den Dokumenterstellungszyklus wird die Mehrsprachigkeit hingegen zu einem von Anfang an mitgeplanten, weiteren Bearbeitungsschritt als Teil des gesamten Textlebenszyklus und der Übersetzer mit seinem Expertenwissen Teil des Teams.

Handelt es sich bei der Konversion für den Übersetzer meist um einen temporären Bearbeitungsschritt, der dann am Ende des Prozesses de facto wieder rückgängig gemacht wird, indem das Ausgangsformat wieder hergestellt wird, so stellt die Transformation eine endgültige Umwandlung der textuellen Daten dar: Eine Auszeichnungsart wird zur weiteren Verarbeitung in eine andere Auszeichnung überführt.

5.1. *Texttechnologie und Translationsprozess*

Im Translationsprozess geht es unter texttechnologischem Blickwinkel zunächst darum, aus einem situationsspezifischen digitalen Text die übersetzungsrelevanten Daten zu extrahieren, diese zu übersetzen bzw. zu lokalisieren und wiederum zu einem situationsspezifischen digitalen Text überzuführen bzw. aufzubereiten, damit ein funktionaler Einsatz des Zieltextes gewährleistet werden kann. Der Begriff *Daten* wird hier zwar aus der Informatik entlehnt, dennoch aber einerseits nach entsprechender Segmentierung und Strukturierung des Textes für die sprachliche Oberfläche des Textes verwendet, andererseits auch für alle mit dem Text verbundenen Metadaten.

Abhängig vom Kontext der Übersetzungsdienstleistung (Translationsauftrag, Funktion des Zieltextes etc.) kann es dabei zu zwei verschiedenen Szenarien kommen: In einem ersten Szenario unterscheidet sich der Kontext und die Funktion des Zieltextes nur geringfügig von der des Ausgangstextes und die Form des Ausgangstextes kann auch für den Zieltext übernommen werden. Auch hier stammt der Begriff *Form* aus der Informatik und beschreibt Dokumententypdefinitionen, Layout-Vorlagen, Metadaten, Navigation u.ä., alles was zur Struktur des digitalen Textes beiträgt. Bei der Extraktion der textuellen Daten und dem damit verbundenen Trennen von Form und Inhalt steht die Berücksichtigung der Spezifika des jeweiligen digitalen Textes im Vordergrund, da dabei unterschiedliche Verfahren angewendet werden können. Bei einem Webauftritt, der eine funktionale Einheit bildet und daher als (Gesamt-)Text zu sehen ist (vgl. Jakobs / Lehnen 2005, 161), werden beispielsweise die einzelnen Teiltexthe (HTML-Dokumente), die jeweiligen Layout-Vorlagen (CSS, XSLT) sowie die multimedialen Inhalte (Audio, Video) erfasst.

Wenn wir jedoch den Text als kulturelles Artefakt sehen, das auf den verschiedenen Ebenen jeweils von fachspezifischem Wissen und von Kultur geprägt ist und jeweils kulturellen Kommunikationsmustern (Textsorten) zuzurechnen ist, und dadurch den Text in seiner Komplexität und in seinen Wechselbeziehungen zwischen außersprachlicher Realität, den Begriffen, Aussagen bzw. Propositionen, kognitiven Modellen (ICM) und der Ausdrucksseite bzw. den verwendeten sprachlichen Mitteln und den eingesetzten Textgestaltungsmitteln sehen, eröffnet sich ein zweites Szenarium, das die pauschale Übernahme der Formgestaltung auf den Zieltext einschränkt bzw. verhindert. Die Einbettung in eine veränderte Kommunikationssituation und in einen neuen kulturellen Rahmen erfordert eine entsprechende Anpassung der Form, die ihrerseits kulturspezifisch ist: Textsorten können eine veränderte Makrostruktur und damit auch eine unterschiedliche Anzahl von Teiltextrn sowie eine unterschiedliche Navigationsstruktur aufweisen, Farben zu Missverständnissen Anlass geben und allgemeine Konventionen abweichen.

Unabhängig von kulturellen Anpassungen oder pauschaler Übernahme der Form steht die Analyse und Zerlegung des digitalen Textes in die Datenkomponenten Form und Inhalt, das Erfassen des digitalen Textes in seiner Struktur, das Verstehen von Textzusammenhängen im Mittelpunkt dieses ersten Schrittes, den wir als Makroebene der texttechnologischen Transformationsarbeit durch den Übersetzer bezeichnen, weil es den gesamten Text als funktionale Einheit betrifft.

Während die Analyse auf der Makroebene zu den übersetzungsvorbereitenden Maßnahmen zählt, sind im eigentlichen Translations- bzw. Lokalisierungsvorgang ebenfalls Textkonversions- und Transformationsvorgänge nötig, um die einzelsprachlichen Textstellen für den Übersetzer bearbeitbar zu machen, bzw. umgekehrt ausgedrückt, um das Text-Markup auf den Zieltext übertragen zu können. Dies geschieht durch Datei- und Konversionsfilter, über die computergestützte Übersetzungswerkzeuge (Translation-Memory-Systeme) verfügen. So wird für die Teiltextrn eines Webauftrittes beispielsweise der HTML-Code durch das TM-System automatisch ausgeklammert und nur der reine Text dargestellt, der bearbeitet und übersetzt werden kann. Hierbei sprechen wir von Textkonversions- und Transformationsvorgängen auf der Mikroebene.

Doch nicht nur der Translationsprozess digitaler Texte erfordert texttechnologisches Wissen, sondern ebenso die Nachbearbeitung, die Weitergabe und das Wiederverwenden der Ergebnisse. Damit die Translationsergebnisse ausgetauscht und als Daten verwendbar gemacht werden können, braucht es die digitale Form wie auch die Nachvollziehbarkeit der Übersetzung, die Daten müssen dazu sowohl den Ausgangstext als auch den Zieltext enthalten. Zu diesem Zweck wurden Standardformate entwickelt, die diese Voraussetzungen erfüllen.

Das wohl bekannteste ist das Translation-Memory-Austauschformat TMX (Translation Memory eXchange), das eine segmentweise Speicherung der Übersetzung gewährleistet. TMX-Dateien stützen sich auf den allgemeinen Markup-Standard XML (eXtended Markup Language) und repräsentieren ein Datenrepositorie an Segmentübersetzungen: Jedes Segment des Ausgangstextes wird mit seiner Übersetzung (= Segment des Zieltextes) in einem sogenannten Translation Unit (TU), das durchaus auch mehrsprachig sein kann, gespeichert. Dadurch ist das TMX-Format nicht auf einen spezifischen Text bezogen, sondern kann Übersetzungssegmente vieler Texte enthalten. Die Organisation der Übersetzungsdaten bleibt dem Anwender überlassen, der TMX-Dateien jeweils für einzelne Fachgebiete, für spezifische Textsorten, Kunden oder nach anderen Kriterien geordnet anlegen kann.

Für den Austausch von Lokalisierungsdaten wurde das XLIFF-Format (XML Localization Interchange File Format) entwickelt, das analog zum TMX-Format die segmentweise Speicherung der Übersetzung zulässt, im Gegensatz zu TMX aber nicht als Datenrepository konzipiert wurde, sondern als projektspezifisches Austauschformat. Damit ist das XLIFF-Format textbezogen, oder zumindest projektbezogen, sofern das Lokalisierungsprojekt mehr als einen Text enthält. Darüber hinaus können XLIFF-Dateien zusätzliche, für die Lokalisierung wichtige Informationen (Kommentare, Statusangaben, etc.) enthalten.

Im Rahmen von Textproduktion wird das XML:TM-Format (XML based Text Memory) eingesetzt, das innerhalb eines Textes unterschiedliche Versionen, historisch auf den Textproduktionsprozess bezogen sowie interlingual als Sprachversionen desselben Textes gesehen, im XML-Code ablegt. Dadurch werden Versionskontrollen und das Nachvollziehen von Übersetzungen möglich. Auch hier können zusätzliche Informationen mitgespeichert werden.

Von diesen drei Formaten hat sich in der Praxis das TMX-Format auf breiter Basis durchgesetzt, da es die Interoperabilität von Translation-Memory-Systemen gewährleistet. Das XLIFF-Format beginnt langsam im Bereich der Softwarelokalisierung Fuß zu fassen (siehe XLIFF-Seminare Irland), während das XML:TM-Format bisher über ganz spezifische Anwendungsszenarien nicht hinausgekommen ist.

Aus einem erfolgreichen Translationsprojekt ergibt sich meist eine mehrsprachige Fachterminologie, die ebenfalls für weitere Verwendung abgespeichert und weitergegeben werden kann. Dafür steht das Austauschformat TBX (TermBank eXchange) zur Verfügung, das seinerseits auf XML basiert und begriffsorientierte terminologische Einträge dokumentiert.

5.2. *Inhaltsaspekte einer translationsspezifischen Texttechnologie*

Der Anwendungsbereich einer allgemeinen Texttechnologie – mit (Carstensen et al. 2010, 554) umschrieben mit Textklassifikation, Textzusammenfassung, Textmining, Informationsextraktion, natürlichsprachige Suche, Frage-/Antwort-Systeme – ist für die Zwecke einer translationswissenschaftlichen Integration enger zu fassen bzw. funktional anzupassen. Aufbauend auf die Erkenntniss verschiedener Projekte (Letrac-Projekt, LISA-LEIT, u.a.) umfasst die translationsspezifische Texttechnologie einerseits zu einem geringen Ausmaß Elemente einer translationsspezifischen IT, andererseits umfasst sie einen wesentlichen Teil der Inhalte der Sprachverarbeitung (Language Engineering). Sie ermöglicht dem Übersetzer den „Umgang mit Formen der Informations- und Kommunikationstechnologie“ (Austermühl 2001, 247) und enthält folgende Elemente, hier dargestellt in aufsteigender Reihenfolge vom Allgemeinen zum Spezifischen:

- a) Konversion und Transformation digitaler Texte:
Jeder digitale Text wird im Computerspeicher in einer bestimmten Form abgespeichert, die Art und Weise der Informationsspeicherung festlegt. Im Translationsprozess wird der professionelle Umgang bzw. die Transformation und Konversion dieser Formate vorausgesetzt, wodurch das Wissen um Dokumentenformate, Darstellung von Sprachen, Schriften und Kodierung notwendig werden. Zu übersetzender bzw. anzupassender Text ist nicht nur in offensichtlichen Textformaten enthalten, sondern kann ebenso in Grafikformaten oder Multimediadateien (z.B. Videoformate) eingebettet sein. Die Unterscheidung zwischen offenen und proprietären Dateiformaten fällt ebenso darunter.
- b) Auszeichnungssprachen:
Markupsprachen, insbesondere XML, sind eine wesentliche Voraussetzung für alle texttechnologischen Anwendungen: Hier steht das grundlegende Verständnis der Funktionsweise im Vordergrund (Dokumentgrammatiken DTD) sowie die Visualisierung und Transformation (CSS, XSLT, XML-Schema).
- c) Digitale Textmodelle:
Darunter subsumieren wir einerseits das Verständnis digitaler Textsorten, andererseits alle XML-basierten Formate für Translation und Lokalisierung. Beide Kompetenzen enthalten jeweils einen theoretisch-strukturellen und einen technisch-praktischen Aspekt.
 - Digitale Textsorten:
Zentral ist hier das Verständnis neuer digitaler Textsorten. Für den Bereich WWW zählen hier die Hypertextlinguistik und die Hyper-textsorten sowie multimodale Texte zu den wesentlichen Inhalten. Auf

der technisch-praktischen Ebene werden Kenntnisse von (X)HTML, CSS und Webapplikationen (Web 2.0), Multimedia- und Untertitelungsformate vermittelt.

Der Übersetzer soll u.a. in die Lage versetzt werden, HTML-Dokumente „lesen“ zu können, d.h. die HTML-Tags vom eigentlichen Text trennen, ihre Funktion innerhalb des Textes verstehen und entsprechend in die Struktur des HTML-Dokumentes eingreifen zu können. Lehrinhalte können exemplarisch mit der Durchführung von praktischen Projekten verbunden werden: z.B. Erstellen einer Homepage, Retrodigitalisierung eines bestehenden Textes (vom analogen zum digitalen Textobjekt).

- XML-basierte Translations- und Lokalisierungsformate
Dazu zählen die oben erwähnten Standardformate TBX, TMX, XLIFF und XML:TM sowie die ergänzenden SRX (Segmentation Rules eXchange) und GMX-V (Global Information Management Metrics eXchange) als auch andere offene XML-Formate wie DocBook, DITA.

d) Multilinguale Korpuslinguistik.

Die Rolle von Korpora ist in der Translationspraxis und -wissenschaft unbestritten: Die Extraktion des in Parallelkorpora enthaltenen praktischen Sprachwissens soll für die teilautomatische Übersetzung sowie für die Analyse und Bewertung von Übersetzungen nutzbar gemacht werden. Inhalte sind hier Methoden und Verfahren des automatischen und nichtautomatisierten Markups anhand von Auszeichnungssprachen, Fragen der Segmentierung und des Alignments zur Produktion mehrsprachiger Korpora, Informationserschließung aus Parallelkorpora.

e) Metadaten.

Das Arbeiten mit digitalen Texten beinhaltet den Umgang mit Metadaten und den dafür entwickelten Formaten: TEI (Text Encoding Initiative), Dublin Core, RDF (Resource Description Framework), Ontologien, insbesondere unter dem Stichwort der semantischen Interoperabilität des WWW sowie der intertextuellen Bezüge digitaler Texte.

f) Informationserschließung (IR).

Unter diesem Punkt fassen wir das sogenannte Textmining (vgl. Mehler 2004, 330) und die eigentliche Informationserschließung (vgl. Krüger-Thielmann / Pajmans 2004, 353) zusammen.

5.3. *Curriculare Integration*

Translationsspezifische Texttechnologie ist damit ein unverzichtbarer Teil des Faches Translationstechnologie und darf in keinem Curriculum für Translatoren fehlen. Die Integration in Studienpläne kann aber durchaus unterschiedlich erfolgen. Wünschenswert wäre ein Modul Translationstechnologie, das einen inhaltlichen Schwerpunkt Texttechnologie enthält, einerseits um die Bedeutung dieser Grundlagen für die Translation digitaler Texte zu unterstreichen, andererseits aber auch um die häufig anzutreffenden Produktschulung im Studium zu vermeiden. Die Vermittlung von Translationstechnologie darf sich keinesfalls darin erschöpfen, exklusive Zertifizierungen im Umgang mit Softwareprodukten zu erwerben, seien es auch Softwareanwendungen des Marktführers: Es wäre wohl nicht im Sinne einer meist öffentlich finanzierten Ausbildungsinstitution in Konkurrenz mit privatwirtschaftlichen, marktorientierten Anbietern zu treten bzw. sich als Marketing- und Absatzinstrument von Softwarehäusern instrumentalisieren zu lassen.

Häufig ist die translationsspezifische Texttechnologie aber auch bereits in anderen bestehenden Ausbildungsgängen oder -modulen zumindest teilweise integriert und bildet einen Teil folgender Kurse:

- Lokalisierung: Software und Websites
- Technische Redaktion
- Terminologie
- Translation allgemein

Wichtig erscheint jedenfalls, dass die oben genannten Inhalte explizit im Translationscurriculum genannt werden.

6. *Schlußbemerkungen*

In diesem Beitrag wurde versucht, die relativ neue Teildisziplin der translationsspezifischen Texttechnologie definitorisch zu erfassen, ihre Bedeutung für die Translationspraxis und Translationsdidaktik zu erläutern und Ansätze für eine inhaltliche und curriculare Einordnung in die Translationsausbildung zu formulieren. Dadurch soll ein wesentlicher Baustein für eine zeitgemäße, moderne Vermittlung der Translationstechnologie gelegt werden.

Literaturverzeichnis

- Appadurai, A. (1998): *Modernity at large*. Minneapolis
- Austermühl, F. (2001): *Übersetzen im Informationszeitalter. Überlegungen zur Zukunft fachkommunikativen und interkulturellen Handelns im Global Village*. Trier
- Carstensen, K-U. / Ebert, Ch. / Endriss, C. / Jekat, S. / Klabunde, R. / Langer, H. (Hg.) (2010): *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Heidelberg

- Holz-Mänttari, J. (1984): *Translatorisches Handeln*. Helsinki (Annales Academiae Scientiarum Fennicae, Series B)
- Jakobs, E-M. / Lehnen, K. (2005): Hypertext – Klassifikation und Evaluation. In: Siever, T. / Schlobinski, P. / Runkehl, J. (Hg.): *Websprache.net. Sprache und Kommunikation im Internet*. Berlin, New York, 159-184
- Krüger-Thielmann, K. / Paijmans, H. (2004): Informationserschließung. In: Lobin / Lemnitzer (Hg.), 353-378
- LEIT – Lisa Education Initiative Taskforce. Online: <http://leit.lisa.org> (Link seit Auflösung der Localization Industry Standards Association 01-2011 nicht mehr erreichbar).
- LETRAC – Language Engineering for Translator Curricula. Reuthen, Ursula, IAI, Saarbrücken. Online: <http://www.uni-saarland.de/fak4/fr46/workshop/LETRAC/sld001.htm> (05-2011)
- Lobin, H. / Lemnitzer, L. (Hg.) (2004): *Texttechnologie. Perspektiven und Anwendung*. Tübingen
- Manecke, H. / Seeger, T. (1997): Zur Entwicklung der Information und Dokumentation in Deutschland. In: Buder, M. / Rehfeld, W. / Seeger, Th. / Strauch, D. (Hg): *Grundlagen der praktischen Information und Dokumentation*. München, 16-60
- McLuhan, M. (1995): *The Global Village. Der Weg der Mediengesellschaft in das 21. Jahrhundert*. Paderborn
- Mehler, A. (2004): Textmining. In: Lobin / Lemnitzer (Hg.), 329-352
- Poster, M. (2006): *Information Please: Culture and Politics in the Age of Digital Machines*. Durham, London
- Schäler, R. (1999): New Media Localisation, LingLink Report for the European Commission DGXII. Dublin. Online: <http://www.localisation.ie/resources/Publications/NewMedia/NewMedia.htm> (05-2011)
- Schmitt, P. A. (1997): Was ist ein „Text“? In: Fleischmann, E. / Kutz, W. / Schmitt, P.A. (Hg.): *Translationsdidaktik. Beiträge zu Grundfragen der Translationswissenschaft*. Tübingen, 15-27
- Storrer, A. (2004): Text und Hypertext. In: Lobin / Lemnitzer (Hg.), 13-49
- Witt, A. / Metzger, D. (2006): Texttechnologie an der Universität Bielefeld. Online: <http://computerphilologie.tu-darmstadt.de/jg05/wittdiewald.html> (29.09.2010)