Open Translation Data: Neue Herausforderung oder Ersatz für Sprachkompetenz?

Peter Sandrini

Zusammenfassung

Durch das ubiquitäre Netz nimmt das Speichern, Bearbeiten und Wiederverwenden von Daten jeglicher Art immer mehr an Bedeutung zu, während Zugang zu und Rechte an Daten zu einer zentralen gesellschaftlichen Frage werden. In diesem Beitrag stehen die verschiedenen Arten und die Beschreibung der im Zuge der Mehrsprachigkeit und der Translation anfallenden Daten im Mittelpunkt, ebenfalls die Frage nach der gesellschaftlichen Relevanz dieser Art von Daten. Der neue Begriff der „Open Translation Data" wird erklärt und seine Auswirkungen auf das Übersetzen und die Mehrsprachigkeit im Allgemeinen sowie auf die Sprach- und Translationskompetenz im Besonderen diskutiert.

Übersetzungsdaten

Wir leben in einem Zeitalter, in dem Daten immer wichtiger werden: Sowohl aus wirtschaftlicher Sicht, wo Daten über unser Kaufverhalten an Bedeutung zunehmen, als auch aus privater bzw. gesellschaftlicher Sicht, wo Daten über unsere Vorlieben, Gewohnheiten und Freunde immer mehr in den Mittelpunkt der öffentlichen Aufmerksamkeit, der Datenschützer und der Geheimdienste rücken. Im folgenden Beitrag stehen die für die Übersetzung verwendeten und durch die Übersetzung entstehenden Daten im Mittelpunkt der Betrachtung, wobei aufgezeigt werden soll, um welche Art von Daten es sich dabei handelt, welche Voraussetzungen dafür bestehen müssen und schließlich welche Chancen und Gefahren dadurch entstehen.

Das Übersetzen ist im digitalen Zeitalter angekommen (vgl. Cronin 2013). Die Digitalisierung hat den Umgang mit Texten, die nunmehr zu Dateien geworden sind, deutlich verändert: Texte können beliebig kopiert und vervielfältigt werden; auch der Transport und die Wiedergabe von elektronischen Texten funktioniert ohne Aufwand durch Tastendruck. Darüber hinaus können umfangreiche Textmengen einfach nach Zeichenfolgen durchsucht sowie ganze Texte und Textabschnitte beliebig oft wiederverwendet werden. Die sogenannten „Digital Humanities" wenden computergestützte Verfahren und digitale Ressourcen systematisch in den Geistes- und Kulturwissenschaften an (Burdick 2012). Für das Übersetzen werden Ausgangstexte ausschließlich als digitale Texte bearbeitet und übersetzt und in digitale Zieltexte überführt, wodurch sich u.a. auch neue Ausbildungsanforderungen in der Form der Texttechnologie ergeben (vgl. Sandrini 2012). Um in digitalen Textwelten arbeiten zu können, bedarf es digitaler Werkzeuge, die zu einer weiteren Veränderung der Translation beigetragen haben. Daraus folgt die weitgehende Technologisierung des Übersetzens durch das computergestützte Übersetzen und die automatische Maschinenübersetzung. Der Einsatz von ‚Translation Environment Tools' (TEnT) (Zetzsche 2014: S. 189) bestimmt heute weitgehend die Vorgangsweise des menschli-

chen Übersetzers und den Ablauf von Übersetzungsprojekten (Choudhury/McConnell 2013). Verwendung und Einsatz dieser Werkzeuge hat sich – wie auch in anderen Bereichen – von einer instrumentellen Komponente der Translationskompetenz (PACTE 2007, S. 238) weitgehend zu einer den gesamten Translationsprozess durchdringenden, allgemeinen Kompetenz gewandelt (Cnyrim/Hagemann/ Neu 2013, S. 11). Wie weit die Bedeutung dieser Werkzeuge für das Übersetzen angestiegen bzw. unentbehrlich geworden ist, zeigt sich auch im Selbstverständnis sowie in den vorausgesetzten Kompetenzen (Koby/Melby 2013, S. 184) des professionellen und institutionellen Translators, der sich u.a. auch über den Einsatz spezifischer Translationstechnologie definiert (EC-DGT 2009).

Der notwendige und allgegenwärtige Einsatz von Translationstechnologie bedingt den Umgang mit digitalen Daten, seien dies nun der digitale Ausgangstext, die Produktion des digitalen Zieltextes, der Rückgriff auf digital gespeicherte Daten als Referenzmaterial oder die Produktion neuer digitaler Daten als künftiges Referenzmaterial. Der Translationsprozess kann damit als ein Datenverarbeitungsprozess dargestellt werden. Natürlich bedeutet dies eine reduktionistische Sicht auf das komplexe Handlungsgefüge der Translation, die offensichtlich weit allgemeiner und umfassender als eine „konventionalisierte, interlinguale und transkulturelle Interaktion […], die in einer Kultur als zulässig erachtet wird“ (Prunč 1997: S. 108) definiert werden muss und handlungs- und kommunikationstheoretische, kulturelle, soziologische und ethische Fragen mit einschließt.

Dennoch wollen wir uns an dieser Stelle mit dem spezifischen Aspekt der Datenverarbeitung beschäftigen und Translation als einen Remix bzw. eine kreative Neugestaltung eines Textes verstehen, in dem bestehende Übersetzungsdaten wiederverwendet und neue Übersetzungsdaten erzeugt werden; Translation ist die Produktion eines Zieltextes mithilfe von Übersetzungsdaten und zugleich die Produktion neuer Übersetzungsdaten in der Kombination aus Ausgangstext und Zieltext. Translation bedeutet Variation des Ausgangstextes über Sprachgrenzen hinweg sowie zugleich die Selektion, Rekombination und Adaptation von Daten. Die Auswahl und das Anpassen von zur Verfügung stehenden Daten, um einen neuen Text zu produzieren, schließt ein automatisches Ersetzen durch die Maschine aus und setzt den menschlichen Eingriff durch den Translator voraus. Translation als aktive und bewusste Datenverarbeitung ist niemals automatische Übersetzung. Der englische Begriff der „datafication“ wurde in der aktuellen Diskussion geprägt, um die Bedeutung der Daten („Big Data“) in allen möglichen Bereichen hervorzuheben, und trifft auch in unserem Zusammenhang für die Translation zu.

Was sind nun Übersetzungsdaten bzw. „Translation Data“? Daten, die Übersetzungen speichern, können aufgrund der Länge der abgespeicherten Textsegmente drei sprachlichen Ebenen zugeordnet werden: Auf Wortebene sind dies lexikographische und terminologische Daten, auf Satzebene sprechen wir von Übersetzungsspeicher oder Translation-Memory und auf Textebene vereinen Parallelkorpora Texte in zwei oder mehreren Sprachen. Während in Terminologiedatenbanken einzelne Fachtermini ihren Äquivalenten in einer oder mehreren Sprachen zugeordnet werden und damit die fremdsprachliche Benennung fachlicher Begriffe des Ausgangstextes angeboten wird, werden in Translation-Memories einzelne Textsegmente,

also Absätze oder Sätze, aus Ausgangstext und Übersetzung einander zugeordnet; man spricht dabei von Übersetzungseinheiten bzw. „Translation Units“.

Ausgangstext(-segment)	+	Zieltext(-segment)	+	Äquivalenzrelation

Parallelkorpora sammeln zwar Texte und ihre Übersetzungen, die Zuordnung besteht aber nur zwischen einem Text und seiner Übersetzung, eine Zuordnung auf Satz- oder Wortebene erfolgt nicht. Eine Zuordnung zwischen einem Text bzw. einem Textsegment und seiner Entsprechung in einer anderen Sprache erfolgt damit zwar auf allen drei Ebenen, durchgesetzt hat sich bei den Übersetzungsdaten aber lediglich die Satzebene, vor allem wegen der allgemeineren Wiederverwendbarkeit gegenüber Paralleltexten – je kleiner ein Textsegment, desto wahrscheinlicher ist seine Wiederholung und Wiederverwendbarkeit – sowie der Berücksichtigung der Syntax gegenüber der Terminologie – nicht nur das Fachwort, sondern auch die Formulierung kann übernommen werden.

Zuordnen schließt das Herstellen einer Äquivalenzbeziehung mit ein, wobei diese Beziehung im Grunde eigentlich nur ausdrückt, dass ein Textsegment in einer früheren Übersetzung bereits einmal so übersetzt und gespeichert wurde. In diesem Sinne liegt dem Abspeichern von Übersetzungsdaten ein vereinfachtes Äquivalenzmodell zugrunde: A wurde als B übersetzt, was aber noch lange nicht darüber informiert, in welchen Kontext, zu welchem Zeitpunkt, für welchen kommunikativen Zweck, von wem und für wen, in welcher Textsorte und in welchem Fachbereich A in dieser Weise als B übersetzt wurde. Die in der Translationswissenschaft lange und heftig debattierte Äquivalenzfrage (vgl. Reiß/Vermeer 1985, Koller 1995) und ihre Auswirkung auf die Translation werden dadurch umgangen.

Eine Möglichkeit, die in den Daten angebotenen Äquivalente als Übersetzungen zu kontextualisieren und damit Translation wieder als kontextabhängiges zweck- und zielgerichtetes transkulturelles Handeln zu sehen, bietet das Hinzufügen von Metadaten wie Fachbereich, Datum, Projekt, Autor u.Ä., was aber in der Praxis noch relativ wenig berücksichtigt wird. Dennoch gibt es Forschungsanstrengungen und Vorschläge, die in diese Richtung gehen, beispielsweise das „Internationalization Tag Set“ (ITS) der „W3 Multilingual Web Working Group“ sowie die innerhalb des Austauschformates Translation Memory Exchange (TMX) vorgesehenen Metadatenkategorien oder die im „Structured Translation Specification Set“ (Melby et al. 2011: S. 3) vorgeschlagenen Zusatzinformationen zu Übersetzungsprojekten.

Auf der von North (1998) vorgestellten Wissenstreppe ist zwischen der in den zur Verfügung stehenden Daten enthaltenen Information und dem in einer professionellen Übersetzung zum Ausdruck kommenden Wissen, Können und Handeln eine klare Stufe eingebaut. Auf der unteren Stufe der Daten geht es um spezielle Einzellösungen, auf den oberen Stufen des Wissens um Vernetzung und Anwendungsbezug. Übersetzungsdaten entpersonalisieren zwar den Übersetzungsprozess zumindest teilweise; sie nehmen dem Menschen aber auch eine Aufgabe ab, nämlich das Abrufen bereits gemachter Übersetzungen, das Sich-Erinnern an vorhergehende

Übersetzungen aufgrund eines Vergleichs mit dem aktuellen Ausgangstext. Wir schließen daraus, dass auch bei extensiver Anwendung von Übersetzungsdaten der Beitrag des Menschen fundamental bleibt und gerade in der Kontextualisierung, Rekombination und Adaptation dieser Daten besteht: Übersetzungsdaten stehen damit im klaren Gegensatz zur automatischen Maschinenübersetzung und sind eindeutig der computergestützten Übersetzung zuzurechnen.

Ein Übermaß an Daten bzw. eine Datenflut kann nur dadurch beherrscht werden, dass sie Regeln unterworfen wird, seien dies Regeln zum sicheren Umgang mit den Daten oder Regeln zum Schutz sensibler Daten. Eine große Rolle spielt dabei die Transparenz sowohl der geltenden gesetzlichen Rahmenbedingungen als auch der Datenformate. In diesem Zusammenhang hat sich die Bewegung der offenen Daten (Open Data) herausgebildet, die den freien Zugang zu Daten öffentlichen Interesses fordert.

Open Data

Offene Daten (Open Data) werden anhand von drei wesentlichen Merkmalen definiert. Entscheidend sind:

1. die Abwesenheit jeglicher Zugangsbarrieren (*permission barriers*) wie Preis, Lizenz, Login, etc.;
2. ihre potenzielle Wiederverwendung (*re-use*) für jeden Zweck, der vom Datenproduzenten vorhergesehen wurde oder auch nicht; und
3. ihre strukturierte und maschinenlesbare Form.

Durch die freie Verfüg- und Nutzbarkeit von Daten können diese zum Vorteil der Allgemeinheit weiter bearbeitet und verwendet werden. Open Data stehen dabei in enger Verflechtung mit zahlreichen anderen Open-Initiativen, wie z.B. der Open Source und der freien Software, deren Grundgedanke (Software als freie, digitale Ressource) ihrerseits wiederzufinden ist in der Bewegung zur Open Education (Wissen ist freies Gut). Gemeinsames Ziel ist die Förderung der Entwicklung freier Software und freier Wissensgüter. Open Educational Resources (OER) zeichnen sich dadurch aus, dass sie analog zu offenen Daten frei austauschbar und zugänglich sind für alle. Dazu müssen sie in Formaten bereitgestellt werden, die offene Standards respektieren, sodass sie mit Open Source Software gelesen und idealerweise auch bearbeitet werden können. Insofern ist Open Source und freie Software als kostenlos verfügbares und frei weiterentwickelbares Lehr- und Lernmittel zu sehen. Mit dem Open Society Institute, gegründet 1993 von George Soros, und der darauffolgenden Entwicklung und Förderung von Open Access mit den Erklärungen von Budapest (2002) und Berlin (2003) hat sich auch der offene Zugang zu wissenschaftlichen Quellen und Informationen sowie zu Quellen in Museen, Archiven und Bibliotheken als kulturelles Erbe durchgesetzt.

In diesem Sinne können auch öffentliche Übersetzungen bzw. Übersetzungen, die mit öffentlichen Mitteln finanziert wurden und öffentliche Texte zum Inhalt haben, als öffentliches Gut gesehen werden. Wenn wir diese Übersetzungen als offene Daten auffassen, müssen dafür auch

dieselben Kriterien angewandt werden: Open Translation Data sind frei zugängliche Datenbestände, die Übersetzungen in strukturierter und maschinenlesbarer Form in einem freien Format speichern.

Ein freies Format ist eine veröffentlichte Spezifikation zum Speichern von Daten in digitaler Form und kann ohne rechtliche Einschränkungen genutzt werden. Freie Formate stehen im Gegensatz zu proprietären Formaten, die aus kommerziellen Interessen und durch proprietäre Software implementiert werden. Das wohl beste und am weitesten verbreitete Beispiel für ein freies Format ist das Open Document Format (ODF), das als ISO-Norm veröffentlicht wurde und von zahllosen freien und nicht-freien Programmen unterstützt wird.

Freie Formate für Übersetzungsdaten sind beispielsweise das Translation-Memory-Austauschformat TMX, das dazu gehörende Dateiformat zum Austausch von Segmentierungsregeln SRX (Segmentation Rules Exchange), das Format zum Austausch von Übersetzungsprojekten im Bereich der Softwarelokalisierung XLIFF (XML Localization Interchange File Format), das offene Format zum Austausch von Terminologiedaten TBX (TermBase eXchange) und das offene Format zur unabhängigen und objektiven Feststellung von Übersetzungsvolumen GMX-V (Global information management Metrics eXchange). Alle genannten Formate basieren auf der allgemeinen Mark-up-Sprache XML und wurden als internationale Standards veröffentlicht. Dazu kommt noch das etwas ältere mehrsprachige PO-Format (Portable Objects): Es gehört zum GNU-Gettext-Framework, das dazu dient, Textstellen aus Programmen des GNU/Linux-Systems in eigenen sprachabhängigen Dateien zu sammeln und damit leichter übersetzbar zu machen, wobei die PO-Dateien mehrsprachige Übersetzungsdaten enthalten.

Freie Formate, die einzelne Textstellen mit ihren Übersetzungen enthalten, sind also die folgenden drei: TMX, XLIFF und PO. Während in TMX-Dateien Übersetzungseinheiten bestehend aus Ausgangstextsegment und Zieltextsegment in beliebiger Reihenfolge abgespeichert werden, enthalten XLIFF-Dateien aufgrund der Projektorientierung dieses Formats einzelne Übersetzungseinheiten nach Texten gruppiert; es wird in der Software- und Weblokalisierung hauptsächlich zum Austausch von Projektdaten eingesetzt. Das PO-Format ist auf das freie Betriebssystem GNU/Linux und freie Software beschränkt und erreichte dadurch keine allgemeine Verbreitung. Damit bleibt für das Abspeichern von offenen Übersetzungsdaten vornehmlich das TMX-Format, das sich in der Praxis auch weitgehend durchgesetzt hat.

Zusammenfassend müssen Open Translation Data die folgenden Voraussetzungen erfüllen:

1. frei zugänglich sein bzw. ohne Zugangsbarrieren (Preis, Lizenz, Login, etc.) downloadbar sein;
2. in jeder Anwendungsumgebung das Wiederverwenden von Übersetzungen ermöglichen, d.h. ein Ausgangstextsegment und dessen Übersetzung sowie möglichst Metadaten enthalten;
3. im TMX-Format (seltener auch XLIFF oder PO) vorliegen.

Open Translation Data

In der Praxis können bereits einzelne Beispiele für offene Übersetzungsdaten genannt werden: So z.B. die Daten der Generaldirektion Übersetzung der Europäischen Kommission, die bereits seit mehreren Jahren öffentlich zur Verfügung gestellt werden, aber erst seit Ende 2013 im offenen Datenportal der Europäischen Union bereitgestellt und interessanterweise erst ab diesem Zeitpunkt als Open Data bezeichnet werden.[1] Die amtlichen Übersetzungen des gesamten Europäischen Vertragswerkes (*acquis communautaire*) in allen 23 Amtssprachen können in einzelnen, nach Jahrgängen getrennten Dateien im TMX-Format heruntergeladen und wiederverwendet werden. Das gewünschte Sprachpaar kann wahlweise mit einem kleinen Zusatztool (TMExtract) extrahiert werden, um die Datenmenge zu reduzieren. Eine detailliertere Filterung nach Fachgebieten, Textsorten o.Ä. wäre dabei sehr sinnvoll und würde das Verhältnis zwischen Trefferquote und Datenmenge deutlich erhöhen.

Innerhalb der Europäischen Union werden auch andere Übersetzungsdaten angeboten, wie beispielsweise vom „European Centre for Disease Prevention and Control" (ECDC), das ein spezifischeres Translation-Memory zum Fachbereich öffentliches Gesundheitswesen in 25 Sprachen ebenfalls im TMX-Format anbietet.[2]

Die Vereinten Nationen veröffentlichen die Übersetzungen der Protokolle ihrer Vollversammlungen unter http://www.uncorpora.org/ als TMX-Datei in den sechs offiziellen Sprachen der UNO Englisch, Arabisch, Chinesisch, Spanisch, Französisch und Russisch.

Open Translation Data sind aber nicht nur bei internationalen Organisationen zu finden, sondern auch bei den Übersetzungsdiensten sprachlicher Minderheiten, wo der Bedarf an offiziellen Übersetzungen besonders stark ist. So auch im Baskenland, wo die „Memorias de traducción del Servicio Oficial de Traductores" vom Open-Data-Portal Euskadi im TMX-Format zum Download bereit stehen.[3] Diese Übersetzungsdaten in den beiden Landessprachen Spanisch und Baskisch waren die ersten, die offiziell als Open Data bezeichnet wurden.

Neben den Übersetzungsdaten öffentlicher Einrichtungen bieten auch private Anbieter TMX-Daten an, so z.B. http://mymemory.translated.net/, wobei hier die Grenzen zwischen Maschinenübersetzung und menschlicher Übersetzung verschwimmen, da häufig maschinell übersetzte Textsegmente eingebunden werden. Reine Online-Portale für Translation-Memories[4] können nicht als Open Data deklariert werden, da eine autonome Wiederverwendung in anderen Programmen nicht möglich ist und kein freies Datenformat verwendet wird, auch wenn die Daten öffentlich und kostenlos angeboten werden.

---

1 https://open-data.europa.eu/de/data/dataset/ dgt-translation-memory.
2 http://ipsc.jrc.ec.europa.eu/?id=782.
3 http://opendata.euskadi.net/w79-contdata/es/contenidos/ds_recursos_linguisticos/memorias_traduccion/es_izo/memorias_traduccion_izo.html.
4 http://glosbe.com/tmem/ oder http://www.linguee.com/.

Von einer allgemeinen Datenflut kann im Bereich Übersetzen bisher noch nicht gesprochen werden, auch wenn Daten eine immer wichtigere Rolle spielen. Im Allgemeinen können Daten, und das gilt insbesondere für Übersetzungsdaten, am besten eingesetzt und wiederverwendet werden, je besser sie strukturiert, vernetzt und mit Zusatzinformationen kontextualisierbar gemacht werden. In dieser Hinsicht müssten die angebotenen Daten noch verbessert und mit mehr Zusatzinformationen und Metadaten ausgestattet werden.

Open Translation Data und Mehrsprachigkeit

Werden Übersetzungsdaten genutzt, entstehen daraus Vorteile wie eine Kostenersparnis durch das Wiederverwenden bereits erstellter Übersetzungen, eine Erhöhung der Konsistenz innerhalb der durchgeführten Übersetzungen sowie eine weitgehende Unterstützung von Sprachplanung und Terminologieharmonisierung. Dabei übernimmt die Maschine einen Teil der Übersetzungsarbeit, indem sie sich an bereits in Form von Daten vorliegende Übersetzungen erinnert und diese wiederverwendet. Der Übersetzer wird dadurch entlastet. Kann auf den Menschen durch diesen Beitrag der Maschine verzichtet werden, bzw. können Übersetzungsdaten die Sprachkompetenz des menschlichen Übersetzers ersetzen?

Mehrsprachigkeit wird unterschieden in individuelle Mehrsprachigkeit, wodurch die Sprachkompetenz des Individuums in den Vordergrund gerückt wird, und institutionelle Mehrsprachigkeit, die das Verwenden mehrerer Sprachen innerhalb einer Institution bzw. Gesellschaft (vgl. Edwards 2007: 448) oder auch eines Fachbereichs (vgl. Sandrini 2006: 110) betrifft. Übersetzungsdaten werden mithilfe individueller Sprach- und vor allem Translationskompetenz produziert und bringen diese in neue, spätere Übersetzungsprojekte durch das Wiederverwenden wieder ein. Dies bedeutet, dass vorhandene Übersetzungsdaten für die institutionelle Mehrsprachigkeit von großer Bedeutung sind, da sie die Sprach- und Translationskompetenz des Einzelnen für die Gesellschaft mit allen damit verbundenen Vorteilen verwertbar machen.

Die Erweiterung des Forschungsinteresses der Sprachwissenschaft sowie die technologische Entwicklung der letzten beiden Jahrzehnte haben zu einer neuen Auffassung von Sprache geführt, die Sprache nicht mehr nur bzw. nicht mehr ausschließlich als ein begrenztes, regelbasiertes System sieht, sondern vielmehr als eine repräsentative Sammlung von sprachlichen Daten. Daraus folgt eine veränderte Ausrichtung der Forschungsanstrengungen in Richtung empirischer Sprachverwendungsforschung – Sprache wird gelernt, indem man beobachtet, wie Sprache konkret verwendet wird. Die Bedeutung sprachlicher Korpora bzw. Sammlungen von konkreten Verwendungsbeispielen für die Entwicklung von Sprachkompetenz, für die Translationsdidaktik und für viele weitere Zwecke hat daher enorm zugenommen.

Die Entwicklung der statistischen Maschinenübersetzung basiert auf genau diesen Voraussetzungen: Zum Übersetzen eines Textes werden nicht mehr ein Lexikon und ein abstraktes Regelsystem verwendet, sondern es wird ein zweisprachiges Korpus zugrunde gelegt, aus dem durch Wahrscheinlichkeitsberechnungen die treffendsten Übersetzungslösungen extrahiert und zusammengesetzt werden. Die Gültigkeit dieses Ansatzes unterstreicht der Erfolg von Google-

Translate[5] und Microsofts Bing Translator[6], beides Online-Versionen solcher statistischen Maschinenübersetzungssysteme, aber auch der Erfolg des freien statistischen Maschinenübersetzungssystems Moses[7], das mittlerweile nach Jahren intensiver Finanzierung regelbasierter Systeme auch von der EU eingesetzt wird.[8]

Analog zu dieser Entwicklung kann – wie oben ausgeführt – Translation als ein fallspezifisches Anwenden von Übersetzungsdaten gesehen werden, wobei die Konsistenz und die Effizienz des Übersetzungsprozesses sowie zum Teil auch die Qualität des Zieltextes von den zur Verfügung stehenden Übersetzungsdaten abhängt.

Kommen Open-Translation-Data zur Anwendung, die frei verfügbar sind und für alle Zwecke eingesetzt werden können, bedarf es jedenfalls der Sprach- und Translationskompetenz des Menschen, um beurteilen zu können, ob diese Daten für die neu anzufertigende Übersetzung überhaupt eingesetzt werden können, ob sie zum neuen Kommunikationskontext passen und ob sie der geplanten Verwendung des Zieltextes entsprechen. Darüber hinaus bleibt die menschliche Sprach- und Translationskompetenz unabdingbar zur fallspezifischen Evaluierung des Outputs von Maschinenübersetzungssystemen. Der Bereich des Post-Editings automatisch erstellter Übersetzungen bildet einen neuen, stetig zunehmenden Arbeitsbereich für ausgebildete Übersetzerinnen und Übersetzer.

Insgesamt betrachtet darf die Datenflut keinesfalls Zukunftsängste auslösen, sondern kann bei einer kritischen Betrachtung im Bereich der Translation durchaus auch positiv gesehen werden. Neben den genannten Vorteilen durch frei verfügbare Übersetzungsdaten erschließen sich auch neue Berufsfelder für ausgebildete Translatorinnen und Translatoren, die gerade darin bestehen, mit diesen Daten umzugehen und ihren Einsatz zu planen, etwa in der Konzeption und Planung von Translationsprozessen (Translationsmanagement), in der Planung des Einsatzes und der Adaptation von Translationstechnologie und Translation-Memory-Systemen oder in der Vorbereitung und dem Zusammenstellen von Korpora für statistische Maschinenübersetzungssysteme. In all diesen Fällen kann die Verwendung von offenen Übersetzungsdaten eine große Rolle spielen.

Open Translation Data stellen eine neue Herausforderung für Übersetzerinnen und Übersetzer dar; sie bedeuten aber zugleich auch eine große Chance für die Bewältigung von Mehrsprachigkeit und rücken die gesellschaftliche Bedeutung der Translation sowie den bewussten Umgang mit Translationstechnologie in den Vordergrund. Offene Übersetzungsdaten sind kein Ersatz für Sprach- und Translationskompetenz, sondern setzen diese für einen verantwortungsvollen und bewusst geplanten Einsatz der Daten, insbesondere in der Evaluierung und Kontextualisierung, voraus.

---

5 http://translate.google.com.
6 http://www.bing.com/translator.
7 http://www.statmt.org/moses.
8 MT@EU http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm

Literatur

Burdick, Anne (2012): *Digital humanities*. Cambridge, Mass.: MIT Press.

Choudhury, Rahzeb & McConnell, Brian (2013): *Translation Technology Landscape Report*. TAUS – Translation Automation User Society.

Cnyrim, Andrea; Hagemann, Susanne & Neu Julia (2013): Towards a Framework of Reference for Translation Competence. In: Kiraly, Donald; Hansen-Schirra, S. & Maksymski K. (Hrsg.): *New prospects and perspectives for educating language mediators*. Tübingen: Gunter Narr, S. 9–34.

Cronin, Michael (2013): *Translation in the digital age*. London: Routledge.

EC-DGT (2009): The size of the language industry in the EU. *Studies on Translation and Multilingualism* 1/2009.

Edwards, John (2007): Societal multilingualism: reality, recognition and response. In: Auer, Peter (Hrsg.): *Handbook of multilingualism and multilingual communication*. Berlin: de Gruyter, S. 447–467.

Koller, Werner (1995): The concept of equivalence and the object of Translation Studies. *Target*.

Link, Michael (2013): *Open Access im Wissenschaftsbereich*. Frankfurt am Main: Lang.

Melby, Alan K.; Lommel, Arle; Rasmussen, Nathan & Housley, Jason (2011): The Container Project. Presented at the First International Conference on Terminology, Languages, and Content Resources, Seoul, South Korea, 10–11 June 2011. www.linport.org/downloads/2011-07-container-project.pdf [Stand vom 05-10-2013].

Melby, Alan & Koby Geoffrey (2013): Certification and Job Task Analysis (JTA): Establishing Validity of Translator Certification Examination. *Translation & Interpreting*, 5 (1), S. 174–210.

North, Klaus (1998): *Wissensorientierte Unternehmensführung – Wertschöpfung durch Wissen*. Wiesbaden: Gabler Verlag.

PACTE (2007): Zum Wesen der Übersetzungskompetenz – Grundlagen für die experimentelle Validierung eines ÜK-Modells. In: Wotjak, Gerd (Hrsg.): *Quo vadis Translatologie? Ein halbes Jahrhundert universitäre Ausbildung von Dolmetschern und Übersetzern in Leipzig*. Berlin: Frank & Timme, S. 327–342.

Prunč, Erich (1997): „Translationskultur (Versuch einer konstruktiven Kritik des translatorischen Handelns). *TEXTconTEXT*, 11.1 = NF 1 (2), S. 99–127.

Reiß, Katharina & Vermeer, Hans J. (1984): *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.

Sandrini, Peter (2006): LSP Translation and Globalization. In: Gotti, Maurizio & Šarčević, Susan (Hrsg.): *Insights into Specialized Translation*. Linguistic Insights 46. Bern, Berlin, Frankfurt am Main: Peter Lang, S. 107–120.

Sandrini, Peter (2012): Texttechnologie und Translation. In: Zybatow, L.; Petrova, A. & Ustaszewski M. (Hrsg.): *Translationswissenschaft interdisziplinär: Fragen der Theorie und Didaktik. Tagungsband der 1. Internationalen Konferenz TRANSLATA Translationswissenschaft: gestern – heute – morgen.* Frankfurt am Main [u.a.]: Peter Lang. S. 21–33.

Zetzsche, Jost (2014): *The Translator's Toolbox. A Computer Primer for Translators*. International Writers' Group, LLC.