

## Vorwort

---

Maschinelle Übersetzung (MÜ) ist in den vergangenen Jahren in der Sprachenindustrie wieder in den Mittelpunkt des Interesses gerückt. Nicht neue bahnbrechende technologische Ansätze, die eine bessere Qualität der maschinell übersetzten Texte liefern, sondern die Verfügbarkeit großer Datenmengen als Trainingskorpora für statistische Systeme sowie die Integration von maschineller Übersetzung in vorhandene Arbeitsabläufe versprechen sowohl verbesserte Kommunikationsabläufe in global agierenden Unternehmen als auch zeitnähere und kostengünstigere Übersetzungen selbst an Stellen, an denen hochwertige professionelle Übersetzungsqualität gefordert ist.

Bill Gates äußerte 2005 die Ansicht, maschinelle Übersetzung von Texten aus dem Computerbereich erreiche bereits die Qualität humanübersetzter Texte dieses Fachgebiets (vgl. [Gates(2005)]), eine These, die allzu leicht widerlegt werden kann. Denn wer Anfang 2009 Hilfe aus der Microsoft Wissensdatenbank über den *Release Candidate 1* des *Internet Explorer 8* in deutscher Sprache benötigte, erhielt einen maschinell übersetzten Text, dessen Qualität dem Ratsuchenden kaum weiterhalf (vgl. [Ries(2009)]). Dennoch sollten derartige Beispiele aber nicht dazu herangezogen werden, um über die grundsätzliche Eignung maschineller Übersetzung in bestimmten Bereichen zu urteilen, wie zahlreiche Einsatzgebiete in Unternehmen zeigen. Auch die Lokalisierungsindustrie setzt seit einiger Zeit auf Maschinelle Übersetzung. Denn durch die Integration in bestehende Übersetzungsabläufe, in denen Translation-Memory-Technologie eingesetzt wird, birgt Maschinelle Übersetzung Einsparpotentiale. Voraussetzung hierfür ist allerdings, dass das MÜ-System auf die Unternehmens- bzw. Fachgebietsterminologie zugreift und Texte verarbeitet werden, die hinsichtlich ihrer syntaktischen Struktur keinen zu hohen Komplexitätsgrad aufweisen. Entsprechend dieser Erwartung veröffentlichen Anbieter von Translation-Memory-Technologie ebenso wie Entwickler von Übersetzungsprogrammen zunehmend Schnittstellen zu verschiedenen Anwendungen: Anbieter von MÜ-Software bieten vermehrt die Integration in Office-Produkte und E-Mail-Dienste an, während Translation-Memory-Hersteller mit Entwicklern von MÜ-Systemen kooperieren oder ihre Schnittstellen zur Integration entsprechender Systeme bereitstellen.

Doch nicht nur die Integration von MÜ in vorhandene Übersetzungsabläufe, auch die Integration verschiedener Ansätze im Bereich der MÜ versprechen Qualitätszugewinne, so dass sich Maschinelle Übersetzung gerade auch im akademischen Umfeld wieder zu einem viel beachteten Gegenstand der computerlinguistischen Forschung entwickelt hat. Hier stehen die verschiedenen Ansätze, regelbasierte MÜ (*Rule Based Machine Translation*, RBMT), statistische MÜ (*Statistical Machine Translation*, SMT) und beispielbasierte MÜ (*Example Based Machine Translation*, EBMT) einander gegenüber. Vor

allem die Verfügbarkeit großer paralleler Korpora, die statistischen MÜ-Systemen als Trainingsdatenbasis dienen, haben dazu beigetragen, dass statistische MÜ gegenüber den in ihrem Entwicklungsaufwand sehr kostspieligen regelbasierten Systemen für Entwickler und Anwender interessant geworden ist. So basiert denn auch das von Google veröffentlichte Übersetzungsangebot Google Translate im Internet auf einem statistischen Ansatz. Als Trainingsdaten dienten hier große Dokumentenmengen der Vereinten Nationen (UN) in den offiziellen Sprachen der UN und Dokumente der Europäischen Union. Für Sprachen, in denen keine großen Datenmengen paralleler Texte vorliegen, verspricht dieses Verfahren allerdings weit weniger schnell zufriedenstellende Ergebnisse, so dass bereits Ansätze einer Integration der verschiedenen MÜ-Methoden erprobt werden.

Maschinelle Übersetzung ist heute in global agierenden Unternehmen wie Volkswagen (vgl. [Porsiel(2008a)] und [Porsiel(2008b)]) für die standortübergreifende weltweite Kommunikation zu einem integrativen Bestandteil der Unternehmenskommunikation geworden. Aus Gründen der Datensicherheit vertrauen solche Unternehmen aber vielfach nicht auf kostenfreie Internetangebote, sondern setzen auf hauseigene Implementierungen, die dem Markt ebenfalls neue Impulse verleihen.

Die im vorliegenden Band unter dem Titel „Maschinelle Übersetzung – Von der Theorie zur Anwendung“ versammelten Beiträge basieren auf Vorträgen, die im Rahmen des gleichnamigen Workshops des Arbeitskreises „Maschinelle Übersetzung“ der GSCL im Juni 2008 an der Hochschule Anhalt in Köthen gehalten wurden. Einleitend beschreibt Daniel Stein in seinem Beitrag „Maschinelle Übersetzung – ein Überblick“ die historische Entwicklung der verschiedenen Ansätze der MÜ. Einen Einblick in die jüngsten Entwicklungen der Integration von computerunterstützter Übersetzung mithilfe von Translation-Memory-Technologie und Maschinellem Übersetzung gewährt Dino Azzano in seinem Beitrag „CAT und MÜ – getrennte Welten?“. Kurt Eberle illustriert in seinem Beitrag „Integration von regel- und statistikbasierten Methoden in der Maschinellen Übersetzung“, wie regelbasierte mit statistischen Verfahren kombiniert und zur Auflösung linguistischer Mehrdeutigkeiten erfolgversprechend eingesetzt werden können. Unter dem Titel „METIS-II: Low-Resource MT for German to English“ erläutert Michael Carl am Beispiel der Implementierung der Übersetzungsrichtung Deutsch-Englisch die Prinzipien der im METIS-II-Projekt implementierten Methoden der maschinellen Übersetzung. Wie bereits METIS-I zielte das Projekt METIS-II darauf ab, maschinelle Übersetzungen auf der Grundlage einsprachiger Textkorpora mit getaggten und lemmatisierten Texten der Zielsprache und zweisprachiger Lexika in den jeweiligen Sprachen der für die MÜ gewünschten Übersetzungsrichtungen zu ermöglichen. Dieser auf Pattern-Matching-Methoden beruhende beispielbasierte Ansatz ist vor allem für Sprachpaare interessant, für die keine großen parallelen Korpora verfügbar sind. Der Beitrag „The Automatic Translation of Film Subtitles. A Machine Translation Success Story“ von Martin Volk zeigt, dass sich eine Textsorte wie Filmuntertitel für die Über-

setzung mit einem MÜ-System eignet und für das Sprachpaar Schwedisch-Dänisch eindrucksvolle Ergebnisse liefert. Um die Eignung einer anderen Textsorte, allerdings für die computergestützte Übersetzung, geht es im letzten Beitrag dieses Bandes. Heribert Härtinger präsentiert die Ergebnisse einer Untersuchung von Patentschriften und deren Eignung für die Übersetzung mit Translation-Memory-Systemen auf der Grundlage eines Korpus deutscher und spanischer Patentschriften.

Bei der Gegenüberstellung von Übersetzungsergebnissen verschiedener MÜ-Verfahren mit von Humanübersetzern übersetzten Texten – wie im Beitrag von Michael Carl und Martin Volk erwähnt – rücken auch Methoden der Evaluierung, die automatisiert erfolgen und in ihren Ergebnissen möglicherweise über bisherige Verfahren wie das von IBM entwickelte Verfahren BLEU ([Papineni(2002)]) oder NIST hinausgehen, in das Interesse der MÜ-Forschung und dürften hier auch künftig von Bedeutung sein.

### Literatur

- [Gates(2005)] Gates, Bill. "Remarks by Bill Gates, Chairman and Chief Software Architect.", 2005, Accessed 25.03.09. <http://www.microsoft.com/presspass/exec/billg/speeches/2005/10-14Princeton.aspx>.
- [Papineni(2002)] Papineni, Kishore. "BLEU: a method for automatic evaluation of machine translation." In *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics*. 2002, 311–318.
- [Porsiel(2008a)] Porsiel, Jörg. "Machine translation at Volkswagen: a case study." In *Multilingual*. 2008a, 58–61.
- [Porsiel(2008b)] ———. "Maschinelle Übersetzung bei Volkswagen. Sprache als betriebswirtschaftlicher Faktor." *MDÜ* 5: (2008b) 44–47.
- [Ries(2009)] Ries, Uli. "Übelsetzt: Microsoft verstört IE8-Nutzer mit wirrem Hilfstext.", 2009. [http://www.computerzeitung.de/articles/uebelsetzt\\_microsoft\\_verstoert\\_ie8-nutzer\\_mit\\_wirrem\\_hilfstext:/2009006/31815447\\_ha\\_CZ.html?thes=](http://www.computerzeitung.de/articles/uebelsetzt_microsoft_verstoert_ie8-nutzer_mit_wirrem_hilfstext:/2009006/31815447_ha_CZ.html?thes=).