

# UNSERE SPRACHEN IN GEFAHR: EINE HERAUSFORDERUNG FÜR DIE SPRACHTECHNOLOGIE

Wir sind Zeugen einer digitalen Revolution, die enorme Auswirkungen auf Kommunikation und Gesellschaft hat – die jüngsten Entwicklungen in der digitalen Informations- und Kommunikationstechnologie werden oft mit Gutenbergs Erfindung der Druckerpresse verglichen. Was sagt uns diese Analogie über die Zukunft der europäischen Informationsgesellschaft und insbesondere der unserer Sprachen?

---

Die digitale Revolution ist mit der Erfindung der Druckerpresse vergleichbar.

---

Nach Gutenbergs Erfindung wurden echte Durchbrüche im Kommunikations- und Wissensaustausch erzielt, z. B. durch Luthers Bibelübersetzung in die Landessprache. In den folgenden Jahrhunderten wurden Kulturtechniken entwickelt, die die sprachliche Kommunikation und den Wissensaustausch erleichterten:

- Die orthografische und grammatikalische Standardisierung großer Sprachen ermöglichte die schnelle Verbreitung neuer wissenschaftlicher und intellektueller Ideen.
- Die Entwicklung von Amtssprachen ermöglichte es den Bürgern, innerhalb oftmals politischer Grenzen miteinander zu kommunizieren.
- Sprachübergreifender Austausch wurde durch Lehren und Übersetzen von Sprachen möglich.

- Die Schaffung redaktioneller und bibliografischer Richtlinien stellte die Qualität und Verfügbarkeit von Druckmaterial sicher.
- Die Entwicklung von Medien wie Buch, Zeitung, Radio und Fernsehen befriedigte ganz unterschiedliche Kommunikationsbedürfnisse.

In den vergangenen zwanzig Jahren trug die Informationstechnologie zu einer Automatisierung und Vereinfachung zahlreicher Prozesse bei:

- Desktop-Publishing-Software ist an die Stelle von Schreibmaschine und Textsatz getreten.
- Microsoft PowerPoint ersetzte Overhead-Folien.
- Per E-Mail lassen sich Dokumente schneller versenden und empfangen als mit einem Faxgerät.
- Skype ermöglicht preiswerte Telefonanrufe über das Internet und virtuelle Konferenzen.
- Digitale Audio- und Video-Formate erleichtern den Austausch von Multimediainhalten.
- Suchmaschinen bieten Zugriff auf Webseiten per Stichwortsuche.
- Onlinedienste wie Google Translate erzeugen schnell Grobübersetzungen.
- Soziale Medien wie Facebook und Twitter erleichtern Kommunikation und Informationsaustausch.

All diese Anwendungen sind zweifellos hilfreich, aber noch lange nicht ausreichend, um eine nachhaltige,

mehrsprachige europäische Gesellschaft zu unterstützen, in der Informationen und Wissen so ungehindert fließen können wie heute schon Waren und Kapital.

## 2.1 SPRACHGRENZEN BREMSEN DIE EUROPÄISCHE INFORMATIONSGESELLSCHAFT

Wir können nicht genau vorhersagen, wie die Informationsgesellschaft der Zukunft aussehen wird. Es ist jedoch sehr wahrscheinlich, dass die Revolution in der Kommunikationstechnologie Menschen mit unterschiedlichen Sprachen auf eine ganz neue Weise zusammenbringen wird. Dadurch ist der Einzelne angehalten, neue Sprachen zu lernen. Insbesondere sind aber Entwickler gefragt, neue Technologien zu schaffen, die das gegenseitige Verstehen und den Zugriff auf gemeinsam nutzbares Wissen sicherstellen.

---

Der globale Wirtschafts- und Informationsraum konfrontiert uns mit mehr Sprachen, Sprechern und Inhalten.

---

Im globalen Wirtschafts- und Informationsraum sorgen digitale Medien dafür, dass immer mehr Menschen, Sprachen und Inhalte an der immer schneller werdenden Kommunikation beteiligt sind. Die Popularität von Plattformen wie Wikipedia, Facebook, Twitter und Google+ ist nur die Spitze des Eisbergs.

Heute können wir in Sekunden gigabyte Weise Textmenügen um die Welt schicken, bevor wir überhaupt merken, dass sie in einer Sprache sind, die wir gar nicht verstehen. Einem Bericht der Europäischen Kommission zufolge erwerben nur 57% der Internetnutzer in Europa Waren und Dienstleistungen über Online-Shops in Sprachen, die nicht ihre Muttersprache sind – von einem gemeinsamen digitalen Markt innerhalb der Europäischen Union kann keine Rede sein! Englisch ist

hierbei die gängigste Fremdsprache, gefolgt von Französisch, Deutsch und Spanisch. 55% der Nutzer lesen Inhalte in einer Fremdsprache, und 35% verwenden eine fremde Sprache, um E-Mails zu schreiben oder online Kommentare zu posten [2]. Vor ein paar Jahren mag Englisch noch die lingua franca des Webs gewesen sein – die meisten Webinhalte waren auf Englisch verfasst – doch das hat sich dramatisch gewandelt. Der Anteil an Onlineinhalten in anderen europäischen Sprachen (wie auch in Sprachen aus Nahost und dem restlichen Asien) hat explosionsartig zugenommen.

Erstaunlicherweise hat diese allgegenwärtige, durch Sprachgrenzen bedingte digitale Kluft bisher kaum öffentliche Aufmerksamkeit erhalten. Es drängt sich eine Frage auf: Welche europäischen Sprachen gewinnen in der vernetzten Informationsgesellschaft an Bedeutung und welche sind dem Untergang geweiht?

## 2.2 UNSERE SPRACHEN IN GEFAHR

Die Druckerpresse hat den Informationsaustausch in Europa zwar mit vorangetrieben, gleichzeitig aber auch zum Aussterben vieler europäischer Sprachen geführt. Regionale Sprachen und Minderheitssprachen fanden sich kaum in Druckform. Sprachen wie Kornisch und Dalmatisch waren auf mündliche Übertragung begrenzt und damit in ihrer Reichweite stark beschränkt. Wird das Internet die gleiche Auswirkung auf unsere modernen Sprachen haben?

---

Die breite Sprachenvielfalt in Europa ist eines unserer reichsten und wichtigsten Kulturgüter.

---

Europas rund 80 Sprachen sind eines unserer reichsten Kulturgüter und ein wichtiger Teil des einzigartigen gesellschaftlichen Modells der europäischen Integration [3]. Sprachen wie Englisch und Spanisch werden auf

dem aufstrebenden digitalen Marktplatz wohl überleben, doch zahlreiche europäische Sprachen könnten in der vernetzten Gesellschaft schnell unbedeutend werden. Dies würde Europas Ruf schwächen und dem strategischen Ziel widersprechen, allen europäischen Bürgern ungeachtet ihrer Sprache gleiche Partizipationsmöglichkeiten zuzusichern.

Einem Bericht der UNESCO zufolge ist Sprache ein wichtiges Medium, um grundlegende Rechte wie politische Meinungsäußerung, Bildung und gesellschaftliche Beteiligung wahrnehmen zu können [4].

## 2.3 SPRACHTECHNOLOGIE IST EINE SCHLÜSSELTECHNOLOGIE

Früher konzentrierten sich Investitionen in den Sprach-erhalt vor allem auf Sprachausbildung und Übersetzungen. Im Jahr 2008 belief sich der europäische Markt für Übersetzungen, Dolmetschen, Softwarelokalisierung und Website-Globalisierung auf 8,4 Mrd. Euro mit einem geschätzten jährlichen Wachstum von 10% [5]. Diese Zahl deckt jedoch nur einen geringen Teil unseres aktuellen und künftigen Kommunikationsbedarfs zwischen den Sprachen ab. Die einzige Lösung, den Sprachgebrauch im Europa von morgen in voller Breite und Tiefe sicherzustellen, ist der Einsatz geeigneter Technologie – so wie wir auch Technologie einsetzen, um unseren Transport- und Energiebedarf zu decken.

Sprachtechnologie für alle Formen geschriebener und gesprochener Sprache ermöglicht es den Menschen, ungeachtet von Sprachbarrieren und Computerfertigkeiten zusammenzuarbeiten, Geschäfte zu tätigen, Wissen auszutauschen und sich an sozialen und politischen Debatten zu beteiligen. Diese Technologie arbeitet oft unsichtbar in komplexen Softwaresystemen und unterstützt schon heute:

- Informationssuche mit Suchmaschinen
- Rechtschreib- und Grammatikprüfung

- Produktempfehlungen in Online-Shops
- Sprachanweisungen von Navigationssystemen
- Online-Übersetzung von Webseiten

Sprachtechnologie besteht aus einer Reihe von Kerntechnologien, die Einsatz in größerer Anwendungssoftware finden. Die META-NET Weißbücher zeigen auf, wie gut die Kerntechnologien bereits für die einzelnen europäischen Sprachen einsetzbar sind.

---

Europa benötigt für alle Sprachen robuste und preiswerte Sprachtechnologie.

---

Damit Europa seine Position an vorderster Front der globalen Innovation behaupten kann, benötigt es eine auf alle europäischen Sprachen abgestimmte Sprachtechnologie, die robust und preiswert ist und sich in größere Softwareumgebungen einbetten lässt. Ohne Sprachtechnologie können wir in der Zukunft keine multimedialen und mehrsprachigen Endanwendungen für alle europäischen Nutzer realisieren.

## 2.4 CHANCEN FÜR DIE SPRACHTECHNOLOGIE

In der Welt des Druckes bestand der technologische Durchbruch in der raschen Duplizierung eines Textbildes mittels einer ausreichend leistungsstarken Druckerpresse. Dem Menschen kam die mühsame Aufgabe zu, Informationen und Wissen zu suchen, zu bewerten, zu übersetzen und zusammenzufassen. Wir mussten bis Edison warten, um gesprochene Sprache aufnehmen zu können – aber auch seine Technologie fertigte nur analoge Kopien an.

Mithilfe von Sprachtechnologie lassen sich heute auch die Prozesse der Übersetzung, der Inhaltsproduktion

und des Wissensmanagements vereinfachen und automatisieren. Auch können mit Sprachtechnologie intuitive sprachbasierte Schnittstellen für Haushaltselektronik, Maschinen, Fahrzeuge, Computer und Roboter entwickelt werden. Gewerbliche und industrielle Anwendungen für den praktischen Einsatz befinden sich in einer frühen Entwicklungsstufe. Dennoch bieten die bisherigen Erfolge in Forschung und Entwicklung eine einmalige Chance. In bestimmten Domänen funktioniert die maschinelle Übersetzung z. B. bereits relativ präzise. Experimentelle Anwendungen ermöglichen Informations- und Wissensverwaltung sowie die Produktion von Inhalten in mehreren europäischen Sprachen.

Wie bei fast allen Technologien wurden die ersten Sprachanwendungen wie Sprachkommando- und -dialogsysteme für hoch spezialisierte Anwendungen entwickelt und wiesen oftmals nur begrenzte Leistungsfähigkeit auf. In Bereichen wie dem Bildungsbereich und der Unterhaltungsbranche bestehen jedoch ungeahnte Marktchancen für die Integration von Sprachtechnologien in Spiele, Edutainment-Pakete, Bibliotheken, Simulationsumgebungen und Schulungsprogramme. Mobile Informationsdienste, computergestützte Sprachlernsoftware, e-Learning-Umgebungen, Selbstbewertungstools und Plagiatserkennungssoftware sind nur einige der Anwendungsbereiche, in denen Sprachtechnologie eine wichtige Rolle spielen wird. Die Popularität von Social Media-Anwendungen wie Twitter und Facebook macht einen weiteren Bedarf an hochentwickelten Sprachtechnologien deutlich, die Postings überwachen, Diskussionen zusammenfassen, Meinungstrends aufspüren, auf emotionale Reaktionen aufmerksam machen, Urheberrechtsverletzungen aufdecken oder Missbrauch nachverfolgen können.

Zudem sei noch der Einsatz von Sprachtechnologie bei Rettungaktionen in Katastrophengebieten genannt, wo deren Leistungsfähigkeit über Leben und Tod entschei-

den kann. Hard- und Softwaresysteme für die Kommunikation und die Einsatzkoordination mit integrierten Übersetzungsfunktionen und intelligente mehrsprachige Roboter können Leben retten.

---

Sprachtechnologie hilft, das „Handicap“  
Sprachenvielfalt zu überwinden.

---

Die Sprachtechnologie stellt für die Europäische Union eine ungeheure Chance dar. Durch sie lässt sich das komplexe Thema Mehrsprachigkeit in Europa bewältigen – die Realität der selbstverständlichen Koexistenz verschiedener Sprachen in Unternehmen, Organisationen und Bildungseinrichtungen. Auf dem gemeinsamen Markt müssen die Europäer über Sprachgrenzen hinweg frei kommunizieren können.

Sprachtechnologie kann dabei helfen, diese letzte Barriere zu überwinden, und dabei gleichzeitig den freien und offenen Gebrauch der einzelnen Sprachen wirksam unterstützen. Blicken wir noch weiter in die Zukunft, so wird innovative europäische Sprachtechnologie unseren globalen Partnern als Maßstab dienen, wenn sie beginnen, ihre eigenen mehrsprachigen Gesellschaften technologisch zu unterstützen. Sprachtechnologie kann als technisches Hilfsmittel angesehen werden, um das „Handicap“ Sprachenvielfalt zu überwinden und die Sprachgemeinschaften für einander zu öffnen.

## 2.5 HERAUSFORDERUNGEN FÜR DIE SPRACHTECHNOLOGIE

In den vergangenen Jahren hat die Sprachtechnologie zwar erhebliche Fortschritte gemacht, doch das Tempo des technologischen Fortschritts und der Produktinnovationen ist noch viel zu gering. Weit verbreitete Technologien wie Rechtschreib- und Grammatikkorrektur in Textverarbeitungsprogrammen sind üblicherweise einsprachig und stehen nur für eine Handvoll

Sprachen zur Verfügung. Mit Online-Übersetzungsdiensten lässt sich ein Dokument zwar schnell und grob übersetzen, doch wenn es um genaue und vollständige Übersetzungen geht, versagt diese Technologie.

---

Der technologische Fortschritt  
muss beschleunigt werden.

---

Wegen der Komplexität der menschlichen Sprache ist deren Modellierung in Software und der Einsatz solcher Programme in der Praxis ein langwieriges und kostspieliges Unterfangen. Europa hat bei der technologischen Bewältigung der Herausforderungen durch seine Vielsprachigkeit eine Pionierrolle übernommen. Die kann und muss es behaupten, indem es neue Methoden entwickelt, um den Fortschritt der Sprachtechnologie an vielen Stellen zu beschleunigen. Dazu gehören sowohl Fortschritte in der Verarbeitung als auch neue Forschungsmethoden wie z. B. das Crowdsourcing.

## 2.6 SPRACHERWERB BEI MENSCH UND MASCHINE

Um zu veranschaulichen, wie Computer Sprache verarbeiten und warum es so schwierig ist, sie so zu programmieren, dass sie verschiedene Sprachen verarbeiten können, wollen wir uns kurz ansehen, wie sich Menschen Erst- und Zweitsprachen aneignen und dann skizzieren, wie Sprachtechnologiesysteme funktionieren.

---

Der Mensch eignet sich Sprachfertigkeiten durch  
das Lernen anhand von Beispielen sowie  
zugrunde liegender Sprachregeln an.

---

Der Mensch erlernt Sprachfertigkeiten auf zweierlei Weise. Babys eignen sich Sprache an, indem sie den Unterhaltungen zwischen ihren Eltern, Geschwistern und

anderen Familienmitgliedern zuhören. Ab einem Alter von etwa zwei Jahren bringen Kinder ihre ersten Worte und kurze Sätze hervor. Das ist nur möglich, weil der Mensch eine genetische Veranlagung besitzt, das Gehörte zu imitieren und dann zu begreifen. Das spätere Erlernen einer zweiten Sprache erfordert deutlich mehr Anstrengungen, vor allem, wenn das Kind nicht in eine Sprachgemeinschaft von Muttersprachlern eingebettet ist. In der Schule eignet man sich Fremdsprachen an, indem grammatikalische Strukturen, Vokabeln und Rechtschreibung mithilfe von Übungen erlernt werden. Dabei wird linguistisches Wissen anhand von abstrakten Regeln, Tabellen und Beispielen erläutert.

Die beiden Hauptarten von Sprachtechnologiesystemen „erwerben“ ihre Sprachkompetenz auf ähnliche Weise. Statistische (oder datengetriebene) Ansätze beziehen Sprachkenntnisse aus riesigen Sammlungen konkreter Beispieltex-te. Zum Trainieren etwa einer Rechtschreibprüfung reichen Texte in einer einzigen Sprache, für maschinelle Übersetzungssysteme müssen parallele Texte in mindestens zwei Sprachen zur Verfügung stehen. Mit Algorithmen des maschinellen Lernens werden in diesen Paralleltexten dann Muster erkannt, die bestimmen, wie Wörter, Wortgruppen und komplette Sätze übersetzt werden.

Dieser statistische Ansatz kann Millionen von Trainingssätzen erfordern, und die Leistungsqualität erhöht sich mit der Menge des analysierten Textmaterials. Dies ist ein Grund, warum Anbieter von Suchmaschinentech-nologien möglichst viel Textmaterial sammeln. Die Rechtschreibkorrektur in Textverarbeitungssystemen und Dienste wie Google Search und Google Translate setzen alle auf statistische Ansätze. Der große Vorteil dieses Verfahrens liegt darin, dass die Maschine in ständigen Trainingszyklen schnell lernt, wobei die Qualität der Ergebnisse sehr unterschiedlich sein kann.

Den zweiten Ansatz in der Sprachtechnologie und der maschinellen Übersetzung bilden regelbasierte Systeme,

bei denen Experten aus Linguistik, Computerlinguistik und Informatik manuell grammatikalische Analysen (Übersetzungsregeln) kodieren und Vokabellisten erstellen. Das ist sehr zeitaufwändig und arbeitsintensiv. Einige der führenden regelbasierten Übersetzungssysteme werden bereits seit über dreißig Jahren ständig weiterentwickelt. Der Vorteil regelbasierter Systeme besteht darin, dass die Sprachverarbeitung von den Fachleuten genau kontrolliert werden kann. So lassen sich Softwarefehler systematisch korrigieren, und der Nutzer erhält eine ausführliche Rückmeldung, z. B. wenn regelbasierte Systeme zum Sprachenlernen genutzt werden. Da dieser Ansatz jedoch sehr kostspielig ist, wurde regelbasierte Sprachtechnologie bisher nur für wenige meist große Sprachen entwickelt.

Da die Stärken und Schwächen statistischer und regelbasierter Systeme oft komplementär sind, konzentriert sich die Forschung heute auf hybride Ansätze, die die

beiden Methoden kombinieren. Bei industriellen Anwendungen haben sich diese Ansätze bisher jedoch als weniger erfolgreich erwiesen als im Forschungslabor.

Auch wenn die Sprachtechnologie in den vergangenen Jahren erhebliche Fortschritte gemacht hat, besteht, zusammenfassend, nach wie vor enormer Verbesserungsbedarf in puncto Qualität, Robustheit und Abdeckung der sprachtechnologischen Systeme und Methoden. Das gilt insbesondere angesichts der Tatsache, dass viele der in der heutigen Informationsgesellschaft auf breiter Ebene eingesetzten Anwendungen stark von Sprachtechnologie abhängig sind – vor allem im europäischen Wirtschafts- und Informationsraum. Im nächsten Kapitel werden wir die Rolle des Deutschen in der europäischen Informationsgesellschaft genauer unter die Lupe nehmen und den aktuellen Stand der Sprachtechnologie für die deutsche Sprache bewerten.

The German Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 81 p. 26 illus., 24 illus. in color., Softcover

ISBN: 978-3-642-27165-6