# Sprachbarrieren durchbrechen: Traum oder Wirklichkeit?

Alexander Waibel (Karlsruhe)

## Zusammenfassung

In einer zunehmend vernetzten Welt trennen nicht mehr kommunikationstechnische Barrieren („digital divide“) die Menschen voneinander, sondern linguistische („language divide“). Bereits 80 % der Weltbevölkerung haben ein mobiles Telefon, und so kann fast jeder Erdenbürger fast jeden erreichen. Jedoch können die wenigsten einander verstehen bei über 6000 Sprachen plus zahlreichen Dialekten und Akzenten. Menschliche Dolmetscher können – wenn sie auch qualitativ besser sind – diesen wachsenden Kommunikationsbedarf nicht mehr alleine decken, und so sind technische Hilfsmittel unabdingbar. Automatische Übersetzung nicht nur von Texten, sondern auch von gesprochener Sprache wird nun möglich. Seit den späten 1980er Jahren widmet sich unsere Forschung dieser Herausforderung. Aus anfänglichen Prototypen, die noch langsam, schwerfällig und domäneneingeschränkt arbeiteten, sind nun Systeme erwachsen, die als ernstzunehmende Verständigungswerkzeuge im Alltag ihren Dienst aufnehmen. Tragbare Taschendolmetscher für Touristen, Dialogübersetzer für Ärzte und Hilfsorganisationen, automatische Simultanübersetzer, die Vorlesungen, TV-Nachrichten oder Reden in Echtzeit dolmetschen, werden verfügbar als Apps auf Smartphones oder über cloudbbasierte Dienste.

Dieser Beitrag erläutert die Herausforderung des Computerdolmetschens, beschreibt die Technologien, die hierbei zum Einsatz kommen, und die Entwicklungsphasen und Einsatzgebiete aktueller Systeme und Dienste.



## Abstract

In an increasingly connected world, the inhabitants of our planet are no longer separated by lack of digital communication technology (the “digital divide”) but by the language barriers between us. With 6,000 languages, not to mention numerous dialects and accents, we can reach everyone, but understand only few of our fellow humans. Human interpreters alone – even though qualitatively still better – cannot cover the growing demand, calling for the assistance from automatic machine interpretation technology. Since the late 1980s our research has been devoted to the problem of translation and not only of text, but also of spoken language. Initial prototypes were still slow, domain limited and inflexible, but have expanded into practical deployable systems today. Mobile pocket interpreters for tourists, dialog interpreters for doctors and for healthcare providers in humanitarian missions, automatic real-time simultaneous translators for lectures, TV-broadcasts, and speeches are possible and become available as cloud-based services.

In this chapter, we explain the scientific challenges of computer interpretation of spoken language, present the technologies that respond to these challenges, and we discuss the development phases and deployments of today’s cross-lingual language communication systems and services.

## 1. Einführung

In einer zunehmend vernetzten Welt ist digitale Kommunikation schon lange nicht mehr die größte Herausforderung, die Menschen voneinander trennt. Die digitalisierte Welt ist bereits durchweg vernetzt, und auch in entlegenen Teilen der Welt wird durch den Gebrauch von Mobiltelefonen fast jeder Erdenbürger erreichbar. Mehr als sechs Milliarden Mobiltelefone sind bereits in Gebrauch, also im Durchschnitt rund ein Telefon pro Erdenbürger. Besonders in Entwicklungsländern wird die Anbindung an Telefonie und Internet als Voraussetzung für eine bessere Zukunft und ein wirtschaftliches Überleben gesehen. Sicher sind viele Regionen der Welt im Hinblick auf Internet und Telefonie noch nicht erschlossen. Aber sinkende Kosten und der Wille des Einzelnen, am Netz teilzuhaben, treiben diesen Prozess mit Eile voran. Es ist rasch möglich, auch an entlegenen Orten der Welt Kommunikationsantennen aufzubauen oder über Satelittenfunk sowie Telefon mit jedem Menschen auf diesem Planeten zu sprechen oder sich auszutauschen. Die Internetgiganten Google, Facebook und andere wetteifern daher bereits auch in fern abgelegenen Regionen der Welt mit Drohnen, Ballons und ähnlichem darum, diese als Erste zu erschließen.

In diesem Umfeld und mit rasch voranschreitender Globalisierung weicht der sogenannte „Digital Divide" daher auch schnell dem „Linguistic Divide" als dem größeren der Kommunikationsprobleme zwischen den Menschen. Die bleibende babylonische Sprachverwirrung erhält auch weiterhin Kommunikationsbarrieren zwischen uns: Wir können zwar jeden Menschen auf dem Planeten erreichen, aber wir verstehen ihn nicht. Rund 6000 bis 7000 Sprachen gibt es auf der Welt. Man bräuchte also über 36 Millionen Übersetzungsrichtungen, um jeden mit jedem kommunizieren zu lassen. Bereits in der Europäischen Union (bei 24 offiziellen Sprachen) gibt es nicht für jedes Sprachenpaar Übersetzer, die beider Sprachen mächtig sind. Auf diese Problematik werden hauptsächlich drei Antworten angeboten:

– mehr Sprachen lernen;
– Englisch (oder Latein?, Chinesisch?…) als gemeinsame Lingua franca nutzen;
– menschliche Übersetzer oder Dolmetscher als Bindeglied einsetzen.

Die erste davon ist wichtig und kulturell gesund, aber mühsam, und realistisch können in der Regel nur einige wenige Sprachpaare persönlich und individuell beherrscht werden. Glücklich, aber rar sind multilinguale Sprecher, die drei, vier oder gar fünf Sprachen fließend beherrschen.

Die zweite Lösung, alle lernen Englisch, ist auch unrealistisch und kulturell bedenklich: Sollten Menschen und Nationen tatsächlich die kulturelle Verschiedenheit, Individualität und Eigenart ihrer Sprache aufgeben? Sollte ein Land langfristig seine eigene Literatur nur noch in Übersetzungen lesen können? Aber abgesehen vom kulturellen Verlust ist diese Lösung auch unrealistisch. Denn auf welche Sprache sollte man sich einigen? Und warum sollte man sich auf Englisch einigen? Warum nicht Französisch, Spanisch, Chinesisch oder Latein? Und wie sollte man gleiche sprachliche Kompetenz für alle sichern können, um nicht soziale Barrieren durch linguistische auszulösen. Allein in Europa, wo die Sprachausbildung vergleichbar gut gefördert und schulisch gefordert wird, sprechen im Durchschnitt nur 34 % der Europäer gut genug Englisch, um damit effektiv arbeiten zu können. Der dritte Lösungsansatz, menschliche Dolmetscher und Übersetzer einzusetzen, muss

leider auch als impraktikabel verworfen werden: Mit Ausnahme weniger kritischer Einsatzgebiete, wo dies möglich und wichtig ist (z. B. das Europaparlament), wäre dies in weiten Teilen der Gesellschaft unbezahlbar.

Es muss daher eine vierte Lösung gefunden und effektiv in die Kommunikationswege eingebunden werden: eine automatische, rechnergestützte und kostengünstige bzw. kostenlose Lösung. Translinguale Kommunikation darf dabei aber nicht nur als Übersetzung von Texten gesehen werden. Denn Sprache wird gesprochen, geschrieben, ge-„textet" und gemalt. Sprache muss also erst einmal in ihrer originellen Ausdrucksform verstanden werden, bevor der Versuch einer Übersetzung unternommen werden kann.

## 2. Technologie

Eine wachsende Gemeinde von Wissenschaftlern widmet sich diesem Thema. Dabei ist zwischen unterschiedlichen Komponenten zu differenzieren, die für Teilaspekte des Kommunikationsproblems zuständig sind (siehe Abb. 1). Damit ein Mensch, der in einer Sprache spricht, einen anderen in einer anderen Sprache verstehen kann, sind drei Teilaufgaben zu lösen: (*1*.) Automatische Spracherkennung: Hier wird das gesprochene Signal in Sprache 1 ($L_a$) *via* Mikrophon aufgenommen, verarbeitet und als Text ausgegeben (*Speech-to-Text*). (*2*.) Maschinelle Übersetzung: Hier wird Text der einen Sprache ($L_a$) in Text der anderen ($L_b$) übersetzt (*Text-to-Text*). Und (*3*.) Sprachsynthese ($L_b$): Hier wird Text in der Zielsprache $L_b$ in gesprochener Sprache ausgegeben (*Text-to-Speech*). Um einen Dialog zwischen Menschen in zwei Sprachen zu ermöglichen, muss dieser Prozess dann auch in der anderen Sprachenrichtung (also von $L_b$ nach $L_a$) möglich sein und erfordert daher analoge Subsysteme in der jeweils anderen Sprache. Eine finale Integration dieser Untersysteme muss dann schließlich auch mit einer komfortablen Benutzerschnittstelle einfach in realen Kommunikationssituationen bedienbar und einsetzbar sein.

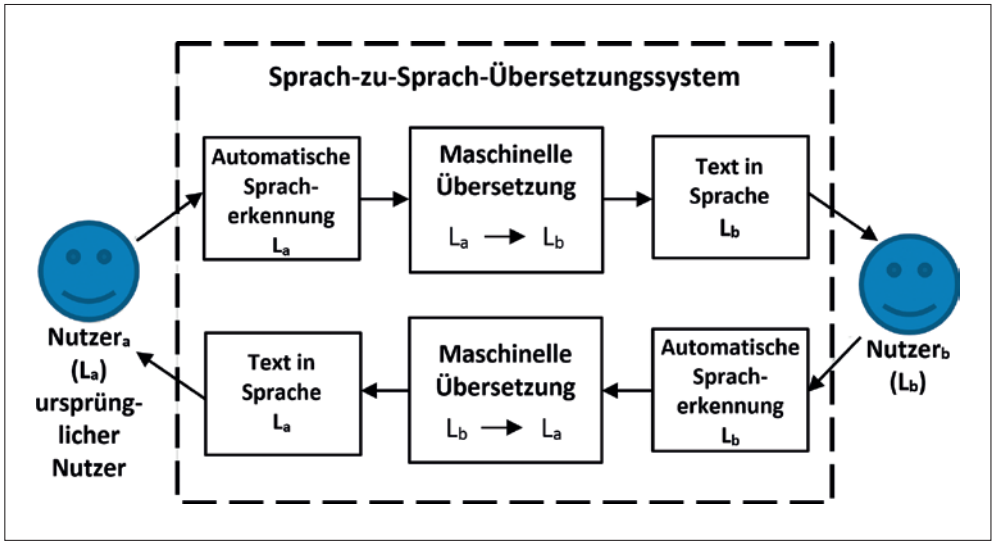


Abb. 1 Übersetzung gesprochener Sprache (*Speech-to-Speech*-Translation) – Übersicht

Jede dieser Teilaufgaben ist ein Forschungsgebiet für sich, das sich aufgrund der Komplexität und Ambiguität menschlicher Sprache als schwieriger als zunächst gedacht, herausstellte und daher die wissenschaftliche Gemeinde bereits mehrere Jahrzehnte beschäftigt und trotz großer Fortschritte weiter herausfordert. Die wichtigsten Lektionen dieser Arbeiten sind dabei, dass (*1*.) wir bei der menschlichen Sprache wegen ihrer Ambiguität bei jeder Wissensquelle nie harte, sondern nur weiche probabilistische Aussagen machen können, und (*2*.) dass wir wegen der Komplexität diese Aussagen und deren Wechselwirkungen nicht manuell kodieren, sondern nur an Daten erlernen können.

### *2.1 Spracherkennung (Automatic Speech Recognition [ASR])*

Für den unvorbereiteten Betrachter mag das Problem der Spracherkennung zunächst nicht als all zu schwierig erscheinen, bewältigen wir es als Mensch ja doch gut und mit Leichtigkeit. Jedoch verstecken sich bereits in der gesprochenen Sprache viele Mehrdeutigkeiten: So könnte z. B. dieselbe englische, akustische Sequenz von Lauten /juθəneʒə/ (in phonetischer Lautschrift) sowohl „Euthenasia“ als auch „Youth in Asia“ bedeuten. Und Sätze wie „This machine can recognize speech“ werden genauso ausgesprochen wie „This machine can wreck a nice beach“. Spracherkennung bedarf daher einer Deutung, welche von mehreren ähnlichen Alternativen in einem gegebenen Kontext die sinnvollere oder wahrscheinlichere Interpretation des Gesagten ist. Dies geschieht in modernen Spracherkennungssystemen durch eine Kombination von akustischen Modellen, die jedem Laut eine Wahrscheinlichkeit zuordnen, einem Aussprachewörterbuch (das jedem Wort eine Aussprache zuordnet) und einem Sprachmodell, das die Wahrscheinlichkeit jeder möglichen Wortsequenz „$w_1, w_2, \ldots$“ des Satzes auswertet. Abbildung 2 zeigt einen solchen typischen Erkenner. Die Auswertung dieser Modelle während der Erkennung sowie die Einstellungen der besten Parameter dieser Modelle sind jedoch nicht mehr manuell bestimmbar, sondern bedürfen automatischer Such- und Optmierungsalgorithmen.

Die Parameter der akustischen und linguistischen Modelle werden mit Hilfe von Lernalgorithmen über große Datenbasen von Sprachproben, deren Transkription bekannt

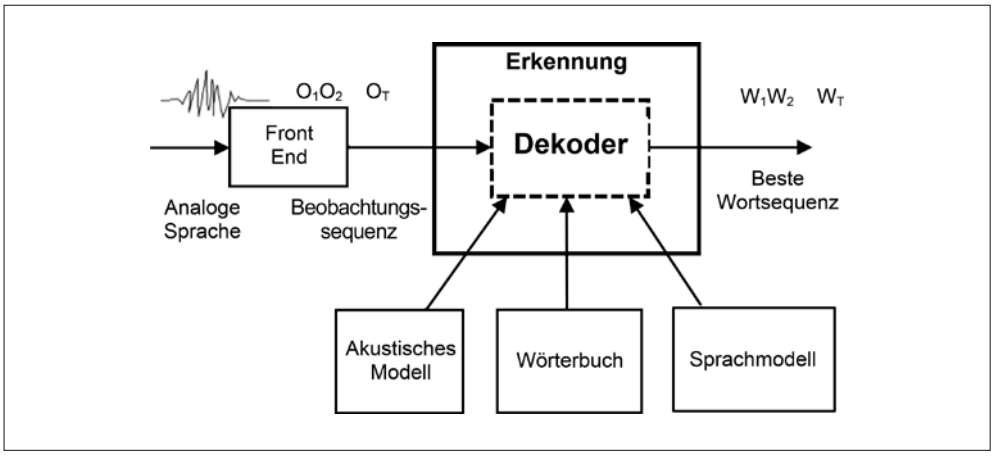


Abb. 2 Ein typischer Spracherkenner (*Speech-to-Text*)

ist, erlernt. Diese arbeiten mit statistischen Optimierungsmethoden oder mit Neuronalen Netzen und erlernen die beste Abbildung zwischen Signalen und Symbolen (kontextabhängige Phoneme und Worte) anhand von bekannten Beispieldaten. Moderne Systeme benutzen dazu inzwischen Neuronale Netze mit vielen Millionen von neuronalen Verbindungen, die der Lernalgorithmus optimiert.

### *2.2 Maschinelle Übersetzung (Machine Translation [MT])*

Erste Versuche mit der maschinellen Übersetzung von Texten (MT = *Machine Translation*) wurden bereits in den Weltkriegen unternommen. Diese und spätere Versuche scheiterten aber immer wieder an der Mehrdeutigkeit der Sprache und der Komplexität, diese mit Hilfe des damit verbundenen Kontextwissens aufzulösen. Fast jedes Wort (Bank, Stuhl, Spitze,…) hat mehrere Bedeutungen und daher Übersetzungen, die sich nur im Kontext richtig deuten lassen. Der biblische Satz „The spirit is willing but the flesh is weak" soll daher angeblich (Folklore der MT) von den ersten Maschinenübersetzern einmal ins Russische als „The vodka is good but the flesh is rotten" übersetzt worden sein. Auch strukturell ist die Interpretation von Sprache mehrdeutig. Worauf bezieht sich z. B. das Pronomen „it" in „If the baby doesn't like the milk, boil it". Sicher meinte der Autor, dass die Milch gekocht werden soll und nicht das Baby!

Der Versuch, das benötigte syntaktische, semantische und lexikalische Wissen mit Hilfe von Regeln manuell zu kodieren, scheiterte jedoch auch hier und musste über die Jahrzehnte der Forschung automatischen Lernverfahren weichen. Ein modernes MT-System benutzt nun daher auch eine ähnliche Systemarchitektur wie die Spracherkennung, die mehrere *erlernte statistische* Wissenskomponenten optimal kombinert (siehe Abb. 3).

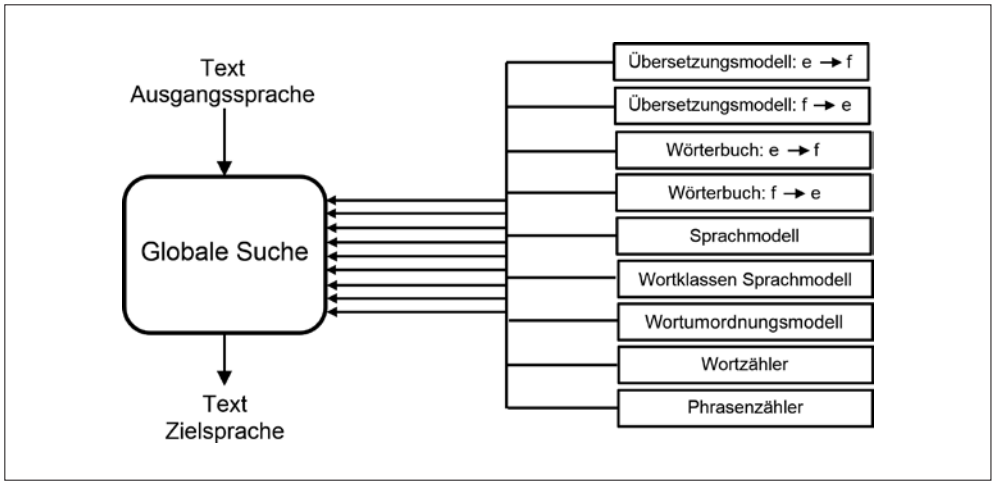


Abb. 3 Maschinelle Übersetzung (*Text-to-Text*)

### *2.3 Sprachsynthese (Text-to-Speech [TTS])*

Die dritte Komponente ist in der Regel die Sprachsynthese, die den übersetzten Satz in der Zielsprache hörbar machen, d. h. aussprechen, soll. Die TTS-Synthese ist dabei im Vergleich vielleicht die einfachere Komponente, da man mit einem vorgegebenen textuellen Satz immer nur genau ein Signal produzieren muss und daher nicht mit vergleichbar vielen Ambiguitäten der anderen Komponenten zu kämpfen hat. Aber auch hier stecken Tücken, die das Problem auch weiterhin als Forschungsproblem gelten lassen. Wie z. B. verwandle ich Text in eine phonetische Lautschrift (z. B. though → „*th oh*“). Und wie sollen diese Zuordnungen in vielen Sprachen getroffen werden, bei denen die Ausspracheregeln sich deutlich unterscheiden? Auch hier kommen automatische Lernalgorithmen zum Tragen, die die Einstellungen und Abbildungen über vorgefertigtes oder bereits existierendes Trainingsmaterial optimieren.

## 3. Evolution der Systeme

Tab. 1 Entwicklungsphasen sprachübersetzender Systeme

| | Jahre | Vokabular | Sprechstil | Domäne | Geschwindigkeit | Plattform | Beispielsysteme |
|---|---|---|---|---|---|---|---|
| Erste Dialog Demonstrationssysteme | 1989–1993 | begrenzt | beschränkt | limitiert | 2–10 × RT | Workstation | JANUS-1, C-STAR-I |
| Einfache Phrasenbücher | seit 1997 | begrenzt, modifizierbar | beschränkt | limitiert | 1–3 × RT | Handheld-Gerät | Phraselator, Ectaco |
| Spontane Zwei-Weg-Systeme | seit 1993 | unbegrenzt | spontan | limitiert | 1–5 × RT | PC/Handheld-Gerät | JANUS-III, C-STAR, Verbmobil, Nespole, Babylon, Transtac |
| Übersetzung von Nachrichtensendungen, politischen Reden | seit 2003 | unbegrenzt | Lesen/ Vorbereitete Rede | offen | offline | PCs, PC-Clusters | NSF-STRDUST, EC TC-STAR, DARPA GALE |
| Simultanübersetzen von Vorträgen | seit 2005 | unbegrenzt | spontan | offen | Echtzeit | PC, Laptop | Lecture Translator |
| Kommerzielle Konsekutivübersetzer per Telefon | seit 2009 | unbegrenzt | spontan | offen | online und offline | Smartphone | Jibbigo, Google, Microsoft |
| Simultandolmetscher-Leistungen | seit 2012 | unbegrenzt | spontan | offen | online | Server, Cloud-basiert | KIT, EU-Bridge, Microsoft |

Die Entwicklung von automatischen Übersetzungssystemen gesprochener Sprache begann Anfang der 1990er Jahre, als erste ASR-, MT- und TTS-Systeme einen minimalen Reifegrad erreichten, um erste Integration zu versuchen. Über die folgenden zwei Jahrzehnte wurden in mehreren Forschungs- und Entwicklungsphasen wichtige Einschränkungen der Technologie überwunden, die heute Sprachübersetzer im Einsatz möglich machen (siehe dazu Tab. 1 zur Übersicht der Systemqualifikationen).

### *3.1 Erste Demonstratoren*

In den USA und Europa wurde 1991 das JANUS-System als erstes derartiges *Speech*-Translation-System der Öffentlichkeit vorgestellt. JANUS wurde an der Universität Karlsruhe und an der Carnegie-Mellon-Universität in Pittsburgh (PA, USA) für Deutsch, Japanisch und Englisch entwickelt. Es entstand in Zusammenarbeit mit den *ATR Interpreting Telephony Laboratories* in Japan, die parallel ähnliche erste Systeme für die japanische Sprache entwickelten. Die Systeme wurden im Verbund in ersten übersetzenden Videokonferenzschaltungen vorgeführt (Waibel et al. 1991, *Handelsblatt* 30. 7. 1991)

Die Systeme stellten erste Gehversuche dar, beherrschten zunächst noch ein kleines Vokabular (< 1000 Wörter) und erforderten eine relativ eingeschränkte Syntax, beschränkten sich auf eine limitierte Gesprächsdomäne (wie z. B. eine Konferenzanmeldung) und waren zu groß und langsam, um in Feldsituationen, z. B. einem Reisenden, tatsächlich zur Seite stehen zu können. Ähnliche Demonstrationssysteme wurden dann 1992 auch von anderen Forschungsgruppen (ATT und NEC) vorgestellt.


Abb. 4 Erste Sprachübersetzungsprototypen in Videokonferenzschaltungen (JANUS in Konferenzschaltung mit Partnern, 1991)

### *3.2 Forschungssysteme und -prototypen*

Um derartige Systeme in den tatsächlichen Gebrauch zu bringen, folgten weitere wichtige Phasen der Entwicklung, die schwierige Probleme sukzessive meistern mussten.

Spontansprache, domänen-beschränkte Forschungssysteme: Um praktische Systeme zu realisieren, muss zunächst die Annahme syntaktischer Richtigkeit gelockert bzw. aufgehoben werden. Menschen sprechen selten syntaktisch richtige und vollständige Sätze. Sie sprechen vielmehr fragmentarische Sprachfetzen, mit Stotterern, Wiederholungen, Füllwörtern und Pausen (Ähs, Ähms, usw.). Diese müssen zunächst akustisch richtig erkannt und dann in der Aufbereitung herausgefiltert oder korrigiert werden, bevor eine Übersetzung stattfindet. Erste spontansprachliche Systeme wurden von 1993 bis 2000 entwickelt (Morimoto et al. 1993, Takezawa et al. 1998). Diese Systeme waren immer noch langsam und benötigten umfangreiche Hardware, und sie waren weiterhin domänenbeschränkt, um mit Hilfe einer Modellierung der Semantik in der eingeschränkten Domäne die für die Übersetzung bedeutungstragenden Sprachfetzen herauszuextrahieren. JANUS-III, C-STAR-Systeme, VERBMOBIL und andere Projekte lieferten hier Fortschritte, waren aber für den Einsatz zunächst noch unbrauchbar (Lavie et al. 1997). Hierzu musste die Domänenbeschränkung noch genommen werden, und die Systeme mussten für den Einsatz schneller und mobiler werden. Auch wurden handprogrammierte Regelwerke (die in eingeschränkten Domänen noch möglich sind) durch automatisch erlernte, statistische Subsysteme ersetzt, um Robustheit und Skalierbarkeit zu verbessern (Brown et al. 1993, Och und Ney 2004, Koehn et al. 2007).

Es zeichneten sich zwei unterschiedliche Einsatzgebiete ab, die weitere technische Herausforderungen mit sich brachten:

- Konsekutives Dolmetschen von Dialogen im mobilen Einsatz: In diesem Einsatz soll ein Dialog mit einem anderssprachigen Dialogpartner ermöglicht werden, bei dem die Sätze der beiden Gesprächspartner zunächst übersetzt werden, damit das Gegenüber dann mit einer Antwort darauf reagieren kann. In der Regel spricht bei dem konsekutiven Übersetzer jeder Gesprächsteilnehmer nur ein bis zwei Sätze, und in den meisten Einsatzgebieten (Tourismus, medizinische Einsätze, humanitäre Hilfe, …) wird nur ein allgemeines Vokabular von ca. 40 000 Wörtern benötigt. Jedoch muss eine Übersetzung schnell geliefert werden (um den Gesprächsfluss nicht einzuschränken), und das System muss mobil (also auf tragbarer kleiner Hardware) verfügbar sein.
- Simultanes Dolmetschen im stationären Einsatz: Bei vielen Einsatzgebieten der Sprachübersetzung wird kein Dialog zwischen zwei Gesprächspartnern benötigt, sondern die rasche, effektive Übersetzung eines Monologs. Beispiele hierfür sind TV-Übertragungen, Internetvideos, Vorlesungen, Ansprachen und Reden. In dieser Klasse der Anwendungen ist Mobilität meist nicht entscheidend, denn die Rechenleistung kann durchaus auf großen Servern online oder offline erbracht werden. Jedoch handelt es sich hierbei meist um komplexere Themen, mit Fachwörtern und Fachjargon. Auch produziert der Sprecher nicht nur einen oder zwei Sätze, sondern eine Rede, d. h. einen kontinuierlichen Strom von Wörtern, bei dem der Anfang und das Ende übersetzbarer Einheiten oder Sätze vom System selbst gefunden werden müssen. Eine derartige Segmentierung in Satzeinheiten oder -fragmente und das automatische Setzen von Interpunktion (Punkte, Kommata, Fragezeichen) müssen unter Einbezug des Kontextes automatisch durchgeführt werden (*The Economist* 12. 6. 2006)

## 4. Übersetzung gesprochener Sprache im Einsatz

Die frühen Forschungssysteme (1990–2005) lösten technische Probleme, damit wurde der Weg geebnet für den Vertrieb und tatsächlichen Gebrauch sprachübersetzender Systeme in der Gesellschaft.

### *4.1 Konsekutiv-Dolmetschen (Consecutive Translation)*

Der Transfer von dolmetschenden Systemen im tatsächlichen Feldeinsatz wurde zunächst in humanitären und logistischen Übungen der US-Regierung erprobt. Dabei kamen zunächst domänenbeschränkte *Speech-to-Speech*-Übersetzungssysteme zum Einsatz (*Mobile Technologies*, Eck et al. 2010) oder auch einfachere sprachbasierte Phrasenbücher (*Voxtec*), die ohne Übersetzungskomponente vorgefertigte Phrasen über Spracheingabe abrufen (http://www.voxtec.com). Frühe Modelle dieser kommerziellen Systeme von *Voxtec* und *Mobile Technologies* sind in Abbildung 5 abgebildet. Diese frühen Systeme waren jedoch in Vokabular und Sprachgebrauch domänenbeschränkt und somit nur für Dialoge in speziellen Einsatzgebieten konzipiert (Einsätze in Krisengebieten, medizinische Hilfseinsätze, Polizei, Hotelrezeption usw.). Vertrieb und Vermarktung waren daher auch auf kleine Abnehmergruppen beschränkt.

Jedoch mit dem Einzug von Smartphones erreichte die Rechenleistung von Mobiltelefonen die kritischen Voraussetzungen, unter denen Spracherkennung und Übersetzung offener unbeschränkter (> 40 000 Wörter) Vokabularien in nahezu Echtzeit ermöglicht wird.


Abb. 5 Erste kommerzielle Systeme: (*A*) Phraselator, (*B*) iPaq PDA-basierter *Speech*-Translator[1] (~2005) sowie (*C*) JIBBIGO, der weltweit erste *Speech-to-Speech*-Übersetzer auf einem Telefon (2009)

1 Phraselator von Voxtec LLC und Speech Translator von Mobile Technologies LLC.

*Mobile Technologies* (eine Start-up-Firmengründung aus unseren Carnegie-Mellon-Laboratorien) konnte so 2009 das weltweit erste domänenunbeschränkte *Speech-to-Speech*-Übersetzungssystem auf einem Telefon auf den Markt bringen: JIBBIGO (Eck et al. 2010). Beflügelt von den einfachen Vertriebsmechanismen des Apple iTunes App-Stores sowie der wachsenden Verbreitung von Smartphones weltweit, konnte JIBBIGO rasch auf 15 Sprachen expandieren und eine große weltweite Verbreitung durchsetzen. JIBBIGO öffnete den Markt für tragbare Sprachübersetzer. Alternative Produkte von *Google* und *Microsoft* kamen auf den Markt, die aber zunächst die Sprachverarbeitung nur serverbasiert über das Internet auf externen Servern durchführten und somit nur bei bestehender Internetverbindung einsatzfähig waren.

Die Mobilität, die niedrigen Kosten, die großen Vokabularien und die allgemeine Verfügbarkeit (auch ohne Netzwerkverbindung) einer Offlinelösung, bei der alle Komponenten lokal auf dem Telefon laufen, sind unabhängig vom Internet und so nützlicher für Reisende (keine Roamingkosten) und humanitäre Einsätze (kein Internet erforderlich). JIBBIGO kam daher als präferierte Plattform in einer Reihe humanitärer Einsätze der US-Regierung sowie karitativer Nichtregierungsorganisationen (NGOs) in Thailand, Kambodscha und Honduras bei der Vermittlung zwischen englischsprechenden Ärzten und fremdsprachigen Patienten zum Einsatz (Abb. 6*A-D*)


Abb. 6 Medizinische Einsätze in Thailand, Kambodscha und Honduras: (*A*) Translinguale Dialoge zwischen amerikanischen Ärzten und Patienten in Thailand, (*B*) Ärztliche Versorgung mit Hilfe des JIBBIGO-Sprach-Dialog-Übersetzers in Thailand, (*C*) Medizinischer Einsatz in Kambodscha und (*D*) humanitäre Einsätze mit JIBBIGO in Honduras.

Die Systeme können auf iPhone- oder Android-Telefonen eingesetzt werden, bieten aber besonders auf einem Tabletcomputer für humanitäre Einsätze eine benutzerfreundliche Interaktionsoberfläche zwischen gegenübersitzenden Gesprächspartnern.

Nach fünf Jahren der Entwicklung in Feldsituationen wurden die Systeme 2013 auch in humanitären Einsätzen (*Medical Civial Action Program*) in Thailand evaluiert (Hourin et al. 2013). In 95 % der Interaktionen bei der Registrierung von Patienten war es möglich, den Dialog mit Hilfe eines JIBBIGO-maschinellen Tabletdolmetschers zu bewältigen, ohne einen menschlichen Dolmetscher zu bemühen.

### *4.2 Simultan-Dolmetschen (Simultaneous Interpretation)*

In einem multilingualen Umfeld ist der Dialog zwischen mehrsprachigen Gesprächspartnern nicht die einzige Herausforderung. Wenn wir an TV-Nachrichten, Filme, Vorträge, Vorlesungen, Ansprachen, Straßenschilder, Vortragsfolien, SMS-Nachrichten denken, sehen wir viele andere Herausforderungen, bei denen translinguale Technologien benötigt werden.

Ein wichtiges Einsatzgebiet ist dabei die Übersetzung von Vorlesungen (Fügen et al. 2007). Besonders deutsche Universitäten sind trotz hervorragender Wissenschaftslandschaft und Fördermittel im internationalen Wettbewerb um Talente häufig benachteiligt, einfach weil viele ausländische Studierende oder wissenschaftliches Personal und Akademiker sich nicht der zusätzlichen Auflage unterziehen wollen, eine neue Sprache zu erlernen (besonders eine so schwere wie die deutsche). Wie sollen deutsche Universitäten oder deutsche Firmen darauf reagieren? Soll eine deutsche Universität alle Lehrveranstaltungen auf Englisch abhalten? Der Autor dieses Beitrags hält dies weder für wünschenswert noch für praktikabel. Eine hybride Lösung mit Hilfe moderner Sprachtechnologien, die sprachliche und kulturelle Vielfalt und Toleranz unterstützt (und nicht in die eine oder andere Richtung unterbindet), erscheint hier deutlich vielversprechender, da sie die Internationalisierung und Völkerverständigung fördert und verbessert.

Am Karlsruher Institut für Technologie (KIT) haben wir ein derartiges System für Studierende im Audimax bereits im Einsatz (Cho et al. 2013). Sprachübersetzung (*Speech*-Translation) bleibt dabei weiterhin noch eine Forschungsaufgabe; nicht alle Probleme sind gelöst. Aber für einen Zuhörer, der die Sprache des Vortragenden nicht beherrscht, ist auch ein imperfekter Computerdolmetscher besser als gar nichts.

Das zwischen der Carnegie-Mellon-Universität (CMU) und dem KIT erstmals 2005 vorgestellte System (siehe Abb.7) beherrscht zunächst nur eine Sprachenrichtung, da eine Vorlesung als Monolog nur von der einen in die andere Sprache übersetzt werden soll. Ein solches System muss auch nicht mobil im Taschenformat, sondern kann cloudbasiert auf Servern laufen und über das Internet abgerufen werden. Im Unterschied zum Dialogsystem benötigt ein Vorlesungsübersetzer allerdings eine Spracherkennungskomponente und den maschinellen Übersetzer. Die Synthese in gesprochene Sprache kann danach erfolgen, aber wenn Untertitel gewünscht sind, ist Sprachsynthese optional. Zusätzlich jedoch wird hierzu noch eine Segmentierungskomponente benötigt, die explizit oder implizit entscheidet, wann im erkannten Redeschwall das Ende eines Satzes oder zumindest eines übersetzbaren Satzfragments erreicht ist. Erschwerend sind auch die Vokabularien, die bei Vorlesungen viele Fachwörter und Jargon enthalten.

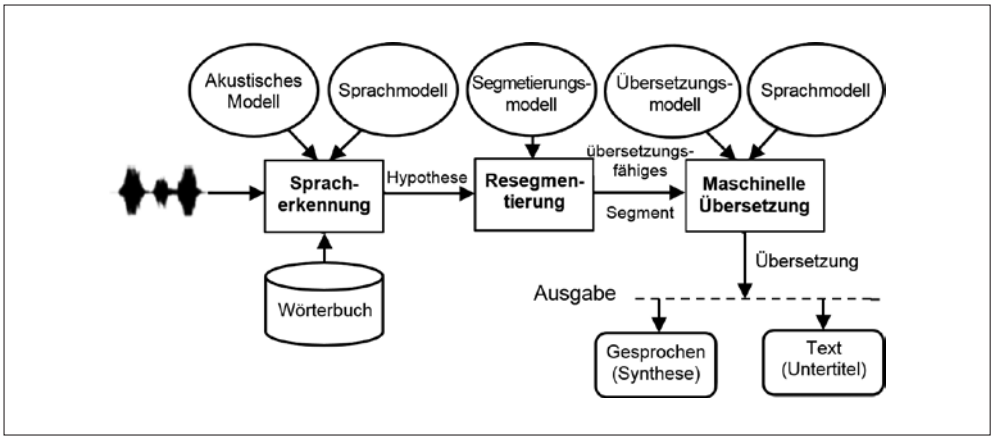


Abb. 7 Sprach- (*Speech*-) Übersetzung von Vorlesungen

Ein Vorlesungsübersetzer kann online oder offline laufen. Ein Onlinesystem ist nötig, wenn ein Zuhörer simultan eine Untertitelung (während des Vortrags) in der einen Sprache (Transkript) oder beiden Sprachen (Übersetzung) mitverfolgen möchte. Ein Onlinegebrauch erfordert echtzeitfähige Erkennung und Übersetzung (d. h., das System muss mit dem Gesagten „mithalten"), und es ist wünschenswert, dass die Latenz (d. h. der zeitliche Abstand zwischen dem gesagten Wort und der übersetzten Ausgabeworte) so gering wie möglich gehalten wird. Andernfalls verliert der Zuhörer den Bezug zum Vortrag und zum Geschehen im Vortragssaal. Diese Anforderungen sind besonders im Deutschen eine Herausforderung, weil im Deutschen das Verb oder wichtige Teile des Verbs am Ende (manchmal auch erst viel später) kommen. Ein Satz wie z. B. „**Ich schlage** Ihnen nach eingehender Prüfung Ihres Antrags, der uns gestern … und … und … nachdem … und … eine neue Vorgehensweise **vor**". Das kleine Wörtchen „vor", dass ja eventuell Minuten später erst gesprochen wird, entscheidet aber darüber, ob der englische Satz in der Übersetzung hier mit „I **propose** …" oder mit „I **hit** …" anfängt.

Im Lehrbetrieb einer Universität, aber auch im multimedialen Sendebetrieb gibt es jedoch auch viele Anwendungsszenarien, in denen eine Offlinebearbeitung der Sprache und der Übersetzung tolerierbar oder wünschenswert ist. Im Offlinebetrieb ist Echtzeitfähigkeit nicht zwingend notwendig (wenn auch eine exzessiv lange Bearbeitungszeit zum Kostenfaktor werden kann), und das System kann eine bessere Transkription und Übersetzung unter Einbezug eines längeren Kontexts bestimmen. So kann z. B. ein Vorlesungsübersetzer im Hörsaal unter Onlinebedingungen laufen und dann archivarisch im Offlinemodus noch einmal nachbearbeitet werden, um einem Zuhörer im Archiv später eine bessere Version zur Verfügung zu stellen.

Ein derartiges Vorlesungsübersetzungssystem wurde am KIT seit 2012 als Internetdienst im Audimax in Betrieb genommen (Abb. 8) (Cho et al. 2013, von Greve-Dierfeld 12. 6. 2012). Studierende, die die sprachliche Unterstützung wünschen, verbinden ihre Telefone, Tablets oder PCs über einen normalen Internet-Webbrowser mit einer dafür eigens eingerichteten Webseite und erhalten simultan die textuelle Transkription ins Deutsche (dies ist also auch bei Hörproblemen nützlich) sowie die Übersetzung ins Englische. Andere Ausgabesprachen sind derzeit in Bearbeitung.


Abb. 8 Der Vorlesungsübersetzer im Einsatz im Audimax am KIT

Die Transkription und Übersetzung von Vorlesungen im Lehrbetrieb einer Universität beherbergt noch weitere offene Fragen, die die Forschung noch in weiterführenden Arbeiten beschäftigen, besonders im Deutschen. Zusätzlich zu den Problemen mit der Wortstellung und den Verben, wie oben diskutiert, beobachten wir andere Schwierigkeiten des Deutschen:

- Komposita: „Wörter" wie „Fehlerstromschutzschalterprüfung" im Deutschen müssen zunächst zerlegt werden, bevor man sie sinnvoll ins Englische übersetzen kann. Unser Institut entwickelt hierzu Algorithmen, die Komposita zerlegen. Aber auch dies ist aufgrund der Ambiguität der Sprache nicht immer einfach. Eine Zerlegung in Fehler-Strom-Schutz-Schalter-Prüfung in unserem Beispiel oben ist sinnvoll, aber eine Zerlegung von „dramatisch" in „Drama-Tisch", oder „Asiatisch" in „Asia-Tisch" je nach Kontext unpassend und verändert die Bedeutung (Koehn und Knight 2003).
- „Agreement": Die Endungen müssen im Deutschen konsistent sein und zu den Nomen passen: „in der wichtigen, interessanten, didaktisch gut vorbereiteten, heute und gestern wiederholt stattfindenden Vorlesung".
- Fachwörter und Jargon: Dies ist besonders bei der Bearbeitung von Vorlesungen an einer Universität ein großes Problem, denn jede Vorlesung hat wiederum ihre eigenen Fachwörter und sprachlichen Eigenarten. Was sind „Cepstral-Koeffizienten", „Walzrollenlager" und „Würfelkalküle", usw.? Um dies teilweise zu lösen, benutzen wir automatische Algorithmen, die die Vortragsfolien des Vortragenden nach unbekannten Wörtern durchforsten, um diese (und im Internet auffindbare verwandte Fachwörter) dann automatisch

in das Erkennervokabular einzubauen. Alternativ nutzen wir auch das Benutzer-Feedback der Studenten und deren spontane Korrekturen, um neue Begriffe automatisch zu erlernen. Nach den Übersetzungen dieser Fachwörter fahnden dann unsere Programme automatisch in öffentlichen Internetquellen wie *Wikipedia* (Niehues und Waibel 2012).

– „Code-Switching“: Oft enthalten Vorlesungen und Vorträge Zitate und Begriffe aus anderen Sprachen. Besonders in Vorlesungen der Informatik findet man häufig englische Begriffe, aber auf Deutsch dekliniert. So spricht man von „iPhone“, „iPad“, „cloudbasiertem Webcastzugriff“ oder „Files, die man downgeloadet hat“. Die englischen Wörter erscheinen also in einem deutschen Redefluss manchmal nach englischen und manchmal nach deutschen Ausspracheregeln ausgesprochen und dann noch deutsch dekliniert und in Komposita verpackt!
– Pronomen: Worauf beziehen sich Pronomen? Auch hier gibt es häufig noch Schwierigkeiten. „Wir freuen uns, Sie heute hier begrüßen zu dürfen“ wird als „we are happy to welcome her here“ übersetzt.
– Lesbarkeit: Wenn Menschen sprechen, sprechen sie nicht die Satzzeichen oder Umbrüche, die lesbarer Text enthalten sollte. Es müssen also Punkte, Kommata, Fragezeichen, Paragraphen, eventuell sogar Überschriften automatisch generiert und eingefügt werden (Cho et al. 2014a).
– Spontansprache: Unterschiedliche Sprecher sprechen mehr oder weniger syntaktisch sauber. Die Hesitationen, Stotterer, Wiederholungen, Abbrüche spontangesprochener Sprache erschweren die Lesbarkeit und machen die Übersetzung schwierig. Ein gesprochener Satz aus einer Vorlesung würde von einem fehlerlosen, also perfekten Spracherkenner, aber ohne Satzzeichen, ohne Korrektur der Spontaneffekte, dann nämlich so aussehen: „Das ist alles was Sie das haben Sie alles gelernt ja und jetzt können Sie es einsetzen und ich erzähle gleich welche Implikationen das hat Ähm das ist auch so ja und äh wenn Sie die Systeme die Sie bauen dann eben auch einsetzen dann sind Sie wenn Sie so wollen ja der der der erste Tablet ähm den es so gab ähm hatte ein Wireless LAN Ähm eine Charakteristik ist dass wir versuchen unsere ähm den den den Zweck und die Funktionalität zu äh einzuschränken ja da gibt’s da gab’s in äh gab’s nur eines“. Auch hier gilt es, die erkannte Sprache zunächst sprachlich aufzubereiten, um sie in der Quellsprache lesbar zu machen und um sie danach in lesbaren Text in der Zielsprache zu übersetzen (Cho et al. 2014b).
– Mikrophone und Geräusche: In der gegenwärtigen Konfiguration unseres Vorlesungsübersetzers trägt der Sprecher ein Nahbesprechungsmikrophon. Dies ist im Vorlesungsbetrieb durchaus akzeptabel, da Vortragende in Hörsälen meist ohnehin Mikrophone tragen. Aber in Seminaren und Besprechungen wäre dies störend. Leider führen offene Tischmikrophone zu Echo und Hall bzw. Geräuschen, zudem führt auch das Sprechen mehrerer Sprecher zu deutlichen Einbußen in der Erkennungsleistung.
– Sprachliche Skalierbarkeit (*Portability*): Wie können wir die entwickelten Technologien nicht nur in ein, zwei Sprachen realisieren, sondern zur Vermittlung zwischen allen Sprachen und Kulturen auf unserem Planeten erweitern? Hierzu müssen die Entwicklungskosten eines Übersetzungssystems deutlich reduziert werden. Sprachenunabhängige Technologien (derzeit an unserem Institut mit Hilfe neuronaler Netze in Bearbeitung), Adaption, Inferenz, Abstraktion, bessere Nutzung monolingualer Ressourcen sowie „Crowdsourcing“ (um das vielsprachige Wissen der Menschheit besser zu extrahieren) bieten alle erfolgsversprechende Ansätze.

Die Architektur für den KIT-Vorlesungsübersetzer für den tatsächlichen Betrieb wurde unter dem Integrated-Projekt der Europäischen Union, EU-Bridge (Abb. 9), am KIT eingeführt und erprobt. Sie kann nun cloudbasiert mehrere Vorlesungen gleichzeitig unterstützen.

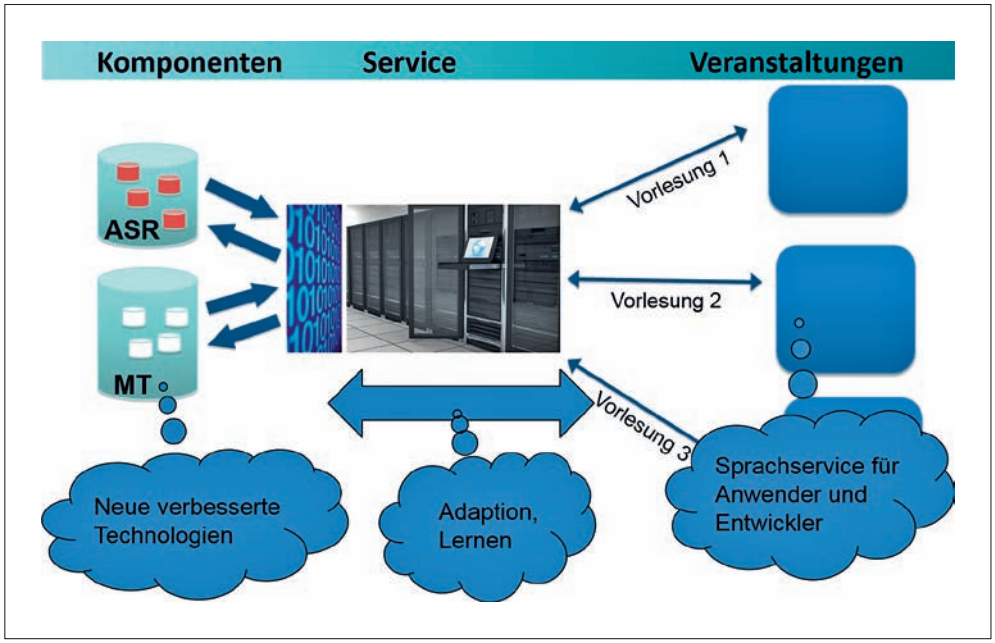


Abb. 9 EU-Bridge: Der maschinelle Dolmetscher als cloud-basierter Dienst. ASR (Automatic Speech Recognition) – Automatische Spracherkennung, MT (Machine Translation) – Maschinelle Übersetzung


Abb. 10 Automatisch übersetzter Vortrag im Europäischen Parlament

Die Sprachdienste können durch eine derartige Serverarchitektur in mehreren Hörsälen und auch in anderen Anwendungsszenarien (also nicht nur Vorlesungen an der Universität) verteilt, genutzt werden. Das Sprachübersetzungssystem für Vorlesungen konnte in der Zwischenzeit (2012, 2013 und 2014) so auch im Europäischen Parlament (Abb. 10), während mehrerer Rektorenkonferenzen sowie bei Schulungen der Dolmetscher vorgeführt werden.

Über den Einsatz an Universitäten (bei denen sonst in der Regel gar keine Sprachunterstützung existiert) hinaus, können automatische Systeme aber auch Experten, wie beispielsweise Humandolmetscher im Parlament, unterstützend zur Seite gegeben werden. Ein erster Test in Übersetzungskabinen des Europäischen Parlaments wurde im Parlament in Straßburg Ende 2014 bereits unter *EC-Integrated-Project* „EU-Bridge" erfolgreich durchgeführt.

Abb. 11 Erster Test beim Einsatz eines maschinellen Dolmetschers im Parlament bei Abstimmungen

Da das Europäische Parlament wesentlich besseres Dolmetschen durch menschliche Experten zur Verfügung stellen kann (dort arbeiten die weltbesten Dolmetscher in mehr Sprachen als in jeder anderen Organisation der Welt), sind derartige Dienste eher als *Back-up*-Lösungen konzipiert, um das Arbeiten der Dolmetscher zu erleichtern. So kann die Technologie z. B. automatisch Terminologielisten erstellen oder Zahlen und Namen erfassen, die dann als „Scratch-Pad" und Erinnerungsstütze den Dolmetschern zur Seite stehen. Auch ist ein „Interpreter's Cruise-Control" denkbar, der bei unkritischen oder repetitiven Sitzungsfragmenten (z. B. dem Verlesen von Abstimmungsergebnissen) eingeschaltet werden könnte.

## 5. Translinguale Kommunikation

In einem multilingualen und multikulturellen Umfeld treten Sprachbarrieren nicht nur bei gesprochenen Dialogen, Vorlesungen oder in Textdokumenten auf. Sie kommen in vielen weiteren Kommunikationssituationen, Modalitäten und Medien vor: Wichtige Information finden wir in Straßenschildern, SMS-Text-Messaging, Fernsehnachrichten, Vortragsfolien, der Gestik und vielem anderem mehr. Um die Vision einer multilingualen, sprachbarrierefreien Welt zu realisieren, darf unser Unterfangen also auch nicht beim Bau besserer Übersetzungssysteme enden. Wir müssen uns viel mehr auch um bessere Benutzerschnittstellen bemühen, die diese Sprachbarrieren gänzlich transparent in den Hintergrund treten lassen. Erfolgreiche translinguale Kommunikation ist erst dann erreicht, wenn Menschen miteinander interagieren können, ohne sich der Barrieren bewusst zu sein.

An unseren InterACT-Laboratorien an der CMU und am KIT beschäftigen wir uns daher schon länger parallel zu den Übersetzungsaufgaben mit multimodalen Benutzerschnittstellen, die eine solche Kommunikation in unterschiedlichen Situationen transparent ermöglichen sollen. Eine Reihe von beispielhaften Prototypen und Einsatzszenarien wurden daher bereits untersucht:

– Ein *Road-Sign*-Translator: Bereits 2001 wurden Systeme vorgestellt, die Straßenschilder mit Hilfe einer mobilen Kamera lesen und übersetzen können (Yang et al. 2001). Die Übersetzungen wurden dabei in das Bild der Szene eingeblendet und das System auf einer (damaligen) PDA-Plattform zuerst erprobt. Ähnliche Anwendungen wurden nun inzwischen auch für iPhones als Apps entwickelt und herausgegeben.


Abb. 12 *Road-Sign*-Translator: Übersetzung von Straßenschildern

– *Speech*-Translation-*Goggles*. Bereits 2005 wurde die Ausgabe unserer Simultanübersetzer auch in *Heads-up-Display*-Brillen vorgenommen, die es einem Benutzer erlauben sollen, einem Gesprächspartner gegenüber zu sitzen und die Übersetzungen seiner Sprache wie Untertitel in der Brille eingeblendet zu bekommen. Während das 2005 noch wie *Science Fiction* erschien, sind derartige Konfigurationen *via* Smartphone, Smartwatch oder Google Glass durchaus realisierbar und teilweise verfügbar.
– Handschrifterkenner, die die Schrift erkennen und auch die Übersetzung liefern. Auch dies ist durch *Road-Sign*-Translator oder -Scanner bereits teilweise gelöst, es bedarf aber auch hier einer einfacheren Bedienung für reelle Umgebungen.
– Übersetzung von Vortragsfolien: Für unsere ausländischen Studierenden sind nicht nur die Vorträge in einer fremden Sprache unverständlich, sondern meist auch die Vortragsfolien des Vortragenden. Daher wurde auch ein Übersetzungssystem für Folien am Institut entwickelt, das den Text einer Folie, der mit der Maus angefahren wird, übersetzt und mit einer Sprechblase über die Folie legt (Abb. 13).

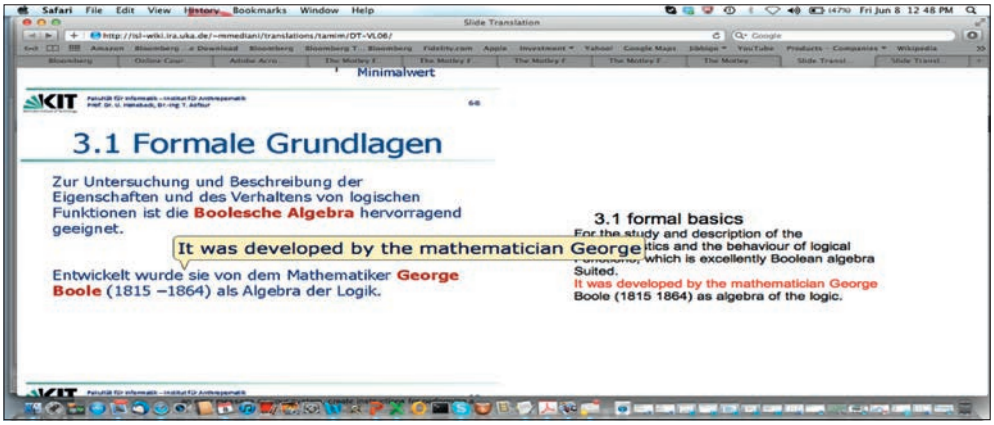


Abb. 13 Übersetzung von Vortragsfolien

– Stille Sprache (*Silent Speech*): Sprache ist immer mit Geräusch verbunden. Es wurde daher auch ein System entwickelt, das *via* Elektromyographie artikulierte Mundbewegungen als Sprache erkennen kann, auch wenn die Sprache nicht laut ausgesprochen wurde. Derartig artikulierte, stille Sprache kann daher erkannt (wenn auch die Erkennung nicht so gut läuft wie bei gesprochener Sprache), übersetzt und *via* Synthese hörbar gemacht werden (Maier-Hein et al. 2005), so dass bei der Artikulation der einen Sprache hörbare Sprache der anderen produziert werden konnte (Abb. 14).
– Sprachübersetzung mit Hilfe von Richtungslautsprechern: Mit Hilfe derart ausgerichteter Lautsprecher ist es möglich, das Übersetzungsergebnis als synthetisierte Sprache nur auf bestimmte Punkte in einem Raum zu fokussieren. So kann hörbare Simultanübersetzung auch ohne Kopfhörer individuellen Zuhörern in unterschiedlichen Sprachen präsentiert werden (Abb. 15).

Abb. 14 Ein Übersetzer lautloser Sprache

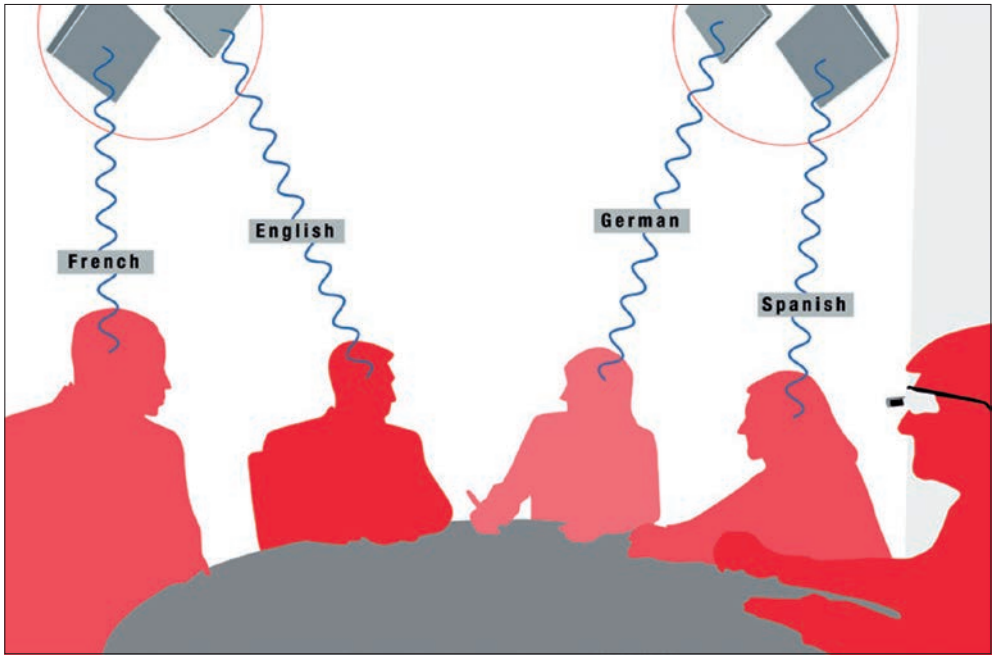


Abb. 15 Individuell angepasste Simultanübersetzung ohne Kopfhörer: mit Hilfe von direktionalen Lautsprechern und *Heads-up-Display*-Brillen

## 6. Zusammenfassung und Ausblick

Moderne Sprachtechnologien reißen schon jetzt Sprachbarrieren ein. Noch vor wenigen Jahren galten diese Visionen als uneinlösbar und als *Science Fiction*, haben Sprachbarrieren doch über Jahrhunderte hinweg die Menschheit mit babylonischer Sprachverwirrung voneinander getrennt.

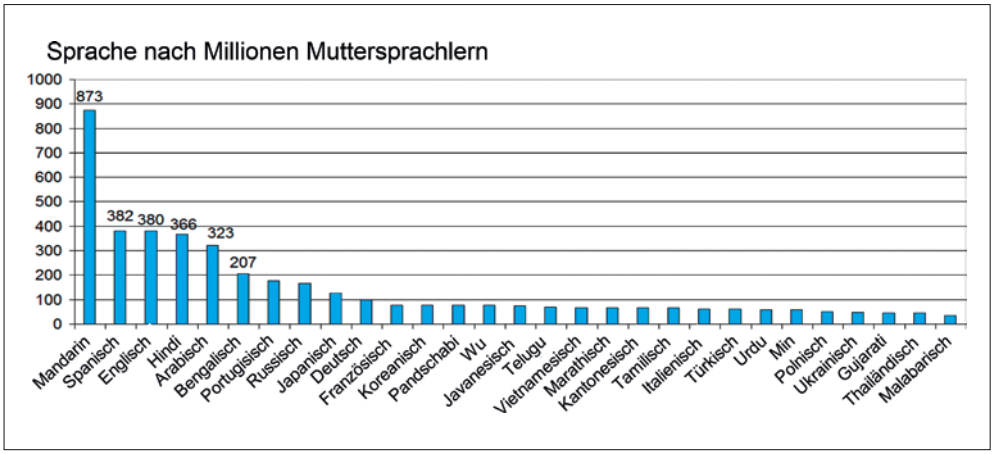


Abb. 16 Verteilung der Sprachen der Welt

Viel ist noch zu tun, denn aktuelle Sprachtechnologien decken bisher nur wenige Sprachen (~50) ab, rund 6000 jedoch gibt es. Um alle Sprachen der Welt einzubeziehen, müssen die hier vorgestellten Technologien noch auf viele Sprachen „portiert" werden, und um das zu erreichen, muss der Bau der Systeme deutlich billiger und einfacher werden. Multilinguale Modellierung (Nguyen et al. 2014), Adaption, sprachenunabhängige Modelle, bessere Interaktion mit den Benutzern und selbstlernende Systeme werden dies realisieren lassen.

Aber auch hier sind die Prognosen gut: Die in diesem Beitrag beschriebenen Technologien sind alle „lernbar", d. h., sie lassen sich anhand von Datenbasen automatisch für jede Sprache neu trainieren. Wichtig ist dabei immer nur wieder die Erhebung hinreichend großer Datenmengen. Das Internet macht auch dieses immer leichter und realistischer: Daten können auch aus der Ferne einfach erhoben werden, oder sie werden indirekt durch *Crowdsourcing*, Aktivitäten, Spiele oder Benutzer-Feedback freiwillig und kostenlos produziert. Moderne Übersetzungssysteme werden bereits heute mit mehr Daten trainiert (>> 1GWords), als ein Mensch im Durchschnitt während seiner Lebenszeit spricht (~0,5 GWords). Und dies wird weiter zunehmen.

Die Chancen sind daher gut, dass neue Algorithmen und Technologien uns nicht nur physisch näher bringen und Kommunikation mit jedem Menschen auf unserem Planeten ermöglichen, sondern auch Sprachbarrieren und Verständigungsprobleme untereinander noch in unserer Generation verschwinden lassen.

## *Dank*

Der vorliegende Beitrag wurde substantiell verbessert durch die Lektüre, Diskussion und editorielle Unterstützung von Jan Niehues, Margit Rödder, Maria Schmidt und Dorothea Schweizer. Der Autor bedankt sich für die Unterstützung und Mitwirkung.

## *Literatur*

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., and Mercer, R. L.: The mathematics of statistical machine translation: Parameter estimation. Computational Linguistics *19*/2, 263–311 (1993)

Cho, E., Fügen, C., Herrmann, T., Kilgour, K., Mediani, M., Mohr, C., Niehues, J., Rottmann, K., Saam, K., Stüker, S., and Waibel, A.: A real-world system for simultaneous translation of german lectures. Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Lyon, France. 2013

Cho, E., Niehues, J., and Waibel, A.: Tight integration of speech disfluency removal into SMT. EACL *43* (2014a)

Cho, E., Niehues, J., and Waibel, A.: Machine translation of multi-party meetings: Segmentation and disfluency removal strategies. IWSLT 2014 (2014b)

Eck, M., Lane, I., Zhang, Y., and Waibel, A.: Jibbigo: Speech-to-speech translation on mobile devices. IEEE Spoken Language Technology Workshop; pp. 165–166. 2010

Fügen, C., Waibel, A., and Kolss, M.: Simultaneous translation of lectures and speeches. J. Machine Translation *21*/4, 209–252 (2007)

*The Economist*: How to build a bable fish. The Economist 12. 6. 2006

Greve-Dierfeld, A. von: Uni-Übersetzungs-Automat: Don't worry about make. Spiegel-Online 12. 6. 2012. http://www.spiegel.de/unispiegel/studium/dolmetscher-fuer-die-vorlesung-kit-entwickelt-uebersetzungsprogramm-a-838409.html

*Handelsblatt*: Übersetzung. Handelsblatt 30. 7. 1991

Hourin, S., Binder, J., Yeager, D., Gamerdinger, P., Wilson, K., and Torres-Smith, K.: Speech-to-Speech Translation Tool Limited Utility Assessment Report. Report OMB No.0704-0188 (2013)

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E.: Moses: Open source toolkit for statistical machine translation. ACL '07 Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions; pp. 177–180. Stroudsburg (PA, USA): Association for Computational Linguistics 2007

Koehn, P., and Knight, K.: Empirical methods for compound splitting. EACL 2003

Lavie, A., Waibel, A., Levin, L., Finke, M., Gates, D., Gavalda, M., Zeppenfeld, T., and Zhan, P.: JANUS III: Speech-to-speech translation in multiple languages. International Conferences on Acoustics, Speech, and Signal Processing (ICASSP) 1997, Munich, Germany, 1. April 1997. ICASSP 1997

Maier-Hein, L., Metze, F., Schultz, T., and Waibel, A.: Session independent non-audible speech recognition using surface electromyography. Proceedings of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), Cancun, Mexiko, November 2005. ASRU 2005

Morimoto, T., Takezawa, T., Yato, F., Sagayama, S., Tashiro, T., Nagata, M., and Kurematsu, A.: ATR's speech translation system: ASURA. Proceedings Eurospeech '93, Geneva, Italy; pp. 1291–1294, September 1993. 1993

Nguyen, B. Q., Gehring, J., Müller, M., Stüker, S., and Waibel, A.: Multilingual Shifting Deep Bottleneck Features for Low-Resource ASR, ICASSP'14. ICASSP 2014

Niehues, J., and Waibel, A.: Using wikipedia to translate domain-specific terms in SMT. Proceedings of the Eight International Workshop on Spoken Language Translation (IWSLT) 2011. IWSLT 2011

Och, F. J., and Ney, H.: The alignment template approach to statistical machine translation. J. Computational Linguistics *30*/4, 417–449 (2004)

Takezawa, T., Morimoto, T., Sagisaka, Y., Campbell, N., Iida, H., Sugaya, F., Yokoo, A., and Yamamoto, S.: A Japanese-to-English speech translation system: ATR-MATRIX. Proceedings ICSLP'98, Sydney, Australia, S. 779–782, November 1998. ICSLP 1998

Waibel, A., and Fuegen, C.: Spoken language translation. IEEE Signal Processing Magazine May 2008

Waibel, A., Jain, A., McNair, A., Saito, H., Hauptmann, A., and Tebelskis, J.: JANUS: A speech-to-speech translation system using connectionist and symbolic processing strategies. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Toronto, May 1991. ICASSP 1991

Yang, J., Gao, J., Zhang, Y., Chen, X., and Waibel, A.: An automatic sign recognition and translation system. Workshop on Perceptual User Interfaces 2001, Orlando, USA, 1 November 2001. PUI 2001

Prof. Dr. Alexander Waibel
Karlsruher Institut für Technologie
Institut für Anthropomatik und Robotik
Interactive Systems Labs (ISL)
Adenauerring 2
76131 Karlsruhe
Bundesrepublik Deutschland
Tel.: +49 721 60 84 47 30
Fax: +49 721 60 77 21
E-Mail: waibel@kit.edu