

Man will mit einer neuen Technik Eindruck machen, und dann so etwas. »Good afternoon, Melanie, how are you?«, fragte Steve Clayton seine deutsche Kollegin Melanie Schoebel. Per Skype war sie zugeschaltet zur Worldwide Partner Conference von Microsoft in Washington. Eine Computerstimme schnarrte ihm hinterher: »Aber nach neuen Melanie, wie geht es dir.« Die Angesprochene, drei Tage nach dem deutschen Sieg bei der Fußball-WM im Juli vergangenen Jahres noch im Nationaltrikot, ließ sich von diesem Kauderwelsch nicht beirren und antwortete, jedes Wort überdeutlich artikulierend: »Mir geht es gut, wie geht es dir?« – »Well how is it me friends?«, übersetzte der Computer.

Man kann sich vorstellen, wie Microsofts PR-Leute, Entwickler und Wissenschaftler hinter den Kulissen in die Tischkante bissen. Eigentlich hatten sie ja vorführen wollen, wie Computer endlich die letzte Barriere für die Völkerverständigung niederreißen: Nachdem die Menschheit global per Telefon und Internet vernetzt ist, sollen Rechner die Grenzen zwischen den Tausenden verschiedenen Sprachen aufheben. Stattdessen geriet ihre Show zu einer Demonstration der Schwierigkeiten, die diesem Menschheitsstraum im Detail entgegenstehen.

Zugestehen muss man dem Computer, der zwischen Steve und Melanie vermittelte, dass die Konversation nach dem holprigen Anfang etwas flüssiger lief. Eine Simultanübersetzung für Telefongespräche oder Videokonferenzen ist im Prinzip möglich. Da-

von kann sich mittlerweile selbst überzeugen, wer Englisch oder Spanisch spricht: Seit Dezember dürfen Nutzer des Online-Telefoniedienstes Skype, den Microsoft im Jahr 2011 gekauft hat, die Übersetzung vom Englischen ins Spanische und umgekehrt testen. Im Lauf dieses Jahres sollen im Zweimonatstakt weitere Sprachpaare hinzukommen – auch ein Deutsch-Englisch-Dolmetscher.

Aus der Science-Fiction kennen wir das schon lange, etwa aus Douglas Adams' Roman *Per Anhalter durch die Galaxis*. Da stecken sich die Leute den sogenannten Babelfisch ins Ohr, der alles Fremdländische in die eigene Sprache dolmetscht.

**»Mir geht es gut, wie geht es dir?« wurde plötzlich zu:
»Well how is it me friends?«**

Der Protokolldroide C-3PO aus *Star Wars* beherrscht sechs Millionen verschiedene Sprachen. Jetzt wird der jahrzehntealte Traum Wirklichkeit, umsonst und für eine breite Masse: 300 Millionen Menschen nutzen den (in der Basisversion kostenlosen) Dienst weltweit. Wie die maschinelle Simultanübersetzung in den Alltag tritt und wie sie funktioniert, das sagt viel über ein Forschungsfeld aus, das wie kein zweites für die Hybris der Informatik steht: KI, die Künstliche Intelligenz.

Sie hat seit den fünfziger Jahren viel versprochen und wenig gehalten, auch in Sachen Digitaldolmetscher. Das Projekt Verbomobil, in dem die

Bundesregierung zwischen 1993 und 2000 umgerechnet 58 Millionen Euro versenkte, brachte außer heißer Luft nicht viel hervor. Sein Ziel war ein elektronisches Kästchen, das zwischen zwei Sprechern als Dolmetscher fungieren sollte. Heraus kam ein rudimentäres Spracherkennungssystem fürs Auto, das einsilbige Befehle entgegennahm.

Das Scheitern der KI-Forscher lehrt Kurioses über intelligente Maschinen: Das scheinbar Schwere ist für sie eher leicht, das vermeintlich Leichte schwer. Schach spielen lernten die Computer schnell, aber erst vergangene Woche konnten Google-Forscher in *Nature* verkünden, dass ihr Rechner sich simple Computerspiele wie *Space Invaders* oder *Pong* selbst beibringen konnte. Auch für die Sprachverarbeitung gilt: Komplexe wissenschaftliche Texte lassen sich leichter verstehen und übersetzen als nuschelige Unterhaltungen unter Teenagern.

Im vergangenen Jahrzehnt aber hat die KI auf einmal auf vielen Gebieten kräftige Fortschritte gemacht. Googles Autos gondeln fahrerlos durch Kalifornien, IBMs Watson gewinnt beim Fernsehquiz.

Auch Smartphones dienen als Babelfisch, besonders in Weltgegenden, wo Englisch als Lingua franca nicht funktioniert. Etwa in den Touristengebieten Südkoreas. Dort streckt die des Englischen nicht mächtige Wirtin einer kleinen Pension dem Gast kurzerhand ihr Samsung-Telefon entgegen. Der kann auf Englisch hineinsprechen, dass er gern zwei Doppelzimmer mit Meerblick hätte, und sofort plärrt sein Wunsch auf Koreanisch aus dem Handy-Lautsprecher.

So eine »Face to Face«-Übersetzung ermöglichen Apps wie Google Translate oder Jibbigo, eine zum Teil in Deutschland entwickelte Software, die inzwischen von Facebook gekauft wurde. Deren Güte ist akzeptabel: Sie missverstehen vieles, produzieren unfreiwillige Komik, aber der Sinn des Gesprochenen kommt meistens rüber.

Allerdings spricht der Nutzer bei solchen Anwendungen einen Satz in den Computer und wartet auf die Übersetzung. Und weil er weiß, dass eine Maschine zuhört, spricht er besonders deutlich und korrekt. »In einer Konversation ist das ganz anders«, erklärt Chris Wendt, bei Microsoft Programm-Manager fürs maschinelle Übersetzen. »Da gibt es falsche Anfänge, Stotterer, »was ich noch, ähm, was ich noch, ähm, sagen wollte, ähm«, kleine und größere Versprecher – all das will man nicht in der Übersetzung hören.«

Um Gesprochenes übersetzen zu können, braucht eine Software drei Teilprogramme: Das erste extrahiert aus dem Tonsignal zuerst Laute und ordnet diese Buchstaben und Wörtern zu. Das zweite übersetzt dieses schriftliche Ergebnis in die Fremdsprache. Und das dritte liest das Ergebnis der Übersetzung vor.

Teil drei ist heute kein Problem mehr. Wir haben uns an Computerstimmen in Callcentern und Siri auf dem iPhone gewöhnt. Spracherkennung und Übersetzung dagegen sind immer noch mit hohen Fehlerraten behaftet. »Und wenn man diese zwei unzuverlässigen Technologien kombiniert, dann multiplizieren sich die Fehler«, sagt Wendt. Anders ausgedrückt: Wird in der Ausgangssprache kein sinnvoller Satz erkannt, kann man von der Übersetzung keine Wunder erwarten.

Fortschritte erzielte die Spracherkennung ausgerechnet mit einer Technik, die in den achtziger Jahren schon einmal sehr populär war: Die sogenannten neuronalen Netzwerke sollen im Computer Strukturen simulieren, die dem Gehirn ähneln. Sie empfangen auf der einen Seite Rohdaten (Tonsignale etwa), die an mehrere Schichten von Neuronen weitergegeben werden, bis die letzte Schicht dann das Resultat ausgibt (etwa ein Transkript). Verarbeitet werden die Daten nicht aufgrund eines ausdrücklichen Programms. Das neuronale Netz lernt stattdessen anhand von Beispielen, die man ihm gibt, und passt daraufhin seine inneren »Nervenbahnen« an. Hat es genug trainiert, dann kann es – so die Hoffnung – selber Wörter erkennen.

ANZEIGE

Stimmt's?

Die Kolumne von Christoph Drösser
können Sie auch hören, täglich 6.50 Uhr.



Praktisch scheiterten die Systeme lange an der schwachbrüstigen Technik. Doch inzwischen existieren Chips, die mit 1000 Prozessoren wirklich große Netze mit vielen Zwischenschichten simulieren können (Deep Neural Networks). Microsoft gelang es, mit so einem System die Zahl falsch erkannter Wörter bei spontanem Gesprochenem von 30 auf 20 Prozent zu senken.

Und anders als Diktiersysteme für den Computer müssen neuronale Netze nicht auf einzelne Sprecher trainiert werden.

Diesen Fortschritt erkennt auch Alex Waibel vom Karlsruher Institut für Technologie (KIT) an. Der 58-Jährige war der Kopf hinter der App Jibbiggo. Schon 1991 präsentierte Waibels Team das erste deutschsprachige Übersetzungsprogramm für gesprochene Sprache, das allerdings mit einem Wortschatz von 500 Wörtern nur ein Gesprächsthema kannte: die Reservierung von Hotelzimmern. »Um allgemeine Gespräche zu übersetzen, brauchen Sie etwa 40 000 Wörter«, sagt Waibel.

Besonders faszinierend findet er, dass die Netze offenbar sehr ähnlich lernen wie ein Mensch. Obwohl man ihnen nur den Input und den gewünschten Output präsentiert, spezialisieren sich einzelne Neuronen bald auf bestimmte Details, etwa die kleinsten Lauteinheiten (Phoneme).

Funktioniert mit der Spracherkennung der erste Schritt, kommt dabei eine Folge von Wörtern heraus, ohne Punkt und Komma, Groß- und Kleinschreibung, inklusive Ähs und Ähms. Als Input für das Übersetzungsmodul taugt das noch nicht. Es ist ein Säuberungsschritt notwendig, und in dieses Reinemachen hat Microsoft besonders viel Energie gesteckt. Die Forscher dort haben ihre Systeme auf informelle Sprache getrimmt, wie sie vor allem in Sozialen Netzwerken gesprochen wird. Sie eliminieren Wiederholungen und Fehlstarts von Sätzen, fügen Groß- und Kleinschreibung ein sowie, noch wichtiger, Satzzeichen. So

werden aus unstrukturierten Wortflüssen Sätze – und der Satz ist die Grundeinheit der nachfolgenden Übersetzung.

Frühe Dolmetscherprogramme haben versucht, das Ausgangsmaterial grammatisch zu analysieren, Subjekt, Prädikat und Objekt zu finden. Bei spontaner Sprache misslingt das häufig, weil kaum einer von uns grammatisch korrekt spricht. Um 1990 entwickelten die IBM-Forscher Peter Brown und Robert Mercer einen radikal anderen Ansatz: Sie analysierten große Mengen von übersetzten Texten statistisch und ordneten so die Wörter der Ausgangssprache denen der Zielsprache zu. Aus der Häufigkeit von Nachbarschaften, in denen Wörter auftauchen, schlossen sie auf deren Bedeutung – etwa ob der »Hahn« nun ein Federvieh oder ein Wasserspender ist. Wenn man sich heute Webseiten mit Google oder Bing übersetzen lässt, ist diese statistische Analyse am Werk.

Während aber die neuronalen Netze bei der Spracherkennung immer noch besser werden, stagniert der Fortschritt beim Übersetzen. Mehr Trainingsbeispiele machen die Statistik nicht mehr viel besser. Wahrscheinlich müssen auch beim Übersetzen bald die tiefen Netze ran – nur weiß wohl noch niemand so richtig, wie.

Ein weiteres Problem ist, dass diese Übersetzungsverfahren nicht mit allen Sprachen gleich gut zurechtkommen. »Wenn ich ein Demo machen sollte, würde ich mir auch Englisch/Spanisch aussuchen.« So erklärt Alex Waibel, warum Microsoft als Debüt des Skype-Dolmetschers

ausgerechnet diese Variante wählte. Deutsch hingegen gilt als notorisch schwierig, vor allem weil bei uns so oft das Verb zerrissen wird. Spöttisch zitiert Waibel ein Beispiel dafür: »Ich schlage Ihnen einen Termin für nächste Woche nachmittags um 14 Uhr, wenn wir alle zusammen sind und sowieso an der Videokonferenz teilnehmen, vor!« Menschliche Dolmetscher wissen aus Erfahrung, dass »schlagen« hier nichts mit Gewalt zu tun hat – ein Computer hingegen übersetzt ins Englische drauflos mit »*I hit you ...*«. Waibel kennt das Problem besonders gut, weil am KIT bereits ein Computer Vorlesungen simultan ins Englische übersetzt: Studenten können die Übersetzung auf einer Website mitlesen wie Untertitel im Fernsehen.

Aber Microsoft will seine Probleme mit dem Deutschen bald beheben. Zwar legt sich Chris Wendt auf kein Datum für einen deutschen Skype-Übersetzer fest, verspricht aber: »Man kann dieses Jahr damit rechnen.« Noch läuft die Dolmetscherfunktion in einer öffentlichen Testphase. Um sie kontinuierlich zu verbessern, setzt Microsoft auf menschliche Helfer. Transkripte müssen überprüft und bewertet werden; dazu braucht man Kräfte, die beide Sprachen beherrschen. Wo die sitzen? Microsoft-Mann Wendt antwortet ausweichend: »Die finden Sie nicht in Kalifornien, in Deutschland oder in Japan.« Sondern? Es gebe Firmen, die so etwas zu einem sehr günstigen Preis machten. Man darf vermuten, dass diese polyglotten Telearbeiter in Ostasien zu finden sind.

Wie bitte – alle automatisch übersetzten Gespräche werden mitgeschnitten und nachher von Menschen abgehört? Exakt. Jeder Skype-Nutzer, der am Betatest teilnimmt, stimmt mit einem Klick dieser Bedingung zu. Später einmal soll die Einwilligung freiwillig sein, sagt Wendt und betont: »Wir nehmen den Datenschutz sehr ernst, wir anonymisieren, soweit es geht.« Und was ist, wenn die NSA vor Microsofts Tür steht und einen Skype-Mitschnitt zweier Verdächtiger haben möchte? »Selbst wenn wir es wollten, könnten wir das nicht für die NSA reproduzieren.« Das kann man nur glauben. Datenerhebung und -auswertung wecken Misstrauen, das gehört genauso zur Welt der digitalen Dolmetscher wie die gewaltige Rechenkraft, die sie erst ermöglicht.

Microsoft hat die Computerübersetzung nicht neu erfunden, traut sich aber als Erster mit seinem Babelfisch für die Internettelefonie an die schwierigste aller Gesprächssituationen heran: die ganz normale Unterhaltung. Weil die Firma es an dem weitgehend kostenlosen Dienst Skype probiert, verringert sie die Sprachbarrieren zwischen Menschen in aller Welt. Und sie trägt dazu bei, dass die lange geschmähte KI wieder Oberwasser bekommt. Viele Experten glauben, dass nun tatsächlich das Zeitalter der klugen Computer anbricht. »Diese Netze kommen dem Traum von intelligenten Maschinen näher«, sagt der Veteran Alex Waibel. Die Bildqualität von Skype beim Interview ist leider zu schlecht, um zu sehen, ob er dabei Tränen in den Augen hat.