

MASCHINENÜBERSETZUNG

Googlish für Anfänger

Wenn Computer Texte übersetzen, ist das Ergebnis oft grotesk und komisch. Wird eine neue, rein statistische Methode künftig menschliche Übersetzer überflüssig machen?

Wissen Sie, was „Knallkonzertkarten“ sind? So etwas kommt heraus, wenn das Sprachprogramm der Internet-Suchmaschine Google den Begriff „Pop concert tickets“ ins Deutsche übersetzt. Derartige Automatik-Übersetzungen sind oft unterhaltsam, meist aber nutzlos. So dichtet die Maschine fröhlich „Töpferende gibt passenschließen“, wenn es um ein wür-

diges Ende („fitting closure“) der Harry-Potter-Saga geht.

„Zurzeit ist die Qualität noch so, dass die Leute zweifeln, ob man das Programm wirklich nutzen kann“, räumt der 35-jährige Deutsche Franz Och ein, der das Übersetzerteam in der Google-Zentrale im Silicon Valley leitet. Computer sind nun mal Rechenmaschinen, keine Sprachexperten. Sie können sich zwar Millionen von Wörtern merken, egal ob auf Deutsch, Spanisch oder Kiswaheli – aber die Bedeutung dieser Wörter und die Welt, die sie beschreiben, bleiben dem Elektronenhirn fremd.

Deshalb kommt der Automat bei Ausdrücken wie „breaking records“ schnell ins Schleudern: Gehen da Schallplatten kaputt, oder werden Rekorde aufgestellt? Menschen orientieren sich am Zusammenhang; der Rechner, für den alles gleichermaßen sinnvoll oder sinnlos ist, kann nur raten.

Doch nun gibt es Hoffnung: Für bisher drei Sprachen – Arabisch, Chinesisch und Russisch – zeigt Google seit kurzem, dass Computer durchaus als Dolmetscher taugen könnten. Die

Ohne Hirn geht es noch nicht. Eine Rechenmaschine, die schwierige Texte korrekt übersetzt, ist bislang Zukunftsmusik

neue Übersetzungsfunktion, bisher noch im unreifen „Beta“-Stadium, erlaubt es etwa, Nachrichten des arabischen Senders Al Jazeera zu lesen. Tatsächlich liefert das System, das bisher nur in Kombination mit Englisch angeboten wird, oft deutlich verständlichere Ergebnisse als die Konkurrenz: Beim Wettbewerb der automatischen Übersetzungsprogramme, den die US-Technikbehörde NIST regelmäßig ausschreibt, landete das Google-Team schon im ersten Anlauf auf Platz eins.

Die neue Methode umgeht die Probleme traditioneller Lösungsansätze. Lange Zeit haben Computerlinguisten versucht, Maschinen Sprache beizubringen, indem sie die Automaten mit Vokabeln und aufwendigen Regelwerken zu Syntax und Grammatik fütterten. So ähnlich lernen Menschen, so sollte es auch bei Computern funktionieren. Das jedenfalls dachten die Urväter des Forschungszweigs, als sie vor einem halben Jahrhundert begannen, mit ihren raumfüllenden Rechenkisten erste Sätze aus dem Russischen ins Englische zu übertragen.

Doch schnell zeigte sich, sagt Franz Och, dass „Sprache viel komplexer ist, als man ursprünglich geglaubt hat“. Kaum eine Regel ohne Ausnahme, kaum ein Satz ohne Zwischentöne,

Doppelbedeutungen, Stolpersteine. Das erklärt, weshalb selbst nach Jahrzehnten ausgiebiger Forschung viele automatisch übersetzten Texte wie das Gestammel eines Wahnsinnigen wirken. „Aus der Sicht der Maschine gibt es einfach zu viele Wahlmöglichkeiten – jedes Wort hat mehrere Bedeutungen, und die Zahl der Kombinationen ist endlos“, sagt Kevin Knight, Informatiker an der University of Southern California.

Weil es die menschliche Welt der Technik so schwer macht, sie zu verstehen, erdachten Forscher in den 1990er Jahren eine Methode, die ideal für Rechenhirne ist, die nur in Einsen und Nullen denken. Sie nahmen von Menschen erstellte Übersetzungen und ließen den Computer die Texte mathematisch vergleichen. So lernt der Rechner allein auf statistischer Basis, welche Begriffe zusammengehören. Wenn das englische „the“ häu-

fig neben „house“ steht und auf der Gegenseite entsprechend das deutsche „das“ neben „Haus“, ergibt sich eine hohe Wahrscheinlichkeit, dass die Übersetzung passt. Mit diesem Wissen kann die Maschine schließlich Texte übersetzen, ohne die geringste Ahnung von deren Bedeutung zu haben.

„Im Vergleich sind die statistischen Systeme haushoch überlegen“, urteilt Alex Waibel, Informatiker in Karlsruhe und an der Carnegie Mellon University in Pittsburgh. „Man kommt erstaunlich weit mit den primitiven Zählmethoden.“ Zwei Trends helfen den Forschern: immer leistungsfähigere Rechenhirne, die sich immer flinker durch gigantische Datenberge wühlen, und das Internet. Viele der Übersetzungen, die sie zum Trainieren ihrer Maschinen benötigen, finden die Wissenschaftler auf den Webseiten der Vereinten Nationen, der Europäischen Union und globaler Konzerne. Auch deshalb, vermutet Waibel, haben Organisationen, denen große Datenmengen zur Verfügung stehen, Erfolg mit statistischen Übersetzernsystemen.

Probleme bereiten allerdings ungewöhnliche Sprachkombinationen, für die wenige Referenzübersetzungen vorliegen, etwa Norwegisch-Thai. Denn statistische Modelle erfordern Datenschätze mit Hunderten von Millionen Einträgen. Auf lange Sicht aber erwarten die Forscher auch hier Fortschritte. Alex Waibel arbeitet an statistischen Übersetzungsalgorithmen, die wesentlich weniger Daten benötigen.

Wird irgendwann der Tag kommen, an dem Maschinenübersetzungen zu gleich guten oder gar besseren Ergebnissen führen als menschliche? Waibels Kollege Kevin Knight sieht selbst bei der Poesie gute Chancen, den Menschen zu übertrumpfen: „Eine Maschine könnte viel besser dazu geeignet sein, nach passenden Reimen zu suchen.“

Was Johann Wolfgang von Goethe wohl dazu sagen würde? Vielleicht dies: „Die Anzeige hören vermutlich I, gleichwohl der Glaube zu mir fehlt.“ Das war Googlish und soll heißen: „Die Botschaft hör ich wohl, allein mir fehlt der Glaube.“ Karsten Lemm

