

Für das Europa der 23 Sprachen: Der Computer lernt das Übersetzen

Das neue europäische Projekt EuroMatrix entwickelt automatische Übersetzungssysteme für alle Amtssprachen der EU

Von Jahr zu Jahr wachsen Zahl und Umfang der Übersetzungen in Wirtschaft, Verwaltung, Politik und Kultur. Hauptauftraggeber sind exportorientierte Wirtschaftszweige, internationale Organisationen einschließlich der Europäischen Union und der Vereinten Nationen sowie multinationale Unternehmen. Der explodierende Bedarf an Textübersetzungen gehört zu den direkten Folgen von Handelsfreiheit, Globalisierung, europäischer Integration und internationaler Zusammenarbeit. Das Heer der Übersetzer wächst.

Mit dem allgemeinen Trend zur Spezialisierung wachsen aber auch die Anforderungen an die Fähigkeiten der Fachübersetzer. So steigen denn auch die Kosten für qualitativ akzeptable Übersetzungen. Technische und juristische Übersetzungen zwischen den Sprachen der Europäischen Union zum Beispiel kann die Europäische Kommission nicht nach Indien, in die Philippinen oder andere Billiglohnländer auslagern.

Natürlich verwenden die Übersetzer, die für große Unternehmen und Organisationen arbeiten, bereits seit Jahren spezialisierte Computerprogramme, die aus großen Mengen früherer Übersetzungen passende Übersetzungsvorschläge fischen, jedoch können diese sogenannten Übersetzungsgedächtnisse keine Sätze übertragen, die sie noch nicht im Speicher haben.

Das kann nur die automatische Übersetzung, von der wir oft lesen, dass sie langfristig die einzige realistische Lösung zur Bewältigung der Kommunikationsprobleme einer multilingualen Weltgesellschaft darstellt. Sie wird auch von immer mehr Experten als eine notwendige Voraussetzung für die nachhaltige Bewahrung der mannigfaltigen Sprachen und Kulturen unseres Planeten gesehen.

Während sich aber alle Analysten über den immensen Bedarf an automatischer Übersetzung einig sind, gehen die Meinungen über die Möglichkeiten der automatisierten Übertragung von anspruchsvollen Texten jedoch stark auseinander. Während viele Experten noch bezweifeln, dass Maschinen brauchbare Übersetzungen liefern können, erzielen einige Unternehmen durch den Einsatz des Computers bereits erhebliche Kosteneinsparungen.

Die Wahrheit liegt zwar nicht ganz in der Mitte und doch haben beide Seiten auf ihre Weise recht. Selbst die völlig veralteten und qualitativ unzureichenden Übersetzungssysteme, die heute sehr begrenzt in Wirtschaft und Verwaltung eingesetzt werden, ermöglichen bereits erhebliche Kosteneinsparungen. Der eigentliche Durchbruch der maschinellen Übersetzung steht aber noch bevor. Und das zu einer Zeit, in der die Wirtschaft die Hoffnung auf Automatisierung fast aufgegeben hat.

Jahrzehntelang waren zu hohe Erwartungen immer wieder enttäuscht worden. Von 1975-1995 hatte die Forschung mehr oder weniger auf der Stelle getreten. Seit 1995

gibt aber es einen bemerkenswerten neuen Trend, der erhebliche Bewegung in das Forschungsgebiet gebracht hat: die Entwicklung statistischer Übersetzungsmethoden. Statt mit handgefertigten komplexen Regelsystemen zu arbeiten, deren Vervollkommnung Jahrzehnte konzentrierter Arbeit bedeutet, konnte man mithilfe automatischer Lernverfahren den Reichtum bestehender Übersetzungen nutzen. Das Ergebnis sind Computerprogramme, die sich das Übersetzen selbst beigebracht haben.

So wie der menschliche Übersetzer die Zielsprache perfekt beherrschen muss, weil es schwerer ist, stilistisch einwandfreie Sprache zu formulieren, als nur Inhalte zu verstehen, so können auch statistische Übersetzungsprogramme für die Sprache perfektioniert werden, in die sie übersetzen. Auch das geschieht durch automatisches Lernen. Hierfür benötigt man lediglich große Vorräte an einsprachigen Texten, die man heute auf dem Internet zuhauf findet. Noch sind diese Autodidakten unter den Übersetzungssystemen bei weitem nicht perfekt, doch sie werden von Jahr zu Jahr besser.

In einem neuen internationalen Forschungsprojekt, das im 6. Rahmenprogramm der Europäischen Union gefördert wird, sollen technologische Durchbrüche in der automatischen Übersetzung erzielt werden. Dafür sollen neue Kombinationen der besten bestehenden Verfahren für regelbasierte und für statistische Übersetzung ausprobiert werden. Außerdem wird untersucht werden, wie sich statistische Übersetzungsverfahren durch linguistische Methoden verbessern lassen. Das ehrgeizige Vorhaben wird Systeme für Übersetzungen zwischen den 23 Amtssprachen der EU-Länder entwickeln und testen. Ein wesentlicher Bestandteil des Vorhabens ist auch die Organisation von Wettbewerben zwischen den besten Übersetzungssystemen, zu denen alle Forschungszentren und Firmen aus Europa aber auch aus anderen Teilen der Welt eingeladen sind. Natürlich wird es nicht möglich sein, alle Sprachpaare gleich gut zu bearbeiten. Aber ein praktisches Ergebnis des Projekts wird auch eine ständig fortgeschriebene Bestandsaufnahme des Technologiestandes für alle Sprachen sein. Dieser Überblick in Form einer großen tabellarischen Matrix hat dem Projekt den Namen gegeben.

Die Führung des Konsortiums liegt bei der Universität des Saarlandes; Koordinator ist Prof. Hans Uszkoreit. Er wird unterstützt durch Dr. Philipp Koehn von der Universität Edinburgh, dem technisch/wissenschaftlichen Koordinator des Projekts. Weitere Partner sind Forschungsgruppen an der Prager Karlsuniversität unter Leitung von Prof. Jan Hajic und dem Zentrum CELCT in Trento unter der Leitung von Prof. Amedeo Capelli.

Wichtig ist die Teilnahme der Industrie: die deutsche Firma GroupWare AG bringt ihr LOGOS System ins Projekt ein, eines der großen und meistgenutzten kommerziellen Übersetzungssysteme, das im vorletzten Jahr in einer Open-Source Version verfügbar gemacht wurde. Die Firma Morphologic in Budapest hat digitale Wörterbücher für viele europäische Sprachen hergestellt und verfügt über ihre eigene Übersetzungstechnologie.

Das gerade begonnene Vorhaben wird 30 Monate dauern und ca. 2,5 Mio Euro kosten.

Kontakt: Dr. Andreas Eisele
Universität des Saarlandes
eisele@dfki.de