

Die Übersetzungsmaschinisten

Coca, Lama, Pampa: Quechua hat Spuren im Deutschen hinterlassen. Doch die Verbreitung der südamerikanischen Sprache nimmt ab. Von Martin Bieri

Quechua ist eine Familie autochthoner Sprachen Südamerikas, die heute noch von etwa zehn Millionen Menschen im Andenraum gesprochen werden. «Langfristig ist Quechua bedroht», sagt Martin Volk, Professor am Institut für Computerlinguistik der Universität Zürich, «es ist die prestigearme Sprache der ländlichen Unterschichten.» Viele Kinder wachsen bilingual auf: Quechua ist die Sprache ihrer Eltern und Grosseltern, Spanisch die Medien- und Unterrichtssprache. Sozialer Aufstieg setzt Spanisch voraus. Volk und sein Team wollen die Grenzen der beiden Sprachwelten durchlässiger machen und damit das indigene Idiom stützen. Sie arbeiten an einem Computerprogramm zur automatischen Übersetzung von Spanisch nach Quechua in Form einer Internetseite, die in der Anwendung ähnlich wie Google-Translate funktionieren und beispielsweise im Journalismus Verwendung finden soll.

Mehr Texte, besseres Programm

Geschrieben wird Quechua erst seit der spanischen Eroberung Südamerikas, noch heute existiert wenig Literatur, weil sich kein Sprachstandard hat durchsetzen können. Das aber wäre für den Erhalt der Sprache sehr wichtig, sagt Volk. Das Team der Universität Zürich macht das Quechua der peruanischen Provinz Cusco zur Grundlage seiner Forschung, weil es die am besten dokumentierte Varietät und die gleichnamige Stadt ein kulturelles Zentrum ist.

Maschinelle Übersetzungssysteme basieren heute oft auf statistischen Methoden. Sie bestimmen sprachübergreifende Wortfolgen nach der Häufigkeit ihres Vorkommens. Dafür braucht es zwei grosse Textsammlungen, die miteinander verglichen werden. Dem gegenüber stehen regel-

basierte Systeme. Sie analysieren grammatische Strukturen, erkennen Muster und stellen sie in so genannten Syntaxbäumen dar. Diese werden dann von der Ausgangs- in die Zielsprache transferiert. Für die Übersetzung von Spanisch nach Quechua kommt nur ein regelbasiertes Verfahren in Frage, weil es zu wenig Paralleltexte in den beiden Sprachen gibt. Doch das Zürcher Team baut die Statistik als nachträgliches und erweiterbares Kontrollinstrument ein, womit eine neuartige Kombination der beiden Ansätze entsteht. Dem Sprachenpaar Spanisch-Quechua stellt es zudem das typologisch ähnlichere Spanisch-Deutsch gegenüber, um die Verbesserung der Übersetzungsqualität bei höherer statistischer Vergleichbarkeit vorauszusagen. Je mehr Texte vorliegen, desto besser wird das Programm. Die Zürcher Computerlinguisten

arbeiten eng mit Wissenschaftlern in Cusco zusammen, denen sie das System zur Weiterentwicklung übergeben werden.

Ein zusätzliches wichtiges Produkt des Projekts ist eine Ausgangssammlung von Texten, die von Menschen übersetzt und ebenfalls über das Internet zugänglich gemacht werden. «Wir zeigen die Sprache in ihrer Verwendung», sagt Volk. Ein bestimmtes Wort wird nicht nur übersetzt, sondern auch in seinen verschiedenen Flexionsvarianten gezeigt. Der Computer nähert sich so menschlicher Sprach- und Übersetzungskompetenz, die der maschinellen nach wie vor überlegen ist. Martin Volk glaubt nicht daran, dass sich das bald ändern wird. Trotzdem kann die Sprachtechnologie schon jetzt dazu beitragen, eine Sprache durch ihre Verknüpfung mit anderen zu stärken.



Zweisprachigkeit im peruanischen Cusco: Das Hotel «Verstecktes Haus» einmal in Spanisch, einmal in Quechua angeschrieben. Bild: Martin Volk