

ÜBERSETZUNG

C-Star II dolmetscht im Internet sechs Weltprachen fließend

Ein Durchbruch in der Übersetzung von Spontansprache ist einem internationalen Forscherteam gelungen. Computer sind damit in der Lage, über das Internet zu dolmetschen.

Derzeit kann das System, das im Rahmen des C-Star-II-Projekts entwickelt wurde, je nach Sprache zwischen 10 000 und 100 000 Wörter in die Sprachen Englisch, Japanisch, Koreanisch, Italienisch, Französisch und Deutsch übersetzen. C-Star steht für Consortium for Speech Translation Advanced Research und besteht aus sechs Partnern aus den Ländern Deutschland, USA, Italien, Frankreich, Korea und Japan. Es beschäftigt sich mit automatischen Systemen zur Übersetzung spontan gesprochener Sprache. „Frühere Demonstrationen

haben gezeigt, daß Sprachübersetzung möglich ist, aber bislang war das Vokabular begrenzt, und die Sätze mußten perfekt formuliert und sauber ausgesprochen sein“, erklärt Alex Waibel, derzeitiger Vorsitzender des C-Star Konsortiums und Professor an der Universität Karlsruhe und der Carnegie Mellon University, Pittsburgh. „Jetzt kann das Spracherkennungssystem den nachlässigen Sprechstil verstehen, in dem Menschen normalerweise miteinander reden.“

Nachlässige Aussprache, Versprecher, Dehnungen, emotional bedingte Veränderungen der Stimme durch Lachen oder Denkpausen mitten im Wort lassen normale Spracherkennungssysteme versagen. Aktuelle Programme zur Textübersetzung scheitern selbst

bei völlig fehlerfreier Erkennung an mangelnder Grammatik und idiomatischen Ausdrücken. Um das Problem in den Griff zu bekommen, müssen alle Teile des Systems auf die Übersetzung gesprochener Sprache ausgelegt werden. Die Kerntechnologie im C-Star-II-Konsortium sind Systeme zur Übersetzung spontan gesprochener Sprache. Zur Spracherkennung werden von der Universität Karlsruhe und der Carnegie Mellon University je nach Bereich Module aus Janus und Verbomobil eingesetzt, die Geräusche und andere spontansprachliche Effekte wie Ähs, Hmms, Unterbrechungen und Verzögerungen explizit übertragen. Die Spracherkennung können dem Übersetzungssystem zudem Informationen zu ähnlich klin-

genden Alternativen geben. Das Übersetzungsmodul dolmetscht dann zunächst in die Zwischensprache Interlingua. Diese Sprache repräsentiert ohne Mehrdeutigkeiten die tatsächliche Bedeutung einer Äußerung. Von da aus wird dann in eine echte Sprache übersetzt. Durch die Arbeitsteilung in C-Star genügt es, wenn jeder Partner einen Übersetzer zwischen Interlingua und seiner Muttersprache entwickelt. Bislang befindet sich die Technik im Prototypenstadium. Bis elektronische Dolmetscher im PC oder am Handgelenk zum Alltag werden, ist weitere Forschung in der Spracherkennung zum Beispiel in Hinblick auf Dialekte der Sprecher und die engere Kopplung an multimediale Eingabegeräte erforderlich. wk