

EINLADUNG

Zeit: Mittwoch, 18. Januar 2012, 10.00 Uhr

Ort: Raum 5054, Ahornstr. 55

Referent: Dipl.-Informatiker Daniel Stein

Thema: Soft Features for Statistical Machine
Translation of Spoken and Signed
Languages

Ziel der statistisch-maschinellen Übersetzung ist eine Überführung von unbekanntem Sätzen aus einer Quellsprache in eine Zielsprache. Zu diesem Zweck werden durch statistische Analyse aus bestehenden zweisprachigen Datensammlungen Regeln zur Übersetzung extrahiert. Mittels eines probabilistischen Ansatzes werden mehrere alternative Sätze generiert, die durch Merkmalsfunktionen bewertet werden. Die Alternative mit der höchsten zugewiesenen Wahrscheinlichkeit wird dann als Übersetzung ausgegeben.

In diesem Vortrag wird der Einfluß von mehreren, meist linguistisch motivierten, Merkmalsfunktionen auf die Qualität der statistisch-maschinellen Übersetzung untersucht. Durch die beschriebenen Funktionen wird keine der Alternativen völlig verworfen, so dass die Variabilität des Übersetzers erhalten bleibt. Analysiert werden Sprachpaare gesprochener Sprachen (z.B. Deutsch, Englisch oder Chinesisch) untereinander.

Der Vortrag beschäftigt sich auch mit der automatischen Gebärdensprachübersetzung als Spezialfall der maschinellen Übersetzung. Es wird zunächst auf die Besonderheiten der Gebärdensprache als Kommunikationsmittel eingegangen. Bestehende Datensammlungen werden analysiert, außerdem werden der RWTH-Phoenix Corpus und der Corpus NGT vorgestellt, die im Rahmen

der Doktorarbeit erweitert wurden bzw. entstanden sind. Dann werden die beiden Datensammlungen eingehend auf Eignung zur maschinellen Übersetzung geprüft. Aufgrund ihrer relativ geringen Größe ist eine Anpassung der bestehenden Verfahren sinnvoll. So kann durch Einsatz von (in der maschinellen Übersetzung eher unüblichen) Kreuzvalidierung eine signifikante Verbesserung der Übersetzungsqualität erreicht werden. Auch durch morpho-syntaktische Vor- und Nachverarbeitung gelingt es in der Übersetzung, einen wohlgeformteren Satz zu finden bzw. mit die zusammengesetzten Wörter im Deutschen besser zu erfassen. Neben dem Vergleich zweier Übersetzungsparadigmen werden auch Verfahren aus der Systemkombination erfolgreich auf diese Sprachrichtung eingesetzt.