

Nachrichten

Habilitation im Fach Computerlinguistik

Dr. Nico Weber, Mitglied der GLDV und Leiter des Arbeitskreises „Lexikographie“, hat am 13. Februar 1998 seine Habilitation vor der Philosophischen Fakultät der Universität Bonn abgeschlossen und die Venia Legendi für das Fach Computerlinguistik erworben. Die Habilitationsschrift hat das Thema „*Die Semantik von Bedeutungsexplikationen*“ und wird in Kürze im Verlag Peter Lang erscheinen.

Die GLDV gratuliert herzlich!

Projekte

Quantitative Verfahren zur Zuordnung von Präpositionalphrasen

Hagen Langer (Universität Osnabrück), Stephan Mehl (Universität Duisburg), Martin Volk (Universität Zürich)

Im September 1996 fand im Rahmen der KONVENS in Bielefeld unter der Leitung von Andreas Mertens und Marion Schulz (FernUniv. Hagen) sowie von Stephan Mehl ein Workshop zum Problem der PP-Zuordnung statt. Im Anschluß an diesen Workshop bildeten Stephan Mehl, Hagen Langer und Martin Volk eine Arbeitsgruppe mit dem Ziel, die Effizienz quantitativer Verfahren der PP-Zuordnung für das Deutsche zu untersuchen. Angeregt durch die Ergebnisse von Hindle/Rooth 1993 sind in den letzten Jahren zahlreiche Arbeiten entstanden, die das Problem der PP-Zuordnung für das Englische mit statistischen Mitteln erfolgreich angehen. Diese Ergebnisse sind aber aufgrund der freieren Wortstellung nicht ohne weiteres auf das Deutsche übertragbar. Außerdem setzen viele einschlägige Arbeiten eine umfangreiche Treebank voraus, die für das Deutsche noch nicht existiert.

Ausgangspunkt der gemeinsamen Untersuchungen ist die Tatsache, daß Valenzerwartungen eine bedeutende Rolle bei der syntaktischen Disambiguierung spielen. In vielen Fällen kann eine Entscheidung über die korrekte PP-Zuordnung dadurch getroffen werden, daß entweder das Verb oder eines der möglichen Bezugsnomina eine Präpositionalphrase mit der betreffenden Präposition erwartet. Unter „Erwartung“ sind dabei in erster Linie obligatorische oder fakultative Valenzstellen zu verstehen, darüber hinaus aber auch statistisch signifikante Kookkurrenzen wie der Ausdruck „Familien mit Kindern“ u. ä. Leider ist für das Deutsche kein maschinenlesbares Lexikon mit vollständigen Valenzangaben verfügbar. Freundlicherweise stellte das Zentrum für elektronische Ressourcen europäischer Sprachen (ZERES) in Bochum uns ein Lexikon zur Verfügung, das für einen Teil der deutschen Verben Valenzangaben auflistet. Die in Nijmegen entwickelte CELEX-Datenbank enthält ebenfalls zahlreiche Valenzangaben. Insbesondere für Nomina fehlen uns solche Angaben jedoch. Darüber hinaus gibt es Sätze, in denen sowohl Verb als auch Nomina konkurrierende Erwartungen an eine PP stellen. Für solche Fälle wäre eine numerische Gewichtung der Valenzangaben interessant.

Unsere bisherigen Resultate bestätigen, daß quantitativ ermittelte Erwartungswerte auch für das Deutsche einen wichtigen Beitrag zur Disambiguierung leisten können. Allerdings ist es für weniger häufige Lexeme und für solche mit fakultativen Valenzstellen schwierig, genügend Belege zu finden, um Valenzstellen automatisch zu bestimmen bzw. quantitativ zu gewichten. Weitere Schwerpunkte unserer Arbeit neben der Valenzproblematik sind u. a. die Ermittlung fester Wendungen und die Einbeziehung grober semantischer Informationen, z. B. die Identifizierung von Zeitangaben. Hagen Langer untersucht darüber hinaus den Einsatz von Bindungswahrscheinlichkeiten in einem probabilistischen Parser. Ein weiterer Aspekt ist die Untersuchung fachtextspezifischer Unterschiede.

Zwischenbericht: Langer, Hagen/Mehl, Stephan/Volk, Martin: „Hybride NLP-Systeme und das Problem der PP-Anbindung.“ In: Workshop „Hybride konnektionistische, statistische und symbolische Ansätze zur Verarbeitung natürlicher Sprache“, KI 97, Freiburg, <http://www.dfki.de/~busemann/ki97/ki97-ws03.html>.

Literatur: D. Hindle/M. Rooth (1993): Structural ambiguity and lexical relations. *Computational Linguistics* 19(1), 103-120.