AIMS

Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung

Lehrstuhl für Theoretische Computerlinguistik Universität Stuttgart

2003 • Vol. 9 • No. 2

Experiments on the Automatic Induction of German Semantic Verb Classes

Sabine Schulte im Walde

AIMS
Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung
Universität Stuttgart

ISSN 1434-0763

AIMS

In January 1994 we initiated a series of working papers: AIMS. AIMS stands for 'Arbeitpapiere des Instituts für Maschinelle Sprachverarbeitung' (Working Papers of the Department of Natural Language Processing), at University of Stuttgart. So far every issue of AIMS is centered around a specific topic, either by presenting individual dissertations or by compiling articles and master theses on this topic written by members of the IMS and associates. The work has pre-publication status and might appear in other form elsewhere. We invite other departments and sites to exchange their reports with us. Otherwise our working papers are also available at the cost of 20 EUR per issue plus costs for shipping and handling. To initiate an exchange, or to obtain the working papers on a regular or issue-wise basis, contact:

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Azenbergstr. 12
70174 Stuttgart, Germany

aims@IMS.Uni-Stuttgart.DE
Tel.: +49 (0)711 121 1379
Fax: +49 (0)711 121 1366
www.ims.uni-stuttgart.de/phonetik/

Previous Issues of AIMS

Vol. 1 (1) 1994: Wigner Distribution in Speech Research. Master thesis by Wolfgang Wokurek (in German), papers by Grzegorz Dogil, Wolfgang Wokurek, and Krzysztof Marasek (in English), and a bibliography.

Vol. 2 (1) 1995: Sprachentstörung. Doctoral Dissertation, University of Vienna, 1994 by Wolfgang Wokurek (in German with abstract in English). Full title: Sprachentstörung unter Verwendung eines Lautklassendetektors (Speech enhancement using a sound class detector).

Vol. 2 (2) 1995: Word Stress. Master thesis by Stefan Rapp (in German) and papers mostly by Grzegorz Dogil, Michael Jessen, and Gabriele Scharf (in English).

Vol. 2 (3) 1995: Language and Speech Pathology. Master theses by Gabriele Scharf and by Jörg Mayer (in German) and papers by Hermann Ackermann, Ingo Hertrich, Jürgen Konczak, and Jörg Mayer (mostly in German).

Vol. 3 (1) 1997: Tense versus Lax Obstruents in German. Revised and expanded version of Ph.D. Dissertation, Cornell University 1996 by Michael Jessen (in English). Full title: Phonetics and phonology of the tense and lax obstruents in German.

Vol. 3 (2) 1997: Electroglottographic Description of Voice Quality. Habilitationsschrift, University of Stuttgart, by Krzysztof Marasek (in English).

Vol. 3 (3) 1997: Aphasie und Kernbereiche der Grammatiktheorie (Aphasia and core domains in the theory of grammar). Doctoral Dissertation, University of Stuttgart, by Annegret Bender (in German with abstract in English).

Vol. 3 (4) 1997: Intonation und Bedeutung (Intonation and meaning). Doctoral Dissertation, University of Stuttgart, by Jörg Mayer (in German with summary in English).

Vol. 3 (5) 1997: Koartikulation und glottale Transparenz (Coarticulation and glottal transparency). Doctoral Dissertation, University of Bielefeld, by Kerstin Vollmer (in German with abstract in English).

Vol. 3 (6) 1997: Der TFS-Repräsentationsformalismus und seine Anwendung in der maschinellen Sprachverarbeitung (The TFS Representation Formalism and its Application to Natural Language Processing). Doctoral Dissertation, University of Stuttgart, by Martin C. Emele (in German with summary in English).

Vol. 4 (1) 1998: Automatisierte Erstellung von Korpora für die Prosodieforschung (Automated generation of corpora for prosody research). Doctoral Dissertation, University of Stuttgart, by Stefan Rapp (in German with abstract in English).

Vol. 4 (2) 1998: Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese (Theory-based modelling of German intonation for speech synthesis). Doctoral Dissertation, University of Stuttgart, by Gregor Möhler (in German with abstract in English).

Vol. 4 (3) 1998: Inducing Lexicons with the EM Algorithm. Papers by M. Rooth, S. Riezler, D. Prescher, S. Schulte im Walde, G. Carroll, F. Beil (in English).

Vol. 5 (1) 1999: Probabilistic Constraint Logic Programming. Doctoral Dissertation, University of Stuttgart, by Stefan Riezler (in English).

Vol. 5 (2) 1999: Timing bei Dysarthrophonie (Timing in Dysarthria). Doctoral Dissertation, University of Stuttgart, by Gabriele Scharf (in German with abstract in English).

Vol. 5 (3) 1999: Semantikkonstruktion (Semantic Construction). Doctoral Dissertation, University of Stuttgart, by Michael Schiehlen (in German with abstract in English).

Vol. 6 (1) 2000: YAP: Parsing and Disambiguation with Feature-Based Grammars. Doctoral Dissertation, University of Stuttgart, by Helmut Schmid (in English).

Vol. 6 (2) 2000: Semantikbasierter Transfer - Maschinelle Übersetzung auf Mengen linguistischer Constraints. Doctoral Dissertation, University of Stuttgart, by Michael Dorna (in German).

Vol. 6 (3) 2000: The Contribution of Intonation to the Perception of Foreign Accent. Doctoral Dissertation, University of Stuttgart, by Matthias Jilka (in English).

Vol. 6 (4) 2000: Research Papers from the Phonetics Lab. Contributions by Kathrin Claßen, Grzegorz Dogil, Michael Jessen, Bernd Möbius, Karin Müller, Detlef Prescher and Wolfgang Wokurek (most papers in English).

Vol. 7 (1) 2001: Automatische Segmentierung, Verwaltung und Abfrage von Korpora gesprochener Sprache (Automated Segmentation, Administration and Retrieval of Spoken Language Corpora). Doctoral Dissertation, University of Stuttgart, by Marcus L. Fach (in German with summary in English).

Vol. 7 (2) 2001: Topics in the Clausal Syntax of German. Doctoral Dissertation, University of Stuttgart, by Judith Berman (in English).

Vol. 7 (3) 2001: Formal and Computational Aspects of Optimality-Theoretic Syntax. Doctoral Dissertation, University of Stuttgart, by Jonas Kuhn (in English).

Vol. 7 (4) 2001: German and Multilingual Speech Synthesis. Habilitationsschrift, University of Stuttgart, by Bernd Möbius (in English).

Vol. 8 (1) 2002: The Automatic Acquisition of Lexical Information from Portuguese Text Corpora with a Probabilistic Context-free Grammar. Doctoral Dissertation, University of Stuttgart, by Erika Fonseca de Lima (in English).

Vol. 8 (2) 2002: EM-basierte maschinelle Lernverfahren für natürliche Sprachen (EM–Based Maschine Learning in Natural Language Processing). Doctoral Dissertation, University of Stuttgart, by Detlef Prescher (in German with summary in English).

Vol. 8 (3) 2002: Probabilistic Syllable Modeling Using Unsupervised and Supervised Learning Methods. Doctoral Dissertation, University of Stuttgart, by Karin Müller (in English).

Vol. 8 (4) 2002: Ein Suchwerkzeug für syntaktisch annotierte Textkorpora. Doctoral Dissertation, University of Stuttgart, by Wolfgang Lezius (in German with summary in English).

Vol. 9 (1) 2003: Implementing and Documenting Large-Scale Grammars – German LFG. Doctoral Dissertation, University of Stuttgart, by Stefanie Dipper (in English with summary in German).

Vol. 9 (2) 2003: Experiments on the Automatic Induction of German Semantic Verb Classes. Doctoral Dissertation, University of Stuttgart, by Sabine Schulte im Walde (in English with summary in German).

Experiments on the Automatic Induction of German Semantic Verb Classes

Von der Philosophisch-Historischen Fakultät der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Philosophie (Dr. phil.) genehmigte Abhandlung

Vorgelegt von

Sabine Schulte im Walde

aus Rheine

Hauptberichter: Prof. Dr. h.c. Hans Kamp, PhD

Mitberichter: Assistant Prof. Chris Brew, D.Phil.

Tag der mündlichen Prüfung: 18. Juni 2003

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart

2003

Acknowledgements

The work on this thesis was performed while I was a member of the DFG-funded PhD program 'Graduiertenkolleg' *Sprachliche Repräsentationen und ihre Interpretation*. I was supervised internally by Hans Kamp at the IMS, and externally by Chris Brew at the Ohio State University. To both of them as well as to Helmut Schmid for his permanent unofficial supervision I am grateful for their interest in my work, their motivation and feedback, and especially for the fact that their doors were always open for me. In addition, Stefan Evert took part in my work by repeatedly discussing mathematical questions and problems with me; and the evaluation of subcategorisation frames in Chapter 3 was performed in an enjoyable collaboration with *Bibliographisches Institut & F. A. Brockhaus AG*.

On various occasions I had the opportunity to receive valuable motivation and feedback from (computational) linguists working in related research. For this I wish to thank Ana Fernández, Mirella Lapata, Diana McCarthy, Paola Merlo, Patrick Saint-Dizier and Gloria Vázquez who followed up my invitations to Stuttgart; Eric Joanis, Anna Korhonen, Lillian Lee and Suzanne Stevenson who were willing to meet with me during my visits, and Uli Heid and John Nerbonne who were around in Stuttgart. Furthermore, I received many valuable comments at various conferences and workshops.

I am grateful to the professors at the IMS, Grzegorz Dogil, Hans Kamp and Christian Rohrer, as well as to the administrative staff, particularly Sabine Dieterle, Edgar Hoch and Sybille Laderer, who all took care that my environment was both working and comfortable. In addition, the 'Mensa-Runde' including Stefanie Dipper, Arne Fitschen, Martin Forst, Josef van Genabith, Effi Georgala, Jonas Kuhn, Wolfgang Lezius, Tina Säuberlich and Jasmin Saric made sure that I could forget about work for at least one hour per day, enjoying an entertaining meal in the university's cafeteria, possibly outside in the sun, but certainly not with a coffee after lunch. For a change in the daily routine, Inga Frese and Michael Klein joined me to explore the 'Mittagstisch' neighbourhood of the IMS.

Special thanks go to my office-mate Heike Zinsmeister with whom I enjoyed to discussing work matters as well as questions and problems concerning social issues. Furthermore, I am so lucky to have a loving family and many friends inside and outside the computational linguistics world who believe in me even if I stop doing so. My volleyball team provided for the physical and for yet another social aspect of my life. And, last but not least, most thanks go to Matthias who shared both the fun and the burden of my work with me.