

## Arbeitsbereich „Natürlichsprachliche Systeme“ (NatS)

Vogt-Kölln-Str. 30, 22527 Hamburg; Tel.: 040 / 42883-2433, Fax: 040 / 42883-2515  
URL: <http://nats-www.informatik.uni-hamburg.de>

### 1. Zusammenfassende Darstellung

#### Mitglieder der Departmenteinrichtung:

##### *ProfessorInnen:*

Dr. Walther von Hahn, Dr.-Ing. Wolfgang Menzel

##### *AssistenInnen/Wiss. MitarbeiterInnen:*

Dipl.-Inf. Kilian Foth (bis 2/2006), Thomas Kopinski (ab 10/2006), Dipl.-Inf. Michael Daum, Dr. Cristina Vertan, Dipl.-Inf. Nguyen Thinh Le, Dipl.-Inf. Monica Gavrilă, Irina Aleksenko, Eelco Mossel (ab 2/2006), Klaus Dalinghaus (ab 2/2006), Dipl.-Inf. Carmen Gámiz Marcos (ab 10/2006)

Natalia Elita - DAAD Stipendiatin, Solomon Teferra Abate – ehem. DAAD Stipendiat aus Äthiopien, Martha Yifiru Tachbelie - Stipendiatin aus Äthiopien, Tian Gan (MSc.) - Internationales Graduiertenkolleg CINACS, Patrick McCrae (MLitt) - Internationales Graduiertenkolleg CINACS

##### *Technisches und Verwaltungspersonal:*

Karin Jarck (Sekretariat)

##### *Gäste:*

Prof. Dr. Galia Angelova (Sofia, BG) 2.-8.12.2006 (BIS21++ Center of Excellence)  
Prof. Dr. Monica Tataram (Bukarest, RO) 2.-21.7.2006 (Univ. Partnerschaft)  
Dr. Svetla Boytcheva (Sofia, BG) 1.2.2006 – 15.8.2006 (UNESCO-Stipendium)  
Krisztina Babos (Budapest, HU), 16.-22.10.2006 (Stipendium Univ. Budapest)  
Dr. Valia Kordoni (Saarbrücken, DE), 1.-7.7.2006 (Vortragender ELSNET Sommerschule)  
Dr. Pontus Svenson (Stockholm, S) 1.-7.2006 (Vortragender ELSNET Sommerschule)  
Dr. Hans van Halteren (Nijmegen, NL), 1.-7.7.2006 (Vortragender ELSNET Sommerschule)  
Dr. Stephen Oepen (OSLO, N), 8.-15.7.2006 (Vortragender ELSNET Sommerschule)  
Dr. Rainer Stiefelhagen (Karlsruhe, DE), 8.-15.7.2006 (Vortragender ELSNET Sommerschule)  
Dr. Luciano Serafini (Trento, IT), 8.-15.7.2006 (Vortragender ELSNET Sommerschule)  
Dr. Ulrich Schäfer (Saarbrücken, DE), 8.-15.7.2006 (Vortragender ELSNET Sommerschule)  
Steven Krauwer (Utrecht, NL) 7.-9.7.2006 (Vortragender ELSNET Sommerschule)  
Prof. Dr. Nicola Guarino (Trento, IT), 6.-9.3.2006 (Vortragender Ontology-Workshop)  
Werner Behrendt (Salzburg, A) 6.-9.3.2006 (Vortragender Ontology-Workshop)  
Prof. Dr. Susanne Johanna Jekat, Zürcher Hochschule Winterthur seit 10/2006

Ivona Profire (Bukarest, RO), 3.-16.3.2006 und 22.11.-3.12.2006 (Univ. Partnerschaft)  
Nelu Adam Altar (Bukarest, RO) 1.-15.7.2006 (Univ. Partnerschaft)  
David Kolovratnik (Prag, CZ), 1.-15.7.2006 (Univ. Partnerschaft)  
Miroslav Tynowsky (Prag, CZ) 1.-15.7.2006 (Univ. Partnerschaft) und 1.-18.11.2007 (Projekt Lt4eL)  
Ivelina Nikolova (Sofia, BG), 1.-15.7.2006 (Univ. Partnerschaft)  
Radoslav Kirkov (Sofia, BG), 1.-15.7.2006 (Univ. Partnerschaft)  
Mihaela Munteanu (Bukarest, RO) 21.11.-2.12.2006 (Univ. Partnerschaft)

Christina von Bremen (geb. Maack) – Projekt PROLIV/Kooperation mit dem FB 07  
Dr. Kerstin Fischer - Universität Bremen

##### *Ehemalige:*

Kilian Foth

#### Allgemeiner Überblick

Ziel der Forschungsarbeiten im Arbeitsbereich NatS ist der Entwurf und die prototypische Realisierung komplexer sprachverarbeitender Systeme mit dem Schwerpunkt auf schriftlich realisierter Sprache

(language) bis hin zur kommunikativen Bewertung. Vor dem Hintergrund unterschiedlicher Anwendungsbeispiele werden Voraussetzungen und Lösungsansätze für die Einbeziehung der jeweils relevanten sprachlichen Ebenen in den Verarbeitungsprozess untersucht, wobei auf integrative Systemarchitekturen besonderer Wert gelegt wird. Daneben werden Fragen der maschinellen Übersetzung und Bedingungen mehrsprachiger Systeme theoretisch untersucht und prototypisch implementiert

### **Forschungsschwerpunkte**

#### *Architektur integrierter Sprachverarbeitungssysteme*

Angesichts der Vielzahl und Verschiedenheit der Komponenten, die bei der Sprachverarbeitung notwendigerweise miteinander interagieren, haben sich vereinfachende und starre Architektorentwürfe wiederholt als entscheidendes Hindernis auf dem Weg zu anspruchsvollen natürlichsprachlichen Systemen herausgestellt. Besonders extreme Anforderungen stellen hierbei Systeme für die Verarbeitung gesprochener Sprache. Die Forschungsarbeiten zielen darauf ab, über eine interaktive und plausibilitätsgestützte Kopplung der einzelnen Systemkomponenten (Syntax, Semantik, Weltwissen, Worterkennung, Prosodie usw.) die Grundlagen für robuste und zeitadaptive Analyseverfahren zu schaffen und auf diesem Wege auch Möglichkeiten zur Erweiterung auf multimodale Kommunikationsformen vorzubereiten.

#### *Konfrontative Sprachverarbeitung*

Neben den international sehr intensiv betriebenen Forschungen zur Maschinellen Übersetzung bietet die Entwicklung von spezialisierten Arbeitsplätzen für Humanübersetzer einen weiteren Erfolg versprechenden Weg zur Befriedigung des ständig steigenden Übersetzungsbedarfs. Dabei sollen erstmalig Ergebnisse der Computerlinguistik und der wissensbasierten Sprachverarbeitung zusammengeführt werden, um eine Bedienoberfläche zu schaffen, die eine weit gehende strukturelle Verwandtschaft zu den kognitiven Aktionen aufweist, wie sie ein Übersetzer subjektiv für gegeben ansieht. Völlig neuartige, bisher noch nie untersuchte Forschungsfelder ergeben sich durch die Arbeit an Systemen zum automatischen Dolmetschen. Erforderlich dazu ist ein systematisches Studium des menschlichen Vorbilds, die Festlegung von geeigneten und gleichzeitig realistischen Szenarien und schließlich auch die Erarbeitung von fundierten Thesen zur Nutzerakzeptanz.

#### *Robustes Parsing natürlicher Sprache*

Die vielfältigen Anstrengungen zur Steigerung der Robustheit sprachverarbeitender Systeme beziehen sich traditionell stets auf isolierte Aspekte, wie die Fähigkeit zur Analyse unrestringierter Texte oder aber die Behandlung der inhärenten Erkennungsunsicherheit bei der Verarbeitung gesprochener Sprache. Im Gegensatz dazu wird hier ein Ansatz verfolgt, der Robustheit als multidimensionales Phänomen begreift. Auf der Grundlage eines einheitlichen Verarbeitungsmechanismus werden neben den bereits genannten Fragen auch Ansätze zur Verarbeitung unter Zeitdruck, sowie die Fähigkeit zur Diagnose von sprachlichen Fehlern untersucht. Eine wesentliche Rolle spielt hierbei der Abgleich unsicherer und partiell widersprüchlicher Information in einem System mit hoher struktureller Redundanz.

#### *Lehrsysteme für den Fremdsprachenunterricht*

Systeme für den Fremdsprachenunterricht sind sowohl ein wichtiges Anwendungsgebiet als auch ein ideales Testfeld für Verfahren zur robusten Sprachverarbeitung. Im Mittelpunkt stehen dabei Techniken zur Fehlerdiagnose in den Bereichen von Phonetik/Phonologie und Syntax. Auf dieser Grundlage soll dem Schüler ein aktives Üben in möglichst realitätsnahen Kommunikationssituationen ermöglicht werden, wobei er gleichzeitig auf ein qualitativ deutlich verbessertes Feedback zu möglichen Schwächen und besonders empfehlenswerten Übungsformen zurückgreifen kann.

#### *Maschinelle Übersetzung*

Systeme zur maschinellen Übersetzung sind eine sehr geeignete Testumgebung verschiedener Architekturen und Verfahren für die Sprachverarbeitung. Robuste Bearbeitung natürlicher Sprache ist ebenfalls eine wichtige Komponente in MT-Systemen. Traditionelle regelbasierte Systeme sowie neue statistische bzw. beispiel-basierte Verfahren können allein allerdings die Vielfältigkeit der Phänomene in der MÜ nicht lösen. Untersucht werden Hybride Systeme, insbesondere Erweiterungen von beispiel-basierten Verfahren mit linguistischer Information und regelbasierten Komponenten, auch Phänomene in wenig untersuchten Sprachpaaren werden dabei berücksichtigt.

### **Wissenschaftliche Zusammenarbeit**

#### *Industrie und industriennahe Forschungseinrichtungen in Deutschland*

- Harman&Becker, Ulm

*Universitäten und Technische Hochschulen in Deutschland*

- Universität Tübingen

*Kooperationspartner im Ausland*

- Bulgarische Akademie der Wissenschaften, Sofia, Bulgarien
- Universität Sofia, Bulgarien
- Universität Bukarest, Rumänien
- Charles Universität Prag, Tschechien
- Universität Granada, Spanien
- Universität Utrecht, Holland
- Macquarie Universität, Sydney, Australien
- Universität Patras, Griechenland
- Universität Manchester, Großbritannien
- Züricher Winterthur Fachhochschule, Schweiz

**Ausstattung**

Der Arbeitsbereich NATS vereinheitlicht sein heterogenes Rechnernetzwerk in Richtung auf die Plattformen x86/Linux und PPC/Mac OS X. Zentrale Dienste wie Filesharing, WWW etc. werden von Server-PCs zur Verfügung gestellt; die Benutzerverwaltung wird von einem Linux-Rechner und Backups werden direkt vom Rechenzentrum erbracht. Ingesamt stehen dem Arbeitsbereich aus Haushaltsmitteln 21 PCs und 3 Laptops sowie 3 Macintosh-Computer zur Verfügung. Im Rahmen von Drittmittelprojekten gibt es darüber hinaus 1 Laptop. Für rechenintensive Forschungsarbeiten kann ein Linux-Cluster mit 14 Zweiprozessormaschinen genutzt werden.

**Drittmittel**

Projekt:	Erwerb semantischer Templates für beispielbasierte Maschinelle Übersetzung
Geldgeber:	DAAD
Gesamtmittel:	36 Monate 1 Promotionsstipendium
Projekt:	Copa: Constraint-basiertes Parsing und psychologische Plausibilität 2005-2007
Geldgeber:	DFG
Gesamtmittel:	24 PM BAT IIA/2 15 PM BAT IIA 24 Monate 2x40h SHK Sachmittel: € 18.200
Projekt:	Dia Learn: Constraint-basierte Diagnose für interaktive Lernsysteme 2005-2007
Geldgeber:	DFG
Gesamtmittel:	24 PM BAT IIA 24 Monate 1x73h SHK Sachmittel: € 7000
Projekt:	High-level information fusion for speech and language 4/2006 bis 11/2010
Geldgeber:	DFG
Gesamtmittel:	Sachmittel und Reisekosten: € 30.600 Personal: € 222.900
Projekt:	Morphologie-basierte Sprachmodellierung für Amharisch
Geldgeber:	Universität Hamburg
Gesamtmittel:	24 Monate 1 Promotions-Stipendium
Projekt:	Constraint-Parsing und psycholinguistische Plausibilität 11/2005 - 10/2007
Geldgeber:	DFG
Gesamtmittel:	27 Monate BAT IIA 24 Monate 2x40h SHK

Sachmittel: € 18.200

Projekt:	LING-Net – Linguistische Einführung im Netz 10/2004 – 3/2006
Geldgeber:	Multimedia Kontor
Gesamtmittel:	€ 5000
Projekt:	EMS - E-Learning - Einführung in die maschinelle Sprachverarbeitung 8/2004 - 2/2006
Geldgeber:	Multimedia Kontor
Gesamtmittel:	€ 5000
Projekt:	SemWeb – Eine eLearning Einführung im Semantic Web 3/2005 – 3/2006
Geldgeber:	Multimedia Kontor
Gesamtmittel:	€ 5000
Projekt:	PROLIV: Animierte Prozessmodelle komplexerer linguistischer Verfahren 6/2005 - 6/2007
Geldgeber:	Hamburger E-Learning Consortium
Gesamtmittel:	€ 98.500
Projekt:	LT4eL – Language Technology for eLearning 12/2005 – 5/2008
Geldgeber:	EU/IST-FP6
Gesamtmittel:	€ 95.000
Projekt:	I*teach – Intelligent Teacher 10/2005 – 9/2007
Geldgeber:	EU/Leonardo da Vinci
Gesamtmittel:	€ 32.400
Projekt:	BIS21* Center of Excellence Kooperation mit der Bulgarischen Akademie der Wissenschaft 10/2005 – 09/2008
Geldgeber:	EU/IST-FP6[Name des Geldgebers]
Gesamtmittel:	€ 10.400
Projekt:	PREiS „International Master in Speech and Language“ 10/2005 – 12/2007
Geldgeber:	Universität Hamburg
Gesamtmittel:	€ 79.000

## 2. Die Forschungsvorhaben der Departmenteinrichtung

### Etatisierte Projekte

#### 2.1. Wissensgewinnung mit maschinellen Constraint-Grammatiken

Foth, Kilian - Dipl.-Inf. (bis 2/2006)

*Laufzeit des Projektes:*

seit 10/1999

*Projektbeschreibung:*

Die automatische Syntaxanalyse ist mittlerweile auf einem Stand angelangt, in dem durch bloße Ähnlichkeitsmaximierung aus voranalysierten Daten ein relativ hoher Grad an Analysegenauigkeit erreicht werden kann. Die verschiedenen Varianten dieser Verfahren unterscheiden sich nur noch durch Feinheiten der Glättungs- und Ähnlichkeitsparameter und erreichen auf den standardisierten Testkorpora in etwa dieselben Ergebnisse. Jedoch erscheinen diese Ergebnisse auch durch Beschaffung noch größerer Datenmengen nicht mehr wesentlich steigerbar zu sein.

In den DFG-Projekten DAWAI und PaPa wurde ein System entwickelt, das in vieler Hinsicht den genau entgegen gesetzten Ansatz verfolgt. Statt Evidenz aus möglichst großen Baumbanken zu sammeln, muss hier der Grammatikentwickler die Regeln der natürlichen Sprache selbständig und ausdrücklich formulieren und in ihrer Wichtigkeit beurteilen; Baumbanken werden vornehmlich zur Verifikation eingesetzt. Der Aufwand sowohl für die Maschine als auch für den Menschen ist hierbei wesentlich höher, jedoch lassen sich aufgrund des ausdrucksstärkeren Formalismus viele sprachliche Prinzipien in viel adäquaterer Weise formulieren. Mit dem Aufkommen der neuesten Prozessorgenerationen ist es erstmals möglich geworden, diesen Formalismus in großem Maßstab zur Analyse einzusetzen; es hat sich herausgestellt, dass hierdurch tatsächlich eine noch höhere Analysegenauigkeit erreicht werden kann als durch rein statistische Verfahren.

In dieser Arbeit wurde nun das bestehende, im wesentlichen von Hand geschriebene Analysesystem für das Deutsche an einzelnen Stellen gezielt um statistische Information angereichert, um spezifische Probleme zu lösen, die durch Handkodierung auch mit großem Aufwand kaum gelöst werden können, aber bereits sehr erfolgreiche statistische Lösungen besitzen (z.B. Kategoriebestimmung oder PP-Attachment). Dadurch konnte die Analysegenauigkeit des Systems nochmals verbessert werden, so dass nun der genaueste veröffentlichte Parser des Deutschen zur Verfügung steht.

*Schlagwörter:*

Wissensakquisition, partielles Parsing, Dependenzgrammatik, machine learning

*Publikationen aus dem Projekt:*

- Foth, K.; Menzel, W.: Robust parsing: More with less. In Proc. EACL Workshop, Robust Methods in Analysis of Natural Language Data, ROMAND-2006, pages 25-32, Trento, Italy, 2006
- Foth, K.; Menzel, W.: Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; Menzel, W.: The benefit of stochastic PP-attachment to a rule-based parser. In Proc. 21st Int. Conf. on computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; By, T. and Menzel, W.: Guiding a constraint dependency parser with supertags. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; Menzel, W. Ingo Schröder. Robust Parsing with Weighted Constraints. Natural Language Engineering, 2005

## **2.2. ManageLex - Ein Lexikonmanagement-Tool zur Transformation und Mischung von maschinenlesbaren Lexika**

v.Hahn, Walther, Prof. Dr.; Vertan, Cristina, Dr.; Gavrilă, Monica

*Laufzeit des Projektes:*

seit 5/2002

*Projektbeschreibung:*

Das Ziel des Projekts ist die Entwicklung eines Systems für das Management unterschiedlicher Lexika, d.h. Lesen, Transformieren, Mischen und Speichern von Lexika unabhängig von ihrer Kodierung, ihrem Format oder linguistischer Struktur. Die Problematik ist sehr aktuell, einerseits weil die Entwicklung neuer Sprachressourcen (in diesem Fall Lexika) sehr zeitaufwendig und teuer ist, andererseits weil die existierenden Lexika meist keiner der inzwischen existierenden Standardisierungen folgen.

Die Managelex-Architektur besteht aus formalen Modellen und den entsprechenden Bearbeitungstools für:

- Kodierung / Dekodierung der Formate und Speicherstruktur,
- Manipulation und Änderung der linguistischen Struktur, und
- Mischung zweier Lexika.

Das Projekt wird überwiegend durch Diplomarbeiten verfolgt.

*Schlagwörter:*

Lexikon, Sprachressourcen, Lexikonmanagement, Computerlinguistik, Standards

*Publikationen aus dem Projekt:*

- v.Hahn, W. : „Merging Computer-Readable Heterogenous Terminological Material: The ManageLex Tool.In“, Berlin 2006, in Picht, Heribert (Ed.), Modern Approaches to Terminological Theories and Applications. S. 235 - 250
- Gavrilă, M., Vertan, C., v. Hahn, W.: „Managelex - a tool for the management of complex lexical structures“, Koethen, 2006, In GLDV-Journal for Computational Linguistics and Language Technology,

- Band 21, Heft 1, Exchange of Lexical and Terminological Resources, Beiträge des Workshops des GLDV-Arbeitskreises Maschinelle Übersetzung, pages 57-65, ISSN 0175-1336 .
- Gavrila, M., Vertan, C.: „Managelex and the semantic web“, Jeju/South Korea, 2005, In IJCNLP-05, OntoLex-Ontologies and Lexical Resources Workshop Proceedings, S. 22-27
- Gius, E.: Vergleich maschinenlesbarer deutscher Lexika nach linguistischem Inhalt, Wertebereichen und Kodifizierung, 2003, Magisterarbeit, Universität Hamburg,
- Gavrila, M.: „LexMod and LexTool, lexical model, acquisition and editing in MANAGELEX“, Hamburg, 2004., Hamburg Univ., Fachbereich Informatik, Diplomarbeit 2004. R36887

### 2.3. High-level information fusion for speech and language

McCrae, Patrick (MLitt); Gan, Tian (MSc); Menzel, Wolfgang (Prof. Dr.-Ing.)

*Laufzeit des Projektes:*

von 04/2006 bis 11/2010

*Projektbeschreibung:*

As part of the research program of the International Research School CINACS (Cross-modal interaction in natural and artificial cognitive systems) this project sets out to investigate cross-modal facilitatory effects in technical systems for speech and language processing. This kind of system-integration differs from other problems of information fusion by the need to deal with highly abstract representations for which no straightforward mapping to the corresponding sensory input exists. Speech input, for instance, needs to be interpreted in terms of word or phrase meaning before it can be related to the visual experience of the hearer. Despite the enormous gap between the representations involved, the human cognitive system is able to make the complementary contributions from other modalities almost instantaneously available to the speech and language processing faculty, establishing the prerequisites for a synergy which contributes considerably to the rapid and robust behavior of human multi-modal communication and sets it apart from any current technical solutions.

To gain deeper insights into the fundamental requirements for high-level information fusion currently two selected problem areas are studied in two PhD-projects, namely audio-visual speech recognition (Gan, Tian) and the (dynamic) influence of context on sentence processing (McCrae, Patrick).

Audio-visual speech recognition makes use of extra video data, in particular lip-reading information, to improve the performance of a traditional acoustic-only speech recognizer. Based on state-of-the-art speech recognition technology, information fusion between acoustic and visual cues is attempted on the level of phones, where the visual stimulus is described by means of an underspecified phone representation. These two representations are then to be combined by a third component for word recognition. Hidden Markov Models (HMM) will be used to represent the acoustic and visual characteristics of phones and different HMM topologies will be designed to improve the system performance.

During sentence processing cross-modal semantic influence often results in an (early) disambiguation, whenever there are several competing structural interpretations for an utterance, which are equally plausible with respect to lexical preferences and static world knowledge. Contextual information, as obtained from sensory input, e. g. via cross-modal perception, can then provide additional cues. While a wide range of studies have systematically investigated the impact of context and world knowledge upon structural disambiguation in human sentence processing, surprisingly few attempts set out to model the integration of dynamic context and world knowledge in natural language processing applications. The projects will study possibilities for integrating extra-sentential context and world knowledge into the decision process on the most plausible structural interpretation of a sentence, using a parsing system based on weighted constraints.

*Schlagwörter:*

information fusion, multi-modal communication, speech recognition, sentence processing

*Publikationen aus dem Projekt:*

- Menzel, W. System architecture as a problem of information fusion. In Proc. Int. Symposium Natural Language Processing between Linguistic Inquiry and Systems Engineering, pages 74-84, Hamburg, 2002.
- Menzel, W. Konfliktlösung als Grundlage intelligenten Handelns. In Rüdiger Valk, editor, Ordnungsbildung und Erkenntnisstrukturen, pages 87-101. Hamburg University Press, 2006.
- Foth K.; Menzel, W. Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006.

*Finanzierung:*

Projekt:	High-level information fusion for speech and language 4/2006 bis 11/2010
Geldgeber:	DFG
Gesamtmittel:	Sachmittel und Reisekosten: € 30.600 Personal: € 222.900

**2.4. Constraint-Parsing und psychologische Plausibilität**

Menzel, Wolfgang, Prof. Dr.-Ing.; Daum, Michael, Dipl.-Inf.; Dalinghaus, Klaus

*Laufzeit des Projektes*

von 9/2006 bis 2/2008

*Projektbeschreibung*

Als Alternative zu den etablierten probabilistischen Modellen wird im Projekt ein Ansatz auf der Basis von gewichteten Constraints im Hinblick auf sein Potenzial zur Erklärung psycholinguistischer Phänomene bei der Satzanalyse untersucht. Über die Möglichkeiten probabilistischer Modelle hinaus gestattet er es, die verschiedenen Informationsbeiträge zum Sprachverstehen so zu isolieren, dass ihr Zusammenwirken einer experimentellen Untersuchung zugänglich wird. Zudem bietet er einen vielversprechenden Erklärungsansatz für das Zusammenspiel von Sprachsystem und Erfahrungswissen bei der Beurteilung gradueller Abweichungen von der sprachlichen Norm.

Die Arbeiten gehen von einem in Vorgängerprojekten entwickelten System zum Parsing natürlichsprachlicher Sätze aus, mit dem auf der Basis einer Constraint-Grammatik für das Deutsche und unter Verwendung von probabilistischen Prädiktorkomponenten eine sehr hohe Analysegenauigkeit für eine Vielzahl unterschiedlicher Textsorten nachgewiesen werden konnte.

*Finanzierung*

Projekt:	Copa: Constraint-basiertes Parsing und psychologische Plausibilität 2005-2007
Geldgeber:	DFG
Gesamtmittel:	24 PM BAT IIa/2 15 PM BAT IIa 24 Monate 2x40h SHK Sachmittel: € 18.200

*Publikationen aus dem Projekt:*

- Daum, M.: Dynamic Dependency Parsing. In Proc. ACL 2004 Workshop on Incremental Parsing, Barcelona, Spain, 2004
- Foth, K.; Menzel, W.: Robust parsing: More with less. In Proc. EACL Workshop, Robust Methods in Analysis of Natural Language Data, ROMAND-2006, pages 25-32, Trento, Italy, 2006
- Foth, K.; Menzel, W.: Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; Menzel, W.: The benefit of stochastic PP-attachment to a rule-based parser. In Proc. 21st Int. Conf. on computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; By, T. and Menzel, W.: Guiding a constraint dependency parser with supertags. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Menzel, W.: Konfliktlösung als Grundlage intelligenten Handelns. In Rüdiger Valk, editor, Ordnungsbildung und Erkenntnisstrukturen, pages 87-101. Hamburg University Press, 2006

**2.5. Constraint-basierte Diagnose für interaktive Lehrsysteme**

Menzel, Wolfgang, Prof. Dr.-Ing.; Nguyen Thinh Le, Dipl.-Inf.

*Laufzeit des Projektes*

von 1/2006 bis 12/2007

*Projektbeschreibung*

Für Aufgabenstellungen aus der Softwareentwicklung bei denen nicht nur die Reproduktion von Wissen bzw. dessen Anwendung für analytische Tätigkeiten, sondern vor allem produktive Fertigkeiten im Mittelpunkt der Ausbildung stehen, soll eine computergestützte Lernumgebung entwickelt werden, die es

dem Lernenden ermöglicht, kreativ an einer Problemlösung zu arbeiten, wobei sein Handlungsspielraum möglichst wenig durch vorgegebene Musterlösungen, bzw. vorgefertigte Auswahlantworten eingeschränkt wird. Um eine solche kreative Tätigkeit zu unterstützen, muss das Lernsystem über präzise und tiefreichende diagnostische Fertigkeiten verfügen, um dem Schüler zu helfen, die Mängel seines derzeitigen Lösungsversuchs zu erkennen und sie zu überwinden, ohne dabei die Lösung selbst vorwegzunehmen.

Für analoge Aufgaben aus dem Bereich des Sprachunterrichts hat sich hierfür ein Diagnoseansatz auf der Basis von Constraints als Erfolg versprechend herausgestellt. Constraints beschreiben Bedingungen für eine akzeptable Schülerlösung in einer Weise, die es gestattet, aus Constraintverletzungen Rückschlüsse auf Mängel der Lösung und die sie verursachenden Fehlkonzeptionen zu ziehen. Über geeignete Interventionen kann der Lernende dann schrittweise an eine verbesserte Lösung herangeführt werden. Als besonderer Vorzug des constraint-basierten Ansatzes erweist sich dabei, dass er verhältnismäßig robust gegenüber Lücken im Wissensbestand des Lernsystems ist und auch partielle Schülerlösungen problemlos bewerten kann.

Im Projekt wird angestrebt, Techniken aus dem Bereich der Constraint-basierten Diagnose natürlichsprachlicher Äußerungen auf die Diagnose von Programmierfehlern zu übertragen. In diesem Zusammenhang soll auch geklärt werden, welche Voraussetzungen für eine erfolgreiche Anwendung des constraint-basierten Paradigmas in interaktiven Lernsystemen gegeben sein müssen und wie die Arbeitsteilung zwischen den grundlegenden Komponenten eines solchen Systems gestaltet werden muss.

#### *Finanzierung*

Projekt:	Dia Learn: Constraint-basierte Diagnose für interaktive Lernsysteme 2005-2007
Geldgeber:	DFG
Gesamtmittel:	24 PM BAT IIa 24 Monate 1x73h SHK Sachmittel: € 7000

#### *Publikationen aus dem Projekt*

- Le, N.-T.: Using prolog design patterns to support constraint-based error diagnosis in logic programming. In Kevin Ashley Vincent Aleven, Niels Pinkwart and Collin Lynch, editors, Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains, the 8th Conference on ITS, pages 38 - 46, 2006
- Le, N.-T.: A constraint-based assessment approach for free-form design of class diagrams using uml. In Kevin Ashley Vincent Aleven, Niels Pinkwart and Collin Lynch, editors, In Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains, the 8th Conference on ITS, 2006, pages 11 - 19, 2006
- Le, N.-T. and Menzel, W.: Problem solving process oriented diagnosis in logic programming. In to be published in Proceedings of the 14th International Conference on Computers in Education, volume 134. IOS Press, 2006
- Le, N.-T. and Menzel, W.: Constraint-based problem generation for a self-assessment system. In ICCE-2006 Workshop Proceedings of Problem-Authoring, -Generation and -Posing in a Computer-Based Learning Environment, pages 33 - 40, 2006
- Menzel, W.: Constraint-based modelling and ambiguity. Int. Journal on Artificial Intelligence in Education, 16:29-63, 2006
- Menzel, W.: Detecting mistakes or finding misconceptions? Diagnosing morpho-syntactic errors in language learning. In Galia Angelova, Kiril Simov, and Milena Slavcheva, editors, Readings in Multilinguality. Incoma Ltd., Shoumen, Bulgaria, 2006

## **2.6. LingNet - Linguistik-Einführung im Netz**

v. Hahn, Walther, Prof. Dr.

#### *Laufzeit des Projektes:*

von 10/2004 bis 2/2007

#### *Projektbeschreibung*

Das Seminar „Einführung in das Studium der Linguistik“ im Institut für Germanistik I der Universität Hamburg wurde mehrfach erfolgreich mit ca. 130 Folien gehalten. Diese waren /sind im Netz als Mitschreib-Vorlage vorhanden. Die Materialien werden im Rahmen eines Fortgeschrittenen-Seminars vom Seminarleiter und den Teilnehmern mit weiteren wissenschaftlichen Texten, Erklärungen, Erläuterungen, Beispielen, Tests, einem Definitionsregister und Index angereichert in einer vernetzten Version internetfähig gemacht. Ein Nebeneffekt ist, dass geisteswissenschaftliche Studenten Grundbegriffe von Web-Publishing lernen.

Das Material wurde zuerst im Wintersemester 04/05 als begleitende Maßnahme für das Seminar "Einführung in das Studium der Linguistik". Die Vorbereitung auf die Abschlussprüfung der Einführung wird durch eine umfangreiche Webdarstellung verbessert  
Zusätzlich wird das Material in der Informatik zum Erwerb der nötigen linguistischen Kenntnisse für die Sprachverarbeitung genutzt.

*Schlagwörter:*

E-learning, Linguistik, Sprachverarbeitung

*Finanzierung:*

Geldgeber:	Multimedia Kontor Hamburg
Laufzeit der Förderung:	10/2004 bis 2/2007
Sachmittel:	-
Personalmittel:	€ 5.000

## **2.7. SemWeb - Technologien für das Semantic Web: eine eLearning Einführung**

Vertan, Cristina, Dr.

*Laufzeit des Projektes:*

von 3/2005 bis 9/2006

*Projektbeschreibung:*  
Semantic Web ist eine wichtige Technologie, die die Funktionalität des aktuellen Webs ergänzt. Semantic Web wird das Web der Zukunft sein, deswegen es ist extrem wichtig, dass die nötige Technologien, Methoden und Sprachen zu Studenten in verschiedene Fächern übermittelt werden. Leider, obwohl bereits viel Literatur zu dem Thema geschrieben wurde, ist diese total unstrukturiert. Besonders für Einsteiger im Gebiet (wie es die Studenten sind) ist es extrem kompliziert, die wichtige Information zu extrahieren und zu interpretieren. Ein zusätzlicher Nachteil ist die fast nicht existierende Einführungsliteratur in deutscher Sprache.

Das Projekt entwirft einen On-line Pool von Materialien und Demos über Semantic Web Technologien. Die Ziele des Projekts sind:

- Realisierung einer einheitlichen Version eines eLearning Moduls über Semantic Web, vollständig in deutscher Sprache
- Erhöhung der effektiven „Hands-on-Zeit“ für Studenten.
- Dauernder Zugriff während der Referat- oder Implementierungsphase zur gesamten Information: Das Material könnte auch von Studenten der Linguistik, die Veranstaltungen wie Computerphilologie besuchen, benutzt werden.

*Schlagwörter:*

E-learning, Semantic Web, Internetanwendungen, Ontologien

*Finanzierung:*

Geldgeber:	Multimedia Kontor Hamburg
Laufzeit der Förderung:	3/2005 bis 9/2006
Sachmittel:	-
Personalmittel:	€ 5.000

## **2.8. EMS: E-learning-Einführung in die maschinelle Sprachverarbeitung**

Vertan, Cristina, Dr.

*Laufzeit des Projektes:*

von 10/2004 bis 5/2006

*Projektbeschreibung:*

Innerhalb des Grundstudiums werden verschiedene Proseminare und Praktika über verschiedene Aspekte der maschinellen Sprachverarbeitung angeboten.

Besonders für Praktika ist eine Einführungsphase (3-4 Sitzungen) obligatorisch, um Grundkenntnisse zum Thema des Praktikums zu vermitteln. Die studentische Evaluation von Praktika und Proseminaren hat gezeigt, dass die Studenten an der Problematik sehr interessiert sind, die Einführungsphase aber zu viel Implementierungszeit wegnimmt. Hierfür wird nun ein Online Pool mit Materialien und Demos über maschinelle Sprachverarbeitung zur Verfügung gestellt: häufig verwendeten Architekturen und Verfahren,

Anwendungen wie maschinelle Übersetzung, natürlichsprachliche Schnittstelle, Informations-Extraktion, Computer-Assisted Language Learning.

*Schlagwörter:*

E-learning, natürlichsprachliche Systeme, Anwendungen der maschinellen Sprachverarbeitung

*Finanzierung:*

Geldgeber:	Multimedia Kontor Hamburg
Laufzeit der Förderung:	10/2004 bis 5/2006
Sachmittel:	-
Personalmittel:	€ 5.000

## 2.9. Erwerb semantischer Templates für Beispielbasierte Maschinelle Übersetzung

Natalia Elita; v. Hahn, Walther, Prof. Dr.; Vertan, Cristina, Dr.

*Laufzeit des Projektes:*

von 10/2005 bis 10/2008

*Projektbeschreibung:*

Das Ziel des gegenwärtigen Forschungsvorhabens ist es, dynamisches automatisches Lernen von semantischen Mustern als Unterstützung für die maschinelle Übersetzung zu erforschen. Das schließt zwei wichtige und zurzeit nicht gelöste Probleme in der beispiel-basierten maschinellen Übersetzung ein:

1. Berücksichtigung semantischer Information in den Übersetzungsmustern. Die zurzeit existierenden Verfahren enthalten lediglich morphologische oder syntaktische Regeln, was offensichtlich die Übersetzungsqualität beeinträchtigt.
2. Das dynamische Lernen von Mustern, d.h., die Möglichkeit des Benutzers, die erzeugten Übersetzungen (und einige erzeugte Parameter) für weitere Übersetzungen wieder zu verwenden. Als Konsequenz wird sich die Größe der Datenbank mit Beispiel-Übersetzungen reduzieren lassen.

Das Forschungsprojekt wird sich im Gegensatz zu dem gros der internationalen Forschung nicht auf Englisch, sondern auf Wirtschaftstexte in zwei hochgradig flektierten Sprachen, Rumänisch und Deutsch, konzentrieren.

*Schlagwörter:*

Maschinelle Übersetzung, semantische patterns, translation memories, maschinelles Lernen

*Publikationen aus dem Projekt*

- Elita N., Gavrilă M.: "Enhancing translation memories with semantic knowledge." Budapest, Hungary, 2006, Proceedings of the First Central European Student Conference in Linguistics .
- Gavrilă M., Elita N.: "RoGER - un corpus paralel aliniat" (Romanian), Iasi, 2006, Workshop proceedings "Romanian Linguistic Resources and Tools for Natural Language Processing", ed. University A.I. Cuza, Iasi.

*Finanzierung:*

Geldgeber:	DAAD
Laufzeit der Förderung:	10/2005 bis 10/008
Sachmittel:	-
Personalmittel:	€ 36.000

## 2.10. Rekombinationsstrategien in der beispielbasierten maschinellen Übersetzung

Gavrilă, Monica; v. Hahn, Walther, Prof. Dr.; Vertan, Cristina, Dr.

*Laufzeit des Projektes:*

seit 10/2005

*Projektbeschreibung :*

Das Projekt untersucht Rekombinationsstrategien für Beispiel-basierte maschinelle Übersetzung. Dieses Thema wird als "der schwierigste Schritt im EBMT Prozess" betrachtet - (H. Somers, "An Overview of EBMT" 2003) und es ist ein "Bereich, der wenig Aufmerksamkeit empfangen hat" - (K. McTait "Translation Pattern Extraction and Recombination for Example-Based Machine Translation, 2001).

Zurzeit verwenden EBMT-Systeme als Zielsprache in der Übersetzung Englisch dadurch treten viele Probleme in der Morphologie bei der Rekombinationsphase nicht auf. Das Projekt konzentriert sich exemplarisch auf zwei stark flektierte Sprachen - Rumänisch und Deutsch.

Obwohl im Projekt die zwei erwähnten Sprachen untersucht werden, ist das Ziel des Forschungsvorhabens die Entwicklung von Sprachunabhängigen Algorithmen.

*Schlagwörter:*

Beispielbasierte maschinelle Übersetzung, Rekombination, computergestützte Morphologie

*Publikationen:*

Elita N., Gavrilă M.: "Enhancing translation memories with semantic knowledge." Budapest, Hungary, 2006, Proceedings of the First Central European Student Conference in Linguistics .

Gavrilă M., Elita N.: "RoGER - un corpus paralel aliniat" (Romanian), Iasi, 2006, Workshop proceedings "Romanian Linguistic Resources and Tools for Natural Language Processing", ed. University Al.I. Cuza, Iasi.

## 2.11. PROLIV - Animierte Prozessmodelle komplexer Linguistischer Verfahren

v. Hahn, Walther, Prof. Dr.; Vertan, Cristina, Dr.; Gavrilă, Monica; Redder, Angelika, Prof. Dr. (FB. 07); von Bremen, Christina (geb. Maaack) (FB. 07) Laufzeit des Projektes:

von 6/2005 bis 6/2007 Projektbeschreibung:

Das Projekt PROLIV soll durch effiziente lernorientierte Information bei Geisteswissenschaftlern die Bereitschaft für die Anwendung komplexer formaler Methoden und bei Informatikern die Einsicht in geisteswissenschaftliche Sachverhalte und Interpretationen erhöhen. Technisches Hauptziel des Projekts PROLIV ist die Entwicklung von interaktiven multimedialen Lernobjekten hauptsächlich für die Vertiefung von Grundstudiumsveranstaltungen in Linguistik und Computerphilologie (FB07), sowie der Sprachverarbeitung (FB18) und die Integration solcher Lernobjekte und entsprechenden eLearning-Contents in eine homogene Oberfläche innerhalb von "WebCT".

Das plattformunabhängige Projekt wird auf solche Methoden fokussieren, die einerseits inhaltlich auf sehr aktuellen linguistischen, mathematischen oder informatischen Theorien basieren und im Grundstudium allein durch textuelle Präsentation nicht leicht zu verstehen sind, deren Kenntnis und Verwendung aber schon in Grundstudium sinnvoll ist.

Daher werden in PROLIV Demonstratoren für folgende Methoden implementiert:

- Latent Semantic Analysis (LSA)
- Lexikonsprachen und -formate
- Fragenmodellierung

*Schlagwörter:*

E-learning, natürlichsprachliche Systeme, linguistische Verfahren, Visualisierung

*Publikationen aus dem Projekt*

Gavrilă, M., Vertan, C.: „Visualization of complex linguistic theories“. Bangkok/Thailand, 2006, Proceedings of the International Conference on Digital Media and Learning Conference Proceedings, pages 158-163

Gavrilă, M., Vertan, C., v. Hahn, W.: „Managelex- a tool for the management of complex lexical structures. In GLDV-Journal for Computational Linguistics and Language Technology, Band 21, Heft 1, Exchange of Lexical and Terminological Resources, Beiträge des Workshops des GLDV-Arbeitskreises Maschinelle Übersetzung, Koethen, Juni 2005, pages 57-65, ISSN 0175-1336, 2006

*Finanzierung:*

Geldgeber:	Multimedia Kontor Hamburg
Laufzeit der Förderung:	6/2005 bis 6/2007
Sachmittel:	€ 63.469
Personalmittel:	€ 2.133

## 2.12. Internationaler Masterstudiengang „Computerlinguistik und Sprachverarbeitung“

v. Hahn, Walther, Prof. Dr.; Menzel, Wolfgang, Prof. Dr.; Vertan, Cristina, Dr. Laufzeit des Projektes: von 9/2005 bis 12/2007 Projektbeschreibung:

Der Masterstudiengang wird für insgesamt 4 Semester d.h. einen Umfang von 120 ECTS Punkten konzipiert, wobei das Studium an mindestens zwei Partneruniversitäten durchgeführt wird und an jeder Universität mindestens 30% der ECTS Credits zu erwerben sind. Voraussetzung für die Verleihung des Titels: **Msc. in**

**Computational Linguistics and Speech Processing** ist die Anfertigung einer Masterarbeit in einem Schwerpunkt des Masterprogramms. Die Masterarbeit sollte gemeinsam von Hochschullehrern verschiedener Partneruniversitäten betreut werden. Die Struktur des Masters folgt den Richtlinien des EU-Masters in Language and Speech (<http://www.cstr.ed.ac.uk/emasters>). Die verschiedenen Themengebiete werden durch Lehrveranstaltungen ggf. auch an den Partneruniversitäten abgedeckt. Die angegebenen „Credit Points“ entsprechen den minimalen Leistungen, die Studierende für ein Themengebiet erbringen müssen, um dieses erfolgreich nachzuweisen. Der Vorteil dieses Systems ist, dass jede Partneruniversität sich auf bereits existierende Lehrveranstaltungen stützen kann, d.h., man braucht in der Regel keine neuen Lehrkräfte. Entsprechend haben wir in Tabelle 1 die Verteilung des aktuellen Lehrangebots an den Partneruniversitäten aufgeführt. Der Masterstudiengang bietet dabei zwei Schwerpunkte an: Wissensrepräsentation und Sprachverstehen und Sprachsignalverarbeitung. Die Studierenden müssen Vertiefungsveranstaltungen in einem Schwerpunkt nachweisen. Die Masterarbeit muss im ausgewählten Schwerpunkt liegen.

*Schlagwörter:*

Sprachverarbeitung, Masterstudiengang, doppeltes Diplom

*Finanzierung:*

Geldgeber:	Universität Hamburg, Abt. Internationales (PREiS Aktion)
Laufzeit der Förderung:	9/2005 bis 12/2007
Sachmittel:	€ 15.600
Personalmittel:	€ 62.400

**2.13. LT4eL - Language Technology for eLearning**

Vertan, Cristina Dr.; v.Hahn, Walther, Prof. Dr.; Mossel, Eelco

*Laufzeit des Projektes:*

von 12/2005 bis 06/2008

*Projektbeschreibung*

Das Projekt entwickelt Technologien für die Verbesserung von eLearning Systemen. Bisher wurden LO's (Learning Objects) nur mit Standard Annotationen versehen. ImLT4eL sollen linguistische Annotationen sowie die Anwendung einer Ontologie und eines Crosslingual Retrieval Prozesses die Suche nach LO's in einem eLearning System deutlich verbessern. Das Projekt wird von der Universität Utrecht koordiniert und hat 12 Institutionen aus 7 Ländern als Teilnehmer. Die Universität Hamburg ist für den Bereich Multilingualität verantwortlich, d.h. Entwicklung einer cross-lingual retrieval engine und die Abbildung von multilingualen Lexika auf der Ontologie.

*Schlagwörter:*

eLearning, Crosslingual Retrieval, Sprachverarbeitung

*Publikationen aus dem Projekt:*

Monachesi, P.; Cristea, D.; Evans, D.; Killing, A.; Lemnitzer, L.; Simov, K.; Vertan, C.: „Integrating Language Technology and Semantic Web techniques in eLearning.“, Villach/Österreich, 2006, in Proceedings of ICL 2006  
 Monachesi, P., Lemnitzer, L., Simov, K., Cristea, D., Killing, A., Evans, D., Vertan, C.: „LT4eL - Language Technology for eLearning.“, Varna/Bulgarien, 2006, Proceedings of NLP4ME 2006

*Finanzierung:*

Geldgeber:	EU- IST/FP6
Laufzeit der Förderung:	12/2005 bis 6/2008
Sachmittel:	€ 12.376
Personalmittel:	€ 66.980

**2.14. I\*Teach - Innovative Teacher**

v. Hahn, Walther Prof. Dr.; Vertan, Cristina Dr.; Debacher, Uwe; Breier, Norbert Prof. Dr. (Erziehungswissenschaft)

*Laufzeit des Projektes:*

von 9/2005 - 9/2007

*Projektbeschreibung*

Das Projekt I\*Teach hat als Ziel die Sammlung bzw. Entwicklung von neuartigen Methoden im Unterricht von "enhanced ICT" Skills. Ein Handbuch sowie ein on-line repository mit allen diesen Methoden wird entwickelt. Lehrfortbildungseinheiten für zirka 20 Lehrer werden regelmäßig organisiert. In Hamburg wird eine „Virtual-Training Center“ implementiert und als Angebot für den Austausch von Lehrinhalte für die Lehrer zur Verfügung gestellt.

*Schlagwörter:*

Informatik für Lehramtstudierende, Informatik in der Schule, eLearning,

*Finanzierung:*

Geldgeber:	EU- Leonardo da Vinci
Laufzeit der Förderung:	9/2005 bis 9/2007
Sachmittel:	€ 16.165
Personalmittel:	€ 27.150

**2.15. Morphology-based language modeling for Amharic**

Yifiru Tachbelie, Martha

*Laufzeit des Projektes:*

4/2006 bis 3/2007

*Projektbeschreibung:*

Language models are fundamental to many natural language applications such as speech recognition, statistical machine translation, spelling correction, etc. A language model is a probability distribution over strings  $P(s)$  that attempts to reflect the frequency with which each string  $s$  occurs as a sentence in natural language text. Since it is impossible to calculate conditional probabilities for all word sequence of arbitrary length in a given language, N-gram language models are generally used. Even, with N-gram models, it is not possible that all word sequences can be found in the training data (data sparseness problem). In particular for morphologically rich languages, there are even individual words that might not be encountered in the training data irrespective of how large it is (Out of Vocabulary problem). In such cases, a promising alternative is to abandon the word as a modeling unit and use sub-word units for the purpose of language modeling.

The goal of this research project is the development of a morpheme-based language model for Amharic, a major language spoken in Ethiopia. Amharic belongs to the semitic language family and is morphologically rich. It exhibits both non-concatenative (root-pattern) and concatenative morphological features. Stems can be derived from roots (radicals) by intercalating patterns (vocalic elements). Moreover, affixes are used to create inflectional and derivational word forms.

*Schlagwörter:*

Language modeling, morpheme-based language modeling, sub-word based language modeling, Amharic

*Finanzierung:*

Geldgeber:	Universität Hamburg - Nachwuchsförderung
Laufzeit der Förderung:	4/2006 bis 3/2007
Mittel:	€ 11.700

**3. Publikationen und weitere Leistungen****Wissenschaftliche Publikationen im Berichtszeitraum**

- Elita N., Gavrilu M.: "Enhancing translation memories with semantic knowledge." Budapest, Hungary, 2006, Proceedings of the First Central European Student Conference in Linguistics
- Foth, K.; Menzel, W.: Robust parsing: More with less. In Proc. EACL Workshop, Robust Methods in Analysis of Natural Language Data, ROMAND-2006, pages 25-32, Trento, Italy, 2006
- Foth, K.; Menzel, W.: Hybrid parsing: Using probabilistic models as predictors for a symbolic parser. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; Menzel, W.: The benefit of stochastic PP-attachment to a rule-based parser. In Proc. 21st Int. Conf. on computational Linguistics, Coling-ACL-2006, Sydney, 2006
- Foth, K.; By, T. and Menzel, W.: Guiding a constraint dependency parser with supertags. In Proc. 21st Int. Conf. on Computational Linguistics, Coling-ACL-2006, Sydney, 2006

- Gavrila M., Elita N.: "RoGER - un corpus paralel aliniat" (Romanian), Iasi, 2006, Workshop proceedings "Romanian Linguistic Resources and Tools for Natural Language Processing", ed. University A.I. Cuza, Iasi
- Gavrila, M., Vertan, C.: „Visualization of complex linguistic theories“. Bangkok/Thailand, 2006, Proceedings of the International Conference on Digital Media and Learning Conference Proceedings, pages 158-163
- Gavrila, M., Vertan, C., v. Hahn, W.: „Managelex- a tool for the management of complex lexical structures. In GLDV-Journal for Computational Linguistics and Language Technology, Band 21, Heft 1, Exchange of Lexical and Terminological Resources, Beiträge des Workshops des GLDV-Arbeitskreises Machinelle Übersetzung, Koethen, Juni 2005, pages 57-65, ISSN 0175-1336, 2006
- v.Hahn, W. : „Merging Computer-Readable Heterogenous Terminological Material: The ManageLex Tool.In“.: Berlin 2006, in Picht, Heribert (Ed.), Modern Approaches to Terminological Theories and Applications. S. 235 – 250
- v. Hahn, W. and G. Angelova: Conceptual Foundations of Terminological Collections in the Internet Age. In: Slavcheva, M., G. Angelova and K. Simov (Eds.), Readings in Multilinguality, Selected papers. Shoumen, Bulgaria 2006, pp. 49-56
- Le, N.-T.: Using prolog design patterns to support constraint-based error diagnosis in logic programming. In Kevin Ashley Vincent Aleven, Niels Pinkwart and Collin Lynch, editors, Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains, the 8th Conference on ITS, pages 38 - 46, 2006
- Le, N.-T.: A constraint-based assessment approach for free-form design of class diagrams using uml. In Kevin Ashley Vincent Aleven, Niels Pinkwart and Collin Lynch, editors, In Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains, the 8th Conference on ITS, 2006, pages 11 - 19, 2006
- Le, N.-T. and Menzel, W.: Problem solving process oriented diagnosis in logic programming. In to be published in Proceedings of the 14th International Conference on Computers in Education, volume 134. IOS Press, 2006
- Le, N.-T. and Menzel, W.: Constraint-based problem generation for a self-assessment system. In ICCE-2006 Workshop Proceedings of Problem-Authoring, -Generation and -Posing in a Computer-Based Learning Environment, pages 33 - 40, 2006
- Menzel, W.: Constraint-based modelling and ambiguity. Int. Journal on Artificial Intelligence in Education, 16:29-63, 2006
- Menzel, W.: Detecting mistakes or finding misconceptions? Diagnosing morpho-syntactic errors in language learning. In Galia Angelova, Kiril Simov, and Milena Slavcheva, editors, Readings in Multilinguality. Incoma Ltd., Shoumen, Bulgaria, 2006
- Menzel, W.: Konfliktlösung als Grundlage intelligenten Handelns. In Rüdiger Valk, editor, Ordnungsbildung und Erkenntnisstrukturen, pages 87-101. Hamburg University Press, 2006
- Menzel, W., Pallotta, V. (Hrsg.): Proc. 4th Int. Workshop on Robust Methods in Analysis of Natural Language Data, ROMAND-2006, Trento, 2006
- Monachesi, P.; Cristea, D.; Evans, D.; Killing, A.; Lemnitzer, L.; Simov, K.; Vertan, C.: „Integrating Language Technology and Semantic Web techniques in eLearning.“, Villach/Österreich, 2006, in Proceedings of ICL 2006
- Monachesi, P., Lemnitzer, L., Simov, K., Cristea, D., Killing, A., Evans, D., Vertan, C.: „LTeL - Language Technology for eLearning.“, Varna/Bulgarien, 2006 , Proceedings of NLP4ME 2006
- Vertan, C. and von Hahn, Walther (eds.): Proceedings of the First Workshop on Ontology based Modelling in Humanities, Hamburg, April 2006, Bericht 264

#### **Wichtige Publikationen aus den vergangenen Jahren 2003-2005**

- Bekele, R. and Menzel, W.: A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students. In Proc. Int. Conf. Artificial Intelligence and Applications, AIA-2005, Vienna, Austria, 2005
- v.Hahn, Walther: Knowledge Representation in Machine Translation. In: Dam, Helle .V., Engberg, Jan and Gerzymisch-Arbogast, Heidrun (Hrsg.), Knowlede Systems and Translation. Berlin 2005 S. 61 - 81
- v.Hahn, W.; Hutchins J. und Vertan, C. (Eds.): Proceedings of the International RANLP 2005 Workshop „Modern Approaches in Translation Technologies. Borovets, Bulgaria 2005
- v.Hahn, W. and Vertan, C.: Computerphilologie in Hamburg: Thematische Module einer computerphilologischen Grundvorlesung, in Jahrbuch für Computerphilologie, v. Georg Braungart, Peter Gendolla und Fotis Jannidis (Eds.), <http://computerphilologie.uni-muenchen.de/jg04/hamburg.html>
- Gavrila, M. and Vertan, C.: Managelex and the semantic web. In IJCNLP-05, OntoLex-Ontologies and lexical Resources Workshop Proceedings, Jeju, South Korea, October 2005, S. 22-27
- Fischer, K. (2005): Discourse Conditions for Spatial Perspective Taking. Proceedings of the Workshop of Spatial Language and Discourse, Hanse Wissenschaftskolleg, October, 23-25th, 2005

- Fischer, K. & Wilde, R. (2005): Methoden zur Analyse interaktiver Bedeutungskonstitution. In: Ueckmann, Natascha (Hrg.): Von der Wirklichkeit zur Wissenschaft: Aktuelle Forschungsmethoden in den Sprach-, Literatur- und Kulturwissenschaften. Hamburg: LIT
- Fischer, K. (2005): Rezension von Weigand, Edda (Hrg.): Emotion in Dialogic Interaction. Advances in the Complex. Amsterdam/Philadelphia: Benjamins, 2004. LinguistList 16.924
- Foth, K.; Menzel, W. and Schröder, I.: Robust parsing with weighted constraints. Natural Language Engineering, 11(1):1-25, 2005
- Foth, K., Daum, M. and Menzel, W.: Parsing unrestricted German text with defeasible constraints. In H. Christiansen, P. R. Skadhauge, and J. Villadsen, editors, Constraint Solving and Language Processing, volume 3438 of Lecture Notes in Artificial Intelligence, pages 88-101. Springer-Verlag, Berlin, 2005
- Le, N.-T. and Menzel, W.: Constraint-based error diagnosis in logic programming. In Proc. 13th International Conference on Computers in Education 2005, Singapore, 2005 -1111
- Teferra Abate, S.; Menzel, W. and Bairu Tafira. An Amharic speech corpus for large vocabulary continuous speech recognition. In Proc. 9th Europ. Conf. on Speech Communication and Technology, Interspeech-2005, Lisbon, Portugal, 2005
- Vertan, C.; v. Hahn, W. and Gavrilă, M.: Designing a parole/simple german-english-romanian lexicon. In Language and Speech Infrastructure for Information Access in the Balkan Countries Workshop Proceedings - RANLP 2005, Borovets, Bulgaria, September 2005. S. 82-87
- Vertan, C.; v. Hahn, W. and Kubon, V. (Eds.). Proceedings of the International MT Summit X 2005, Phuket, Thailand, 2005
- Vertan, C.; v. Hahn, W.: Challenges for the Multilingual Semantic Web, in Proceedings of the International MT Summit X 2005, Phuket, Thailand, S. 5-9
- Vertan, C.: Crosslingual Retrieval in Semantic Web, in Proceedings of the International MT Summit X 2005, Phuket, Thailand, S. 2-4
- Daum, M.: Dynamic Dependency Parsing. In Proc. ACL 2004 Workshop on Incremental Parsing, Barcelona, Spain, 2004
- Fischer, K.: Addressee-oriented Blending in Spatial Object Localisation Tasks. In Proceedings of Spatial Cognition '04, Frauenchiemsee, Oct. 2004, 2004
- Foth, K.: Writing Weighted Constraints for Large Dependency Grammars. In Recent Advances in Dependency Grammar, Workshop COLING 2004, 2004.
- Foth, K.: Daum, M. and Menzel, W.: Interactive grammar development with WCDG. In Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics, Interactive Posters and Demonstrations, Barcelona, Spain, 2004.
- Knoblauch, D.: Data Driven Number-of-States Selection in HMM Topologies. In Proc. 8th Int. Conf. of Spoken Language Processing, ICSLP, Jeju Island, Korea, 2004. Seiten 825 - 829
- Menzel, W.: Errors, Intentions, and Explanations: Feedback Generation for Language Tutoring Systems. In Proceedings International Conference InSTIL/ICALL-2004, pages 75-82, Venice, Italy, 2004.
- Vertan, C.: Querying Multilingual Semantic Web in Natural Language, in Proceedings of W3C International workshop "Semantic Web Applications and Perspectives (SWAP)", Ancona Italy, pg. 98-100.
- Vertan, C.: Introduction to Machine Translation, Course Material of the 16th European Summer School in Logic Language and Information, 9-20 August, 2004, Nancy, France, Edited by FoLLI (26 pages).
- Menzel, W. und Vertan, C. (Hrsg.): Natural Language Processing between Linguistic Inquiry and System Engineering, Ed. Universitatii "Al. I. Cuza", Iasi, Rumänien, 2003, 95 S.
- Menzel, W.: Dependenzgrammatik und Semantische Netze. In V. Agel, L. M. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer, and H. Lobin, editors, Dependenz und Valenz - Ein Internationales Handbuch der zeitgenössischen Forschung. Walter de Gruyter, Berlin, New York, 2003.
- Vertan, C.: Understanding texts across languages without translation, In: Proceedings of LSP'03 (Language for Special purposes conference), Guildford, August 2003

### Begutachtungen und abgeschlossene Betreuungen am Department

#### Dissertationen

DoktorandIn	GutachterIn	Thema	Datum
Kilian Foth	W. Menzel; J. Nivre; Chr. Habel	Hybrid Methods of Natural Language Analysis	Okt. 2006
Dirk Knoblauch	W. Menzel; G. Fink; J. Zhang	Topologieoptimierung phonetischer Hidden-Markov-Modelle mittels Erkennungssimulation	Nov. 2006

*Diplomarbeiten*

<b>DiplomandIn</b>	<b>BetreuerIn</b>	<b>Thema</b>	<b>Datum</b>
Karsten Kowalewski	W. Menzel	Referenzauflösung für Personalpronomen	Jun. 2006

**Begutachtungen und abgeschlossene Betreuungen außerhalb des Departments**

<b>DoktorandIn</b>	<b>GutachterIn</b>	<b>Thema</b>	<b>Datum</b>
Kristina Spranger	U.Heid; W. Menzel; H. Kamp;	Combining Deterministic Processing with Ambiguity-Awareness – The case of Quantifying Noun Groups in German	Jul. 2006

**Mit-Begutachtungen und abgeschlossene Mit-Betreuungen***Diplomarbeiten*

<b>DiplomandIn</b>	<b>BetreuerIn</b>	<b>Thema</b>	<b>Datum</b>
Michael Christian Jabbusch	W. Menzel	Inhaltsbasierte Erschließung von Datenquellen in Grid-Umgebungen	Jan. 2006
Jorge Homann	W. Menzel	Spezifikation und Ausführung transaktionaler Abläufe in Grid-Umgebungen	Mai 2006
Sören Fandrich	W. Menzel	Schema Integration and Query Processing in Dynamic Environments	Jul. 2006
Björn Waldau	W. Menzel	SQL-Anfragebearbeitung im SQXML-Projekt	Aug. 2006

**Wissenschaftliche Vorträge**

v. Hahn, Walther

- 6. 1. 2006 „Werthaltung von mehrsprachigem lexikalischem Material“ Zürcher Hochschule Winterthur
- 16.3.2006 „Language Technology- Study and Research at University of Hamburg-“, Universität Patras
- 27.4.2006 „Language Technology in the Information Society - Hot issues and open questions –“, Akademie der Wissenschaften Sofia

24. -26. 5. Seminar „Maschinelle Übersetzung“ Universität Granada, Informatik

Menzel, Wolfgang

3.4.2006 Robust parsing: More with less, EACL-Workshop Trento/Italien

Foth, Kilian und Menzel, Wolfgang

7.11.2006 Linguistische Datenstrukturen:Theoretische und empirische Grundlagen der Grammatikforschung Parsing mit gewichteten Constraints Kolloquium des SFB 441 Tübingen

3.-7.7.2006 Parsing mit gewichteten Constraints, 12th European ELSNET Summer School Course: Hamburg

Vertan, Cristina

21.1.2006 „Electronic Dictionaries -management and reusability-“, Conference on Digital Philology, Hamburg

21.1.2006 „Computerphilologie in Hamburg“, Pressekonferenz zur Digital Philology Konferenz, Hamburg

14.3.2006 Building Example Based Machine Translation Systems with few linguistic resources - a perspective for on-line MT, Universität Patras, Griechenland

3.4.2006 International Master in Computational Linguistics and Speech Technology, CHE Tagung Internationale Masterstudiengänge erfolgreich planen

5.-7.4.2006 „Activities in Digital Humanities at the University of Hamburg“, Workshop Ontology based modelling in the Humanities, Hamburg

4.5.2006 Mapping Multilingual Lexical Material on the Ontology in LT4eL Project, OS Natürlichsprachliche Systeme

25.5.2006 Basic principles of Example based Machine Translation, Universität Granada

12.7.2006 Activities in multilingual retrieval and translation at Natural Language Systems division, an der ELSNET Summer School, Hamburg

14.7.2006 Language Technology for eLearning, ELSNET Summer School, Hamburg

13.10.2006 Research directions in the Natural Language Systems Division, Seminar Language Technology for eLearning, state of art and future perspectives, Charles University Prag

4.11.2006 Texttechnologie, Workshop Hamburg Digital Humanities

## 4. Wichtige weitere Aktivitäten

### Mitarbeit in universitären Gremien

- Menzel, Wolfgang
  - Studienreformausschuss
  - IT-Strategiekommission
- v. Hahn, Walther
  - 2. Vorsitzender Hamburg Digital Humanities
- Vertan, Cristina
  - Vorsitzende der Arbeitsstelle Computerphilologie
  - Gründungsmitglied Hamburg Digital Humanities

### Begutachtungstätigkeit

- Menzel, Wolfgang
  - Deutsche Forschungsgemeinschaft
  - MIT-Press
  - 4th Int. Workshop Romand 2006
  - 21st Int. Conf. On Computational Linguistics
  - 44th Meeting of the Assoc. for Computational Linguistics
- Vertan, Cristina
  - „Linguistik online“
  - International Workshop „Natural Language Processing for eLearning“, Varna, September 2006
- v. Hahn, Walther
  - Gutachter mehrerer Konferenzen, darunter COLING06

### Kongressorganisation/-ausrichtung durch Mitglieder der Departmenteinrichtung

- Vertan, Cristina; v. Hahn, Walther; Menzel, Wolfgang
  - 12th ELSNET Summer School of the European Network of Excellence in Speech and Language - Information Fusion in natural languages systems (€ 15.000 von Elsnet Network)
- v. Hahn, Walther; Vertan, Christina
  - Workshop: Ontology based Modelling in Humanities, 5.-7. März 2006 (€ 3000 von Hansische Stiftung)

### Sonstige Aktivitäten

- v. Hahn, Walther; Vertan, Christina
  - Seminar zur Lehrerfortbildung „Neue Ansätze in CS Lehre“, 27-29 November 2006 (Mittel aus EU-Projekt I\*Teach)
- Vertan, Cristina, Menzel, Wolfgang:
  - Projekt Mensch Maschine Kommunikation mit natürlicher Sprache - Schnupperstudium 2006