

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:

Irina Aleksenko, Klaus Dalinghaus M.A., Dipl.-Inf. Michael Daum (bis 9/2007), Dipl.-Inf. Carmen Gámiz Marcos (bis 9/2007), Dipl.-Inf. Thomas Kopinski (ab 10/2006), Dipl.-Inf. Nguyen Thinh Le, Eelco Mossel, Dr. Cristina Vertan

Natalia Elita - Promotionsstipendiatin (DAAD), Tian Gan (MSc.) - Promotionsstipendiat (Internationales Graduiertenkolleg CINACS), Dipl.-Inf. Monica Gavrilă - Promotionsstipendiatin (Universität Hamburg), Patrick McCrae (MLitt) – Promotionsstipendiat (Internationales Graduiertenkolleg CINACS), Martha Yifiru Tachbelie - Promotionsstipendiatin (Universität Hamburg)

Technisches und Verwaltungspersonal:

Karin Jarck (Sekretariat) (bis 9/2007)

Gäste:

Prof. Dr Monica Tataram - University of Bucharest
Prof. Dr. Voichita Ghenghea - Politechnica University of Bucharest
Prof. Dr. Galia Angelova - Bulgarian Academy of Sciences
Prof. Dr. Antonio Rubio - University of Granada

Myroslav Tynovsky, Gaststudent, Charles University Prague
Catalina Oancea, Gaststudentin, University of Bucharest
Mihaela Munteanu, Gaststudentin, University of Bucharest

Christina von Bremen (geb. Maack) – Projekt PROLIV/Kooperation mit dem FB 07
Dr. Kerstin Fischer – University of Southern Denmark
Dr. Solomon Teferra Abate – ehem. DAAD Stipendiat aus Äthiopien

Ehemalige:

Michael Daum, Carmen Gámiz Marcos

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.

Zum 1.10.2007 hat der langjährige Arbeitsbereichsleiter, Prof. Dr. Walther von Hahn, seinen Ruhestand angetreten. In diesem Zusammenhang wurde der Arbeitsbereich formal aufgelöst. Die laufenden Forschungsarbeiten werden jedoch weiterhin fortgeführt.

Forschungsschwerpunkte

Architektur integrierter Sprachverarbeitungssysteme

Angesichts der Vielzahl und Verschiedenheit der Komponenten, die bei der Sprachverarbeitung notwendigerweise miteinander interagieren, haben sich vereinfachende und starre Architekturentwürfe 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. beispielbasierte Verfahren können allein allerdings die Vielfältigkeit der Phänomene in der MÜ nicht lösen. Untersucht werden Hybride Systeme, insbesondere Erweiterungen von beispielbasierten 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
- Universität Potsdam

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
- University of Southern Denmark, Dänemark

Ausstattung

Der Arbeitsbereich NATS hat in den letzten Jahren sein heterogenes Rechnernetzwerk in Richtung auf die Plattformen x86/Linux und PPC/Mac OS X vereinheitlicht. 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. Insgesamt stehen dem Arbeitsbereich aus Haushaltsmitteln 19 PCs und 3 Laptops sowie 3 Macintosh-Computer zur Verfügung. 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:	Constraint-Parsing und psycholinguistische Plausibilität 11/2005 - 2/2008
Geldgeber:	DFG
Gesamtmittel:	27 Monate BAT IIa 24 Monate 2x40h SHK Sachmittel: €18.200
Projekt:	Dia Learn: Constraint-basierte Diagnose für interaktive Lernsysteme 1/2005-1/2010
Geldgeber:	DFG
Gesamtmittel:	48 PM BAT IIa 24 Monate 1x73h, 24 Monate 1x40h SHK Sachmittel: €13.600
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
Projekt:	LING-Net – Linguistische Einführung im Netz 10/2004 – 2/2007
Geldgeber:	Multimedia Kontor
Gesamtmittel:	€5000
Projekt:	Texttechnologien 10/2005 - 12/2007
Geldgeber:	Multimedia-Kontor
Projekt:	PROLIV: Animierte Prozessmodelle komplexerer linguistischer Verfahren

	6/2005 - 5/2008
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:	TAPEJA - Text Analyse mit PERL und JAVA - Ein Programmierungskurs für Linguisten und Literaturwissenschaftler 10/2007 - 10/2008
Geldgeber:	Universität Hamburg
Gesamtmittel:	27 Monate BAT IIa 24 Monate 2x40h SHK Sachmittel: €5.000
Projekt:	BIS21* Center of Excellence Kooperation mit der Bulgarischen Akademie der Wissenschaft 10/2005 – 09/2008
Geldgeber:	EU/IST-FP6
Gesamtmittel:	€10.400

2. Die Forschungsvorhaben der Departmenteinrichtung

Etatisierte Projekte

2.1 Machine Learning für die Strukturanalyse natürlicher Sprache

Thomas Kopinski

Laufzeit des Projektes:

seit 10/2006

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 dieser Arbeit soll untersucht werden, inwiefern durch die Kombination verschiedener komplementärer Ansätze dennoch signifikante Verbesserungen erreicht werden können. Von besonderem Interesse ist dabei die Kopplung regelbasierter Verfahren mit Modellen, die aus Korpusdaten extrahiert wurden, insbesondere dann, wenn beide auch unabhängig voneinander die Analyseaufgabe bereits mit hoher Präzision erfüllen können.

Schlagwörter:

Dependenzgrammatik, Constraints, Support-Vektor-Maschinen

Publikationen aus dem Projekt:

Khmylko, L. Hybrid Parsing with a maximum spanning tree predictor. Masterarbeit TU Hamburg-Harburg, 2007

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 .

Gavrilă, 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,

Gavrilă, M.: „LexMod and LexTool, lexical model, acquisition and editing in MANAGELEX“, Hamburg, 2004., Hamburg Univ., Fachbereich Informatik, Diplomarbeit 2004. R36887

Drittmittelprojekte

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. Articulatory Features are used as an intermediate representation, capturing relevant characteristics of the speech production information.

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.
 McCrae, P. Integrating Cross-Modal Context for PP Attachment Disambiguation. In Proceedings 3rd International Conference on Natural Computation, 2007, Vol. 3, pp. 292 - 296.
 McCrae, P. Menzel, W. Towards a System Architecture for Integrating Cross-Modal Context in Syntactic Disambiguation, Proceedings NLPCS 2007 - 4th. International Workshop on Natural Language Processing and Cognitive Science, pp. 228 - 237.
 Gan, T.; Menzel, W.; Yang S. An Audio-Visual Speech Recognition Framework Based on Articulatory Features. In Proc. Int. Conf. Auditory-Visual Speech Processing, AVSP 2007, pp. 137-141, Tilburg, 2007.

Finanzierung:

Projekt:	High-level information fusion for speech and language 4/2006 bis 11/2010
Sachmittel:	€30.600
Personalmittel:	€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 11/2005-2/2008
Geldgeber:	DFG
Personalmittel:	24 PM BAT IIa/2, 15 PM BAT IIa 24 Monate 1x72h SHK

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 Lernsysteme

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 1/2005-1/2010
Personalmittel:	48 PM BAT IIa 24 Monate 1x73h, , 24 Monate 1x40h 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 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
- Le, N.-T.; Menzel, W. Using constraint-based modelling to describe the solution space of ill-defined problems in logic programming. In *Proceedings 6th International Conference on Web-based Learning*, 2007.
- Le, N.-Th. Diagnosing errors in logic programming - The case of an ill-defined domain. Technical Report, University of Hamburg, Department of Informatics, FBI-HH-B-280/07, 2007.

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 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.I. Cuza, Iasi.

Finanzierung:

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

2.8 Rekombinationsstrategien in der beispielbasierten maschinellen Übersetzung

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

Laufzeit des Projektes:

10/2005 bis 07/2009

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 .

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.

Elita, N.; Gavrila, M.; Vertan, C.: Experiments with String Similarity measures in the EBMT Framework. In Proceedings Europ. Conf. on Recent Advances in Natural Language Processing, RANLP-2007, Borovets, Bulgaria, 2007.

2.9 PROLIV - Animierte Prozessmodelle komplexer Linguistischer Verfahren

v. Hahn, Walther, Prof. Dr.; Vertan, Cristina, Dr.; Gavrila, Monica; Redder, Angelika, Prof. Dr. (FB. 07); von Bremen, Christina (FB. 07)

Laufzeit des Projektes:

von 6/2005 bis 5/2008

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:

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

Vertan, C.; Lemnitzer, L.; Killing, A.; Simov, K.; Evans, D.; Osenova, P.; Monachesi, P. Crosslingual retrieval in an eLearning environment, in proceedings of Artificial Intelligence and Human Oriented Computing, AIIA 2007 (erscheint 2008 auch in Lecture Notes of Computer Science)

Lemnitzer, L.; Killing, A.; Kiril Simov, Vertan, C.; Evans, D.; Cristea, D.; Monachesi, P. Improving the search for learning objects with keywords and ontologies. In Proceedings of Technologically enhanced learning conference 2007 (Best paper award).

Lemnitzer, L.; Mossel, E.; Vertan, C., Language Technology for eLearning - a multilingual approach from the German perspective, in Datenstrukturen für linguistische Ressourcen und ihre Anwendungen, Proceedings of the Biennial GLDV Conference, Gunter Narr Verlag Tübingen, 2007, S. 135-144

Finanzierung:

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

2.10 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.: „LTel - 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

2.11 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 ein „Virtual-Training Center“ implementiert und als Angebot für den Austausch von Lehrinhalten für die Lehrer zur Verfügung gestellt.

Schlagwörter:

Informatik für Lehramtstudierende, Informatik in der Schule, eLearning

Publikationen:

Dodero, G.; Vertan, C.; Ratcheva, D.; Stefanova, E.; Miranowicz, M.; Musakoviene, V. The Virtual Training Center: A support for Teachers Community, in Proceedings of the 3rd Balcan Conference in Informatics, 27-29 September 2007, Sofia, pp. 349-362

Vertan, C.; v. Hahn, W.; InnovativeTeacher - Enhanced IT Skills in Education, in Proceedings of the German e-Science Conference, Mai 2007, Baden-Baden.

Finanzierung:

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

2.12 Morphology-based language modeling for Amharic

Yifiru Tachbelie, Martha

Laufzeit des Projektes:

4/2006 bis 3/2008

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.

Publikationen:

Yifiru Tachbelie, M.; Menzel, W. Sub-word based language modeling for Amharic. In *Proc. Europ. Conf. on Recent Advances in Natural Language Processing, RANLP-2007*, Borovets, Bulgaria, 2007.

Schlagwörter:

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

Finanzierung:

Geldgeber:	Universität Hamburg - Nachwuchsförderung
Laufzeit der Förderung:	4/2006 bis 3/2008
Mittel:	€23.400

2.13 Texttechnologien

v. Hahn, Walther

Laufzeit des Projektes:

10/2005 - 05/2008

Projektbeschreibung

Texttechnologie umreißt den inhaltlichen Umgang mit Texten auf einem Computer, eine Kompetenz, die heute von jedem Absolventen einer Sprach- und Literaturwissenschaft erwartet wird, der in einen Medienberuf eintritt oder ihn anstrebt. Es setzt voraus, dass elementare Computerkenntnisse im Betriebssystem und Büroanwendungen bereits vorhanden sind. Weitere maschinelle Auswertungen der Semantik, Syntax oder Wortstruktur eines Textes sind aber an Kenntnisse aus der Informatik und Computerlinguistik gebunden, die Fächer die solche Texttechnologien entwickelt haben. Das Projekt entwickelt ein eLearning Modul über nicht-elementare Auswertung von textuellen Ressourcen auf dem Computer für sprach- literatur- und medien-wissenschaftliche Studierende als technologische Zusatzqualifikation.

Schlagwörter :

E-learning, Linguistik, Sprachverarbeitung, Computerphilologie

Finanzierung:

Geldgeber:	multimedia Kontor Hamburg
Laufzeit der Förderung:	10/2005 bis 5/2008
Personalmittel:	€5.000

2.14 TAPEJA - Text Analyse mit PERL und JAVA - Ein Programmierungskurs für Linguisten und Literaturwissenschaftler

Vertan, Cristina

Laufzeit des Projektes:

10/2007 - 10/2008

Projektbeschreibung:

Das Projekt richtet sich an Studierende der Linguistik und Literaturwissenschaften, die sich mit neuartigen computergestützten Textanalysen vertraut machen wollen. In erster Linie geht es um die Benutzung von existierenden Modulen, geschrieben in Perl oder Java, die aber ohne ein grobes Verständnis der Programmierverfahren sehr schwierig ist. Das Projekt wird auch elementare Programmierkenntnisse für Studierende in der Geisteswissenschaft vermitteln. Die Studierenden sollten am Ende des Moduls in der Lage sein, Textanalysen mit Hilfe von gegebenen Programmen durchzuführen und auch leichte Programme selbst zu schreiben.

Das Modul bietet sich als eine notwendige Ergänzung zu den bereits existierenden Modulen zur Computerphilologie an.

Schlagwörter:

E-learning, Linguistik, Programmierung

Finanzierung:

Geldgeber:	Universität Hamburg
Laufzeit der Förderung:	10/2007 bis 10/2008
Personalmittel:	€4.500
Sachmittel:	€500

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

- Dodero, G.; Vertan, C.; Ratcheva, D.; Stefanova, E.; Miranowicz, M.; Musakoviene, V. The Virtual Training Center: A support for Teachers Community, in Proceedings of the 3rd Balcan Conference in Informatics, 27-29 September 2007, Sofia, pp. 349-362
- Elita, N.; Gavrilă, M.; Vertan Experiments with String Similarity measures in the EBMT Framework. In Proceedings Europ. Conf. on Recent Advances in Natural Language Processing, RANLP-2007, Borovets, Bulgaria, 2007
- Fetzer, A.; Fischer, K. (Hrsg.) Lexical Markers of Common Grounds. Number 3 in Studies in Pragmatics. Amsterdam: Elsevier, 2007
- Fischer, K.; Lohse, M. Shaping naive users' models of robots' situation awareness. In Proceedings of the 16th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2007), Jeju, Korea, August 2007
- Gan, T.; Menzel, W.; Yang S. An Audio-Visual Speech Recognition Framework Based on Articulatory Features. In Proc. Int. Conf. Auditory-Visual Speech Processing, AVSP 2007, pp. 137-141, Tilburg, 2007
- Le, N.-Th. Diagnosing errors in logic programming - The case of an ill-defined domain. Technical Report, University of Hamburg, Department of Informatics, FBI-HH-B-280/07, 2007
- Le, N.-Th.; Menzel, W. Using constraint-based modelling to describe the solution space of ill-defined problems in logic programming. In Proceedings 6th International Conference on Web-based Learning, 2007
- Lemnitzer, L.; Killing, A.; Kiril Simov, Vertan, C.; Evans, D.; Cristea, D.; Monachesi, P. Improving the search for learning objects with keywords and ontologies. In Proceedings of Technologically enhanced learning conference 2007, Crete, Greece (Best paper award)
- Lemnitzer, L.; Mossel, E.; Vertan, C., Language Technology for eLearning - a multilingual approach from the German perspective, in Datenstrukturen für linguistische Ressourcen und ihre Anwendungen, Proceedings of the Biennial GLDV Conference, Gunter Narr Verlag Tübingen, 2007, S. 135-144
- McCrae, P. Integrating Cross-Modal Context for PP Attachment Disambiguation. In Proceedings 3rd International Conference on Natural Computation, 2007, Vol. 3, pp. 292 - 296
- McCrae, P. Menzel, W. Towards a System Architecture for Integrating Cross-Modal Context in Syntactic Disambiguation, Proceedings NLPCS 2007 - 4th. International Workshop on Natural Language Processing and Cognitive Science, pp. 228 - 237
- Teferra Abate, S.; Menzel, W. Syllable-Based Speech Recognition for Amharic. In Proc. Workshop on Computational Approaches to Semitic Languages: Common Issues and Resources, ACL, Prague, Czech Republic, 2007
- Teferra Abate, S.; Menzel, W. Automatic Speech Recognition for an Under-Resourced Language - Amharic. In Proc. 8th Annual Conference of the International Speech Communication Association, Interspeech 2007, Antwerp, Belgium, 2007

- Vertan, C.; Lemnitzer, L.; Killing, A.; Simov, K.; Evans, D.; Osenova, P.; Monachesi, P. Crosslingual retrieval in an eLearning environment, in proceedings of Artificial Intelligence and Human Oriented Computing, AIIA 2007, Rom (erscheint 2008 auch in Lecture Notes of Computer Science)
- Vertan, C.; v. Hahn, W.; InnovativeTeacher - Enhanced IT Skills in Education, in Proceedings of the German e-Science Conference, Mai 2007, Baden-Baden
- Yifiru Tachbelie, M.; Menzel, W. Sub-word based language modeling for Amharic. In Proc. Europ. Conf. on Recent Advances in Natural Language Processing, RANLP-2007, Borovets, Bulgaria, 2007

Wichtige Publikationen aus den vergangenen Jahren 2004-2006

- Fischer, K. What Computer Talk Is and Is not: Human-Computer Conversation as Intercultural Communication, volume 17 of Linguistics - Computational Linguistics. Saarbrücken: AQ, 2006
- Fischer, K. The role of users' preconceptions in talking to computers and robots. In Fischer, K. (Hrsg.) Proceedings of the Workshop on 'How People Talk to Computers, Robots, and Other Artificial Communication Partners', Hansewissenschaftskolleg, Delmenhorst, April 21-23, 2006, pages 112-130. SFB/TR8 Report 010-09_2006, 2006
- Fischer, K.; Bateman, J. A. Keeping the initiative: an empirically motivated approach to predicting user-initiated dialogue contributions in HCI. In Proceedings of EACL '06, 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., 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
- 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
- Le, N.-T. and Menzel, W.: Problem solving process oriented diagnosis in logic programming. In Proceedings of the 14th International Conference on Computers in Education, volume 134. IOS Press, 2006
- Menzel, W.: Constraint-based modelling and ambiguity. Int. Journal on Artificial Intelligence in Education, 16:29-63, 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
- 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.), Knowledge Systems and Translation. Berlin 2005 S. 61 - 81
- 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. & 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.: 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
- Fischer, K.: Addressee-oriented Blending in Spatial Object Localisation Tasks. In Proceedings of Spatial Cognition '04, Frauenchiemsee, 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
- 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

Begutachtungen und abgeschlossene Betreuungen im Department

Diplomarbeiten

DiplomandIn	BetreuerIn	Thema	Datum
Jochen Maaß	W. Menzel (N. Ritter)	Funktionsweise und Manipulierbarkeit von Suchmaschinen	Jan. 2007
Julia Nabel	W. Menzel (M. Sonnenschein)	Data Mining Verfahren zur Erzeugung von Regelsystemen am Beispiel von Wildunfalldaten	Mai 2007
Timo Baumann	W. Menzel (G. Möhler)	Automatische Erkennung von Akzentuierungen und Phrasierungen in Sprachsynthesekorpora	Mai 2007

Baccalaureatsarbeiten

Name	GutachterIn	Thema	Datum
Gerhard Reuss	W. von Hahn	Einsatz und Einfluß von EBMT-Methoden auf Crosslingual Information Retrieval	Aug. 2007
Felix Dobslaw	W. von Hahn	An Architecture for Crosslingual Retrieval; Software-Design & Evaluation	Feb. 2007

Mit-Begutachtungen und abgeschlossene Mit-Betreuungen

Diplomarbeiten

DiplomandIn	BetreuerIn	Thema	Datum
Robertino Solanas	N. Ritter (W. Menzel)	Untersuchung automatisierter Schema-Matching-Verfahren für dynamische Datenintegration in Grid-Umgebungen	Mai 2007
Ning Li	S. Voß (W. Menzel)	Eine Evaluation der Core Vector Maschine zur Lösung betriebswirtschaftlicher Klassifikationsprobleme	Jun. 2007
Christoph Walters	Ch. Habel (W. Menzel)	Using Dynamics in Sketch Map Understanding	Jun. 2007
Alexander Holbreich	W. Lamersdorf (W. Menzel)	Transaktionsunterstützung für verteilt ausgeführte mobile Prozesse	Aug. 2007
Marina Svagusa	J. Zhang (W. Menzel)	Entwurf einer Architektur eines natürlichsprachlichen Mensch-Maschine-Dialogs für einen Serviceroboter zur Manipulation bekannter Objekte	Sep. 2007
Jörn Schumann	B. Neumann (W. Menzel)	Umsetzung einer Groupware-Lösung anhand eines Prototypen für den Hamburger Schul-Support-Service	Sep. 2007
Tihomir Magdic	N. Ritter (W. Menzel)	Dienstkomposition mit Hilfe semantischer Service-Templates	Nov 2007
Feng Wang	B. Neumann (W. Menzel)	Probabilistische Inferenz in einer Abstraktionshierarchie	Dez. 2007

Wissenschaftliche Vorträge

- Aleksenko, Irina; von Bremen, Christina; Gavrilu, Monica; von Hahn, Walther; Redder, Angelika; Vertan, Cristina:

- 30.8.2007 Proliv - Animated Models for Visualizing Complex Linguistics Theories, LSP Conference, Hamburg
 Elita, Natalia; Gavrilă, Monica:
 28.8.2007 Semantics for Translation Memories, LSP Conference, Hamburg
 von Hahn, Walther:
 6.3.2007 Probleme und Lösungen der Maschinellen Übersetzung, Politechnica University, Bukarest
 Menzel, Wolfgang:
 5.-7.9.2007 Language Processing, Course at the 2nd CINACS Summer School, Beijing, China (zusammen mit Carola Eschenbach und Christopher Habel)
 25.11.2007 Important application areas of NLP in the Ethiopian languages/context, Int. Workshop on Language Resource Development for Language Technology Research in Ethiopian Languages, Addis Ababa, Ethiopia
 Vertan, Cristina:
 13.3.2007 Korpus-basierte Verfahren in der Maschinellen Übersetzung, Politechnica University, Bukarest
 10.9.2007 Example-based Machine Translation for cross-lingual retrieval, Tutorial Machine Translation Summit, Copenhagen

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
 Europ. Conf. on Recent Advances in Natural Language Processing, RANLP 2007
 Conference on Empirical Methods in Natural Language Processing, EMNLP 2007
 Conference on Computational Natural Language Learning, CoNLL 2007
 International Journal on Artificial Intelligence in Education
 Vertan, Cristina :
 Zeitschrift für Sprachwissenschaften
 Linguistik onLine
 Machine Translation Summit 2007
 Int. Conference on Language Resources and Evaluation, LREC 2008

Konferenzorganisation

- Vertan, Cristina; von Hahn, Walther:
 Internationales Symposium für Fachsprachen, Hamburg, 23-31 August 2007
 Vertan, Cristina:
 Mitorganisation des Workshops: Natural Language Processing and Knowledge Representation for eLearning, at RANLP 2007, Borovets, Bulgarien

Preisverleihungen an Mitglieder der Departmenteinrichtung

- Lemnitzer, L.; Killing, A.; Kiril Simov, Vertan, C.; Evans, D.; Cristea, D.; Monachesi, P.:
 Best paper award der Conference on Technologically Enhanced Learning, Crete, Greece, 2007