

# Curriculum Vitae

*PD Dr. Günter Neumann*



## Address

Name: PD Dr. Günter Neumann  
Office address: DFKI GmbH  
Language Technology Laboratory  
Stuhlsatzenhausweg 3  
D-66123 Saarbrücken  
Phone: +49 6897 85775 5298  
URL: <http://www.dfki.de/~neumann>  
E-mail: [neumann@dfki.de](mailto:neumann@dfki.de)

## Research interests

The core of my research interest is in the field of natural language driven human computer interaction, where natural language is the common interface for extracting, classifying, and talking about information from large-scale unstructured text sources. Specific topics of interest are: natural language parsing and generation, Question answering, Information extraction, Textual inference, Machine Learning for language technology.

## Education

- 10th March 1989: Diploma degree in Computer Science with a grade at 1.0 from the Technical Faculty of the Saarland University. Thesis title: “POPEL-HOW: Parallel and incremental generation of natural language sentences from conceptual representations”. Supervisor: Prof. Dr. W. Wahlster.
- 3rd November 1994: Doctoral degree in computer science (magna cum laude) from the Technical Faculty of the Saarland University. Thesis title: “A Uniform Computational Model for Natural Parsing and Generation”. Joint supervisors: Prof. Dr. H. Uszkoreit and Prof. Dr. W. Wahlster.

- May 1999: DFKI research fellow, the decision procedure is similar to that for the appointment of an associate professor, C3/W2.
- 14th July 2004: Habilitation in “Computational Linguistics” with the topic “Domain-adaptive Core Technology for Natural Language Analysis and Generation” at the Saarland University, Germany. Committee: Prof. Dr. Pinkal, Prof. Dr. Uszkoreit, Prof. Dr. Wahlster (all: Univ. of the Saarland), Prof. Dr. Nerbonne (Univ. of Groningen).

## Professional experience

**January 2012 – December 2014** I am the site project leader of the project “EXCITEMENT: EXploring Customer Interactions through Textual EntailMENT” which is funded by the EC as part of FP7-ICT. I am the work package leader and responsible for the interaction between research and development.

**March 2011 – September 2011** I am the project leader of “DILIA-IE” funded by researchGate GmbH in the area of unsupervised and domain independent keyphrase extraction from a large scale corpus of scientific abstracts (more than 40 million abstracts from natural science and medicine). Our methods are integrated into the social platform of researchGate GmbH.

**January 2010 – December 2012** I am involved in the project “THESEUS\_ORDO\_TechWatch - Ontologies and Language Technologies for Technology Observation in Theseus-ORDO” where I am working as a project leader.

**March 2009 – December 2009** I am involved in the project “THESEUS-CTC\_IUI - Theseus Core Technology Cluster” where I am working as a conceptual lead for the research and development of novel semantic search technology on basis of the ODP (Ontology-based Dialogue Platform) framework.

**March 2009 – December 2009** I am involved in an industrial project funded by the Deutsche Telekom. The topic is on the improvement of semantic search on basis of computational morphology.

**June 2008 – December 2010** I am the site project leader of the DiLiA project that is funded by the Investitionsbank Berlin. The main goal is the development of hybrid approaches to information extraction for digital scientific libraries. Our partners are the IUI-Lab of DFKI (headed by Prof. Dr. Wolfgang Wahlster) and the Max Planck Digital Library (MPDL).

**June 2008 – September 2009** I am the project leader of the ConQA project funded by Saarland. The main goal is the research and development in domain-specific ontology-based question answering for a restricted but scalable knowledge area by exploiting web-based ontology population.

**April 2007 – November 2007** I am the project leader of the DFKI project IDEX funded by the Investitionsbank Berlin. The main goal is the research and development of novel methods for interactive dynamic information extraction in the domain of risk analysis.

**Oct. 2006 – September 2009** I am the site project leader of the QALL-ME project, funded by EC. The main goal of the project is the development of question answering learning technologies in a multiLingual and Multimodal environment. DFKI’s work package responsibilities concern the design and development of a Service-Oriented Architecture (SOA) and the integration of the QALL-ME system.

**Jan. 2006 – December 2008** I am co-project leader of HyLaP (together with Prof. Dr. Hans Uszkoreit), a BMBF funded project. The main goal of the project is the research and development of *Hybrid Language Processing Technology for a Personal Associative Information Access and Management Application*. My main responsibilities are in the area of web-based question answering with a scientific focus on textual inference and Machine Learning for QA. Project reviewers: Dr. O. Stock (IRST) and Prof. Dr. D. Scott (University of Brighton).

**October 2004 – April 2007** I am actively involved in the project SmartWeb, a large BMBF-funded research project about *Mobile Broadband Access to the Semantic Web*. I am responsible for the development of open-domain question answering and its interface to mobile speech-based input devices, Web-based search engines and answer selection.

**January 2003 – December 2005** I am the co-project leader of Quetal (together with Dr. Anette Frank), a BMBF funded project. The main goal is the development of a hybrid, web-based and multilingual question-answering system. The focus of my research is on open-domain cross-language question-answering, deep answering processing, and QA-relevant semantic web services. Project reviewers: Dr. O. Stock (IRST) and Prof. Dr. D. Scott (University of Brighton).

**January 2003 – August 2003** I am the co-project leader (together with Bernd Kiefer) of the PAW (parsing of word graphs) project funded by DaimlerChrysler. The main task was to implement a robust and flexible parser for Daimler's speech recognizers.

**June 2001 – December 2001** I am involved in the EU-funded project "ONTOWEB: Ontology-based information exchange for knowledge management and electronic commerce" as part of DFKI's participation.

**January 2000 – December 2002** I am the project leader of the BMBF-funded project WHITEBOARD ("Multilevel Annotation for Dynamic Free Text Processing"). Research focus: shallow robust text understanding, Machine learning for information extraction, text Mining, XML-based text IO specifications. The project is annual reviewed by DFKI's international scientific advisory board. Project's main reviewer: Prof. Dr. Jaime Carbonell, CMU, USA.

**September 1999 – September 2000** I am a technical consultant in the project TIM funded by Deutsche Telekom in the area of intelligent information extraction.

**May 1999 – July 1999** I am the project leader of the project TMB funded by Dresdner Bank in the area of text mining and intelligent information extraction.

**June 1998 – June 2001** I am the site project leader of the project GETESS (*German Text Exploitation and Search System*). GETESS was a joint project between the universities Karlsruhe and Rostock, DFKI and the SME Gecko mbH. DFKI's responsibilities are in the area of multi-lingual abstract generation, user query analysis and text summarization. From Jan. 1999 to June 1999 our research group hosted Prof. Tsunenori Mine from the Kyushu University, Japan.

**January 1997 – March 2000** I am the project leader of the project PARADIME (*Parameterizable Domain-adaptive Information and Message Extraction*). PARADIME's responsibilities are in the area of domain-adaptive information extraction, shallow text processing, and Machine Learning. Project's main reviewer: Prof. Dr. Jaime Carbonell, CMU, USA.

**October 1996 – April 1997** I am responsible for the technical coordination of the development of an information extraction prototype for a commercial customer (ESG Munich).

**May 1992 – December 1996** I am a full-time researcher at the Language Technology Laboratory of DFKI. I was a member of the projects DISCO (Dialogue System for Cooperating Agents),

PARADICE (Parameterizable NL Discourse Core Engine) and COSMA (Cooperative Schedule Management Agents). I was responsible for the development and implementation of an object-oriented system architecture, uniform algorithms for parsing and generation, development of Machine Learning techniques for intelligent data analysis, and the development of intelligent methods for information extraction.

**January 1991 – April 1992** I am the project leader of the DFG (German Science Foundation) funded project BILD (Bidirectional Linguistic Deduction) of the SFB 314 (Artificial Intelligence and Knowledge-based Systems) headed by Prof. Dr. H. Uszkoreit. The main research focus was in the area of grammar reversibility, chart-based uniform algorithms for parsing and generation, and computational models for self-monitoring and revision. During the project phase I worked together with Dr. Gertjan van Noord, now University of Groningen, the Netherlands.

**March 1989 – April 1992** I am a full time researcher at the Department of Computational Linguistics of the Saarland University headed by Prof. Dr. H. Uszkoreit.

**October 1985 – February 1989** I am a research assistant in the project Xtra (Expert Translator) of the SFB 314 (Artificial Intelligence and Knowledge-based Systems) headed by Prof. Dr. W. Wahlster. I was responsible for the implementation of the German morphological package Morphix and SB-PATR, a re-implementation and extension of the PATR-II constraint-based grammar development tool. During my diploma thesis I developed a parallel and incremental natural language generation system following a distributed AI approach. All these implementations have been used in the daily research work of the AI-lab's projects.

## Research visits

**March 2007** research visit at the Ecole Nationale de l'Industrie Minérale (Rabat, Morocco) working together with Prof. Dr. Abdelhadi Soudi in the area of multilingual morphology.

**July 2002** research visit at M.I.T. AI Laboratory working together with Dr. J. Mallery in the area of uniform parsing and generation of constraint-based natural language grammars.

**June – July 2001** guest researcher at the Language Technologies Institute (LTI) of the School of Computer Science at Carnegie Mellon University, Pittsburgh, PA. Working area: open-domain question answering working together with Prof. Dr. Eric Nyberg and Prof. Dr. Jaime Carbonell. The focus of my visit was on exploring inter-lingua-based representations in order to support answer validation, and on the integration of shallow and deep text analysis.

**July 1996, August 1997 and November 1998** Guest researcher at the Center for the Study of Linguistics and Information (CSLI), Stanford University working together with Dr. Dan Flickinger and Dr. Ann Copstake in the area of efficient parsing and generation of high-level NL grammars (viz. HPSG). The main focus was the adaptation of a speed-up learning algorithm on basis of Explanation-based Learning (EBL) which I invented at DFKI for the efficient processing of HPSG-based grammars developed at CSLI.

## Project acquisition

I was actively involved in the acquisition of the following projects:

### Research

- BiLD (Bidirectional Linguistic Deduction), SFB 314, funded by DFG (German Science Foundation), 1991.

- Twenty-One (Disclosure and dissemination of documents on sustainable development), funded by the European Commission (EC), 1996.
- Mulindex (Multilingual Indexing, Editing and Navigation Extensions for the World Wide Web) (Co-authored with Dr. G. Erbach), funded by the European Commission (EC), 1997.
- Paradime (Parameterizable Domain-adaptive Information and Message Extraction) (I was the principal author), funded by BMBF (German department for education, science, research and technology), 1997.
- Getess (German Text Exploitation and Search System) (I was the principal author for DFKI), funded by BMBF, 1998.
- Whiteboard (Multilevel Annotation for Dynamic Free Text Processing), together with Dr. A. Bredenkamp; funded by BMBF, 2000.
- Ontoweb (Ontology-based information exchange for knowledge management and electronic commerce); funded by EU, 2001.
- Quetal (A Multilingual Hybrid Question-Answering System for Exploring Large Collections of Annotated Free Text Documents) (I am the co-principal investigator and proposal writing coordinator); funding by BMBF, 2003.
- HyLaP (Hybrid Language Processing Technology for a personal associative information access and management application ) (I am the co-principal investigator together with Hans Uszkoreit and proposal writing coordinator); funding by BMBF, 2006.
- QALL-ME (Question Answering Learning Technologies in a multilingual and multimodal Environment). I am the principal investigator on DFKI's side; funded by the European Commission (EC), 2006.
- IDEX (Interactive Dynamic Information Extraction). I am the principal investigator; funded by Investitionsbank Berlin, 2007.
- ConQA (Controlled Question Answering). I am the co-principal investigator together with Hans Uszkoreit; funded by Saarland, 2008.
- DiLiA (Digital Library Assistant). Co-principal investigator together with Dr. Norbert Reithinger (IUI-lab, DFKI) and Dr. Laurent Romary (MPDL, Max Planck Society), funded by Investitionsbank Berlin, 2008.
- EXCITEMENT (EXploring Customer Interactions through Textual EntailMENT). I am the principal investigator on DFKI's side; funded by the European Commission (EC), 2011.

## Industry

- ESG-SMES (intelligent message extraction) funded by an industrial partner; I was technical project leader, 1996.
- TMB (Text Mining in the Banking area) funded by Dresdner Bank; I was project lead (together with Dr. R. Schäfer), 1999.
- TIM (Telecommunication and Information Management) funded by German Telecom; I held several oral presentations, 1999.
- PAW (Parsing of word graphs) funded by DaimlerChrysler; I was the principal investigator (together with B. Kiefer and H.-U. Krieger), 2003.
- GelbeSeiten (Semantic search) funded by the Deutsche Telekom. I am the co-principal investigator (together with Dr. Markus Loeckelt, IUI-lab, DFKI), 2009.

- EVE (Dialogues with the virtual character Eve of the YelloStrom Web page); I was technical project leader (together with Dr. A. Ndiaye, IUI-lab, DFKI), 2010.
- “Improving the search functionality of OMQ diagnostic” funded by OMQ GmbH, 2010.
- “Automatic term extraction” funded by OMQ GmbH, 2011.
- “Unsupervised key phrase extraction from scientific publications” funded by researchGate GmbH, 2011.

## Teaching experience

- June 2005 – June 2012: invited lecture in *Current Topics in Computational Linguistics* at the University of Klagenfurt as part of the Faculty of business and computer science (25 lecture hours as block course).
- Winter 2010, Winter 2011: lecture on *Information Extraction* (2SWS)
- Winter 2010, Summer 2011: software project *Named Entity Processing*.
- Winter 2009: seminar on *Wikipedia Mining* (2 SWS).
- Winter 2005 – Winter 2009, Winter 2011: selected lectures in the Master course on *Language Technology I* (covering Chunk Parsing, Machine Learning for Named Entity Recognition, Question-Answering, Information Retrieval and Textual Inference).
- Summer 2008: seminar about “Computational Linguistics in economy and society” (2 SWS).
- Summer 2008: seminar on *Information Extraction in the Biomedical Domain* (2 SWS).
- Summer 2007: software project “Data-driven web-based QA”.
- Winter 2006: seminar on *Current Trends in Information Extraction* (2 SWS).
- Summer 2005: seminar on *Data-Oriented Parsing and Generation* (2 SWS).
- August 2004: tutorial on *Intelligent Information Extraction* given at the 16th European Summer School in Logic, Language and Information, ESSLI 2004.
- Winter 2003: software project: *Cross-language question-answering* (2 SWS).
- Winter 2001: lecture course on *Information Extraction and Question-Answering Systems: Foundations and methods* (4 SWS).
- Winter 2001: software course *Information extraction and Text Mining* (2 SWS).
- Summer 2001: seminar on *Machine Learning for Information Extraction and Text Mining* (3 SWS).
- Summer 1999: software course *Shallow parsing with finite state grammars* (2 SWS).
- Winter 1997: seminar *Multi-lingual information retrieval and extraction* (2 SWS).
- Winter 1995: seminar *Parsing and generation with constraint-based grammars* (2 SWS).
- Summer 1992: programming course on *Introduction to functional programming* (4 SWS).
- Summer 1991: programming course on *Introduction to functional programming* (4 SWS).
- Summer 1990: programming course on *Symbolic and object-oriented programming: Common Lisp und CLOS* (4 SWS).

- Winter 1989: lecture course on *Introduction to natural language systems* (4 SWS).
- Summer 1989: programming course on *Introduction to functional programming* (4 SWS).

The courses and seminars have been taught at the departments of computer science and computational linguistics of the Saarland University, if not otherwise specified. For most of them online material is available from my DFKI home page.

## Supervised Theses

1. Ludwig Dickmann  
Flache Analyseverfahren zur Identifikation und Extraktion von Personennamen aus elektronischen Texten, Diploma Thesis, 1998, Computational Linguistics, Saarland University.
2. Christian Braun  
Flaches und robustes Parsen deutsche Satzgefüge, Diploma Thesis, 1999, Computational Linguistics, Saarland University.
3. Volker Morbach  
A Hybrid Machine Learning Approach for Information Extraction, Diploma Thesis, 2002, Computer Science, Saarland University.
4. Holger Neis  
Cross-linguale textuelle Frage-Antwort-Verarbeitung: Algorithmische Methoden, Diploma Thesis, 2003, Computational Linguistics, Saarland University.
5. Hubert Schlarb  
Cross-linguale textuelle Frage-Antwort-Verarbeitung: Linguistische Methoden, Diploma Thesis, 2003, Computational Linguistics, Saarland University.
6. Rob Basten  
Answering Open-domain Temporally Restricted Questions in a Multilingual Context, Master Thesis, 2005, Computer Science, University of Twente together with Prof. Franziska de Jong.
7. Alejandro Figueroa  
Genetic Algorithms for syntactic and data-driven Question Answering on the Web, Master Thesis, 2006, Computational Linguistics, Saarland University.
8. Christian Spurk  
Ein minimal überwacht Verfahren zur Erkennung generischer Eigennamen in freien Texten, Diploma Thesis, 2006, Computational Linguistics, Saarland University.
9. Rui Wang  
Textual Entailment Recognition — A Data-Driven Approach, Master Thesis, 2007, Computational Linguistics, Saarland University.
10. Alexander Volokh  
Datenbasiertes Dependenzparsing, Bachelor Thesis, 2008, Computational Linguistics, Saarland University.
11. Andrea Horbach  
Relation Extraction from Topic Related Web Texts, Diploma Thesis, 2009, Computational Linguistics, Saarland University.
12. Alexander Volokh  
Data-Driven Transition-Based Dependency Parsing with Maximum Entropy Models, Master Thesis 2009, Computational Linguistics, Saarland University.

13. Alejandro Figueroa  
Finding Answers to Definition Questions on the Web, PhD thesis, 2010, Saarland University.
14. Carolin Shiadeh  
Named Entity Recognition from Arabic Text, Bachelor Thesis, 2010, Computer Science, Saarland University.

### Membership in academic boards

- I was the scientific supervisor of Markus Löckelt (Computer Science department of Saarland University) and Daniel Sonntag (Computer Science department of Saarland University).
- I was member of the **2005 PhD board of examiners** of Dr. Tianfang Yao (Computer Science department of Saarland University).
- I was member of the **2009 PhD board of examiners** of David Tomas (Computer Science department of University of Alicante, Spain).
- I was a member of the **2009 Habilitation board** of the Faculty of Philosophy of the Saarland University for the examination of Dr. Sabine Schulte im Walde.

### Membership in program committees

Annual Meeting of the Association for Computational Linguistics (American Chapter, European Chapter and HLT) — 1997, 1998, 1999, 2003, 2007, 2008, 2010, 2011, 2012

Biologically Inspired Cognitive Architectures (BIAC) – 2011

Conference on Multilingual and Multimodal Information Access Evaluation (Question Answering lab CLEF-QA) — 2007, 2008, 2010

Conference Empirical Methods in Natural Language Processing (EMNLP) — 2001, 2002, 2009, 2011

European Conference on Artificial Intelligence (ECAI) — 2008

ECAI Workshop on Semantic Authoring, Annotation & Knowledge Markup (SAAKM) — 2002

Extended Semantic Web Conference (ESWC) — 2012

IJCAI workshop on Ontology Learning, OL-2001

International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning (AKRR) — 2005

International Conference on Computational Linguistics (Coling) — 2000, 2004, 2008, 2010

International Conference on Management of Emergent Digital EcoSystems (MEDES) — 2009, 2010, 2011

International Conference on Applications of Natural Language to Information Systems (NLDB) — 2007, 2008, 2009, 2010, 2011, 2012

International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2004) — 2004

International Conference on Language Resources and Evaluation (LREC) — 2002, 2004, 2008, 2010, 2012

International Conferences on Pervasive Patterns and Applications (PATTERNS) — 2010, 2011

Workshop on “The Peoples Web meets NLP: Collaboratively Constructed Semantic Resources and their Applications to NLP” — 2009, 2010, 2012

Workshop on “Logic, Learning, and Language” — 2000, 2001

Workshop on Question Answering over Linked Data (QALD-1) — 2011

Workshop on Systems and Frameworks for Computational Morphology (SFCM) — 2009, 2011

Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA) — 2010, 2011, 2012

Workshop on Interacting with Linked Data (ILD) — 2012

Workshop Information Retrieval of the special group in IR of the German Informatics Society — 2008, 2010

## Organizations of Workshop, Symposium and other events

**September 2008** Organization of the workshop “ONTOLOGY-BASED INFORMATION EXTRACTION SYSTEMS (OBIES 2008)” as part of the 31th German Conference on Artificial Intelligence (KI’2008), together with Benjamin Adrian (DFKI; chair), Alexander Trousov (IBM) and Birslav Popov (Ontotext).

**June 2008** Organization of the symposium on “Human Language Technologies in Africa: Status and Prospects” funded by the Alexander von Humboldt foundation together with Prof. Abdelhadi Soudi (Rabat, Morocco; chair), Prof. Dafydd Gibbon (Bielefeld, Germany) and Prof. Justus Roux (Stellenbosch, South Africa).

**August 1996** Organization of the workshop “corpus-oriented semantic analysis” during the “12th European Conference on Artificial Intelligence, ECAI-96”, Budapest, 1996, together with Daniel Kayser, Francois Lvy, Laboratoire d’Informatique de Paris-Nord and Joachim Quantz (workshop chair), technical university Berlin.

## Reviewing of Scientific Journal and Book Chapter Publications

- Semantic Web Journal (2011)
- Journal of Intelligent Information Systems (2008)
- Journal of Language Resources and Evaluation (2006, 2011)
- Journal of Data and Knowledge Engineering (2006)
- Two chapters in the book “Swarm Intelligence and Data Mining” (eds. A. Abraham, C. Grosan and V. Ramos), 2005.
- Journal of Automata, Languages and Combinatorics (2005)
- Journal of Speech Communication (2005)

- Journal of Artificial Intelligence (2004)
- Journal of Applied Artificial Intelligence (2004)
- Journal of Artificial Intelligence in Medicine (2003)
- Journal of Computational Linguistics (1995, 2000, 2002)
- Journal of Natural Language Engineering (2002, 2003)
- Journal “Künstliche Intelligenz” (2008)
- Iranian Journal of Science and Technology (2003)
- A chapter in the book “Robustness in language and speech technology” (eds. Junqua, van Noord), 1998.

## Reviewing of Research Proposals

**July 2007** Project proposal reviewer for the Czech Science Foundation (area: Data-Mining)

**October 2006** Project proposal reviewer for the Council of Physical Sciences of the Netherlands Organization for Scientific Research (NWO) (area: Unsupervised Data-Oriented Parsing)

**May 2006** Project proposal reviewer for the “Schweizerische Nationalfonds zur Förderung der Wissenschaftlichen Forschung” (FNSNF)(area: Text Mining)

**January 2006** Invited participation to the workshop about “Interactive Systems and New Interface Technology” organized by the Interfaces Unit of the Directorate General Information Information Society and Media, EC, Luxembourg, 30/31.1.2006. The main workshop’s task was to review state-of-the-art and to define the key research priorities for FP7, in the areas of natural, intuitive and easy-to-use interfaces. I presented ideas about “implicit personal information access” and was member of a small working group brainstorming about “Implicit Interfaces” (together with Dr. Philipp Koehn, Prof. Chabane Djeraba, Prof. Maurizio Omologi, Prof. Hans Uszkoreit, Prof. Alex Waibel, Prof. Yorik Wilks).

**2006 – 2008** Member of the advisory board for the BMBF-funded project “Ontoverse - the pilot project for cooperative network-supported knowledge management in the field of Life Sciences.”

**May 2004** Active participation in the construction of the Roadmap for Language and Speech Technology Resources, created under the auspices of ELSNET, ENABLER and ELRA. Area of expertise: question answering.

**April 2003** Project proposal reviewer for the Council of Physical Sciences of the Netherlands Organization for Scientific Research (NWO) (area: Natural Language Generation)

**January 2003** Reviewer for the Alexander von Humboldt Foundation (reviewing a proposal for a Georg Forster-Research grant).

**November 2002** Project proposal reviewer for the Engineering and Physical Sciences Research Council, United Kingdom in the area of Natural Language Generation (area: Induction of Natural Language grammars).

**January 2001** Reviewer of the EU-funded IST project LLAVES (area: information extraction, text summarization).

## Other academic and industrial experience

**2011** I received one of the Best Reviewer Awards from the “Conference on Empirical Methods in Natural Language Processing (EMNLP)”

**2011** I am a member of the Research/Industry Chairs of the Third International Conferences on Pervasive Patterns and Applications (PATTERNS 2011).

**19th June — 19th July, 2010** I was the scientific host of Prof. Dr. Adnan Yahya (CS department, University of Birzeit, Palestine) who obtained a research grant from the Alexander von Humboldt Foundation.

**July – Dec. 2003, July – Sept. 2004, July. – Sept. 2005** I was the scientific host of Prof. Dr. Abdelhadi Soudi (University of Morocco, Rabat) who obtained a Georg Forster–Research grant from the Alexander von Humboldt Foundation.

**Since 1994** Several presentations of DFKI’s research activities in the area of intelligent information extraction (e.g., Deutsche Bank, Dresdner Bank, Media Control, Debis, Microsoft USA, German Defense Department, Siemens, SAP AG, CEA Capital Partners Management GmbH, M.I.T, IBM global services).

**August 1995** offer for a post doc. position at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology (MIT). I had to decline.

**1991** Member of a committee whose task was to define the hard and software infrastructure for a new computer centre at the faculty of philosophy of the Saarland University.

**1989** Traveling award from the US Scholarship Program for active participation at IJCAI 1989, Detroit, USA.

## Participation at Scientific Challenges

**Multiword keyword extraction** ACL SigLex event SemEval-2 Evaluation Exercises on Semantic Evaluation, 2010.

**Recognizing Textual Entailment** Active participation in the workshop on Recognizing Textual Entailment RTE-3 (2007, 5th best result) and RTE-4 as part of the Textual Analysis Conference (TAC) (2008, 3rd best result in RTE-4; 2009, 2nd best result in RTE-5; 2010, 4th best result in the RTE-6 Novelty Detection Subtask with new team members); 2011, 5th best result out of 13 participating teams in the RTE-7 Main Task; active participation in PETE 2010 (as part of SemEval-2 2010; 4th best result)

**Cross-Lingual Evaluation Forum** Active participation in track on open-domain cross-lingual question answering (QA at CLEF) tasks 2003 until 2008. We achieved best results for monolingual German and cross-lingual German–English and English–German.

**Cross-Lingual Evaluation Forum** Active participation in the Answer Validation Exercise 2007 and 2008. We achieved best results for English (2007, 2008) and German (2008).

**Cross-Lingual Evaluation Forum** Active participation in the track “Cross-Language Geographic Information Retrieval”, 2008. We achieved best results for monolingual English.

## Programming experience

Professional experience in object-oriented and functional programming.

## Selected developed software packages

- **QALL-ME Framework** A Service-Oriented Framework for entailment-based and ontology-driven question answering. Open source link: <http://qallme.sourceforge.net/>.
- **Quantico**: open-domain QA system based on **BiQue** and **SMES**. Handles questions in English and German and searches for answers in English and German from large text documents and the Web. **We obtained the best results in the main qa@clef task with German and English as target languages, 2005 – 2008.**
- **SMES**: multilingual information extraction core system (English, German). Includes novel components for robust and efficient parsing of free text documents and sophisticated interfaces to domain-specific ontologies, as well as high-performance Named Entity recognizers. Parameterizable XML output interfaces. Rich lexical and grammatical databases. Good run time performance: 2000 words per second. Used in a number of DFKI projects as well as in a number of outside research institutes (mostly in areas like information extraction, Text Mining, Ontology Learning from texts, e.g., European Media Lab Heidelberg, Uni. Dortmund, Karlsruhe, Rostock, Koblenz-Landau, TU-Darmstadt, Humboldt-Uni. Berlin). Developed in Common Lisp.
- **SPPC**: a purely C++-based version of SMES. Focus here was on speed and integration into industrial platforms. Less powerful linguistic functionality than SMES, but much faster: 12.000 words per second.
- **HPSG-DOP**: a system for learning and parsing of stochastic lexicalized tree adjoining grammars (SLTAG). SLTAG are learned from parsing results using a deep NL grammar based on Head-driven Phrase Structure Grammar and from large treebanks such as the Penn Treebank. The Machine Learning approach is similar to explanation-based learning (EBL). The version was developed together with researchers from CSLI Stanford.
- **Morphix**: a module for morphological analysis and generation including online compounding and handling of unknown words. Version for CommonLisp and C++. Very robust and fast (more than 10.000 words in a second on standard PC hardware based on a stem lexicon of 120.000 entries). High number of dissemination to outside universities and institutes (among others CMU, CSLI Stanford, DEC, Siemens, Eurospider, Uni. Karlsruhe, Osnabrück, Zürich). It was also used for morphological analysis in the Verbmobil project, and it is part of KBML, a development tool for NL generation systems (University of Bremen) and the LKB development system for HPSG-like grammars (Stanford University and University of Cambridge, UK).

## Invited talks

(This list contains talks I have held in addition to the talks I had to give as part of accepted conference and workshop papers.)

- “TechWatch - Technologie- und Marktbeobachtung powered by SMILA.”, invited talk at the Theseus conference about “SMILA - Das Framework für semantische Applikationen - Überblick und Einsatz”, June 2011.  
Dr. Gnter Neumann, Deutsches Forschungszentrum für künstliche Intelligenz
- “Hybrid Information Extraction - A Survey of Methods and Techniques”, Conference: 10 Years Verbmobil - Review and Outlook - Multilingual processing of spontaneous speech, Saarbrücken, Germany, November, 2010.

- “Textual Entailment - Methoden und Anwendungen” invited talk, Department of Psychology, University of Würzburg, Germany, January, 2010.
- “Textual Entailment - Methoden und Anwendungen” invited talk, Philosophical Faculty of Heinrich–Heine University of Düsseldorf, October, 2009.
- “Data-oriented Parsing with Lexicalized Tree Insertion Grammars” invited talk, Nancy NLP seminar, CNRS, Nancy, France, March, 2009.
- “Web based Multilingual Question Answering” invited talk given during the computer science colloquium of the Technical University of Darmstadt, Germany, November, 2008.
- “Ontology-based Information Extraction and Question Answering — Coming Together” invited talk presented at the 1st Workshop on Ontology-based Information Extraction (OBIS–2008), Kaiserslautern, Germany, September, 2008.
- “Best Research Practice” invited talk given at the Doctoral Symposium of the NLDB-08, London, UK, June, 2008.
- “The Multilingual and Cross-lingual Web” invited talk presented at the symposium of *Human Language Technologies in Africa: Status and Prospects*, Rabat, Morocco, June, 2008
- “Towards Dynamic Interactive Information Extraction” invited talk at the 2nd workshop of the Ontoverse project, Heinrich–Heine University of Düsseldorf, November, 2007.
- “DFKI at Clef 2007”, invited talk at the Clef workshop, Budapest, Hungary, September, 2007.
- “Interactive Systems and New Interface Technology: A Language Technology Perspective” short oral presentation of a white paper I was invited to prepare for the workshop “Interactive Systems and New Interface Technology” organized by the Interfaces Unit of the Directorate General Information Information Society and Media, EC, Luxembourg, January, 2006
- “Open-domain Cross-lingual Question Answering from Unstructured Documents” talk given during the LCT Colloquium as part of the European Masters Program in Language and Communication Technologies at the Free University of Bozen, June, 2005.
- “Open-domain Cross-lingual Question Answering from Unstructured Documents” talk given during the OEFAI Colloquium, Vienna, June, 2005.
- “Open-domain Cross-lingual Question Answering from Unstructured Documents” talk given during the Computer Science Colloquium of the University of Klagenfurt, June, 2005.
- “Interleaving Parsing and Generation through Uniform Processing” talk given during the AI Lab Colloquium, MIT Artificial Intelligence Laboratory, Monday, July, 2002.
- “Language Technology and Intelligent Information Extraction for Structured Web–Search Engines”, talk given during Computer Science Colloquium, TU Darmstadt, Monday, June, 2002.
- “Language Technology and Intelligent Information Extraction” talk given during the Information Retrieval Colloquium, Department of Computer Science, University Dortmund, Prof. Norbert Fuhr, May, 2002.
- “Language Technology and Intelligent Information Extraction for Structured Web–Search Engines”, talk given during *Fachtagung: E-BUSINESS: TECHNOLOGIEN UND ANWENDUNGEN*, Umwelt-Campus Birkenfeld, April, 2002.

- “Information extraction from real-world German text”, talk given during the Computational Linguistics Colloquium, joined seminar organized by the department of Computational Linguistics and Phonetics and the European Post-Graduate College in Language Technology and Cognitive Systems, Saarbrücken, April, 2001.
- “Towards parameterizable domain-adaptive information extraction”, invited talk at the international workshop on innovative language technology and Chinese information processing, ILT&CIP-2001, Shanghai Jiao Tong University, April, 2001.
- “PARADIME – Parametrizable Domain-Adaptive Information and Message Extraction. Adapting the SMES System to the Soccer Domain”, University of Karlsruhe, February, 2000.
- “Learning Stochastic Lexicalized Tree Grammars from HPSG”, Dagstuhl-Seminar on “Efficient Language Processing with High-level Grammar Formalisms”, October, 1999.
- “Intelligent Information Extraction - Technology and Applications”, University of Osnabrück, June, 1999.
- “Intelligent Information Extraction - Technology and Applications”, University of Koblenz/Landau, June, 1999.
- “Learning Stochastic Lexicalized Tree Grammars from HPSG”, German Natl. Research Center for Information Technology (GMD), St. Augustin, January, 1999.
- “Text Mining und Informationsextraktion im Bankenbereich”, Dresdner Bank, Frankfurt, May, 1999.
- “Intelligente Methoden der Informationsextraktion aus freien Textdokumenten”, talk given during the “Informatikforum 98”, a forum for knowledge transfer from research to industry, Saarland University, October, 1998.
- “Data-driven extraction of Lexicalized Tree Substitution Grammars from HPSG”, University of Groningen, Netherlands, November, 1997.
- “Application of Explanation-based Learning to Natural Language Generation”, CSLI, Stanford, August, 1997.
- “Application of Explanation-based Learning to Efficient Processing of Constraint-based Grammars”, CSLI, Stanford, Juli, 1996.
- “Ein uniformes Verarbeitungsmodell für Parsing und Generierung natürlicher Sprache”, University of Koblenz/Landau, June, 1996.
- “Ein uniformes Verarbeitungsmodell für Parsing und Generierung natürlicher Sprache”, University of Tübingen, November, 1994.
- “Reversible Grammatiken und Natürlich-sprachliche Systeme”, *invited talk given at the “35sten Tagung experimentell arbeitender Psychologen, TEAP-35”*, Münster, Germany, 1993.
- “BiLD: Bidirectional Linguistic Deduction”, *talk given at the “1st International Workshop on Constraint-based Generation”*, Bad Teinach, Germany, November 1990.
- “Morphix: Ein hochportabler Lemmatisierungsmodul für das Deutsche.”, AEG, Konstanz, Mai, 1987.

## Publications<sup>1</sup>

### PhD Thesis and Habilitation

1. **G. Neumann** A Uniform Computational Model for Natural Parsing and Generation. In Saarbrücken Dissertations in Computational Linguistics and Language Technology, Volume 1, 1998, 1-203, ISBN 3-933218-00-4.
2. **G. Neumann** Domänenadaptive Kerntechnologien für die Analyse und Synthese von Natürlicher Sprache. Habilitation in Computational Linguistics, University of the Saarland, July, 2004 (German survey paper - English title is "Domain-adaptive Core Technology for the Analysis and Synthesis of Natural Language", and 8 English papers, all of them listed below).

### Books and edited collections

3. **A. Soudi, A. van den Bosch, and G. Neumann (Edt.)** *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Series: Text, Speech and Language Technology, Vol. 38, 2007, 305 p., Hardcover, ISBN: 978-1-4020-6045-8.
4. **B. Adrian, G. Neumann, A. Trousov, B. Popov (Edt.)** *Proceedings 1st International and KI-08 Workshop on Ontology-based Information Extraction Systems*. CEUR-WS volume 400, Kaiserslautern, Germany, DFKI, 2008. Online at <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-400/proceedings.pdf>.

### Journal articles

5. **O. Ferrandez, C. Spurk, M. Kouylekov, I. Dornescu, S. Ferrandez, M. Negri, R. Izquierdo, D. Tomas, C. Orasan, G. Neumann, B. Magnini and J. Luis Vicedo** The QALL-ME Framework: A specifiable-domain multilingual Question Answering architecture. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, online version, doi:10.1016/j.websem.2011.01.002.
6. **A. Figueroa, G. Neumann, and J. Atkinson** Searching for Definition Answers on the Web using Surface Patterns. *IEEE Computer* volume 42 number 4, Pages 68–76, IEEE, 4/2009.
7. **G. Neumann** A Computational Linguistics Perspective on the Anticipatory Drive. Commentary on the target article by Martin V. Butz. In *Constructivist Foundations*, November 2008, Vol. 4., No. 1, 26-28.
8. **A. Figueroa and G. Neumann** Genetic Algorithms for data-driven Web Question Answering. *Evolutionary Computation Journal*, Spring 2008, Vol. 16, No. 1: 127–147.
9. **G. Neumann** Strategien zur Webbasierten Multilingualen Fragenbeantwortung – Wie Suchmaschinen zu Antwortmaschinen werden. In *Journal "Computer Science - Research and Development"*, Volume 22, Number 2 / February 2008, pages 71–84.
10. **G. Neumann and F. Xu** Mining Natural Language Answers from the Web. In *International Journal Web Intelligence and Agent Systems*, Volume 2. Number 2, 2004.
11. **G. Neumann and J. Piskorski** A Shallow Text Processing Engine. In *International Journal Computational Intelligence*, Volume 18, Number 3, August 2002, 451–476.
12. **G. Neumann and S. Schmeier** Shallow Natural Language Technology and Text Mining. In *KI: German Journal on Artificial Intelligence*, Special Issue on Text Mining, T. Joachims, E. Leopold (Eds.), ISSN: 0933-1875, 16 (2002) 2.

---

<sup>1</sup>The papers can be retrieved from: <http://www.dfki.de/~neumann/publications.html>

13. **M. Klettke, G. Neumann, M. Bietz, I. Bruder, A. Heuer, D. Priebe, M. Becker, J. Bedersdorfer, H. Uszkoreit, A. Maedche, S. Staab and R. Studer** GETESS - Ontologien, objektrelationale Datenbanken und Textanalyse als Bausteine einer semantischen Suchmaschine. In *Journal Datenbank-Spektrum* 1, 2001, 14–24.
14. **G. Neumann** Interleaving Natural Language Parsing and Generation Through Uniform Processing. In *Journal of Artificial Intelligence* 99, 1998, 121-163.
15. **G. Neumann** Grammatikformalismen in der Generierung und ihre Verarbeitung. In *KI: German Journal of Artificial Intelligence, Special Issue on Natural Language Generation*, W. Hoepfner (Ed.), 1993, 22–30.

#### Book articles

16. **G. Neumann and B. Crysmann** *Extracting Supertags from HPSG-based Tree Banks*. In S. Bangalore and A. Joshi (eds): *Supertagging - Using Complex Lexical Descriptions in Natural Language Processing*, Pages 313-335, MIT Press, Cambridge, Massachusetts, 2010.
17. **G. Neumann** *Text-basiertes Informationsmanagement*. In Carstensen, K.-U.; C. Ebert; C. Endriss; S. Jekat; R. Klabunde & H. Langer (Hrsg.) *Computerlinguistik und Sprachtechnologie. Eine Einführung*. 3., überarbeitete und erweiterte Auflage, Heidelberg: Spektrum Akademischer Verlag, Pages 576-615, 2009.
18. **Soudi, A., Neumann. G., and A. Van den Bosch** *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. In A. Soudi, A. van den Bosch, and G. Neumann (eds.), *Arabic Computational Morphology: Knowledge-based and Empirical Methods*, Springer, 2007, pages 1-14.
19. **B. Sacaleanu and G. Neumann** *A Cross-Lingual German-English Framework for Open-Domain Question Answering*. In *Evaluation of Multilingual and Multi-modal Information Retrieval - 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006*, Alicante, Spain, September 20-22, 2006, Revised Selected Papers; Peters, C., Clough, P., Gey, F.C., Karlgren, J., Magnini, B., Oard, D.W., de Rijke, M., Stempfhuber, M. (eds.), Vol. 4730, 2007, ISBN 978-3-540-74998-1, Softcover.
20. **A. Figueroa and G. Neumann** *Language Independent Answer Prediction from the Web*. In *Advances in Natural Language Processing*, Springer series "Lecture Notes in Computer Science", pages 423-434, Springer-Verlag Berlin, Heidelberg, New-York, 2006.
21. **G. Neumann** *A Hybrid Machine Learning Approach for Information Extraction from Free Texts*. In Spiliopoulou et al. (eds.). *From Data and Information Analysis to Knowledge Engineering*. Springer series "Studies in Classification, Data Analysis, and Knowledge Organization", pages 390-397, Springer-Verlag Berlin, Heidelberg, New-York, 2006.
22. **G. Neumann and B. Sacaleanu** *Experiments on Cross-Linguality and Question-type driven Strategy Selection for Open-Domain Question Answering*. In: C. Peters et al. (eds.): *Clef 2005*, LNCS 4022, Springer-Verlag Berlin Heidelberg, pp. 429-438, 2006.
23. **G. Neumann and B. Sacaleanu** *Experiments on Robust NL Question Interpretation and Multi-layered Document Annotation for a Cross-Language Question/Answering System*. In: C. Peters et al. (eds): *Clef 2004*, LNCS 3491, pp. 411-422, 2005, Springer Berlin Heidelberg.
24. **G. Neumann** *Informationsextraktion*. In Carstensen, K.-U.; C. Ebert; C. Endriss; S. Jekat; R. Klabunde & H. Langer (Hrsg.) (2004) *Computerlinguistik und Sprachtechnologie. Eine Einführung*. 2., bearbeitete und erweiterte Auflage. Heidelberg: Elsevier.
25. **G. Neumann** *A Data-driven Approach to Head-driven Phrase-Structure Grammar*. In R. Bod, R. Scha and K. Sima'an (eds.) *Data-Oriented Parsing*, CSLI Publications, 2003.

26. **G. Neumann** *A Uniform Method for Automatically Extracting Stochastic Lexicalized Tree Grammars from Treebanks and HPSG*. In A. Abeille (ed.) Building and using syntactically annotated corpora, Kluwer, 2003.
27. **A. Maedche, G. Neumann and S. Staab** *Bootstrapping an Ontology-based Information Extraction System*. In Szczepaniak, Piotr S.; Segovia, Javier; Kacprzyk, Janusz; Zadeh, Lotfi A. (eds.), Intelligent Exploration of the Web, Springer, ISBN 3-7908-1529-2, 2002, pages 345–360.
28. **G. Neumann** *Programming Languages in Artificial Intelligence*. In Bidgoli (ed) Encyclopedia of Information Systems, Academic Press, San Diego, CA, ISBN 0-12-227240-4, 2002, pages 31-45.
29. **G. Neumann** *Informationsextraktion*. Carstensen, K.-U.; C. Ebert; C. Endriss; S. Jekat; R. Klabunde & H. Langer (Hrsg.) (2001) Computerlinguistik und Sprachtechnologie. Eine Einführung. Heidelberg: Spektrum.
30. **G. Erbach, G. Neumann and H. Uszkoreit** *MULINEX — Multilingual Indexing, Editing and Navigation Extensions for the World Wide Web*. In: David Hull and Doug Oard (eds.) Cross-Language Text and Speech Retrieval – Papers from the 1997 AAAI Spring Symposium, AAAI Press, Menlo Park, 1997.
31. **G. van Noord and G. Neumann** *Syntactic Generation*. In: “Survey of the State of the Art in Human Language Technology” edited by R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen and V. Zue. Cambridge University Press and Giardini, 1997.
32. **G. Neumann and G. van Noord** *Reversibility and Self-Monitoring in Natural Language Generation*. In: Strzalkowski (ed.) Reversible Grammars in Natural Language Processing, Kluwer, 1994, 59-96.

#### Reviewed conference articles

33. **Günter Neumann and Sven Schmeier** (2011) *A Mobile Touchable Application for Online Topic Graph Extraction and Exploration of Web Content*. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.
34. **Alexander Volokh and Günter Neumann** (2011) *Automatic Detection and Correction of Errors in Dependency Treebanks*. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011.
35. **Alexander Volokh and Günter Neumann** (2011) *Using MT-Based Metrics for RTE.*, The Fourth Text Analysis Conference, Gaithersburg, MD, USA, 2011.
36. **Alexander Volokh, Günter Neumann and Bogdan Sacaleanu** (2010) *Combining Deterministic Dependency Parsing and Linear Classification for Robust RTE*. Third Text Analysis Conference, Gaithersburg, MD, USA.
37. **Eichler, K. and Neumann, G.** *Bootstrapping Noun Groups Using Closed-Class Elements Only*. In KDML 2010: Knowledge Discovery, Data Mining, and Machine Learning, Kassel, Germany, 2010.
38. **Eichler, K. and Neumann, G.** *DFKI KeyWE: Ranking keyphrases extracted from scientific articles*. In K. Erk, C. Strapparava (eds.): ACL SigLex event SemEval-2 Evaluation Exercises on Semantic Evaluation, Uppsala, Sweden, 2010.
39. **Volokh, A. and Neumann, G.** *Comparing the Benefit of Different Dependency Parsers for Textual Entailment Using Syntactic Constraints Only*. In K. Erk, C. Strapparava (eds.): ACL SigLex event SemEval-2 Evaluation Exercises on Semantic Evaluation, Uppsala, Sweden, 2010.

40. **Eichler, K., Hensen, H., Neumann, G., Reithinger, N., Schmeier, S., Schumacher, K. and Seifert, S.** *DiLiA - The Digital Library Assistant*. In M. Lalmas J. Jose, A. Rauber, F. Sebastiani, I. Frommholz (eds.): Proceedings of the European Conference on Research and Advanced Technology for Digital Libraries (ECDL) 2010, Glasgow, United Kingdom, Springer.
41. **I. Seifert, K. Eichler, H. Hensen, S. Schmeier, M. Kruppa, N. Reithinger, and G. Neumann** *DILIA — A DIGITAL LIBRARY ASSISTANT: A new approach to information discovery through information extraction and visualization* In Proceedings of IC3K - The International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Madeira, October, 2009.
42. **R. Wang and G. Neumann** *An Accuracy-Oriented Divide-and-Conquer Strategy for Recognizing Textual Entailment* In Text Analysis Conference (TAC 2008), National Institute of Standards and Technology (NIST).
43. **A. Volokh and G. Neumann** *A Puristic Approach for Joint Dependency Parsing and Semantic Role Labeling*. In Proceedings of the CoNLL-2008 shared task, 2008, Manchester.
44. **K. Eichler, H. Hensen, M. Löckelt, G. Neumann, and N. Reithinger** *Interactive Dynamic Information Extraction*. In Proceedings of KI'2008, Kaiserslautern, Sept. 2008.
45. **A. Figueroa and G. Neumann** *Finding Distinct Answers in Web Snippets*. In Proceedings of the 4th International Conference on Web Information Systems and Technologies, pages 26-33, 2008.
46. **E. Cabrio, M. Kouylekov, B. Magnini, M. Negri, L. Hasler, C. Orasan, D. Tomas, J. L. Vicedo, G. Neumann, and C. Weber** *The QALL-ME Benchmark: a Multilingual Resource of Annotated Spoken Requests for Question Answering*. The 6th Edition of Language Resources and Evaluation Conference, LREC 2008, Marrakech, Morocco, 2008.
47. **K. Eichler, H. Hensen, and G. Neumann** *Unsupervised Relation Extraction From Web Documents*. The 6th Edition of Language Resources and Evaluation Conference, LREC 2008, Marrakech, Morocco, 2008.
48. **A. Figueroa and G. Neumann** *A multilingual framework for searching definitions on web snippets*. 30th Annual German Conference on Artificial Intelligence (KI-2007), Osnabrück, Germany, 2007.
49. **R. Wang and G. Neumann** *Recognizing Textual Entailment Using a Subsequence Kernel Method*. In Proceedings of AAAI-2007 conference, Vancouver, 2007.
50. **G. Neumann and B. Crysmann** *Exploring HPSG-based Treebanks for Probabilistic Parsing*. In Proceedings of 5th International Conference on Language Resources and Evaluation, Genoa, Italy, 2006.
51. **G. Neumann, F. Xu and B. Sacaleanu** *Strategies for Web-based Cross-Language Question Answering*. In proceedings of 2nd CoLogNET-ElsNET Symposium on *Questions and Answers: Theoretical and Applied Perspectives*, Amsterdam, 18 December 2003.
52. **G. Neumann and F. Xu** *Mining Answers in German Web Pages*. In proceedings of The International Conference on Web Intelligence (WI 2003), Halifax, Canada, Oct. 2003.
53. **G. Neumann, Dan Flickinger** *HPSG-DOP: data-oriented parsing with HPSG*. In proceedings of the 9th International Conference on HPSG (HPSG-2002), Kyung Hee University in Seoul, South Korea, August 2002.

54. **G. Neumann and M. Kappes** *A simple base-line text-categorizer for evaluating the effect of feature extraction in text mining applications.* In: Program and Abstract booklet accompanying the 26th Annual Conference of the Gesellschaft für Klassifikation (GfKl) July 22-24, 2002, University of Mannheim, Germany.
55. **B. Crysmann, A. Frank, B. Kiefer, H. Krieger, S. Müller, G. Neumann, J. Piskorski, U. Schäfer, M. Siegel, H. Uszkoreit, and F. Xu** *An Integrated Architecture for Shallow and Deep Processing.* In proceedings of Association for Computational Linguistics 40th Anniversary Meeting (ACL-02), University of Pennsylvania, Philadelphia, July 2002.
56. **G. Neumann, U. Schäfer** *WHITEBOARD – Eine XML-basierte Architektur für die Analyse natürlichsprachlicher Texte.* In Proceedings of 25th European Congress Fair for Technical Communication, Düsseldorf, Germany, 2002, pp. 635.01–635.12.
57. **T. Declerck and G. Neumann** *A Distributed Shallow Approach to Reference Resolution in the Context of Information Extraction.* In proceedings of EuroConference on Recent Advances in NLP, RANLP-2001.
58. **A. Düsterhöft, G. Neumann, M. Becker, J. Bedersdorfer, I. Bruder** *GETESS: Constructing a Linguistic Search Index for an Internet Search Engine.* In Proceedings of the 5th International Conference on Applications of Natural Language to Information Systems (NLDB'2000), Versailles (France), June, 2000.
59. **G. Neumann, C. Braun and J. Piskorski** *A Divide-and-Conquer Strategy for Shallow Parsing of German Free Texts.* In the Proceedings of the 6th International Conference on Applied Natural Language Processing (ANLP), Seattle, Washington, 2000, pages 239–247.
60. **J. Piskorski and G. Neumann** *An Intelligent Text Extraction and Navigation System.* In the 6th International Conference on Computer-Assisted Information Retrieval (RIA0-2000), Paris, April 2000, 18 pages.
61. **S. Staab, C. Braun, I. Bruder, A. Düsterhof, A. Heuer, M. Klettke, G. Neumann, B. Prager, J. Pretzel, H.-P. Schnurr, R. Studer, H. Uszkoreit, B. Wrengers** *A System for Facilitating and Enhancing Web Search.* In Proceedings of Engineering Applications of Bio-Inspired Artificial Neural Networks, International Work-Conference on Artificial and Natural Neural Networks, IWANN '99, Lecture Notes in Computer Science, Vol. 1607, Springer, 1999, pages 706-714.
62. **T. Declerck, J. Klein, G. Neumann** *Evaluation of the NLP Components of an Information Extraction System for German.* In Proceedings of the first international Conference on Language Resources and Evaluation (LREC), Granada, 1998, 293–297.
63. **G. Neumann** *Applying Explanation-based Learning to Control and Speeding-up Natural Language Generation.* In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics ACL, Madrid, 1997, 214–221.
64. **G. Neumann, R. Backofen, J. Baur, M. Becker, C. Braun** *An Information Extraction Core System for Real World German Text Processing.* In Proceedings of 5th International Conference on Applied Natural Language Processing, Washington, March, 1997, 208-215.
65. **G. Neumann** *Methoden zur intelligenten Informationsextraktion im Internet.* In Proceedings of 20th European Congress Fair for Technical Communication, ONLINE '97, Hamburg, 1997.
66. **G. Neumann** *Application of Explanation-based Learning for Efficient Processing of Constraint-based Grammars.* Proceedings of the Tenth Conference on Artificial Intelligence for Applications, San Antonio, Texas, March, 1994, 208–214.

67. **H. Uszkoreit, R. Backofen, S. Busemann, A. K. Diagne, E. A. Hinkelman, W. Kasper, B. Kiefer, H. Krieger, K. Netter, G. Neumann, S. Oepen, S. Spackman** *DISCO - An HPSG-based NLP System and its Application for Appointment Scheduling*. In: Proceedings of the 15th International Conference on Computational Linguistics, Kyoto, Japan, 1994.
68. **G. Neumann, G. van Noord** *Self-Monitoring with Reversible Grammars*. In: Proceedings of the 14th International Conference on Computational Linguistics, Nantes, 1992, 700–706.
69. **G. Neumann** *A Bidirectional Model for Natural Language Processing*. In: Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics, Berlin, 1991, 245-250.
70. **G. Neumann, W. Finkler** *A Head-Driven Approach to Incremental and Parallel Generation of Syntactic Structures*. In Proceedings of the 14th International Conference on Computational Linguistics, Helsinki, 1990, 288–293.
71. **W. Finkler, G. Neumann** *POPEL-HOW: A Distributed Parallel Model for Incremental Natural Language Production with Feedback*. In: Proceedings of the 11th International Joint Conference on Artificial Intelligence, Detroit, Michigan USA, 1989, Volume 2, 1518-1523.
72. **W. Finkler, G. Neuman** *MORPHIX. A Fast Realization of a Classification-Based Approach to Morphology*. In: Trost, H. (ed.): 4. Österreichische Artificial-Intelligence-Tagung. Wiener Workshop - Wissensbasierte Sprachverarbeitung. Proceedings. Berlin etc.: Springer, 1988, 11-19.

#### Reviewed workshop articles

73. **K. Eichler, H. Hemsén, and G. Neumann** *Unsupervised and domain-independent extraction of technical terms from scientific articles in digital libraries* In Proceedings of the Workshop “Information Retrieval 2009” organized as part of LWA, Darmstadt, 2009.
74. **K. Eichler, H. Hemsén, G. Neumann and S. Schmeier** *A Hybrid Approach for Information Extraction from Scientific Documents in Digital Libraries*. Proceedings of the First Natural Language Processing for Digital Libraries (NLP4DL) Workshop, in Viareggio, Italy (15 June 2009).
75. **R. Wang and G. Neumann** *Relation Validation via Textual Entailment*. In Proceedings of ONTOLOGY-BASED INFORMATION EXTRACTION SYSTEMS (OBIES 2008), KI-2008, Sept. 2008.
76. **B. Sacaleanu, G. Neumann and C. Spurk** *DFKI at QA@CLEF 2008* In Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark, 2008.
77. **R. Wang and G. Neumann** *Information Synthesis for Answer Validation*. In Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark, 2008.
78. **R. Wang and G. Neumann** *Ontology-based Query Construction for GeoCLEF* In Working Notes for the CLEF 2008 Workshop, 17-19 September, Aarhus, Denmark, 2008.
79. **A. Figueroa and G. Neumann** *Mining Web Snippets to Answer List Questions*. In Proceedings of the 2nd International Workshop on Integrating AI and Data Mining (AIDM 2007), December, 2007.
80. **B. Sacaleanu, G. Neumann and C. Spurk** *DFKI-LT at QA@CLEF 2007*. In online proceedings of CLEF 2007 Working Notes, Budapest, September, 2007, ISBN: 2-912335-31-0.

81. **R. Wang and G. Neumann** *DFKI-LT at AVE 2007: Using Recognizing Textual Entailment for Answer Validation*. In online proceedings of CLEF 2007 Working Notes, Budapest, September, 2007, ISBN: 2-912335-31-0.
82. **G. Neumann and R. Wang** *DFKI-LT at QAST 2007: Adapting QA Components to Mine Answers in Speech Transcripts*. In online proceedings of CLEF 2007 Working Notes, Budapest, September, 2007, ISBN: 2-912335-31-0.
83. **R. Wang and G. Neumann** *Recognizing Textual Entailment Using Sentence Similarity based on Dependency Tree Skeletons*. Workshop proceedings of the RTE-3 challenge, 2007, Association for Computational Linguistics.
84. **A. Heyl and G. Neumann** *An Information Extraction based Approach to People Disambiguation*. Articles, J., Gonzalo, J. and Sekine, S. (2007). The SemEval-2007 WePS Evaluation: Establishing a benchmark for the Web People Search Task. In Proceedings of Semeval 2007, Association for Computational Linguistics.
85. **A. Figueroa and G. Neumann** *Identifying Protein-Protein interactions in Biomedical publications*. Workshop Proceedings of the BioCreAtIvE Challenge - Critical Assessment of Information Extraction in Molecular Biology, Madrid, April, 2007.
86. **A. Ankolekar, P. Buitelaar, P. Cimiano, P. Hitzler, M. Kiesel, M. Krötzsch, H. Lewen, G. Neumann, M. Sintek, T. Tserendorj, R. Studer** *SmartWeb: Mobile Access to the Semantic Web*. In: Proc. of the Demo Session at the International Semantic Web Conference, Athens GA, USA, Nov. 2006.
87. **B. Sacaleanu and G. Neumann** *DFKI-LT at the CLEF 2006 Multiple Language Question Answering Track*. In Working Notes for the CLEF 2006 Workshop, August, Alicante, Spain, 2006.
88. **B. Sacaleanu and G. Neumann** *Cross-Cutting Aspects of Cross-Language Question Answering Systems*. In Proceedings of the EACL workshop on Multilingual Question Answering - MLQA'06, Trento, Italy, 2006.
89. **G. Neumann and B. Sacaleanu** *DFKI's LT-lab at the CLEF 2005 Multiple Language Question Answering Track*. In the proceedings of the CLEF 2005 working notes of the QAatCLEF, Vienna, September, 2005.
90. **G. Neumann and B. Sacaleanu** *Experiments on Robust NL Question Interpretation and Multi-layered Document Annotation for a Cross-Language Question/Answering System*. In the proceedings of the CLEF 2004 working notes of the QAatCLEF, Bath, August, 2004.
91. **G. Neumann and B. Sacaleanu** *A Cross-Language Question/Answering-System for German and English*. In the proceedings of the CLEF 2003 working notes of the QAatCLEF, Trondheim, August, 2003.
92. **G. Neumann and T. Declerck** *Domain Adaptive Information Extraction*. In Proceedings of International Workshop on Innovative Language Technology & Chinese Information Processing, ILT&CIP Workshop 2001, Shanghai.
93. **M. Siegel, F. Xu and G. Neumann** *Customizing GermaNet for the Use in Deep Linguistic Processing*. In North American Chapter of ACL 2001, workshop for *WordNet and Other Lexical Resources: Applications, Extensions and Customizations* organized by Dan Moldovan, Sanda Harabagiu, Wim Peters, Mark Stevenson, and Yorick Wilks, Pittsburgh, 2001.
94. **T. Declerck and G. Neumann** *Using a parameterisable and domain-adaptive information extraction system for annotating large-scale corpora?* In Proceedings of the LREC'2000 Workshop Information Extraction meets Corpus Linguistics, Athen 2000.

95. **G. Neumann and S. Schmeier** *Combining Shallow Text Processing and Machine Learning in Real World Applications*. In Proceedings of the IJCAI-99 workshop on Machine Learning for Information Filtering, Stockholm, Sweden, 1999, 55–60.
96. **S. Staab, C. Braun, A. Düsterhof, A. Heuer, M. Klettke, G. Neumann, B. Prager, J. Pretzel, H.-P. Schnurr, R. Studer, H. Uszkoreit, B. Wrengers** *Getess - searching the Web exploiting German texts*. In CIA 99 - Proceedings of the 3th workshop on Cooperative Information Agents, LNCS, Berlin, Springer, 1999, 12 pages.
97. **G. Neumann** *Automatic Extraction of Stochastic Lexicalized Tree Grammars from Tree-banks*. In Proceedings of the 4th workshop on tree-adjointing grammars and related frameworks, Philadelphia, PA, USA, August, 1998, 17–23.
98. **J. Klein, T. Declerck, G. Neumann** *Evaluation of the Syntactic Analysis Component of an Information Extraction System for German*. In Proceedings of the Workshop on the Evaluation of Parsing Systems (LREC 1998), Granada, 1998.
99. **R. Dale, W. Finkler, R. Kittredge, N. Lenke, G. Neumann, C. Peters, M. Stede** *Report from Working Group 2: Lexicalization and Architecture*. In: W. Hoepfner, H. Horacek, J. Moore (eds.) Dagstuhl-Seminar-Report; 93: Principles of Natural Language Generation, Schloß Dagstuhl, Saarland, Germany, 1994, 30-39.
100. **G. Neumann** *Design Principles of the Disco System*. In: Proceedings of the Twente Workshop on Language Technology, Twente, Netherlands, 1993.
101. **G. Neumann** *Reversibility and Modularity in Natural Language Generation*. In: Proceedings of the ACL Workshop on Reversible Grammar in Natural Language Processing, Berkeley, CA, 1991, 31-39.

#### Selected technical reports

102. **R. Schäfer, G. Neumann and W. Wahlster** *TMB - Text Mining und Informationsextraktion im Bankenbereich*. Confidential final report, Dresdner Bank, 1999, 1-143.
103. **G. Erbach, G. Neumann, S. Schmeier and H. Uszkoreit** *Textkategorisierung und Informationsextraktion für Nachrichtmeldungen aus dem Bereich Telekommunikation*. Confidential final report, Deutsche Telekom, 1999, 1-90.
104. **G. Mazzini and G. Neumann** *Domain-Adaptive Information Extraction*. Technical report, DFKI, Saarbrücken, 1999.
105. **G. Neumann** *An On-line Learning Method to Speed-up Natural Language Processing*. Technical report, DFKI, Saarbrücken, 1997.
106. **B. Kiefer, K. Netter, G. Neumann and H. Uszkoreit** *Paradice – Parametrizable Discourse Core Engine*, Final Project Report, DFKI, Saarbrücken, 1997.
107. **G. Neumann** *A Uniform Computational Model for Natural Language Parsing and Generation*. PhD thesis, Saarland University, Saarbrücken, 1994.
108. **S. Busemann, S. Oepen, E. A. Hinkelman, G. Neumann, H. Uszkoreit** *COSMA - Multi-Participant NL Interaction for Appointment Scheduling*. Research report DFKI GmbH, RR-94-34, Germany, 1994.
109. **G. Neumann** *The DISCO development shell and its application in the cosma system*. DFKI Workshop on Natural Language System: Modularity and Re-usability DFKI GmbH, Germany, D-93-03, 1993.

110. **G. Neumann** *POPEL-HOW: Parallele, inkrementelle Generierung natürlichsprachlicher Sätze aus konzeptuellen Einheiten*. Diplomarbeit, FB Informatik, Universität des Saarlandes, Saarbrücken, 1989.
111. **W. Finkler, G. Neumann** *MORPHIX - Ein hochportabler Lemmatisierungsmodul für das Deutsche*. FB Informatik, KI-Labor, Memo Nr. 8, Juli 1986.