Markus Dreyer

CONTACT
INFORMATION

Markus Dreyer
Center for Language and Speech Processing
Johns Hopkins University
Baltimore, MD 21218 USA

Mobile: (443) 366-5100
Fax: (410) 516-5050
E-mail: markus.dreyer@gmail.com
Web: www.clsp.jhu.edu/~markus

AREAS OF
INTEREST

Natural language processing, machine learning, machine translation, finite-state modeling, parsing, computational morphology.

EDUCATION

Johns Hopkins University, Baltimore, Maryland USA

Ph.D. Candidate in Computer Science, Advisor: Jason Eisner.
Expected graduation date: Summer/Fall 2010
M.S., Computer Science, May 2007

University of Heidelberg, Heidelberg, Germany

M.A., Computational Linguistics, 2002

RESEARCH
EXPERIENCE

Johns Hopkins University, Baltimore, Maryland USA
Finite-State Graphical Models over Two or More Strings   **2007 - Present**
(Advisor: Jason Eisner)
- Presented a novel technique for joint prediction of multiple related strings using networks of interacting finite-state machines, a new form of graphical models that are finite-state-based and whose variables range over strings.
- Presented a conditional log-linear model for general string-to-string transductions, such as transliteration, pronunciation or morphology modeling. Showed how latent variables can greatly improve performance in lemmatization and inflectional morphology.
- Used similar techniques to create a fuzzy name-matching module for the JHU-COE entity linking system, submitted to the TAC 2009 Knowledge Base Population competition.

Johns Hopkins University, Baltimore, Maryland USA
Using Prosody to Help Parse Speech (Collaborator: Izhak Shafran)   **Spring 2007**
- Developed a new way to utilize prosodic markers in speech transcripts to help parsing performance. Used prosody as soft constraints on parsing and as initializer for an EM-style learning of split nonterminals.

Johns Hopkins University, Baltimore, Maryland USA
Reordering Constraints for Machine Translation
(Advisors: Keith Hall and Sanjeev Khudanpur)   **Summer 2006**
- Presented an algorithm to extract oracle BLEU scores from translation lattices. Studied existing reordering constraints for machine translation and investigated the effect of allowing long-range reorderings in German.

Johns Hopkins University, Baltimore, Maryland USA
Improving Parsing Performance by Adding Latent Variables
(Advisor: Jason Eisner)   **2005**
- Studied unsupervised methods for learning refinements of the nonterminals in a treebank, from which better parsers can be estimated. We investigated linguistic constraints and constraints on parameter optimization and showed substantial gains in parsing accuracy using the refined nonterminals.

Johns Hopkins University, Baltimore, Maryland USA
Syntactic Models for Machine Translation   **Summer 2005**
- Member of the CLSP Summer Workshop team, integrated a language model into a synchronous parser.

Johns Hopkins University, Baltimore, Maryland USA
Declarative Language for Dynamic Programming (Dyna)   **2004**
- Member of the Dyna team (2004–present); implemented program transformations for the Dyna language.

Heidelberg University, Heidelberg, Germany
Using Syntactic Knowledge for Prosodic Feature Generation   **2001,2002**
- Implemented a transformation-based learning system to annotate written text with prosodic phrases.

INDUSTRY EXPERIENCE

SAP AG, Walldorf, Germany
Software Engineer   **2003**
- Developed testing frameworks for web applications.

IBM Research, Speech Group, Heidelberg, Germany
Research Programmer   **1999–2002**
- Trained grapheme-to-phoneme models for speech synthesis, developed classifier to predict prosodic markers to improve text-to-speech output.

TEACHING EXPERIENCE

Johns Hopkins University, Baltimore, Maryland, USA
Teaching Assistant   **July 2007, 2008, 2009**
- Assisted in the Natural Language Processing Laboratory at the JHU Summer School on Human Language Technology.

University of Heidelberg, Heidelberg, Germany
Teaching Assistant   **2000, 2001**
- Developed and taught one-semester courses on *Programming NLP Parsers* and *Perl Programming.*

PUBLICATIONS

[1] **Markus Dreyer** and Jason Eisner. 2009. Graphical Models Over Multiple Strings. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*

[2] P. McNamee, M. Dredze, A. Gerber, N. Garera, T. Finin, J. Mayfield, C. Piatko, D. Rao, D. Yarowsky, **Markus Dreyer**. 2009. HLTCOE Approaches to Knowledge Base Population. In *Proceedings of the NIST Text Analysis Conference (TAC 2009) Knowledge Base Population Track*

[3] **Markus Dreyer**, Jason Eisner and Jason Smith. 2008. Finite-State Modeling of Log-Linear String Transductions With Latent Variables and Backoff Features. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*

[4] Damianos Karakos, Jason Eisner, Sanjeev Khudanpur and **Markus Dreyer**. 2008. Machine Translation System Combination using ITG-based Alignments. In *Proceedings of the Conference of the Association for Computational Linguistics (ACL).*

[5] **Markus Dreyer** and Izhak Shafran. 2007. Exploiting Prosody for PCFGs with Latent Annotations. In *Proceedings of Interspeech.*

[6] **Markus Dreyer**, Keith Hall, Sanjeev Khudanpur. 2007. Comparing Reordering Constraints for SMT Using Efficient BLEU Oracle Computation. In *Proceedings of the HLT-NAACL Workshop on Syntax and Structure in Statistical Translation (SSST)*.

[7] **Markus Dreyer** and Jason Eisner. 2006. Better Informed Training of Latent Syntactic Features. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[8] **Markus Dreyer**, David A. Smith, Noah A. Smith. 2006. Vine Parsing and Minimum Risk Reranking for Speed and Precision. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*.

[9] A. Burbank, M. Carpuat, S. Clark, **Markus Dreyer**, P. Fox, D. Groves, K. Hall, M. Hearne, I. D. Melamed, Y. Shen, A. Way, B. Wellington, and D. Wu. 2005. Statistical Machine Translation by Parsing. *CLSP Technical Report*.

[10] **Markus Dreyer**. 2002. Combining Hand-Coded and Machine-Learned Chunking Rules. In *Proceedings of the Student Conference on Computational Linguistics (TaCoS)*.

PROFESSIONAL ACTIVITIES AND SERVICE	Program committee member of: - EMNLP 2009, machine translation track (Empirical Methods in Natural Language Processing), - NAACL 2009, machine learning track (North American Association for Computational Linguistics), - ACL 2007, student research workshop (Association for Computational Linguistics); Reviewer for TMI 2006 (Theoretical and Methodological Issues in Machine Translation). Member of Association for Computational Linguistics (ACL). Member of the German Association for Language Technology and Computational Linguistics (GSCL).
INVITED TALKS	**Johns Hopkins University**, Baltimore, Maryland, USA, Cognitive Science Dept. *Graphical Models Over Multiple Strings* **Dec 2009** **University of Saarbrücken** Germany, Dept. for Computational Linguistics and Phonetics *Finite-state Transductions for String Pairs and Multiple Strings* **Dec 2008** **University of Zürich**, Switzerland, Department for Computational Linguistics *Transformation-based Learning of Syntactic-Prosodic Chunks* **Mar 2002**
HONORS/AWARDS	AMTA SSST Scholarship 2007, NSF Post-Workshop Research Award 2005, Wolman Fellowship 2003.
COMPUTER SKILLS	Languages: C/C++, Java, Perl, Python, R, SQL Toolkits: OpenFst, Moses, Joshua, SRILM, Mallet, and others
PERSONAL DETAILS	Date of Birth: January 3, 1977 Citizenship: German Languages: German (native), English (fluent), French (reading), Latin