

Chris Dyer

CONTACT INFORMATION

712 Ivy St.
Pittsburgh, PA 15232
USA

Phone: +1 (240) 328-4439
E-mail: cdyer@cs.cmu.edu
WWW: <http://www.cs.cmu.edu/~cdyer/>

RESEARCH INTERESTS

Statistical machine translation, unsupervised learning, computational morphology and phonology, large-scale data processing, probabilistic models of natural language processing, Bayesian techniques, machine learning.

EDUCATION

University of Maryland, College Park, Maryland, USA

Ph.D., Linguistics, 2010

- Dissertation: “A Formal Model of Ambiguity and its Applications in Machine Translation”
- Advisor: Prof. Philip Resnik

Duke University, Durham, North Carolina USA

A.B., Computer Science, May, 2000

A.B., Music

PROFESSIONAL EXPERIENCE

Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Postdoctoral research associate

August, 2010 – present

Supervisor: Prof. Noah A. Smith

Research areas include machine translation, computational social science, and general natural language processing. Other responsibilities include supervisory duties and teaching.

Google, Mountain View, California, USA

Research intern

May, 2008 – August, 2008

Conducted research on improving word alignment for structurally divergent language pairs and discriminative training in the Machine Translation group.

Amazon.com, Inc., Seattle, Washington USA

Software development engineer

October, 2001 – July, 2005

Responsibilities included design and development of software in the online and back-end order-processing systems; analysis of developer productivity bottlenecks

PUBLICATIONS

P. Simianer, C. Dyer, and S. Riezler. Joint Feature Selection in Distributed Stochastic Learning for Large-Scale Discriminative Training in SMT. Submitted to ACL, 2012.

A. Levenberg, C. Dyer, and P. Blunsom. A Bayesian Model for Learning SCFGs with Discontiguous Rules. Submitted to ACL, 2012.

K.-M. Hermann, C. Dyer, P. Blunsom, and S. Pulman. A Probabilistic Account of Noun-Modifier Semantics. Submitted to EACL, 2012.

D. Yogatama, M. Heilman, B. O’Connor, C. Dyer, B. Routledge, and N. A. Smith. Predicting a Scientific Community’s Response to an Article. In Proceedings of EMNLP, 2011.

D. Chen, C. Dyer, S. B. Cohen, and N. A. Smith. Unsupervised Bilingual POS Tagging with Markov Random Fields. In Proceedings of the First Workshop on Unsupervised Learning in NLP, 2011.

- C. Dyer, K. Gimpel, J. H. Clark, and N. A. Smith. The CMU-ARK German-English Translation System. In Proceedings of the Sixth Workshop on Machine Translation, 2011.
- J. H. Clark, C. Dyer, A. Lavie, and N. A. Smith. Better Hypothesis Testing for Machine Translation: Controlling for Optimizer Instability. In Proceedings of ACL, 2011.
- C. Dyer, J. H. Clark, A. Lavie, and N. A. Smith. Unsupervised Word Alignment with Arbitrary Features. In Proceedings of ACL, 2011.
- A. Arun, B. Haddow, P. Koehn, A. Lopez, C. Dyer, and P. Blunsom. Monte Carlo techniques for phrase-based translation. In *Machine Translation* 24(2):103-121, 2010.
- H. Setiawan, C. Dyer, and P. Resnik. Discriminative Word Alignment with a Function Word Re-ordering Model. In Proceedings of EMNLP, 2010.
- V. Eidelman, C. Dyer, and P. Resnik. The University of Maryland statistical machine translation system for the Fifth Workshop on Machine Translation. In Proceedings of the Association for Computational Linguistics (ACL) Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR, 2010.
- C. Dyer, A. Lopez, J. Ganitkevitch, J. Weese, F. Ture, P. Blunsom, H. Setiawan, V. Eidelman, and P. Resnik. cdec: A Decoder, Alignment, and Learning Framework for Finite-State and Context-Free Translation Models. In Proceedings of ACL, 2010.
- C. Dyer. Two monolingual parses are better than one (synchronous parse). In Proceedings of NAACL, Los Angeles, 2010.
- C. Dyer and P. Resnik. Context free reordering, finite-state translation. In Proceedings of NAACL, Los Angeles, 2010.
- J. Lin and C. Dyer. *Data-intensive text processing with MapReduce*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers. 2010.
- P. Blunsom, T. Cohn, C. Dyer, and M. Osborne. A Gibbs sampler for phrasal synchronous grammar induction. In Proceedings of ACL, 2009.
- S. Kumar, W. Macherey, C. Dyer, and F. Och. Efficient minimum error rate training and minimum Bayes-risk decoding for translation hypergraphs and lattices. In Proceedings of ACL, 2009.
- C. Dyer. Using a maximum entropy model to build segmentation lattices for MT. In Proceedings of NAACL, 2009.
- A. Arun, C. Dyer, B. Haddow, A. Lopez, P. Blunsom, and P. Koehn. Monte Carlo inference and maximization for phrase-based translation. In Proceedings of CoNLL, 2009.
- C. Dyer, H. Setiawan, Y. Marton, and P. Resnik. The University of Maryland Statistical Machine Translation System for the Third Workshop on Machine Translation. In Proceedings of the Fourth Workshop on Statistical Machine Translation (WMT), 2009.
- Z. Li, C. Callison-Burch, C. Dyer, J. Ganitkevitch, S. Khudanpur, L. Schwartz, W. Thornton, J. Weese, and O. Zaidan. 2009. Joshua: Open Source Toolkit for Parsing-based Machine Translation. In Proceedings of the Fourth Workshop on Statistical Machine Translation (WMT), 2009.

- C. Dyer, A. Cordova, A. Mont, and J. Lin. Fast, Cheap, and Easy: Construction of Statistical Machine Translation Models with MapReduce. In Proceedings of the Third Workshop on Statistical Machine Translation (WMT), 2008.
- C. Dyer, S. Muresan, and P. Resnik. Generalizing Word Lattice Translation. In Proceedings of ACL, 2008.
- C. Dyer. The University of Maryland Translation System for IWSLT 2007. In Proceedings of the International Workshop on Spoken Language Translation (IWSLT), 2007.
- C. Dyer. The ‘noiser channel’: Translation from morphologically complex languages. In Proceedings of the Second Workshop on Statistical Machine Translation (WMT), 2007.
- P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin and E. Herbst. Moses: Open Source Toolkit for Statistical Machine Translation. In Proceedings of ACL, 2007.
- A. Omaki, C. Dyer, S. Malhotra, J. Sprouse, J. Lidz, and C. Phillips. The time-course of anaphoric processing and syntactic reconstruction. Presented at CUNY Sentence Processing Conference, 2007.

OTHER TALKS

- C. Dyer. Generate-and-Test Models for Alignment and Machine Translation. Presented at ISI / USC, Marina del Rey, CA. December, 2011.
- C. Dyer. Statistical Translation as Constrained Optimization. Presented at University College, London. December, 2011.
- C. Dyer. Using MapReduce to Learn Language and Translation at Scale. Presented at the University of Texas, Austin. December, 2011.
- C. Dyer. Statistical Translation as Constrained Optimization. Presented at the University of Texas, Austin. December, 2011.
- C. Dyer. Introduction to Statistical Machine Translation. Presented at the 6th MT Marathon, Trento Italy. September, 2011.
- C. Dyer. Unsupervised Undirected Models for Word Alignment and Part of Speech Induction. Presented at the Gatsby Computational Neuroscience Unit, University College London, London, UK. September, 2011.
- C. Dyer. Introduction to Statistical Machine Translation. Presented at Johns Hopkins University Center for Language and Speech Processing Summer School. June, 2011.
- C. Dyer. Integrating Morphology in Probabilistic Translation Models. Presented at the Workshop on Machine Translation in Morphologically Rich Languages. Haifa, Israel, January, 2011.
- C. Dyer. Scaling NLP and MT with MapReduce. Presented at Raytheon BBN Technologies, Cambridge, MA, November, 2010.
- C. Dyer and O. Bojar. Syntax in machine translation. Invited talk at the 4th MT Marathon, Dublin, Ireland, January 2010.
- C. Dyer. Improving machine translation by propagating uncertainty. Presented at the Carnegie Mellon University School of Computer Science, October 2009.

- C. Dyer. Improving machine translation by propagating uncertainty. Presented at the Center for Language and Speec Processing (CLSP) Colloquium at The Johns Hopkins University, September 2009.
- C. Dyer. Improving machine translation by propagating uncertainty. Presented at the IBM TJ Watson Research Center, Yorktown Heights, NY, August 2009.
- J. Lin and C. Dyer. Data-Intensive Text Processing with MapReduce. Tutorial presented at the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT) 2009 Conference.
- C. Dyer. Fast word alignment and phrase-table construction with Hadoop. Presented at Carnegie Mellon University, Pittsburgh, Pennsylvania, April 2008.
- C. Dyer. Decoder guided backoff: using word lattices to improve translation from morphologically complex languages. Presented at the MT Marathon, Edinburgh University, April 2007.
- C. Dyer and P. Resnik. Confusion network decoding in Hiero. Presented by P. Resnik at GALE 2007 PI Meeting, San Fransisco, March 2007.
- C. Dyer and J. Lin. Computing in the Clouds: Applications of MapReduce in “Web-Scale” Information Processing. Presented at the University of Maryland CLIP Colloquium, October 17, 2007.
- C. Dyer. Decoder guided backoff: using word lattices to improve translation from morphologically complex languages. Presented at the MT Marathon, Edinburgh University, April 2007.
- C. Dyer and P. Resnik. Confusion network decoding in Hiero. Presented by P. Resnik at GALE 2007 PI Meeting, San Fransisco, March 2007.
- C. Dyer. The Syntax of Tonal Music, Syntax Lunch Talk, University of Maryland, October 2006.

GRANTS

N. A. Smith and C. Dyer. *Probabilistic Models for Recovering Latent Structure in Natural Language from Text*. NSF TeraGrid grant TG-DBS110003 for 627,600 CPU hours, 2010.

SOFTWARE

`cdec` open source decoder and training platform for statistical machine translation and other structured prediction models. <http://cdec-decoder.org/>

HONORS AND AWARDS

JHU Summer Workshop Participant, Summer, 2010
 JHU Summer Workshop Participant, Summer, 2006
 University Fellowship, University of Maryland, 2005–2006
 Fulbright Scholar, Technical University, Berlin, Germany, 2000–2001
 Phi Beta Kappa, Duke University, 2000
 Louis Sudler Prize in the Arts, Duke University, 2000
 Horace Trumbauer Engineering Scholarship, Duke University, 1997

SERVICE

Grant reviewing: National Science Foundation (NSF; 2010), Netherlands Organisation for Scientific Research (2011)

Journal reviewing: *ACM Transactions on Asian Language Information Processing*, *Journal of Artificial Intelligence Research (JAIR)*, *IEEE Transactions on Audio, Speech and Language Processing*, *IEEE Software*, and *Machine Translation*

Conference organization: EMNLP (2011, Area chair for “Phonology Morphology Tagging Chunking and Segmentation”)

Conference reviewing: AAAI (2011–2012), ACL (2007–2012), EMNLP (2008–2011), EACL (2009,2012), NAACL(2009, 2010, 2012), IJCAI (2011), AMTA (2008, 2009), COLING (2008, 2010), CoNLL (2009,2010), Workshop on Machine Translation (2007–2011)

Departmental service: Linguistics Colloquium Committee, 2006–2007; Co-organizer, Mayfest 2006 Workshop, Department of Linguistics, University of Maryland

LANGUAGES

Native speaker of English, fluent in German, basic proficiency in Modern Standard Arabic.