Julia Hockenmaier

Department of Computer Science,
3324 Siebel Center, MC-258
201 N Goodwin Ave,
Urbana, IL 61801-2302, USA
Phone: +1 (217) 265-6855   Fax: +1 (217) 265-6591
juliahmr@cs.uiuc.edu
http://www.cs.uiuc.edu/~juliahmr

Current Employment

Assistant Professor, Department of Computer Science, University of Illinois at Urbana-Champaign
> since August 2007

Past Employment

Postdoctoral Fellow, Institute for Research in Cognitive Science, University of Pennsylvania
> With Prof. Aravind K. Joshi (University of Pennsylvania)
> Also a Visiting Postdoctoral Scholar with Prof. Ken A. Dill (UC San Francisco)
> *Employed on National Science Foundation Information Technology Research grant*
> *"Language, Learning, and Modeling Biological Sequences"*
> Feb 2003–August 2007

Education

PhD (Informatics), University of Edinburgh
> Thesis: *Data and Models for Statistical Parsing with Combinatory Categorial Grammar*
> Supervisor: Prof. Mark Steedman
> Feb 1999 – Jan 2003 (degree awarded July 2003)
> **Shortlisted for the British Computer Society's Distinguished Dissertation Award 2004**

MSc (Cognitive Science and Natural Language), University of Edinburgh
> Thesis: *Rule-based Word-Segmentation*
> Supervisor: Dr. Chris Brew
> Oct 1996—Sept 1997

Diplom (Computer)linguistik (minor: Computer Science), Universität Stuttgart
> *Degree in computational linguistics*
> Thesis: *Parsing unsegmented Chinese text with a head-lexicalised PCFG*
> Supervisor: Prof. Mats Rooth
> Final grade: 1 (*"sehr gut"*, best out of 1–6)
> Oct 1993—Jan 1999

Chinese language studies, Nanjing University, People's Republic of China
> Sept 1995—Aug 1996

Other research experience

Graduate student researcher at CLSP Summer Workshop
Center for Language and Speech Processing, Johns Hopkins University
Summer 2002
Student research assistant in computational semantics
IMS, Universität Stuttgart
May—Aug 1995, Aug—Sept 1996 and Oct 1997—Sept 1998
Internship, Natural Language Processing Group, Research Centre of Daimler Benz AG
Research Centre of Daimler Benz AG, Ulm, Germany, April 1995

Teaching experience

Classes taught

CS 598: Advanced Natural Language Processing
University of Illinois at Urbana-Champaign, Spring 2009, Spring 2010
CS 498: Introduction to Natural Language Processing
University of Illinois at Urbana-Champaign, Fall 2008, Fall, 2009, Fall 2010
CS 591: Semantics, Knowledge Representation and Language Understanding in Context
University of Illinois at Urbana-Champaign, Fall 2008 *(With Prof. Dan Roth)*
CS 598: Integrative Intelligent Information Systems
University of Illinois at Urbana-Champaign, Spring 2008
With Profs. David Forsyth and ChengXiang Zhai
CS 498: Expressive Grammars for Natural Language Processing
University of Illinois at Urbana-Champaign, Fall 2007

Classes taught as teaching assistant

Data Intensive Linguistics
School of Informatics, University of Edinburgh
Winter 1999, Spring 2001, Spring 2002
Introduction to Computational Linguistics
School of Informatics, University of Edinburgh, Winter 2001

Tutorials and summer schools

"Wide-Coverage NLP with Linguistically Expressive Grammars"
Tutorial at ACL 2010
With Prof. Josef van Genabith (Dublin City University)
and Yusuke Miyao (National Institute of Informatics)
Uppsala, Sweden, July 2010
"Treebank-Based Acquisition of LFG, HPSG and CCG Resources"
European Summer School in Logic, Language and Information
With Prof. Josef van Genabith (Dublin City University) and Yusuke Miyao (U.Tokyo)
Malaga, Spain, Summer 2006.

Student supervision

Graduate students
> Yonatan Bisk, PhD (Computer Science), since Fall 2009
> Micah Hodosh, PhD (Computer Science), since Spring 2009
> Peter Young, PhD (Computer Science), since Spring 2008

Senior theses
> Yanchuan Sim, Spring 2010
> Cyrus Rashtchian, Spring/Summer 2010

Independent studies
> Matt Garley, PhD student (Linguistics), Fall 2010
> Parikshit Sondhi, PhD student (Computer Science), Summer 2010
> Mahmoud Abunasser, PhD student (Linguistics), Fall 2009 (Qualifying paper)
> Ryan Cunningham, PhD student (Computer Science), Spring 2009
> Micah Hodosh, PhD student (Computer Science), Spring 2009
> Oscar Sanchez Plazas, PhD student (Computer Science), Spring 2009
> Cyrus Rashtchian, Freshman (Computer Science), Summer 2008

Thesis committees
> Michael Connor, PhD (Computer Science)
> Hannaneh Hajishirzi, PhD (Computer Science)
> Alexander Sorokin, PhD (Computer Science)
> Matt Garley, PhD (Linguistics)
> Kevin Small, PhD (Computer Science, defended Summer 2009)

Qualifying exam committees (PhD Computer Science)
> Prateek Jindhal, Rajhans Samdhani (Fall 2010)
> Gourab Kundu, Binbin Liao (Spring 2010)
> Dan Goldwasser (Fall 2009)
> Jingjin Yu, Abner Rivera (Spring 2009)
> Vivek Srikumar (Fall 2008)
> Lev Ratinov (Fall 2007)

Grants

NSF IIS medium grant: Understanding the Meaning of Images
> $550,000 for August 2008–July 2011
> PI: David Forsyth, Co-PI: Julia Hockenmaier

Awards

Engineering and Physical Sciences Research Council Quota Award
> 1999—2002

Division of Informatics studentship, University of Edinburgh
> 1999—2002

Fellow of the German National Academic Foundation (Studienstiftung des deutschen Volkes)
> *(Scholarship awarded to top 0.5% of undergraduates across all disciplines in Germany)*
> 1995—1999

Fellow of the China programme, Alfried Krupp von Bohlen und Halbach Stiftung
　　　　1995—1996

Membership in professional societies

Association for Computational Linguistics (ACL)

Professional service

Editorial Board:
　　　　Revue Traitement Automatique des Langues (TAL);
　　　　(Journal of the Association pour le Traitement Automatique des Langues (ATALA),
　　　　the French association for natural language processing)
Conference and Workshop Organization:
　　　　Faculty advidsor, Student Research Workshop, NAACL-HLT 2010
　　　　(with Prof. Diane Litman, U. Pittsburgh);
　　　　Organizer for ACL 2007 Workshop on Deep Linguistic Processing, Prague
　　　　Senior program committee (co-chair for syntax/parsing area), HLT-NAACL 2007
　　　　(Joint Human Language Technology Conference / North American Chapter of the ACL meeting);
　　　　Co-program chair, Student Research Workshop, EACL '03;
Reviewing (Journals), regular:
　　　　Computational Linguistics;
Reviewing (Journals), ad-hoc:
　　　　Journal of Logic, Language and Information
　　　　Proceedings of the National Academy of Sciences (PNAS)
　　　　Journal of Computer Speech and Language
　　　　Journal of Natural Language Engineering
　　　　Language Resources and Evaluation
　　　　IEEE Intelligent Systems
　　　　Entropy
Reviewing (Conferences):
　　　　ACL 2004, 2005, 2007, 2008 and 2010*(Annual Meeting of the Association for Computational Linguistics)*
　　　　ACL/IJCNLP 2009 *(47th ACL and 4th International Joint Conference on NLP of the Asian Federation of NLP)*
　　　　ACL/COLING 2006 *(Joint Meeting of the ACL and International Conference on Computational Linguistics)*
　　　　COLING 2004, 2008, 2010 *(International Conference on Computational Linguistics)*
　　　　EMNLP 2003, 2006 and 2008*(Conference on Empirical Methods in Natural Language Processing)*
　　　　EMNLP-CoNLL 2007 *(Joint EMNLP and CoNLL)*
　　　　CoNLL 2009 and 2010*(Conference on Computational Natural Language Learning.)*
　　　　HLT-NAACL 2006 , 2010*(Joint Human Language Technology Conference / North American Chapter of the ACL meeting)*
　　　　HLT-EMNLP 2005 *(Human Language Technology /Conf. on Empirical Methods in Natural Language Processing)*
　　　　EACL 2009 *(12th Conference of the European Chapter of the Association for Computational Linguistics)*
　　　　ICML 2009 *(26th International Conference on Machine Learning)*
　　　　AAAI 2008 *(AAAI Conference on Artificial Intelligence)*
　　　　ICCPOL 2006 *(International Conference on Computer Processing of Oriental Languages)*
　　　　IJCNLP 2005 *(International Joint Conference on Natural Language Processing)*
　　　　IWPT 2007*(10th International Conference on Parsing Technologies)*
Reviewing (Workshops):

ACL Student Resarch Workshop 2010
ACL Workshop on NLP and Linguistics 2010
NAACL Workshop on Statistical Parsing of Morphologically-rich Languages 2010
EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition
International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+) 2008
Workshop on Automated Syntactic Annotations for Interoperable Resources 2008
ACL Workshop on Parsing German 2008
COLING Workshop on Cross-Framework and Cross-Domain Parser Evaluation 2008
ACL Workshop on Issues in Teaching Computational Linguistics, 2008
Midwest Computational Linguistics Colloquium 2008
ACL SIGSEM Workshop on Prepositions 2005 and 2006;
"Beyond PARSEVAL" Workshop, LREC 2002;
Student Research Workshop at ACL 2002, 2003 and 2005;
NASSLLI Student Session 2003;
Women in Machine learning workshop, 2006

Reviewing (Grants):
Science Foundation Ireland Basic Research Programme review
National Science Foundation Panelist 2008, 2010

North-American Computational Linguistics Olympiad (NACLO) site coordinator
University of Illinois, 2009, 2010, 2011 *(with Roxana Girju)*
Speech and Natural Processing Lunch coordinator,
University of Illinois, 2008-2009
School of Engineering and Applied Sciences Postdoc Association Representative:
University of Pennsylvania, 2005-2007 *(with Dana Pirone)*
Member of Student Board of the European Chapter of ACL, 2001–2003
Nominated for membership. (together with Jon Herring and Malte Gabsdil)
Seminar Series coordinator, 2000–2002
Institute for Communicating and Collaborative Systems
and Human Communication Research Centre, University of Edinburgh

Other qualifications

Programming: Java, C, Matlab, Perl, Prolog
Languages: German (native); English, French (near-native); Mandarin Chinese, Italian (proficient); Latin;
HSK Chinese Proficiency Test, May 1996 *(language exam, qualifies for university entrance in PR China)*

Publications

PhD thesis

J. Hockenmaier (2003a). *Data and models for statistical parsing with Combinatory Categorial Grammar*, Informatics, University of Edinburgh.

Corpora

[1] J. Hockenmaier and Mark Steedman (2005a). *CCGbank*,
Linguistic Data Consortium, LDC2005T13 (ISBN 1-58563-340-2), Philadelphia, PA.

Journal publications

[1] J. Hockenmaier and M. Steedman (2007). CCGbank - a corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics* 33(3), pp 355–396.

[2] K.A. Dill, A. Lucas, J. Hockenmaier, L. Huang, D. Chiang, A.K. Joshi (2007). Computational Linguistics – a new tool for exploring biopolymer structures and statistical mechanics. Invited contribution (feature article), *Polymer*, 48(15), pp 4289-4300.

[3] J. Hockenmaier, A.K. Joshi and K.A. Dill (2007). Routes are trees: the parsing perspective on protein folding, *Proteins: Structure, Function, and Bioinformatics*, 66(1), pp 1–15

[4] J. Hockenmaier, G. Bierner and J. Baldridge (2004). Extending the Coverage of a CCG System, *Research in Language and Computation, Special Issue on Linguistic Theory and Grammar Implementation 2(2)*, pp 165–208, June 2004.

[5] J. Hockenmaier and C. Brew (1998). Error Driven Segmentation of Chinese, *Communications of COLIPS 8(1)*, pp 69—84, June 1998.
Also in J. Guo, K. T. Lua, and J. Xu, editors, *Proceedings of 12th Pacific Conference on Language and Information*, pp 218—229, Singapore, February 1998.

Refereed conference publications

[1] J. Hockenmaier and Y. Bisk (2010). Normal-form parsing for Combinatory Categorial Grammars with generalized composition and type-raising. *Proceedings of COLING 2010*, Beijing China. (Acceptance rate: 19% oral)

[2] P. Sondhi, M. Gupta, CX Zhai and J. Hockenmaier (2010). Shallow Information Extraction from Medical Forum Data. in *Proceedings of COLING 2010*, Beijing China. (Poster; acceptance rate: 42% oral or poster)

[3] Y. Tu, N. Johri, D. Roth and J. Hockenmaier (2010). Citation Author Topic Model in Expert Search. in *Proceedings of COLING 2010*, Beijing China. (Poster; acceptance rate: 42% oral or poster)

[4] A. Farhadi, M. Hejrati, M.A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, D. Forsyth (2010). Every picture tells a story: generating sentences from images. *Proceedings of the 11th European Conference on Computer Vision (ECCV 2010)*, pp. 15-29, Greece, 2010. (Acceptance rate: 27.7%)

[5] M. Hodosh, P. Young, C. Rashtchian and J. Hockenmaier (2010). Cross-Caption Coreference Resolution for Automatic Image Understanding. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL)*, pp. 162–171, Uppsala, Sweden. (Acceptance rate: 30.8%)

[6] J. Hockenmaier (2006). Creating a CCGbank and a wide-coverage CCG lexicon for German, *Proceedings of COLING/ACL 2006*, pp. 505–512, Sydney, 2006

[7] J. Hockenmaier, A.K. Joshi and K.A. Dill (2006). Protein folding and chart parsing, *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, pp. 293–300, Sydney, 2006.

[8] D. Reitter, J. Hockenmaier and F. Keller (2006). Priming effects in Combinatory Categorial Grammar, *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*, pp. 308–316, Sydney, 2006.

[9] J. Bos, S. Clark, M. Steedman, J. Curran, J. Hockenmaier (2004). Wide-Coverage Semantic Representations from a CCG Parser, *Proceedings of the 20th International Conference on Computational Linguistics (COLING '04)*, Geneva, Switzerland, August 2004.

[10] J.Hockenmaier (2003b). Parsing with Generative Models of Predicate-Argument Structure, *Proceedings of the 41th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp.359–366, Sapporo, Japan, July 2003.

[11] D. Gildea and J. Hockenmaier (2003). Identifying Semantic Roles Using Combinatory Categorial Grammar, *Proceedings of 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 57–64, Sapporo, Japan, July 2003.

[12] M. Steedman, R. Hwa, S. Clark, M. Osborne, A. Sarkar, J. Hockenmaier, P. Ruhlen, S. Baker and J. Crim (2003). Example Selection for Bootstrapping Statistical Parsers, *Proceedings of the Annual Meeting of the North American Chapter of the ACL (NAACL)*, pp.236–243, Edmonton, Canada, May 2003

[13] M. Steedman, M. Osborne, A. Sarkar, S. Clark, R. Hwa, J. Hockenmaier, P. Ruhlen, S. Baker and J. Crim (2003). Bootstrapping Statistical Parsers from Small Datasets, *Proceedings of the Annual Meeting of the European Chapter of the ACL (EACL)*, pp.331–338, Budapest, Hungary, April 2003.

[14] J. Hockenmaier and M. Steedman (2002a). Generative Models for Statistical Parsing with Combinatory Categorial Grammar, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp.335–342, Philadelphia, PA, July 2002.

[15] S. Clark, J. Hockenmaier and M. Steedman (2002). Building Deep Dependency Structures Using a Wide-Coverage CCG Parser, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp.327–334, Philadelphia, PA, July 2002.

[16] J. Hockenmaier and M. Steedman (2002b). Acquiring Compact Lexicalized Grammars from a Cleaner Treebank, *Third International Conference on Language Resources and Evaluation (LREC)*, Vol. V, pp. 1974–1981, Las Palmas, Spain, May 2002.

Refereed workshop publications

[1] C. Rashtchian, P. Young, M. Hodosh and J. Hockenmaier (2010). Collecting Image Annotations Using Amazon's Mechanical Turk, *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 139–147, June 2010, Los Angeles.

[2] J. Hockenmaier and P. Young (2008). Non-local scrambling: the equivalence of TAG and CCG revisited, *Proceedings of The Ninth International Workshop on Tree Adjoining Grammars and Related Formalisms*, pp. 41–48, Tübingen, Germany. June, 2008

[3] S. Clark and J. Hockenmaier (2002). Evaluating a Wide-Coverage CCG Parser,
Proceedings of Beyond PARSEVAL workshop, pp. 60–66, Las Palmas, Spain, June 2002.

[4] J. Hockenmaier (2001). Statistical Parsing for CCG with Simple Generative Models,
*Proceedings of Student Research Workshop, 39th Annual Meeting of the Association for
Computational Linguistics and 10th Meeting of the European Chapter*, pp 7–12,
Toulouse, France, July 2001.

[5] J. Hockenmaier, G. Bierner and J. Baldridge (2000). Providing Robustness for a CCG
System, *Proceedings of Workshop on Linguistic Theory and Grammar Implementation,
ESSLLI 2000*, pp 97—112, Birmingham, UK, August 2000.

Refereed abstracts

[1] D. Reitter, F. Keller and J. Hockenmaier (2007). Corpus-based evidence against sequence prim-
ing. Poster presented at the *20th annual CUNY conference on human sentence processing*, La
Jolla, March 2007.

[2] J. Hockenmaier, A. Joshi and K. A. Dill (2005). Routes are trees: The parsing
view on protein folding. Poster presented at *13th Annual International Conference
on Intelligent Systems for Molecular Biology (ISMB 2005)*, Detroit, MI, June 2005.
Among 50 (out of >600 submissions) selected for oral presentation

Unrefereed publications

[1] J. Hockenmaier, Mark Steedman (2005). *CCGbank: Users' manual*.
Department of Computer and Information Science Technical Report MS-CIS-05-09,
University of Pennsylvania, Philadelphia, PA.

[2] J. Hockenmaier, K.A. Dill, A.K. Joshi (2003). Protein folding as chart parsing,
Proceedings of Greater Philadelphia Bioinformatics Alliance Retreat, Haverford, PA,
October 2003.

[3] M.Steedman, S. Baker, S. Clark, J. Crim, J. Hockenmaier, R. Hwa, M. Osborne,
P. Ruhlen, A. Sarkar (2003). *CLSP WS-02 Final Report: Semi-Supervised Training
for Statistical Parsing*

Invited seminars, keynotes and panel discussions

Keynotes

" The future role of language resources for natural language parsing"
 Symposium on Language Resources
 Pacific Asia Conference on Language, Information and Computation (PACLIC 24)
 Tohoku University, Sendai, Japan, November 2010
"Translating the Tiger corpus to CCG: Why yet another German corpus?"
 Invited talk (keynote) at the ACL 2008 Workshop on Parsing German,
 Columbus, OH, June 2008
"What extracting grammars from treebanks can tell us about linguistic theory"
 Keynote lecture at the 6th International Workshop on Treebanks and Linguistic Theories,
 Bergen, Norway; December, 2007

Panel discussions

Workshop on NLP and Linguistics: Finding the Common Ground
 ACL, Uppsala, Sweden, July 2010
First Workshop on Statistical Parsing of Morphologically-Rich Languages
 NAACL HLT, Los Angeles, June 2010

Invited seminars

" NLP and computer vision "
 Tsujii lab, University of Tokyo, Japan, November 2010
"Scrambling in CCG: a theoretical and empirical evaluation"
 Colloquium, Department of Linguistics, Indiana University,
 Bloomington IN, March 2009
"Protein folding and parsing"
 Living texts: interdisciplinary approaches and methodological commonalities
 in biology and textual analysis (Invitation-only workshop),
 E-science Institute, Edinburgh, UK, October 2008
"Protein folding and parsing"
 Department of Computer Science,
 Union College, Schenectady, NY, February 2008
"What extracting grammars from treebanks can tell us about linguistic theory"
 Colloquium, Department of Linguistics,
 The University of Illinois at Urbana-Champaign, February 2008
"Protein folding and parsing"
 Seminar Series, The Center for Language and Speech Processing,
 Johns Hopkins University, Baltimore, April 2007
"Protein folding and parsing"
 Colloquium, Computer Science Department,
 The University of Illinois at Urbana-Champaign, April 2007
"Protein folding and parsing"
 Colloquium, Computer Science Department,
 The University of California, Irvine, March 2007
"Statistical parsing with expressive wide-coverage grammars"
 Linguistics Department, Northwestern University, February 2007
"Routes are Trees: The Parsing Perspectives on Protein Folding"
 Computational Biology Seminar, University of Pittsburgh, December 2005
"Protein folding as chart parsing"
 Language Technology Institute, Carnegie Mellon University, December 2005
"Statistical Parsing with CCG, or: Don't fire your linguist"
 Google, New York City, June 2005
"Protein folding as parsing"
 NLP group, Stanford University, October 2004
"Data and models for statistical parsing with CCG"
 Dublin Computational Linguistics Research Seminar Series,
 Trinity College, Dublin, December 2003
"Statistical Parsing with CCG",
 NLP group, Computer Laboratory, University of Cambridge, January 2001
"Extracting Combinatory Categorial Grammars from the Penn Treebank",
 IMS, Universität Stuttgart, December 2000