

Seth Kulick

Curriculum Vitae

Contact

Institute for Research in Cognitive Science
University of Pennsylvania
3401 Walnut Street, Suite 400A
Philadelphia, PA 19104
skulick@seas.upenn.edu
<http://www.cis.upenn.edu/~skulick/>

work: (215) 898-0358
cell: (267) 357-4558

Research Interests

Arabic Natural Language Processing, utilizing Natural Language Processing techniques for corpus development and search, statistical models of morphology, syntax, and semantics of language, categorial grammar and Tree Adjoining Grammar.

Education

Ph.D., Computer Science

University of Pennsylvania, August 2000

Dissertation: *Constraining Non-Local Dependencies in Tree Adjoining Grammar:
Computational and Linguistic Perspectives*

Co-adviser: Aravind K. Joshi (Computer Science) and Anthony Kroch (Linguistics)

Committee: Mitchell Marcus, Beatrice Santorini, Robin Clark, Ed Stabler (UCLA)

B.S.E., Computer Science, December 1986

University of Pennsylvania

Additional: Three years of Arabic instruction

University of Pennsylvania 2007–2009

ACTFL scores: Intermediate Medium to Advanced Low

Research Experience

7/2003–present

Associate Director, Annotation and Information Extraction

Institute for Research in Cognitive Science, University of Pennsylvania

Researcher in multiple projects in computational linguistics:

- Improving corpus annotation: Current research focuses on uniting and extending previous work utilizing techniques from Natural Language Processing (tree grammar extraction) for new approaches leading to significant improvements in various aspects of annotation, including quality-control and corpus searching.
- Arabic Treebank: Crucial role in multiple aspects of annotation production. Improved syntactic parser resulting in increased speed of treebank annotation. Designed and implemented new methods for relating levels of morphological and syntactic annotation resulting in improved quality-control for corpus revision. Currently overseeing redesign and implementation of annotation software infrastructure for increased speed and improved quality-control.

- Machine Translation evaluation: Co-developed an improved version of machine translation evaluation by utilizing syntactic annotation and tree grammar extraction to add flexibility to judgement of the accuracy of a translation compared with a single reference translation.
- "Mining the Bibliome": Integration of machine learning techniques (part-of-speech taggers and syntactic parsing) with annotation workflow for a biomedical corpus. Played key role in analysis relating multiple levels of annotation (syntax, semantic, named entities, relations) for information extraction.
- Co-Principal Investigator, National Science Foundation Grant SGER 0527116, "Enriching Parser Output for Treebank Construction." (April 2005–March 2006). Developed techniques for improving syntactic parser performance focusing on needs of treebank annotation. Currently being used in multiple annotation projects for increased annotation speed.

7/02–7/03

Research Associate

Dept. of Computer and Information Science, University of Pennsylvania

Researcher with the TIDES ("Translingual Information Detection, Extraction and Summarization") program at Penn. Crucial role in development of Penn's first contribution to the Automatic Content Extraction evaluation (pipelined statistical system for Information Extraction), including development of infrastructure and supervising other contributors. Assisted with management of program. Co-developed software for extraction of semantic information (predicate argument structure) from Chinese Treebank.

1993–2000

Research Fellow

Institute for Research in Cognitive Science, University of Pennsylvania

Doctoral Dissertation: Investigated Romance and German reduced constructions within the framework of Tree Adjoining Grammar (TAG). Showed how such constructions are related to previously unnoticed problems for TAG. The formal system proposed as a solution produced an analysis of these constructions such that constraints on different types of non-local dependencies are reduced to differences in syntactic "building blocks" of a very small size. Also produced a computational analysis of the processing complexity of German sentences with long scrambling.

Professional Activities

Reviewer, Association for Computational Linguistics (ACL), 2001, 2007, 2008, 2009

Program Committee: Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+6), Venice, Italy, May 2002

Program Committee: Logical Aspects of Computational Linguistics 1997, Nancy, France

Reviewer, Journal of Logic, Language, and Information, 2002

Reviewer, Mathematics of Language 1999

Reviewer, Fourth Workshop on Tree Adjoining Grammar (TAG+4) 1998

Reviewer, Penn Working Papers in Linguistics 6 1999

Reviewer, Penn Working Papers in Linguistics 5 1998

Publications

Book Chapters and Contributions

Mohamed Maamouri, Ann Bies, and Seth Kulick. Upgrading and enhancing the Penn Arabic treebank: A GALE challenge. In Joseph Olive, editor, *in progress for publication (book describing work in GALE program)*. 2009.

Seth Kulick and Mitch Marcus. Description of the KM machine translation evaluation system. In Joseph Olive, editor, *in progress for publication (book describing work in GALE program)*, chapter Machine Translation Evaluation and Optimization. 2009. Presented as MULCH system at GALE PI Meeting, May 2009.

Aravind K. Joshi, Seth Kulick, and Natash Kurtonina. An LTAG perspective on categorial inference. In M. Moortgat, editor, *Logical Aspects of Computational Linguistics 1998*, Lecture Notes in Artificial Intelligence. Springer-Verlag, 2001.

Seth Kulick and Aravind K. Joshi. Implementational aspects of a categorial grammar based on partial proof trees. In Ginzburg et. al, editor, *The Tbilisi Symposium on Logic, Language, and Computation: Selected Papers*, Studies in Logic, Language and Information. CSLI Publications and FoLLI, 1998.

Aravind K. Joshi and Seth Kulick. Partial proof trees, resource sensitive logics, and syntactic constraints. In C. Retoré, editor, *Logical Aspects of Computational Linguistics 1996*, Lecture Notes in Artificial Intelligence 1328. Springer-Verlag, 1997.

Journal articles

Aravind Joshi, Seth Kulick, and Natasha Kurtonina. Partial proof trees, hybrid logic, and quantifier scope. *Research on Logic and Computation*, 1.3, 2002.

Robert Frank, Seth Kulick, and K. Vijay-shanker. Monotonic c-command: A new perspective on Tree Adjoining Grammar. *Grammars*, 3:151–173, 2000.

Aravind Joshi and Seth Kulick. Partial proof trees as building blocks for a categorial grammar. *Linguistics and Philosophy*, 20:636–667, 1997.

Corpora

Mohamed Maamouri, Ann Bies, Seth Kulick, and Fatma Gaddeche. Arabic treebank part 5 - v1.0. Linguistic Data Consortium LDC2009E72, June 16 2009.

Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, and Basma Bouziri. Arabic treebank part 3 - v3.1. Linguistic Data Consortium LDC2008E22, January 20 2009.

Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, and Basma Bouziri. Arabic treebank part 2- v3.0. Linguistic Data Consortium LDC2008E62, January 20 2009.

Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, and Basma Bouziri. Arabic treebank part 1 - v4.0. Linguistic Data Consortium LDC2008E61, December 4 2008.

Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, and Basma Bouziri. Arabic treebank part 3 - v3.0. Linguistic Data Consortium LDC2008E22, August 20 2008.

David Graff, Mohamed Maamouri, Basma Bouziri, Sondos Krouna, Seth Kulick(contributions), and Tim Buckwalter. Standard Arabic morphological analyzer (sama) version 3.1. Linguistic Data Consortium LDC2009E73, 2009.

Conference proceedings

Ann Bies Mohamed Maamouri and Seth Kulick. Creating a methodology for large-scale correction of treebank annotation: the case of the Arabic treebank. In Khalid Choukri and Bente Maegaard, editors, *Proceedings of the Second International Conference on Arabic Language Resources and Tools*, Cairo, Egypt, April 2009. The MEDAR Consortium.

- Mohamed Maamouri, Seth Kulick, and Ann Bies. Diacritic annotation in the Arabic treebank and its impact on parser evaluation. In *Proceedings of LREC2008: Sixth International Conference on Language Resources and Evaluation*, Morocco, 2008.
- Mohamed Maamouri, Ann Bies, and Seth Kulick. Enhancing the Arabic treebank: A collaborative effort toward new annotation guidelines. In *Proceedings of LREC2008: Sixth International Conference on Language Resources and Evaluation*, Morocco, 2008.
- Mohamed Maamouri, ann Bies, and Seth Kulick. Enhanced annotation and parsing of the Arabic treebank. In *Proceedings of INFOS*, Cairo, Egypt, March 2008.
- Ryan Gabbard and Seth Kulick. Construct state modification in the Arabic treebank. In *Proceedings of ACL-08: HLT, Short Papers*, pages 209–212, Columbus, Ohio, June 2008. Association for Computational Linguistics.
- Nizar Habash, Ryan Gabbard, Owen Rambow, Seth Kulick, and Mitch Marcus. Determining case in Arabic: Learning complex linguistic behavior requires complex linguistic features. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1084–1092, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- Seth Kulick, Ryan Gabbard, and Mitch Marcus. Parsing the Arabic treebank: Analysis and improvements. In *Proceedings of Treebanks and Linguistic Theories*, Prague, 2006.
- Seth Kulick, Daniel Bikel, and Anthony Kroch. Treebank construction by levels using constrained chart parsing. In *Proceedings of Treebanks and Linguistic Theories*, Prague, 2006.
- Ann Bies, Stephanie Strassel, Haejoong Lee, Kazuaki Maeda, Seth Kulick, Yang Liu, Mary Harper, and Matthew Lease. Linguistic resources for speech parsing. In *Proceedings of LREC2006: Fifth International Conference on Language Resources and Evaluation*, 2006.
- Mohamed Maamouri, Ann Bies, and Seth Kulick. Diacritization: A challenge to Arabic treebank annotation and parsing. In *Proceedings of Arabic NLP/MT Conference*. The British Computer Society, Natural Language Translation Specialist Group, October 2006.
- Olga Babko-Malaya, Ann Bies, Ann Taylor, Szuting Yi, Martha Palmer, Mitch Marcus, Seth Kulick, and Libin Shen. Issues in synchronizing the English treebank and propbank. In *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006*, pages 70–77, Sydney, Australia, July 2006. Association for Computational Linguistics.
- Ryan Gabbard, Seth Kulick, and Mitchell Marcus. Fully parsing the Penn treebank. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 184–191, New York City, USA, June 2006. Association for Computational Linguistics.
- Ryan McDonald, Fernando Pereira, Seth Kulick, Scott Winters, Yang Jin, and Pete White. Simple algorithms for complex relation extraction with applications to biomedical IE. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 491–498, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
- Ann Bies, Seth Kulick, and Mark Mandel. Parallel entity and treebank annotation. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 21–28, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
- Seth Kulick, Ann Bies, Mark Liberman, Mark Mandel, Ryan McDonald, Martha Palmer, Andrew Schein, Lyle Ungar, Scott Winters, and Pete White. Integrated annotation for biomedical information extraction. In Lynette Hirschman and James Pustejovsky, editors, *HLT-NAACL 2004 Workshop: BioLINK 2004, Linking Biological Literature, Ontologies and Databases*, pages 61–68, Boston, Massachusetts, USA, May 6 2004. Association for Computational Linguistics.

O. Babko-Malaya, M. Palmer, N. Xue, A. Joshi, and S. Kulick. Proposition bank II: Delving deeper. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 17–23, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.

Nianwen Xue and Seth Kulick. Automatic predicate argument structure analysis of the Penn Chinese treebank. In *Proceedings of Machine Translation Summit IX*, New Orleans, Louisiana, 2003. AMTA.

Seth Kulick, Mark Liberman, Martha Palmer, and Andrew Schein. Shallow semantic annotation of biomedical corpora for information extraction. In *Special Interest Group Meeting on Text Mining (Biolink)*, Brisbane, Australia, June 2003.

Seth Kulick. A unified account of locality constraints for clitic climbing and long scrambling. In *Proceedings of the 24th Penn Linguistics Colloquium*, Philadelphia, 2000.

Aravind K. Joshi, Seth Kulick, and Natasha Kurtonina. Semantic computation for partial proof trees. In *Proceedings of the Twelfth Amsterdam Colloquium*, Amsterdam, 1999.

Aravind K. Joshi, Seth Kulick, and Natasha Kurtonina. Partial proof trees, hybrid logic, and quantifier scope. In *Proceedings of the Workshop on Resource Logics and Derivational Minimalism*, Utrecht, ESSLI 1999.

Seth Kulick. Long transformations in TAG and HPSG. In *Proceedings of Formal Grammar Conference*, Utrecht, 1999.

Aravind K. Joshi, Seth Kulick, and Natasha Kurtonina. Cross-serial dependencies and partial proof trees. In *Proceedings of Formal Grammar Conference*, Utrecht, 1999.

Robert Frank, Seth Kulick, and K. Vijay-Shanker. C-command and extraction in tag. In *Mathematics of Language*, 1999.

Seth Kulick. Generalized transformations and restructuring in romance. In *Proceedings of Eastern States Conference on Linguistics (ESCOL)*, 1997.

Seth Kulick. Using higher-order logic programming for semantic interpretation of coordinate constructs. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pages 213–219, Cambridge, Massachusetts, USA, June 1995. Association for Computational Linguistics.

B. Srinivas, Christine Doran, and Seth Kulick. Heuristics and parse ranking. In *Proceedings of the 4th Annual International Workshop on Parsing Technologies*, pages 224–233, Prague, 1995.

Working Papers and Technical Reports

Seth Kulick, Robert Frank, and K. Vijay-Shanker. Defective complements in Tree Adjoining Grammar. In *Penn Working Papers in Linguistics*, Philadelphia, 1999.

Seth Kulick. Exceptional case marking in the XTAG system. In *Penn Working Papers in Linguistics*, Philadelphia, 1997.

Seth Kulick. Process algebra, CCS, and bisimulation decidability. Technical Report 94-06, Institute for Research in Cognitive Science, 1994.

Other Conference Presentations

Seth Kulick and Mitch Marcus. The MULCH MT evaluation system. May 2009. Presented at GALE PI Meeting. Written up as KM system in (Kulick and Marucs, 2009).

Robert Frank and Seth Kulick. Locality, phases and Tree Adjoining Grammar. In *Workshop on EPP and Phases*, MIT, January 2003.

Seth Kulick. Tree composition and generalized transformations. In *Tools in Linguistic Theory*, Utrecht, July 2001.

Seth Kulick. Scrambling in German: Relating processing difficulty to syntactic constraints. In *14th Annual CUNY Conference on Human Sentence Processing*, Philadelphia, March 2001.

Seth Kulick. Locality domains and reduced constructions. In *Linguistic Society of America*, Washington, D.C., January 2001.

Seth Kulick. Complex predicates in Tree Adjoining Grammar. In *Workshop on Complex Predicates*, DGfS, Halle, Germany, 1998.

Heidi Harley and Seth Kulick. TAG and raising in VSO languages. In *Proceedings of the TAG+ Conference*, Philadelphia, 1998.

Seth Kulick. TAG and clitic climbing in romance. In *Proceedings of the TAG+ Conference*, Philadelphia, 1998.

Seth Kulick. An exploration of some extensions of Tree Adjoining Grammar. In *Mathematics of Language*, Saarbrücken, Germany, 1997.

Aravind Joshi and Seth Kulick. An application of typed label-selective lambda-calculus to formal grammar. In *Mathematics of Language*, Philadelphia, 1995.

Aravind Joshi and Seth Kulick. Partial proof trees as building blocks for a categorial grammar. In *Proceedings of the Formal Grammar Conference*, Barcelona, 1995.