

STEPHANIE M. STRASSEL

215.863.1813 | strassel@ldc.upenn.edu | <https://www.ldc.upenn.edu/staff/stephanie-strassel>

PROFILE

Senior manager, linguist with 15+ years professional experience managing creation, collection, annotation and distribution of linguistic resources for a diverse set of technologies including machine translation, speech recognition, question answering, handwriting recognition, document classification, topic detection, information extraction, knowledge base population, natural language understanding, speaker identification, dialect recognition, multimodal event detection and related areas.

PROFESSIONAL EXPERIENCE

Linguistic Data Consortium, University of Pennsylvania 1998 — present
Philadelphia, PA

Senior Associate Director, Annotation and Collection 2011—present

Manage all aspects of data collection and creation, human subject research, linguistic annotation and corpus distribution for a diverse set of externally sponsored human language technology research programs and evaluation campaigns. Act as PI or co-PI for multiple, concurrent, revenue-generating grants/contracts, comprising several million dollars in annual funding. Lead Annotation/Collection Group, consisting of 13+ full-time and 75+ part-time linguists, managers, researchers and programmers. Plan and direct linguistic resource development activities among internal and external senior researchers and administrators, technical managers, programmers, linguists, annotators, subcontractors and vendors. Provide leadership for LDC senior staff, developing and disseminating infrastructure for activities that support multiple functional areas. Represent LDC activities and best practices to external communities via international publications, conferences and workshops; serve as external expert on language resource creation activities on review panels and oversight committees.

Senior Associate Director, Annotation 2009—2011

Associate Director, Annotation 2004— 2009

Associate Director, Data Creation 2002—2004

Manager, Annotation Group 1998—2002

Johns Hopkins University Human Language Technology Center of Excellence 2013 — present
Baltimore, MD
Research Associate (Adjunct)

Law School Admissions Council 2001 — 2005

Princeton, NJ

Provided linguistic consultation on dialect diversity of American English; conducted feasibility study for inclusion of diverse dialects in a high-stakes graduate-level listening comprehension test of cognitive abilities.

University of Pennsylvania, College of General Studies 2001 — 2005

Philadelphia, PA

Instructor: Introduction to Sociolinguistics

Instructor: American Dialects

Teaching Assistant: Introduction to Language 1 & 2

Swarthmore College, Linguistics Department 1999, 2007

Swarthmore, PA

Lecturer: Language Culture and Society

SELECTED RESEARCH GRANTS

Linguistic Data Consortium, University of Pennsylvania

Principal Investigator

- SonicViper, October 2015-September 2020. \$2,378,007.
- SIREN: Specialized Intra/Interlingual Resources for Emergent News (DARPA), August 2015-September 2019, HR0011-15-C-0123, \$9,140,117.
- Data Resources Aladdin Video Evaluations (IARPA ALADDIN), May 2014-October 2015, 2104-14042400005, \$749,928.
- Linguistic Resources for NIST Evaluations - TRECVID MED/MER, October 2013-September 2014. SB1341-13-CQ-0021, \$1,018,230
- ReORIENT: Resources for Operationally Relevant Information Extraction from Non-Explicit Text (DARPA DEFT), November 2012-December 2017, FA8750-13-2-0045. \$11,106,157
- Linguistic Resources for Multilingual, Genre-Independent Language Technologies (DARPA BOLT), September 2011-October 2015, HR0011-11-C-0145. \$13,691,165
- Linguistic Resources for Robust Automatic Transcription of Speech (DARPA RATS), August 2010 - June 2014, D10PC20016. \$3,271,926.
- DARPA Machine Reading, SAIC, July 2009-June 2014. \$1,477,713.
- Multilingual Automatic Document Classification Analysis and Translation (DARPA MADCAT), December 2007-April 2014, HR0011-08-1-004. \$6,041,454.

Co-Principal Investigator

- ViperToxin, September 2012-September 2015, H98230-10-D-0041. \$2,223,157.
- Indefinite-Delivery/Indefinite Quantity Provision of Linguistic Resources, (NIST Open Evaluations), September 2007-January 2013, SB134107CQ0024. \$4,797,799.
- Integrated Linguistic Resources for Language Exploitation Technologies, (DARPA GALE), October 2005-June 2011, HR0011-03-1-0003. \$29,411,033.

EDUCATION

University of Pennsylvania, Philadelphia, PA 1992 — 2003

Graduate level coursework toward PhD in Linguistics

George Washington University, Washington, DC 1998 — 1992

B.A. summa cum laude, Sociology

SELECTED PROFESSIONAL ACTIVITIES

Review Panels

- National Science Foundation - various
- European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies (CHIST-ERA)
- CASL (Center for Advanced Study of Language) - Language Variation Library

Training

- Advanced Management Program at the University of Pennsylvania (AMP-UP) completed 2013

Program Committee/Reviewer

Conferences

Language Resources and Evaluation Conference (LREC), Interspeech, Association for Computational Linguistics (ACL), Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), Conference on Empirical Methods in Natural Language Processing (EMNLP), International Conference on Computational Linguistics (COLING), American Machine Translation Association (AMTA), International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (COCOSDA), New Ways of Analyzing Variation (NWAV), North East Linguistic Society (NELS), Penn Linguistics Colloquium (PLC)

Journals

Language Resources and Evaluation, Speech Communication, Machine Translation, Language Variation and Change, Computers and the Humanities, Penn Review of Linguistics, Swarthmore Linguistics Review

Thesis Adviser

Linguistics Department, Swarthmore College

- Christine Lattin, *The potential of data sharing in quantitative sociolinguistics: /t,d/ deletion as a case study*. 2000.
- Jennifer Tyson, *Acquiring a common dialect: the speech of military brats*. 2000.

SELECTED VOLUNTEER EXPERIENCE

Historic Lansdowne Theater Corporation

2015 — present

Lansdowne, PA

Member, Board of Directors

Chair, Events and Community Outreach Committee

Co-Chair, 2015 Crowdfunding Campaign

Designed and executed crowdfunding campaign resulting in donations from 317 individuals totaling over \$45,000.

Historic Architecture Review Board

2016 — present

Lansdowne, PA

Member, Board of Directors

SELECTED PUBLICATIONS

2017

Stephanie Strassel, Ann Bies and Jennifer Tracey. *Situational Awareness for Low Resource Languages: the LORELEI Situation Frame Annotation Task*. First Workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP). Aberdeen, Scotland, April 2017.

Karen Jones, Stephanie Strassel, Kevin Walker, David Graff, Jonathan Wright. *Call My Net Corpus: A Multilingual Corpus for Evaluation of Speaker Recognition Technology*. Interspeech 2017. Stockholm, Sweden, August 2017.

2016

Christopher Cieri, Mike Maxwell, Stephanie Strassel, Jennifer Tracey. *Selection Criteria for Low Resource Language Programs*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

Ann Bies, Zhiyi Song, Jeremy Getman, Joe Ellis, Justin Mott, Stephanie Strassel, Martha Palmer, Teruko Mitamura, Marjorie Freedman, Heng Ji, Tim O'Gorman. *A Comparison of Event Representations in DEFT*. NAACL 2016 - The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation, San Diego, CA, June 2017.

Zhiyi Song, Ann Bies, Stephanie Strassel, Joe Ellis, Teruko Mitamura, Hoa Dang, Yukari Yamakawa, Sue Holm. *Event Nugget and Event Coreference Annotation*. NAACL 2016 - The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation, San Diego, CA, June 2017.

Ian Soboroff, Kira Griffitt, Stephanie Strassel. *The BOLT IR Test Collections of Multilingual Passage Retrieval from Discussion Forums*. SIGIR 2016: 39th Annual International ACM Special Interest Group on Information Retrieval Conference, Pisa, Italy, July 2017.

Stephanie Strassel, Jennifer Tracey. *LORELEI Language Packs: Data, Tools, and Resources for Technology Development in Low Resource Languages*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

Justin Mott, Ann Bies, Zhiyi Song, Stephanie Strassel. *Parallel Chinese-English Entities, Relations and Events Corpora*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

Xuansong Li, Martha Palmer, Nianwen Xue, Lance Ramshaw, Mohamed Maamouri, Ann Bies, Kathryn Conger, Stephen Grimes, Stephanie Strassel. *Large Multi-lingual, Multi-level, Multi-genre Annotation Corpus*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

Xuansong Li, Jennifer Tracey, Stephen Grimes, Stephanie Strassel. *Uzbek-English and Turkish-English Morpheme Alignment Corpora*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

Karen Jones, David Graff, Jonathan Wright, Kevin Walker, Stephanie Strassel. *Multi-language Speech Collection for NIST LRE*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

Kira Griffitt, Stephanie Strassel. *The Query of Everything: Developing Open-Domain, Natural-Language Queries for BOLT Information Retrieval*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference, Portoroz, May 2016.

2015

Teruko Mitamura, Yukari Yamakawa, Susan Holm, Zhiyi Song, Ann Bies, Seth Kulick, Stephanie Strassel. *Event Nugget Annotation: Processes and Issues*. NAACL HLT 2015 - 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, Denver, CO June 2015.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant and Xiaoyi Ma. *From Light to Rich ERE: Annotation of Entities, Relations, and Events*. NAACL HLT 2015 - 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, Denver, CO June 2015.

Mariona Taulé, M Antonia Martí, Ann Bies, Aina Garí, Montserrat Nofre, Zhiyi Song, Stephanie Strassel and Joe Ellis. *Spanish Treebank Annotation of Informal Non-Standard Web Text*. NLPIT 2015 - 1st International Workshop on Natural Language Processing for Informal Text, Rotterdam, Netherlands, June 2015.

2014

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, Joe Ellis. *A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards*. 2nd Workshop on Events: Definition, Detection, Coreference, and Representation. ACL 2014, Baltimore, June 2014.

Ann Bies, Zhiyi Song, Mohamed Maamouri, Stephen Grimes, Haejoong Lee, Jonathan Wright, Stephanie Strassel, Nizar Habash, Ramy Eskander, Owen Rambow. *Transliteration of Arabizi into Arabi Orthography: Developing a Parallel Annotated Arabizi-Arabic Script SMS/Chat Corpus*. ANLP Workshop: Arabic Natural Language Processing Workshop. EMNLP 2014, Doha, October 2014.

Christopher Cieri, Denise DiPersio, Mark Liberman, Andrea Mazzucchi, Stephanie Strassel, Jonathan Wright. *New Directions for Language Resource Development and Distribution*. LREC 2014, Reykjavik, May 2014.

David Graff, Kevin Walker, Stephanie Strassel, Xiaoyi Ma, Karen Jones, Ann Sawyer. *The RATS Collection: Supporting HLT Research with Degraded Audio Data*. LREC 2014, Reykjavik, May 2014.

Zhiyi Song, Stephanie Strassel, Haejoong Lee, Kevin Walker, Jonathan Wright, Jennifer Garland, Dana Fore, Brian Gainor, Preston Cabe, Thomas Thomas, Brendan Callahan, Ann Sawyer. *Collecting Natural SMS and Chat Conversations in Multiple Languages: The BOLT Phase 2 Corpus*. LREC 2014, Reykjavik, May 2014.

2013

Joe Ellis, Jeremy Getman, Justin Mott, Xuansong Li, Kira Griffitt, Stephanie Strassel, Jonathan Wright. *Linguistic Resources for 2013 Knowledge Base Population Evaluations*. TAC KBP 2013 Workshop, Gaithersburg, November 2013.

2012

Joe Ellis, Xuansong Li, Kira Griffitt, Stephanie Strassel, Jonathan Wright. *Linguistic Resources for 2012 Knowledge Base Population Evaluations*. TAC KBP 2012 Workshop, Gaithersburg, November 2012.

Jennifer Garland, Stephanie Strassel, Safa Ismael, Zhiyi Song, Haejoong Lee. *Linguistic Resources for Genre-Independent Language Technologies: User-Generated Content in BOLT*. LREC 2012, Istanbul, May 2012.

Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Mohamed Maamouri, Ann Bies, Nianwen Xue. *Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures*. LREC 2012, Istanbul, May 2012.

Xuansong Li, Stephanie Strassel, Heng Ji, Kira Griffitt, Joe Ellis. *Linguistic Resources for Entity Linking Evaluation: from Monolingual to Cross-lingual*. LREC 2012, Istanbul, May 2012.

Zhiyi Song, Safa Ismael, Stephen Grimes, David Doermann, Stephanie Strassel. *Linguistic Resources for Handwriting Recognition and Translation Evaluation*. LREC 2012, Istanbul, May 2012.

Stephanie Strassel, Amanda Morris, Jonathan Fiscus, Christopher Caruso, Haejoong Lee, Paul Over, James Fiumara, Barbara Shaw, Brian Antonishek, Martial Michel. *Creating HAVIC: Heterogeneous Audio Visual Internet Collection*. LREC 2012, Istanbul, May 2012.

Stephanie Strassel, Kevin Walker, Karen Jones, Dave Graff, Christopher Cieri. *New Resources for Recognition of Confusable Linguistic Varieties: The LRE11 Corpus*. Odyssey 2012: The Speaker and Language Recognition Workshop, Singapore, June 2012.

Kevin Walker and Stephanie Strassel. *The RATS Radio Traffic Collection System*. Odyssey 2012: The Speaker and Language Recognition Workshop, Singapore, June 2012.

Jonathan Wright, Kira Griffitt, Joe Ellis, Stephanie Strassel, Brendan Callahan. *Annotation Trees: LDC's customizable, extensible, scalable annotation infrastructure*. LREC 2012, Istanbul, May 2012.

2011

Data Acquisition and Linguistic Resources, Chapter in Olive, Joseph, Caitlin Christianson and John McCary (eds.) *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*, New York, Springer (Editor with Caitlin Christianson, John McCary, William Staderman and Joseph Olive).

Xuansong Li, Joe Ellis, Kira Griffitt, Stephanie Strassel, Robert Parker, Jonathan Wright. *Linguistic Resources for 2011 Knowledge Base Population Evaluation*. TAC KBP 2011, Gaithersburg, November 2011.

2010

Marianna Di Paolo, Malcah Yaeger-Dror, Christopher Cieri, Stephanie Strassel, Zsuzsanna Fagyal. *Towards Best Practices in Sociophonetics Workshop*. NWAV 39, San Antonio, November 2010.

Meghan Lammie Glenn, Stephanie M. Strassel, Haejoong Lee, Kazuaki Maeda, Ramez Zakhary, Xuansong Li. *Transcription Methods for Consistency, Volume and Efficiency*. LREC 2010, Valletta, May 2010.

Stephen Grimes, Xuansong Li, Ann Bies, Seth Kulick, Xiaoyi Ma, Stephanie Strassel. *Creating Arabic-English Parallel Word-Aligned Treebank Corpora at LDC*. Workshop on Language Resources and Human Language Technologies for Semitic Languages. LREC 2010, Valletta, May 2010.

Xuansong Li, Niyu Ge, Stephen Grimes, Stephanie M. Strassel, Kazuaki Maeda. *Enriching Word Alignment with Linguistic Tags*. LREC 2010, Valletta, May 2010.

Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Xiaoyi Ma, Niyu Ge, Ann Bies, Nianwen Xue, Mohamed Maamouri. *Parallel Aligned Treebank Corpora at LDC: Methodology, Annotation and Integration*. TLT9: The 9th International Workshop on Treebanks and Linguistic Theories, University of Tartu, December 2010.

Paul McNamee, Hoa Trang Dang, Heather Simpson, Patrick Schone, Stephanie Strassel. *An Evaluation of Technologies for Knowledge Base Population*. LREC 2010, Valletta, May 2010.

Heather Simpson, Stephanie Strassel, Robert Parker, Paul McNamee. *Wikipedia and the Web of Confusable Entities: Experience from Entity Linking Query Creation for TAC 2009 Knowledge Base Population*. LREC 2010, Valletta, May 2010.

Zhiyi Song, Stephanie Strassel, Gary Krug, Kazuaki Maeda. *Enhanced Infrastructure for Creation and Collection of Translation Resources*. LREC 2010, Valletta, May 2010.

Stephanie Strassel, Dan Adams, Henry Goldberg, Jonathan Herr, Ron Keesing, Daniel Oblinger, Heather Simpson, Robert Schrag, Jonathan Wright. *The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks*. LREC 2010, Valletta, May 2010.

2009

Christopher Cieri and Stephanie Strassel. *Closer Still to a Robust, All Digital, Empirical, Reproducible Sociolinguistic Methodology*. NWAV 38, Ottawa, October 2009.

Lauren Friedman, Stephanie Strassel, Haejoong Lee. *A Quality Control Framework for Gold Standard Reference Translations: The Process and Toolkit Developed for GALE*. *Translating & The Computer* 2009, London, November 2009.

Meghan Lammie Glenn, Stephanie Strassel, Haejoong Lee. *XTrans: a speech annotation and transcription tool*. In INTERSPEECH-2009, 2855-2858.

Niklas Paulsson, Khalid Choukri, Djamel Mostefa, Denise DiPersio, Meghan Lammie Glenn, Stephanie Strassel. *A Large Arabic Broadcast News Speech Data Collection*. MEDAR Second International Conference on Arabic Language Resources and Tools, Cairo, April 2009.

Stephanie Strassel. *Linguistic Resources for Arabic Handwriting Recognition*. A Large Arabic Broadcast News Speech Data Collection. MEDAR Second International Conference on Arabic Language Resources and Tools, Cairo, April 2009.

2008

Chomicha Bendahman, Meghan Lammie Glenn, Djamel Mostefa, Niklas Paulsson, Stephanie Strassel. *Quick Rich Transcriptions of Arabic Broadcast News Speech Data*. LREC 2008, Marrakech, May 2008.

Christopher Cieri, Stephanie Strassel, Meghan Glenn, Reva Schwartz, Wade Shen, Joseph Campbell. *Bridging the Gap Between Linguists and Technology Developers: Large-Scale Sociolinguistic Annotation for Dialect and Speaker Recognition*. LREC 2008, Marrakech, May 2008.

Lauren Friedman and Stephanie Strassel. *Identifying Common Challenges for Human and Machine Translation: A Case Study from the GALE Program*. The Eighth Conference of the Association for Machine Translation in the Americas (AMTA) 2008, Waikiki, October 2008.

Lauren Friedman, Stephanie Strassel, Meghan Lammie Glenn. *Explicit and Implicit Requirements of Technology Evaluations: Implications for Test Data Creation*. ELRA Workshop on Evaluation: Looking into the Future of Evaluation: when automatic metrics meet task-based and performance-based approaches. LREC 2008, Marrakech, May 2008.

Meghan Lammie Glenn, Stephanie Strassel, Lauren Friedman, Haejoong Lee, Shawn Medero. *Management of Large Annotation Projects Involving Multiple Human Judges: a Case Study of GALE Machine Translation Post-editing*. LREC 2008, Marrakech, May 2008.

Kazuaki Maeda, Haejoong Lee, Shawn Medero, Julie Medero, Robert Parker, Stephanie Strassel. *Annotation Tool Development for Large-Scale Corpus Creation Projects at the Linguistic Data Consortium*. LREC 2008, Marrakech, May 2008.

Kazuaki Maeda, Xiaoyi Ma, Stephanie Strassel. *Creating Sentence-Aligned Parallel Text Corpora from a Large Archive of Potential Parallel Text using BITS and Champollion*. LREC 2008, Marrakech, May 2008.

Zhiyi Song and Stephanie Strassel. *Entity Translation and Alignment in the ACE-07 ET Task*. LREC 2008, Marrakech, May 2008.

Stephanie Strassel, Lauren Friedman, Safa Ismael, Linda Brandschain. *New Resources for Document Classification Analysis and Translation Technologies*. LREC 2008, Marrakech, May 2008.

Stephanie Strassel, Mark A. Przybocki, Kay Peterson, Zhiyi Song, Kazuaki Maeda. *Linguistic Resources and Evaluation Techniques for Evaluation of Cross-Document Automatic Content Extraction*. LREC 2008, Marrakech, May 2008.

Stephanie Strassel. *Beyond Corpora: The Role of Data Centers in Technology Research and Development*. Invited speaker, Oriental COCODA (International Conference on Speech Databases and Assessments), Kyoto, November 2008.

2007

Christopher Cieri, Stephanie Strassel, Meghan Lammie Glenn, Lauren Friedman. *Linguistic Resources in Support of Various Evaluation Metrics*. Workshop on Automatic Procedures in MT Evaluation. MT Summit XI, Copenhagen, September 2007.

2006

Ann Bies, Stephanie Strassel, Haejoong Lee, Kazuaki Maeda, Seth Kulick, Yang Liu, Mary Harper, Matthew Lease. *Linguistic Resources for Speech Parsing*. LREC 2006, Genoa, May 2006.

Kazuaki Maeda, Haejoong Lee, Julie Medero, Stephanie Strassel. *A New Phase in Annotation Tool Development at the Linguistic Data Consortium: The Evolution of the Annotation Graph Toolkit*. LREC 2006, Genoa, May 2006.

Julie Medero, Kazuaki Maeda, Stephanie Strassel, Christopher Walker. *An Efficient Approach for Gold-Standard Annotation: Decision Points for Complex Tasks*. LREC 2006, Genoa, May 2006.

Stephanie Strassel and Andrew W. Cole. *Corpus Development and Publication*. LREC 2006, Genoa, May 2006.

Stephanie Strassel, Christopher Cieri, Andy Cole, Denise DiPersio, Mark Liberman, Xiaoyi Ma, Mohamed Maamouri, Kazuaki Maeda. *Integrated Linguistic Resources for Language Exploitation Technologies*. LREC 2006, Genoa, May 2006.

2005

Meghan Lammie Glenn and Stephanie Strassel. *Linguistic Resources for Meeting Speech Recognition*. MLMI 05 (2nd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms), Edinburgh, July 2005.

Emmanuele Grosicki, Kazuaki Maeda, Stephanie Strassel, Christopher Walker. *Prosodic Analysis of Disfluencies, Fillers and Semantic Units in American English Telephone Conversations*. IDP05: Interface Discours-Prosodie. (International Symposium on Discourse and Prosody as Complex Interface), Aix-en-Provence, 2005.

Jachym Kolar, Jan Svec, Stephanie Strassel, Christopher Walker, Dagmar Kozlakova, Josef Psutka. *Czech spontaneous speech corpus with structural metadata*. In INTERSPEECH-2005, 1165-1168.

Stephanie Strassel, Jachym Kolar, Zhiyi Song, Leila Barclay, Meghan Lammie Glenn. *Structural Metadata Annotation: Moving Beyond English*. Interspeech 2005 (9th European Conference on Speech Communication and Technology), Lisbon, September 2005.

2004

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, Ralph Weischedel. *Automatic Content Extraction (ACE) program - task definitions and performance measures*. LREC 2004, Lisbon, May 2004.

Shudong Huang, Stephanie Strassel, Alexis Mitchell, Zhiyi Song. *Shared Resources for Multilingual Information Extraction and Challenges in Named Entity Annotation*. Named Entity Recognition for NLP Applications Workshop. IJCNLP 04, Hainan Island, March 2004.

Kazuaki Maeda and Stephanie Strassel. *Annotation Tools for Large-Scale Corpus Development: Using AGTK at the Linguistic Data Consortium*. LREC 2004, Lisbon, May 2004.

Douglas Oard, Dagobert Soergel, G. Craig Murray, David Doermann, Jianqiang Wang, Bhuvana Ramabhadran, Martin Franz, James Mayfield, Samuel Gustman, Stephanie Strassel. *Building an Information Retrieval Test Collection for Spontaneous Conversational Speech*. 27th Annual International ACM SIGIR Conference (SIGIR2004), Sheffield, July 2004.

Stephanie Strassel. *Linguistic Resources for Effective, Affordable, Reusable Speech-to-Text*. LREC 2004, Lisbon, May 2004.

2003

Christopher Cieri and Stephanie Strassel. *Robust Sociolinguistic Methodology: Tools, Data and Best Practices Workshop* co-leader and organizer. NWAV 32, Philadelphia, 2003.

Christopher Cieri and Stephanie Strassel. *Robust Sociolinguistic Methodology: Tools, Data and Best Practices*. NWAV 32, Philadelphia, 2003.

Christopher Cieri, Mike Maxwell, Stephanie Strassel. *Core Linguistic Resources for the World's Languages*. ELSNET, ENABLER, ICWLR Joint Workshop, Paris, 2003.

Stephanie Strassel, Mike Maxwell, Christopher Cieri. *Linguistic Resource Creation for Research and Technology Development: A Recent Experiment*. Association for Computing Machinery Transactions on Asian Language Information Processing (TALIP). Volume 2, Issue 2, 101-117.

Stephanie Strassel, David Miller, Kevin Walker, Christopher Cieri. *Shared Resources for Robust Speech-to-Text Technology*. Eurospeech 2003, Geneva, September 2003.

Stephanie Strassel, Alexis Mitchell, Shudong Huang. *Multilingual Resources for Entity Extraction*. 41st Annual Meeting of the Association for Computational Linguistics Workshop on Multilingual and Mixed-language Named Entity Recognition: Combining Statistical and Symbolic Models, Sapporo, July 2003.

Stephanie Strassel. *Corpus Creation for Disfluency Research*. Disfluency in Spontaneous Speech Conference, Gothenburg, 2003.

2002

Christopher Cieri and Stephanie Strassel. *The DASL Project: a Case Study in Data Re-Annotation and Re-Use*. LREC 2002, Canary Islands, May 2002.

Christopher Cieri, Stephanie Strassel, David Graff, Nii Martey, Kara Rennert, Mark Liberman. *Corpora for Topic Detection and Tracking*. James Allan, ed. Topic Detection and Tracking: Event-based Information Organization, Kluwer International Series on Information Retrieval, Bruce Croft, series editor, Boston, Kluwer Academic Publishers.

Christopher Cieri, Stephanie Strassel, William Labov. *Sharable Resources for Sociolinguistic Research*. NNAV 31, Stanford, October 2002.

Horacio Saggion, Dragomir Radev, Simone Teufel, Wai Lam, Stephanie Strassel. *Developing Infrastructure for the Evaluation of Single and Multi-document Summarization Systems in a Multi-lingual Environment*. LREC 2002, Canary Islands, May 2002.

Stephanie Strassel. *Annotating Temporal Attributes in the ACE Corpora*. Invited panelist, Annotation Standards for Temporal Information in Natural Language Workshop. LREC 2002, Canary Islands, May 2002.

2001

Christopher Cieri, Andy Cole, Dave Graff, Nii Martey, Stephanie Strassel, Cristina Tofan. *SPINE 2001 Data Preparation and Annotation and the SPINE Corpora*. Communicator, SPINE, ROAR Workshop, Orlando, November 2001.

Stephanie Strassel and Christopher Cieri. *Data and Annotations for SocioLinguistics: A Corpus- Based Approach to Sociolinguistic Research*. Penn Linguistic Colloquium, Philadelphia, March 2001.

Stephanie Strassel, Christopher Cieri, Steven Bird. *Shared Resources and Community Building for Corpus Linguistics and Language Teaching*. Corpus Linguistics and Language Teaching Workshop, Boston, March 2001.

Stephanie Strassel. *Corpus-Based Sociolinguistic Research: A -t/d Deletion Case Study*. NNAV 30, Raleigh, October 2001.

2000

Charles Boberg and Stephanie Strassel. *Short-a in Cincinnati: A Change in Progress*. Journal of English Linguistics, June 2000.

Christopher Cieri, Dave Graff, Mark Liberman, Nii Martey, Stephanie Strassel. *Large Multilingual Broadcast News Corpora for Cooperative Research in Topic Detection and Tracking: The TDT2 and TDT3 Corpus Efforts*. LREC 2000, Athens, May 2000.

Christopher Cieri, David Graff, Nii Martey, Stephanie Strassel. *The TDT-3 Text and Speech Corpus*. Topic Detection and Tracking Workshop, Vienna, Virginia, February 28 - March 1, 2000.

Dave Graff, Stephanie Strassel, Christopher Cieri. *Resources, New and Forthcoming, from LDC*. 2000 Speech Transcription Workshop, University of Maryland, May 2000.

Stephanie Strassel, Dave Graff, Nii Martey, Christopher Cieri. *Quality Control in Large Annotation Projects Involving Multiple Judges: The Case of the TDT Corpora*. LREC 2000, Athens, May 2000.

Stephanie Strassel. *Recent Developments in Annotated Speech Corpora at the Linguistic Data Consortium*. Invited speaker, Microsoft Corporation Speech Products Group, October 2000.

1999

Steven Bird and Stephanie Strassel. *Annotated Corpora in Linguistic Research*. North American Symposium on Corpora in Linguistics and Language Teaching, University of Michigan, May 1999.

Christopher Cieri, David Graff, Mark Liberman, Nii Martey, Stephanie Strassel. *The TDT-2 Text and Speech Corpus*. DARPA Broadcast News Workshop, Washington, DC, February 1999.

Stephanie Strassel and Christopher Cieri. *Corpus Sociolinguistics: Issues, Data and Tools*. NNAV 28, Toronto, October 1999.

Stephanie Strassel. *Corpus Creation and Quality Control at the LDC*. Invited speaker, Corpus of Spoken Dutch Workshop, Tilburg, November 1999.

1998

Kevin Dougherty and Stephanie Strassel. *A New Look at Variation in and Perception of American English Quotatives*. NWAV 27, University of Georgia, October 1998.

Stephanie Strassel. *The Phonology of Variable English Flapping*. NWAV 27, University of Georgia, October 1998.

Stephanie Strassel. *Variation in American English flapping*. In Paradis, Claude, Diane Vincent, Denise Deshaies, and Marty Laforest (eds.), *Papers in Sociolinguistics: NWAV 26 l'Universite Laval*. Quebec: Nota Bene, 125-135.