

ANN BIES

Linguistic Data Consortium, University of Pennsylvania
3600 Market St., Suite 810, Philadelphia PA 19104
(215) 746-0245 bies@ldc.upenn.edu

EDUCATION

M.A. in Linguistics, University of Pennsylvania, 1996

Syntax and Discourse Factors in Early New High German: Evidence for Verb-Final Word Order.
Anthony Kroch, advisor. IRCS Technical Report 96-24

B.A. in English, Concentration in Medieval Studies, Certificate of Advanced Study in German,
Carleton College, 1989, *magna cum laude*

ANNOTATION AND TREEBANK EXPERIENCE

Senior Research Associate

Linguistic Data Consortium, University of Pennsylvania, 2002-present:

- Penn Arabic Treebank, English Treebank, Biomedical Information Extraction Project, English parallel of Chinese and Arabic Treebanks; Project manager and content developer for treebank annotator training (English and Arabic)
- Developing and adapting annotation systems, guidelines and procedures
- Contributing to project planning, reporting, tool development, papers, grants, data release packages
- Training, evaluating, monitoring and supervising annotators
- Managing English Treebank team
- Contributing to DEFT Entities Relations Events team, supervising lead annotators, developing guidelines and quality control annotation
- Contributing to AIDA team, ontology, guidelines, and annotation development

Consultant, FUSE English Treebank

CCLS, Columbia University, 2012-2013:

- Managing treebank team

Linguist and Treebank advisor, Penn Arabic Treebank

Linguistic Data Consortium / Institute for Research in Cognitive Science, University of Pennsylvania, 2001-2002:

- Developing and writing guidelines for Arabic part-of-speech and syntactic structure annotation
- Training, evaluating and monitoring annotators
- Adapting and developing annotation systems and procedures
- Assisting with project planning, reporting and tool development

Consultant, Penn Chinese and Korean Treebanks

Institute for Research in Cognitive Science, University of Pennsylvania, 1998-present:

- Consulted on guidelines and project planning
- Training annotators for associated English parallel text annotation

Language Analyst, Penn Parsed Corpus of Middle English

Linguistics Department, University of Pennsylvania, 1996-1998:

- Annotated historical text corpus for part-of-speech and syntactic structure

Annotator, Penn Treebank

Computer and Information Science Department, University of Pennsylvania, 1993-1996:

- Annotated text and speech corpora for part-of-speech, syntactic structure and dysfluency structure
- Developed and wrote guidelines for syntactic structure annotation
- Trained annotators

Research Assistant, Early New High German comparative syntax project

Linguistics Department, University of Pennsylvania, 1990-1993:

- Annotated historical text corpus for syntactic and discourse structure

- Developed annotation guidelines
- Carried out linguistic analysis of the corpus

OTHER WORK EXPERIENCE

Outreach Coordinator, Institute for Research in Cognitive Science, University of Pennsylvania, 1998-2002:

- Writing, producing and editing promotional and outreach materials
- Working with faculty as needed on grant proposals, project management, progress reports, final reports
- Creating and maintaining websites for IRCS and associated research projects (ACE, TIDES)

Special Projects Coordinator, Institute for Research in Cognitive Science, University of Pennsylvania, 1997-1998:

- Organized and ran summer undergraduate program, workshops and other special events
- Maintained websites for summer program and workshops
- Developed and maintained *IRCS Resource Manual*

PUBLICATIONS AND PRESENTATIONS

- “Simple Semantic Annotation and Situation Frames: Two Approaches to Basic Text Understanding in LORELEI.” Kira Griffitt, Jennifer Tracey, Ann Bies, Stephanie Strassel. In *Proceedings of LREC 2018: 11th Edition of the Language Resources and Evaluation Conference*.
- “Cross-Document, Cross-Language Event Coreference Annotation Using Event Hoppers.” Zhiyi Song, Ann Bies, Justin Mott, Xuansong Li, Stephanie Strassel, Christopher Caruso. In *Proceedings of LREC 2018: 11th Edition of the Language Resources and Evaluation Conference*.
- “Overview of Linguistic Resources for the TAC KBP 2016 Evaluations: Methodologies and Results.” Joe Ellis, Jeremy Getman, Neil Kuster, Zhiyi Song, Ann Bies, & Stephanie M. Strassel. In *Proceedings of TAC KBP 2016 Workshop*, National Institute of Standards and Technology, Gaithersburg, MD, November 14-15.
- “Situational Awareness for Low Resource Languages: the LORELEI Situation Frame Annotation Task.” Stephanie M. Strassel, Ann Bies, Jennifer Tracey. First workshop on: *Exploitation of Social Media for Emergency Relief and Preparedness (SMERP)*, 9th April, 2017. Co-located with: ECIR 2017, Aberdeen, Scotland UK.
- “A Comparison of Event Representations in DEFT.” Ann Bies, Zhiyi Song, Jeremy Getman, Joe Ellis, Justin Mott, Stephanie Strassel, Martha Palmer, Teruko Mitamura, Marjorie Freedman, Heng Ji, Tim O’Gorman. The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2016). *4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. 2016.
- “Event Nugget and Event Coreference Annotation.” Zhiyi Song, Ann Bies, Stephanie Strassel, Joe Ellis, Teruko Mitamura, Hoa Dong, Yukari Yamakawa, Sue Holm. The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2016). *4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. 2016.
- “Large Multi-lingual, Multi-level and Multi-genre Annotation Corpus.” Xuansong Li, Martha Palmer, Nianwen Xue, Lance Ramshaw, Mohamed Maamouri, Ann Bies, Kathryn Conger, Stephen Grimes and Stephanie Strassel. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 2016.
- “Rapid Development of Morphological Analyzers for Typologically Diverse Languages.” Seth Kulick and Ann Bies. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 2016.
- “Parallel Chinese-English Entities, Relations and Events Corpora.” Justin Mott, Ann Bies, Zhiyi Song and Stephanie Strassel. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 2016.

- “Overview of Linguistic Resources for the TAC KBP 2015 Evaluations: Methodologies and Results.” Joe Ellis, Jeremy Getman, Dana Fore, Neil Kuster, Zhiyi Song, Ann Bies, Stephanie Strassel. In *Proceedings of TAC KBP 2015 Workshop*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, November 16-17, 2015.
- “Balancing the Existing and the New in the Context of Annotating Non-Canonical Language.” Ann Bies. *The 9th Linguistic Annotation Workshop (LAW IX)*, at the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015), invited paper. 2015.
- “Spanish Treebank Annotation of Informal Non-Standard Web Text.” Mariona Taulé, M Antonia Martí, Ann Bies, Aina Garí, Montserrat Nofre, Zhiyi Song, Stephanie Strassel and Joe Ellis. *1st International Workshop on Natural Language Processing for Informal Text (NLPIT 2015)*, at the International Conference on Web Engineering (ICWE 2015). 2015.
- “From Light to Rich ERE: Annotation of Entities, Relations, and Events.” Zhiyi Song, Ann Bies, Tom Riese, Justin Mott, Jonathan Wright, Seth Kulick, Neville Ryant, Stephanie Strassel, Xiaoyi Ma. *3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, at the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015). 2015.
- “Event Nugget Annotation: Processes and Issues.” Teruko Mitamura, Yukari Yamakawa, Sue Holm, Zhiyi Song, Ann Bies, Seth Kulick, Stephanie Strassel. *3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, at the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015). 2015.
- “Transliteration of Arabizi into Arabic Orthography: Developing a Parallel Annotated Arabizi-Arabic Script SMS/Chat Corpus.” Ann Bies, Zhiyi Song, Mohamed Maamouri, Stephen Grimes, Haejoong Lee, Jonathan Wright, Stephanie Strassel, Nizar Habash, Ramy Eskander, Owen Rambow. *Arabic Natural Language Processing Workshop*, EMNLP, Doha, Qatar. 2014.
- “Parser Evaluation Using Derivation Trees: A Complement to evalb.” Seth Kulick, Ann Bies, Justin Mott, Anthony Kroch, Beatrice Santorini and Mark Liberman. *The 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*. 2014.
- “Inter-annotator Agreement for ERE annotation.” Seth Kulick, Ann Bies, Justin Mott. *The 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation* at the 52nd Annual Meeting of the Association for Computational Linguistics (ACL). 2014.
- “Incorporating Alternate Translations into English Translation Treebank.” Ann Bies, Justin Mott, Seth Kulick, Jennifer Garland, Colin Warner. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. 2014.
- “Developing an Egyptian Arabic Treebank: Impact of Dialectal Morphology on Annotation and Tool Development.” Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Nizar Habash, and Ramy Eskander. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*. 2014.
- “Automatic Correction and Extension of Morphological Annotations.” Ramy Eskander, Nizar Habash, Ann Bies, Seth Kulick, Mohamed Maamouri. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, pages 1–10, Sofia, Bulgaria, August 8-9, 2013. © 2013 Association for Computational Linguistics.
- “Using Derivation Trees for Informative Treebank Inter-Annotator Agreement.” Seth Kulick, Ann Bies, Justin Mott, Mohamed Maamouri, Beatrice Santorini, Anthony Kroch. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2013)*. 2013.
- “Linguistic resources for Arabic machine translation: The Linguistic Data Consortium (LDC) catalog.” Ann Bies, Mohamed Maamouri, Denise DiPersio. In Soudi, Abdelhadi, Ali Farghaly, Günter Neumann and Rabih Zbib (Eds.), *Challenges for Arabic Machine Translation [NLP 9]*. John Benjamins Publishing Company. 2012.
- “Expanding Arabic Treebank to Speech: Results from Broadcast News.” Mohamed Maamouri, Ann Bies, Seth Kulick. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. 2012.

- “Further Developments in Treebank Error Detection Using Derivation Trees.” Seth Kulick, Ann Bies, Justin Mott. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. 2012.
- “Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures.” Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Mohamed Maamouri, Ann Bies, Nianwen Xue. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*. 2012.
- “Using Supertags and Encoded Annotation Principles for Improved Dependency to Phrase Structure Conversion.” Seth Kulick, Ann Bies, Justin Mott. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. 2012.
- “Using Derivation Trees for Treebank Error Detection.” Seth Kulick, Ann Bies, and Justin Mott. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2011.
- “Arabic Treebanking.” Mohamed Maamouri, Ann Bies, Seth Kulick, Nizar Habash, Reem Faraj and Ryan Roth. In Joseph Olive, Caitlin Christianson and John McCary (Eds.), *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*. Springer. 2011.
- “Parallel Aligned Treebank Corpora at LDC: Methodology, Annotation and Integration.” Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Xiaoyi Ma, Niyu Ge, Ann Bies, Nianwen Xue, Mohamed Maamouri. In *Proceedings of the Workshop on Annotation and Exploitation of Parallel Corpora (AEPC)*. 2010.
- “The Penn Arabic Tree Bank.” Mohamed Maamouri and Ann Bies. In Ali Farghaly (Ed.), *Arabic Computational Linguistics*. CSLI Studies in Computational Linguistics. 2010.
- “A TAG-derived Database for Treebank Search and Parser Analysis.” Seth Kulick and Ann Bies. In *Proceedings of TAG+10: The 10th International Conference on Tree Adjoining Grammars and Related Formalisms (TAG+10)*. 2010.
- “A Treebank Query System Based on an Extracted Tree Grammar.” Seth Kulick and Ann Bies. In *Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2010)*.
- “Consistent and Flexible Integration of Morphological Annotation in the Arabic Treebank.” Seth Kulick, Ann Bies, Mohamed Maamouri. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*.
- “From Speech to Trees: Applying Treebank Annotation to Arabic Broadcast News.” Mohamed Maamouri, Ann Bies, Seth Kulick, Wajdi Zaghouni, David Graff, Mike Ciul. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*.
- “Creating Arabic Parallel Word-Aligned Treebank Corpora at LDC.” Stephen Grimes, Xuansong Li, Ann Bies, Seth Kulick, Xiaoyi Ma, Stephanie Strassel. In *Proceedings of the Workshop on Language Resources (LRs) and Human Language Technologies (HLT) for Semitic Languages Status, Updates, and Prospects*, held in conjunction with the *Seventh International Conference on Language Resources and Evaluation (LREC 2010)*.
- “Exploiting Closed-Class Categories for Arabic Segmentation.” Seth Kulick, Ann Bies, and Mohamed Maamouri. Georgetown University Round Table 2010: Arabic Language and Linguistics. March 12-14, 2010.
- “A Quantitative Analysis of the Idafa Construction in the Arabic Treebank.” Seth Kulick, Ann Bies, and Mohamed Maamouri. Georgetown University Round Table 2010: Arabic Language and Linguistics. March 12-14, 2010.
- “Treebank Analysis and Search Using an Extracted Tree Grammar.” Seth Kulick and Ann Bies. In *Proceedings of Treebanks and Linguistic Theories 8 (TLT8)*. Milan, Italy, December 3-5, 2009.
- “Creating a Methodology for Large-Scale Correction of Treebank Annotation: The Case of the Arabic Treebank.” Mohamed Maamouri, Ann Bies and Seth Kulick. In *Proceedings of MEDAR 2nd International Conference on Arabic Language Resources and Tools*, April 22-23. 2009.

- “Diacritic Annotation in the Arabic Treebank and its Impact on Parser Evaluation.” Mohamed Maamouri, Seth Kulick, Ann Bies. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*.
- “Enhancing the Arabic Treebank: A Collaborative Effort toward New Annotation Guidelines.” Mohamed Maamouri, Ann Bies, Seth Kulick. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*.
- “A Pilot Arabic PropBank.” Mona Diab, Ann Bies, Mohamed Maamouri, Wajdi Zaghrouani, Olga Babko-Malaya, Aous Mansouri, Martha Palmer. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*.
- “Enhanced Annotation and Parsing of the Arabic Treebank.” Mohamed Maamouri, Ann Bies and Seth Kulick. In *Proceedings of the 6th International Conference on Informatics and Systems (INFOS2008)*.
- Arabic Treebank Morphological and Syntactic Annotation Guidelines*. Arabic Treebank Project. Linguistic Data Consortium, University of Pennsylvania. <http://projects ldc.upenn.edu/ArabicTreebank> (2008).
- “Diacritization: A Challenge to Arabic Treebank Annotation and Parsing.” Mohamed Maamouri, Ann Bies and Seth Kulick. *Machine Translation SIG of the British Computer Society Conference*, 2006.
- “Issues in Synchronizing the English Treebank and PropBank.” Olga Babko-Malaya, Ann Bies, Ann Taylor, Szuting Yi, Martha Palmer, Mitch Marcus, Seth Kulick and Libin Shen. In *Proceedings of Frontiers in Linguistically Annotated Corpora Workshop 2006*, Association for Computational Linguistics (Coling-ACL 2006).
- “Linguistic Resources for Speech Parsing.” Ann Bies, Stephanie Strassel, Haejoong Lee, Kazuaki Maeda, Seth Kulick, Yang Liu, Mary Harper, Matthew Lease. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*.
- “Developing and Using a Pilot Dialectal Arabic Treebank.” Mohamed Maamouri, Ann Bies, Tim Buckwalter, Mona Diab, Nizar Habash, Owen Rambow, Dalila Tabessi. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*.
- “Automated information extraction from biomedical text.” R. Scott Winters, Yang Jin, Ryan McDonald, Seth Kulick, Ann Bies, Mark A. Mandel, Eric Pancoast, Fernando C. N. Pereira, Mark Y. Liberman, Peter S. White. *Transactions of the Integrated Bio-medical Informatics and Enabling Technologies Symposium (TIBETS)*, in press (2005).
- “Parallel Entity and Treebank Annotation,” Ann Bies, Seth Kulick and Mark Mandel. *ACL2005 Conference Workshop: Frontiers in Corpus Annotation II*, 2005.
- “The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus,” Mohamed Maamouri, Ann Bies, Tim Buckwalter and Wigdan Mekki. In *Proceedings of NEMLAR 2004*.
- “Developing an Arabic Treebank: Methods, Guidelines, Procedures, and Tools,” Mohamed Maamouri and Ann Bies. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*.
- Addendum to the Penn Treebank II Style Bracketing Guidelines: BioMedical Treebank Annotation*. Colin Warner, Ann Bies, Christine Brisson, and Justin Mott. Ms., November 2004.
- “Integrated Annotation for Biomedical Information Extraction,” Seth Kulick, Ann Bies, Mark Liberman, Mark Mandel, Ryan McDonald, Martha Palmer, Andrew Schein, Lyle Ungar, Pete White and Scott Winters. *BioLINK 2004: Linking Biological Literature, Ontologies, and Databases*. Association for Computational Linguistics. pp. 61-68, 2004.
- Penn Arabic Treebank Guidelines*, Ann Bies and Mohamed Maamouri. Ms., Linguistic Data Consortium, University of Pennsylvania, 2003.
- “Using rich dependency structures as the basis of Korean/English Machine Translation,” Martha Palmer, Chung-hye Han, Anoop Sarkar, Ann Bies. Ms., University of Pennsylvania, 2002.
- PropBank Annotation Guidelines*, Paul Kingsbury, Martha Palmer and Ann Bies. Ms., University of Pennsylvania, 2001.

- Bracketing Guidelines for Treebank II Style*, Eds: Ann Bies, Mark Ferguson, Karen Katz, Robert MacIntyre, Penn Treebank Project, University of Pennsylvania, CIS Technical Report MS-CIS-95-06, 1995.
- “The Penn Treebank: Annotating predicate argument structure,” M. Marcus, G. Kim, M. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz & B. Schasberger, *Proceedings of the Human Language Technology Workshop*, San Francisco, 1994.
- “NP postposition in Early New High German: VO word order as focus movement,” *Proceedings of the Interdisciplinary Conference on Focus and Natural Language Processing in Celebration of the 10th Anniversary of the Journal of Semantics*, 1994.
- “Free Relatives and Case in Modern Greek,” *Georgetown University Round Table on Languages and Linguistics (GURT) Pre-session: Issues in Greek Linguistics*, Georgetown University, March 12, 1994.
- “Extrapolation in the Old English of *Beowulf*,” *New Ways of Analyzing Variation (NWAY) XX*, Georgetown University, October 3-5, 1991.

INVITED TALKS

- Syntactic Annotation of Non-Canonical Language*. Linguistic Annotation Workshop (LAW) Discussion Panel 2015.
- Treebank Analysis Using Derivation Trees*. Seth Kulick, Ann Bies, Anthony Kroch, Mark Liberman, Justin Mott, Beatrice Santorini. CLUNCH talk. March 27, 2014.
- Derivation Trees with Reduced Elementary Trees for Treebank Analysis and Parsing*. Seth Kulick, Ann Bies, Justin Mott. CLUNCH talk. February 5, 2013.
- The Penn Arabic Treebank*. Mohamed Maamouri, Ann Bies, Seth Kulick. CLUNCH talk. December 6, 2011.
- Google English Web Treebank*. Ann Bies. LDC Lunch Series. October 28, 2011.
- Parallel Aligned Treebanks*. Xuansong Li, Mohamed Maamouri, Ann Bies. BOLT Kickoff. October 24, 2011.
- Treebank Consistency: A TAG-based Examination of Locality and Context*. Seth Kulick, Ann Bies, Justin Mott. CLUNCH talk. November 4, 2010.
- An Introduction to the Penn Arabic Treebank and the Pilot Arabic PropBank*. Ann Bies. IBM T.J. Watson Research Center, Yorktown Heights, NY. November 3, 2006.
- An Introduction to Penn Treebanking*. Short course in syntax and treebanking. Ann Bies. University of Colorado at Boulder. June 15-21, 2006.

DATA PUBLICATIONS

- BOLT Egyptian Arabic Treebank - Discussion Forum*, CatalogID: LDC2018T23, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- Arabic Treebank – Weblog*, CatalogID: LDC2016T02, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- DEFT Spanish Treebank*, CatalogID: LDC218T01, Linguistic Data Consortium. Authors: Mariona Taulé, Maria Antonia Martí, Ann Bies, Aina Garí, Montserrat Nofre, Zhiyi Song, Stephanie Strassel, Joe Ellis.
- English News Text Treebank: Penn Treebank Revised*, CatalogID: LDC2015T13, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner.
- BOLT Phase 3 English Treebank CTS Part 4 – EATB*, CatalogID: LDC2015E51, Linguistic Data Consortium.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 8 V2.0*, CatalogID: LDC2015E26, Linguistic Data Consortium.
- DEFT Spanish Treebank V1*, CatalogID: LDC2015E66, Linguistic Data Consortium.
- DEFT International Spanish NW Treebank V2.0*, CatalogID: LDC2014E130, Linguistic Data Consortium.
- DEFT International Spanish NW Treebank*, CatalogID: LDC2014E130, Linguistic Data Consortium.

- DEFT DF Data for Spanish Treebank Annotation*, CatalogID: LDC2014R71, Linguistic Data Consortium.
- DEFT NW Data for Spanish Treebank Annotation V2.0*, CatalogID: LDC2014R62, Linguistic Data Consortium.
- DEFT NW Data for Spanish Treebank Annotation*, CatalogID: LDC2014R62, Linguistic Data Consortium.
- BOLT Phase 3 English Treebank CTS Part 3 – EATB*, CatalogID: LDC2015E30, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 3 English Treebank CTS Part 2 – ECTB*, CatalogID: LDC2015E25, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 3 English Treebank CTS Part 1 – ECTB*, CatalogID: LDC2015E15, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 2 English Treebank SMS/Chat Part 6 – EATB*, CatalogID: LDC2014E126, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 2 English Treebank SMS/Chat Part 5 – EATB*, CatalogID: LDC2014E107, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 2 English Treebank SMS/Chat Part 4 – ECTB*, CatalogID: LDC2014E78, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 2 English Treebank SMS/Chat Part 3 – ECTB*, CatalogID: LDC2014E44, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 8 V1.0*, CatalogID: LDC2015E26, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 3 Egyptian Arabic Treebank CTS Part 3 V2.0*, CatalogID: LDC2015E16, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 3 Egyptian Arabic Treebank CTS Part 3 V1.0*, CatalogID: LDC2015E16, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 3 Egyptian Arabic Treebank CTS Part 2 V2.0*, CatalogID: LDC2015E04, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 3 Egyptian Arabic Treebank CTS Part 2 V1.0*, CatalogID: LDC2015E04, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 3 Egyptian Arabic Treebank CTS Part 1 V2.0*, CatalogID: LDC2014E120, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 3 Egyptian Arabic Treebank CTS Part 1 V1.0*, CatalogID: LDC2014E120, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 7 V2.0*, CatalogID: LDC2014E95, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 7 V1.0*, CatalogID: LDC2014E95, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 6 V2.0*, CatalogID: LDC2014E77, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 6 V1.0*, CatalogID: LDC2014E77, Release date: September 16, 2014. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.

- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 5 V2.0*, CatalogID: LDC2014E63, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 5 V1.0*, CatalogID: LDC2014E63, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 4 V2.0*, CatalogID: LDC2014E43, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 4 V1.0*, CatalogID: LDC2014E43, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 3 V2.0*, CatalogID: LDC2014E17, Linguistic Data Consortium. Authors: Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 3 V1.0*. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2014E17.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 2 V2.0*, Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E133.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 2 V1.0*. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E133.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 1 V2.0*. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E120.
- BOLT Phase 2 Egyptian Arabic Treebank SMS/Chat Part 1 V1.0*. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E120
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 8 V2.0*. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E21.
- BOLT Phase 2 English Treebank SMS/Chat Part 2*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2014E03.
- BOLT Phase 2 English Treebank SMS/Chat Part 1*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2013E127.
- BOLT Phase 1 English Treebank DF Part 7 V1.0 – ECTB*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2013E76.
- BOLT Phase 1 English Treebank DF Part 6 V1.0 – ECTB*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2013E50.
- BOLT Phase 1 English Treebank DF Part 5 V1.0*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2013E40.
- Google WSJ Treebank Update V1.0*. Ann Bies, Justin Mott, Colin Warner. *BOLT Phase 1 English Treebank DF Part 5 V1.0*, CatalogID: LDC2013E40, Linguistic Data Consortium. Authors: Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2013R26.
- QuestionBank with Manually Revised Treebank Annotation V1.0*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2012R121.
- BOLT Phase 1 English Treebank DF Part 4 V1.0*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2013E17.
- Arabic Treebank - Broadcast News V1.0*. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012T07.
- English Web Treebank*. 2012. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2012T13.
- English Translation Treebank: An-Nahar Newswire*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2012T02.
- BOLT Phase 1 English Treebank DF Part 3 V1.0*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2012E114.

- BOLT Phase 1 English Treebank DF Part 2 V1.0.* Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2012E97.
- BOLT Phase 1 English Treebank DF Part 1 V1.0.* Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2012E92.
- Google Web Treebank with Coreference Nodes V1.0.* Ann Bies, Justin Mott, Colin Warner. LDC2012R43.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 8 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E21.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 8 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E21.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 7 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E12.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 7 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2013E12.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 6 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E125.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 6 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E125.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 5 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E107.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 5 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E107.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 4 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E99.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 4 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E99.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 3 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E89.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 3 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E89.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 2 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E98.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 2 V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E88.
- BOLT Phase 1 Egyptian Arabic Treebank DF Part 1 V2.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E93.
- Arabic Treebank ARZ Part 1, V1.0.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E28.
- Arabic Treebank Part 20 V1.0 - BOLT Pilot ARZ Email, 2012.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul. LDC2012E25.
- Arabic Treebank Part 16 V2.0 - GALE Phase 5 BC.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul, Wajdi Zaghouni. LDC2011E116.
- Arabic Treebank Part 15 V2.0 - GALE Phase 5 WB.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul, Wajdi Zaghouni. LDC2012E10.
- Arabic Treebank Part 14 V2.0 - GALE Phase 5 BC.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul, Wajdi Zaghouni. LDC2011E52.
- Arabic Treebank Part 13 V1.0 - GALE Phase 4 WB.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Michael Ciul, Wajdi Zaghouni. LDC2011E18.
- Arabic Treebank Part 11 V2.0 - GALE Phase 4 WB.* Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Wajdi Zaghouni. LDC2011E16.

- Arabic Treebank Part 1 - V4.1 (EATB Part 5)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, Basma Bouziri, Wajdi Zaghouni. LDC2010T13.
- Arabic Treebank Part 16 V1.0 - GALE Phase 5 BC*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Wajdi Zaghouni. LDC2010E74.
- Arabic Treebank Part 14 V1.0 - GALE Phase 5 BC (EATB Part 13)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Wajdi Zaghouni. LDC2010E73.
- Arabic Treebank Part 12 V1.0 - GALE Phase 4 BN (EATB Part 12)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Wajdi Zaghouni. LDC2010E72.
- Arabic Treebank Part 9 V2.0 - GALE Phase 4 BN (EATB Part 10)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Wajdi Zaghouni. LDC2010E42.
- Arabic Treebank Part 8 V2.0 - GALE Phase 4 BN (EATB Part 9)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Sondos Krouna, Dalila Tabassi, Wajdi Zaghouni. LDC2010E41.
- Arabic Treebank Part 11 V1.0 - GALE Phase 4 WB (EATB Part 11)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Wajdi Zaghouni. LDC2010E50.
- LDC Standard Arabic Morphological Analyzer (SAMA) Version 3.1*. 2010. Mohamed Maamouri, Dave Graff, Basma Bouziri, Sondos Krouna, Ann Bies, Seth Kulick. LDC2010L01.
- English Translation Treebank, EATB Part 13 V1.1 - GALE Phase 5 BC*. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2011E81.
- English Translation Treebank -- EATB Part 12 v1.0 (POS) – GALE Phase 5 BN (ATB12)*. 2011. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2011E05.
- English Translation Treebank -- EATB Part 11 v1.0 – GALE Phase 5 WB (ATB11)*. 2010. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2010E71.
- English Translation Treebank -- EATB Part 10 v1.0 – GALE Phase 5 BN (ATB9)*. 2010. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2010E60.
- English Translation Treebank -- EATB Part 9 v1.0 – GALE Phase 4 BN (ATB8/10)*. 2010. Ann Bies, Justin Mott, Colin Warner, Seth Kulick. LDC2010E40.
- English Translation Treebank -- EATB Part 6 v2.0 (Annahar/ATB3)*. 2010. Ann Bies, Justin Mott, Colin Warner. LDC2010E21.
- English Translation Treebank -- EATB Part 5 v2.0 (AFP/ATB1)*. 2010. Ann Bies, Justin Mott, Colin Warner. LDC2010E20.
- English Translation Treebank Part 8 V1.0 - GALE Phase 4 BN (ATB7)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E115.
- English Translation Treebank Part 7 V1.0 - GALE Phase 4 dev09 (ATB6)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E109.
- English Translation Treebank -- Part 4 v1.0 (WB/ATB11)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E59.
- English Translation Treebank -- EATB Part 3 v2.0 (ATB5)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E55.
- English Translation Treebank -- ECTB Part 2 v2.1 (Sinorama)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E19.
- English Translation Treebank -- ECTB Part 2 v2.0 (Sinorama)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E19.
- English Translation Treebank -- ECTB Part 1 v2.0 (Xinhua)*. 2009. Ann Bies, Justin Mott, Colin Warner. LDC Catalog ID: LDC2009E21.
- Arabic Treebank part 3 - v3.2 (EATB part 6)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, Basma Bouziri, Wajdi Zaghouni. LDC2010T08.
- Arabic Treebank Part 10 V1.0 - GALE Phase 4 BN (EATB part 11)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Sondos Krouna, Wajdi Zaghouni. LDC2010E22.
- Arabic Treebank Part 9 V1.0 - GALE Phase 4 BN (EATB part 10)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Wajdi Zaghouni. LDC2010E19.

- Arabic Treebank Part 8 V1.0 - GALE Phase 4 BN (EATB part 9)*. 2010. Mohamed Maamouri, Ann Bies, Seth Kulick, Wajdi Zaghouani. LDC2010E11.
- Arabic Treebank Part 7 V1.0 - GALE Phase 4 BN (EATB part 8)*. 2009. Mohamed Maamouri, Ann Bies, Seth Kulick. LDC Catalog ID: LDC2009E114.
- Arabic Treebank Part 6 V1.0 - GALE Phase 4 dev09 (EATB part 7)*. 2009. Mohamed Maamouri, Ann Bies, Seth Kulick. LDC Catalog ID: LDC2009E108.
- Arabic Treebank part 5 - v1.0 (EATB part 3)*. 2009. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche. LDC Catalog ID: LDC2009E72.
- Arabic Treebank part 3 - v3.1*. 2009. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, Basma Bouziri. LDC Catalog ID: LDC2008E22.
- Arabic Treebank part 2 - v3.0*. 2009. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, Basma Bouziri. LDC Catalog ID: LDC2008E62.
- Arabic Treebank part 1 - v4.0*. 2008. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki, Sondos Krouna, Basma Bouziri. LDC Catalog ID: LDC2008E61.
- Arabic Treebank: Part 3(b) v. 2.7*. To appear, 2008. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki.
- Arabic Treebank: Part 3(a) v. 2.6*. 2007. Mohamed Maamouri, Ann Bies, Seth Kulick, Fatma Gaddeche, Wigdan Mekki. LDC Catalog ID: LDC2007E65.
- English Chinese Translation Treebank v 1.0 (Xinhua)*, 2007. Ann Bies, Martha Palmer, Justin Mott, Colin Warner. LDC Catalog No.: LDC2007T02.
- English Newswire Translation Treebank, Version 1.0 (Annahar)*, 2006. Ann Bies. LDC Catalog No.: LDC2006E36.
- Arabic Broadcast News Treebank, Version 1.0: VOA Subcorpus*, 2006. Mohamed Maamouri, Ann Bies, Tim Buckwalter, Hubert Jin, Wigdan Mekki. LDC Catalog No.: LDC2006E35.
- English/Chinese Treebank 2.0 (Sinorama)*, 2006. LDC Catalog No. pending.
- English Newswire Translation Treebank, Version 1.0*, 2005. Ann Bies. LDC Catalog No.: LDC2005E85.
- Arabic Broadcast News Treebank, Version 1.0: ACE and VOA Subcorpora*, 2005. Mohamed Maamouri, Ann Bies, Tim Buckwalter, Hubert Jin. LDC Catalog Nos. pending.
- JHU Speech Parsing Workshop English Treebank Data v1.7*, 2005. LDC Catalog No.: LDC2005E15.
- Levantine Arabic Treebank v1.0*, 2005. LDC Catalog No. pending.
- Arabic Treebank: Part 3 v. 2.0*, 2005. Mohamed Maamouri, Ann Bies, Hubert Jin, Tim Buckwalter. LDC Catalog No.: LDC2005T20.
- Mining the Bibliome: PennBioIE Release 0.9*, 2004. UPenn Biomedical Information Extraction Project. <http://bioie ldc.upenn.edu/>
- Arabic Treebank: Part 1 v. 3.0*, 2005. Mohamed Maamouri, Ann Bies, Hubert Jin, Tim Buckwalter. LDC Catalog No.: LDC2005T02.
- English/Arabic Treebank 1.0 (AFP)*, 2006. LDC Catalog No.: LDC2006T10.
- Arabic Treebank: Part 2 v 2.0*, 2004. Mohamed Maamouri, Ann Bies, Tim Buckwalter, Hubert Jin. LDC Catalog No.: LDC2004T02.
- Arabic Treebank: Part 1 - 10K-word English Translation*, 2003. Moussa Bamba, Mohamed Maamouri, Hubert Jin, Ann Bies, Xiaoyi Ma. LDC Catalog No.: LDC2003T07

PROFESSIONAL SERVICE

- Reviewer/Program Committee frequently for conferences such as ACL, NAACL-HLT, LREC, COLING, TLT, LAW, EventStory.
- Programme Committee, Workshop on Computational Approaches to Semitic Languages, EACL09. Athens, Greece, April 2009.
- Program Committee, Computational Approaches to Semitic Languages Workshop, Association for Computational Linguistics (ACL) 2007. Prague, Czech Republic.

Program Committee, Treebanks and Linguistic Theories (TLT) 2006 and Treebanking and Advanced Processing of Arabic (TAPA) 2006. Charles University, Prague, Czech Republic.

NONPROFIT SERVICE

President, Board of Directors, Tempesta di Mare, Philadelphia Baroque Orchestra and Chamber Players.
November 2005 to September 2011.

Member, Board of Directors, Tempesta di Mare, June 2001 to November 2005, September 2011 to present.

LANGUAGES

- English – native speaker
- German – written and spoken competence
- Spanish, Russian, Serbian/Croatian – reading knowledge
- Modern Standard Arabic, Egyptian Arabic, Middle English, Old English, Early New High German, Modern Greek, Korean – annotation knowledge