

Curriculum Vitae

Dr. Amir Zeldes

Postal address: Department of Linguistics
Georgetown University
1421 37th St. NW
Poulton Hall 243
Washington, DC 20057
USA

E-Mail: amir.zeldes@georgetown.edu
Website: <https://gucorpling.org/amir>

Education

- 12/2007 – 1/2012 Doctorate in General Linguistics (*summa cum laude*) at Humboldt University (HU), Berlin. Supervisors: Prof. Anke Lüdeling, HU Berlin, Prof. Stefan Th. Gries, UC Santa Barbara. Thesis (2011): “Productivity in Argument Selection: A Usage-Based Approach to Lexical Choice in Syntactic Slots”.
- 10/2006 – 12/2007 Magister Artium (1st honors, *mit Auszeichnung*) at Humboldt University, Berlin (major in German Linguistics, minor in Comparative Historical Linguistics) and at Potsdam University (minor in Computational Linguistics). Thesis: “Data-Based Methods for Historical Grammar and Lexicon Extraction in a Diachronic Corpus”. Final grade of 1,0.
- 10/2005 – 9/2006 DAAD Exchange Scholarship Program, Humboldt University, Berlin, Institute for German Language and Linguistics. Research project: Application of Machine Translation Techniques to Historical Corpora.
- 2002 – 2005 B.A. (*summa cum laude*) at the Hebrew University of Jerusalem, Israel, Majors in General Linguistics and Cognitive Science. GPA of 97.42%.

Employment

- 8/2019 – present Associate Professor of Computational Linguistics, Department of Linguistics, Georgetown University, Washington, DC.
- 8/2014 – 7/2019 Assistant Professor of Computational Linguistics, Department of Linguistics, Georgetown University, Washington, DC.
- 3/2014 – 7/2014 Director of the BMBF eHumanities Junior Research Group KOMeT (corpus linguistics methods for e-Philology with TEL) at HU Berlin.
- 1/2012 – 2/2014 Post-doc at Collaborative Research Center 632 on Information Structure, Project D1 “Linguistic Database for Information Structure” (Prof. A. Lüdeling, HU Berlin and Prof. M. Stede, Potsdam University).
- 12/2007 – 1/2012 Doctoral researcher at Collaborative Research Center 632 on Information Structure, Project D1 “Linguistic Database for Information Structure” (Prof. A. Lüdeling, HU Berlin and Prof. M. Stede, Potsdam University).
- 10/2006 – 12/2007 Student Assistant at Humboldt University, Berlin, Institute fir German Language and Linguistics, Chair of Corpus Linguistics and Morphology (Prof. A. Lüdeling).

- 10/2004 – 7/2005 Student Assistant, Hebrew University of Jerusalem, Israel, Department of Egyptology and Ancient Near Eastern Studies (Prof. O. Goldwasser).
- 2/2004 – 7/2005 Research Assistant, Hebrew University of Jerusalem, Israel, Institute of Asian and African Studies (Prof. Em. P. Shinar).
- 10/2003 – 10/2004 Cataloguer of Arabic and Japanese Books, Israeli National Library, Jerusalem, Israel. Department of Asian and African Studies.

Awards

- 2021 DISRPT Shared Task 1st place in all 5 shared task tracks on discourse relation segmentation, classification and connective detection
- 2020 university-wide *Provost's Extraordinary Merit Award* (awarded)
- 2020 *Gerald M. Mara Faculty Mentoring Award for Graduate Student Mentoring* (nominee)
- DH Awards 2019 (<http://dhawards.org/dhawards2019/>), *Best DH Tool or Suite of Tools* (awarded, for the KELLIA project's Coptic Dictionary Online, see Grants)
- 2019 DISRPT Shared Task 1st place in the category 'Connective detection – treebanked', 2nd places in the categories 'EDU Segmentation - treebanked' and 'plain', as well as 'Connective detection - plain' at NAACL 2019.
- 2019 *Gerald M. Mara Faculty Mentoring Award for Graduate Student Mentoring* (nominee)

Grants

External funding agencies

- *Expanding Coptic Digital Online Collections* (PW-290519). A level 3 (Implementation grade) Humanities Collections and Reference Resources Grant funded by the NEH (with myself as co-PI, co-directed with Prof. Caroline T. Schroeder, University of Oklahoma). [2023-2026, \$349,887]
- *EU COST Action OC-2021-1-25588 UniDive: Universality, diversity and idiosyncrasy in language technology* (Secondary Proposer, PIs Agata Savary, Carlos Ramisch, Daniel Zeman and Joakim Nivre) [2022-2025, funding based on network costs]
- *A Linked Digital Environment for Coptic Studies: Integrating Heterogeneous Data with Machine Learning and Natural Language Processing* (HAA-261271-18). A Stage III Digital Humanities Advancement Grant funded by the NEH, (with myself as lead PI, co-directed with Prof. Caroline T. Schroeder, University of the Pacific). [2018-2022, \$348,267, incl. \$24,500 in matching funds]
- *KELLIA* (Koptische/Coptic Electronic Language and Literature International Alliance), Bilateral NEH/DFG DH grant (HG-229371). An NEH/DFG funded American/German bilateral project, co-PI with Prof. Caroline T. Schroeder, University of the Pacific, in collaboration with Prof. Heike Behlmer, Georg-August University of Göttingen. [2015-2018, \$192,500 for US partners] The Coptic Dictionary Online created in this project won the DH Awards 2019 first place (<http://dhawards.org/dhawards2019/>), in the category *Best DH Tool or Suite of Tools*
- *Coptic SCRIPTORIUM: Digitizing a Corpus for Interdisciplinary Research in Ancient Egyptian* – NEH Preservation and Access grant (PW5167214), co-PI with Prof. Caroline T. Schroeder (University of the Pacific). An NEH Digital Humanities project on digitizing Coptic texts from the first millennium for online search and visualization [2014-2016, \$40,000].
- *Coptic SCRIPTORIUM: A Corpus, Tools, and Methods for Corpus Linguistics and Computational Historical Research in Ancient Egypt*, ODH Start-Up grant (HD51907), co-PI with Prof. Caroline

T. Schroeder (University of the Pacific). An NEH Digital Humanities pilot grant for the creation of tools to process data in Coptic [2014-2016, \$60,000].

- *KOMeT – Corpus Linguistics Methods for e-Philology with TEI* – PI and director of a BMBF funded Junior Researcher Group (01UG1406) on Corpus Linguistics methods in the Digital Humanities at HU Berlin [2014, €80,000].
- *Textual Revision in Second Language Writing* – A joint DAAD project (#54368646) on English and German learner language in cooperation with City University of Hong Kong (together with Marc Reznicek, HU Berlin and Prof. John Lee, City University of Hong Kong, and hosted by Prof. Anke Lüdeling, HU Berlin) [2012-2013, €7,390]

Internal grants

- *Research Infrastructure Award 2024* – Georgetown University internal grant to fund high performance computing for LLM Research in Computational Linguistics [2024-2025, \$15,000]
- *Research Infrastructure Award 2019* – Georgetown University internal grant to acquire GPU computing resources for Computational Linguistics research and teaching [2019-2021, \$24,016]
- *Competitive Summer Academic Grant 2017* – Georgetown University internal summer grant to overhaul corpus resources at Georgetown University [2017, \$10,000]
- *Competitive Summer Academic Grant 2015* – Georgetown University internal summer grant to support writing a book proposal and sample chapters [2015, \$10,000]

Publications

Books

- [1] **Zeldes, Amir** (2018) *Multilayer Corpus Studies*. (Routledge Advances in Corpus Linguistics 22.) London: Routledge, 274 pp.
- [2] **Zeldes, Amir** (2012) *Productivity in Argument Selection. From Morphology to Syntax*. (Trends in Linguistics: Studies and Monographs 260.) Berlin & Boston: De Gruyter, xx+282 pp.

Edited Proceedings

- [3] Braud, Chloé, Liu, Yang Janet, Metheniti, Eleni, Muller, Philippe, Rivière, Laura, Rutherford, Attapol and **Zeldes, Amir** (eds.) (2023) *Proceedings of the 3rd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2023)*. Collocated with ACL 2023, 14 July 2023, Toronto, Canada.
- [4] **Zeldes, Amir**, Liu, Yang Janet, Iruskieta, Mikel, Muller, Philippe, Braud, Chloé and Badene, Sonia (eds.) (2021) *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*. Collocated with EMNLP 2021, 11 November 2021, Punta Cana, Dominican Republic.
- [5] Dipper, Stefanie and **Zeldes, Amir** (eds.) (2020) *Proceedings of the 14th Linguistic Annotation Workshop (The LAW XIV)*. Collocated with COLING 2020, 12 December 2020, Barcelona, Spain.
- [6] **Zeldes, Amir**, Das, Debopam, Galani, Erick Maziero, Antonio, Juliano Desiderato and Iruskieta, Mikel (eds.) (2019) *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*. Collocated with NAACL 2019, 6 June 2019, Minneapolis, MN.
- [7] **Zeldes, Amir** and Lüdeling, Anke (eds.) (2011) *Proceedings of Quantitative Investigations in Theoretical Linguistics 4 (QITL-4)*. 29-31 March 2011, Humboldt-Universität zu Berlin.

Journal Articles

- [8] **Zeldes, Amir** (2022) "Can we Fix the Scope for Coreference? Problems and Solutions for Benchmarks beyond OntoNotes". *Dialogue and Discourse* 13(1), 41–62.
- [9] Sanguinetti, Manuela, Cassidy, Lauren, Bosco, Cristina, Çetinoğlu, Özlem, Cignarella, Alessandra Teresa, Lynn, Teresa, Rehbein, Ines, Ruppenhofer, Josef, Seddah, Djamelé and **Zeldes, Amir** (2022) "Treebanking User-Generated Content: A UD Based Overview of Guidelines, Corpora and Unified Recommendations". *Language Resources and Evaluation*.
- [10] Schroeder, Caroline T. and **Zeldes, Amir** (2020) "A Collaborative Ecosystem for Digital Coptic Studies". *Journal of Data Mining and Digital Humanities*. Special Issue on Collecting, Preserving, and Disseminating Endangered Cultural Heritage for New Understandings through Multilingual Approaches, 1–9.
- [11] **Zeldes, Amir** and Liu, Yang (2020) "A Neural Approach to Discourse Relation Signal Detection". *Dialogue and Discourse* 11(2), 1-33.
- [12] **Zeldes, Amir** (2017) "The GUM Corpus: Creating Multilayer Resources in the Classroom". *Language Resources and Evaluation* 51(3), 581-612.
- [13] Gaeta, Livio and **Zeldes, Amir** (2017) "Between VP and NN: On the Constructional Types of German *-er* Compounds". *Constructions and Frames* 9(1), 1-40.
- [14] Spalek, Katharina* and **Zeldes, Amir*** (2017) "Converging Evidence for the Relevance of Alternative Sets: Data from NPs with Focus Sensitive Particles in German". *Language and Cognition* 9(1), 24-51. (* Joint first authors)
- [15] Odebrecht, Carolin, Belz, Malte, **Zeldes, Amir**, Lüdeling, Anke and Krause, Thomas (2017) "RIDGES Herbology - Designing a Diachronic Multi-Layer Corpus". *Language Resources and Evaluation* 51(3), 695-725.
- [16] Schroeder, Caroline T. and **Zeldes, Amir** (2016) "Raiders of the Lost Corpus". *Digital Humanities Quarterly* 10(2).
- [17] **Zeldes, Amir** (2016) "Probabilistic Pragmatics and Probabilistic Experience. A Response to Franke and Jäger". *Zeitschrift für Sprachwissenschaft [=German Journal of Linguistics]* 35(1), 109-116.
- [18] Krause, Thomas and **Zeldes, Amir** (2016) "ANNIS3: A New Architecture for Generic Corpus Query and Visualization". *Digital Scholarship in the Humanities* 31(1), 118-139.
- [19] Lee, John, Yeung, Chak Yan, **Zeldes, Amir**, Reznicek, Marc, Lüdeling, Anke and Webster, Jonathan (2015) "CityU Corpus of Essay Drafts of English Language Learners: A Corpus of Textual Revision in Second Language Writing". *Language Resources and Evaluation* 49(3), 659-683.
- [20] **Zeldes, Amir** and Schroeder, Caroline T. (2015) "Computational Methods for Coptic: Developing and Using Part-of-Speech Tagging for Digital Scholarship in the Humanities". *Digital Scholarship in the Humanities* 30(1), 164-176.
- [21] Romary, Laurent, **Zeldes, Amir** and Zipser, Florian (2015) "<tiger2/> - Serialising the ISO SynAF Syntactic Object Model". *Language Resources and Evaluation* 49(1), 1-18.
- [22] **Zeldes, Amir** (2013) "Productive Argument Selection: Is Lexical Semantics Enough?". *Corpus Linguistics and Linguistic Theory* 9(2), 263-291.
- [23] **Zeldes, Amir** (2013) "Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch [=Compounding as a network of constructions in advanced L2 German]". *Zeitschrift für germanistische Linguistik [=Journal of German Linguistics]* 41(2), 240-276.
- [24] **Zeldes, Amir** (2013) "Is Modern Hebrew Standard Average European? The View from European". *Linguistic Typology* 17(3), 439-470.

- [25] Hirschmann, Hagen, Lüdeling, Anke and **Zeldes, Amir** (2012) "Measuring and Coding Language Change: An Evolving Study in a Multilayer Corpus Architecture". *ACM Journal on Computing and Cultural Heritage (JOCCH)* 5(1), 4:1-4:16.
- [26] Chiarcos, Christian, Fiedler, Ines, Grubic, Mira, Haida, Andreas, Hartmann, Katharina, Ritz, Julia, Schwarz, Anne, **Zeldes, Amir** and Zimmermann, Malte (2011) "Information Structure in African Languages: Corpora and Tools". *Language Resources and Evaluation* 45(3), 361-374.
- [27] Crane, Gregory, Babeu, Alison, Bamman, David, Breuel, Thomas, Cerrato, Lisa, Deckers, Daniel, Lüdeling, Anke, Mimno, David, Singhal, Rashmi, Smith, David A. and **Zeldes, Amir** (2009) "Classics in the Million Book Library". *Digital Humanities Quarterly* 3(1).
- [28] Petrova, Svetlana, Solf, Michael, Ritz, Julia, Chiarcos, Christian and **Zeldes, Amir** (2009) "Building and Using a Richly Annotated Interlinear Diachronic Corpus: The Case of Old High German Tatian". *Traitement automatique des langues* 50(2), 47-71.
- [29] Lüdeling, Anke and **Zeldes, Amir** (2008) "Three Views on Corpora: Corpus Linguistics, Literary Computing, and Computational Linguistics". *Jahrbuch für Computerphilologie* [=Yearbook of Computational Philology] 9, 149-178.
- [30] **Zeldes, Amir** (2007) "Abstracting Suffixes: A Morphophonemic Approach to Polish Morphological Analysis". *Zeitschrift für Sprachwissenschaft* [=German Journal of Linguistics] 26(2), 347-370.

Peer Reviewed Book Chapters

- [31] **Zeldes, Amir** (2021) "Corpus Architecture". In: Paquot, Magali & Gries, Stefan Th. (eds.), *A Practical Handbook of Corpus Linguistics*. Heidelberg: Springer, 49–73.
- [32] **Zeldes, Amir** (2018) "The Case for Caseless Prepositional Constructions with *voller* in German". In: Boas, Hans C. and Ziem, Alexander (eds.), *Constructional Approaches to Syntactic Structures in German*. (Trends in Linguistics: Studies and Monographs.) Berlin: De Gruyter, 245-285.
- [33] **Zeldes, Amir** (2018) "Compounds and Productivity in Advanced L2 German Writing: A Constructional Approach". In: Ortega, Lourdes, Tyler, Andrea, Uno, Mariko and Park, Hae In (eds.), *Usage-inspired L2 Instruction: Researched Pedagogy*. Amsterdam: Benjamins, 237-265.
- [34] Lüdeling, Anke, Ritz, Julia, Stede, Manfred and **Zeldes, Amir** (2016) "Corpus Linguistics and Information Structure Research". In: Féry, Caroline and Ichihara, Shinichiro (eds.), *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press, 599-617.
- [35] Petrova, Svetlana and **Zeldes, Amir** (2015) "How exceptional is CP recursion in Germanic OV languages? Corpus-based evidence from Middle Low German". In: Gippert, Jost and Gehrke, Ralf (eds.), *Historical Corpora: Challenges and Perspectives*. (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 5.) Tübingen: Narr, 151-164.
- [36] **Zeldes, Amir** and Kanbar, Ghazwan (2014) "Arabisch und Hebräisch [=Arabic and Hebrew]". In: Krifka, Manfred, Błaszczak, Joanna, Leßmöllmann, Annette, Meinunger, André, Stiebels, Barbara, Tracy, Rosemarie and Truckenbrodt, Hubert (eds.), *Das mehrsprachige Klassenzimmer* [=The Multilingual Classroom]. Heidelberg: Springer, 135-174.
- [37] Hirschmann, Hagen, Lüdeling, Anke, Rehbein, Ines, Reznicek, Marc and **Zeldes, Amir** (2013) "Underuse of Syntactic Categories in Falko - A Case Study on Modification". In: Granger, Sylviane and Meunier, Fanny (eds.), *20 Years of Learner Corpus Research. Looking Back, Moving Ahead*. (Corpora and Language in Use.) Louvain: Presses universitaires de Louvain, 223-234.
- [38] Gaeta, Livio and **Zeldes, Amir** (2012) "Deutsche Komposita zwischen Syntax und Morphologie: Ein korpusbasierter Ansatz [=German Compounds between Syntax and Morphology: A Corpus Based Approach]". In: Gaeta, Livio and Schlücker, Barbara (eds.), *Das Deutsche als kompositionsfreudige Sprache: Strukturelle Eigenschaften und systembezogene*

Aspekte [=German as a Compound-Happy Language: Structural Properties and System-Related Aspects]. (Linguistik - Impulse & Tendenzen 46.) Berlin: De Gruyter, 197-217.

- [39] Lüdeling, Anke, Hirschmann, Hagen and **Zeldes, Amir** (2011) "Variationism and Underuse Statistics in the Analysis of the Development of Relative Clauses in German". In: Kawaguchi, Yuji, Minegishi, Makoto and Viereck, Wolfgang (eds.), *Corpus Analysis and Diachronic Linguistics*. (Tokyo University of Foreign Studies, Studies in Linguistics 3.) Amsterdam: John Benjamins, 37-57.

*ACL Conference and Related Workshop Papers

- [40] Zhu, Yilun, Peng, Siyao, Pradhan, Sameer, and **Zeldes, Amir** (2024) "SPICE: A Singleton-Enhanced Pipeline for Coreference Resolution". In: *Proceedings of LREC-COLING 2024*. Turin, Italy.
- [41] Braud, Chloé, **Zeldes, Amir**, Rivière, Laura, Liu, Yang Janet, Muller, Philippe, Sileo, Damien and Aoyama, Tatsuya (2024) "DISRPT: A Multilingual, Multi-domain, Cross-framework Benchmark for Discourse Processing". In: *Proceedings of LREC-COLING 2024*. Turin, Italy.
- [42] Weissweiler, Leonie, Böbel, Nina, Guiller, Kirian, Herrera, Santiago, Scivetti, Wesley, Lorenzi, Arthur, Melnik, Nurit, Bhatia, Archana, Schütze, Hinrich, Levin, Lori, **Zeldes, Amir**, Nivre, Joakim, Croft, William and Schneider, Nathan (2024) "Typologically-Informed Annotation of Constructions atop Universal Dependencies". In: *Proceedings of LREC-COLING 2024*. Turin, Italy.
- [43] Poesio, Massimo, Ogrodniczuk, Maciej, Ng, Vincent, Pradhan, Sameer, Yu, Juntao, Moosavi, Nafise, Paun, Silviu, **Zeldes, Amir**, Nedoluzhko, Anna, Novák, Michal, Popel Martin, and Žabokrtský, Zdeněk (2024) "Universal Anaphora: The First Three Years". In: *Proceedings of LREC-COLING 2024*. Turin, Italy.
- [44] Lin, Jessica, and **Zeldes, Amir** (2024) "GUMsley: Evaluating Entity Salience in Summarization for 12 English Genres". In: *Proceedings of EACL 2023*. St. Julian's, Malta, 2575–2588.
- [45] Zhu, Yilun, Peng, Siyao, Pradhan, Sameer, and **Zeldes, Amir** (2023) "Incorporating Singletons and Mention-based Features in Coreference Resolution via Multi-task Learning for Better Generalization". In: *Proceedings of IJCNLP-AACL 2023*, Bali, Indonesia, 121–130.
- [46] Liu, Yang Janet, Aoyama, Tatsuya and Zeldes, Amir (2023) "What's Hard in English RST Parsing? Predictive Models for Error Analysis". In: *Proceedings of SIGDIAL 2023*, Prague, 31–42.
- [47] Braud, Chloé, Liu, Yang Janet, Metheniti, Eleni, Muller, Philippe, Rivière, Laura, Rutherford, Attapol and **Zeldes, Amir** (2023) "The DISRPT 2023 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification". In: *Proceedings of the 3rd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2023)*, Canada, 1–21.
- [48] Aoyama, Tatsuya, Behzad, Shabnam, Gessler, Luke, Levine, Lauren, Lin, Jessica, Liu, Yang Janet, Peng, Siyao, Zhu, Yilun and **Zeldes, Amir** (2023) "GENTLE: A Genre-Diverse Multilayer Challenge Set for English NLP and Linguistic Evaluation". In: *Proceedings of the Seventeenth Linguistic Annotation Workshop (LAW-XVII 2023)*, 166–178. Toronto, Canada.
- [49] Behzad, Shabnam, Sakaguchi, Keisuke, Schneider, Nathan and **Zeldes, Amir** (2023) "ELQA: A Corpus of Metalinguistic Questions and Answers about English". In: *Proceedings of ACL 2023*, 2031–2047. Toronto, Canada.
- [50] Liu, Yang Janet and **Zeldes, Amir** (2023) "GUMSum: Multi-Genre Data and Evaluation for English Abstractive Summarization". In: *Findings of ACL 2023*, 9315–9327. Toronto, Canada.
- [51] Liu, Yang Janet and **Zeldes, Amir** (2023) "Why Can't Discourse Parsing Generalize? A Thorough Investigation of the Impact of Data Diversity". In: *Proceedings of EACL 2023*. Dubrovnik, Croatia.

- [52] **Zeldes, Amir** and Schneider, Nathan (2023) "Are UD Treebanks Getting More Consistent? A Report Card for English UD". In: *Proceedings of Universal Dependencies Workshop (UDW 2023)* co-located with GURT 2023. Washington, DC.
- [53] Gessler, Luke and **Zeldes, Amir** (2022) "MicroBERT: Effective Training of Low-resource Monolingual BERTs through Parameter Reduction and Multitask Learning". In: *Proceedings of MRL at EMNLP 2022*. Abu Dhabi, UAE.
- [54] **Zeldes, Amir**, Howell, Nick, Ordan, Noam and Ben Moshe, Yifat (2022) "A Second Wave of UD Hebrew Treebanking and Cross-Domain Parsing". In: *Proceedings of EMNLP 2022*. Abu Dhabi, UAE.
- [55] Peng, Siyao, Liu, Yang Janet and **Zeldes, Amir** (2022) "GCDT: A Chinese RST Treebank for Multigenre and Multilingual Discourse Parsing". In: *Proceedings of ACL-IJCNLP 2022*. Taipei, 382–391.
- [56] Nedoluzhko, Anna, Novák, Michal, Popel, Martin, Žabokrtský, Zdeněk, **Zeldes, Amir** and Zeman, Daniel (2022) "CorefUD 1.0: Coreference Meets Universal Dependencies". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022)*. Marseille, France, 4859–4872.
- [57] Gessler, Luke, Levine, Lauren and **Zeldes, Amir** (2022) "Midas Loop: A Prioritized Human-in-the-Loop Annotation for Large Scale Multilayer Data". In: *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI)*. Marseille, France, 103–110.
- [58] **Zeldes, Amir**, Liu, Yang Janet, Iruskieta, Mikel, Muller, Philippe, Braud, Chloé and Badene, Sonia (2021) "The DISRPT 2021 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification". In: *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*. Punta Cana, Dominican Republic, 1–12.
- [59] Gessler, Luke, Behzad, Shabnam, Liu, Yang Janet, Peng, Siyao, Zhu, Yilun and **Zeldes, Amir** (2021) "DisCoDisCo at the DISRPT2021 Shared Task: A System for Discourse Segmentation, Classification, and Connective Detection". In: *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*. Punta Cana, Dominican Republic, 51–62.
*** This paper one first place in all 5 tracks of the DISRPT 2021 Shared Task**
- [60] Zhu, Yilun, Pradhan, Sameer and **Zeldes, Amir** (2021) "Anatomy of OntoGUM – Adapting GUM to the OntoNotes Scheme to Evaluate Robustness of SOTA Coreference Algorithms". In: *Fourth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2021)*. Punta Cana, Dominican Republic, 141–149.
- [61] Lin, Jessica and **Zeldes, Amir** (2021) "WikiGUM: Exhaustive Entity Linking for Wikification in 12 Genres". In: *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*. Punta Cana, Dominican Republic, 170–175.
- [62] Zhu, Yilun, Pradhan, Sameer and **Zeldes, Amir** (2021) "OntoGUM: Evaluating Contextualized SOTA Coreference Resolution on 12 More Genres". In: *Proceedings of ACL-IJCNLP 2021*. Bangkok, Thailand, 461–467.
- [63] Manning, Emma, Schneider, Nathan and **Zeldes, Amir** (2021) "A Balanced and Broadly Targeted Computational Linguistics Curriculum". In: *Fifth Workshop on Teaching NLP at NAACL 2021*. Mexico City, Mexico, 65–69.
- [64] **Zeldes, Amir**, Martin, Lance and Tu, Sichang (2020) "Exhaustive Entity Recognition for Coptic: Challenges And Solutions". In: *Proceedings of the SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2020)*. Barcelona, Spain, 19–28.
- [65] Sanguinetti, Manuela, Bosco, Cristina, Cassidy, Lauren, Çetinoğlu, Özlem, Cignarella, Alessandra Teresa, Lynn, Teresa, Rehbein, Ines, Ruppenhofer, Josef, Seddah, Djamel and **Zeldes, Amir** (2020) "Treebanking User-Generated Content: A Proposal for a Unified

- Representation in Universal Dependencies". In: *Proceedings of LREC 2020*. Marseille, France, 5240–5250.
- [66] Gessler, Luke, Peng, Siyao, Liu, Yang, Zhu, Yilun, Behzad, Shabnam and **Zeldes, Amir** (2020) "AMALGUM – A Free, Balanced, Multilayer English Web Corpus". In: *Proceedings of LREC 2020*. Marseille, France, 5267-5275.
- [67] Sanguinetti, Manuela, Bosco, Cristina, Cassidy, Lauren, Çetinoğlu, Özlem, Cignarella, Alessandra Teresa, Lynn, Teresa, Rehbein, Ines, Ruppenhofer, Josef, Seddah, Djamé and **Zeldes, Amir** (2020) "Treebanking User-Generated Content: A Proposal for a Unified Representation in Universal Dependencies". In: *Proceedings of LREC 2020*. Marseille, France, 5240-5250.
- [68] Behzad, Shabnam and **Zeldes, Amir** (2020) "A Cross-Genre Ensemble Approach to Robust Reddit Part of Speech Tagging". In: *Proceedings of the 12th Web as Corpus Workshop (WAC-XII)*. Marseille, France, 50–56.
- [69] Slaughter, Laura, Morgado Da Costa, Luis, Miyagawa, So, Büchler, Marco, **Zeldes, Amir** and Behlmer, Heike (2019) "The Making of Coptic Wordnet". In: *Global Wordnet Conference (GWC 2019)*. Wrocław, Poland, 166-175.
- [70] **Zeldes, Amir**, Das, Debopam, Galani, Erick Maziero, Antonio, Juliano Desiderato and Iruskieta, Mikel (2019) "Introduction to Discourse Relation Parsing and Treebanking (DISRPT): 7th Workshop on Rhetorical Structure Theory and Related Formalisms". In: *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*. Minneapolis, MN, 1-6.
- [71] **Zeldes, Amir**, Das, Debopam, Galani, Erick Maziero, Antonio, Juliano Desiderato and Iruskieta, Mikel (2019) "The DISRPT 2019 Shared Task on Elementary Discourse Unit Segmentation and Connective Detection". In: *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*. Minneapolis, MN, 97-104.
- [72] Gessler, Luke, Liu, Yang and **Zeldes, Amir** (2019) "A Discourse Signal Annotation System for RST Trees". In: *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*. Minneapolis, MN, 56-61.
- [73] Yu, Yue, Zhu, Yilun, Liu, Yang, Liu, Yan, Peng, Siyao, Gong, Mackenzie and **Zeldes, Amir** (2019) "GumDrop at the DISRPT2019 Shared Task: A Model Stacking Approach to Discourse Unit Segmentation and Connective Detection". In: *Proceedings of Discourse Relation Treebanking and Parsing (DISRPT 2019)*. Minneapolis, MN, 133-143.
*** This paper one first place in 1 track and second place in 3 tracks of the DISRPT 2019 Shared Task**
- [74] **Zeldes, Amir** (2018) "A Characterwise Windowed Approach to Hebrew Morphological Segmentation". In: *Proceedings of the 15th SIGMORPHON Workshop at EMNLP 2018*. Brussels, Belgium.
- [75] **Zeldes, Amir** and Abrams, Mitchell (2018), "The Coptic Universal Dependency Treebank". In: *Proceedings of the 2018 Universal Dependencies Workshop (UDW 2018)*. Brussels, Belgium.
- [76] Peng, Siyao and **Zeldes, Amir** (2018), "All Roads Lead to UD: Converting Stanford and Penn Parses to English Universal Dependencies with Multilayer Annotations". In: *Proceedings of the XII Linguistic Annotation Workshop (LAW) at COLING2018*. Santa Fe, NM, 167-177. [This paper was also selected for **cross-presentation** at the 2018 Universal Dependencies Workshop at EMNLP in Brussels, Belgium]
- [77] Feder, Frank, Kupreyev, Maxim, Manning, Emma, Schroeder, Caroline and **Zeldes, Amir** (2018), "A Linked Coptic Dictionary Online". In: *Proceedings of the SIGHUM Workshop on Language Technology for Cultural Heritage (LaTeCH) at COLING2018*. Santa Fe, NM, 12-21.
- [78] **Zeldes, Amir** (2018) "A Predictive Model for Notional Anaphora in English". *NAACL 2018 Workshop on Computational Models of Reference, Anaphora, and Coreference (CRAC)*. New Orleans, LA, 34-43.

- [79] MacAvaney, Sean and **Zeldes, Amir** (2018) "A Deeper Look into Dependency-Based Word Embeddings". *NAACL 2018 Student Research Workshop*. New Orleans, LA, 40-45.
- [80] **Zeldes, Amir** (2017) "A Distributional View of Discourse Encapsulation: Multifactorial Prediction of Coreference Density in RST". *6th Workshop on Recent Advances in RST and Related Formalisms at INLG*. Santiago de Compostela, Spain, 20-28.
- [81] **Zeldes, Amir** and Simonson, Dan (2016) "Different Flavors of GUM: Evaluating Genre and Sentence Type Effects on Multilayer Corpus Annotation Quality". In: *Proceedings of LAW X - The 10th Linguistic Annotation Workshop at the Annual Meeting of the ACL*. Berlin, 68-78.
- [82] **Zeldes, Amir** and Schroeder, Caroline T. (2016) "An NLP Pipeline for Coptic". In: *Proceedings of LaTeCH 2016 - The 10th SIGHUM Workshop at the Annual Meeting of the ACL*. Berlin, 146-155.
- [83] **Zeldes, Amir** (2016) "rstWeb - A Browser-based Annotation Interface for Rhetorical Structure Theory and Discourse Relations". In: *Proceedings of NAACL-HLT 2016 System Demonstrations*. San Diego, CA, 1-5.
- [84] **Zeldes, Amir** and Zhang, Shuo (2016) "When Annotation Schemes Change Rules Help: A Configurable Approach to Coreference Resolution beyond OntoNotes". In: *Proceedings of the NAACL2016 Workshop on Coreference Resolution Beyond OntoNotes (CORBON)*. San Diego, CA, 92-101.
- [85] **Zeldes, Amir** (2009) "Quantifying Constructional Productivity with Unseen Slot Members". In: *Proceedings of the NAACL HLT Workshop on Computational Approaches to Linguistic Creativity*. Boulder, CO, 47-54.
- [86] Chiarcos, Christian, Fiedler, Ines, Grubic, Mira, Haida, Andreas, Hartmann, Katharina, Ritz, Julia, Schwarz, Anne, **Zeldes, Amir** and Zimmermann, Malte (2009) "Information Structure in African Languages: Corpora and Tools". In: *Proceedings of the EACL Workshop on Language Technologies for African Languages (AFLAT), 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*. Athens, Greece, 17-24.

Other Peer Reviewed Conference Papers

- [87] **Zeldes, Amir**, Schroeder, Caroline T., Wagner, Nicholas, Bremer-McCollum, Lydia and Takla, Hany (2024) "Digital Developments for Dialects in Coptic". In: *Proceedings of the Digital Humanities Conference 2024*. Washington, DC.
- [88] Gessler, Luke, Peng, Siyao, Liu, Yang, Zhu, Yilun, Behzad, Shabnam and **Zeldes, Amir** (2021) "Overview of AMALGUM - Large Silver Quality Annotations across English Genres". In: *Proceedings of the Society for Computation in Linguistics (SCiL)*. Vol. 4, 434-437.
- [89] Slaughter, Laura, Morgado Da Costa, Luis, Miyagawa, So, Büchler, Marco, Zeldes, Amir and Behlmer, Heike (2019) "The Making of Coptic Wordnet". In: *Global Wordnet Conference (GWC 2019)*. Wrocław, Poland.
- [90] Liu, Yang and **Zeldes, Amir** (2019) "Discourse Relations and Signaling Information: Anchoring Discourse Signals in RST-DT". In: *Proceedings of the Society for Computation in Linguistics (SCiL 2018)*. New York, NY.
- [91] Miyagawa, So, **Zeldes, Amir**, Behlmer, Heike and Büchler, Marco (2018). "Building Linguistically and Intertextually Tagged Coptic Corpora with Open Source Tools". *Proceedings of the Japanese Society of Digital Humanities (JADH)*. Tokyo, Japan.
- [92] Peng, Siyao and **Zeldes, Amir** (2018) "Validating and Merging a Growing Multilayer Corpus – the Case of GUM". *14th American Association of Corpus Linguistics Conference (AACL 2018)*. Atlanta, GA.
- [93] **Zeldes, Amir** (2018) "A Multi-Dimensional Analysis of RST Discourse Relations in Eight Genres". *14th American Association of Corpus Linguistics Conference (AACL 2018)*. Atlanta, GA.

- [94] Miyagawa, So and **Zeldes, Amir** (2018) "A Semantic Map of the Coptic Complementizer *čē*-based on Corpus Analysis: Grammaticalization and Areal Typology in Africa". *International Workshop on Semantic maps: Where do we stand and where are we going?* Liège, Belgium.
- [95] **Zeldes, Amir** (2018) "A Neural Approach to Discourse Relation Signaling". *Georgetown University Round Table (GURT) 2018: Approaches to Discourse*. Washington, DC.
- [96] Zhang, Shuo and **Zeldes, Amir** (2017) "GitDOX: A Linked Version Controlled Online XML Editor for Manuscript Transcription". In: *Proceedings of FLAIRS 2017, Special Track on Natural Language Processing of Ancient and other Low-resource Languages*. Marco Island, FL, 619-623.
- [97] **Zeldes, Amir** (2016) "A Quantitative Approach to Syntactic Alternations in Sahidic". In: *11th Congress of the International Association of Coptic Studies (IACS 2016)*. Claremont, CA.
- [98] **Zeldes, Amir** (2015) "Duplicitous Diabolos: Parallel witness encoding in quantitative studies of Coptic manuscripts". In: *Proceedings of Balisage, Symposium on Cultural Heritage Markup*. (Balisage Series on Markup Technologies, vol. 16.) Washington, DC.
- [99] **Zeldes, Amir** (2015) "Tagging the Desert Fathers: Part of Speech Analysis in Sahidic Coptic Corpora". *43rd Annual North American Conference on Afroasiatic Linguistics (NACAL2015)*. Washington, DC.
- [100] **Zeldes, Amir** (2014) "German *voller* as a Productive Argument Structure *sui generis*". In: *36th Annual Meeting of the DGfS, Workshop on Problems of Syntactic Categorisation*. Marburg, Germany.
- [101] Bosch, Sonja, Eckart, Kerstin, Faaß, Gertrud, Heid, Ulrich, Lee, Kiyong, Pareja-Lora, Antonio, Pretorius, Laurette, Romary, Laurent, Witt, Andreas, **Zeldes, Amir** and Zipser, Florian (2014) "From <tiger2> to ISOTiger - Community Driven Developments for Syntax Annotation in SynAF". In: Henrich, Verena, Hinrichs, Erhard, de Kok, Daniël, Osenova, Petya and Przepiórkowski, Adam (eds.), *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13)*. Tübingen, Germany 258-264.
- [102] Krause, Thomas, Lüdeling, Anke, Odebrecht, Carolin and **Zeldes, Amir** (2012) "Multiple Tokenizations in a Diachronic Corpus". In: *Exploring Ancient Languages through Corpora*. Oslo, Norway, 1-3.
- [103] Bosch, Sonja, Choi, Key-Sun, de la Clergerie, Éric, Chengyu Fang, Alex, Faass, Gertrud, Lee, Kiyong, Pareja-Lora, Antonio, Romary, Laurent, Witt, Andreas, **Zeldes, Amir** and Zipser, Florian (2012) "<tiger2> as a standardised serialisation for ISO 24615 - SynAF". In: Hendrick, Iris, Kübler, Sandra and Simov, Kiril (eds.), *TLT11 - 11th international workshop on Treebanks and Linguistic Theories*. Lisbon, Portugal, 37-60.
- [104] Krause, Thomas, Ritz, Julia, **Zeldes, Amir** and Zipser, Florian (2011) "Topological Fields, Constituents and Coreference: A New Multi-layer Architecture for TüBa-D/Z". In: Hedeland, Hanna, Schmidt, Thomas and Wörner, Kai (eds.), *Multilingual Resources and Multilingual Applications. Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011*. (Working Papers in Multilingualism 96.) Hamburg: Universität Hamburg, 259-262.
- [105] **Zeldes, Amir** (2011) "On the Productivity and Variability of the Slots in German Comparative Correlative Constructions". In: Konopka, Marek, Kubczak, Jacqueline, Mair, Christian, Štícha, František and Waßner, Ulrich H. (eds.), *Grammar & Corpora, Third International Conference, Mannheim, 22.-24.09.2009*. Tübingen: Narr, 429-449.
- [106] Zipser, Florian, **Zeldes, Amir**, Ritz, Julia, Romary, Laurent and Leser, Ulf (2011) "Pepper: Handling a Multiverse of Formats". In: *33. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*. Göttingen: University of Göttingen, 235.
- [107] Lüdeling, Anke, **Zeldes, Amir**, Reznicek, Marc, Rehbein, Ines and Hirschmann, Hagen (2010) "Syntactic Misuse, Overuse and Underuse: A Study of a Parsed Learner Corpus and its Target Hypothesis". In: Dickinson, Markus, Müürisep, Kaili and Passarotti, Marco (eds.), *Proceedings*

- of the Ninth International Workshop on Treebanks and Linguistic Theories (TLT-9). (NEALT Proceedings Series 9.) University of Tartu, Estonia, 1-3.
- [108] **Zeldes, Amir**, Ritz, Julia, Lüdeling, Anke and Chiarcos, Christian (2009) "ANNIS: A Search Tool for Multi-Layer Annotated Corpora". In: *Proceedings of Corpus Linguistics 2009*. Liverpool, UK.
- [109] **Zeldes, Amir**, Lüdeling, Anke and Hirschmann, Hagen (2008) "What's Hard? Quantitative Evidence for Difficult Constructions in German Learner Data". In: Arppe, Antti, Sinnemäki, Kaius and Nikanne, Urpo (eds.), *Proceedings of Quantitative Investigations in Theoretical Linguistics 3 (QITL-3)*. Helsinki, 74-77.
- [110] **Zeldes, Amir** (2007) "Machine Translation between Language Stages: Extracting Historical Grammar from a Parallel Diachronic Corpus of Polish". In: *Proceedings of Corpus Linguistics 2007*. Birmingham: University of Birmingham.
- [111] **Zeldes, Amir** (2006) "Abstracting Suffixes: A Morphophonemic Approach to Polish Morphological Analysis". In: *Proceedings of KONVENS'06*. Konstanz: University of Konstanz, 151-158.

Poster Presentations

- [112] Krause, Thomas, Weißenfels, Benjamin, **Zeldes, Amir** and Zipser, Florian (2014) "ANNIS3: Towards Generic Corpus Search and Visualization". *36th Annual Meeting of the DGfS*, Poster Session of the SIG on Computational Linguistics. Marburg, 409-410.
- [113] **Zeldes, Amir** (2012) "Novel Argument Realization: Semantic, Pragmatic and Conventional Productivity Effects". *Proceedings of Linguistic Evidence 2012. Empirical, Theoretical and Computational Perspectives*. Tübingen, 327-331 (*received the best poster award).
- [114] Gaeta, Livio and **Zeldes, Amir** (2011) "German Synthetic Compounds and the Architecture of the Grammar: A Behavioral Analysis". *Mediterranean Morphology Meeting 8*, 14-17 September 2011, Cagliari.
- [115] Krause, Thomas, Ritz, Julia, **Zeldes, Amir** and Zipser, Florian (2011) "Topological Fields, Constituents and Coreference: A New Multi-layer Architecture for TüBa-D/Z". *Conference of the German Society for Computational Linguistics and Language Technology (GSCL) 2011*. Hamburg.
- [116] Reznicek, Marc, Krummes, Cedric, Hirschmann, Hagen, Lüdeling, Anke, Ensslin, Astrid, Chan, Jia Wei, **Zeldes, Amir**, Krause, Thomas and Zipser, Florian (2010) "'Dass wenn man etwas will, muss man dafür arbeiten'- Zielhypothesen im Lernerkorpus Falko". *31. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*, Berlin, 25 February 2010.
- [117] **Zeldes, Amir** (2010) "Novel Argument Selection: Is Lexical Semantics Enough?". *Doktorandentag 2010*, Humboldt-Universität zu Berlin.
- [118] Chiarcos, Christian, Krause, Thomas, Lüdeling, Anke, Ritz, Julia, Rosenfeld, Viktor, Stede, Manfred, **Zeldes, Amir** and Zipser, Florian (2009) "Search and Visualization of Richly Annotated Corpora with ANNIS2". In: *31. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Sektion CL*. Osnabrück: University of Osnabrück, 265.
- [119] **Zeldes, Amir** (2009) "Applying Morphological Productivity Measures to Syntactic Constructions: German Comparatives and the *je ... desto* Constructions". *Doktorandentag 2009*, Humboldt-Universität zu Berlin.

Technical Reports

- [120] **Zeldes, Amir** and Martin, Lance (2020) *Coptic Scriptorium – Entity Annotation Guidelines*. Version 1.1.0-2020-07–20. Technical Report, Georgetown University.
- [121] **Zeldes, Amir** (2016) *Coptic Scriptorium Lemmatization Guidelines*. Version 1.0.5-2017-09-15. Technical Report, Georgetown University.

- [122] **Zeldes, Amir** (2016) *Coptic Universal Dependency Guidelines*. Version 1.1.0-2016-10-19. Technical Report, Georgetown University.
- [123] **Zeldes, Amir** and Schroeder, Caroline (2016) *SCRIPTORIUM Part-of-Speech Tagsets for Sahidic Coptic*. Version: 1.1.7_2016.03.14. Technical Report, Georgetown University and University of the Pacific.
- [124] **Zeldes, Amir**, Zipser, Florian and Neumann, Arne (2013) *PAULA XML Documentation: Format Version 1.1*. Technical report, Humboldt-Universität zu Berlin and Universität Potsdam.

Invited Talks

- [125] **Zeldes, Amir** and Howell, Nick (2022) "IAHLT NLP Innovations for Hebrew and Arabic". AI Data Science Summit 2022, College of Management, Rishon LeTsiyon, Israel, 10 May 2022.
- [126] **Zeldes, Amir** (2022) "Multilayer and Multi-Genre Models of Discourse. Data and Insights beyond the Sentence Level". Invited talk at Pittsburgh University NLP Seminar, 5 March 2022.
- [127] **Zeldes, Amir** (2022) "Multilayer and Multi-Genre Annotation of Natural Languages: Recent Advances and Applications". Invited talk at IAHLT Achievements and Innovations, Online, 3 January 2022.
- [128] **Zeldes, Amir** (2021) "UD Treebanking for Coptic DH. Low Resource NLP Technologies for NER, Lexicography and Linked Open Data". Invited talk at Digital Transformation in the Humanities, KUDH International Conference, Kyoto University, 2 October 2021.
- [129] **Zeldes, Amir**, Schroeder, Caroline T. and Martin, Lance (2021) "Leveraging Non-Named Entities in Coptic antiquity". Digital Classicist Seminar 2021, online, 10 September 2021.
- [130] **Zeldes, Amir** (2021) "Issues in UD Consistency, Phrases, and Semantics". Dagstuhl Seminar 21351, Universals of Linguistic Idiosyncrasy in Multilingual Computational Linguistics . Online, 30–31 August 2021.
- [131] **Zeldes, Amir** (2021) "A Second Wave of Gold Standard Hebrew Treebanks". *AI Data Science Summit 2021*, Amdocs Raanana, Israel, 3 August 2021.
- [132] Martin, Lance, Schroeder, Caroline T. and **Zeldes, Amir** (2021) "Digitally Linking People and Places in Coptic Literature". *St. Shenouda Foundation Annual Conference 2021*, UCLA, 16 July 2021.
- [133] **Zeldes, Amir** (2021) "IAHLT Treebanking & NLP – Progress report". *IAHLT Meeting, Israeli Association of Human Language Technology*. Online, 6 June 2021.
- [134] **Zeldes, Amir** and Peng, Siyao (2021) "Universal Anaphora - Collapsed format data tests and tricky coreference guidelines". *Universal Anaphora Meeting*. Online, 9 April 2021.
- [135] **Zeldes, Amir** (2021) "Building Blocks for Discourse Models". *Georgetown University Interdisciplinary Text Analysis Research (GUITAR) at the Massive Data Institute* (10 February 2021).
- [136] **Zeldes, Amir** (2020) "Coreference at Georgetown: GUM, ANNIS, and Amalgum". In: *Invited panel on Universal Anaphora, at CRAC 2020* (12 December 2020).
- [137] **Zeldes, Amir** (2020) "Multilayer GUM – Computational Discourse Models in the CL Classroom". Invited talk at the *DALI Seminar*, Queen Mary University London (10 November 2020).
- [138] **Zeldes, Amir** (2020) "Chewing GUM: On Constructing Multilayer Corpora in the Linguistics Classroom". Invited talk at *WocDit 2020, International Seminar Working with Corpora and Digital Tools in Language and Translation Studies: Practical Issues and Methodological Challenges*. Universidad Complutense Madrid (UCM) (27 October 2020).
- [139] **Zeldes, Amir**, Schroeder, Caroline T. and Martin, Lance (2020) "Exposing Coptic entities: automation, search and visualization". *Digital Coptic 3, 3rd workshop on Digital Methods for Coptic Studies*. Held online during Covid-19.

- [140] Slaughter, Laura, Miyagawa, So, Morgado da Costa, Luis, **Zeldes, Amir**, Behlmer, Heike and Lundhaug, Hugo (2020) "An Overview of the Coptic Wordnet Project". *Digital Coptic 3, 3rd workshop on Digital Methods for Coptic Studies*. Held online during Covid-19.
- [141] **Zeldes, Amir** (2019) "Discourse Structure and Crystal Balls in NLP". Invited talk at *DC NLP*. Washington, DC (14 November 2019).
- [142] **Zeldes, Amir** (2019) "Coreference and Coherence Revisited". Invited talk at *CRAC 2019, Second Workshop on Computational Models of Reference, Anaphora and Coreference*, co-located with NAACL 2019. Minneapolis, MN (7 June 2019).
- [143] **Zeldes, Amir** (2018) "Referentiality and Genre in Computational Discourse Models". Invited talk at the Linguistics Research Center. UT Austin, Texas (30 November 2018)
- [144] Krawiec, Rebecca, Luckritz-Marquis, Christine, Platte, Elizabeth, Schroeder, Caroline T. and **Zeldes, Amir** (2018) "A Linked Digital Environment for Coptic Studies". Invited talk at the *Friday Speaker Series*. Georgetown University, Washington DC (14 September 2018)
- [145] **Zeldes, Amir** (2018) "Feature Rich Models for Discourse Signaling". Invited talk at the *Center for Language and Speech Processing*. Baltimore, Johns Hopkins University (23 February 2018).
- [146] **Zeldes, Amir** (2017) "A Multilayer View of Discourse Relation Graphs". Invited talk at the *CMU Language Technologies Institute Colloquium*. Pittsburgh, Carnegie Mellon University (10 November 2017).
- [147] **Zeldes, Amir** (2017) "Predicting Notional Agreement in Anaphora Resolution". Invited talk at the workshop on *Headedness and/or Grammatical Anarchy*. Berlin, Free University (11 May 2017).
- [148] **Zeldes, Amir** (2016) "Status and New Developments in NLP for Coptic". *KELLIA Workshop*, Claremont Graduate University (23 July 2016).
- [149] **Zeldes, Amir** (2015) "Applying Universal Dependencies to Coptic: First Tests and Some Applications for Entity Recognition". *Treebanking Ancient Languages: Current and Prospective Research*, Leipzig University (14 December 2015).
- [150] **Zeldes, Amir** (2015) "Digital Coptic. What, Why and How?". Friday Speaker Series, Georgetown University (18 September 2015).
- [151] **Zeldes, Amir** (2015) "Developing and Using Multilayer Corpora". Computational Linguistics and Information Processing (CLIP) Talks, University of Maryland, College Park, MD (11 March 2015).
- [152] **Zeldes, Amir** (2014) "Corpus Architectures for Multilingual Studies". Advanced Research in Multilingualism Talk Series, Georgetown University (19 November 2014).
- [153] **Zeldes, Amir** (2014) "Compounding in Advanced L2 German Writing - A Corpus Study". Georgetown Linguistics Friday Speaker Series (12 September 2014).
- [154] **Zeldes, Amir** (2014) "Productivity and Grammar". Key note address at Grammar and Corpora 2014, University of Warsaw (26 June 2014).
- [155] Schroeder, Caroline and **Zeldes, Amir** (2014) "Digital Coptic. Building an Online Environment for the Study of Coptic Literature". UC Berkeley, Center for Tebtunis Papyri (15 May 2014).
- [156] **Zeldes, Amir** (2014) "Mehrebenenkorpora in ANNIS, PAULA und SaltNPepper". Invited talk at Neofonie GmbH, Berlin (20 May 2014).
- [157] **Zeldes, Amir** (2014) "Towards Digital Coptic: Searching and Visualizing Coptic Manuscript Data". Talk at the Berlin Digital Classicist Seminar (14 January 2014).
- [158] Krause, Thomas, Odebrecht, Carolin, **Zeldes, Amir** and Zipser, Florian (2013) "Unary TEI Elements and the Token Based Corpus". Workshop Perspectives on Querying TEI-annotated data, TEI Conference 2013, Rome, 1 October 2013.

- [159] Neumann, Arne, **Zeldes, Amir** and Zipser, Florian (2013) "ANNIS 3: Challenges and Innovations for Corpora in SFB632". 17th Internal Workshop of the SFB 632 / Information Structure, Lutherstadt Wittenberg, 13 June 2013.
- [160] **Zeldes, Amir** (2013) "Corpus Linguistics Tools for Sahidic Coptic". Talk at the Leipzig eHumanities Seminar (18 December 2013).
- [161] **Zeldes, Amir** (2013) "Viel mehr als eine Liste: Was uns Lernerdaten über das mentale Lexikon als produktives Netzwerk zeigen [=Much more than a list: What learner data tells us about the mental lexicon as a productive network]". Invited talk at the Interdisciplinary Centre for Research on Lexicography, Valency and Collocation at FAU Erlangen-Nuremberg (29 October 2013).
- [162] **Zeldes, Amir** (2013) "The State of the Art in Collaborative Multilayer Historical Treebanks of German". Invited talk at the Workshop on Syntactic Change and Information Structure, University of Manchester (13 April 2013).
- [163] **Zeldes, Amir** (2013) "Zur Wortart und Kasusrektion des Wortes voller [=On the part-of-speech and case government of the word voller]". Research Colloquium Corpus Linguistics, HU Berlin (12 February 2013).
- [164] **Zeldes, Amir** (2012) "New Developments in Multilayer Corpora for Information Structure". 15th Internal Workshop of the SFB 632, Wandlitz (8 June 2012).
- [165] **Zeldes, Amir** (2012) "Compounding as a Network of Constructions in Advanced Learner German". Università degli Studi di Napoli Federico II (17 May 2012).
- [166] **Zeldes, Amir** (2012) "Productivity in Argument Selection". Università degli Studi di Napoli Federico II (16 May 2012).
- [167] **Zeldes, Amir** (2012) "Komposition als Konstruktionsnetzwerk in L2-Deutsch. Über Lebensblatt, Frauenbedruck, Lehrerkrankenversicherungsamt und andere FALKOMposita". 2nd Kobalt Network Meeting, HU Berlin (1 March 2012).
- [168] **Zeldes, Amir** (2012) "Productivity in Argument Selection: Extensibility Effects of Semantic Classes in Usage". Talk at the Seminar für Englische Philologie, Georg-August-Universität Göttingen (17 January 2012).
- [169] **Zeldes, Amir** (2011) "Getan - gesagt? Pragmatische und lexikalisierte Erklärungen zur Besetzung von Argumentstellen mit neuem Material". IZ Colloquium Europäische Sprachen, Freie Universität Berlin (10 May 2011).
- [170] Krummes, Cedric, Reznicek, Marc, Jia Wei, Chan, Hirschmann, Hagen, Krause, Thomas, **Zeldes, Amir**, Astrid, Ensslin and Lüdeling, Anke (2010) "'What's Hard in German?': Touching the Void of Over- and Underuse". Paper given at the Forum for Germanic Language Studies, Gregynog/Aberystwyth (8-9 January 2010).
- [171] **Zeldes, Amir** (2010) "Mehrebenenkorpora in ANNIS. Datenrepräsentation, Abfrage und Visualisierung". Computerlinguistik-Kolloquium, Universität Zürich, Institut für Computerlinguistik (18 May 2010).
- [172] **Zeldes, Amir** (2010) "ANNIS, PAULA & SaltNPepper. Open Source Resources for Multilayer Corpora". Service-oriented Architectures for Language Technology, Universität Tübingen (11 February 2010).
- [173] Hirschmann, Hagen, **Zeldes, Amir** and Lüdeling, Anke (2009) "Interaction between Colligation, Register and Surface Variability in German Learners and Natives". 31. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft. Osnabrück, Germany (4-6 March 2009).
- [174] Hirschmann, Hagen, **Zeldes, Amir** and Lüdeling, Anke (2009) "Multilevel Learner Corpora". Workshop on Automatic Analysis of Learner Language 2009 (AALL'09), CALICO '09. Arizona State University (10-14 March 2009).

- [175] **Zeldes, Amir** (2009) "Multi-Layer Resources in ANNIS: Historical Corpora and Information Structural Annotation". Workshop Annotating and Analysing Information Structure in Historical Corpus Texts, HU Berlin (13 November 2009).
- [176] **Zeldes, Amir** (2009) "Variabilität in der Produktivität von je-desto-Konstruktionen". Invited talk at the Lecture Series on Variability and Invariability in Language, Potsdam University (25 June 2009).
- [177] **Zeldes, Amir** (2006) "Design and Applications of Polimatth - a Small Parallel Diachronic Bible Corpus of Polish". Lecture at Dagstuhl seminar 06491 Digital Historical Corpora. International Conference and Research Center for Computer Science, Schloss Dagstuhl, Wadern, Germany.
- [178] **Zeldes, Amir** (2006) "Verb Formation in Biblical and Modern Hebrew: Continuity and Innovation". International Conference Morphology and the Digital World, Freie Universität, Berlin.

Book Reviews

- [179] **Zeldes, Amir** (2014) "Review of: Schmidt, Thomas & Wörner, Kai, 'Multilingual Corpora and Multilingual Corpus Analysis'". *Languages in Contrast* 14(2), 317-321.
- [180] **Zeldes, Amir** (2010) "Review of: McEnery, Tony, Xiao, Richard & Tono, Yukio, 'Corpus-Based Language Studies: An Advanced Resource Book'". *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft* [=Journal of Reviews in German Linguistics] 2(2), 268-274.

Corpora

Lead editor

- Coptic Universal Dependency Treebank (2016-), the first syntactically annotated corpus of Coptic, part of the Universal Dependencies (<http://universaldependencies.org>) project. <http://copticcriptorium.org/treebank.html>.
- The Georgetown University Multilayer Corpus (GUM, 2014-), a multilayer, class-sourced corpus of English Web genres, <https://gucorpling.org/gum/>. An additional Universal Dependencies version of this corpus has been published by LINDAT/CLARIN, see <http://universaldependencies.org/>.
- AMALGUM – A Machine Annotated Lookalike of GUM (2020-), automatically annotated but larger scale balanced corpus of English Web genres, modeled after GUM.
- Georgetown University Historical Reddit Corpus (2017), an automatically acquired corpus sampled from monthly reddit dumps since 2005.
- GENTLE (2023), Genre Tests for Linguistic Evaluation, a richly annotated corpus of unusual English genres, <https://github.com/gucorpling/gentle/>

Collaborative projects

- Coptic Scriptorium – a collection of tagged, morphologically and paleographically annotated corpora (2013-, see <http://copticcriptorium.org>). Current corpora I maintain include:
 - A range of corpora from the discourses and canons of Saint Shenoute of Atripe (*Not Because a Fox Barks, I See your Eagerness, Acephalous Work 22, Abraham our Father, Some Kinds of People Sift Dirt, God says through those who are his, Unknown Work 5.1*)

- Letters of Abbot Besa of the White Monastery
- Sayings of the Desert Fathers
- The Canons of Apa Johannes
- The Martyrdom of Victor the General
- Pseudo-Theophilus on the Cross and the Thief
- Saints Lives: Cyrus, Onnophrius, Phib, Aphou, Paul of Tamma, Longinus and Lucius
- Discourses of Pseudo Athanasius
- Documentary Papyri corpus
- Biblical corpora (multiple books of the New and Old Testaments)
- IAHLT Hebrew treebanks – collaboration with the Israeli Association for Human Language Technology, see https://github.com/universalDependencies/UD_Hebrew-IAHLTWiki and <https://aclanthology.org/2022.emnlp-main.292/>
- RIDGES Herbology (2012-), a richly annotated multilayer corpus of historical German scientific writing in the 16th-19th centuries, including early print facsimiles. Supported by two Google Digital Humanities Research Awards (grants awarded to my supervisor, and co-authored by me): https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/research/ridges-projekt?set_language=en.
- Hong Kong City University Corpus of English Learner Academic Drafts (2014-), a learner corpus of Chinese learners of English, containing word and sentence aligned textual revisions and repeated corrective feedback and error annotations. Available at: <https://gucorpling.org/annis/>.

Software

Lead developer

- xrenner – a highly configurable multilingual (non-)named entity recognizer and coreferencer (<https://gucorpling.org/xrenner/>, also available on GitHub and via PyPI)
- rstWeb – web-based interface for Rhetorical Structure Theory annotation (<https://gucorpling.org/rstweb/info/>, available on GitHub)
- DepEdit – a configurable tool for manipulating dependency parses (<https://gucorpling.org/depedit/>, also available on GitHub and via PyPI)
- GitDox – a linked version controlled online xml editor for manuscript transcription and multilayer annotation (<http://gucorpling.org/gitdox/>)
- HebPipe – a SOTA NLP pipeline for Hebrew (<https://github.com/amir-zeldes/HebPipe/>)
- Coptic Dictionary Online – an online dictionary interface for the Coptic language, featuring data collected by KELLIA partners at the BBAW, Göttingen and FU Berlin (<https://coptic-dictionary.org/>). The online dictionary was the winner of the **2019 Digital Humanities Awards** in the category Best Tool or Suite of Tools.
- Coptic NLP tool suite (automatic normalization, multilevel tokenizer, morphological analysis, POS tagging, lemmatization and parsing, <https://tools.copticscriptorium.org/coptic-nlp/>)
- RFTokenizer – state of the art morphological segmentation for Modern Hebrew, Arabic and other morphologically rich languages (<https://github.com/amir-zeldes/RFTokenizer>)

Contributor

- ANNIS – a browser-based search and visualization architecture for complex multilayer corpora (<http://corpus-tools.org/annis/>)
- SaltNPepper – a highly extensible platform for conversion and manipulation of linguistic data between multiple formats (<http://corpus-tools.org/pepper/>)

Teaching Experience

Spring Semester 2024	LING-5444 Machine Learning for Linguistics (Mon+Wed 11:00-12:15, Car Barn 170)
	LING-8415 Computational Discourse Modelling (Mon+Wed 2:00-3:15, Healy 105)
Fall Semester 2023	LING-4427 Computational Corpus Linguistics (Mon+Wed 11:00-12:15, St. Mary's 107)
	LING-4400 Introduction to Natural Language Processing (Mon+Wed 2:00-3:15, Walsh 490)
Spring Semester 2023	LING-469 Analyzing Language Data with R (Mon+Wed 2:00-3:15, ICC 104)
Fall Semester 2022	LING-367 Computational Corpus Linguistics (Mon+Wed 11:00-12:15, Walsh 399)
Spring Semester 2022	LING-504 Machine Learning for Linguistics (Mon+Wed 11:00-12:15, Car Barn 303)
	LING-765 Computational Discourse Modelling (Mon+Wed 2:00-3:15, Car Barn 317)
Fall Semester 2021	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, Maguire 102)
	LING-362 Introduction to Natural Language Processing (Mon+Wed 11:00-12:15, ICC 213)
Fall Semester 2020	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, online due to COVID-19)
	LING-362 Introduction to Natural Language Processing (Mon+Wed 11:00-12:15, online due to COVID-19)
Spring Semester 2020	LING-504 Machine Learning for Linguistics (Mon+Wed 11:00-12:15)
	LING-469 Analyzing Language Data with R (Mon+Wed 2:00-3:15)
Fall Semester 2019	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, Healy 103)
	LING-362 Introduction to Natural Language Processing (Mon+Wed 11:00-12:15, Walsh 497)
Spring Semester 2019	LING-765 Computational Discourse Models (Mon+Wed 11:00-12:15, Walsh 497)
	LING-469 Analyzing Language Data with R (Mon+Wed 2:00-3:15, ICC 117)

Fall Semester 2018	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, Walsh 495)
	LING-362 Introduction to Natural Language Processing (Mon+Wed 11:00-12:15, Maguire 103)
Fall Semester 2017	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, Reiss 261A)
	LING-362 Introduction to Natural Language Processing (Mon+Wed 11:00-12:15, WGR311)
Spring Semester 2017	Corpus Linguistics (2017 Linguistic Institute, Language Across Space and Time, together with Nathan Schneider) (July 5-August 1, Tuesday/Friday 11:00 AM-12:50 PM, Lexington, KY)
	LING-264 Multilingual and Parallel Corpora (Mon+Wed 11:00-12:15, ICC 201)
	LING-469 Analyzing Language Data with R (Mon+Wed 2:00-3:15, ICC 201)
Fall Semester 2016	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, Healy 104)
	LING-765 Computational Discourse Models (Mon+Wed 11:00-12:15, ICC 208A)
Spring Semester 2016	LING-362 Introduction to Natural Language Processing (Mon+Wed 11:00-12:15, WGR 206)
	Statistical Methods for Corpora Using R (Indiana University, Bloomington, June 6-10, 2016)
Fall Semester 2015	LING-367 Computational Corpus Linguistics (Mon+Wed 2:00-3:15, Healy 104)
Spring Semester 2015	LING-465 Non-Standard Syntax Annotation (Mon+Wed 2-3:15 ICC 219B)
	LING-469 Analyzing Language Data with R (Mon+Wed 11-12:15 ICC 223A)
Fall Semester 2014	LING-367 Computational Corpus Linguistics (Mon+Wed 9:30-10:45 CBN 302)
Summer Semester 2014	52 20063 SE Introduction to Coptic Grammar (Wed 14-16 DOR24)
	52 20061 KO Research Colloquium Corpus Linguistics (Wed 18-20 DOR24, together with Hagen Hirschmann)
	MGK Workshop Mixed-effects Models and Logistic Regression with R (Universität Potsdam, Campus Golm, Bdg. 14, R. 2.15/16, 3.7.2014)
Winter Semester 2013/14	Multilayer and Parallel Corpora: An Introduction Using ANNIS (Hong Kong City University, Block Seminar, 13.11.2013)

Summer Semester 2013	Theoretical and Empirical Approaches to Argument Structure (HU Berlin, Seminar 52 20015) Corpus Linguistics and Non-Compositionality (Minerva Non-Compositionality Summer School, Jerusalem 23-28 June 2013) Statistical Methods in Linguistics (1-5 July 2013, Hebrew University of Jerusalem intensive MA course 10817)
Winter Semester 2012/13	Syntactic Annotation (HU Berlin, Seminar 52 20077) Multifactorial Methods with R (HU Berlin, 4 th Emp. Methods Workshop) Introduction to Statistics (Tutorial at 35 th Annual Conference of the German Association for Linguistics [DGfS], Potsdam University)
Summer Semester 2012	Revival of the Hebrew Language (HU Berlin, Seminar 52 20004) Statistics for the Digital Humanities (NEH Institute ‘Working with Text in a Digital Age’, Tufts University)
Winter Semester 2011/12	R II - Advanced Course in Statistical Methods with R (HU Berlin, Empirical Methods Workshop 2012)
Summer Semester 2011	R Workshop, Binary Logistic Regression & Linear Mixed Effects Models (Centre for General Linguistics [ZAS], Berlin)
Winter Semester 2010/11	Introduction to Linguistics (HU Berlin, Intro Course 52 20104) R II - Advanced Course in Statistical Methods with R (HU Berlin, Empirical Methods Workshop 2011)
Summer Semester 2010	Statistical Methods in Linguistics (HU Berlin, Tutorial 52 20146) Corpus Linguistics (HU Berlin, Empirical Methods Workshop 2010)
Summer Semester 2009	Introduction to Morphology (HU Berlin, Intro Course 52 20120) Corpus Linguistics (University of Oldenburg, block seminar together with Hagen Hirschmann)
Winter Semester 2008/09	Parallel Corpora (HU Berlin, Seminar 52 20154)
Summer Semester 2008	Introduction to Corpus Linguistics (HU Berlin, Tutorial 52 185)

Service and Unfunded Voluntary Project Work

Departmental Service

- Concentration Director of Computational Linguistics at the GU Linguistics Department (2015–2018 [fall], 2019–2020, 2021 [spring]–2024)
- Tenure-line Search Committee chair (2015, 2023)
- GU Linguistics Department DEI committee (2022)
- GU Linguistics Department Director of Corpus Resources (since 2014)
- Organizer of reading group on Usage Based Approaches to Grammar (UBAG) at GU (2015-2018)
- GU Linguistics Department Committee to reexamine departmental MS programs (2016)

- GU Linguistics Department merit review committee (2015, 2018-2019, 2022-2023)
- Co-organizer of the departmental reading circle on usage-based grammar models at the Institute for German Language and Linguistics, Humboldt University in Berlin: <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/events-en/reading-circle-ubg/> (2013–2014)

University Service

- Massive Data Institute affiliate (2023–2024)
- Provost’s Internal Grants Review Committee (2021)
- GU Main Campus Executive Faculty Steering Committee (2019–2020)
- Georgetown University Main Campus Executive Faculty (MCEF, 2015–2020)
- Georgetown University executive faculty constitution revision committee (2017–2018)
- Committee to revise the doctoral regulations of Philosophical Faculty II, HU Berlin (January 2009)

Service to the Field

- President of ACL SIGANN, the Special Interest Group on Annotation of the Association for Computational Linguistics (since 2021)
- Universal Dependencies Guidelines Core Group (since 2021)
- Associate editor of the journal *Dialogue and Discourse* (since 2019)
- Editorial board of the journal *Corpus Linguistics and Linguistic Theory* (since 2016)
- Editorial board of the book series *Trends in Linguistics: Studies and Monographs* (De Gruyter, since 2013)
- Member of CLARIN (<http://www.clarin.eu/>) Working Group 5.7 on Interoperability and Standards for language resources (2010–2014)
- Co-lead of <tiger2/>, an effort to standardize the XML serialization of the ISO SynAF syntactic model for ISO/TC37/SC4: <http://korpling.german.hu-berlin.de/tiger2/> (2009–2014)
- Pro bono counseling for the Arbeitskreis neue Erziehung (Association for New Education) Project on Multilingual Letters for New Parents: <http://www.ane.de/homepage-english/letters-to-parents-media-of-parent-education.html> (2009)

Reviewing, Conference and Workshop Organization

- Reviewing for scholarly journals:
 - *Association for Computing Machinery (ACM) Journal on Computing and Cultural Heritage*
 - *Cognitive Linguistics*
 - *Cognitive Science*
 - *Computational Linguistics*
 - *Corpus Linguistics and Linguistic Theory*
 - *Diachronica*
 - *Dialogue and Discourse*
 - *Interdisciplinary Studies in Information Structure*
 - *International Journal of Applied Linguistics*
 - *International Journal of Corpus Linguistics*
 - *Journal of Language Contact*

- *Journal of Language Modelling*
- *Language Resources and Evaluation Journal*
- *Languages in Contrast*
- *Studies in Language*
- *Zeitschrift für Sprachwissenschaft* [=German Journal of Linguistics]
- *Zeitschrift für germanistische Linguistik* [=Journal of German Linguistics]
- *Zeitschrift für Rezensionen zur Germanistischen Sprachwissenschaft* [=Review Journal of German Linguistics]
- Reviewing for publishing houses (book proposals and chapters):
 - Cambridge University Press
 - De Gruyter Mouton
 - Springer

Conference and Workshop Organizing Committee

- LAW 2024, The ACL SIGANN *Linguistic Annotation Workshop*.
- LREC-COLING 2024, Workshops Co-Chair, Turin, Italy, 20-25 May, 2024.
- GURT 2023, *Georgetown University Roundtable on Linguistics conference* (hosting the ACL SIGPARSE sponsored TLT, UDW, Depling and CxGs+NLP workshops), 16-19 March 2023, Washington, DC.
- LAW 2023, The ACL SIGANN *Linguistic Annotation Workshop*, co-located with ACL 2023 in Toronto, 13 July 2023.
- CODI 2022, *Workshop on Computational Approaches to Discourse*, co-located with COLING 2022. 16-17 October 2022, Gyeongju, Republic of Korea.
- CODI 2021, *Workshop on Computational Approaches to Discourse*, co-located with EMNLP 2021 / organizer for the DISRPT 2021 Shared Task on Discourse Relation Classification, Segmentation, and Connective Detection. November 2021, Dominican Republic.
- LAW 2020, the fourteenth *Linguistic Annotation Workshop*, Program Committee Co-Chair (with Stefanie Dipper), 2020.
- Digital Coptic 3 - 3rd workshop on Digital Methods for Coptic Studies, held online due to the COVID-19 pandemic, July 12-13, 2022.
- NAACL2019 Workshop on *Discourse Relation Parsing and Treebanking (DISRPT 2019)*, June 6, 2019, Minneapolis, MN.
- Faculty mentor for the Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL), George Washington University, May 6, 2017.
- Coptic SCRIPTORIUM Workshop and Tutorial, World Congress of the International Association of Coptic Studies (IACS), Claremont, CA, July 23-29, 2016.
- Digital Coptic 2 - Second Workshop on Coptic in the digital age, Department of Linguistics, Georgetown University, March 12-13, 2015.
- Conference *Big Data in a Transdisciplinary Perspective*, Volkswagen Stiftung Herrenhausen Symposium on Big Data, Hanover 2015.
- Workshop *Coptic Resources in the Digital Age*, Humboldt University Berlin, 14 May 2013.
- Workshop on *Usage-Based Approaches to Morphology* at the 35th Annual Meeting of the German Society for Linguistics (DGfS), Potsdam University 13–15 March 2013.

- International conference on *Information Structure: Empirical Perspectives on Theory*. Potsdam University, 2–3 December 2011.
- International conference *Quantitative Investigations in Theoretical Linguistics 4 (QITL-4)*, Humboldt University Berlin, 28 February – 31 March 2011.
- *ANNIS/PAULA User and Developer Workshop* at the Annual Meeting of the German Society for Computational Linguistics (GSCL), Potsdam University, 2 October 2009.

Conference Program Committee / Reviewing / ARR

- NAACL 2024, Senior Area Chair for Discourse and Pragmatics. Mexico City, 16-21 June 2024.
- *LREC-COLING 2024*, main conference Discourse and Pragmatics track, Turin, Italy, 20-25 May, 2024.
- *SIGLEX MWE-UD Workshop 2024*. At LREC 2024, Turin, 25 May 2024.
- *EMNLP 2023*, Area Chair for Discourse and Pragmatics. Singapore, 6-10 December, 2023.
- *Discourse Studies and Linguistic Data Science (DisLiDas 2023)*, Vienna, September 2023.
- *ACL 2023*, Area Chair for Discourse and Pragmatics. Toronto, Canada, 9-14 July 2023.
- *EACL 2023*, Dubrovnik, Croatia, 2-6 May 2023.
- *EMNLP 2022*, Area Chair for Discourse and Pragmatics. Abu Dhabi, 7-11 December 2022.
- *COLING 2022, the 29th International Conference on Computational Linguistics*. 12-17 October 2022, Gyeongju, Republic of Korea.
- *Discourse Studies and Linguistic Data Science: Addressing Challenges in Interoperability, Multilinguality and Linguistic Data Processing (DisLiDas)*, Jerusalem, 24 May 2022.
- *ACL Rolling Review (ARR)*, Action Editor (Feb. 2022).
- *Language Resources and Evaluation Conference (LREC 2022)*, Marseille, 20-25 June 2022.
- *ACL Student Research Workshop (SRW) 2022*, 22 May 2022.
- *SyntaxFest 2021*, Sofia, Bulgaria, 21-25 March 2022.
- *EMNLP 2021*, Dominican Republic, 7-11 November 2021.
- *ACL-IJCNLP 2021*, Bangkok, Thailand, 1-6 August 2021.
- *CODI at EMNLP 2020, First Workshop on Computational Approaches to Discourse*. Online, 19 November 2020.
- *COLING 2020*, Barcelona, Spain, 15-18 September 2020.
- *Universal Dependencies Workshop 2020 (UDW 2020)*, Barcelona, Spain, September 14 2020.
- *Trebanks and Linguistic Theories 2020 (TLT 2020)*, Düsseldorf, Germany, 17-18 August 2020.
- *Integrating Perspectives on Discourse Annotation*, workshop at ESSLLI 2020, Utrecht, 3 August 2020.
- *Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France, 11-16 May 2020.
- *Language Technologies for Historical and Ancient Languages (LT4HALA 2020)*, Marseille, France, 12 May 2020.
- *Explicit and implicit coherence relations: Different, but how exactly?* Workshop at Humboldt University, Berlin, 17-18 January 2020.

- *Workshop on Discourse Structure in Neural NLG (DSNNLG) at INLG-19*. Tokyo, 1 November 2019.
- *SyntaxFest-2019 (=TLT+DepLing+UDW+Quasy 2019)*. Paris, 26-30 August 2019.
- *Challenges in the management of large corpora (CMLC-7)*. Cardiff, 22 July 2019.
- *Thirteenth Linguistic Annotation Workshop (LAW XIII) at ACL 2019*. Florence, 1-2 August 2019.
- *DATeCH 2019, 3rd Conference on Digital Access to Textual Cultural Heritage*. Brussels, Belgium, May 2019.
- *Second Workshop on Universal Dependencies at EMNLP 2018*. Brussels, Belgium, 31 October 2018.
- *Workshop on Annotation in Digital Humanities (annDH) at ESSLLI 2018*, Sofia, Bulgaria, August 2018.
- *ACL 2018, 56th Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, 15-20 July 2018.
- *LAW-MWE-CxG, 12th SIGANN Linguistic Annotation Workshop / Multiword Expression Workshop*. Santa Fe, NM, 25-26 August 2018.
- *CMLC-2018, Challenges in the Management of Large Corpora*, Miyazaki, Japan, 7 May 2018.
- *Language Resources and Evaluation Conference (LREC 2018)*, Miyazaki, Japan, 7-12 May 2018.
- *Corpus-based Research in the Humanities (CRH) 2018*, Vienna, 25-26 January 2018.
- *Teaching NLP for Digital Humanities (Teach4DH)*, 12 September 2017, Berlin, Germany.
- *6th Workshop Recent Advances in RST and Related Formalisms*, in conjunction with INLG 2017, September 4th, 2017, Santiago de Compostela, Spain.
- *ACL 2017, 55th Annual Meeting of the Association for Computational Linguistics*. Vancouver, 30 July – 4 August 2017.
- *11th Web as Corpus workshop (WAC-XI)*, Birmingham, 24 July 2017.
- *CMLC-2017, Challenges in the Management of Large Corpora, Big Data and Natural Language Processing*. Birmingham, 24 July 2017.
- *NoDaLiDa 2017 Workshop on Universal Dependencies*, 22 May 2017, Gothenburg, Sweden.
- *FLAIRS 2017 Workshop on NLP of Ancient and other Low-resource Languages*, 22-24 May 2017, Marco Island, Florida.
- *EACL Workshop on Coreference Resolution beyond OntoNotes (CORBON 2017)*, 4 April 2017, Valencia, Spain.
- *EMNLP 2016, Conference on Empirical Methods in Natural Language Processing*, November 2-6, 2016, Austin, TX.
- *KONVENS2016, 13th Conference on Natural Language Processing*, Bochum, 19-21 September 2016.
- *17th International Morphology Meeting (IMM 17)*, Vienna, February 18-21, 2016.
- *Quantitative Investigations in Theoretical Linguistics 6 (QITL-6)*, Tübingen, November 4-6, 2015.
- *COGSCI 2015, Annual Meeting of the Cognitive Science Society*. July 23-25, Pasadena, CA.
- *Fall School in Computational Linguistics 2015*, German Society for Linguistics (DGfS), Special Interest Group on Computational Linguistics. Trier, 2015.

- *KONVENS2014*, 12th Conference on Natural Language Processing, Hildesheim, 8-10 October 2014.
- *Tenth International Tbilisi Symposium on Language, Logic and Computation*, Gudauri, Georgia, September 23-27, 2013.
- *Quantitative Investigations in Theoretical Linguistics 5 (QITL-5)*, University of Leuven, 12–14 September 2013.
- *Fall School in Computational Linguistics 2013*, German Society for Linguistics (DGfS), Special Interest Group on Computational Linguistics. Heidelberg, September, 2013.
- *KONVENS2012*, 11th Conference on Natural Language Processing, Vienna, 19-21.9.2012.
- LREC 2012, Workshop *Challenges in the Management of Large Corpora*, Istanbul, 22 May 2012.
- *Fall School in Computational Linguistics 2011*, German Society for Linguistics (DGfS), Special Interest Group on Computational Linguistics. Zurich, 29 August –9 September, 2011.

Membership in Professional Societies

- Association for Computational Linguistics (ACL, since 2009)
- Linguistic Society of America (since 2015)
- International Academy of Humanities Computing and Digital Humanities (since 2015)
- Deutsche Gesellschaft für Sprachwissenschaft (DGfS) – The German Society for Linguistics, and the Special Interest Group on Computational Linguistics (DGfS-CL) (since 2008)
- International Association for Coptic Studies (IACS, since 2015)

Computer Skills

- **Programming languages:** Python, Java, JavaScript, R, Perl, PHP, VB
- **Operating systems:** Experience with both Windows and Linux desktop and server environments (esp. Ubuntu and RHEL).
- **Databases:** Extensive experience with PostgreSQL, SQLServer, MySQL, SQLite, etc.
- **Web design/markup:** HTML/CSS, XML/XSLT, SVG, jQuery, Graphviz, Photoshop/GIMP, oXygen.
- **Office etc.:** Excellent knowledge of MS Office, including VBA API, L^AT_EX.

Languages

Native:	Hebrew, English
Near-native fluency:	German
Advanced knowledge:	Arabic, French, Japanese
Intermediate knowledge:	Mandarin Chinese, Polish
Elementary knowledge:	Italian, Spanish
Ancient languages (reading):	Ancient Egyptian, Church Slavonic, Classical Greek, Coptic, Sanskrit