

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](mailto:amir.zeldes@georgetown.edu)

**Website:** <http://corpling.uis.georgetown.edu/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 (1<sup>st</sup> 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 TEI) 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

---

- 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 *Gerald M. Mara Faculty Mentoring Award for Graduate Student Mentoring* (nominee)

## Grants

---

### External funding agencies

- *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). [2018-2020, \$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 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] **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.
- [4] **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

- [5] Schroeder, Caroline T. and **Zeldes, Amir** (to appear) "A Collaborative Ecosystem for Digital Coptic Studies". *Journal of Data Mining and Digital Humanities*.
- [6] **Zeldes, Amir** and Liu, Yang (2020) "A Neural Approach to Discourse Relation Signal Detection". *Dialogue and Discourse* 11(2), 1-33.
- [7] **Zeldes, Amir** (2017) "The GUM Corpus: Creating Multilayer Resources in the Classroom". *Language Resources and Evaluation* 51(3), 581-612.
- [8] 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.
- [9] 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)
- [10] 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.
- [11] Schroeder, Caroline T. and **Zeldes, Amir** (2016) "Raiders of the Lost Corpus". *Digital Humanities Quarterly* 10(2).
- [12] **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.
- [13] 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.

- [14] 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.
- [15] **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.
- [16] Romary, Laurent, **Zeldes, Amir** and Zipser, Florian (2015) "<tiger2/> - Serialising the ISO SynAF Syntactic Object Model". *Language Resources and Evaluation* 49(1), 1-18.
- [17] **Zeldes, Amir** (2013) "Productive Argument Selection: Is Lexical Semantics Enough?". *Corpus Linguistics and Linguistic Theory* 9(2), 263-291.
- [18] **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.
- [19] **Zeldes, Amir** (2013) "Is Modern Hebrew Standard Average European? The View from European". *Linguistic Typology* 17(3), 439-470.
- [20] 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.
- [21] 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.
- [22] 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).
- [23] 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.
- [24] 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.
- [25] **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

- [26] **Zeldes, Amir** (to appear) "Corpus Architecture". In: Paquot, Magali & Gries, Stefan Th. (eds.), *The Practical Handbook of Corpus Linguistics*. Heidelberg: Springer.
- [27] **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.
- [28] **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.
- [29] 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.

- [30] 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.
- [31] **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.
- [32] 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.
- [33] 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.
- [34] 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 / Workshop Papers

- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] **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.
- [40] **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.

- [41] 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.
- [42] 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.
- [43] **Zeldes, Amir** (2018) "A Characterwise Windowed Approach to Hebrew Morphological Segmentation". In: *Proceedings of the 15<sup>th</sup> SIGMORPHON Workshop at EMNLP 2018*. Brussels, Belgium.
- [44] **Zeldes, Amir** and Abrams, Mitchell (2018), "The Coptic Universal Dependency Treebank". In: *Proceedings of the 2018 Universal Dependencies Workshop (UDW 2018)*. Brussels, Belgium.
- [45] 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]
- [46] 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.
- [47] **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.
- [48] MacAvaney, Sean and **Zeldes, Amir** (2018) "A Deeper Look into Dependency-Based Word Embeddings". *NAACL 2018 Student Research Workshop*. New Orleans, LA, 40-45.
- [49] **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.
- [50] **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.
- [51] **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.
- [52] **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.
- [53] **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.
- [54] **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.
- [55] 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

- [56] Gessler, Luke, Peng, Siyao, Liu, Yang, Zhu, Yilun, Behzad, Shabnam and **Zeldes, Amir** (to appear) "AMALGUM – A Free, Balanced, Multilayer English Web Corpus". In: *Proceedings of LREC 2020*. Marseille, France.
- [57] Sanguinetti, Manuela, Bosco, Cristina, Cassidy, Lauren, Çetinoğlu, Özlem, Cignarella, Alessandra Teresa, Lynn, Teresa, Rehbein, Ines, Ruppenhofer, Josef, Seddah, Djamé and **Zeldes, Amir** (to appear) "Treebanking User-Generated Content: A Proposal for a Unified Representation in Universal Dependencies". In: *Proceedings of LREC 2020*. Marseille, France.
- [58] 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.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] **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.
- [63] Miyagawa, So and **Zeldes, Amir** (2018) "A Semantic Map of the Coptic Complementizer *č*-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.
- [64] **Zeldes, Amir** (2018) "A Neural Approach to Discourse Relation Signaling". *Georgetown University Round Table (GURT) 2018: Approaches to Discourse*. Washington, DC.
- [65] 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.
- [66] **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.
- [67] **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.
- [68] **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.
- [69] **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.
- [70] 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.

- [71] 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.
- [72] 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.
- [73] 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.
- [74] **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.
- [75] 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.
- [76] 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.
- [77] **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.
- [78] **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.
- [79] **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.
- [80] **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

- [81] 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.
- [82] **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).
- [83] 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.



- [84] 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.
- [85] 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.
- [86] **Zeldes, Amir** (2010) "Novel Argument Selection: Is Lexical Semantics Enough?". Doktorandentag 2010, Humboldt-Universität zu Berlin.
- [87] 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.
- [88] **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

- [89] **Zeldes, Amir** (2016) *Coptic Scriptorium Lemmatization Guidelines*. Version 1.0.5-2017-09-15. Technical Report, Georgetown University.
- [90] **Zeldes, Amir** (2016) *Coptic Universal Dependency Guidelines*. Version 1.1.0-2016-10-19. Technical Report, Georgetown University.
- [91] **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.
- [92] **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

- [93] **Zeldes, Amir**, Schroeder, Caroline T. and Martin, Lance (2020) "Exposing Coptic entities: automation, search and visualization". *Digital Coptic 3, 3<sup>rd</sup> workshop on Digital Methods for Coptic Studies*. Held online during Covid-19.
- [94] 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, 3<sup>rd</sup> workshop on Digital Methods for Coptic Studies*. Held online during Covid-19.
- [95] **Zeldes, Amir** (2019) "Discourse Structure and Crystal Balls in NLP". Invited talk at *DC NLP*. Washington, DC (14 November 2019).
- [96] **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).
- [97] **Zeldes, Amir** (2018) "Referentiality and Genre in Computational Discourse Models". Invited talk at the Linguistics Research Center. UT Austin, Texas (30 November 2018)
- [98] 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)
- [99] **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).

- [100] **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).
- [101] **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).
- [102] **Zeldes, Amir** (2016) "Status and New Developments in NLP for Coptic". *KELLIA Workshop*, Claremont Graduate University (23 July 2016).
- [103] **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).
- [104] **Zeldes, Amir** (2015) "Digital Coptic. What, Why and How?". Friday Speaker Series, Georgetown University (18 September 2015).
- [105] **Zeldes, Amir** (2015) "Developing and Using Multilayer Corpora". Computational Linguistics and Information Processing (CLIP) Talks, University of Maryland, College Park, MD (11 March 2015).
- [106] **Zeldes, Amir** (2014) "Corpus Architectures for Multilingual Studies". Advanced Research in Multilingualism Talk Series, Georgetown University (19 November 2014).
- [107] **Zeldes, Amir** (2014) "Compounding in Advanced L2 German Writing - A Corpus Study". Georgetown Linguistics Friday Speaker Series (12 September 2014).
- [108] 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).
- [109] **Zeldes, Amir** (2014) "Mehrebenenkorpora in ANNIS, PAULA und SaltNPepper". Invited talk at Neofonie GmbH, Berlin (20 May 2014).
- [110] **Zeldes, Amir** (2014) "Towards Digital Coptic: Searching and Visualizing Coptic Manuscript Data". Talk at the Berlin Digital Classicist Seminar (14 January 2014).
- [111] 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.
- [112] 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.
- [113] **Zeldes, Amir** (2013) "Corpus Linguistics Tools for Sahidic Coptic". Talk at the Leipzig eHumanities Seminar (18 December 2013).
- [114] **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).
- [115] **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).
- [116] **Zeldes, Amir** (2013) "Zur Wortart und Kasusreaktion des Wortes voller [=On the part-of-speech and case government of the word voller]". Research Colloquium Corpus Linguistics, HU Berlin (12 February 2013).
- [117] **Zeldes, Amir** (2012) "New Developments in Multilayer Corpora for Information Structure". 15th Internal Workshop of the SFB 632, Wandlitz (8 June 2012).

- [118] **Zeldes, Amir** (2012) "Compounding as a Network of Constructions in Advanced Learner German". Università degli Studi di Napoli Federico II (17 May 2012).
- [119] **Zeldes, Amir** (2012) "Productivity in Argument Selection". Università degli Studi di Napoli Federico II (16 May 2012).
- [120] **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).
- [121] **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).
- [122] **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).
- [123] 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).
- [124] **Zeldes, Amir** (2010) "Mehrebenenkorpora in ANNIS. Datenrepräsentation, Abfrage und Visualisierung". Computerlinguistik-Kolloquium, Universität Zürich, Institut für Computerlinguistik (18 May 2010).
- [125] **Zeldes, Amir** (2010) "ANNIS, PAULA & SaltNPepper. Open Source Resources for Multilayer Corpora". Service-oriented Architectures for Language Technology, Universität Tübingen (11 February 2010).
- [126] 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).
- [127] 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).
- [128] **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).
- [129] **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).
- [130] **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.
- [131] **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

- [132] **Zeldes, Amir** (2014) "Review of: Schmidt, Thomas & Wörner, Kai, 'Multilingual Corpora and Multilingual Corpus Analysis'". *Languages in Contrast* 14(2), 317-321.
- [133] **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://copticSCRIPTORIUM.org/treebank.html>.
- The Georgetown University Multilayer Corpus (GUM, 2014-), a multilayer, class-sourced corpus of English Web genres, <http://corpling.uis.georgetown.edu/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.

### Collaborative projects

- Coptic Scriptorium – a collection of tagged, morphologically and paleographically annotated corpora (2013-, see <http://copticSCRIPTORIUM.org>). Current corpora I maintain include:
  - o 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*)
  - o Letters of Abbot Besa of the White Monastery
  - o Sayings of the Desert Fathers
  - o The Canons of Apa Johannes
  - o The Martyrdom of Victor the General
  - o Pseudo-Theophilus on the Cross and the Thief
  - o Saints Lives: Cyrus, Onnophrius, Phib, Aphou, Paul of Tamma, Longinus and Lucius
  - o Discourses of Pseudo Athanasius
  - o Documentary Papyri corpus
  - o Biblical corpora (multiple books of the New and Old Testaments)
- RIDGES Herbology (2012-), a richly annotated multilayer corpus of historical German scientific writing in the 16<sup>th</sup>-19<sup>th</sup> 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](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://corpling.uis.georgetown.edu/annis/>.

## Software

---

### Lead developer

- xrenner – a highly configurable multilingual (non-)named entity recognizer and coreferencer (<https://corpling.uis.georgetown.edu/xrenner/>, also available on GitHub and via PyPI)

- rstWeb – web-based interface for Rhetorical Structure Theory annotation (<https://corpling.uis.georgetown.edu/rstweb/info/>, available on GitHub)
- DepEdit – a configurable tool for manipulating dependency parses (<https://corpling.uis.georgetown.edu/depedit/>, also available on GitHub and via PyPI)
- GitDox – a linked version controlled online xml editor for manuscript transcription and multilayer annotation (<http://corpling.uis.georgetown.edu/gitdox/>)
- 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://corpling.uis.georgetown.edu/coptic-dictionary/>). 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://corpling.uis.georgetown.edu/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 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 <sup>th</sup> Emp. Methods Workshop)

	Introduction to Statistics (Tutorial at 35 <sup>th</sup> 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–)
- GU Linguistics Department Director of Corpus Resources (since 2014)
- Organizer of reading group on Usage Based Approaches to Grammar (UBAG) at GU (since 2015)
- GU Linguistics Department Committee to reexamine departmental MS programs (2016)
- GU Linguistics Department merit review committee (2015)
- 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

- GU Main Campus Executive Faculty Steering Committee (2019–)
- Georgetown University Main Campus Executive Faculty (MCEF, 2015–2019)
- 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

- Editorial board of the journal *Corpus Linguistics and Linguistic Theory* (since 2016)
- Editorial board of the journal *Dialogue and Discourse* (since 2019)

- 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 2020, the fourteenth *Linguistic Annotation Workshop*, Program Committee Co-Chair (with Stefanie Dipper), 2020.
- 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 35<sup>th</sup> 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

- *COLING 2020*, Barcelona, Spain, 15-18 September 2020.
- *Universal Dependencies Workshop 2020 (UDW 2020)*, Barcelona, Spain, September 14 2020.
- *Treebanks 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, 3<sup>rd</sup> 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, 12<sup>th</sup> 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, R, Perl, JavaScript, 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<sup>A</sup>T<sub>E</sub>X.

## Languages

---

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