

Kyle Gorman

Curriculum vitae

April 1, 2024

365 5th Ave, New York, NY 10016 Tel. (212) 817-8502.

Education

Postdoctoral fellow, Center for Spoken Language Understanding, Oregon Health & Science University, Portland, OR, 2013.

Ph.D., Linguistics, University of Pennsylvania, Philadelphia, PA, 2013.

B.A., Linguistics, University of Illinois at Urbana-Champaign, Urbana, IL, 2006.

Positions

Associate professor (with tenure), Graduate Program in Linguistics, CUNY Graduate Center, New York, NY, 2023–present.

Assistant professor, Graduate Program in Linguistics, CUNY Graduate Center, New York, NY, 2018–2023.

Director, Master's Program in Computational Linguistics, CUNY Graduate Center, New York, NY, 2018–present.

Software engineer, Google, Inc., New York, NY, 2015–present.

Assistant professor, Center for Spoken Language Understanding, Oregon Health & Science University, Portland, OR, 2014–2015.

Publications

Books

2021. K. Gorman & R. Sproat. *Finite-State Text Processing*. Morgan & Claypool.

Refereed journal articles

2019. H. Zhang, R. Sproat, A. Ng, F. Stahlberg, X. Peng, K. Gorman, & B. Roark. Neural models of text normalization for speech applications. *Computational Linguistics* 45(2): 293–337.

2017. H. MacFarlane, K. Gorman, R. Ingham, A. Presmanes Hill, K. Papadakis, G. Kiss, & J. van Santen. Mazes in children with autism spectrum disorders or language impairment. *PLOS ONE* 12(3): e0173936.

2016. G. Fergadiotis, K. Gorman, & S. Bedrick. Algorithmic classification of five characteristic types of paraphasias. *American Journal of Speech-Language Pathology* 25(4S): S776–S787.

2016. K. Gorman & R. Sproat. Minimally-supervised number normalization. *Transactions of the Association for Computational Linguistics* 4: 507–519.
2016. K. Gorman, L. Olson, A. Presmanes Hill, R. Lunsford, P. Heeman, & J. van Santen. *Uh* and *um* in children with autism spectrum disorders or language impairment. *Autism Research* 9(8): 854–865.
2015. A. Presmanes Hill, J. van Santen, K. Gorman, B. Langhorst, & E. Fombonne. Memory in language-impaired children with and without autism. *Journal of Neurodevelopmental Disorders* 7: 19.
2013. L. Hinrichs, A. Bohmann, & K. Gorman. Real-time trends in the Texas English vowel system: F2 trajectory in GOOSE as an index of a variety's ongoing delocalization. *Rice Working Papers in Linguistics* 4.
2011. K. Gorman, J. Howell, & M. Wagner. Prosodylab-Aligner: A tool for forced alignment of laboratory speech. *Journal of the Canadian Acoustical Association* 39(3): 192–193.
2011. J. Fruehwald & K. Gorman. Cross-derivational feeding is epiphenomenal. *Studies in the Linguistic Sciences* 2011: 36–50.

Invited chapters

2022. K. Gorman. Computational morphology. In M. Aronoff & K. Fudeman, *What is Morphology?*, pages 246–273. 3rd edition. John Wiley & Sons.
2019. K. Gorman & C. Yang. When nobody wins. In F. Rainer, F. Gardani, W. U. Dressler & H. C. Luschützky (ed.), *Competition in inflection and word formation*, 169–193. Springer.
2013. K. Gorman & D. E. Johnson. Quantitative analysis. In R. Bayley, R. Cameron & C. Lucas (ed.), *The Oxford handbook of sociolinguistics*, 214–240. Oxford University Press.

Refereed conference proceedings

2024. K. Gorman & C. Allauzen. A* shortest-string decoding for non-idempotent semirings. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 732–739.
2024. A. Wiemerslage, K. Gorman, and K. von der Wense. Quantifying the hyperparameter sensitivity of neural networks for character-level sequence-to-sequence tasks. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 674–689.
2023. K. Gorman & R. Sproat. Myths about writing systems in speech & language technology. In *Proceedings of the Workshop on Computation and Writing Systems*, pages 1–5.
2021. K. Gorman, C. Kirov, B. Roark, and R. Sproat. Structured abbreviation expansion in context. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 995–1005.
2021. L. F.E. Ashby, T. M. Bartley, S. Clematide, L. Del Signore, C. Gibson, K. Gorman, Y. Lee-Sikka, P. Makarov, A. Malanoski, S. Miller, O. Ortiz, R. Raff, A. Sengupta, B. Seo, Y. Spektor, and W. Yan. Results of the second SIGMORPHON shared task on multilingual grapheme-to-phoneme conversion. In *18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 115–125.
2020. A. Waller & K. Gorman. Detecting objectifying language in online professor reviews. In *Proceedings of the Sixth Workshop on Noisy User-generated Text*, pages 171–180. **Best Paper Award.**
2020. P. Szymański & K. Gorman. Is the best better? Bayesian statistical comparison for natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2203–2212.
2020. K. Gorman, L. F.E. Ashby, A. Goyzueta, A. D. McCarthy, S. Wu, and D. You. The SIGMORPHON 2020 shared task on multilingual grapheme-to-phoneme conversion. In *17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 40–50.

2020. J. L. Lee, L. F.E. Ashby, M. E. Garza, Y. Lee-Sikka, S. Miller, A. Wong, A. D. McCarthy, and K. Gorman. Massively multilingual pronunciation mining with WikiPron. in *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4223–4228.
2020. A. D. McCarthy, C. Kirov, M. Grella, A. Nidhi, P. Xia, K. Gorman, E. Vylomova, S. J. Mielke, G. Nicolai, M. Silfverberg, T. Arkhangelskij, N. Krizhanovsky, A. Krizhanovsky, E. Klyachko, A. Sorokin, J. Mansfield, V. Ernštreits, Y. Pinter, C. L. Jacobs, R. Cotterell, M. Hulden, and D. Yarowsky. UniMorph 3.0: universal morphology. in *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3922–3931.
2019. K. Gorman, A. D. McCarthy, R. Cotterell, E. Vylomova, M. Silfverberg, and M. Markowska. Weird inflects but OK: making sense of morphological generation errors. In *Proceedings of the 23rd Conference on Computational Natural Language Learning*, pages 140–151.
2019. S. Ritchie, R. Sproat, K. Gorman, D. van Esch, C. Schallhart, N. Bampounis, B. Brard, J. F. Mortensen, M. Holt, and E. Mahon. Unified verbalization for speech recognition & synthesis across languages. In *Proceedings of INTERSPEECH*, pages 3530–3534.
2019. K. Gorman & S. Bedrick. We need to talk about standard splits. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2786–2791. **Outstanding Paper Award.**
2019. S.J. Mielke, R. Cotterell, K. Gorman, B. Roark & J. Eisner. What kind of language is hard to language-model? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4975–4989.
2018. K. Gorman, G. Mazovetskiy, & V. Nikolaev. Improving homograph disambiguation with machine learning. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pages 1349–1352.
2017. A. Ng., K. Gorman, & R. Sproat. Minimally supervised written-to-spoken text normalization. In *Proceedings of ASRU*, pages 665–670.
2017. J. Adams, S. Bedrick, G. Fergadiotis, K. Gorman, & J. van Santen. Target word prediction and paraphasia classification in spoken discourse. In *Proceedings of the ACL Workshop on Biomedical Natural Language Processing*, pages 1–8.
2016. K. Gorman. Pynini: A Python library for weighted finite-state grammar compilation. In *Proceedings of the ACL Workshop on Statistical NLP and Weighted Automata*, pages 75–80.
2015. K. Gorman, S. Bedrick, G. Kiss, M. Mohammed, R. Ingham, K. Papadakis, & J. van Santen. Automated morphological analysis of clinical language samples. In *Proceedings of ACL Workshop on Computational Linguistics and Clinical Psychology*, pages 108–116.
2014. M. Lehr, K. Gorman, & I. Shafran. Discriminative pronunciation model for dialectal speech recognition. In *Proceedings of INTERSPEECH*, pages 1458–1462.
2014. C. Lignos & K. Gorman. Revisiting frequency and storage in morphological processing. In *Proceedings of the 48th Annual Meeting of the Chicago Linguistic Society*, pages 447–461.
2014. K. Gorman. Rhotacism in Latin. In *Proceedings of the 48th Annual Meeting of the Chicago Linguistic Society*, pages 279–293.
2014. K. Gorman. A program for phonotactic theory. In *Proceedings of the 47th Annual Meeting of the Chicago Linguistic Society: the Main Session*, pages 79–93.
2010. K. Gorman. The consequences of multicollinearity among socioeconomic predictors of negative concord in Philadelphia. In *University of Pennsylvania Working Papers in Linguistics 16.2: Selected Papers from NWAV 38*, pages 66–75.
2007. C. Lai, K. Gorman, J. Yuan, & M. Liberman. Perception of disfluency: Language differences and listener bias. In *Proceedings of INTERSPEECH*, pages 2345–2348.

Technical reports

2009. K. Gorman. Hierarchical regression for language research. Institute for Research in Cognitive Science Technical Report 09-02.

In preparation

G. Kiss, K. Gorman, & J. van Santen. Group-matching for subjects and items.

K. Gorman. CityLex: a free English lexical database.

K. Gorman. A tutorial on pair n-gram models.

Edited volumes

2010. K. Gorman & L. MacKenzie (ed.). *University of Pennsylvania Working Papers in Linguistics, vol. 15.2: Selected Papers from NAWAV 37*.

2010. K. Gorman (ed.). *University of Pennsylvania Working Papers in Linguistics, vol. 14.2: Selected Papers from NAWAV 36*.

Grants

2023–2025. BSF 7E02: Diacritization for the world's scripts; \$150,000 (with Y. Pinter and M. Elhadad).

2023–2024. PSC-CUNY 66667-00 54: Multi-source grapheme-to-phoneme conversion; \$5,980.

2022–2026. NSF BCS 2214137: Collaborative research: deconstructing wordlikeness judgments; \$166,990 (with K. Durvasula and J. Kahng.)

2022–2023. PSC-CUNY 65587-00 53: Neural diacritization; \$5,417.

2021–2023. PSC-CUNY 64142-00 52: Defectivity and indirect negative evidence; \$3,490.

Awards

2020. Workshop on Noisy User-generated Text, Best Paper Award (with A. Waller).

2019. Association for Computational Linguistics, Outstanding Paper Award (with S. Bedrick); \$1,000.

Patents

2019. R. Sproat, K. Wu, & K. Gorman. Semiotic class normalization. U.S. Patent 9,852,123.

Advising

(Graduate Program in Linguistics, Graduate Center, City University of New York unless otherwise noted.)

PhD

Dissertations chaired

In progress. Enas Albasiri. Computational approaches to linguistic challenges in Arabic speech recognition and (inverse) text normalization.

Dissertation committee

2023. Adam Wiemerslage (Department of Computer Science, University of Colorado Boulder). (Title TBD).

2023. Arya McCarthy (Department of Computer Science, Johns Hopkins University). Thousand-language processing: Translation, learning, and projection.

2023. Benjamin Shavitz. Memetic epidemiology.

2022. Xiaomeng Ma. Evaluating neural networks as cognitive models for learning quasi-regularities in language.

2022. Soumik Dey (Graduate Program in Computer Science). Grammar competition explored in two case studies: NS state and Old English.

2022. Qihui Xu (Graduate Program in Psychology). Linguistic abstractions of children's very early utterances.

2021. Jennifer Seale. Label imputation for homograph disambiguation: theoretical and practical approaches.

2021. Soumi Maiti (Graduate Program in Computer Science). Speech enhancement using speech synthesis techniques.

2019. Rachel Rakov. Analyzing prosody with Legendre polynomial coefficients.

2019. Pablo Gonzalez Martinez. Do it like a syntactician: using binary grammaticality judgements to train sentence encoders and assess their sensitivity to syntactic structure.

2015. Brian Bush (Graduate Program in Computer Science and Electrical Engineering, Oregon Health & Science University). Modeling coarticulation in continuous speech.

MA

Theses supervised

2024. Martine Harrison. Consonant (de)gradation in Ingrian?

2024. Daniel Yakubov. How do we learn what we cannot say?

2023. Maria Karamihaylova. Neural network vs. rule-based G2P: A hybrid approach to stress prediction and related vowel reduction in Bulgarian.

2023. Wen Zhang. Pronunciation ambiguities in Japanese *kanji*.

2022. Reuben Raff. From Sesame Street to beyond: multi-domain discourse relation classification with pre-trained BERT.

2022. Elizabeth Garza. Does a neural network inflect Spanish verbs in a human-like way?

2022. John Schriener. Predicting stress in Russian using modern machine-learning tools.

2022. Yuying Ren. Methods in reverse transliteration of English loanwords in Japanese.

2022. Lara Novic. Sarcasm detection with machine learning.

2021. Natalia Tyulina. Computational representation of Russian aspectual morphology with a focus on perfective prefixation.

2021. Yulia Spektor. Detection and morphological analysis of novel Russian loanwords.

2021. Jonathan Manczur. From an art to a science: features and methodology in computational authorship identification.

2021. Andrew Kirby. Predicting stock price movements using sentiment and subjectivity analyses.

2021. Yohamy Polanco. A computational study in the detection of English-Spanish code-switches.

2021. Sean Miller. When misclassification is misgendering: gender prediction in the context of trans identities.

2020. Emily Campbell. Does the word “chien” bark? Representation learning in neural machine translation encoders.

2020. Alan Wong. Mitigating gender bias in neural machine translation using counterfactual data.

2020. Magdalena Markowska. Tones in Shupamem reduplication.

2020. Alina Korovatskaya. Genderlects in social media.

2020. Michael C. Stern. Testing the perceptual magnet effect in monolinguals and bilinguals. (Winner of the linguistics annual master’s thesis prize.)

2020. Anthony J. Vicario. Phonologically-informed speech coding for automatic speech recognition-based foreign language pronunciation training.

2020. Angie Waller. Ghost peppers: using ensemble models to detect professor attractiveness commentary on RateMyProfessors.com. (Winner of the computational linguistics annual master’s thesis prize.)

2019. Sara Morini. Demographic factors as domains for adaptation in linguistic preprocessing.

Abstract-reviewed talks

[2022 onwards only.]

2023. K. Gorman & C. Reiss. A simple alternative to SPE’s non-deterministic, intractable evaluation metric. Morris Halle @ 100, Massachusetts Institute of Technology, September 8–10.

2023. K. Gorman. Asymmetries in Latin glide formation. North American Phonology Conference, Concordia University, May 12.

2023. K. Gorman & C. Reiss. Maximal feature specification is feasible; minimal feature specification is not. Generative Linguistics in the Old World, University of Vienna, April 12.

2023. K. Gorman & C. Reiss. Maximal feature specification is feasible; minimal feature specification not so much. Penn Linguistics Colloquium, March 18, University of Pennsylvania.

2023. J. Chang & K. Gorman. Automated analysis of phonological variation in Latin poetry. Annual Meeting of the Linguistics Society of America, January 7, Denver.

2022. K. Gorman. Black box morphophonology. Workshop on Model Theoretic Representations, Stony Brook University, September 23.

2022. K. Gorman & R. Sproat. The persistent conflation of writing and language. Grapholinguistics in the 21st Century, June 8.

Invited talks

2024. Careers in computational linguistics. City College, City University of New York, February 27.

2023. Get rich quick. Rutgers Linguistics Colloquium, Rutgers University, November 3.

2023. Notes on morphological defectivity. Surrey Linguistic Circle, University of Surrey, October 18.

2023. Features in computational phonology. 20th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, Toronto, July 14.

2023. "Massive multilingualism" in speech & language technology. University of Colorado Boulder, April 5.

2023. "Massive multilingualism" in speech & language technology. Stone Center on Socio-Economic Inequality multidisciplinary seminar, Graduate Center, City University of New York, March 3.

2022. Black box phonology. Integrated Language Science and Technology seminar, University of Pennsylvania, October 14.

2022. Weighted finite-state transducers: the later years. Information Sciences Institute, University of Southern California, June 23.

2022. Finite-state text processing. North American Summer School for Logic, Language, and Information (NASSLLI), University of Southern California, June 19.

2022. Weighted finite-state transducers: the later years. Center for Language and Speech Processing, Johns Hopkins University, April 1.

2022. Phonotactics in phonological theory. Department of Linguistics, Michigan State University, March 3.

2021. On the role of phonotactics in phonology. SYNC, Graduate Center, City University of New York, December 4.

2021. Multilingual linguistic resource development at the Graduate Center. Department of Computational Linguistics, University of Zurich, November 15.

2021. Multilingual linguistic resource development at the Graduate Center. SIGTYP Workshop series, September 3.

2021. On productivity in phonology. North Atlantic Phonology Conference, June 27.

2021. Multilingual linguistic resource development at the Graduate Center. Ph.D. Program in Speech-Language-Hearing Sciences, Graduate Center, City University of New York, April 28.

2020. Anatomy of an analogy. Department of Linguistics, Stony Brook University, November 6.

2020. Abstractness, arbitrariness, and productivity in Polish declension. Graduate Program in Linguistics, Graduate Center, City University of New York, October 8.

2020. Weird inflects but OK: making sense of morphological generation errors. Linguistics Department, University of California, San Diego, January 30.

2020. A tutorial on finite-state text processing. Society for Computation in Linguistics, January 2.

2019. Massively multilingual resources for speech and language technology. Department of Linguistics and Communication Disorders, Queens College, December 12.

2019. Weird inflects but OK: making sense of morphological generation errors. Northeast Computational Phonology Circle (NECPhon), Rutgers University, November 16.

2019. A tutorial on finite-state text processing. Linguistic Data Consortium Institute, November 5.

2019. We need to talk about standard splits. Fondazione Bruno Kessler, July 25.

2019. Two little pieces on statistical system evaluation and error analysis. Sapienza University of Rome, July 24.

2019. Internationalization challenges for speech and language technologies. Documentation Division, United Nations, February 19.

2019. Statistically informed reproduction analysis. Linguistics Department, Swarthmore College, February 12.

2019. Statistically informed reproduction analysis. Institute for Research in Cognitive Science, University of Pennsylvania, February 11.

2018. Grammar engineering for speech and language processing. Center for Data Science, New York University, November 29.

2018. Grammar engineering for speech and language processing. Linguistics Department, McGill University, November 14.

2018. Finite-state grammar development in Pynini. Linguistics Department, McGill University, November 12.

2018. Grammar engineering for speech and language processing. Program in Linguistics, Princeton University, September 26.

2018. Minimally supervised number normalization. RIKEN Center for Advanced Intelligence, May 16.

2018. Minimally supervised number normalization. Information Processing Society of Japan, Tokyo, May 12.

2018. Minimally supervised number normalization. University of California, Irvine, May 4.

2018. Computational approaches to language. Graduate Program in Linguistics, Graduate Center, City University of New York, March 23.

2017. Minimally supervised number normalization. Institute for Advanced Computational Science, Stony Brook University, November 29.

2016. Processing informal text. Department of English, University of Texas at Austin, November 28.

2014. Forced alignment with Prosodylab-Aligner. Experimental and Theoretical Advances in Prosody, McGill University, May 10.

2013. Computational characterizations to language in children with autism spectrum disorders. University of Connecticut, Storrs, October 16.

2013. What phonotactics might not be. Department of Linguistics, University of Oregon, April 2.

2013. What phonotactics might not be. Department of Linguistics, Reed College, January 31.

2012. Words vs. rules in irregular acquisition. Center for Spoken Language Understanding, Oregon Health & Science University, June 13.

2011. Forced alignment with Prosodylab-Aligner. Chicago Linguistics Society, University of Chicago, October 14.

2011. Quantitative analysis for linguists. Hunter College, September 9.

2011. What phonotactics might not be. Department of Linguistics, New York University, March 4.

2011. What phonotactics might not be. Department of Linguistics, University of California, Santa Cruz, February 10.

2011. Latin rhotacism for real. Department of Linguistics, Concordia University, February 4.

Teaching

Methods in Computational Linguistics I, Graduate Center, City University of New York.

Methods in Computational Linguistics II, Graduate Center, City University of New York.

Corpus Analysis, Graduate Center, City University of New York.

Language Technology, Graduate Center, City University of New York.

Statistics for Linguistics Research, Graduate Center, City University of New York.

Seminar in Computational Phonology, Graduate Center, City University of New York.

Seminar in Writing Systems, Graduate Center, City University of New York.

Defectivity, Summer Institute, Linguistic Society of America, University of Massachusetts, Amherst.
Natural Language Processing, Oregon Health & Science University.
Research Programming, Oregon Health & Science University.

Internal service

2018–present. Director, Speaker Series in Computational Linguistics.
2019–present. Member, Executive Committee, Graduate Program in Linguistics.
2023–present. Senator, University Faculty Senate.
2019–2023. Member, Admissions Committee, Graduate Program in Linguistics.
2019–2021. Member, Graduate Council, the Graduate Center.
2022–2023. Member, Curriculum & Degree Committee, Graduate Program in Linguistics.

External service

2023. Co-founder and interim treasurer, Association for Computational Linguistics Special Interest Group on Writing Systems and Written Language (SIGWrit).
2023. Organizer, Association for Computational Linguistics workshop on Computation and Written Language (CAWL), ACL 2023.
2020–2022. Area Chair, Conference on Natural Language Learning, Association for Computational Linguistics Special Interest Group on Natural Language Learning (SIGNLL).
2019–2021. Executive Committee, Association for Computational Linguistics Special Interest Group on Computational Morphology and Phonology (SIGMORPHON).
2018. Chair, Program Committee, Language Resources & Evaluation Industry Track.
Reviewer for ACL Rolling Reviews (ACL, EMNLP, etc.), ICASSP, INTERSPEECH, LREC, the Society for Computation in Linguistics, the National Science Foundation, the Natural Sciences and Engineering Research Council of Canada, Oxford University Press, *Cognition*, *Cognitive Science*, *Computational Linguistics*, *Glossa*, *Journal of Linguistics*, *Journal of Autism & Developmental Disorders*, *Language*, *Language Variation & Change*, *Lingua*, *Nature Communication*, *Phonology*, *PLOS ONE*, and *Transactions of the Association for Computational Linguistics*.

Professional affiliations

Linguistics Society of America
Association for Computational Linguistics
Institute of Electrical and Electronics Engineers