

Dr Jonathan Dunn (he/him)

University of Canterbury
Department of Linguistics
Christchurch, New Zealand

www.jdunn.name
www.github.com/jonathandunn
jonathan.dunn@canterbury.ac.nz

Specialisation

Computational Linguistics; Cognitive Linguistics; Syntax/Semantics

Overview

I am a computational linguist working on models of language learning and language change using [large multi-lingual corpora](#). The challenge is to model how [speakers converge onto similar grammars](#) while at the same time modelling [how grammars diverge into distinct dialects and registers](#).

Appointments held

2021-Present	Senior Lecturer , Linguistics, University of Canterbury
2021-Present	Leader of Language Technology Theme, New Zealand Institute for Language, Brain and Behaviour
2018-2020	Lecturer, Linguistics, University of Canterbury
2015-2018	Research Assistant Professor, Computer Science, Illinois Institute of Technology
2015-2018	Visiting Scientist, National Geospatial-Intelligence Agency
2014-2015	Research Fellow, Computer Science, Illinois Institute of Technology

Education

2013	PhD, Purdue University (Discipline: Linguistics) Dissertation: <i>Automatic Identification of Metaphoric Utterances</i>
2010	MA, Purdue University (Discipline: Linguistics) Thesis: <i>Towards a Computational Model of Metaphor</i>
2008	BA, Classics, Hillsdale College

Books

- Dunn, J. (2022). *Natural Language Processing for Corpus Linguistics*. Elements in Corpus Linguistics. Cambridge University Press. [Interactive Labs](#). [Overview](#).
- Dunn, J. (Under Contract). *Working with Construction Grammars: A Computational Approach*. Elements in Cognitive Linguistics. Cambridge University Press. [Overview](#).

MOOCs on edX

Now over 10,000 students

- [Text Analytics 1: Introducing Natural Language Processing](#)
- [Text Analytics 2: Visualizing Natural Language Processing](#)

Visiting Positions

- 2022 Visiting Scholar, Department of Cognitive Science, Case Western Reserve University
2022 Visiting Scholar, Center for Spatial Data Science, University of Chicago
2015 Faculty, LSA Linguistic Institute, University of Chicago

Grants, honours, awards

- 2021 Early Career Researcher Development Fund, University of Canterbury
Towards Equitable Language Models
- 2020 Science for Technological Innovation, National Science Challenges, MBIE
Domain Adaptation to Support Polynesian Language Technology
- 2020 Teaching Development Grant, University of Canterbury
Syntactic Annotations for Māori Corpora
- 2019 Support Grant, University of Canterbury
Automating Tourist Profiles By Integrating Spatial and Textual Artificial Intelligence
- 2015-2018 Visiting Scientist Research Fellowship, Oak Ridge Institute for Science and Education
Geolinguistics
- 2017 Certificate of Merit, Oak Ridge Institute for Science and Education
- 2014-2015 IC Postdoctoral Research Fellowship, Office of the Director of National Intelligence
Computational Cognitive Stylistics For Multi-Modal Identity Analytics
- 2013 Bilsland Dissertation Fellowship, Purdue University Graduate School
Automated Identification of Metaphoric Utterances
- 2011-2012 Bilsland Strategic Initiatives Grant, Purdue University Graduate School
Evaluating Cross-Listed Graduate Courses

Papers

- In Press 1. Li, H.; Dunn, J.; & Nini, A. (In Press). "Register Variation Remains Stable Across 60 Languages." *Corpus Linguistics and Linguistic Theory*.
- 2022 2. Li, H. & Dunn, J. (2022). "Corpus Similarity Measures Remain Robust Across Diverse Languages." *Lingua*, 275: 103377. [Link](#)
3. Dunn, J.; Li, H.; & Sastre, D. (2022). "Predicting Embedding Reliability in Low-Resource Settings Using Corpus Similarity Measures." In *Proceedings of the 13th International Conference on Language Resources and Evaluation*. 6461-6470. European Language Resources Association. [Link](#)
4. Dunn, J. & Nijhof, W. (2022). "Language Identification for Austronesian Languages." In *Proceedings of the 13th International Conference on Language Resources and Evaluation*. 6530-6539. European Language Resources Association. [Link](#).
5. Dunn, J. (2022). "Cognitive Linguistics Meets Computational Linguistics: Construction Grammar, Dialectology, and Linguistic Diversity." In Tay, D. & Xie Pan, M. (eds.), *Data Analytics in Cognitive Linguistics: Methods and Insights*. 273-308. Berlin: De Gruyter. [Link](#)

- 2021
6. Dunn, J. & Tayyar Madabushi, H. (2021). "Learned Construction Grammars Converge Across Registers Given Increased Exposure." In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*. Association for Computational Linguistics. 268-278. [Link](#)
 7. Dunn, J. (2021). "Representations of Language Varieties Are Reliable Given Corpus Similarity Measures." In *Proceedings of the EACL 2021 Workshop on NLP for Similar Languages, Varieties and Dialects*. Association for Computational Linguistics. 28-38. [Link](#)
 8. Dunn, J. & Nini, A. (2021). "Production vs Perception: The Role of Individuality in Usage-Based Grammar Induction." In *Proceedings of the NAACL 2021 Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics. 149-159. [Link](#)
- 2020
9. Dunn, J. (2020). "Mapping Languages: The Corpus of Global Language Use." *Language Resources and Evaluation*, 54(4): 999-1,018. doi: 10.1007/s10579-020-09489-2. [Link](#)
 10. Dunn, J. & Adams, B. (2020). "Geographically-Balanced Gigaword Corpora for 50 Language Varieties." In *Proceedings of the Conference on Language Resources and Evaluation*. European Language Resources Association. 2,521-2,529. [Link](#)
 11. Dunn, J.; Coupe, T.; & Adams, B. (2020). "Measuring Linguistic Diversity During COVID-19." In *Proceedings of the 4th Workshop on Natural Language Processing and Computational Social Science*. Association for Computational Linguistics. doi: 10.18653/v1/P17. [Link](#)
 12. Dunn, J. (2020). "Ontological and Grammatical Constraints on Metaphor Productivity." In Attardo, S. (ed.), *Script-Based Semantics: Foundations and Applications. Essays in Honor of Victor Raskin*. Berlin: De Gruyter. 55-76. doi: 10.1515/9781501511707-004. [Link](#)
 13. Mohammadhassan N., Mitrovic A., Neshatian K. & Dunn J. (2020). "Automatic Quality Assessment of Comments in Active Video Watching Using Machine Learning Techniques." In So H-J; Rodrigo M; Mason J; Mitrovic A (Eds). *Proceedings of the 28th International Conference on Computers in Education*. I: 1-10. Taiwan: Asia-Pacific Society for Computers in Education. [Link](#)
 14. Mohammadhassan N., Mitrovic A., Neshatian K. & Dunn J. (2020) "Developing Personalized Nudges to Improve Quality of Comments in Active Video Watching." In *Proceedings of 28th International Conference on Computers in Education 2*: 766-769. Taiwan: Asia-Pacific Society for Computers in Education. [Link](#)
- 2019
15. Dunn, J. (2019). "Global Syntactic Variation in Seven Languages: Towards a Computational Dialectology." *Frontiers in Artificial Intelligence*. doi: 10.3389/frai.2019.00015. [Link](#)
 16. Dunn, J. (2019). "Modeling Global Syntactic Variation in English Using Dialect Classification." In *Proceedings of the NAACL 2019 Sixth Workshop on NLP for Similar Languages, Varieties and Dialects*. Association for Computational Linguistics. 42-53. doi: 10.18653/v1/W19-1405. [Link](#)
 17. Dunn, J. (2019). "Frequency vs. Association for Constraint Selection in Usage-Based Construction Grammar." In *Proceedings of the NAACL 2019 Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics. 117-128. doi: 10.18653/v1/W19-2913. [Link](#)

18. Dunn, J. & Adams, B. (2019). "Mapping Languages and Demographics with Georeferenced Corpora." In *Proceedings of GeoComputation 19*. doi: 10.17608/k6.auckland.9869252.v2. [Link](#)
- 2018 19. Dunn, J. (2018). "Multi-Unit Association Measures: Moving Beyond Pairs of Words." *International Journal of Corpus Linguistics*, 23(2): 183-215. doi: 10.1075/ijcl.16098.dun. [Link](#)
20. Dunn, J. (2018). "Finding Variants for Construction-Based Dialectometry: A Corpus-Based Approach to Regional CxGs." *Cognitive Linguistics*, 29(2): 275-311. doi: 10.1515/cog-2017-0029. [Link](#)
21. Dunn, J. (2018). "Modeling the Complexity and Descriptive Adequacy of Construction Grammars." In *Proceedings of the Society for Computation in Linguistics*. Stroudsburg, PA: Association for Computational Linguistics. 81-90. doi: 10.7275/R59P2ZTB. [Link](#)
- 2017 22. Dunn, J. (2017). "Computational Learning of Construction Grammars." *Language and Cognition*, 9(2): 254-292. doi: 10.1017/langcog.2016.7. [Link](#)
- 2016 23. Dunn, J.; Argamon, S.; Rasooli, A.; & Kumar, G. (2016) "Profile-Based Authorship Analysis." *Digital Scholarship in the Humanities*, 31(4): 689-710. doi: 10.1093/llc/fqvo19. [Link](#)
- 2015 24. Dunn, J. (2015) "Modeling Abstractness and Metaphoricity." *Metaphor & Symbol*, 30(4): 259-289. doi: 10.1080/10926488.2015.1074801. [Link](#)
25. Dunn, J. (2015). "Three Types of Metaphoric Utterances That Can Synthesize Theories of Metaphor." *Metaphor & Symbol*, 30(1): 1-23. doi: 10.1080/10926488.2015.980694. [Link](#)
- 2014 26. Dunn, J. (2014). "Measuring Metaphoricity." In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics. 745-751. doi: 10.3115/v1/P14-2121. [Link](#)
27. Dunn, J. (2014). "Multi-Dimensional Abstractness in Cross-Domain Mappings." In *Proceedings of the ACL 2014 Second Workshop on Metaphor in NLP*. Stroudsburg, PA: Association for Computational Linguistics. 27-32. doi: 10.3115/v1/W14-2304. [Link](#)
28. Dunn, J.; Beltran de Heredia, J.; Burke, M.; Gandy, L.; Kanareykin, S.; Kapah, O.; Taylor, M.; Hines, D.; Frieder, O.; Grossman, D.; Howard, N.; Koppel, M.; Morris, S.; Ortony, A.; & Argamon, S. (2014). "Language-Independent Ensemble Approaches to Metaphor Identification." In *Proceedings of the 28th Conference on Artificial Intelligence: Workshop on Cognitive Computing for Augmented Human Intelligence*. 6-12. [Link](#)
- 2013 29. Dunn, J. (2013). "How Linguistic Structure Influences and Helps To Predict Metaphoric Meaning." *Cognitive Linguistics*, 24(1): 33-66. doi: 10.1515/cog-2013-0002. [Link](#)
30. Dunn, J. (2013). "Evaluating the Premises and Results of Four Metaphor Identification Systems." In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics, Vol. 1*. Heidelberg: Springer. 471-486. doi: 10.1007/978-3-642-37247-6_38. [Link](#)

31. Dunn, J. (2013). "What Metaphor Identification Systems Can Tell Us About Metaphor-in-Language." In *Proceedings of the NAACL 2013 First Workshop on Metaphor in NLP*. Stroudsburg, PA: Association for Computational Linguistics. 1-10. [Link](#)

2011 32. Dunn, J. (2011). "Gradient Semantic Intuitions of Metaphoric Expressions." *Metaphor & Symbol*, 26(1): 53-67. doi: 10.1080/10926488.2011.535416. [Link](#)

Teaching

ONLINE ONLY (MOOCs ON EDX)

[Text Analytics 1: Introducing Natural Language Processing](#)

[Text Analytics 2: Visualizing Natural Language Processing](#)

GRADUATE-LEVEL

Natural Language Processing (COSC 471), University of Canterbury

English Structures (LING 400), University of Canterbury

Data-Driven Computational Pragmatics, LSA Linguistic Institute, University of Chicago

Classroom Communication for International Graduate Students, Purdue University

SENIOR-LEVEL

Topics in Syntactic Theory (LING 306), University of Canterbury

MID-LEVEL

Grammatical Structure (LING 217), University of Canterbury

Text Analytics (LING 223), University of Canterbury

Forensic Linguistics (LING 225), University of Canterbury

ENTRY-LEVEL

The English Language (LING 101), University of Canterbury

Introduction to Linguistics (LING 227), Purdue University

First Year Composition for International Students (ENGL 101), Purdue University

First-Year Composition (ENGL 101), Purdue University

Residential Learning Community (ENGL 101), Purdue University

Teaching Awards

2012 Quintilian Award (Top 10% of Instructor Evaluations), Purdue University, Department of English

2012 Graduate Teacher Certificate, Center for Instructional Excellence, Purdue University

Supervisions

- 2021-Present Sidney Wong (Ph.D. Student, Linguistics)
2021-Present Matthew Durward (Ph.D. Student, Linguistics)
2020-2022 Negar Mohammadhassan (Ph.D. Student, Computer Science; Primary supervisor: Tanja Mitrovic)
2021-2022 Haipeng Li (Post-doc, New Zealand Institute for Language Brain and Behaviour)

Service to the Discipline

EDITORIAL WORK

- 2018-Present Associate Editor, *Frontiers in Artificial Intelligence* (Language and Computation)

AD HOC REVIEWING

- 2022 Conference on Computational Natural Language Learning (CoNLL)
Asia-Pacific Association for Computational Linguistics (AACL)
Fund for Scientific Research (F.R.S.-FNRS), Belgium
- 2021 North American Association for Computational Linguistics (NAACL)
Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (ACL-IJCNLP)
Empirical Methods in Natural Language Processing (EMNLP)
Conference on Computational Natural Language Learning (CoNLL)
Joint 15th Linguistic Annotation Workshop and 3rd Designing Meaning Representations
- 2020 Association for Computational Linguistics (ACL)
Conference of the European Chapter of the Association for Computational Linguistics (EACL)
Conference on Empirical Methods in Natural Language Processing (EMNLP)
Conference on Computational Natural Language Learning (CoNLL)
Asia-Pacific Association for Computational Linguistics (AACL)
Linguistic Annotation Workshop (LAW XIV)
Workshop on Figurative Language Processing
Workshop on NLP and Computational Social Science
Transactions on Audio, Speech and Language Processing
- 2019 *Journal of Cognitive Science*
IEEE Transactions on Audio, Speech and Language Processing
Conference on Computational Natural Language Learning (CoNLL)
Association for Computational Linguistics (ACL)
Linguistic Annotation Workshop (LAW XIII)
North American Association for Computational Linguistics (NAACL)
- 2018 Linguistic Society of America, 2019 Annual Meeting
Joint LAW-MWE-CxG Workshop (@ COLING)
IEEE Transactions on Audio, Speech and Language Processing (2x)
- 2017 Social Sciences and Humanities Research Council of Canada
Cognitive Linguistics
Language & Cognition
Linguistic Society of America, 2018 Annual Meeting
Corpus Linguistics and Linguistic Theory

- The Computer Journal*
- 2016 Linguistic Society of America, 2017 Annual Meeting
Fourth Workshop on Metaphor in NLP (@ NAACL)
- 2015 Linguistic Society of America, 2016 Annual Meeting
Synthesis Lectures in Human Language Technologies, Book Series (Morgan Claypool)
Journal of Artificial Intelligence Research
Language Resources and Evaluation
Third Workshop on Metaphor in NLP (@ NAACL)
- 2014 *IEEE Transactions on Audio, Speech and Language Processing*
Croatian Science Foundation
Second Workshop on Metaphor in NLP (@ ACL)
- 2012 International Conference on Computational Linguistics (COLING)
Purdue Linguistics Association Annual Symposium

COMPUTATIONAL LINGUISTIC ATLAS

www.earthLings.io

SOFTWARE AND DATA

C2xG: Computational Construction Grammar

Python package for learning, evaluating, and annotating CxGs

<https://www.github.com/jonathandunn/c2xg>

idNet: Identifying minority languages and regional varieties of majority languages

Python package for state-of-the-art LID and DID

<https://www.github.com/jonathandunn/idNet>

pacificLID: Identifying languages and code-switching with a focus on Austronesian languages

Python Version: https://github.com/jonathandunn/pacific_CodeSwitch

fastText Version: <https://jdunn.name/corpora/>

text_analytics: Package for teaching computational linguistics in Python

https://www.github.com/jonathandunn/text_analytics

corpus_similarity: Package for measuring corpus similarity and homogeneity

https://www.github.com/jonathandunn/corpus_similarity

Common Crawl Corpus: Corpus building from the common crawl data

https://www.github.com/jonathandunn/common_crawl_corpus

earthLings.io: Language mapping from digital datasets

Corpus of Global Language Use, 4.2: A 440 billion word geo-referenced web corpus

<https://www.earthLings.io>

<https://www.github.com/jonathandunn/earthLings>

University Service

COMMITTEES

2020-2021	Board of Studies, College of Arts, University of Canterbury
2018-2021	Board of Studies, College of Science, University of Canterbury
2019-2020	Research Committee, School of Language, Social & Political Sciences, University of Canterbury
2020-Present	Steering Group for Bachelors of Data Science Degree, University of Canterbury

ADMINISTRATIVE ACHIEVEMENTS

2020	Developed and delivered one of the flagship MOOCs for the University of Canterbury
2020	Developed the <i>Computational Linguistics</i> undergraduate major
2020	Helped to develop the undergraduate degree, <i>Bachelors of Data Science</i>
2019	Developed <i>Text Analytics</i> , 200-level course in Linguistics and Digital Humanities
2019	Developed <i>Language & Space</i> , 200-level course in Linguistics, Digital Humanities, and Geography

Partially Peer-Reviewed Publications

2018	1. Dunn, J. (2018). "Recursively Emerging Structure: A Discovery-Device CxG." In <i>Proceedings of the Chicago Linguistics Society</i> , 53: 71-86.
2017	2. Dunn, J. (2017). "Learnability and Falsifiability of Construction Grammars: A Learning-Based Approach." In <i>Proceedings of the Linguistic Society of America Annual Meeting</i> , 2(1): 1-15.
	3. Dunn, J. (2017). <i>Automating Human Geography With Dialectology</i> . Technical Report for National Geospatial-Intelligence Agency. Washington, D.C.
2015	4. Dunn, J. (2015). "Review of Frames and Constructions in Metaphoric Language (Constructional Approaches to Language, 14)." <i>Cognitive Linguistics</i> , 26(2): 371-375.
	5. Dunn, J. (2015). "Review of The Semantic Representation of Natural Language." <i>Studies in Language</i> , 39(2): 492-500.
2013	6. Dunn, J. (2013). "Review of Converging Evidence: Methodological and Theoretical Issues for Linguistic Research. (Human Cognitive Processing, 33)." <i>Cognitive Linguistics</i> , 24(4): 711-717.