

Dr Jonathan Dunn

(he/him)

University of Canterbury
Christchurch, New Zealand

Email jonathan.dunn@canterbury.ac.nz
Papers www.jdunn.name
Code www.github.com/jonathandunn
Data www.earthLings.io
Citations [Google Scholar](#)
MOOC [Text Analytics and Natural Language Processing with Python](#)

Current position

Senior Lecturer, Linguistics, University of Canterbury

Areas of specialisation

Computational Linguistics; Cognitive Linguistics; Syntax/Semantics

Appointments held

2021-Present Senior Lecturer, Linguistics, University of Canterbury
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
2015 Faculty, LSA Linguistic Institute, University of Chicago
2014-2015 Research Fellow, Computer Science, Illinois Institute of Technology

Education

2013 PhD, Linguistics, Purdue University
2008 BA, Classics, Hillsdale College

Grants, honours, awards

2020 Science for Technological Innovation, National Science Challenges, MBIE
Domain Adaptation to Support Polynesian Language Technology (\$200k NZD)
2020 Teaching Development Grant, University of Canterbury
Syntactic Annotations for Māori Corpora (\$5k NZD)
2019 Support Grant, University of Canterbury
Automating Tourist Profiles By Integrating Spatial and Textual Artificial Intelligence (\$5k NZD)

- 2015-2018 Visiting Scientist Research Fellowship, Oak Ridge Institute for Science and Education
Geolinguistics (approx \$515k USD)
- 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 (approx \$50k)
- 2011-2012 Bilsland Strategic Initiatives Grant, Purdue University Graduate School
Evaluating Cross-Listed Graduate Courses (approx \$50k)

Publications

BOOKS

- Under Contract 1. Dunn, J. *Computational Linguistics for Corpus Analysis*. Cambridge University Press.

ARTICLES

- 2021 1. Dunn, J. (2021). "Representations of Language Varieties Are Reliable Given Corpus Similarity Measures." In *Proceedings of the EACL 2021 Eighth Workshop on NLP for Similar Languages, Varieties and Dialects*. Association for Computational Linguistics.
2. Dunn, J. (In Press). "Cognitive Linguistics Meets Computational Linguistics: Construction Grammar, Dialectology, and Linguistic Diversity." In Tay, D. & Xie Pan, M. (eds.), *Data Analytics for Cognitive Linguistics: Methods and Insights*. Berlin: De Gruyter.
- 2020 3. 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
4. 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.
5. 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
6. 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
7. 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.
- 2019 8. Dunn, J. (2019). "Global Syntactic Variation in Seven Languages: Towards a Computational Dialectology." *Frontiers in Artificial Intelligence*. doi: 10.3389/frai.2019.00015

9. Dunn, J. & Adams, B. (2019). "Mapping Languages and Demographics with Georeferenced Corpora." In *Proceedings of GeoComputation 19*. doi: 10.17608/k6.auckland.9869252.v2
10. 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
11. 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
- 2018 12. 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
13. 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
14. 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
- 2017 15. Dunn, J. (2017). "Computational Learning of Construction Grammars." *Language and Cognition*, 9(2): 254-292. doi: 10.1017/langcog.2016.7
- 2016 16. 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/fqw019
- 2015 17. Dunn, J. (2015) "Modeling Abstractness and Metaphoricity." *Metaphor & Symbol*, 30(4): 259-289. doi: 10.1080/10926488.2015.1074801
18. 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
- 2014 19. 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
20. 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
21. 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.
- 2013 22. 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

23. 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
24. 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.
- 2011 25. Dunn, J. (2011). "Gradient Semantic Intuitions of Metaphoric Expressions." *Metaphor & Symbol*, 26(1): 53-67. doi: 10.1080/10926488.2011.535416

NON-PEER-REVIEWED

- 2018 1. Dunn, J. (2018). "Recursively Emerging Structure: A Discovery-Device CxG." In *Proceedings of the Chicago Linguistics Society*, 53.
- 2017 2. Dunn, J. (2017). "Learnability and Falsifiability of Construction Grammars: A Learning-Based Approach." In *Proceedings of the Linguistic Society of America Annual Meeting*, 1: 1-15.
3. Dunn, J. (2017). *Automating Human Geography With Dialectology*. 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)." *Cognitive Linguistics*, 26(2): 371-375.
5. Dunn, J. (2015). "Review of The Semantic Representation of Natural Language." *Studies in Language*, 39(2): 492-500.
- 2013 6. Dunn, J. (2013). "Review of Converging Evidence: Methodological and Theoretical Issues for Linguistic Research. (Human Cognitive Processing, 33)." *Cognitive Linguistics*, 24(4): 711-717.

CONFERENCE PRESENTATIONS (NOT PUBLISHED)

- 2018 1. Dunn, J. (2018, November). "Bootstrapping Corpora for Under-Represented Languages: The Case of Māori." *Annual Meeting of the New Zealand Linguistic Society*. Wellington, NZ.
2. Dunn, J. (2018, January). "Corpus-Based Dialectometry Using Construction Grammars." *Annual Meeting of the Linguistic Society of America*. Salt Lake City, UT.
- 2017 3. Dunn, J. (2017, November). "Finding Variants in a Dynamic Feature Space: Classification as Validation." *New Ways of Analyzing Variation* 46. Madison, WI.
4. Dunn, J. (2017, September). "Automating Human Geography With Dialectology." *IC Academic Research Symposium*. Washington, D.C.
5. Dunn, J. (2017, May). "Recursively Emerging Structure: A Discovery-Device CxG." *53rd Annual Meeting of the Chicago Linguistic Society*. Chicago, IL.
6. Dunn, J. (2017, January). "Learnability and Falsifiability of Construction Grammars: A Learning-Based Approach." *Annual Meeting of the Linguistic Society of America*. Austin, TX.

- 2016 7. Dunn, J. (2016, September). "Finding Variants for Unsupervised Computational Dialect Mapping." *IC Academic Research Symposium*. Washington, D.C.
8. Dunn, J. Argamon, S. (2016, January). "Evaluating Unsupervised Profile Bundles for Authorship Analysis." *The Association for Linguistic Evidence Winter Meeting*. Washington, D.C.
- 2015 9. Dunn, J.; Rasooli, A.; Argamon, S. (2015, May). "Profile-based Authorship Matching." *Midwest Speech and Language Days*. Chicago, IL.
10. Dunn, J. Argamon, S. (2015, January). "The Linguistic Status of Predictions and Feature Ranks from SVM Text Classifiers." *Annual Meeting of the Linguistic Society of America*. Portland, OR.
11. Dunn, J. Argamon, S. (2015, January). "Evaluating Profile-Based Authorship Analysis for Distinguishing Similar Individuals." *Association for Linguistic Evidence Winter Meeting*. Portland, OR.
- 2014 12. Argamon, S. Dunn, J. (2014, October). "Predicting Political Ideology from Congressional Speeches." *New Directions in Analyzing Text as Data Conference*. Chicago, IL.
13. Dunn, J. (2014, April). "Computational Cognitive Stylistics for Multi-Modal Identity Analytics." *Annual IC Postdoctoral Research Fellowship Colloquium*. Washington, D.C.
14. Dunn, J. (2014, January). "Computational Evidence for Direct Metaphoric Meaning." Annual Meeting of the Linguistic Society of America. Minneapolis, MN.
- 2013 15. Dunn, J. (2013, June). "Three Kinds of Metaphor-in-Language and How We Can Use Them to Synthesize Theories of Metaphor." *International Cognitive Linguistics Conference*. Edmonton, Canada.
- 2012 16. Dunn, J. (2012, November). "Does Humor Interfere with Machine Learning Approaches to Metaphor Identification?" *Association for the Advancement of Artificial Intelligence, Fall Symposium: The Artificial Intelligence of Humor*. Washington, D.C.
- 2009 17. Dunn, J. (2009, October). "Abstract Semantic Structure Versus Conceptual Metaphor Theory." *Mid-America Linguistics Conference*. Columbia, Missouri.
18. Dunn, J. (2009, November). "A Deeper Dialect: The Linguistic Atlas of Late Medieval English, Morphophonology, and Word Formation in the Pearl-Poet." *Midwest Modern Language Association*. St. Louis, Missouri.

Teaching

ONLINE ONLY (MOOCs)

[Introduction to Text Analytics and Natural Language Processing in Python, edX](#)
[Visualizing Text Analytics and Natural Language Processing in Python, edX](#)

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

2020-Present Negar Mohammadhassan (Ph.D. Student, Computer Science)
Primary supervisor: Tanja Mitrovic
2021-Present 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

- 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)
- 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 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 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

SOFTWARE AND DATA

- 2016-2019 **C2xG**: Computational Construction Grammar
Python package for learning, evaluating, and annotating CxGs
Grammars for ara, deu, eng, fra, por, rus, spa, zho
<https://www.github.com/jonathandunn/c2xg>
- 2017-2018 **idNet**: Identifying minority languages and regional varieties of majority languages
Python package for state-of-the-art LID and DID
LID: Coverage for 464 languages with sample sizes > 50 chars
DID: Coverage for over 100 varieties of eight major languages
<https://www.github.com/jonathandunn/idNet>
- 2017-2019 **Common Crawl Corpus**: Corpus building from the common crawl data
https://www.github.com/jonathandunn/common_crawl_corpus
- 2019-2020 **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>
- 2020-2021 **text_analytics**: Package for teaching computational linguistics in Python
https://www.github.com/jonathandunn/text_analytics

University Service

COMMITTEES

- 2020-Present Board of Studies, College of Arts, University of Canterbury
- 2018-Present Board of Studies, College of Science, University of Canterbury
- 2019-Present 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 *Computational Linguistics* undergraduate major
- 2020 Helped to develop the undergraduate degree, *Bachelors of Data Science*
- 2019 Developed *Text Analytics*, 200-level course in Linguistics and Digital Humanities
- 2019 Developed *Language & Space*, 200-level course in Linguistics, Digital Humanities, and Geography

Last updated: April 3, 2021