

Terrence D. Szymanski, PhD

terry.szymanski@gmail.com
+61 451690474
www.affrication.org

Summary

I am a **data scientist** and **computational linguist**, with a research background in both Linguistics and Computer Science. I use machine learning and natural language processing to build models from unstructured text data to advance the state of the art in my field, and to solve business problems.

Education

2012. University of Michigan. Ph.D. Linguistics.

2003. University of Michigan. B.S.E. Electrical Engineering *magna cum laude*.

2003. University of Michigan. B.A. Classical Languages and Literature *with distinction*.

Experience

2016-present **Data Scientist**, Australia and New Zealand Banking Group (ANZ). Melbourne, Australia

- Designed and developed a financial news analytics platform in Python using AWS cloud infrastructure, custom algorithms, and third-party APIs.
- Support text data integration across multiple data silos within ANZ.
- Helped develop a Python-based customer analytics pipeline.
- Experience with an agile Kanban project management environment.

2014-2016 **Postdoctoral Researcher**, Insight Centre for Data Analytics, UCD, Ireland

- Researched text analytics both independently and within collaborative industry partnerships
- Delivered software and technical reports to industry partners on strict deadlines
- Managed teams of students and researchers on multiple collaborative research projects
- Supervised one Masters thesis and mentored two PhD students

2012-2014 **CTO and Cofounder**, Savvy Languages LLC, USA

- Cofounded a language education startup, raised initial funding and completed the TechTown Venture Accelerator program at Wayne State University in Detroit, Michigan.
- Developed a Javascript / Python LAMP web application for online language learning
- Recruited, hired, and managed a software development intern.

2006-2012 **PhD Student**, Department of Linguistics, University of Michigan, USA

- Thesis: *Morphological Inference from Bitext for Resource-Poor Languages*, Advisor: Steven Abney
- Performed independent and collaborative research on multiple projects
- Taught undergraduate courses as teaching assistant and as primary instructor

Skills

- Strong Python skills: Jupyter notebooks, pandas, sklearn, numpy, gensim, nltk, matplotlib
- Strong Linux skills: bash, grep, find, git, emacs, latex
- Web development: AWS EC2, S3, apache, flask, django, jQuery, AngularJS
- Other programming: SQL (MS SQL, Teradata), NoSQL, Java, .NET, R, C/C++, et al.
- I read Ancient Greek and Latin, and I've studied Arabic and Irish as a beginner
- I have a strong background in formal / theoretical linguistics
- I have extensive university-level teaching experience as a lecturer and teaching assistant

Recent Projects

- **Finance and Business News Analytics:** I created a system to collect and analyze finance and business news from the web, running each news article through a NLP pipeline and then performing unsupervised machine learning (e.g. topic modeling and clustering of named entities). The results are presented in a web application for users to discover insights about companies, commodities, etc. mentioned in the news. (*Amazon Web Services, OpenCalais, Python, sklearn, flask*)
- **Irish Beef Markets:** I analyzed a 10-year corpus of 140k farming news stories: data cleaning, parsing, topic modeling, event detection and time-series term alignment with market prices. I also performed exploratory data analysis and visualization of 5 years of numeric cattle mart transaction data. (*Python, Jupyter Notebook, pandas, matplotlib, gensim, dynamic-NMF, CoreNLP, numpy, sklearn, R, stl*)
- **Headline Optimization:** I collaborated on a system to recommend keywords for online news stories, using named entity recognition, keyword scoring, and supervised regression to predict Facebook/Twitter traction. I also built and deployed a cloud-based web interface for the system. (*Python, flask, sklearn, pandas, Java, Stanford CoreNLP, corenlp-watcher, AWS EC2, Apache*).
- **Diachronic Word Embeddings:** I developed a linear regression method to align independently-trained word2vec models, to study lexical substitutions and changes in word meanings over time. (*Python, gensim, word2vec, numpy, sklearn, word embeddings, vector space transformations*)
- **Diachronic Text Evaluation:** I submitted the top-scoring system on a *SemEval* 2015 task for predicting the date (between 1700 and 2015) when a short sample of news text was written, using a SVM multi-class classifier with multiple ngram features (character, word, and syntactic). (*Weka, SVM, multi-class classification, cross-validation, feature selection, dependency parsing, Google N-Grams*)

Academic Publications

- T. Szymanski. 2017. *Temporal Word Analogies: Identifying Lexical Replacement with Diachronic Word Embeddings*. To appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL).
- Y. Gurin, T. Szymanski and M. T. Keane. 2017. *Discovering News Events That Move Markets*. To appear in Proceedings of Intelligent Systems Conference 2017 (IntelliSys).
- T. Szymanski, C. Orellana-Rodriguez, and M. T. Keane. 2016. *Helping News Editors Write Better Headlines: A Recommender to Improve the Keyword Contents & Shareability of News Headlines*. In Proceedings of Natural Language Processing Meets Journalism.
- T. Szymanski and G. Lynch. 2015. *UCD: Diachronic Text Classification with Character, Word, and Syntactic N-grams*. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval).
- T. Szymanski. 2013. *Automatic Extraction of Linguistic Data from Digitized Documents*. Proceedings of the Berkeley Linguistics Society 39.
- T. Szymanski. 2012. *Morphological Inference from Bixtext for Resource-Poor Languages*. Ph.D. thesis. University of Michigan.
- E. Keshet, T. Szymanski, and S. Tyndall. 2011. *BALLGAME: A Corpus for Formal Semantics*. Proceedings of the Ninth International Conference on Computational Semantics.

Supervision and Mentorship

2014-16	mentor	PhD Student	Yuri Guerin, University College Dublin
2014-16	mentor	PhD Student	Claudia Orellana-Rodriguez, University College Dublin
2016	co-supervisor	MSc Student	Jinghui Lu, University College Dublin
2015	supervisor	research intern	Mengtong Li, Fudan University and University College Dublin
2013	supervisor	software intern	Jacob Howcroft, Savvy Languages
2011	co-supervisor	research interns	BALLGAME annotation team, University of Michigan