on the basis of external criteria such as formality and genre, rather than simply date, and selectively applying specific lemmatizer models to subsets of a corpus.

The main contribution of this paper is its illustration of the importance of targeting machine learning tools toward specific datasets. Through attempting to target Hellenistic Greek, we identified errors and issues for lemmatizing Hellenistic Greek texts, provided evidence that annotations of Ancient Greek texts is less adequate for model training than the Greek New Testament, and provided an initial foray into the use of word space tools in this area of research.

References

[1] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, volume 3, pages 993–1022.

[2] Louw, Johannes P. and Eugene A. Nida. (Eds.). 1988. *Greek-English Lexicon of the New Testament Based on Semantic Domains* (Vols. 1–2). New York: United Bible Society.

[3] Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR, 2013*, pages 1–12, Scottsdale, AZ.

[4] Müller, Thomas, Helmut Schmid, and Hinrich Schütze. 2013. Efficient Higher-Order CRFs for Morphological Tagging. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 322–332, Seattle.

[5] O'Donnell, Matthew B. 2005. *Corpus Linguistics and the Greek of the New Testament*, pages 136, 164–65, Sheffield: Sheffield Phoenix.

[6] Pang, Francis G. H. 2016. *Revisiting Aspect and Aktionsart: a Corpus Approach to Koine Greek Event Typology*, pages 6-35, Leiden: Brill.

[7] Řehůřek, Radim and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta: ELRA.

[8] Sahlgren, Magnus. 2006. *The Word-Space Model: Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations Between Words in High-Dimensional Vector Spaces*. Ph.D. thesis, Stockholm University, Stockholm.

[9] Sievert, Carson and Kenneth E. Shirley. 2014. LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pages 63–70, Baltimore, MD: Association for Computational Linguistics.