**A Computational Approach to Analysing the Corpus of Oz Early English**

Martin Schweinberger

This study takes a machine-learning approach to analyse the Corpus of Oz Early English (COOEE). The COOEE includes 1353 samples of texts written in Australia, New Zealand or Norfolk Island, or by native Australians on travels, between 1788 and 1900. It includes a selection of unpublished letters and published books, and historical texts. The aim of the study is to exemplify how computational methods and natural language processing can assist historical linguists, dialectologists, and corpus linguists in unearthing patterns and topics that would be hard to identify using traditional methods.

By using keyword extraction, topic modelling, and error analysis, this study shows which themes were prominent in early Oz English, it highlights changes in the prominence across time, and it allows drawing inferences about linguistic characteristics of early Oz English. A Latent Dirichlet Allocation-based topic model reports that five topics figure prominent in the texts ranging from descriptions of new homes and living situations, over love letters, and military action, to family life, and contacts with indigenous people.

An additional case study on the use and use of adjective amplifiers finds remarkable stability in this domain. While the stability in the amplifier system aligns with recent research on Irish English (cf. Schweinberger 2021), the results reported here contrast with claims stating that amplifier systems are prone to change (Brinton and Arnovik 2006: 441).

The results presented here exemplify how dialectology, historical linguistics, and studies describing the content of collections of electronic texts can profit from adopting computational methods developed in natural language processing.

**References**

Brinton, Laurel J. and Leslie K. Arnovik. 2006. *The English language. A linguistic history*. Oxford: Oxford University Press.

Schweinberger, Martin. 2021. On historical developments in the Irish English intensifier system. *Anglistik – International Journal of English Studies*: 32(1): 139-158. (Special Issue: Focus on English Linguistics: Varieties meet Histories, Eds. Daniela Kolbe-Hannah & Ilse Wischer)

Words: 320 (with references), 266 (without references)