

## **Exploring Annotation Methods: Enriching a Corpus of 18th and 19th-Century Scottish Pauper Letters**

Moragh S. Gordon (Leiden)

The newly compiled corpus of 18th and 19th-century Scottish pauper letters and petitions (ScotPP) offers a unique window into the linguistic practices of the lower classes. While the linguistic shift from written Scots to written (Scottish) Standard English has been extensively documented (e.g., Romaine, 1982; Devitt, 1989; Meurman-Solin, 1993; Dossena, 2007), the role of lower-class writers in this process remains largely unexplored, particularly in the so-called post-anglicization period, i.e. the 18th and 19th centuries (Van Eyndhoven, 2023). The ScotPP corpus aims to address this gap, but its analysis presents some challenges, particularly in the domain of automated corpus annotation.

This talk explores the prospects and challenges of applying automated annotation techniques to a corpus characterized by highly variable and idiosyncratic spelling. Through a case study focusing on past tense and past participle forms, I aim to demonstrate how such variation impacts annotation processes, complicating tasks such as part-of-speech tagging and morphological analysis. At the same time, despite the challenges posed by variability and non-standardized features, the aim of this talk is to explore the benefits that automated methods may offer for advancing the study of language variation and change from below.

### *References*

- Devitt, A. (1989). *Standardising written English: Diffusion in the case of Scotland, 1520–1659*. Cambridge University Press.
- Dossena, Marina. 2007. "As this leaves me at present"—Formulaic usage, politeness and social proximity in nineteenth-century Scottish emigrants' letters. In Stephan Elspaß, Nils Langer, Joachim Scharloth & Wim Vandenbussche (Eds.), *Germanic Language Histories 'from Below' (1700–2000)*. Berlin: Walter de Gruyter, 13-29  
<https://doi.org/10.1515/9783110925463.13>
- van Eyndhoven, S. (2024). "Quhen I am begun to write I really knou not what to say": Inter- and Intra-writer Variation in the Use of <quh-> in Early Modern Scots. In T. Porck, M.S. Gordon & L. Caon (Eds.) *Conference Proceedings of the 21st International Conference of English Historical Linguistics (ICEHL21)*, Amsterdam: John Benjamins, 225-250
- Meurman-Solin, A. (1993). *Variation and change in early Scottish prose: Studies based on the Helsinki Corpus of Older Scots*. Suomalainen Tiedeakatemia.
- Romaine, S. (1982). *Socio-historical linguistics: Its status and methodology*. Cambridge University Press
- ScotPP = *Scottish Pauper Letters and Petitions Corpus*. Compiled by Moragh Gordon & Hester Groot [under construction].