The Devil's in the Difference: Tracking Phonological Change in Witch Trial Papers

Christa Schneider || University of Bern / Digital Humanities

One of my studies on premodern handwritten criminal records explores the evolution of vernacular orthography in Early Modern Bernese German through the lens of historical phonology and computational sociolinguistics. Drawing on a corpus of witch trial papers from the city-state of Bern (Switzerland), dated between 1551 and 1688 and comprising approximately 90,000 tokens, I employ Levenshtein distance as a quantitative method to measure the degree of phonological divergence between attested vernacular spellings and lemma forms documented in the *Frühneuhochdeutsches Wörterbuch* (FnhdWb = Early Modern German dictionary). Levenshtein distance is a string metric that calculates the minimum number of single-character edits - insertions, deletions, or substitutions - needed to convert one word into another. In this study, it serves as a proxy for measuring the orthographic and, by extension, phonological distance between vernacular spellings in the corpus and their lemma forms in the *Frühneuhochdeutsches Wörterbuch*.

The corpus, composed of scribal renderings of spoken testimony in high-stakes judicial settings, preserves a rich record of non-standard and regionally marked spellings. These forms are shaped not only by phonological features of the Bernese German dialect but also by scribal practice, orthographic norms (and particularly the lack thereof as there were only little to no written language norms in use), and the broader sociopolitical context of post reformation-era Bern. By aligning word forms in the corpus with their corresponding FnhdWb lemmas and calculating Levenshtein distances, I track proximity as a proxy for phonological (and orthographic) convergence or divergence from supraregional norms.

I will present a diachronic analysis that examines whether scribal practices became more or less aligned with dictionary-attested forms over the 137-year period. Across this timespan, the education and professionalisation of scribes changed drastically, a shift which likely influenced their approach from phonologically motivated, vernacular orthography towards linguistic standardisation. I will also investigate the potential influence of social, institutional, and political factors on phonologically motivated spelling variation. These include the role of religious reform, literacy policy, bureaucratic training, and the interaction between oral testimony and written record.

Methodologically, the paper contributes to the use of distance metrics in historical (socio-)linguistics by demonstrating how measuring distance can serve not just as a tool for OCR correction or spelling normalisation, but as an empirical measure of linguistic and sociolinguistic change. Additionally, I will discuss the challenges of aligning vernacular and lemma forms in a low-resource, dialect-rich setting.

By quantifying distance between phonologically-grounded vernacular writing and lexicographic standardisation, this study offers new insight into the historical development of scribal education and practice in a multilingual, transitional phonological landscape. It highlights the value of computational metrics in tracing the social embedding of phonological change, and provides a framework for extending such methods to other historical corpora and language varieties.

Bibliography available upon request, has been deleted due to restriction in length.