

“Designing a metrically annotated corpus of historical English verse”

Andrew Cooper

Stockholm university, Sweden

This talk presents a corpus under development, the intended purpose of which is to demonstrate variation and change in the metre of English verse throughout its history. This is an unusual goal and requires the development of a phonologically annotated corpus of historical verse texts. Creating a corpus for a new purpose requires careful thought, planning and consultation. This talk covers the design of this corpus, starting with the text selection and sampling procedure and a summary of the annotation principles. This talk covers the design and text selection processes of this corpus, along with a summary of the annotation principles.

To produce a corpus of this kind, a number of design criteria must be followed, as described by Lass (2004). Good design is essential if the corpus is to be useful for a variety of purposes. Because of this, it is important that the collection and annotation criteria are properly justified and prioritised. Multiple corpora of English historical verse already exist, but they were collected without regard to phonological factors. The development of a metrically annotated phonological corpus of historical English is an entirely new area. Phonological corpora have become more common in the last decade, and annotation techniques are well-established (Durand, Gut, & Kristoffersen, 2014). These corpora have been annotated for segment quality, stress patterns and syllabic weight, but it is not usual to incorporate all three features (Durand, 2017). The only attempt to create a large corpus of metrically annotated *verse* has been for contemporary Czech (Plecháč & Kolár, 2015).

Texts will be sampled from the time range 950-1600 CE. These must be divided into two major traditions – alliterative and rhyming. In the earlier alliterative form there are around 30,000 surviving lines of Old English verse, but only a few long texts. Middle English alliterative verse diverged from this form, disappeared and underwent a revival, and leaves several surviving long poems by a small number of authors. A few tens of thousands of lines of Middle English alliterative poetry are extant. There is an overlap with the more common rhyming tradition, which has a much more structured line, and which continues into the modern era. The end date 1600 coincides with the beginning of the English renaissance and the development of individual metrical style, which is not an intended area of study for the project. The annotation will cover line length, stress, metrical weight and rhyme, including alliteration. The annotation of these features should be enough to allow the structure of each line of verse to be revealed automatically, without its metrical form needing to be specified.

In addition to presenting the existing ideas for the design of this corpus, I would like to consult the conference for advice on its development.

Works cited

Durand, J. (2017). Corpus Phonology. *Oxford Research Encyclopedia of Linguistics*.

<https://doi.org/10.1093/acrefore/9780199384655.013.145>

Durand, J., Gut, U., & Kristoffersen, G. (2014). Introduction. *The Oxford Handbook of Corpus Phonology*. <https://doi.org/10.1093/oxfordhb/9780199571932.013.032>

Lass, R. (2004). Ut custodiant litteras: Editions, Corpora and Witnesshood,. In M. Dossena & R. Lass (Eds.), *Methods and data in English historical dialectology* (pp. 21–48). Bern: Peter Lang.

Plecháč, P., & Kolár, R. (2015). The Corpus of Czech Verse. *Studia Metrica et Poetica*, 2(1), 107–118. <https://doi.org/10.12697/smp.2015.2.1.05>