Phonological and metrical variation across genres

Arto Anttila and Ryan Heuser
Stanford University

Problem. Speech and writing are rhythmically structured in ways that vary across individuals, styles, and genres. In metrical verse, the natural rhythm of speech is set against a conventional meter that is recognized by hearers and readers, creating a tension the poet manipulates for artistic effect. For example, the ten-syllable sentence *I can't believe that I forgot my keys* is easily recognizable as iambic pentameter ws/ws/ws/ws/ws/ whereas another ten-syllable sentence *It rains almost always when I visit* is not (Steele 1999). In this study, we asked whether standard phonological and metrical constraints proposed by phonologists and metricists on independent grounds can reliably identify arbitrary lines of text as (metrical) verse vs. (non-metrical) prose. We focused on two unrelated languages, English and Finnish.

Data. Our data come from nine English and nine Finnish authors (https://www.gutenberg.org/): Keats, Shelley, Whitman, Wordsworth, Yeats (both prose and verse); Hopkins, Milton, Pope, Shakespeare (only verse); Erkko, Kaatra, Leino, Lönnrot, Siljo (both prose and verse); Hellaakoski, Kailas, Koskenniemi, Kramsu (only verse). We converted all texts to versions of themselves in which each line has exactly five words, with no punctuation, in order to guarantee that any phonological or metrical difference between prose and verse that might emerge would have nothing to do with line length, but only with the local phonological and metrical arrangement of words. Our dataset consists of 500 randomly sampled lines for each author-genre pair, totaling approximately 14,000 lines.

Analysis. To analyze the dataset phonologically and metrically we used PROSODIC (Heuser, Falk, and Anttila 2010-2011, https://github.com/quadrismegistus/prosodic), a software package that provides a phonological analysis and metrical scansion for raw text. While less accurate than hand-coding (see, e.g., Hayes, Wilson, and Shisko 2012 for a recent example), machine analysis yields a reasonable baseline and opens up much larger datasets. The phonological analysis syllabifies the data and annotates it for stress and weight using the CMU Pronouncing Dictionary (Weide 1998) and OpenMary (http://mary.dfki.de/), allowing for stress ambiguity in monosyllabic function words (e.g., *have* vs. *háve*) based on a classification informed by Hirschberg 1993. The metrical analysis provides a scansion based on constraints from Hanson and Kiparsky 1996 (H&K). These constraints regulate the correspondence between metrical positions (s vs. w) and their phonological realization, governing position size (syllable vs. foot), prominence site (s vs. w), and prominence type (stress vs. strength vs. weight). H&K make the interesting claim that mainstream metrical traditions in Finnish and English differ in prominence site and type: in Finnish iambic-anapestic (trochaic-dactylic) meters a strong metrical position may not contain an unstressed syllable; in Shakespeare's iambic pentameter a weak metrical position may not contain a strong syllable, where strength is defined as follows: a constituent is strong iff it is the head of a branching constituent and weak iff it is the non-head of a branching constituent. Thus, in *mány* the stressed syllable is strong, a.k.a. "peak", and the unstressed syllable is weak, a.k.a. "trough", whereas *kéen* is neither. These metrical constraints are observed to varying degrees by individual poets.

Given a line of text, PROSODIC starts from a candidate space of possible s/w scansions; for a line of 10 syllables the upper bound is 2^{10} = 1,024 scansions. PROSODIC assigns each scansion a constraint violation vector, discards harmonically bounded scansions in the sense of Optimality Theory (Prince and Smolensky 1993/2004), allowing for resolution in weak positions which may contain up to two syllables, and returns the remaining scansions, with violation patterns for each phonological and metrical constraint. Stress ambiguities in monosyllabic function words (e.g., *have* vs. *háve*) are resolved by scansion. Violation counts are normalized by dividing the sum of violations by the number of scansions and the number of syllables in the line. We assumed four phonological constraints: PEAKPROM 'No stressed lights', WSP 'No unstressed heavies', NOCLASH 'No adjacent stressed syllables', and NOLAPSE 'No adjacent unstressed syllables' (see, e.g., Prince 1990, Prince and Smolensky 1993), and four metrical constraints drawn from H&K, which crucially include *S/UNSTRESSED 'A strong position may not contain an unstressed syllable' and *W/PEAK 'A weak position may not contain a peak'. For example, the first foot of the line *Néver/ cáme pói/son fróm/ so swéet/ a pláce/* violates *W/PEAK on the weak beat and *S/UNSTRESSED on the strong beat (inversion). These two constraints embody the key difference between Finnish and English meters noted by H&K. Since PROSODIC blindly analyses any text, verse or prose, the resulting violation profiles yield rich information about the phonological and metrical differences among texts. This information is interesting because it allows us to figure out how verse differs from prose and to put H&K's claim to empirical test.

Results. We modeled the data using mixed-effects logistic regression using the R lme4 package (Bates et al. 2013), with genre (prose vs. verse) as the dependent variable, the four phonological and two metrical constraints as independent variables, with violation counts normalized and centered, and author as a random variable. Three main discoveries emerged.

First, the purely phonological constraints register the same differences between prose and verse in both languages, suggesting that in this sense phonology is universal. Violations of PEAKPROM, WSP, and NOLAPSE are highly predictive of prose in both languages ($p = 0.01$-0.001) showing that such violations are avoided in verse. In contrast, violations of NOCLASH are highly predictive of verse in both languages ($p = 0.001$) showing that such violations are avoided in prose (Shih 2014). We note that the number of PEAKPROM and WSP violations almost completely depends on word choice (up to stress ambiguity) whereas the number of NOCLASH and NOLAPSE violations is sensitive to both word choice and linearization. This suggests that word choice and possibly linearization are sensitive to genre in the same way in both languages.

Second, the metrical constraints register the difference between prose and verse differently in the two languages, presumably because metrical traditions are language-specific. We found H&K's claim about the difference between Finnish and English to be supported: in Finnish, violations of *S/UNSTRESSED 'A strong position may not contain an unstressed syllable' are predictive of prose ($p = 0.05$) while violations of *W/PEAK 'A weak position may not contain a peak' do not reach significance. In contrast, in English violations of *W/PEAK are highly predictive of prose ($p = 0.001$) and violations of *S/UNSTRESSED are highly predictive of verse ($p = 0.001$). This is consistent with the view that English verse controls weak positions and cares about strength whereas Finnish verse controls strong positions and cares about stress, hence the prose vs. verse difference is most clearly visible in exactly these prominence sites and types.

Third, we observe that on an average the number of possible scansions is larger in prose than verse. This is not surprising: one would expect prose which by definition does not have meter to be metrically more ambiguous and allow more scansions than metrical verse.