

Inferring stress placement (in)stability from a poetic corpus

Alexander Piperski
Russian State University for the Humanities
Higher School of Economics
apiperski@gmail.com

Anton Kukhto
Massachusetts Institute of Technology
kukhto@mit.edu

In this talk, we pursue two goals: (a) to show how diachronic corpora can be used in assessing stress placement stability and (b) to address the importance of variability in defining and operationalising phonological stability. The empirical testing grounds is lexical stress in Russian, a system that, needless to say, has received a lot of attention in the literature (Halle 1973, Zalizniak 1985, Melvold 1990, a.m.o.). The variability of stress assignment in Russian is less well understood, however, and this talk aims to contribute to the study of this topic.

Our data come from the Poetic subcorpus of the Russian National Corpus (www.ruscorpora.ru), which contains 11m tokens and which is the only resource providing large amounts of accentuated (and annotated for author) texts from various time periods. Even though poetic text imposes certain limitations on projects of the type we undertake, we intend to show that it can be a useful and revealing source especially in the situation when other sources are lacking.

So far, we analyzed the works by 12 Russian poets of the 19th and the 20th centuries and extracted all word forms with variable stress from the texts of each poet. The quantitative summary of our data is presented in Table 1.

Poet	Years of life	Texts	Tokens	Word form types	Word form types with variable stress
Küchelbecker	1797–1846	81	13,420	5,924	10
Pushkin	1799–1837	855	182,014	35,377	187
Yazykov	1803–1846	354	59,008	15,904	41
Lermontov	1814–1841	441	125,883	23,122	120
A. Maykov	1821–1897	553	107,696	26,196	135
Vyach. Ivanov	1866–1949	1025	103,357	28,717	122
Kuzmin	1872–1936	553	57,745	17,860	63
Gumilev	1886–1921	446	57,390	17,245	61
Tvardovsky	1910–1971	306	101,448	21,272	43
Simonov	1915–1979	204	51,332	14,505	30
Samoylov	1920–1990	751	58,179	18,907	15
Gandlevsky	1952–	103	13,890	7,172	5

For instance, the list of word forms with variable stress for Küchelbecker is as follows: *vozvýšennyx* × 1 ~ *vozvyšennyx* × 2 ‘elevated (GEN.PL)'; *dáli* × 1 ~ *dalí* × 1

'distance (GEN.SG)'; *Kamóens* × 1 ~ *Kamoéns* × 1 'Camões'; *króvi* × 2 ~ *kroví* × 1 'blood (GEN.SG)'; *núždý* × 3 ~ *nuždý* × 1 'need (GEN.SG)'; *prízrak* × 9 ~ *prizrák* × 2 'ghost (NOM/ACC.SG)', *rodílsja* × 6 ~ *rodílsjá* × 1 'was born (MASC.SG)', *skórbej* × 1 ~ *skorběj* × 1 'griefs (GEN.PL)', *ščástliv* × 3 ~ *ščástlív* × 2 'happy (MASC.SG.SHORT)', *ščástlívye* × 1 ~ *ščástlívye* × 1 'happy (NOM/ACC.PL)'. However, the figures in the rightmost column are not directly comparable with counts for other poet because of the differences in corpus size for individual poets. A simple proportion of word form types with variable stress is not meaningful for the same reason.

To solve this problem, we propose a more sophisticated metric capturing the amount of variability in stress placement for a given speaker. We randomly sample 10,000 tokens (these tokens need not be adjacent) from the corpus of a poet and count the number of word form types with stress variation in this sample. This procedure is repeated 1000 times, and the mean value is taken to be an indicator of the amount of variation in the corpus of a poet.

We demonstrate that the amount of intra-speaker variability tends to reduce in the 20th century compared to the 19th century (as shown in the Figure below, where the average number of word form types with variable stress per 10,000 tokens is plotted against year of birth). This might be due to the stabilization of the prescriptive standard over the course of the past two centuries. Furthermore, we argue that variability within a sound pattern can be interpreted as a derivative of its stability in a language: the more variability there is, the less stable the pattern is (cf. the discussion in D'Arcy 2015 and references therein).

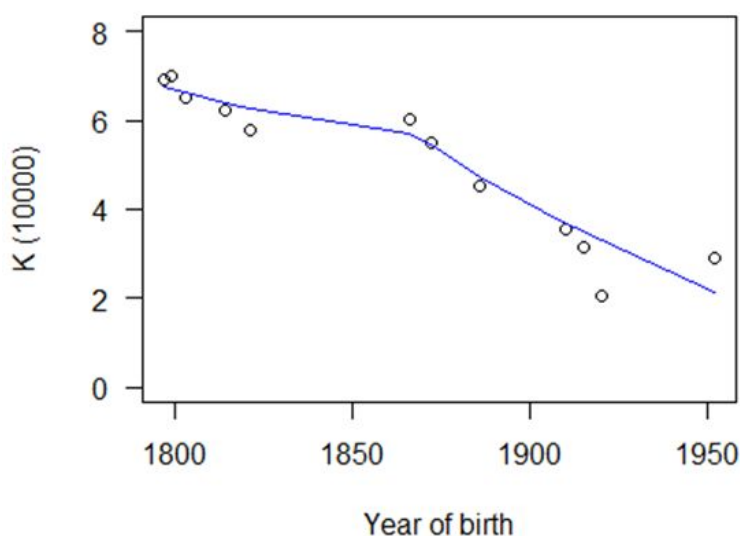


Figure. Average number of word types with variable stress per 10,000 tokens for 12 Russian poets.

References

- D'Arcy, Alexandra. 2015. Variation, transmission, incrementation. In Patrick Honeybone and Joseph Salmons (eds.), *The Oxford handbook of historical phonology*. Oxford: OUP.
- Halle, Morris. 1973. The accentuation of Russian words. *Language* 49(2), 312–348.
- Melvold, Janis. 1990. *Structure and stress in the phonology of Russian*. MIT dissertation.
- Zalizniak, Andréi. 1985. *Ot praslavjanskoj akcentuacii k russkoj* [From Proto-Slavic to Russian accentuation]. Moscow: Nauka.