
Determining the relationship between uninflectedness, overabundance and defectiveness

Dunstan Brown¹, Harald Baayen², Neil Bermel³, Yu-Ying Chuang², Roger Evans¹ & Alex Nikolaev⁴

¹University of York, ²Universität Tübingen, ³University of Sheffield,

⁴University of Eastern Finland

dunstan.brown@york.ac.uk, Harald.Baayen@protonmail.com,

n.bermel@sheffield.ac.uk, yu-ying.chuang@uni-tuebingen.de,

roger.evans@nltg.org.uk, alexandre.nikolaev@helsinki.fi

We aim to understand whether differences in the three phenomena are reflected in their distributional behaviour, using a distributional semantics approach (see, e.g., Firth, 1968; Landauer and Dumais, 1997; Mikolov et al., 2013). We look at phenomena in Czech, Polish, Russian, Finnish.

Czech: The Czech noun *rande* ‘date’ is uninflected. However, it (and others) can adopt t-stem inflections (e.g. like *kotě* ‘kitten’). This leads to overabundance, but it appears that neither the uninflected form *rande* nor the t-stem form *randat* is entirely acceptable in the gen. pl.

Polish: Uninflectedness may be partial, as with *muzeum* ‘museum’ (Kotyczka 1980: 95, 105-6; Tokarski 1993: 257).

Russian: defectiveness (gen. pl. of nouns) may be associated with peculiar distributional semantics. The relevant class uses the stem as exponent of the gen. pl (Chuang et al 2022).

Finnish: pronoun forms such as *tuolla* (abl. sg.) may be used as uninflected adverbial forms.

Defectiveness in the Czech example suggests that emergence of paradigmatic contrasts may change the status of a bare stem from being considered merely uninflected. However, emergence of paradigmatic contrasts seems to be unproblematic in the case of the partially uninflected Polish example. The Russian data suggests that defectiveness may involve an issue with associating distributional features with bare stems, while the Finnish data suggests that distributional divergence that does not impinge on certainty around paradigm structure of the same lexeme may be less of a problem. An understanding of the distributional properties of the features associated with these phenomena will provide a clearer picture beyond the basic assumption of featural inertness usually associated with uninflectedness.

References: • Chuang, Y., Brown, D., Baayen, R. H., & Evans, R. Paradigm gaps are associated with weird "distributional semantics" properties: Russian defective nouns and their case and number paradigm. <https://doi.org/10.31234/osf.io/t7xba> • Firth, J. R. (1968). *Selected papers of J. R. Firth, 1952–59*. Indiana University Press. • Kotyczka, J. (1980). *Kurze polnische Sprachlehre*. Volk und Wissen. • Landauer, T. and Dumais, S. (1997). A solution to Plato’s problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2), 211–240. • Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–9. • Tokarski, J. (1993). *Schematyczny indeks a tergo polskich form wyrazowych*. Opracowanie i redakcja: Zygmunt Saloni. Wydawnictwo Naukowe PWN.