

Processing the inflectional complexity of Russian nouns: features, information and frequency

Understanding how inflectional complexity affects lexical processing is challenging because similar questions have been addressed with different theoretical assumptions (e.g. inferential vs. lexical theories) and with different ways of quantifying complexity (e.g. primitive features vs. information content). In this paper I contrast two analyses of the complexity of Russian nouns and show that despite making distinct predictions for some inflectional suffixes, both accounts are similarly significant predictors of response times in a visual lexical decision task. I discuss implications for morphological theories and future experimental work.

Many lexical-incremental theories posit that primitive features are part of the lexical representation of inflectional affixes. Clahsen et al. (2001) use experimental data from German to argue that lexical decision times are correlated with the number of specified features of inflectional suffixes. The more features an affix contains, the longer it takes to process. In support of a lexical(-incremental) theory of morphology, they argue that features are an indispensable part of lexical representation and inflectional systems more generally. However, the idea that features are necessary to predict the processing cost of inflectional affixes presupposes that no other measure of inflectional complexity could account for the same results. Independent research, informed by inferential-(realizational) theories of morphology, suggests that lexical processing times reflect the amount of information carried by a word form (Kostić 1991; Milin et al. 2009). In these studies it is the paradigmatic relationships between word forms that determine their information content. Word forms that carry more information take longer to process, making an information-theoretic approach a competing view of how to understand the processing cost of inflectional complexity.

I make a direct comparison between these two approaches, using Russian nouns as a test case. I take an existing feature-based account of Russian nouns (Müller 2004) and operationalize Shannon entropy (Shannon 1948) to develop an information-based account using the same parameters, e.g. number of classes, patterns of syncretism, etc. For the information-based account, I weight suffixes by the frequency of word types in each inflection class (extracted from *Grammatičeskij slovar' russkogo jazyka* (Zaliznjak 1977)) and the token frequency of each morphosyntactic property set (taken from the disambiguated sub-section of the Russian National Corpus (www.ruscorpora.ru)). I developed a lexical decision task that focuses on a group of 11 suffixes that vary in the case(s)/number(s) they realize, the inflection class(es) they belong to, and the number of features and information they exhibit.

Despite the fact that each account makes distinct predictions for various suffixes, both features and information-content are similarly statistically significant predictors of reaction times in the visual lexical decision task (in addition to expected predictors, e.g. form frequency, word length, etc.). Interestingly, the frequency of the affixes also predicts reaction times well, leaving it unclear whether features or information are necessary to account for the behavioral data. These results suggest that experimental validation of theoretical constructs can be problematic and brings up questions about the processing and representation of inflectional structure.