Effects of relative frequency on morphological processing in Russian and English

Michelle McKenzie, Andrea D. Sims
The Ohio State University

Since Taft (1979), evidence has suggested that lexical access is sensitive to the distributional properties of a speaker’s lexicon, with words differing in the likelihood that their meanings will be accessed directly via whole word lexical entries (ex. government) vs. via morphological constituents (‘parsing’; ex. govern, -ment). This raises questions about whether there are cross-linguistic differences in lexical access pathways (Frost & Grainger 2000; Marslen-Wilson 2001). Languages differ in the extent to which they make use of morphology, resulting in differences in their lexical-distributional properties. Does this lead some languages to rely more overall on whole word access and others to rely more on parsing? In this paper we test a prediction from Sims and Parker (2015), who compare the distributional properties of Russian and English words with derivational suffixes. They show that Russian derived words, more so than English ones, have properties (growth rate of the vocabulary, productivity, relative frequency) associated with parsing. We test their prediction that Russian derived words are more likely than English ones to be parsed during lexical access by comparing the results of semantic transparency rating experiments conducted in each language.

We focus on word frequency, as a distributional property of the lexicon and independent variable that is known to affect lexical access. Most classes of processing models posit competition between direct storage of words and extrapolated morphological patterns, with the competition mediated in part by word frequency. For convenience we assume a parallel dual-route model (Baayen et al. 1997) in which both whole word and parsed affixal lexical entries are posited. In this class of models, a derived word’s meaning is accessed via whichever type of entry is faster to access. As the token frequency of surface derived words (derived frequency) goes up, the likelihood of whole word access goes up; as the token frequency of derived words’ base forms (base frequency) goes up, the likelihood of parsing goes up.

We conducted parallel experiments in Russian and English that were designed to test the effect of derived frequency and base frequency as predictors of semantic transparency. Semantic transparency is a correlate to parsing (Hay 2001; Schirmeier et al. 2004). In each experiment, participants rated the similarity in meaning of word pairs in their native language using a continuous visual analog scale. Each stimulus consisted of a complex word (ex. government) and its base (ex. govern) in a question frame. We created 120 target stimuli for each experiment/language, representing 20 conditions (10 affixes x 2 values of relative frequency) and 6 items per condition. Relative frequency is derived frequency divided by base frequency (Hay 2001). Half of the stimuli had relative frequency values greater than 1; half had relative frequency values less than 1. To the extent possible we controlled for morphophonological alternations, absolute base and derived frequency, and other factors. In each experiment, stimuli were divided into two lists, each with 3 target items per condition plus 20 control items. So far we have collected data from 24 native English speakers and 5 native Russian speakers. Russian data collection is ongoing.

Stimuli in the two experiments have the same frequency properties in aggregate. This allows us to examine whether frequency has a different qualitative or quantitative effect on the semantic transparency (and by inference, parsing) in the two languages. If we find a different effect of frequency on semantic transparency across the experiments, this would be consistent with the idea that any difference in rate of parsing is not an incidental by-product of the languages having different distributional properties, and would instead suggest a more nuanced interaction between word frequency and lexical access pathways.

As expected, results show a negative correlation between relative frequency and mean semantic transparency ratings by item in English (p < 0.05, based on simple linear regression of z-transformed data), with Russian results trending in the same direction (p = 0.077). At the same time, some interesting differences seem to be emerging in the experiment results (subject to verification once the full set of Russian responses has been collected). First, semantic transparency ratings are higher overall in Russian than in English (means 775.7 vs. 697.8 on a scale to 1000; t = 5.13, p < 0.001), even though the Russian
stimuli are disproportionately of high relative frequency due to a list rotation problem that we are fixing. This is surprising since high relative frequency lowers ratings, all else being equal. If this generalization holds, it is consistent with the idea of greater semantic transparency (→ more parsing) in Russian.

Second, and most interestingly, derived frequency seems to have a larger magnitude effect on semantic transparency ratings in English than in Russian (Figure 1). As log derived frequency increases, semantic transparency ratings decrease exponentially in English (p < 0.001, R² = 0.128 in polynomial regression). Low semantic transparency equates to more whole-word access. By contrast, in the current Russian results derived frequency is not predictive of semantic transparency ratings (p = 0.767).

This difference, if it holds up, would be consistent with the finding of Sims and Parker (2015) that parsing rates are predicted to drop more quickly in English than in Russian as derived frequency increases, and that Russian derivational morphology is overall more biased towards parsing than English is.

The experiment results suggest that lexical access route is more sensitive to word frequency in English than Russian – even when the frequency properties of words are held constant. This points to cross-linguistic differences in lexical access. Specifically, it suggests that the strength of influence of word frequency is not constant cross-linguistically but rather, is determined in a dynamic way by speakers’ experiences with morphological structures.

References


