Abstract:

The grammatical profile method was introduced by Janda & Lyashevskaya (2011) as a subtype of behavioral profiles (e.g. Gries & Divjak 2009). A grammatical profile is defined as the relative frequency of all inflected forms of a lemma in a corpus. This method provides a means to study the relation between the distribution of grammatical properties of words and meaning on a relatively large scale. Janda (2017) shows in a pilot analysis that grammatical profiles of Czech nouns is correlated with animacy. I follow up on this finding and focus on the role of both grammatical and natural gender. In this paper, I present the results of two analyses of the relationship between cases and gender.

In the first analysis, I show the difference between masculine and feminine animate nouns. Czech nouns have four genders, masculine animate and inanimate, feminine, and neuter. Only masculine animate correlates with natural gender. I manually extracted all feminine animate nouns that appear at least 100 times in a corpus of modern written Czech (Křen & al. 2015). These were complemented by their masculine counterparts (e.g. herečka ‘actress’ - herec ‘actor’). I extracted all singular forms of those lemmas and obtained grammatical profiles for 2521 lemmas. I performed hierarchical clustering analysis to identify lemmas with similar profiles. There are two clear findings. First, there is a difference between proper and generic nouns regardless of gender. Second, there is a difference between masculine and feminine generic nouns with masculines appearing more frequently in the cluster characterized with a high proportion of nominatives and relatively low proportion of accusatives.

In the second analysis, I studied the relationship between the grammatical profiles of adjective modifiers and the gender of their nominal heads. I analyzed a total of 747 modifiers that appear at least 100 times with masculine animate, inanimate, and feminine heads in the corpus. I identified four clusters in the data which show that modifiers pattern based both on head gender and meaning. Modifiers with masculine animate heads are again more agentive, while inanimate masculines show relatively high proportions of accusative, locative, and instrumental. Feminine headed modifiers fall between the two categories.

The results show that it is indeed possible to “go from form to meaning via statistics” (Kuznetsova 2015) and that the distribution of relatively abstract grammatical categories can point at meaningful differences in use and meaning.
References:


