Modeling Aspectual Variation in Slavic: A Grammaticometric Approach

Beata Trawiński
(IDS Mannheim)

In this talk, I will argue that there is a correlation between semantic and purely grammatical patterning in language, yet the nature of this correlation has to be explored. This assumption is based on the results of my corpus-driven study of the Slavic aspect, drawing on the so-called Distributional Hypothesis.

According to the East-West Theory of the Slavic aspect, there is a broad east-west isogloss dividing the Slavic languages into an eastern group and a western group. There are also two transitional zones in the north and south, which share some properties with each group (Dickey 2000; Barentsen 1998, 2008). The East-West Theory uses concepts of cognitive grammar such as totality and temporal definiteness, and is based on various parameters of aspectual usage in discourse, including contexts such as habitu als, general factuals, historical (narrative) present, performatives, sequenced events in the past etc.

The purpose of my study was to challenge the semantic approach to the Slavic aspect by comparing the perfective and imperfective verbal aspect on the basis of purely grammatical co-occurrence patterns (see also Janda and Lyashevskaya 2011). The study focused on three Slavic languages: Russian, which, following the East-West Theory, belongs to the eastern group, Czech, which belongs to the western group, and Polish, which is considered as transitional in its aspectual patterning.

Drawing on the so-called Distributional Hypothesis, stating that linguistic entities (conventionally words) that occur in the same contexts tend to have similar meanings / functions (Wittgenstein 1953, Firth 1957, Harris 1954) and using corpora of Russian, Polish and Czech (here, the original texts from the parallel corpus InterCorp, Čermák und Rosen 2012, were used as a data source), I determined the distribution (in terms of relative frequencies) of the perfective and imperfective forms in the respective language with the following grammatical contexts: singular and plural, first, second and third person, feminine, masculine and neuter, past and non-past, imperative and infinitive. This distributional information was collected for each language in a context vector. I compared the perfective and imperfective forms across Russian, Polish and Czech by comparing the respective context vectors in a single multidimensional vector space. The similarities or distances between the vectors were calculated and evaluated by single numbers computed by using different similarity or distance measures (such as cosine or Euclidian distance). Based on these metrics, hierarchical cluster analyses were performed in order to cluster the perfective and imperfective aspects in Russian, Polish and Czech according to the degrees of (dis)similarity and to identify systematic relationships between the co-occurrence contexts.

The results of this study perfectly correlate with the semantically informed East-West Theory: The biggest distance (lowest similarity) was found between Russian and Czech; the distances between Russian and Polish and Czech and Polish were respectively smaller than that between Russian and Czech.

If these findings can be confirmed in further studies including more data and other (Slavic) languages, they might have interesting implications for cross-linguistic research in general, in particular, when multiple languages are compared. Namely, grammatical profiles generated on the basis of corpus data and the comparison of these profiles using methods of quantitative analysis could serve as a point of departure for qualitative semantic analysis, offering a first orientation of how the languages being compared are clustered and, possibly, providing indications of what features particularly influence the clustering.
References


