

A Grammaticometric Approach to Modeling Functional Variation

Beata Trawiński

IDS Mannheim

trawinski@ids-mannheim.de

1 Introduction

This paper argues that there is a correlation between functional and purely grammatical patterning in language, yet the nature of this correlation has to be explored. This claim is based on the results of a 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 habituais, general factuais, historical (narrative) present, performatives, sequenced events in the past etc.

The purpose of the above-mentioned study is 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 & 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.

2 The corpus study

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 & Rosen 2012, were used as a data source), we determined the distribution (in terms of relative frequencies) of the perfective and imperfective forms in the respective language in 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. We 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 / distances between the vectors were calculated and evaluated by single numbers computed by using different similarity / 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. Moreover, we determined that the Polish perfective aspect patterns closer to the Russian perfective aspect (Figure 1), while the Polish imperfective aspect patterns closer to the Czech one (Figure 2).

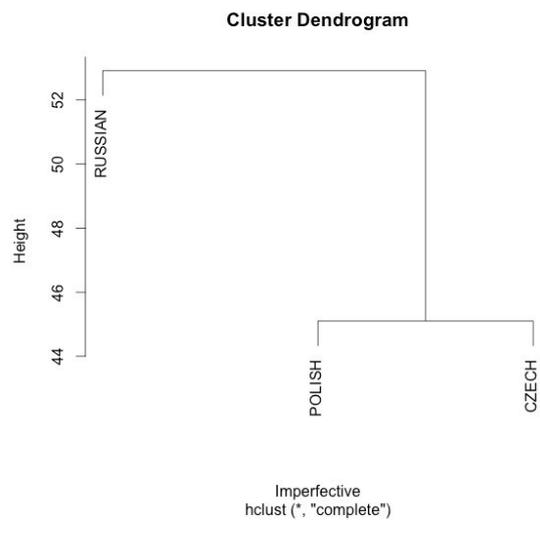
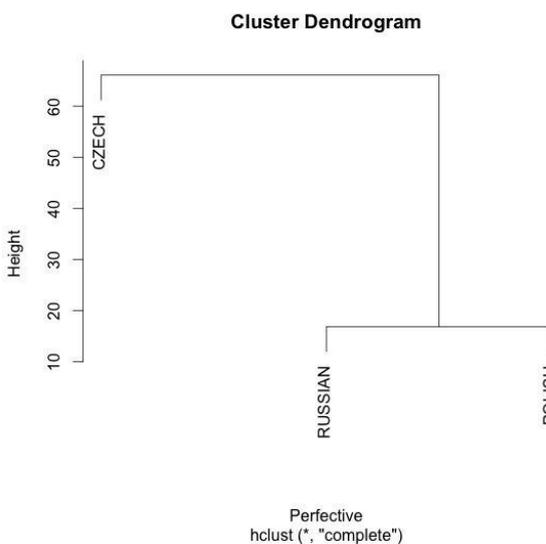


Figure 1: Hierarchical clustering of perfectives

Figure 2: Hierarchical clustering of imperfectives

The analysis of the co-occurrence contexts has shown no significant differences in the distribution of the perfective aspect over all the examined grammatical contexts except for the category of person. More precisely, the frequency of the third person marking was substantially greater than expected in Czech, while it was lower than expected in Polish and Russian. This fact may correlate with the differences in specific aspectual usage in discourse, in particular with contexts such as instructions or performatives (cf. Dickey 2000).

The distribution of the imperfective aspect is visualized in Figure 3. As this association plot shows, there are significant differences in the distribution of the imperfective aspect with third person, past and non-past. Polish imperfectives occur significantly more frequently with past (than with other contexts), while Russian imperfectives occur significantly more frequently with non-past (than with other contexts). The distribution of the imperfective aspect with non-past in Polish is substantially lower.

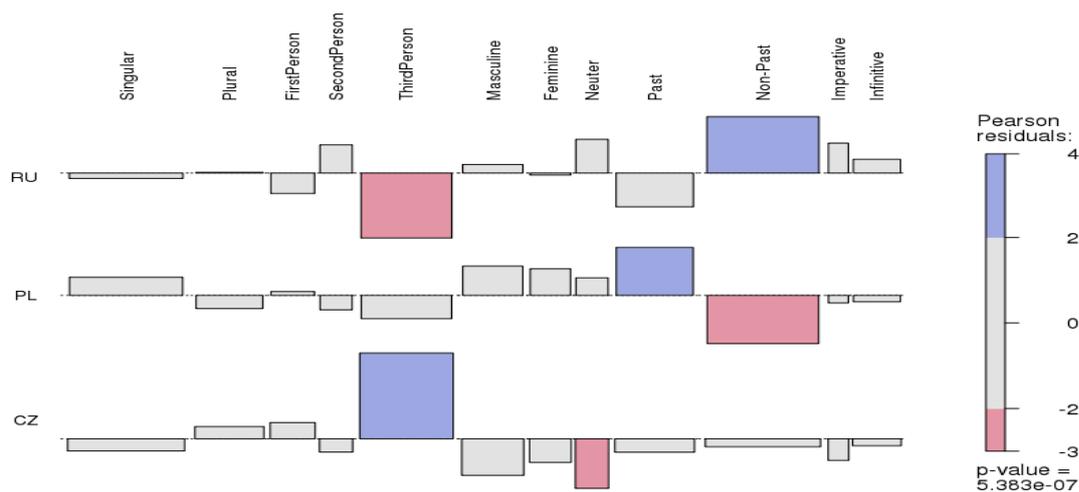


Figure 3: Association plot for Russian, Polish, Czech and the grammatical contexts of the imperfective aspect

3 Conclusion

Our study may have interesting implications for cross-linguistic research, particularly when multiple languages are compared. Grammatical profiles generated on the basis of corpus data and the comparison of these profiles using methods of quantitative analysis are a promising point of departure for qualitative semantic / functional analysis. In further research, broadened to encompass more data and more Slavic languages, we intend to further clarify the pattern(s) of clustering and to determine which features influence the final result.

References

- Barentsen, A. (1998) Признак «секвентная связь» и видовое противопоставление в русском языке. In М.Ю. Черткова, ed., *Типология вида: проблемы, поиски, решения*, 43-58. Языки русской культуры, Москва.
- Barentsen, A. (2008) Выражение последовательности действий при повтораемости в прошлом в современных славянских языках. In *Dutch Contributions to the Fourteenth International Congress of Slavists, Ohrid: Linguistics*, 1-36. Rodopi, Amsterdam / New York.
- Čermák, F. & A. Rosen (2012) The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 17(3): 411-427.
- Dickey, S. M. (2000) *Parameters of Slavic Aspect: A Cognitive Approach*. CSLI Publications, Stanford.
- Firth, J. R. (1957) *Papers in Linguistics 1934-1951*. Oxford University Press, London.
- Harris, Z. (1954) Distributional structure. *Word*, 10(2-3): 146-162.
- Janda, L. & O. Lyashevskaya (2011) Grammatical profiles and the interaction of the lexicon with aspect, tense and mood in Russian. *Cognitive Linguistics*, 22(4): 719-763.
- Wittgenstein, L. (1953) *Philosophische Untersuchungen*. Werkausgabe. Bd. 1, Suhrkamp.