

Tania Avgustinova, Andrea Fischer, Klara Jagrova, Dietrich Klakow, Roland Marti & Irina Stenger
Saarland University
avgustinova@coli.uni-saarland.de; afischer@lsv.uni-saarland.de; kjagrova@coli.uni-saarland.de;
dietrich.klakow@lsv.uni-saarland.de; rwmslav@mx.uni-saarland.de; istenger@coli.uni-saarland.de

The empirical basis of Slavic intercomprehension

The possibility of intercomprehension between related languages is a generally accepted fact suggesting that mutual intelligibility is systematic. Of particular interest are the Slavic languages, which are “sufficiently similar and sufficiently different to provide an attractive research laboratory” (Corbett 1998). They exhibit practically all typologically attested means of encoding grammatical information, ranging from extremely dense to highly redundant constructions, and their development is the result of various language contact scenarios (Balkansprachbund, German influence on West Slavic languages, Finno-Ugric substratum in East Slavic languages etc.).

Similarities observable at all levels of linguistic description and grammatical differences, which are well-studied in Slavic linguistics, allow us to design tests with calibrated degrees of deviation. As the number of pairs among the Slavic languages is large, we select a few representative language pairs where a certain threshold level of mutual intelligibility can be expected. Czech and Polish (both West-Slavic) appear to be very close languages. Russian (East-Slavic) is generally perceived to be closer to Polish than to Czech, while each of them is quite distant from Bulgarian (South-Slavic).

The research questions we address in this contribution are: What linguistic distances can be established in the respective language pair at different linguistic levels (orthography, vocabulary, morphology, syntax)? To what extent are the linguistic distances predictors of mutual intelligibility? What explanations can be found for asymmetric intelligibility? The rules of orthographic correlates are worked out and sound correspondences are established, taking into consideration historically conditioned cross-linguistic variation. The cognate vocabulary is approximated on the basis of available standardised word lists for the selected language pairs. To avoid negative cognates influencing intercomprehension in unpredictable ways, stop-lists of known “false friends” are compiled to be used in the processing of parallel corpora. To account for the intelligibility of morphological forms and syntactic constructions, the respective correspondences in the grammatical subsystems are formulated. Both the nature of the phenomena and the effect of frequency are relevant at these levels.

References

- Corbett, G. 1998. Agreement in Slavic (position paper). *Workshop on Comparative Slavic Morphosyntax*. Indiana University.