

## The linguistic complexity of South Slavic dialects: A quantitative study

**Goal of the study.** The paper presents the results of a quantitative study of South Slavic dialects, based on their linguistic complexity. We make an attempt to measure the level of complexity of different varieties, subdialects, and dialects and analyse these data using various methods of quantitative analysis. Our goal is to examine the complexity variance throughout the South Slavic territory and to explain if and how complexity levels correlate with the (non)involvement of the different dialects in past and present language contact situations.

**Methodology.** In our study we measure what Johanna Nichols (2009: 111) calls *grammatical complexity* — “complexity of the strictly linguistic domains of phonology, morphosyntax, lexicon, etc. and their components” and follow “the usual understanding of a complex system as one consisting on many different elements each with a number of degrees of freedom”. We survey phonological and morphological traits (elements of the systems), and a range of variant forms of inflectional morphemes (paradigmatic variants, or degrees of freedom). All of them are represented as binary features (presence or absence of a certain trait) rather than mere numbers of vowels, consonant phonemes, cases, tenses, and allomorphs. In our selection of features and assessment of dialects, the Proto-Slavic (Schenker 1993) served as a historical reference system, as in the traditional Slavic dialectology, which refers to Proto-Slavic in order to “reconstruct the development of dialects and to interpret them as the result of language change” (Girnth 2010: 116).

**Data description.** The overall sample consists of 735 observations, each of them representing a single Balkan Slavic variety spoken in a certain locality (village or town). The data were taken from various dialectological monographs and dialect text collections, and from major dialectological atlases (see data sources below). The data comprises 14 macro-dialect groups and 62 dialects. The Serbo-Croatian dialect continuum is represented by those dialects which are assumed to have developed in a contact situation with non-Slavic Balkan languages and which are characterized by several innovative Balkan linguistic features: Torlak (eastern Kosovo and eastern Serbia close to the Bulgarian border, western Bulgaria), Kosovo-Resava (central and western Kosovo) and Zeta-Lovčën (Montenegro, south-western Serbia, close to the boarder of Montenegro) macro-dialects. Balkanized South Slavic is represented by its major macro-dialect groups: northern, western, southern and eastern dialects of Macedonian, northern, western, southern and eastern dialects of Bulgarian, Aegean Slavic (Greece).

For measuring complexity scores and distance between the varieties, 29 binary features from the linguistic domains of phonology and morphology were selected. The features can be classified as complexifications (often contact-induced, e. g. presence of postpositive article) or simplifications (e. g. absence of gender differentiation in 3PL personal pronoun). The complexity score for each variety was calculated as a simple sum of the true-values.

**Preliminary results.** The complexity of all considered varieties ranges from 6 to 15 (of maximal 29), the median value is 9. So the South Slavic dialects demonstrate the tendency to simplification rather than to complexification.

There are two major areas with relatively high total complexity scores, which belong to genetically different branches of South Slavic. The first area — with the north-westernmost part of North Macedonia, Montenegro, South Serbia, and Kosovo — includes the varieties of the Serbo-Croatian dialect continuum. The second area is a narrow belt of peripheral western Macedonian varieties extending along the Albanian-Macedonian border. The most complex varieties (total complexity score >12) are those adjacent to the border.

This correlation between the high complexity score and the border/contact areas is interesting from the point of view of both Balkan and general contact linguistics, because the contact zones are usually associated with simplification processes (see, e. g., Rusakov & Morozova 2017 on the complexity of the Albanian dialects). The situation observed in Balkan Slavic may be explained in the different but, maybe, complementary ways. On the one hand, the reason for such a result may lie in the very features chosen for our study. The great part of

complexifying features we have in our survey are contact-related and it is quite natural that they accumulate in the areas where most intensive language contacts have been taking place in the past and present. On the other hand, the different (socio)linguistic types of language contact situations may be instructive here. For example, the mountainous areas on the Albanian-Montenegrin border are characterized by strong isolation and long coexistence of the Albanian and Slavic speakers. These factors may have promoted the development of contact induced complexification, as described by Peter Trudgill (2011). Another situation is characteristic for the western part of North Macedonia and the Western Kosovo (maybe for the Timok valley as well), where the (grammatical) complexification might be spread via language shift from Arumanian and/or Albanian to Slavic.

### **References**

- Nichols, Johanna. 2009. Linguistic complexity: a comprehensive definition and survey. In Sampson, Geoffrey, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 110–125, Oxford.
- Girnth, Heiko. 2010. Mapping language data. In Lameli, Alfred, Roland Kehrein & Stefan Rabanus (eds.), *Language and Space. An International Handbook of Linguistic Variation. Vol. 2. Language Mapping. Part I*, 98–120, De Gruyter Mouton.
- Rusakov, Alexander & Maria Morozova. 2017. Linguistic complexity: What do Albanian dialects show? Paper presented at the Workshop on Balkan linguistics “Macro and micro variables across the Balkans”, 26 October 2017, University of Helsinki, Finland.
- Schenker, Alexander M. 1993. Proto-Slavonic. In Comrie, Bernard & Greville G. Corbett (eds.), *The Slavonic languages*, 60–121, Routledge.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*, Oxford University Press.

### **Data Sources**

- Balgarski dialekten atlas [Atlas of Bulgarian dialects]. 1964–1981. Vol. I–IV. Sofia: BAN.
- Kochev, Ivan (ed.) et al. 2001. *Balgarski dialekten atlas: Obobshchavashch tom. I–III: Fonetika. Aktsentologija. Leksika* [Atlas of Bulgarian dialects: Generalizing volume: I–III: Phonetics. Accentology. Lexicology], Sofia: Trud.
- Srpski dijalektološki zbornik. Rasprave i građa. Vol. XIII–LXIII. 1933–2016.
- Tetovska-Troeva, Margarita (ed.) et al. 2016. *Balgarski dialekten atlas: Obobshchavashch tom. IV: Morfologija* [Atlas of Bulgarian dialects: Generalizing volume. IV: Morphology]. Sofia: Izdatelstvo na BAN “Prof. Marin Drinov”.
- Vidoeski, Bozhidar. 2000. *Tekstovi od dijalektite na makedonskiot jazik*. Skopje: Institut za makedonski jazik “Krstev Misirkov”.