

## Areal Distribution of the Post-positive Article in Timok Dialect of Torlak\*

Teodora Vuković

Slavisches Seminar/Language and Space Lab, University of Zurich  
teodora.vukovic2@uzh.ch

Tanja Samardžić

Language and Space Lab, University of Zurich  
tanja.samardzic@uzh.ch

The paper presents an ongoing research on the areal distribution of the post-positive article in Timok dialect in South-eastern Serbia. The morpheme in question is an unstressed particle derived from demonstrative pronouns that is used as definiteness marker, while still keeping its demonstrative meaning. In Timok dialect, the eastern part of Prizren-timok dialect of Serbian, it forms part of the noun phrase and follows either a noun (e.g. *bratat* – ‘**the** brother’, *ženava* – ‘**the** woman’, *deteno* – ‘**the** child’) or a noun modifier (e.g. *moav starac* – ‘my (**the**) old man’). (Belić, 1905; Tomic, 2006; Ivić, 1985; Sobolev, 1998).

Post-positive article is one of the most salient features of Timok dialect that differentiates it from standard Serbian and other Serbian dialects. Contrary to Serbian, article is used in standard Bulgarian and Bulgarian dialects to the east, as well as Macedonian in the south. The frequency of the use of the article in Timok varies in the territory of the whole dialect, there are big differences between villages and speakers. This study analyses in which parts of the Timok area the article is used more frequently and the geographical factors behind that.

The sample of 30.000 tokens is selected from the corpus of transcripts of ethnolinguistic fieldwork interviews which were collected between 2015 and 2017 by the team of researchers from the Institute for Balkan Studies of Serbian Academy of Sciences and Arts lead by Biljana Sikimić. Texts were annotated for PoS and lemmatized using ReLDI tagger (Ljubešić et al., 2016). Tags for words containing post-positive articles were adapted manually, since they were not included in the original tagger.

The study aims at identifying the influence of surrounding languages on the use of the post-positive article. It analyses the effect of areal and geographic factors such as the proximity of Bulgarian border, longitude and latitude, altitude and the proximity of the administrative center of the area, the city of Knjaževac. The analysis applies standard statistical correlation tests. The contact with these languages is expressed with measures of spatial distance and altitude, while the use of article is presented using normalized frequency counts. Data was extracted using a Python script and GIS software.

Figure 1 shows the frequency of the post-positive article in the area. The size and the intensity of the black circles indicate the frequency of article use in each given location.

---

\* Vukovic and Samardžic (2018).

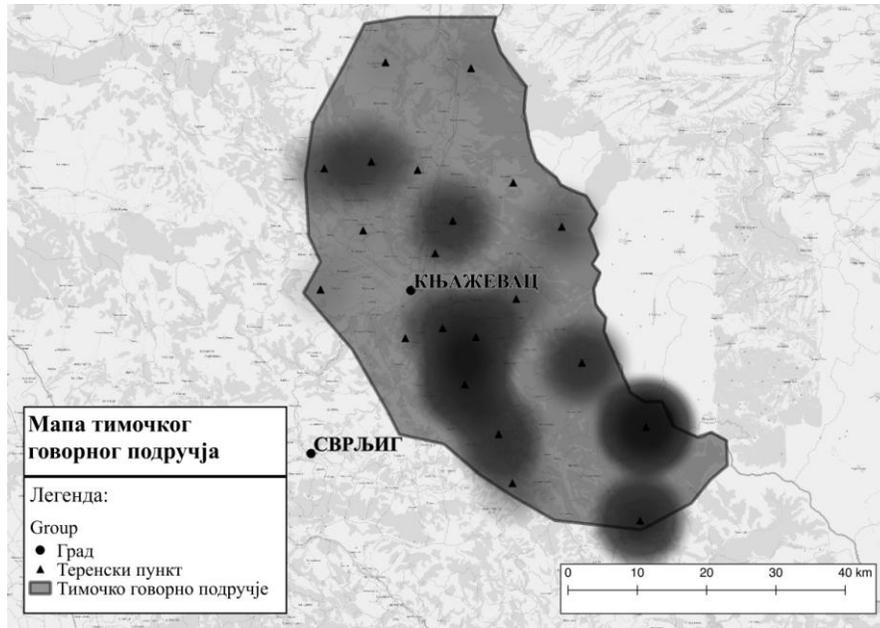


Figure 1: Areal distribution of the article in Timok dialect

The results reveal a strong effect of altitude as an indicator of remoteness, but it is not determines whether there is an influence of Bulgarian. This implies that the use of post-positive article is more affected by a cultural contact with socially more prestigious standard Serbian, than with typologically closer Bulgarian and Macedonian.

Further research on Timok and neighboring dialects is being continued within the Horizon 2020 project TraCeBa, a collaboration between Slavisches Seminar at the University of Zurich, Institute of Linguistic Studies of the Russian Academy of Sciences and the Institute for Balkan Studies of Serbian Academy for Sciences and Arts. The project has the goal of producing an elaborate large-scale corpus of Torlak dialect containing synchronic and diachronic, spoken and written language samples, that will be made publicly available on-line. In the future, the team intends to employ methods of dialectometry and use more complex statistical models to draw a more detailed picture of variation in Torlak, and distributional semantics and sentiment analysis to study cultural phenomena over time and space.

#### References

- Aleksandar Belić. *Dijalekti istočne i južne Srbije*. Srpski dijalektološki zbornik. Sprska Kraljevska Akademija, 1905.
- Pavle Ivić. *Dijalektologija srpskohrvatskog jezika: Uvod i štokavsko narečje*. Matica srpska, Novi Sad, 1985.
- Nikola Ljubešić, Filip Klubička, Željko Agić, and Ivo-Pavao Jazbec. New inflectional lexicons and training corpora for improved morphosyntactic annotation of croatian and serbian. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, 2016. European Language Resources Association (ELRA).
- A.N. Sobolev. *Sprachatlas Ostserbiens und Westbulgariens*. Scripta Slavica. Bibliion, 1998.
- Olga Mišeska-Tomić. *Balkan Sprachbund Morpho-Syntactic Features*. Studies in Natural Language and Linguistic Theory. Springer Netherlands, 2006.
- Teodora Vuković and Tanja Samardžić. Prostorna raspodela frekvencije postpozitivnog člana u timočkom govoru. In Ćirković, Svetlana (main) and Andrej N. Sobolev, Barbara Sonnenhauser, Maja Miličević, Jelenka Pandurević, editor, *Timok. Terenska istraživanja 2015–2017*. Narodna Biblioteka “Njegoš”, 2018.