

## INFLECTIONAL DESCRIPTION OF BULGARIAN MULTIWORD EXPRESSIONS: AN ANALYSIS OF TWO-COMPONENT GEOGRAPHICAL NAMES

SVETLOZARA LESEVA

INSTITUTE FOR BULGARIAN LANGUAGE, BULGARIAN ACADEMY OF SCIENCES  
zarka@dcl.bas.bg

This article explores the grammatical and semantic characteristics of two-component compound proper names in Bulgarian, with a particular focus on geographical names, aiming to define their inflectional types and integrate them into the respective dictionary entries. The study is based on lexical units automatically extracted from internet sources, including structured resources such as Wikipedia. These units were manually filtered and categorised according to the type of referent, e.g. *town*, *village*, *mountain*, *spring*, etc. As a result, the initial dataset of 4,000 candidates was reduced to approximately 2,600 entries.

The selected names were automatically POS-tagged and analysed in terms of their syntactic structure, lexical and grammatical categories, inflectional properties, defectivity, etc. These features collectively determine the inflectional type of each named entity.

The inflectional types enable both the generation of inflected forms and their recognition in free text.

*Keywords:* inflection types, geographical names, multiword expressions, Bulgarian

### REFERENCES

- Baziotis et al. 2023: *Baziotis, C., P. Mathur, E. Hasler*. Automatic evaluation and analysis of idioms in neural machine translation. – In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp 3682–3700, Dubrovnik, Croatia. Association for Computational Linguistics.
- Courtois, Silberztein 1990: *Courtois, B., M. Silberztein*. Dictionnaires 'electroniques du francais. – *Langue française*, 87 (1), pp. 11–22.
- Dimakis et al. 2024: *Dimakis, A. S. Markantonatou, A. Anastasopoulos*. Dictionary-aided translation for handling multi-word expressions in low-resource languages. – In: *Findings of the Association for Computational Linguistics (ACL) 2024*. Bangkok, Thailand and virtual meeting, Association for Computational Linguistics, pp. pp. 2588–2595.
- Gross 1997: *Gross, M.* The construction of local grammars. – In: Roche, E., Y. Schabs (eds.). *Finite State Language Processing*, pp. 329–354.
- Koeva 1998: *Koeva, S.* Gramatichen rechnik na balgarskiya ezik. Opisanie na

- kontseptsiyata za organizatsiyata na lingvistichnite dannii. – *Balgarski ezik*, N 6, s. 49–58.
- Koeva, Stoyanova, Todorova, Leseva 2016: Koeva, S., I. Stoyanova, M. Todorova, S. Leseva. Semi-automatic compilation of the Dictionary of Bulgarian Multiword Expressions. – In: *Proceedings of the GLOBALEX 2016 Workshop: Lexicographic Resources for Human Language Technology*, LREC, pp. 86–95.
- Sag et al. 2002: Sag, I. A., T. Baldwin, F. Bond, A. A. Copestake, D. Flickinger. Multiword expressions: A pain in the neck for NLP. – In: *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, CICLing '02*. Springer-Verlag, pp. 1–15.
- Savary et al. 2019: Savary, A., S. Cordeiro, C. Ramisch. Without lexicons, multiword expression identification will never fly: A position statement. – In: *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*. Florence, Italy. Association for Computational Linguistics, pp. 79–91.
- Todorova, Leseva, Stoyanova 2017: Todorova, M., Sv. Leseva, Iv. Stoyanova. Rečnik na sastavnite dumi v balgarskiya ezik – razvitie i perspektivi. – In: *Dokladi ot Mezhdunarodnata yubileyna konferentsiya na Instituta za balgarski ezik "Prof. Lyubomir Andreychin" (Sofia, 15–16 may 2017 godina)*. T. 1, Sofia, Institut za balgarski ezik "Prof. Lyubomir Andreychin", s. 311–319.
- Zaimov 2012: Zaimov, Y. Balgarski vodopis. Geografsko opisanie, stroezh i proizhod na imenata. Veliko Tarnovo, St Cyril and St Methodius University Press, t. 1.
- Zaimov 2021: Zaimov, Y. Balgarski vodopis. Geografsko opisanie, stroezh i proizhod na imenata. Sofia, Prof. Marin Drinov Publishing House, t. 3.

✉ Assist. Prof. Dr. Svetlozara Leseva  
Department of Computational Linguistics  
Institute for Bulgarian Language "Prof. Lyubomir Andreychin"  
Bulgarian Academy of Sciences  
52 Shipchenski prohod, Bl. 17, Sofia 1113, Bulgaria