

PRINCIPLES OF THE INFLECTIONAL DESCRIPTION OF VERBAL MULTIWORD EXPRESSIONS IN BULGARIAN

SVETLA KOEVA

INSTITUTE FOR BULGARIAN LANGUAGE, BULGARIAN ACADEMY OF SCIENCES

svetla@dcl.bas.bg

This article presents the general principles for the inflectional description of verbal multiword expressions in Bulgarian. Two main criteria are applied: the maintenance of the relation to the inflectional description of simple words in Bulgarian and the presentation of the minimal grammatical information necessary for a standardised description of the forms of verbal multiword expressions. The paper also provides a brief outline of the formalism used to define local grammars, which are compiled into finite-state transducers that recognise and generate the forms of verbal multiword expressions, along with the associated grammatical information.

Keywords: inflectional types; verbal multiword expressions; verbal phraseologisms; Bulgarian

REFERENCES

- Baldwin, Kim 2010: *Baldwin, T., S. N. Kim*. Multiword expressions. – In: Indurkha, N., F., J. Damerau (eds.). *Handbook of Natural Language Processing*, 2nd ed. Boca Raton, CRC Press, pp. 267–292.
- Bally 1909: *Bally, Ch.* *Traité de stylistique française*. Volume 1. Paris, C. Winter.
- Barkalova 1997: *Barkalova, P.* *Balgarskiyat sintaksis: poznat i nepoznat*. Plovdiv, Universitetsko izdatelstvo “Paisiy Hilendarski”.
- Benveniste 1974: *Benveniste, É.* *Formes nouvelles de la composition nominale*. – In: Benveniste, É. (ed.). *Problèmes de linguistique générale*. Vol. 2. Paris, Gallimard, pp. 163–176.
- Blagoeva et al. 2022: *Blagoeva, D., S. Koeva, V. Sosonovski, M. Yaskot*. *Rechnik na aktivnata frazeologiya v balgarskiya i polskiya ezik*, Sofia, Izdatelstvo na Balgarskata akademiya na naukite “Prof. Marin Drinov”.
- Breidt et al. 1996: *Breidt, E., F. Segond, G. Valetto*. Formal description of multi-word lexemes with the finite-state formalism IDAREX. – In: *16th International Conference on Computational Linguistics. Proceedings of the Conference, COLING 1996*. Copenhagen, Center for Sprogteknologi, pp. 1036–1040.
- Chrobot et al. 1999: *Chrobot, A., B. Courtois, M. Herman, M. Gross, K. Zellagui*. *Dictionnaire Electronique DELAC anglais: noms composés*. Rapport technique n°59, LADL, Université Paris 7.

- Constant et al. 2017: *Constant, M., G. Eryğit, J. Monti, L. van der Plas, C. Ramisch, M. Rosner, A. Todirascu*. Multiword expression processing: A survey. – *Computational Linguistics* 43:4, pp. 837–892.
- Courtois, Silberztein 1990: *Courtois, B., M. Silberztein*. Dictionnaires électroniques du français. – *Langue française*, 87 (1), pp. 11–22.
- Fellbaum 1993: *Fellbaum, C*. The determiner in English idioms. – In: Cacciari, C., P. Tabossi (eds.). *Idioms: Processing, structure and interpretation*. Hillsdale, NJ, Lawrence Erlbaum Associates, pp. 271–296.
- Grégoire 2010: *Grégoire, N*. DuELME: a Dutch electronic lexicon of multiword expressions. – *Language Resources and Evaluation*, 44, pp. 23–39.
- Gross 1990: *Gross, G*. Définition des noms composés dans un lexique-grammaire. – *Langue Française*, 87, pp. 84–90.
- Gross 1997: *Gross, M*. The construction of local grammars. – In: Roche E., Y. Schabès (eds.). *Finite-State Language Processing*. MIT Press, pp. 329–354.
- Kaldieva-Zaharieva 2013: *Kaldieva-Zaharieva, S*. Balgarska frazeologiya. Sofia, Akademichno izdatelstvo “Prof. Marin Drinov”.
- Koeva 1998: *Koeva, S*. Gramatichen rechnik na balgarskiya ezik. Opisane na kontseptsiyata za organizatsiyata na lingvistichnite dannii. – *Balgarski ezik*, N 6, s. 49–58.
- Koeva 1999: *Koeva, S*. Sintaksis i punktuatsiya. Plovdiv, Makros-2000.
- Koeva 2006: *Koeva, S*. Inflection morphology of Bulgarian multiword expressions. – In: *Computer Applications in Slavic Studies*. Sofia, Boyan Penev Publishing House, pp. 201–216.
- Koeva 2008: *Koeva, S*. Balgarskiyat FrameNet. Semantiko-sintaktichen rechnik na balgarskiya ezik – kontseptualen model. – In: Koeva, S. (sast.). *Balgarskiyat FrameNet: Semantiko-sintaktichen rechnik na balgarskiya ezik*. Sofia, Institut za balgarski ezik “Prof. Lyubomir Andreychin”, s. 5–57.
- Koeva 2024: *Koeva, S*. Rechnik na aktivnata frazeologiya v balgarskiya i polskiya ezik: osobenosti na flektivnite tipove na balgarskite leksikalni edinitsi. – In: Kolkovska, S. (sast. i red.). *Frazeologichni studii*. Sofia, Izdatelstvo na Balgarskata akademiya na naukite “Prof. Marin Drinov”, s. 60–78.
- Koeva et al. 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.
- Kutsarov 2007: *Kutsarov, I*. Teoretichna gramatika na balgarskiya ezik. Morfologiya. Plovdiv, Koala pres.
- Kyuvlieva-Mishaykova 1986: *Kyuvlieva-Mishaykova, V*. Frazeologizmite v balgarskiya ezik. Sofia, Narodna prosveta.
- Lion-Bouton et al. 2023: *Lion-Bouton, A., A. Savary, J.-Y. Antoine*. A MWE lexicon formalism optimised for observational adequacy. – In: *Proceedings of the 19th Workshop on Multiword Expressions (MWE 2023)*. Dubrovnik, Association for Computational Linguistics, pp. 121–130.
- Maldzhieva 2017: *Maldzhieva, V*. Printsipi na formalното opisane na polisegmetnite ezikovi edinitsi v Normativnata gramatika na balgarskiya ezik. – *Balgarski ezik*, N 2, s. 7–19.

- Maldžieva, Wójtowicz 1994: *Maldžieva, V., J. Wójtowicz*. Polskie i bułgarskie związki frazeologiczne. Warszawa, Wydawnictwo Uniwersytetu Warszawskiego.
- Nicheva 1987: *Nicheva, K.* Balgarska frazeologija. Sofia, Nauka i izkustvo.
- Nicheva et al. 1974: *Nicheva, K., S. Spasova-Mihaylova, K. Cholakova*. Frazeologičen rechnik na balgarskiya ezik. T. 1. Sofia, Izdatelstvo na Balgarskata akademiya na naukite.
- Nunberg et al. 1994: *Nunberg, G., I. Sag, T. Wasow*. Idioms. – *Language*, 70, pp. 491–538.
- O’Grady 1998: *O’Grady, W.* The syntax of idioms. – *Natural Language and Linguistic Theory*, 16, pp. 279–312.
- Oflazer et al. 2004: *Oflazer, K., Ö. Çetinoğlu, B. Say*. Integrating morphology with multi-word expression processing in Turkish. – In: *Proceedings of the Workshop on Multiword Expressions: Integrating Processing*. Barcelona, Spain, Association for Computational Linguistics, pp. 64–71.
- Paskaleva 2007: *Paskaleva, E.* Kompyutarna morfologija. Resursi i instrumenti. Sofia, Institut za paralelna obrabotka na informatsiyata.
- Penchev 1998: *Penchev, Y.* Sintaksis na savremenniya balgarski knizhoven ezik. Ase novgrad, Izdatelska kashta “Vechernik”.
- Przepiórkowski et al. 2017: *Przepiórkowski, A., J. Hajič, E. Hajnicz, Z. Urešová*. Phraseology in two Slavic valency dictionaries: Limitations and perspectives. – *International Journal of Lexicography*, 30 (1), pp. 1–38.
- Sag et al. 2002: *Sag, I. A., T. Baldwin, F. Bond, A. Copestake, D. Flickinger*. Multiword expressions: A pain in the neck for NLP. – In: Gelbukh, A. (ed.). *Computational Linguistics and Intelligent Text Processing. CICLing 2002. Lecture Notes in Computer Science*. Vol. 2276, Berlin/Heidelberg, Springer, pp. 189–206.
- Sailer, Trawiński 2006: *Sailer M., B. Trawiński*. The collection of distributionally idiosyncratic items: A multilingual resource for linguistic research. – In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06), Genoa, Italy*. European Language Resources Association (ELRA), pp. 471–474.
- Savary 2005: *Savary, A.* A formalism for the computational morphology of multiword units. – *Archives of Control Sciences*, 15(3), pp. 437–449.
- Savary 2008: *Savary, A.* Computational inflection of multi-word units. – In: *Linguistic Issues in Language Technology*, CSLI Publications, Vol. 1, Issue 2, pp. 1–53.
- Savary et al. 2018: *Savary, A., S. Petitjean, T. Lichte, L. Kallmeyer, J. Waszczuk*. Object-oriented lexical encoding of multiword expressions: Short and sweet. arXiv preprint arXiv:1810.09947.
- Silberztein 1993: *Silberztein, M.* Les groupes nominaux productifs et les noms composés lexicalisés. – *Linguisticae Investigationes*, 17(2), pp. 405–425.
- Silberztein 2005: *Silberztein, M.* NooJ’s dictionaries. – In: *Proceedings of Language and Technology Conference, Poznań*. Poznań, Wydawnictwo Poznańskie, pp. 291–295.
- Stoyanova 2012: *Stoyanova, I.* Avtomatichno razpoznavane i tagirane na sastavni leksikalni ediniti v balgarskiya ezik. Doktorska disertatsiya. Sofia, Institut za balgarski ezik “Prof. Lyubomir Andreychin”.

- Stoyanova, Todorova 2014: *Stoyanova, I., M. Todorova*. Razrabotvane na rechnitsi na sastavnite leksikalni edinitsi v balgarskiya ezik za tselite na kompyutarnata lingvistika. – In: Koeva, S. (sast.). *Ezikovi resursi i tehnologii za balgarski ezik*. Sofia, Akademichno izdatelstvo “Prof. Marin Drinov”, s. 185–202,
- Todorova 2015: *Todorova, M.* Tipologiya i svoystva na ustoychivi slovosachetaniya v balgarski. Glagolni frazeologizmi. Doktorska disertatsiya. Sofia, Institut za balgarski ezik “Prof. Lyubomir Andreychin”.
- Vinogradov 1947: *Vinogradov, V. V.* Russkii iazyk. Grammaticheskoe uchenie o slove. Leningrad, Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo.
- Vondřička 2019: *Vondřička, P.* Design of a multiword expressions database. – In: *The Prague Bulletin of Mathematical Linguistics*, 112, pp. 83–101.
- Zgusta 1967: *Zgusta, L.* Multiword lexical units. – *Word*, 23 (1–3), pp. 578–587.

✉ *Prof. Dr. Svetla Koeva*

Department of Computational Linguistics
Institute for Bulgarian Language, Bulgarian Academy of Sciences
52 Shipchenski prohod, Bl. 17, 1113 Sofia, Bulgaria