REVIEW

by Assoc. Prof. Petya Nikolova Barkalova, DSc

Paisii Hilendarski **University of Plovdiv**

on the materials submitted

for the acquisition of the academic position of *Associate Professor*

at the Institute of Bulgarian Language "Prof. Lyubomir Andreychin"

Area of professional qualification 2.1. Philology (General and Comparative Linguistics)

Following Order No. RD-09-20/21.02.2023 of the Director of the Institute of Bulgarian Language "Prof. Lyubomir Andreychin" at the Bulgarian Academy of Sciences (IBL, BAS), I have been appointed a member of the Academic Review Board in the competition for the academic position of associate professor in area of professional qualification 2.1. Philology (General and Comparative Linguistics), which was opened for the needs of the Computational Linguistics Department at IBL, BAS.

The single applicant for the Associate Professor open position, announced in the State Gazette, Issue 102 of 23.12.2022, for the needs of the Department of Computational Linguistics, is Ch. Asst. Prof. Tsvetana Ivanova Dimitrova, PhD from the Department of Computational Linguistics at IBL, BAS.

1. General overview of the submitted materials

The set of materials submitted on paper as well as on a digital source by Ch. Asst. Prof. Tsvetana Ivanova Dimitrova, PhD, is in accordance with the Law on the Development of the Academic Staff in the Republic of Bulgaria (LDASRB), the Rulebook for the Application of LDASRB and the corresponding Rulebook on the Conditions and Regulations for the Acquisition of Scientific Degrees and Academic Positions of IBL, BAS. It includes the following documents: curriculum vitae; PhD dissertation abstract; certificate for occupying the *chief assistant professor* academic position at IBL, BAS, issued on 09.11.2012; a complete list of publications and a list of publications for participation in the competition: one monograph (habilitation thesis) and 22 articles and papers accompanied by summaries; a list of the habilitation contributions; a list of citations; attestation documents; a list of international and national research projects in whose teams the candidate participated; a list of participations in scientific forums; other documents certifying awards received, as well as a two-month stay at *Ca Foscari* University in Venice as a visiting scholar of the Department of Linguistics and Comparative Cultural Studies from 15.09.2022 to 15.11.2022. Co-authored publications are accompanied by agreement statements from the co-authors that indicate what the candidate's contribution is, or indications on this matter are provided in the footnotes and abstracts.

2. Brief biographical data

Ch. assistant professor Tsvetana Ivanova Dimitrova, PhD graduated from *St. Kliment Ohridski* Sofia University with a major in Bulgarian philology and a second major in Turkish

philology. She obtained her PhD degree in General Linguistics at the Norwegian University of Science and Technology in Trondheim. She has been working as a researcher at the Department of Computational Linguistics since 2008 and she has held the position of Chief Assistant Professor since 2012. She teaches the Corpus Linguistics lecture course in the Master's Degree Program "Computer Linguistics. Internet technologies in humanities" at the Faculty of Slavic Philology of the *St. Kliment Ohridski* Sofia University. Her scientific interests and scientific output are in the following fields: theoretical linguistics, historical and modern syntax, language dynamics from Old Bulgarian through Middle Bulgarian to Modern Bulgarian, corpus linguistics and linguistic annotation. She holds the post of a Scientific Secretary of IBL, BAS. She was awarded the *Prof. Marin Drinov* Award for young scientists in the field of humanities for 2011. Since 2017, she has been the Secretary of the Commission on Computer Processing of Manuscripts and Old Printed Books of the International Committee of Slavists. She is a member of the Slavic Linguistics Society and the Global Wordnet Association.

3. General characteristics of the applicant's professional activities

Tsvetana Dimitrova is a prolific author with her own distinctive style, recognizable both nationally and internationally. The research experience accumulated by her is proven in a categorical way in her impressive habilitation thesis published in the form of the monograph "Bulgarian Clitics. History and Present" (2023) as well as in the presented articles and papers, divided thematically into two areas: 1. Language resources, computer modelling of language (14 titles) and 2. Historical corpora, description of historical linguistic data (8 titles).

The complete list of publications of Tsvetana Dimitrova (with individual and joint authorship) contains 67 titles, of which the monograph and 22 articles and papers are included in the competition. She has participated in 34 scientific forums over the period, and of these 34 19 were in Bulgaria and 15 – abroad. The total number of noted citations is 125, of which 112 appeared after the acquisition of the scientific position of assistant professor (in 2012). She has taken part in 7 international and 7 national projects, being the head of one of the international projects (identifier: No. - ABR: 06-1 EBR Automatic recognition of named objects in Bulgarian and Czech) herself. Of the 22 articles and papers presented, 6 are individually authored, and 16 are co-authored with leading scholars in the field of computational linguistics, generative grammar, medieval studies, Balkan, Slavic and general linguistics, and the publications are in prestigious peer-reviewed refereed editions in Bulgaria and abroad – national journals and thematic proceedings of international scientific conferences (14 with one co-author, 2 with two co-authors; 6 in Bulgarian, 1 in Russian and 15 in English).

Scientific contributions

Tsvetana Dimitrova's monograph "Bulgarian Clitics. History and Present" (Sofia: Institute of Bulgarian Language "Prof. Lyubomir Andreychin" at BAS, 2023. 256 pp. ISBN 978-619-908-845-6) consists of an introduction, three parts (I. Clitics: essence and features; II. Clitics in the history of the Bulgarian language; III. Analysis), conclusion, sources used, sources cited and an appendix. The study is the result of many years of research work, which provides a comprehensive scientific understanding of the positions and behavior of the Bulgarian (paradigmatic) clitics and

the mechanisms that govern them in diachronic and synchronic terms. The pronominal and verbal clitics, the lost discourse clitics "же/zhe" and "бо/bo" and the interrogative "ли/li" were observed in a deep diachronic cut: from the Old Bulgarian period (Codex Marianus, Codex Zographensis, Codex Assemanianus, Sava's book, Codex Suprasliensis), through the Middle Bulgarian period (Troyan parable, Vlach-Bulgarian letters) to the early New Bulgarian period (Trojan Damascene and Lovech Damascene). The state of the language today is traced through the records in dialect and colloquial speech corpora and examples from the BNC. The very empirical basis of the research, in terms of its scale and the manner of excerption, is a rare example of combining a high technological culture with general philological and specialized linguistic knowledge.

The rich source material made it possible to observe, on the one hand, the changes in the clitic "inventory" itself (the loss of the discourse clitics "же/zhe" and "бо/bo"), on the other – to confirm the positional stability of the clitic complex within the syntactic (and prosodic) structure of the verb phrase and the sentence as a whole. The study proves the relative stability of the syntactic positions of clitics in the sentence from the Old Bulgarian period to the present day. Analytical procedures are based on the so-called phase theory. Although the conceptual apparatus of this theory is presented after the systematic observations of the data (in part III Analysis), it is evident that the knowledge of the philosophy and the conceptual apparatus of the generative theory has facilitated both the orientation in the non-uniform nature of the clitics, and the very observations of their positioning. The observations "see" the exact segmentation of the phrases and the clitics in them, the constituents are clearly identifiable, and an explanation for the positions of the clitics in the structures is sought through the governing mechanisms. In this section (6.2. Sentence structure, pp. 199 – 202), after the theoretical framework, the theoretical model of the sentence structure to which the author adheres is presented, as well as scenarios for the derivation of clitic positions chosen by authors with a similar understanding (R. Pancheva, St. Franks) in order to substantiate the vision of a "dual" position of clitics in Bulgarian, because they appear in two different phases of the minimal sentence. Drawing on the mechanisms and principles behind the "movement" operation Tsvetana Dimitrova explores the left sentence boundary and the boundaries of the verb phrase (including the internal relations with the negative "не (no)", the question "ли (whether)", the copula and the auxiliary "съм (be)", "да (to)"), noticing activation of accusative and dative pronominal clitics. The assumption is made that accusative pronominal clitics show dependence on the verb, probably because they enter into the composition of its argument structure, while other clitics are realized rather at the level of the sentence (discourse clitics in Old Bulgarian, some datives and possessive clitics that have no relation to the verb). Thus, the author confirms her ability to formulate and justify new hypotheses, as well as her ability to prove essential new aspects of already existing scientific fields with the help of new means. It is especially worth noting that all the sources used in the study are available in computer-processed collections and corpora, which is a sign of a switch to a new technological level of grammatical-philological work demonstrated by the author.

In the Appendix, 73 tables with summarized quantitative data and descriptions of the positional placement of clitics in simple verb forms (+ copular verbs) and in complex verb forms are provided. All types of clitics are covered: discourse, interrogative, dative clitics and quasi-clitics, accusative first-person, second-person, third-person and reflexive clitics, dative reflexive

clitics and verbal clitics derived from 'съм, е/am, is' in copular and auxiliary function. The described models have matrix organization and in the linear order present the action of Wackernagel's Law and especially the Tobler-Mussafia Effect in the Bulgarian language from the Old Bulgarian period to the present day. The linear word-order patterns of clitics (2P – not 2P – 1P, or second position – non-second position – first position) correspond to their place in the phrasal organization: pre-verbal, post-verbal or distant. According to the author, the mere presentation of the data "gives an opportunity for verification, correction", and the data can be used "for the purposes of other research" (p. 10). This is a sign of high scientific ethics.

The state of clitics and the clitic complex from the Old Bulgarian literary language to the present stage of the developmental dynamics of the Bulgarian language has not been studied so comprehensively so far. The work is innovative in its combination of both a diachronic and a synchronic approach. Knowledge of the old (Greek sources, Old Bulgarian, Middle Bulgarian, Damascene texts) and the new state of the studied linguistic phenomenon, combined with the most modern syntactic knowledge, is demonstrated. Thus, the author gives more proof of her ability to formulate and confirm new hypotheses, as well as her ability to prove essential new aspects of already existing scientific fields by using new means.

The subject matter of the monograph is relevant, so far unexplored so comprehensively and significant for the history and the current state of Bulgarian and Slavic linguistics. Set against the background of the debate on clitics in the latest generativist studies, the research contributes to a better typological characterization of the Bulgarian language. The author's capacity for broad and in-depth theoretical thinking, as well as her sound linguistic training, become evident. The ability to systematize empirical data meticulously and to adhere to the chosen theoretical framework is clearly visible. A knack for following paradigms and traditions is also shown. For example, the succinct and clear overview of the stages in the generative theory presented in Ch. III (6.2. Theoretical framework, pp. 194-199) is contributive compared to similar problem-oriented reviews in the scientific literature in Bulgaria of the last 40 years, especially with its explanations of the conceptual apparatus of N. Chomsky's Minimalist Program, which, unlike the theory of Government and Binding (Penchev 1993), does not have a complete transfer to Bulgarian language material.

The presented articles from the thematic list 2. *Historical corpora, description of historical linguistic data* ([16] – [23]) are related to the habilitation thesis. In some of them, certain issues concerning the analysis of the clitic complex are solved, not only in terms of the historical changes in the heterogeneous system of clitics ([16] discourse clitics, [18] the particle "ще/will" as a verb quasi-clitic element), but also in relation to the analytical procedure itself. For instance, in [17] there is evidence that the observed pronominal clitics occupy two positions – a sentence one and a second one (2P), which has been inherited in the modern Bulgarian language. A good preliminary result is the "capture" of an extraphrasal position for dative pronominal clitics, called the "external possessor construction". The obtained confirmatory facts ([20], Dimitrova, Krapova) correspond with the idea of "dual position" of clitics further developed in the monograph. Paper [19] presents the concept of the clitic complex as a cluster with two levels – primary, which consists of a (pronominal) core and (verb) periphery, and secondary, which contains (pseudoclitic) elements

that have different behavior both diachronically and synchronically. Convincing solutions for the detection of sentence boundaries in Old Bulgarian texts are those relying on the systematic behavior of clitics, are convincing, e.g. the discourse clitics "же/zhe" and "бо/bo" as reliable reference points for the left border of the sentence [22]. From a syntactic point of view, this article is of particular merit also because of the review of a number of definitions of the sentence, across a wide range of criteria, with the aim delineating the minimum number of criteria. This approach is also helpful in annotating anaphorically related elements ([23], Dimitrova, Boyadzhiev) in parchment manuscripts and in manuscripts from different periods of language development for the purposes of a future annotated corpus of medieval Slavic texts in the form of an electronic edition.

The articles from thematic list 1. *Language resources, computer modeling of languages* ([2] – [15]) correspond to the tasks implemented on a project basis at the Department of Computational Linguistics at IBL, BAS. Their implementation leads to the creation of new classifications, methods and technologies related to the Bulgarian WordNet, the Bulgarian National Corpus, the parallel corpora and other newly created resources. [6], [7] and [8] contain a semantic classification of adjectives intended for implementation in WordNet, with [8] paying particular attention to the principles of applying the classification. In [9], [10], [11] and [12] the semantic web is further developed by delving into a wider range of semantic relations between nouns and verbs [9] and between nouns in WordNet [10]. Seven relations (Agent, Actor, Experiencer, Patient, Recipient, Causer, Theme) related to the Agent and Undergoer relations are highlighted in the predicate-argument structure of verbs [9]. The implemented strategy of transferring the word-formative and morphosemantic relations between nouns and verbs and nouns and adjectives to (new) relations between nouns is interesting. The idea of hierarchical inheritance is productive and captures the systemic property of each "subordinate" word/concept (species) to inherit the distinguishing features and functions from the "superordinate" word/concept (genus) [10]. Following this principle 17 new morphosemantic relations between nouns in WordNet have been added, connected by relations with verbs and adjectives [10]. The article [11] explores in the same vein participles, whose number among the sets of adjectives in Wordnet is small. An approach for detecting participle forms in adjective synonym sets related to verb synonym sets is proposed. The semantic types and classes of nouns referring to food and drink and occurring in Wordnet are the subject of detailed investigation in [12]. A comparison is made with the semantic types in another resource – the Corpus Pattern Analysis (CPA) ontology. On the basis of connectivity through hyperonym-hyponym relations, the study seeks reliable validation of the types and classes attributed there. The validation is related to another important task of the Deparment of Computational Linguistics – the transfer to verb synonym sets.

In the paper ([13], Tarpomanova, Dimitrova) theoretical and practical aspects of syntactic annotation are described, and the Bulgarian-English corpus with correlated (simple) sentences (BulEnSI) is presented. Automatically divided and correlated sentences (including within complex sentences) are verified by experts. In the paper ([14], Koeva, Dimitrova) the structural and semantic classification of the syntactic combinations of two nouns, in particular the combinations that name persons, geographical entities and organizations, has been enriched greatly. The collective study [15] presents a comparative study of predicative state constructions in Russian and Bulgarian. Tsvetana Dimitrova provides observations on the Bulgarian-Russian parallels

through the extraction of data from the Bulgarian-Russian and Russian-Bulgarian parallel (sub)corpora within the National Corpus of the Russian language, as well as through the descriptive part of the analysis related to the overall corpus study of the predicative state constructions in both languages.

Articles [2], [3], [4] and [5] are dedicated to WordNet. In [2] Tsvetana Dimitrova presents the description of the annotation of derivational relations in Bulgarian WordNet, and in [3], [4] and [5] – the description of the work with WordNet's creation and editing system, created by Borislav Rizov and called Hydra for Web.

Quoted

The recognition of Ch. Asst. Prof. Tsvetana Dimitrova, PhD in the scientific community in Bulgaria and abroad is indisputable. The noted citations are a total of 125, and 112 of them are after the acquisition of the scientific position of Chief Assistant Professor. The total number of citing references listed in Scopus/WoS is 46. The geography of the citing publications (journals and proceedings of scientific forums) with high representativeness includes all parts of the world: Japan, China, India; Brazil, USA, France, Germany; Estonia, Poland, Iceland; Spain, Italy, etc.

Among the citing Bulgarian refereed indexed journals included in Scopus/WoS are: Balkan Linguistics (1), Reports from the International Annual Conference of IBL, BAS (6), as well as the Bulgarian Language Journal (7), included in ERIH Plus and etc. The author is cited in 23 monographs, as well as by doctoral students in 9 PhD dissertations (4 in Bulgaria, 5 abroad, of which 1 on the topic of Bulgarian clitics – at MIT (Massachusetts Institute of Technology, 2019).

Educational activity and scientific supervision

At the Faculty of Slavic Philology at the Sofia University "St. Kliment Ohridski" of Ch. Asst. Prof. Tsvetana Dimitrova, PhD has been appointed to lead the lecture course in Corpus Linguistics in the *Computer Linguistics* Master's program (with a duration of 30 lecture hours). From 15.09.2022 to 15.11.2022, she was received as a visiting scholar at the Department of Linguistics and Comparative Cultural Studies at *Ca Foscari* University in Venice. In the previous years she delivered three lecture cycles (8 hours each) there under the Erasmus program: "Bulgarian Language Resources Online", April 2018; "Ordering phenomena in Bulgarian", May 2019; "Bulgarian language as a Slavic Balkan language. Bulgarian as an analytic language", November 2019. In the project "Semantic classification of adjectives in the Bulgarian WordNet" Tsv. Dimitrova was V. Stefanova's scientific supervisor.

4. Assessment of the candidate's personal contribution

Ch. Asst. Prof. Tsvetana Dimitrova, PhD presents in her monograph and related publications another successful attempt (after the PhD dissertation on the topic "The Old Bulgarian Noun Phrase: towards Annotation Specificity", 2010) to apply the generative approach together with modern approaches for syntactic annotation to the analysis and processing of linguistic data from earlier stages in the development of the Bulgarian language.

Tsvetana Dimitrova proves with her overall scientific creativity, on the one hand, her ability to set high scientific goals and to tackle them on a large scale through her own research strategies, using her high linguistic erudition, and on the other hand, to successfully partner in teamwork projects whose results are peak achievements. Some of the publications are the result of the collective implementation of significant research tasks in the field of computational linguistics, corpus linguistics, digitization of written manuscripts, general linguistics, Slavic studies and Bulgarian studies. Tsv. Dimitrova proves that she is able to solve both fundamental and practical tasks with a sustainable effect on the methodological and resource enrichment of the linguistic potential of the research field in Bulgaria as a whole.

5. Critical remarks and recommendations

I would recommend to Ch. Asst. Prof. Tsvetana Dimitrova, PhD to find a wider audience of young scholars to whom she can convey her specific experience of working with Old Bulgarian and Middle Bulgarian manuscripts founded in digital collections through the lens of generative syntax and the techniques of computational linguistics. I am convinced that work in this research field has yet to reveal more unknown aspects of the developmental dynamics and structural richness of the Bulgarian language. Here I also include the preparation of teaching aids, lecture courses, work with undergraduates, graduates and doctoral students.

6. Personal impressions

I know Tsvetana Dimitrova as a cheerful and responsive young person with respectable discipline, erudition and extensive systematic thinking, enhanced by modern analytical skills and attention to detail. The way she uses high-tech resources and tools in her work as a philologist and a linguist is both respectable and inspiring.

CONCLUSION

The documents and materials submitted by Ch. Asst. Prof. Tsvetana Ivanova Dimitrova, PhD, meet all the requirements of the Law on the Development of the Academic Staff in the Republic of Bulgaria (LDASRB) and the Rulebook for the Application of LDASRB. The achievements of Ch. Asst. Prof. PhD Tsvetana Dimitrova's results in the research activity fully correspond to the minimum national and additional requirements of the Regulations for the conditions and procedures for acquiring scientific degrees and for holding academic positions in the IBL at the BAS.

Having acquainted myself with the materials and scientific publications submitted for the purposes of this position, having analyzed their significance and the contributions entailed in them with regard to their theoretical, scholarly, and pedagogical merits, I hereby confidently give my positive assessment and recommend to the Scientific Jury to prepare a report proposal to the Scientific Council of the Institute of Bulgarian Language "Prof. Lyubomir Andreychin" for the selection of Tsvetana Ivanova Dimitrova for the academic position of *Associate Professor* at the Institute of Bulgarian Language in the area of professional qualification 2.1. Philology (General and Comparative Linguistics).

17. 04. 2023   Reviewer: ....

                                          (Assoc. Prof. Petya N. Barkalova, DSc)