

## STATEMENT

on the call for the academic position of associate professor in direction 2.1. Philology, scientific speciality General and Comparative Linguistics for the needs of the Computational Linguistics Section at the Institute of Bulgarian Language "Prof. L. Andreychin", following the announcement in State Gazette of Bulgaria 102/23.12.2022,

with a single candidate: Assistant Professor Tsvetana Ivanova Dimitrova, PhD.

by: Associate Professor Milena Petrova Dobрева, Institute of Mathematics and Informatics - BAS

### 1. General description of the presented materials

The candidate has presented one monograph and 22 articles and studies (14 in the field of building linguistic resources and 8 in the field of historical corpora). Of these 22 publications, 7 are journal articles, 4 are series studies, and 11 are full-text reports in peer-reviewed conference proceedings with international visibility.

Eight publications submitted in response to the call are visible in SCOPUS, i.e. two-thirds of all the candidate's publications that are visible in this database, which shows a steady increase in the presence of the candidate in publications with international high scientometric indicators. One hundred twenty-five citations of the candidate's works were collected (174 are visible on Google Scholar).

The documentation for participation in the competition also includes a certificate of participation in national and international projects (14 in total), a certificate of participation in scientific forums (34 in total), documents from the candidate's last certification for the period 2017-2020, as well as a certificate of one award and confirmation of the visit as a visiting researcher at the Ca ' Foscari in University of Venice, Italy. The application pack includes a description of the candidate's contributions and her biography, as well as the full texts of all publications on the competition as required.

2. General characteristics of the candidate's scientific, scientific-applied and pedagogical activities. Participation in scientific projects and project management

The main and long-lasting interests of the candidate are in the field of computational linguistics; her visible presence since the early years of her scientific path in this field was marked by the awarding of a Young Scientist Award in the name of Marin Drinov in 2011, only three years after the defense of her dissertation "The Old Bulgarian Name Phrase: Towards Annotational Specification " at the Norwegian University of Science and Technology, Trondheim .

It is very satisfying to observe the candidate develop and mature in this area with deepening work in two areas, the issues surrounding the preparation and analysis of historical corpora and the broader area of language resources and computer language modelling.

In the field of preparation of historical corpora, the candidate combines work on formalized descriptions of linguistic phenomena with practical applicability in the Bulgarian National Corpus and in WordNet for the Bulgarian language.

There is no data for prepared graduates and doctoral students; however, there is a profound contribution to scientific projects in cooperation with teams from Bulgaria and foreign countries. In the period 2012-2023, the candidate participated in 7 projects (for one of these projects was the principal investigator) with funding from various international programs, as well as in 7 projects with funding from Bulgarian national and institutional research programs.

The candidate's contribution to the work of the project "Semantic classification on adjectives in Bulgarian WordNet " (funded by DFNP) and " Semantic network with a wide spectrum of semantic relations", financed by the National Scientific Research Fund of Bulgaria provided the space for some of the reported research and publications.

The candidate contributed actively to international and national scientific forums. In 2022 she was a visiting researcher at the Ca ' Foscari University of Venice. Since 2020, she is also giving a lecture course on Corpus Linguistics in Master's degree programme "Computational linguistics. Internet technologies in humanitarian studies " of the Slavic philologies faculty of Sofia University " St. Kliment Ohridski".

The candidate's profile demonstrates a good combination of project-driven research, publications and teaching activity. She is an internationally recognized



researcher with participation in program committees of forums in the field of computational linguistics.

### 3. Analytical assessment of the main scientific and/or scientific-applied contributions of the candidate.

In the vast field of computational linguistics Dr. Tsvetana Dimitrova has been dedicating her efforts on a few mutually related and complementary topics which are applicable in the development of digital resources for the Bulgarian language.

Part of her work is devoted **to the study of clitics in different periods of the development of the Bulgarian language**. Her research monograph (1) is dedicated to clitics; she draws from thorough observations over excerpted examples from Old Bulgarian, Middle Bulgarian and Early modern Bulgarian periods, contemporary literary examples and observations from dialects and colloquial speech. The collection of materials from these diverse sources is a challenge in its own right. It is particularly commendable that Dr. Dimitrova is following and applying the Open Science principles and made the primary materials used in her study available in the public domain. In Bulgaria, sharing in the spirit of open science is rarely applied in the Humanities. The analytical results of Dr. Dimitrova in the analysis on the clitics elements is presented in the monograph (1) and publications (16, 17, 18, 19, 20), which as a set, give a comprehensive and deeper view of the use of clitics.

About 1/3 of the publications of Dr. Dimitrova are dedicated on problems around **the organization of data and structuring on different linguistic resources** aimed at the addition of Bulgarian content to WordNet (see publications 2, 3, 4, 5; and Bulgarian-English parallel body with relative ( simple ) sentences ( BulEnAC ) – publication 13), as well as structuring on linguistic material which reflects development on Bulgarian language in different periods from time (publications 22 and 23). The combination of theoretical and practical focus in the presented works is one of the merits of her research outcomes which are instrumental in the coordinated development of the methodology and the delivery of digital linguistic resources.

In addition, the applicant explores problems around **the practical issues of creating digital linguistic resources**. In this area, her most substantial contribution is to the development of the Bulgarian WordNet and the Bulgarian national corpus. This is the area where she also actively contributed to research projects.

In the project " Semantic classification on adjectives in Bulgarian WordNet" Dr. Dimitrova contributed scientific guidance, preparation of data description

formalisms and detailed argumentation of multilevel classification principles. These achievements are presented in publications 6 – 8.

In the project "Semantic network with a wide spectrum of semantic relations" (OG 10-3 / 14.12.2016), financed by the National Scientific Research Fund, the candidate works by formulation and attribution on additional semantic relations between nouns (10) and nouns and verbs (9) as well as inclusion on participles in synonymous multiples on adjectives WordNet names for the Bulgarian language (11). Formulated are also steps for checking on the validity of attributed semantic types by attitude on compatibility with the existing ones semantic classes in WordNet for the Bulgarian language (12).

The work of Dr. Dimitrova shows a relatively rare combination in the local research community: she brings together excellent knowledge of traditional methods for the analysis of language development with the ability to contribute formal linguistic models describing language phenomena for use in digital resources. In addition, the candidate demonstrates an attitude and contribution to the principles of open science, which is particularly useful for expanding the scope and use of available resources for research and its expansion by other scholars.

It is noteworthy that the candidate shows excellent results both in independent research and in working with diverse teams. This is another sign of scientific maturity and the ability to achieve one's achievements, combined with successfully leading and contributing to teamwork.

#### 4. Critical notes and recommendations on the works presented

Personally, I find the applicant's works in the field of historical corpora of Old Bulgarian texts particularly interesting (16, 23). The knowledge of the available resources is impressive, and the comparisons of their approaches and elements are very useful. What would help a holistic approach is also to include observations of what scientific questions the various available corpora were built around. This aspect has not been addressed so far and could be considered to add to the work in the field of historical corpora in conjunction with clarifying how a corpus should be annotated in order to be of maximum use in teaching Old Bulgarian language.

The candidate's papers note the different approaches to selecting what elements to annotate, but these choices are usually a function of the corpus authors' intentions for this resource to be used for a particular type of research. It would be interesting to relate the annotation structures to the creators' intentions and research questions.



I also find of potential interest to expand the research into the direction of applicability of machine learning and other artificial intelligence methods in the annotation of historical corpora.

## 5. Conclusion

*In conclusion, I confirm that the presented materials fully comply with the requirements of the Bulgarian Academy of Sciences, the Regulations for the Implementation of the Bulgarian Academy of Sciences, the Regulations for the Terms and Conditions for the Acquisition of Scientific Degrees and for Holding Academic Positions in the Bulgarian Academy of Sciences, the Regulations for the Conditions and the procedure for obtaining scientific degrees and for holding academic positions at the Institute of Bulgarian Language "Prof. Lyubomir Andreychin "and I strongly recommend the esteemed commission to confirm Dr. Tsvetana Ivanova Dimitrova to be awarded with the academic position of associate professor in direction 2.1. Philology, scientific speciality General and Comparative Linguistics for the needs of the Computational Linguistics Section at the Institute of Bulgarian Language "Prof. L. Andreychin ".*

Date April 20, 2023.

Prepared the opinion: .....

(signature)