



REVIEW

by Prof. Yovka Tisheva, PhD
Department of Bulgarian Language, Faculty of Slavic Studies
Sofia University "St. Kliment Ohridski"

for a dissertation for the acquisition of the educational and scientific degree "Doctor"
Field of higher education: 2. Humanities; Professional direction 2.1. Philology
Scientific specialty: General and comparative linguistics

Author of the dissertation: Victoria Krasimirova Petrova-Lyubenova

Dissertation topic: "Semi-automatic development of multilingual terminological resources"

I. Data about the author of the dissertation and the doctoral program

Viktoria Petrova graduated from the bachelor's program "Italian Philology" at the Faculty of Classical and Modern Philology at Sofia University. In 2016, she successfully completed her studies in the master's program "Computer Linguistics. Internet Technologies in the Humanities" at the Faculty of Slavic Studies, Sofia University. After completing the master's program, V. Petrova increased her qualifications through a specialization at the Institute of Bulgarian Language "Prof. Lyubomir Andrejcin" (from May 1 to July 31, 2017) and three participations in summer schools: International Terminology Summer School, organized by TermNet and the Center for Modern Languages, Budapest University of Technology and Economics (from July 1 to 4, 2020 . and from 1 to 4 July 2021) and Tradumàtica Summer School: Translation Technologies organized by the Universitat Autònoma de Barcelona (from 13 July to 31 July 2020). In connection with the professional growth of V. Petrova, it should be noted her internship as a linguist at the Center for Translations of the EU Bodies in Luxembourg from April 1, 2015 to August 20, 2017, where her work was related to creation and testing of specialized translation engines, and her work as a junior project manager and translation quality expert at BiConnected Bulgaria in the period 2018 – 2019. This experience, as well as the work as a freelance translator, prepare a large part of the practical aspects and contributions in the presented dissertation.

From January 1, 2019 till July 1, 2022, Viktoria Petrova is a full-time doctoral student in the "General and Comparative Linguistics" PhD program at the Department of Computational

Linguistics, Institute for Bulgarian Language "Prof. Lubomir Andreychin". Scientific supervisor is Prof. Svetla Koeva, PhD.

For the period of her studies, Viktoria Petrova collected a total of 300 credits. She successfully passed the exams in specialized subjects, language training and computer skills. She passed her doctoral minimum (grade: very good 5) and the two exams in the BAS: in Western language (English language; grade: 5.75) and in presentation skills (grade: 6.00). In 2019 and 2020, Victoria Petrova also attended two courses at the Doctoral School at the Faculty of Slavic Studies: "Academic Writing" and "Electronic Resources in the Humanities, Part 1", which ends with excellent grades. I know the doctoral student as a participant in these courses, and I can state that during the training she showed a serious and responsible attitude to her duties in preparing assignments aimed at mastering the genres of written academic communication and deepening the skills of working with electronic resources. V. Petrova gave a report before a scientific seminar of the primary scientific unit. She participated with reports on the topic of the dissertation in several scientific events in our country and one abroad: the International Annual Scientific Conference of the Institute for Bulgarian Language in 2020 and in 2021; Second Workshop on Human-Informed Translation and Interpreting Technology, 2019 in Varna, as well as Translation and Interpreting Forum, 2019 in Olomouc.

The data for the period of PhD program show that V. Petrova expanded her general theoretical training in research methodology, in the field of machine translation and translation quality assessment methods. The dissertation, the abstract, the publications on the subject of the dissertation and administrative documentation for the procedure have been drawn up correctly and in compliance with all the requirements of the Law on the Development of the Academic Staff in the Republic of Bulgaria, the Regulations for its Application and the Regulations for the Conditions and Procedures for Acquisition of scientific degrees and for occupying academic positions at the Institute of the Bulgarian Language in connection with the procedure for acquiring the educational and scientific degree "Doctor".

II. Data for the dissertation and the abstract

Victoria Petrova-Lyubenova's dissertation "Semi-automatic development of multilingual terminological resources" (with a volume of the main part of 176 pages) consists of an introduction, four chapters, a conclusion, three appendices, a list of some terms and a bibliography. The appendices show the results of the survey conducted among translators in Bulgaria, the automatically extracted terms in English in the field of computer terminology and

the English-Bulgarian terminological base in the same field are presented. Analytical information and conclusions are adequately illustrated through 23 figures and 6 tables.

- Relevance of the problem developed in the dissertation work

Globalization and dynamic contacts in all social spheres require rapid access by users with different levels of language competence to content created in different languages. This leads to the increasingly active use of various applications and tools for automatic translation. To improve the quality of machine translation, both theoretical research and experimental work are needed to reach effective practical solutions for the automatic processing of linguistic information and the translation of different types of text. This determines the topicality of the issue in scientific applied terms and its importance for expanding theoretical knowledge in the interdisciplinary field between computational linguistics and translation studies, related to automatic and computer-assisted translation. Last but not least, the topic has social significance due to the possibility that the obtained results can optimize the translation work on texts with a high degree of standardization and terminology of the content.

The subject of the study is the terminological resources that are used in modern computer-aided translation tools. The subject of the work is the creation of complete and consistent terminological resources with the help of modern technologies. The main goal of the dissertation is related to the semi-automatic development of multilingual terminological resources, proposing a methodology for creating terminological resources for a specific specialized area. The expected result is limited to the creation of an English-Bulgarian terminological base in the field of computer terminology. The realization of the scientific goal is related to four groups of scientific tasks, which are specified for each of the chapters in the dissertation:

1. Overview of existing computer-aided translation systems and their components: translation memory, terminological base, machine translation.

This task has been successfully implemented in the first chapter of the dissertation. The first section outlines the specifics of computer-assisted and machine translation. The concepts of translation memory - a database that stores expressions and phrases from documents already translated by a person, and terminological base - a database containing terms specific to a given thematic area are presented. The author also comments on the requirements for the skills of translators today in connection with the introduction of new technologies in their work, on the one hand, and the expectations of translators in connection with the functionalities of computer-

assisted translation systems, on the other. The first chapter ends with a presentation of the results of a survey conducted with professional Bulgarian translators, related to the level of use of computer-aided translation systems and their components, the ways of creating and using terminological resources by translators in their practice and the methodologies for measuring the quality of translation.

2. Detailed analysis of the structure of terminological bases

The second chapter of the dissertation deals with this task. Different definitions of the concept *term* related to different contexts and types of text are commented. The doctoral student motivates the approach that will be applied in the third chapter when working with the corpus data. It is based on concepts and one on individual lexical units, with all terminological information relating to a concept being processed as a single terminological record. I appreciate as a correct and reliable principle accepted in the work that the data is organized around the meaning, and not around the specific term. The given data on the nature of the terminological bases build on what was already said in the first part of the text, in the comparison of the translation memory with the terminological base. The main characteristic of terminological bases is the possibility that they are personal and subject to customization for any institution, organization, company or translator. The connections between terminological bases and translation memory are presented. Since the criteria for selecting terms depend on the requirements of the specific activity or text, V. Petrova proposes six steps for determining whether a given term is appropriate for a given project or text.

3. Development and application of a methodology for semi-automatic creation of multilingual terminological bases, more specifically an English-Bulgarian terminological base in the field of computer terminology

In the third chapter of the text, V. Petrova comments on the features of computer terminology, the available monolingual and bilingual (English-Bulgarian) specialized dictionaries and the methods for automatic extraction of terms. Automatic term extraction technologies allow more entities to be identified by increasing the correspondence between terms in the terminology base and in the textual corpus with which they are associated. According to the author, such an approach largely reduces subjectivity, because in this way terms that are actually used in the corpora are included in the terminological base. However, according to the author, manual verification can still be seen as most reliable way to ensure a quality terminological base.

Contribution in a practical-applied plan is stated in the second and third sections of the chapter. The methodology developed by the doctoral student for the semi-automatic creation of an English-Bulgarian terminological base for a specific field is presented in order to propose a working procedure that can be used by translators in Bulgaria.

4. Comparison of the possibilities for evaluating the quality of the translation and their application in the evaluation of the created terminological base

The results included in the last chapter of the text are related to this task. They represent the possibilities of measuring translation accuracy with computer-aided translation systems, as these programs have built-in quality control functionalities. They check the terms used, the length and segmentation of the text, as well as the correspondence of numbers, punctuation marks, spaces, capital letters, and specific characters from the source language in the target segment. Two experiments to test the performance of these programs on Bulgarian data were done. Errors related to terminology that occur in a real working environment are introduced in the source texts. The experiments show that the term base can be successfully used in order to check translation quality.

The last part of the dissertation contains summaries that show that the main objective has been successfully met. Based on the analysis and comparison of the technological means used in translation (computer-aided translation programs, their components: translation memory, terminological base, machine translation, tools for automatic term extraction and translation quality verification), a methodology for semi-automatic creation of terminological resources for the specific specialized area is proposed.

- Presentation of the state of the problem and the scientific literature

In her work, Victoria Petrova cites 134 sources of scientific and practical information. The bibliography list mainly includes titles in English. The review of the used scientific literature, as well as the introductory parts of the four chapters of the dissertation, shows that the doctoral student knows the general theoretical models related to the organization and operation of computer-assisted translation systems and their application to texts with different stylistic and genre characteristics. In the review of the literature on the essence and definition of the terms as a specific layer of the lexical system, the terminological dictionaries, the terminological bases and the translation memory, knowledge of the specifics of each of the commented objects is evident. The author is familiar with the structure and specifics of the currently existing terminological resources for Bulgarian.

- Comment on the chosen methodology

The methodology applied in the characterization of terminological dictionaries, terminological databases and translation memory, as well as the specific methods in creating a terminological database in the field of computer terminology and in assessing the accuracy of the translation is adequate to the set scientific tasks. Their application leads to an objective and non-contradictory presentation of the specifics of each of the objects in view of the peculiarities of the source and target languages and the tools used during translation. Victoria Petrova uses a complex methodology, including traditional general scientific methods (description, analysis, comparison, classification) and empirical qualitative and quantitative methods (experiment, statistical analysis, computer processing, etc.). The methodology for the compilation of the parallel bilingual term corpus, for the semi-automatic creation of a terminological base from this corpus and for the development of a bilingual terminological base follow the criteria established in practice for the construction of this type of language resources. The databases created for the purposes of the work are sufficient and allow the implementation of the specific tasks formulated for each chapter of the text.

- Brief description of scientific and/or scientific-applied contributions

The presented work shows the indisputable achievements of Victoria Petrova in solving a scientific-applied problem using methods corresponding to the modern achievements of science and practice. The two groups of contributions mentioned in the dissertation are the personal work of the doctoral student and were achieved as a result of her independent work on the realization of the scientific goals. A theoretical contribution is the analysis of the possibilities of computer-aided translation systems and of systems for checking the quality of translation, tested on data from a specific thematic area. A significant scientific-applied contribution is the presented original solution to an actual problem related to the mechanisms for compiling a parallel bilingual corpus of terms, for semi-automatic creation of a terminological base from this corpus and for the development of a bilingual terminological base.

The dissertation abstract comprehensively and accurately presents the structure of the work, the results of the studies, practical and experimental activities, as well as the scientific contributions of the dissertation work.

- Publications on the topic of the dissertation work

The presented publications meet the minimum requirements of the Institute of the Bulgarian Language for the acquisition of a scientific degree. Victoria Petrova has three articles on the subject of her dissertation. All are published in proceedings of international scientific events. Two of the texts are in English. It is important to note that the first of the articles included in the list of V. Petrova's publications has already been cited in a study by a foreign scientist.

Conclusion

Evaluating the topicality of the dissertation "Semi-automatic development of multilingual terminological resources", the results achieved and the conclusions drawn, I propose to the respected Scientific Jury to award Victoria Petrova-Lyubenova the educational and scientific degree "Doctor" in the field of higher education 2. Humanities, professional direction 2.1. Philology, scientific specialty General and comparative linguistics.

December 12, 2022

Prof. Yovka Tisheva, PhD