REVIEW

By prof. Maxim Ivanov Stamenov, DSc, Institute for Bulgarian Language "Prof. Lyubomir Andreychin", BAS

of a dissertation for obtaining a PhD degree

Author of dissertation: **Viktoria Krasimirova Petrova-Lyubenova**, doctoral student at the Section of Computational Linguistics, Institute for Bulgarian Language "Prof. L. Andreychin", Bulgarian Academy of Sciences

Topic of dissertation: **Semi-automatic development of multilingual terminology resources**.

Scientific adviser: prof. Svetla Koeva, PhD from the Institute for Bulgarian Language "Prof. Lyubomir Andreychin", BAS.

The overall volume of the dissertation is 237 pages, which includes 176 pages of main text and 61 pages of appendices and bibliography. It contains an introduction, four chapters, a conclusion, contributions, 3 appendices and a list of terms. The work includes 23 figures and 6 tables.

1. Relevance of the problem developed in the dissertation in theoretic and-applied terms and main results

The problem to which the proposed dissertation is dedicated is very relevant in the context of the latest directions of information technology development with a view to expanding the possibilities of computer-assisted translation (based on translation memory systems, machine translation and terminological bases). The work is a comprehensive study, the purpose of which is to propose a methodology for creating terminological resources for a given specialized field. In order to achieve this, the possibilities of computer-aided translation programs for the creation of terminological bases, of term extraction programs, as well as of currently available terminological resources for the Bulgarian language have been analysed.

The Introduction to the dissertation indicates the relevance of the research, its motivation, its subject, goals and objectives, methods used in the work and outlines its structure. The main goal is to propose a methodology for creating terminological resources for a given specialized field. The subject of the research is terminological bases and their use in computer-aided translation systems, as well as the ways in which modern information technologies enable and facilitate their creation and use.

The first chapter provides an overview of modern computer-aided translation systems, emphasizing their main components (translation memory, machine translation, terminological bases). Here, the new knowledge and skills required of translators in relation to the use of computer-aided translation technologies are adequately described. For this purpose, the current state of use of computer-aided translation systems, their components and new functionalities are outlined. For an appropriate orientation of current practices to date, a survey is being conducted among translators in Bulgaria (73 professional translators) regarding the attitudes towards the use of computer-assisted translation systems.

The second chapter presents the types of terminological dictionaries and their structure in comparison with the structure of the terminological bases. The aim is to outline the

similarities and differences, as well as the advantages of using terminological bases in the process of working with computer-aided translation systems. The structure and features of terminological bases are also discussed, and for this purpose, world-renowned terminological bases that can be used inside or outside computer-aided translation systems are outlined. Special attention is paid to the creation and management of terminology bases in computer-aided translation systems. The main ways of developing and managing term bases in systems for this type of translation are explained. In addition, applications that can be used together with computer-aided translation systems as well as stand-alone terminology handling systems that can be used in parallel are considered. Use of the terminological bases is demonstrated depending on the different types of text for translation and the application of controlled languages. Main attention in this regard is paid to the requirements for coherence and consistency in the use of terms and the possibility of imposing specific restrictions.

In the third chapter, features of computer terminology are interpreted, with the orientation towards generalization from the point of view of the inclusion of computer terms in terminological resources. It describes how to create terminological resources, starting with traditional printed terminologies, moving through online terminologies and reaching automatic term extraction and presenting existing tools for this purpose. The description of the tools is accompanied by a comparison of their functionalities within a certain period of time of testing them. A methodology for the semi-automatic creation of a terminological base in the field of computer terminology has also been created, and the specific stages of the work on the development of a bilingual terminological base for English and Bulgarian are described. The characteristics of the thus created bilingual terminological base are presented. A unified way of working in the creation of terminological bases is proposed, which is based on the best practices and takes into account the capabilities and the way of work of the translators.

In the fourth chapter, an assessment of translation accuracy with computer-aided translation systems without or with the use of term bases is offered. Issues related to measuring translation accuracy with functionalities built into or external to computer-aided translation tools are discussed. A classification of translation errors with computer-aided translation systems is proposed in terms of their relevance to the content of the text. A comparison of the effectiveness of the translation evaluation programs was made according to a predetermined methodology.

As a summary of the achieved results, it can be stated that the dissertation proposes a methodology for the semi-automatic creation of a bilingual terminological base with the direction of translation from English to Bulgarian, and the field to which it refers is computer terminology. Also discussed are some possibilities for measuring translation quality in computer-aided translation systems that have built-in functionalities for this. For the purpose in question, the dissertation presents the results of two experiments aimed at verifying the behavior of quality programs when translating from English into Bulgarian.

2. Orientation of the applicant in the state of the problem and the scientific literature dedicated to it

The author displays a good orientation into the literature on the topic of the work. In the bibliography of the dissertation, a total of 134 publications of our and foreign scientists, with whom the dissertation student worked in the realization of the activities related to the research and writing of her work, and 18 sources from the Internet are indicated.

3. Nature of the chosen methodology

In developing the dissertation, the author uses both modern methods (statistical, for computer processing, corpus) and some traditional ones (descriptive, analytical, experimental). Analysis and description are perceived as fundamental to achieving research goals. Analysis is applied to the selection of appropriate language resources and tools for their processing. Hypotheses are verified experimentally through the use of such tools as an online survey, work with automatic word processing programs, automatic term extraction, automatic sentence alignment of parallel documents, and automatic creation of bilingual term bases. When working with all programs for computer-aided translation, modern means are used for computer processing of texts, for development of terminological bases, for machine translation, and for generation of translation memory. Quantitative methods are also applied to summarize the collected information, as is the case with the survey of translators and with the automatic extraction of terms. The corpus approach involves the annotation of translated texts with errors in order to test the functionalities of translation quality checking programs.

4. Brief analytical description of the scientific and / or applied contributions of the dissertation

We acknowledge the presence of the following contributions in the paper. They are oriented in two main directions - theoretical and with applied potential.

Among the theoretical contributions, it is appropriate to mention the following. First of all, the possibilities of computer-aided translation systems and the changes they bring with them are presented and analyzed in a comprehensive review. This is based on a review of current research in the field of computer-aided translation. Second, available ways of creating term bases for computer-aided translation systems are explored and discussed. Thirdly, a description and analysis of quality control systems is offered.

Among the contributions with an applied orientation, the following should be highlighted. A survey was conducted, with which the researcher addresses translators in Bulgaria who work with computer-aided translation systems, to explore aspects of their practices in this regard. A methodology has been developed for the semi-automatic creation of terminological resources for the Bulgarian language for computer-aided translation systems, taking into account some limitations (free software, support in Bulgarian and in the field of computer terminology). Last but not least, a unifying classification of translation errors is proposed.

5. Evaluation of the publications of the applicant on the dissertation subject: number, nature of the articles in which they are printed, citations

The dissertation student reports a total of three articles as publications related to the topic of the dissertation. Two of them were presented at the IBE Annual Conferences and published in the respective Proceedings, and the third was published in the Proceedings of the Computer Technologies for Translation Workshop. The publications described meet the minimum requirements for obtaining the scientific degree "Doctor of Philology". A list of two citations to her publications is also attached.

6. Critical remarks

The presence of a Conclusion (pp. 170-175) in the Dissertation itself is not noted in the Contents of the Dissertation (pp. 3-5), nor in the description of the structure of the dissertation in the Autoreferate (pp. 6-7).

7. Does the abstract correctly reflect dissertation's content

The Autoreferate is detailed (61 pp.) and correctly reflects the main points and scientific contributions of the dissertation. It corresponds to the content of the work under evaluation.

8. Conclusion

Having in mind the above presented review, I propose to the Scientific Jury to award the applicant **Viktoria Krasimirova Petrova-Lyubenova** the scientific and educational degree "Doctor" in professional field 2.1 Philology, scientific specialty General and Comparative Linguistics (Computational Linguistics) for her dissertation "**Semi-automatic development of multilingual terminology resources**".

10.12.2022 г. Signed:

(Signature)