



## REVIEW

for dissertation for obtaining the scientific and educational degree *doctor*

Author of the dissertation: Viktoria Petrova-Lyubenova, Department of Computational Linguistics, Institute of Bulgarian Language Prof. Lubomir Andreychin, Bulgarian Academy of Sciences

Topic of the dissertation: Semiautomatic development of multilingual terminological resources

Reviewer: assoc. prof. Ekaterina Tarpomanova, Sofia University Saint Kliment Ohridski

The dissertation *Semi-automatic development of multilingual terminological resources* by Viktoria Petrova-Lyubenova examines an important and actual issue in the field of computational linguistics, related to the broad topic of the creation and processing of language resources. A number of authors raise the question of the application of machine translation in the daily work of translators, with opinions leaning towards the development of computer-aided translation rather than the direct application of machine translation, which often turns out to be of no direct benefit. It is namely the question of the development of resources applicable in computer-aided translation that is the focus of the dissertation.

The PhD student is well acquainted with the state of affairs in the field she is researching, as well as with the existing literature on the matter. In the bibliography and correspondingly reflected in the text of the dissertation, important studies are present – both more general and key works in the field of computational linguistics, computer processing of natural languages, the history of machine translation, computer-aided translation, as well as more concrete and specific works that are related to the topic of the dissertation – translation memory, development of terminological resources, term extraction, etc. At the same time, we must note that this topic has not been developed in Bulgarian science, and the doctoral student can rely only on a few

more general lexicological and terminological studies, as well as on a few articles on term extraction. In this sense, it can be said that Victoria Petrova-Lyubenova's dissertation is innovative for the Bulgarian language.

The research methodology applied in the dissertation work is suitable for achieving the goals and objectives. The main goal of the dissertation is to propose a methodology for creating terminological resources for a given specialized field, in the present case – in the field of computer terminology. The methodology used can be defined as complex. On the one hand, in the theoretical part, the methods of descriptive linguistics are applied, where the scientific literature on the subject, the review of the existing terminological bases and the tools for computer-assisted translation are presented. In the applied part of the dissertation – the compilation and evaluation of the bilingual terminological database, methods from computer linguistics, statistical and corpus methods were used. I believe that the methodology is well constructed and appropriately applied, which contributes to the achievement of the goals of the dissertation.

Victoria Petrova-Lyubenova's work follows the established structure for a dissertation. It consists of an introduction, four substantive chapters, a conclusion (which for some reason is not noted in the table of contents), contributions, three appendices, a list of some terms, and a bibliography.

The introduction is structured in a standard way – it presents the relevance of the research, the motivation behind it, the object and subject of the research, the goals and tasks, the methods used and the composition of the work. All introductory components are written concisely, succinctly and to the point. The main objective of the study is supported by more specific objectives for each of the chapters.

The first chapter examines computer-aided translation systems and their components. Computer-aided translation is specialized software that facilitates translation from one language to another. Unlike machine translation, in computer-aided translation the translator manages the entire process through various functions. In this section, several definitions of computer-aided translation are given, as well as illustrations of the interface of various programs. The main components of computer-aided translation are the translation memory and the terminological base. Additionally, some systems integrate machine translation into their functionalities, which is why it is also considered as a component in the dissertation. Translation memory has been described as an important part of computer-aided translation. It is a database in which translations already made are stored. The doctoral student very well describes the structure and functioning of the translation memory – the segment as the basic unit, the

calculation of similarity, the necessary software that is applied. Machine translation is quite appropriately considered in the first chapter as a possible component of computer-aided translation. The two most common models of machine translation – statistical and neural – are briefly described. The possibilities for the integration of machine and computer-assisted translation are analytically presented, with the doctoral student demonstrating knowledge both in the development of this issue and in the practical implementation and its results in some modern systems. Terminology bases are the third component presented, and this part shows their essence, the benefits of their application and the mechanism by which they function.

The first chapter ends with a survey of translators' attitudes towards computer-aided translation technologies. Victoria Petrova-Lyubenova compiled a survey of 26 questions and conducted it among 73 professional translators. Her observations of translators' attitudes are a good motivation for her research in both the theoretical and applied parts, as they show an interest in technology among a fairly good percentage of respondents.

The second chapter focuses specifically on termbases as part of computer-aided translation systems. The chapter begins with definitions and descriptions of the concept of term, distinguishing between terminology in a broad and narrow sense. This is followed by a description of a terminological dictionary and a termbase and the important distinctions between the two types of resources. The specifics of terminological dictionaries are presented from the point of view of the thematic domain, as well as the information they contain. The doctoral student rightly points out the role of termbases in computer-aided translation systems, as they can be managed by the translator, the company or the institution that uses them. Several of the open access termbases are described. An analytical comparison between terminological dictionaries and termbases follows, highlighting the advantages of both types of resources in terms of reliability, accessibility, coverage, etc. The author did not fail to point out some shortcomings of each type of resource, for example, the unclear authorship of some internet dictionaries, which results in the lack of information about their reliability. The relationship between termbases and translation memory is also examined. Special attention is paid to the development and management of termbases in computer-aided translation systems with specific examples from several systems, as well as the possibilities of integrating additional programs to the main ones, or add-ons. Other important aspects are terminology management systems, which can be used independently or be part of a computer-aided translation system, and also the so-called controlled languages, which are particularly suitable for the readability of specialized and technical texts. At the end of the chapter the importance of terminology-related activities is emphasized.

The third chapter deals with the semi-automatic creation of multilingual termbases, which is the real part of the work. The chapter begins with more general questions – an introduction to the subject of computer terminology on which the term base focuses, an introduction to existing dictionaries, methods for automatic term extraction, and tools for this. What follows is the presentation of the methodology developed by the doctoral student for the semi-automatic creation of a terminological base in the field of computer terminology for a pair of languages – Bulgarian and English. As a resource for term extraction, the best choice was made, namely the bilingual corpus. As the most suitable corpus for the purpose, the author focuses on the English-Bulgarian corpus with computer terminology QTLeap. Experiments were conducted to select a term extraction tool that met pre-set criteria, ultimately weighing in favor of a combination of programs. After the compilation of the terminological base, a manual check was made. The newly created database contains terms in English and their Bulgarian equivalents, as well as several fields for additional information. From the presented method of compiling the termbase, I object only to the exclusion of synonyms. This means that when using the resource for reverse translation (from Bulgarian to English) the synonymous word will not be recognized as a term. There are some inaccuracies in the linguistic analysis of the terms (e.g. the origin of the term *плейлист* is not by calquing).

The fourth chapter deals with the quality measurement in computer-aided translation systems and their components. It is quite right to start with the proviso that translation quality is a relative and subjective concept, which is why it is difficult to be measured. The quality measurement experiments are excellently conducted and described by the PhD student. In this way, the benefit of the application of the termbase in real situations is seen. Quality measurement programs show different types of errors, and the idea of classifying and grading errors to detect their translation consequences is a very good one. At the end of the fourth chapter, the international translation quality standards are presented, which were adopted to unify the translation activity and guarantee the quality of the translation.

The conclusion of the dissertation is too narrative and is more of a concise retelling of the text than presenting the results of the research. Substantial conclusions are made sporadically, for example in the third paragraph it is stated that termbases are important because they save searching for terms, but the most valuable component of computer-aided translation is the translation memory.

The three appendices following the text present research-relevant data in a convenient format and are very useful. In fact, in the text of the dissertation itself, there are tables and figures that visualize various data and statistics and facilitate the perception of information.

My overall impression of the dissertation is very good, as it shows Victoria Petrova-Lyubenova's in-depth view of the subject, as well as her experience as a translator. The text is well structured, the content development follows the pattern from a general theoretical approach to a specific and concrete description of the information. The competence of the PhD student in the methods of computer-assisted translation, her good orientation in the systems that are used and in the general theory of translation and terminology is indisputable. A good balance has been achieved between the theoretical and applied parts of the dissertation. In the final version of the work, some typographical errors remain, which should be corrected, especially if a decision is made the work to be published.

Приносите на дисертацията са добре представени от авторката, като са разделени на теоретични и практически. Аз бих изтъкнала, че подобна тема се разработва за първи път за българския език, а също така в приложен аспект за първи път се създава и описва научно този тип терминологична база. Към приносите отнасям и класификацията на грешките при работа със системи за компютърнопомогнат превод.

Виктория Петрова-Любенова представя три публикации по темата на дисертацията, като всички са в сборници от конференции. Първото ѝ участие е в конференция по превод във Варна, а следващите две в годишната конференция на Института за български език, която се наложи като един от престижните форуми в българската лингвистика.

Авторефератът отговаря на всички изисквания, правилно и добре отразява съдържанието на дисертацията.

В заключение на гореизложеното убедено подкрепям присъждането на образователната и научна степен „доктор“ на Виктория Петрова-Любенова.

The contributions of the dissertation are well presented by the author, being divided into theoretical and practical. I would point out that such a topic is being developed for the first time for the Bulgarian language, and also, in an applied aspect, this type of termbase is being created and described scientifically for the first time. Among the contributions, I include the classification of errors when working with computer-aided translation systems.

Victoria Petrova-Lyubenova presents three publications on the topic of the dissertation, all of which are in conference proceedings. Her first participation was in a conference on translation in Varna, and the next two in the annual conference of the Institute for the Bulgarian Language, which is established as one of the prestigious forums in Bulgarian linguistics.

The abstract meets all the requirements, correctly and well reflects the content of the dissertation.

In conclusion of the above, I strongly support the awarding of the educational and scientific degree "doctor" to Victoria Petrova-Lyubanova.

December 5, 2022

assoc. prof. Ekaterina Tarpomanova