

РЕЦЕНЗИЯ

от проф. д-р Йовка Тишева
Катедра по български език, Факултет по славянски филологии
СУ „Св. Климент Охридски“

за дисертационен труд за придобиване на образователната и научна степен **“доктор“**
Област на висше образование: 2. Хуманитарни науки
Професионално направление 2.1. Филология
Научна специалност: Общо и сравнително езиковедие

Автор на дисертационния труд: Виктория Красимилова Петрова-Любенова

Тема на дисертационния труд: „Полуавтоматично разработване на многоезикови терминологични ресурси“

I. Данни за автора на дисертационния труд и за докторантурата

Виктория Петрова е завършила бакалавърската програма „Италианска филология“ във Факултета по класически и нови филологии на СУ „Св. Климент Охридски“. През 2016 г. тя е успешно приключва обучението си в магистърската програма „Компютърна лингвистика. Интернет технологии в хуманитаристиката“ във Факултета по славянски филологии на СУ. След завършване на магистърската програма В. Петрова повишава квалификацията си чрез една специализация в Института за български език „Проф. Любомир Андрейчин“ (от 1 май до 31 юли 2017 г.) и три участия в летни училища: International Terminology Summer School, организирано от TermNet and the Centre for Modern Languages, Budapest University of Technology and Economics (от 1 до 4 юли 2020 г. и от 1 до 4 юли 2021 г.) и Tradumàtica Summer School: Translation Technologies, организирано от Universitat Autònoma de Barcelona (от 13 юли до 31 юли 2020 г.). Във връзка с професионалното израстване на В. Петрова трябва да се отбележи стажът ѝ като лингвист към Центъра за преводи на органите на ЕС в Люксембург от 1 април 2015 г. до 20 август 2017 г., където работата ѝ е била свързана със създаване и тестване на специализирани мотори за превод, и работата ѝ като младши проект мениджър и експерт качество на преводи в БиКонектед България ООД в периода 2018 – 2019 г. Този опит, както и работата като преводач на свободна практика подготвят голяма част от практикоприложните аспекти и приноси в представената дисертация.

От 1 януари 2019 г. до 1 юли 2022 г. Виктория Петрова е редовен докторант по програма „Общо и сравнително езиковедие“ в Секцията по компютърна лингвистика на Института за български език „Проф. Любомир Андрейчин“. Научен ръководител е проф. д-р Светла Коева. За периода на обучението си докторантката е събрала общо 300 кредита. Успешно е положила изпитите по специализираните предмети, езикова подготовка и компютърни умения. Взела е докторантския си минимум (оценка мн. добър 5) и двата изпита в ЦО на БАН: по западен език (английски език; оценка: 5.75) и по презентационни умения (оценка: 6.00). През 2019 г. и 2020 г. Виктория Петрова посещава и два курса в Докторантското училище при Факултета по славянски филологии на СУ „Св. Климент

Охридски“ – „Академично писане“ и „Електронни ресурси в хуманитаристиката, 1. част“, които приключва с отлични оценки. Познавам докторантката като участник в тези курсове и мога да посоча, че по време на обучението тя показва сериозно и отговорно отношение към задълженията си при подготовка на задания, насочени към овладяване на жанровете в писмената академична комуникация и задълбочаване на уменията за работа с електронни ресурси. В. Петрова е изнесла доклад пред научен семинар на първичното научно звено. Участвала е с доклади по темата на дисертацията в няколко научни събития у нас и едно в чужбина: Международната годишна научна конференция на Института за български език през 2020 г. и през 2021 г.; Second Workshop on Human-Informed Translation and Interpreting Technology през 2019 г. във Варна и Translation and Interpreting Forum през 2019 г. в Оломоуц.

Данните за периода на обучението показват, че докторантката е разширила своята общотеоретична подготовка в методологията на научните изследвания, в областта на машинния превод и методите за оценка на качеството на превода. Дисертационният труд, авторефератът, публикациите по темата на дисертацията и цялата административна документация по процедурата са оформени коректно и при спазване на всички изисквания на Закона за развитието на академичния състав в Република България, на Правилника за приложението му и на Правилник за условията и реда за придобиване на научни степени и за заемане на академични длъжности в Института за български език във връзка с процедурата за придобиване на образователната и научна степен „доктор“.

II. Данни за дисертационния труд и за автореферата

Дисертационният труд на Виктория Петрова-Любенова „Полуавтоматично разработване на многоезикови терминологични ресурси“ (с обем на същинската част от 176 стр.) се състои от увод, четири глави, заключение, три приложения, списък на някои термини и библиография. В приложенията са показани резултатите от направената анкета сред преводачите в България, представени са автоматично извлечените термини на английски в областта на компютърната терминология и английско-българската терминологична база в същата област. Аналитичната информация и изводите от анализите са подходящо онагледени чрез 23 фигури и 6 таблици.

- Актуалност на разработвания в дисертационния труд проблем

Глобализацията и динамичните контакти във всички социални сфери налагат бърз достъп на потребители с различни нива на езикова компетентност до съдържание, създадено на различни езици. Това води до все по-активното използване на различни приложения и инструменти за автоматичен превод. За подобряване на качеството на машинния превод са необходими както теоретични изследвания, така и експериментална работа, чрез която да се достигне до ефективни практически решения за автоматичното обработване на езиковата информация и превода на различни типове текст. Това определя актуалността на проблематиката в научноприложно отношение и важноста ѝ за разширяване на теоретичните знания в интердисциплинарната област между компютърната лингвистика и преводознанието, свързана с автоматичния и компютърно подпомогнатия превод. Не на последно място, темата има и социална значимост поради възможността получените резултати да оптимизират преводаческата

работа върху текстове с висока степен на стандартизираност и терминологизация на съдържанието.

Обект на проучването са терминологичните ресурси, които се използват в съвременните инструменти за компютърнопомогнат превод. Предмет на работата е създаването на пълни и непротиворечиви терминологични ресурси с помощта на съвременните технологии.

Основната цел на дисертацията е свързана с полуавтоматичното разработване на многоезикови терминологични ресурси, като се предложи методика за създаване на терминологични ресурси за конкретна специализирана област. Очакваният резултатът е ограничен до създаване на английско-българска терминологична база в областта на компютърната терминология.

Реализирането на научната цел е свързано с четири групи научни задачи, които са конкретизирани за всяка от главите в дисертацията:

1. Обзор на съществуващите системи за компютърнопомогнат превод и техните компоненти: преводна памет, терминологична база, машинен превод.

Тази задача е реализирана успешно в първата глава от дисертацията. В първия раздел са изведени спецификите на компютърнопомогнатия и на машинния превод. Представени са понятията преводна памет – база от данни, която съхранява изрази и фрази от вече преведени от човек документи, и терминологична база – база от данни, съдържаща специфични за дадена тематична област термини. Авторката коментира и изискванията към уменията на преводачите днес във връзка с навлизането на новите технологии в тяхната работа, от една страна, и очакванията на преводачите във връзка с функционалностите на системите за компютърнопомогнат превод, от друга. Първа глава приключва с представяне на резултатите от проведена анкета с професионални български преводачи, свързана с нивото на използване на системите за компютърнопомогнат превод и техните компоненти, начините на създаване и употреба на терминологични ресурси от преводачите в практиката им и методологиите за измерване на качеството на превода. В. Петрова обобщава, че отговорите на българските преводачи не се различават в значителна степен от тези при анкети по темата, проведени в други страни. Тъй като новите технологии, напр. преводна памет или терминологични бази, могат да допринесат за подобряване на бързината и донякъде на консистентността на превода, преводачите се стремят да ги използват. Негативно влияние върху нагласите обаче оказват два факта: липсата на публичнодостъпни терминологични бази в много области от човешкото познание и недостатъчно качественият машинен превод от и на „малки“ езици и езици, различни от английски.

2. Подробен анализ на структурата на терминологичните бази

С тази задача се занимава втората глава от дисертацията. В нея са коментирани различни дефиниции на понятието *термин*, свързани с различни контексти и типове текст, в резултат на което докторантката мотивира подхода, който ще се прилага в третата глава при работата с корпусните данни. Той е базиран на понятието, а не на отделни лексикални единици, като цялата терминологична информация, отнасяща се до едно понятие, се обработва като един терминологичен запис. Оценявам като правилен и надежден приетия в работата принцип данните да са организирани около значението, а не около конкретния термин. Приведените данните за същността на терминологичните

бази надграждат вече казаното в първата част на текста, при сравнението на преводната памет с терминологичната база. Като основна характеристика на терминологичните бази е изведена възможността да са персонални и подлежат на персонализация за всяка институция, организация, фирма или преводач. Представени са връзките между терминологичните бази и преводната памет. Тъй като критериите за подбор на термини зависят от изискванията на конкретната дейност или текст, В. Петрова предлага шест стъпки за определяне на това дали даден термин е подходящ за даден казус.

3. Разработване и приложение на методика за полуавтоматично създаване на многоезикови терминологични бази, по-конкретно на английско-българска терминологична база в областта на компютърната терминология

В третата глава от текста В. Петрова коментира особеностите на компютърната терминология, достъпните едноезични и двуезични (английско-български) специализирани речници и методите за автоматично извличане на термини. Технологиите за автоматично извличане на термини позволяват да се идентифицират повече единици, като се повишава съответствието между термините в терминологичната база и в текстовия корпус, с чието съдържание са свързани те. Според авторката при такъв подход до голяма степен се намалява субективността, защото така в терминологичната база се включват термини, които действително се употребяват в корпусите. Според авторката обаче ръчната проверка все още може да се смята за най-добрия и сигурен начин за гарантиране на качествена терминологична база.

Принос в практико-приложен план е заявен във втория и в третия раздел на главата. Представена е разработената от докторантката методика за полуавтоматично създаване на английско-българска терминологична база за конкретна област с цел да се предложи работеща процедура, която да може да бъде използвана от преводачите в България (без амбициите да се решат всички проблемни казуси, свързани с терминологията). Създадената от докторантката терминологична база се състои от извлечените от корпус термини на английски език и техните еквиваленти на български. Добавени са три допълнителни полета в схемата (за синоними, бележки и части на речта), но друг вид информация, напр. дефиниции, изображения, препратки между термините или към външни източници, липсва. Към момента на един термин от изходния език съответства само един термин в целевия. Всички извлечени термини са съществителни, оформени като прости или като съставни думи. В обобщение авторката отбелязва, че като първи опит да се подпомогне разработването на специализиран софтуер за извличане на термини и създаването на двуезикови терминологични бази, разработената методика не предлага универсално решение за подобряване и улесняване на работата на преводачите с цел уместното използване на термини. Като най-съществени ограничения на методиката са отбелязани ръчният подбор на корпуси и ръчното търсене на преводни еквиваленти.

4. Сравнение на възможностите за оценка на качеството на превода и приложението им при оценката на създадената терминологична база

С тази задача са свързани резултатите, включени в последната глава на текста. Те представят възможностите за измерването на точността на превода със системите за компютърноподпомогнат превод, тъй като тези програми имат вградени функционалности за проверка на качеството. Те проверяват използваните термини,

дължината и сегментирането на текста, както и съответствието на цифрите, препинателните знаци, интервалите, главните букви, специфичните символи от изходния език в целевия сегмент. Направени са два експеримента, за да се провери работата на тези програми върху български данни. В изходните текстове са въведени грешки, свързани с терминологията, които се срещат в реална работна среда. Експериментът показва, че терминологичната база може успешно да бъде използвана за проверка на качеството на превода.

Направено е сравнение и на ефективността на програми за оценка (проверка) на превода. В. Петрова прави класификация на грешките, като ги степенува според техното влияние върху качеството на превода. Резултатите от използването на различни програми за проверка на качеството в текстове с изкуствено въведени грешки са сходни, независимо от начина на създаване на терминологичната база. По повод бъдещето на преводаческата работа В. Петрова отбелязва, че програмите за автоматична проверка на качеството (вградени в системите за компютърнопомогнат превод или самостоятелни) са полезни, защото (до известна степен) улесняват работата на преводачите, но не могат да заместят хората при проверката.

Заклчението на дисертацията съдържа обобщения, които показват, че основната цел е изпълнена успешно. На основата на анализа и сравнението на технологичните средства, използвани при превод (програмите за компютърнопомогнат превод, техните компоненти: преводна памет, терминологична база, машинен превод, инструментите за автоматично извличане на термини и за проверка на качеството на превода), е предложена методика за полуавтоматично създаване на терминологични ресурси за конкретната специализирана област.

- **Представяне на състоянието на проблема и научната литература по темата**

В работата си Виктория Петрова цитира 134 източника на научна и практикоприложна информация. В списъка с библиографията са включени основно заглавия на английски език. Прегледът на използваната научна литература, както и на уводните части на четирите глави на дисертацията показва, че докторантката познава общотеоретичните модели, които са свързани с организацията и работата на системите за компютърнопомогнат превод и приложението им върху текстове с различна стилистична и жанрова характеристика. В прегледа на литературата за същността и дефинирането на термините като специфичен пласт от лексикалната система, на терминологичните речници, терминологичните бази и на преводната памет личи познаване на спецификите на всеки от коментираните обекти. Авторката е запозната със структурата и спецификите на съществуващите към момента терминологични ресурси за българския език.

Виктория Петрова познава актуалните модели и методи за създаване и управление на терминологични бази в системите за компютърнопомогнат превод, както и средствата за автоматично извличане на определени лексикални единици (в случая елементи от компютърната терминология) от двуезични паралелни корпуси. С оглед на конкретните задачи в дисертацията е направен задълбочен анализ на инструментите за измерване на точността на превода.

- **Коментар на избраната методика**

Методиката, приложена при характеристиката на терминологичните речници, терминологичните бази данни и на преводна памет, както и конкретните методи при създаване на терминологична база данни в областта на компютърната терминология и при оценка на точността на превода са адекватни на поставените научни задачи. Прилагането им води до обективно и непротиворечиво представяне на спецификите на всеки от обектите с оглед особеностите на изходния и на целевия език и на използваните инструменти при превод. Виктория Петрова използва комплексна методика, включваща традиционните общонаучни методи (описание, анализ, сравнение, класификация) и емпирични качествени и количествени методи (експеримент, статистически анализ, компютърна обработка и др.). Методиката за съставянето на паралелния двуезичен корпус с термини, за полуавтоматично създаване на терминологична база от този корпус и за разработване на двуезична терминологична база следват установените в практиката критерии за изграждане на такъв тип езикови ресурси. Създадените за целите на работата бази данни са достатъчни и позволяват изпълнението на конкретните задачи, формулирани за всяка глава от текста.

- **Кратка характеристика на научните и/или научно-приложните приноси**

Представеният труд показва безспорните постижения на Виктория Петрова при решаването на научно-приложен проблем чрез методи, съответстващи на съвременните постижения на науката и практиката. Двете групи приноси, посочени в дисертацията, са лично дело на докторантката и са постигнати в резултат на нейната самостоятелна работа по реализирането на научните цели. Теоретичен принос е заявен в анализа на възможностите на системите за компютърноподпомогнат превод и на системите за проверка на качеството на превода, апробирана върху данни от конкретна тематична област. Значим научно-приложен принос е представеното оригинално решение на актуален проблем, свързан с механизмите за съставяне на паралелен двуезичен корпус с термини, за полуавтоматично създаване на терминологична база от този корпус и за разработване на двуезична терминологична база. Принос с практическа насоченост е разработването на методология за полуавтоматично създаване на терминологични ресурси за български език за системите за компютърноподпомогнат превод. Данните от проведената анкета сред български преводачи за установяване на нагласите им към новите технологии, които използват в работата си, свързани с подготовката на преводачи, актуализирането на курсове в бакалавърски и магистърски програми, ориентирани към превода и т.н.

Авторефератът на дисертацията изчерпателно и точно представя структурата на труда, резултатите от проучванията, практическите и експерименталните дейности, както и научните приноси на дисертационния труд.

- **Публикации по темата на дисертационния труд**

Представените публикации отговарят на минималните изисквания на Института за български език за придобиване на научна степен. Виктория Петрова има три статии по темата на дисертацията. Всички са отпечатани в сборници с доклади от

международни научни прояви, като една от публикациите е в сборник с материали от международна конференция по компютърна лингвистика. Два от текстовете са на английски език. Важно е да се отбележи, че първата от статиите, включени в списъка с публикации по темата на дисертацията, вече е цитирана в проучване на чуждестранен учен.

Заклучение

Оценявайки актуалността на темата на дисертационния труд „Полуавтоматично разработване на многоезикови терминологични ресурси“, постигнатите резултати и направените изводи, предлагам на уважаваното Научно жури да присъди на Виктория Петрова-Любенова образователната и научна степен „доктор“ в областта на висшето образование 2. Хуманитарни науки, професионално направление 2.1. Филология, научна специалност Общо и сравнително езикознание.

12.12.2022 г.

проф. д-р Йовка Тишева