

РЕЦЕНЗИЯ

на дисертационен труд за придобиване на образователната и научната степен доктор

Автор на дисертационния труд: Виктория Петрова-Любенова, докторант към Секцията по компютърна лингвистика, Институт за български език „Проф. Любомир Андрейчин“, БАН

Тема на дисертационния труд: Полуавтоматично разработване на многоезикови терминологични ресурси

Рецензент: доц. д-р Екатерина Търпоманова, СУ „Св. Климент Охридски“

Дисертационният труд „Полуавтоматично разработване на многоезикови терминологични ресурси“ на Виктория Петрова-Любенова разглежда важен и актуален въпрос от областта на компютърната лингвистика, свързан с обширната тема за създаването и обработването на езикови ресурси. Редица автори поставят въпроса за приложението на машинния превод в ежедневната работа на преводачите, като мненията клонят към развитието на компютърноподпомогнатия превод, а не директното прилагане на машинен превод, което често се оказва без директна полза. Именно въпросът за развитието на ресурси, приложими в компютърноподпомогнатия превод, е във фокуса на дисертацията.

Дисертантката е добре запозната със състоянието на проблема, който изследва, както и с научната литература по въпроса. В библиографията и съответно отразени в текста на дисертацията присъстват важни изследвания – както по-общи и ключови трудове в областта на компютърната лингвистика, компютърното обработване на естествени езици, историята на машинния превод, компютърноподпомогнатия превод, така и по-конкретни и специфични работи, които имат отношение към темата на дисертацията – преводна памет, създаване на терминологични ресурси, извличане на термини и т.н. Същевременно трябва да отбележим, че тази тема не е разработена в

българската наука и дисертантката може да се опре само на няколко по-общи лексикологични и терминологични изследвания, както и на няколко статии за извличане на термини. В този смисъл може да се каже, че дисертацията на Виктория Петрова-Любенова е новаторска за българския език.

Методиката на изследване, приложена в дисертационния труд, е подходяща за постигането на целите и задачите. Дисертантката поставя като основна цел на изследването си да предложи методика за създаване на терминоголични ресурси за дадена специализирана област, в настоящия случай – в областта на компютърната терминология. Използваната методика може да се определи като комплексна. От една страна в теоретичната част се прилагат методите на описателната лингвистика, където са представени научната литература по въпроса, прегледът на съществуващите терминологични бази и инструментите за компютърноподпомогнат превод. В приложната част на дисертацията – съставянето и оценката на двуезиковата терминологична база данни, са използвани методи от компютърната лингвистика, статистически и корпусни методи. Смятам, че методологията е добре изградена и уместно приложена, което допринася за постигането на целите и задачите на дисертацията.

Трудът на Виктория Петрова-Любенова следва утвърдената структура за дисертация. Състои се от увод, четири съдържателни глави, заключение (което по някаква причина не е отбелязано в съдържанието), приноси, три приложения, списък на някои термини и библиография.

Уводът е изграден стандартно – представя актуалността на изследването, мотивацията, която стои зад него, обекта и предмета на изследването, целите и задачите, използваните методи и композицията на работата. Всички уводни компоненти на написани стегнато, лаконично и по същество. Основната цел на изследването е подкрепена от по-конкретни цели за всяка от главите.

В първа глава се разглеждат системите за компютърноподпомогнат превод и техните компоненти. Компютърноподпомогнатият превод представлява специализиран софтуер, който улеснява превода от един език на друг. За разлика от машинния превод, при компютърноподпомогнатия преводачът управлява целия процес чрез различни функции. В тази част са дадени няколко дефиниции за компютърноподпомогнат превод, както и илюстрации на интерфейса на различни програми. Основните компоненти на компютърноподпомогнатия превод са преводната памет и терминологичната база, като някои системи интегрират във функционалностите си и машинния превод, поради което

в дисертацията той също е разгледан като компонент. Преводната памет е описана като важна част от компютърноподпомогнатия превод. Тя представлява база от данни, в която се съхраняват вече направени преводи. Дисертантката много добре описва структурата и функционирането на преводната памет – сегмента като основна единица, изчисляването на подобие, необходимия софтуер, който се прилага. Машинният превод съвсем уместно се разглежда в първата глава като възможен компонент от компютърноподпомогнатия превод. Накратко са описани двата най-разпространени модела на машинен превод – статистическия и невронния. Аналитично са представени възможностите за интеграция на машинен и компютърноподпомогнат превод, като дисертантката демонстрира познания както в развитието на този въпрос, така и в практическата реализация и резултатите от нея в някои съвременни системи. Терминологичните бази са третият представен компонент, като в тази част е показана същността им, ползите от приложението им и механизмът, по който функционират.

Първата глава завършва със собствено проучване на отношението на преводачите към технологиите за компютърноподпомогнат превод. Виктория Петрова-Любенова е съставила анкета от 26 въпроса и я е провела сред 73 професионални преводачи. Наблюденията ѝ за нагласите на преводачите са една добра мотивация за нейното изследване както в теоретичната, така и в приложната му част, тъй като показват интерес към технологиите сред доста добър процент от анкетиранияте.

Втората глава се спира конкретно върху терминологичните бази като част от системите за компютърноподпомогнат превод. Главата започва с дефиниции и описания на понятието термин, като се прави разграничение между терминология в широк и тесен смисъл. По-нататък следва описание на терминологичен речник и терминологична база и важните разграничения между двата типа ресурси. Представени са спецификите на терминологични речници от гледна точка на тематичния обхват, както и информацията, която съдържат. Дисертантката правилно изтъква ролята на терминологичните бази в системите за компютърноподпомогнат превод, тъй като те могат да бъдат управлявани от преводача, компанията или институцията, която ги ползва. Описани са няколко от терминологичните бази със свободен достъп. Следва аналитична съпоставка между терминологичните речници и терминологичните бази, като са посочени предимствата и на двата типа ресурси от гледна точка на надеждността, достъпността, обхвата и т.н. Авторката не е пропуснала да отбележи и някои недостатъци на всеки тип ресурси, например неясното авторство при някои интернет речници, от което произтича и липсата на информация за тяхната надеждност. Разгледана е и връзката между терминологичните

бази и преводната памет. Обърнато е специално внимание на създаването и управлението на терминологични бази в системите за компютърноподпомогнат превод с конкретни примери от няколко системи, както и възможностите за интегриране на допълнителни програми към основните, или притурки. Други важни аспекти са системите за управление на терминология, които могат да се използват самостоятелно или да бъдат част от система за компютърноподпомогнат превод, а също и т. нар. контролирани езици, които са особено подходящи за четимостта на специализирани и технически текстове. Главата завършва с подчертаване на важноста на дейностите, свързани с терминологията.

Третата глава се занимава с полуавтоматичното създаване на многоезикови терминологични бази, което представлява същинската част от работата. Главата започва с по-общи въпроси – въведение в темата за компютърната терминология, върху която се фокусира терминологичната база, представяне на съществуващите речници, методите за автоматично извличане на термини, както и инструментите за това. Следва представянето на разработената от дисертантката методология за полуавтоматично създаване на терминологична база в областта на компютърната терминология за двойка езици – български и английски. Като ресурс за извличането на термини е направен най-добрият избор, а именно двуезиковият корпус. Като най-подходящ корпус за целта авторката се спира на английско-българския корпус с компютърна терминология QTLear. Проведени са експерименти за избор на инструмент за извличане на термините, който да отговаря на предварително зададени критерии, като в крайна сметка решението натежава в полза на комбинация от програми. След съставянето на терминологичната база е направена ръчна проверка. Новосъздадената база съдържа термини на английски език и техните български еквиваленти, както и няколко полета за допълнителна информация. От представения метод за съставяне на терминологичната база имам възражение само към изключването на синонимите. Това означава, че при използването на ресурса за обратен превод (от български на английски) синонимната дума няма да бъде разпозната като термин. В езиковия анализ на термините има някои неточности (напр. възникването на термина *плейлист* не е чрез калкиране).

Четвъртата глава е посветена на измерването на качеството в системите за компютърноподпомогнат превод и техните компоненти. Много правилно се започва с уговорката, че качеството на превода е относително и субективно понятие, поради което измерването му е трудно. Експериментите за измерване на качеството са отлично проведени и описани от дисертантката. По този начин се вижда ползата от приложението

на терминологичната база в реални ситуации. Програмите, с които се измерва качеството, показват различни типове грешки и идеята за класификация и степенуване на грешките, за да бъдат открити последствията от тях при превод, е много добра. В края на четвъртата глава са представени международните стандарти за качество на превода, които са приети, за да се унифицира преводаческата дейност и да се гарантира качеството на превода.

Заклучението на дисертацията е твърде разказвателно и представлява по-скоро сбит преразказ на текста, отколкото да представя резултатите от изследването. Изводи по същество са направени спорадично, например в третия абзац се констатира, че терминологичните бази са важни, тъй като спестяват търсенето на термини, но най-ценният компонент на компютърноподпомогнатия превод е преводната памет.

Трите приложения след текста представят важни за изследването данни в удобен вид и са много полезни. Всъщност и в самия текст на дисертацията има таблици и фигури, които визуализират различни данни и статистики и улесняват възприемането на информацията.

Цялостното ми впечатление от дисертацията е много добро, личи задълбоченият поглед на Виктория Петрова-Любенова по темата, както и опитът ѝ като преводач. Текстът е добре структуриран, съдържателното развитие следва модела от общотеоретичен подход към специфично и конкретно описание на информацията. Безспорна е компетентността на дисертантката в методите на компютърноподпомогнатия превод, добрата ѝ ориентация в системите, които се използват и в общата теория на превода и терминологията. Постигнат е добър баланс между теоретичната и приложната част на дисертацията. В окончателния вариант на работата са останали някои печатни грешки, които е добре да бъдат отстранени, особено ако се вземе решение за публикация.

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 компютърноподпомогнат 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.

5 декември 2022 г.

доц. д-р Екатерина Търпоманова